The BRECVEMA Model of Musical Emotions and Theories of Embodied Cognition:

A Practice-Based Approach including an Experimental Music Theatre Composition Portfolio

Oli Chun-Yang Jan

November 2022

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy in Music

School of Culture & Creative Arts

College of Arts

University of Glasgow

Abstract

The BRECVEMA Model of Musical Emotions and Theories of Embodied Cognition:
A Practice-Based Approach including an Experimental Music Theatre Composition Portfolio

by
Oli Chun-Yang Jan

This thesis follows the framework of the BRECVEMA model proposed by Juslin (2008, 2010, 2019) in investigating the emotive effects and embodied aspects of music perception. It is structured according to the categories proposed in the model. Thus, each chapter or section will discuss a mechanism in the framework.

I attempt to do develop the discussion in three approaches: the theoretical, the analytical, and the practical. The theoretical approach will review relevant literature in fields of study such as aesthetics, phenomenology, linguistics, cognitive science, neuroscience, and biology. The analytical approach will look at existing EMT (experimental music theatre) pieces and explore how the mechanisms work in these pieces. The practical approach will, at the end of each chapter or section on the mechanisms, discuss my own composition of EMT pieces that are composed on the basis of the mechanisms and relevant theories. In doing so, I seek to present a practice-based research methodology that can have contribution to the topic by demonstrating a mutually beneficial relation between theory and practice, where the former provides bases for the latter, and the latter produce examples of extended interpretation, concrete implementation, and complex interaction for the former.

Acknowledgements

I am deeply thankful for my family, whose unconditional love and support has provided the foundation for my academic pursuit. I remember one afternoon my parents were listening to Beethoven, when they suddenly got curious and asked me why people sway to the beat of the music. This happened to be one of the topics from this thesis which I was writing at that moment, and the serendipity has encouraged me to carry on.

My sincere thanks also go to my colleagues and friends from University of Glasgow, University of Edinburgh, and Royal Musical Association, who provided constant companionship and inspiration for my research. I wish to thank Marielle Brie de Lagerac too, for generously helping me with questions regarding French text. Moreover, my gratitude is extended to Sound Festival in Aberdeen, Edinburgh Fringe Festival, Tongyeong International Music Festival, and all the colleagues that have organised them. It has been an honour and pleasure to be a part of these events, and the experiences have led to significant developments of this thesis.

I need to express my appreciation to the examiners for my progress reviews and viva voce, Nick Fells, Jane Stanley, Iain Findlay-Walsh, Mariusz Kozak, and the convener for my viva voce, John Williamson. They kindly offered me the reassurance and feedback which were crucial for me to continue studying the topic. In addition, I benefited greatly from the discussions with Edward Campbell, Martin Iddon, Raymond Macdonald, Nikki Moran, and Jonathan De Souza, who are members from the research network of Somatic Music, which has been a valuable learning experience for me. My deepest gratitude also goes to my teachers from National Taiwan University, Tien-Yi Chao, Keng-Chen Liang, Pao-Hsiang Wang, and Shan-Shan Wang, and my supervisor from University of Edinburgh, Yati Durant, who have enlightened me in my early stage of research and continued to support me.

I would like to acknowledge the support of LKAS Interdisciplinary PhD Scholarship from University of Glasgow, which funded the study. Throughout the process of my studentship, I am deeply indebted to my co-supervisors, Frank Pollick and Michael Bachmann, who have always been generous to share their expertise in cognitive science and theatre study. Their guidance has been pivotal for this interdisciplinary research.

But above all, I am immensely grateful to my supervisor, Björn Heile. As a distinguished academic and compassionate person, he has supported me in every aspect throughout this journey. His open mind, knowledge, and professionalism has continued to inspire, guide and motivate me. He has truly been a mentor I am privileged to have and wish to emulate.

Table of Contents

1. Introduction

> And yet music is effectively nothing more than a series of audio sensation. […] There is nothing to do here except for listening, without returning to ourselves, our memories, our feelings, without thinking about the composer, as the perception looks at the things themselves without blending our dreams therein.
>
> - Maurice Merleau-Ponty, *Causeries*[1]

In this thesis, I will attempt to combine theoretical, analytical, and practical approaches in demonstrating that music perception is, contrary to the quote above, more than merely listening to the audio. The synthesis of three approaches will provide a methodology model where the concepts of embodied music cognition and practices of music composition engage in meaningful interaction where the former informs the latter, while the latter also enriches the former.

In his discussion on arts and the perceived world, Merleau-Ponty (1948) has argued that it is impossible to separate the things in the world from one's cognitive experience thereof, because it is from the perceptual details that the meanings of things arise. He believes that the same should be said for artworks. Only when one accepts to be guided by perception is one ready to comprehend the work of art. A painting, for instance, is not to be understood as an attempt to imitate the world, but

---

[1] The original in French:

> Et cependant il s'en faut bien qu'une musique ne soit qu'un agrégat de sensations sonores […] Rien à faire ici que d'écouter, sans retour sur nous-même, nos souvenirs, nos sentiments, sans mention de l'homme qui a créé cela, comme la perception regarde les choses mêmes sans y mêler nos rêves. (2017, p.49)

An English translation by Davis can be found in *The World of Perception*, published in 2004. However, the translation by Davis is on the freer side and I believe a more literal translation would better associate with the issues to be discussed here, so I made the translation here myself.

as a world itself (2017, p.47).

Music, considered primarily as a non-representational art by Merleau-Ponty, is only briefly discussed because it offers too straightforward an example by this definition. Acknowledging program music as an exception, he has described music primarily as an artform that consists of series of sound stimuli, which does not refer to anything other than itself, and, as quoted above, is only to be perceived by listening to it.

In this thesis, I will show that music perception is heavily embodied, and therefore involves many more aspects than merely the reception of sounds. This will be approached from several perspectives, not limited to Merleau-Ponty's phenomenology[2]. I chose to open the introduction by quoting the above text because, interestingly, almost every aspect Merleau-Ponty has excluded is actually involved in the process. A listener's memory, senses, background knowledge, and subconscious activities, be it psychological or physiological, all contribute to the perception of music, and it is through these various paths that listeners often have an emotive response to the music perceived.

Structurally, this thesis will discuss several routes through which music generates an emotive response. Each of these mechanisms is to be covered in a corresponding section. In each section, I will firstly review the mechanism involved in terms of philosophy, cognitive psychology, neurophysiology, and other relevant fields. Following the theorical discussion, I will analyse pieces of experimental music theatre where the mechanism is involved to demonstrate how they function. Lastly, I will refer to pieces from my own composition portfolio, consisting mostly of experimental music theatre pieces, where I apply these mechanisms. The goal is to present an embodied approach

[2] which actually also places great emphasis on embodiment, but the embodiment in this study is closer to what Ward & Stapleton describe as 4E's (2012). This will be elaborated in the next chapter.

of music perception and emotive response, which will provide valuable insights that are pertinent to the theory, analysis and composition output involved.

While such responses in this study tend to be discussed in readily defined and observable aspects, such as particular self-reports or biological measurements from the listeners, the general nature of emotionality itself in music has been an issue of debate. One of the controversies is whether "true" emotion occurs in music perception. Kivy (1989), for instance, has dismisses the notion that music can induce actual emotions, due to the belief that emotions emerge on the basis of cognitive appraisals. With no real agent/object/event involved in the process of listening to music, the target of such appraisals is absent, therefore what is evoked is different from actual emotions. Rather, Kivy attributes the listeners' responses simply to other factors such as musical preference. An earlier scholar, Meyer (1956) has attempted to describe listeners' experience from the aspects of their expectancy to music, but also simply considered such experiences as the reactions that occur when the expectations are violated or fulfilled. More recently, some have argued that in daily listening activities, music is very often only a "sonic wallpaper" and therefore emotionally inconsequential (Konečni, 2008), although when it does achieve aesthetic response such as awe, it can have effects on mental and even physical health (Monroy & Keltner, 2023).

To some extent, I will argue for the emotionality in music from two perspectives. Firstly, in terms of observable reactions in listeners, plenty of evidence indicates that behavioural as well as biological patterns typically interpreted as emotive are present when an individual listen to music, as the discussion will show in the following chapters. Psychological models describing the mechanisms have been proposed by early scholars such as James (1884) and Lange (1887), as well as later ones such as Lazarus (1991). They have pointed out that cognitive processes are activated to attribute these patterns to a source, and what has emerged as a result is defined as emotion. In a sense, as long as the process begins, appraisal will search for targets, regardless of how proximate they are to the "real" purpose and consequence of initial reaction. For instance, the increased heart rate has biological functions directly related to fight or flight responses. While the fear when seeing a bear in the wild could indeed get the individual ready for such responses (although which, as we now know, may not be the safest thing to do), the anxiety while watching a sport game hardly contribute to anything functional in the reality of the event. However, it would not seem entirely correct to say that the emotion in the game is "less true" than that in the wild when seeing a bear. Similarly, emotive responses that occur while listening to music, although not immediately related to actual consequences in reality, are still indicated by observable processes that are present. The content of the subsequent emotion may be derived from appraisal processes related to listeners' own embodied experience, or from contagion effects as a result of certain properties in the music. These mechanisms will be discussed in detail in the following chapters.

Secondly, while music may indeed serve as background in certain fashions of music listening nowadays, in many cultures or activities, music had and still has participatory nature. Thus, such music is often associated with a collection of listeners' embodied experience. Later, the same music can also still evoke emotive response, even when they are perceived as "sonic wallpapers", as Konečni would put it.

The above two points involve respectively the internal biological responses and external embodied experience of the listeners, both of which will be discussed in detail in the following chapters. They will indicate that at least certain reactions that are typically reported or regarded as emotive indeed often emerge due to exposure to music, whatever the form of the listening is. Beyond these aspects, I will avoid discussing the ontological nature of emotion, or whether a specific emotive response is "true" compared with others, as it would not be within the scope of this study to do so. For instance, Kivy has argued that "music provides neither the objects nor, therefore, the belief-opportunities that would make it possible for musical works to arouse such emotions as anger, sadness, joy" (1990, p.165). The argument is probably established on a basis which already excludes music or musical experiences with factors that could eventually result in some associations or consequences in particular contextual settings, such as music in a ritual or live concert. It may also exclude emotions that are even slightly vicarious in nature. It is obviously not to say that such distinctions are pointless. On the contrary, they may have important implications in many other aspects, but they simply will not be the focus of this study. Rather, this study will predominantly discuss observable details and significances of the processes whereby listeners perceive and react to music.

The term "emotive response", therefore, will be often used interchangeably with "emotion", as defined by Juslin & Sloboda in their introduction to the 2011 *Handbook of Music and Emotion: Theory, research, applications*. In the terminological guideline provided for the contributors of the book, they define emotion as "a quite brief but intense affective reaction that usually involves a number of sub-components—subjective feeling, physiological arousal, expression, action tendency, and regulation—that are more or less 'synchronized'" (Juslin & Sloboda, 2011, p.10). This definition highlights the reactive nature of emotion, and is therefore appropriate for the discussion in this study.

Unless otherwise indicated, this study will follow the same source for definitions of other terms, such as affect, mood, feeling, and so on. Also, as Juslin and Sloboda have acknowledged, different authors may have their own preference for terminology and definition. It is natural that all sources quoted and discussed in this study may not be using the terms in fashions consistent with the working definitions quoted above.

Whenever this is the case, additional clarifications will be included.

Several models exist to illustrate the process and structure of listeners' emotive response to music. A detailed review of these models will be engaged in section 2.2, where I discuss the BRECVEMA model proposed by Juslin (2008, 2019), the Multifactorial Process Approach by Scherer & Coutinho (2013), and the Constructivistly-Organised Dimensional-Appraisal model by Lennie & Eerola (2022). These frameworks all have their respective advantages in different aspects, but structurally, this thesis will primarily follow the BRECVEMA model. The reason I have decided to apply this specific framework as the basis of this thesis is simply a practical one, since it covers better what I try to explore: firstly, its categories, such as contagion, rhythm, volume, harmony, expectancy, memory, allows the discussion to be organised into that of musical components, such as harmony and rhythm, and that of performers, listeners, and context. This allows me to conveniently arrange the discussion, especially, when I describe a certain methodology in my composition that is derived from processes related to a specific mechanism. On the other hand, the other models propose either routes that simultaneously cover these aspects (Scherer & Coutinho, 2013) or structures that focus on the appraisal of the listener (Lennie & Eerola, 2022). It should be pointed out, however, that this is not to say that the particular chosen framework is more valid than others, but merely a decision to unfold the discussion in a structure that is practically most fitting to this study, which involves analysis and portfolio of music pieces, and would benefit from a framework that has separate sections for features such as rhythms and harmonies to engage in the discussion. Developing examples of analyses and compositions corresponding to respective components in the framework, I can have meaningful methodological results in terms of analysis and composition, even if the framework itself faces challenges or undergoes modifications in the future. Secondly, the study underlying the BRECVEMA model has more readily available support in neurophysiological studies, while more recent ones such as the CODA model by Lennie and Eerola either has only a preliminary blueprint in terms of a biological basis (Brosch & Sander, 2013) or empirical study results that till need further refinement to fully support the model (Grandjean & Scherer, 2008). Likewise, this is not a statement that a framework with less available evidence on biological basis at current stage is less valid, but merely a strategy to involve a context with relatively more pre-established bases where discussions on relevant empirical observations can be enabled. Detailed discussions on these models will be found in 2.2.2.

This thesis is organised into 5 chapters. Chapter 1 introduces the aim of the study and structure of the thesis. Chapter 2 reviews past studies on embodied cognition and music (2.1), emotive effects of music (2.2) and experimental music theatre (2.3). The remaining chapters will cover the components in Juslin's BRECVEMA model. The seven major mechanisms in the model are emotional contagion, visual imagery, rhythmic entrainment, brain stem reflex, evaluative conditioning, episodic memory and musical expectancy, through which music evokes emotive response in listeners.[3] In this thesis, chapter 3 will discuss emotional contagion; chapter 4 will explore visual imagery, rhythmic entrainment and brain stem reflex; and chapter 5 will mention evaluative conditioning, episodic memory and musical expectancy, before the concluding remarks of this thesis. In discussing each mechanism, I will firstly review relevant theoretical and empirical studies, and subsequently analyse pieces of experimental music theatre or another practice where the mechanism can be observed, and finally describe the parts in the pieces from my composition portfolio that apply the mechanism in question.

[3] The most recent addition of "aesthetic attitude" will also be briefly discussed in the final chapter.

These compositions are mostly experimental music theatre pieces in each of which I seek to explore the mechanism in question, developing composition methodologies based on the relevant theories. They are, while still conceived as artistically integral works by myself, created with the specific purpose of investigating and demonstrating the potential methodologies that I try to devise in this study.

As a result, they have certain functional properties which are specific to the written components of this thesis. For instance, I may employ specific timbres, structures, and even instruments/props in a piece in attempt to invoke emotive response through the mechanism of emotion contagion, which is the focus of that specific chapter. The result is a methodological process where the theory can inspire the composition, while the composition interpreting and extending the theory, which is one of the main purposes of this study. However, this does not imply that I deny the possibilities of other processes of musical emotions[4], and the interaction between them and the particular process discussed in each chapter of this study. They simply are not highlighted so the discussion can stay focused. I also attempt to work in a way so that I still making numerous aesthetic decisions in writing the music independent of theoretical exploration, so they also possess the roles of my artistic creation in addition to serving the study.

Combining these approaches, I will attempt to create in this thesis an organic structure where theories of embodied cognition are applied in analytical and compositional output, while also being refined and extended through the process application. The main contribution of this study will be a methodological model that

[4] As defined by Juslin's and Sloboda's guideline, "[t]his term is used only as a short term for 'emotions that were somehow induced by music', without any further implications about the precise nature of these emotions." (p.10)

exhibits the potential of embodied cognition theories in analysing and composing music. The theoretical frameworks discussed in this study will lead to the development of particular interpretational aspects compositional design, as the following chapters will show, but additionally, this study will also serve as an example of the integration of cognitive science theories into musical study, with outcomes that are based on the solid understanding and application of the theoretical components. Such an integrational process can hopefully be transferrable for other researchers and practitioners interested in related fields.

2. Literature Review

This chapter has three sections. Section 2.1 reviews past studies on embodiment cognition and music, which is a core aspect of music perception, as this thesis will argue. I will approach the discussion from the angle of “pitch height” problem, which will be involving theories of embodiment in cognitive science and metaphor in linguistics. My analysis will include more branches of metaphor theories than what current literature takes into consideration, and eventually attempt to show that pitch height connects conceptual, motor, and perceptual mechanisms, and is actually a pertinent phenomenon that exemplifies extended embodiment in music cognition.

Section 2.2 reviews literature on the emotive effect of music, which has been reported as a major purpose of music listening (Juslin & Laukka, 2004; van Gothem & Slobada, 2011; Lonsdale & North, 2011; Thomas et al., 2012) and is largely based on the underlying principle of embodiment cognition. I will outline different basic types of biological measurements used in empirical studies and their applications in the study of emotive response in music cognition. Several theoretical models of the topic will also be discussed, and I will explain why BRECVEMA model better fits the purpose of this thesis.

Section 2.3 discusses experimental music theatre (EMT), a setting which allows the mechanisms related to embodiment and emotion to flexibly play important roles. I will describe the properties of EMT from the angle of conceptual blending[5], which is

[5] Here refers to the term used by McConachie describing the nature of theatricality. The same term is used by some linguists in metaphor theories, which is to be discussed in 2.1.

a notion that serves as the basis for theatre to be interpreted as a mixture of illusion and reality, thus deconstructing the conventional structure where theatrical elements are arranged in a hierarchy. I will also discuss two EMT pieces, *?Corporel* by Globokar and *Tennei-ji* by Udow, and briefly analyse how some of the mechanisms discussed above operate in these settings.

2.1 Music and Embodied Cognition

> The sound like a raging sea was a mountain downpour and above it the sharp rasp of the wind came nearer.
>
> In the depths of the storm there was a roaring.
>
> - Yasunari Kawabata, *The Sound of the Mountain*[6]

In the description of the mysterious and ominous sound of the mountain, which lends itself to the title of the novel, Kawabata seems to place the sound elements in vertical arrangement according to their qualities. There is “the sound like a raging sea”, above which is “the sharp rasp”, and below which, at the bottom, “a roaring”[7].

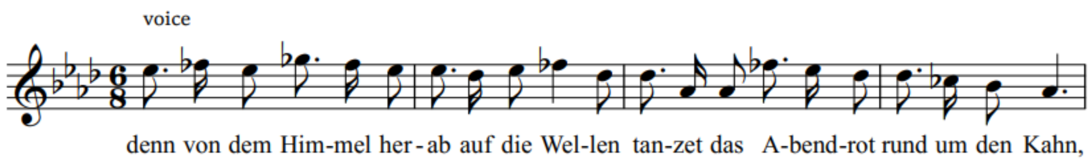
Such an arrangement appears to be intuitive. Sounds with certain properties tend to be inherently described as being in a higher position, while those with opposite

---

[6] The original in Japanese:
海鳴りのようなのは山の嵐の音で、その音の上をまた雨風の尖がこする音が近づいた。
そういう嵐の音の底にごおうっと遠い音が聞こえて来た。

[7] The verticality according to sound quality is not as obvious in the English translation. For the first two elements, the English text would suggest that the vertical position is assigned based on the physical height of the actual events that produces the sounds (“downpour” and “wind”), but in the original, the two sound sources are represented by two words that are less disparate (嵐 *arashi* ‘storm’, and 雨風 *amekaze* ‘rain-wind’) and it would appear that it is indeed their sound qualities (“sound like a raging sea” and “sharp rasp”) that decide their relative height.
The third sound is “in the depth” of the storm in the translation, while in the original it is in the “bottom” (底 *soko*), which again emphasises the verticality more.
The translation quoted here, which is by Edward Seidensticker, however, is literary and convincing. It would definitely be beyond my capability to produce another beautiful and masterful version of English translation that does the original text justice, so I choose to quote Seidensticker anyway, but elaborate on the verticality here in this footnote.

properties lower. In music, this association can sometimes be directly observed in songs with semantic texts.



Example 2.1
'For from the heavens down on the waves / Dances the evening glow round about the boat.'

Example 2.2[8]
'Wind of the night. Wind of the mountains. Wind that falls from Mt. Fuji.'

Example 2.1 is the last two lines from the first stanza of *Auf dem Wasser zu Singen*, an art song by Schubert (1823) set to the text by Friedrich Leopold. Example 2.2 is the introductory phrases of *Kaminari Sendou*, a piece of Tokiwazu-bushi[9]. These are some of the instances which demonstrate obvious correspondences between notes in high pitch and height in the physical world. For instance, in example 2.1, it is at the word

[8] tayuu （太夫） is the singer-storyteller in Johruri (see the next footnote). The rhythm in this transcription is only an approximation. The accurate transcription of the genre remains a challenging task, and a detailed discussion would be beyond the scope of this thesis. To summarise briefly, a few attempts have been made, such as those by Asada Masayuki. Yet as Hoshino (2011) has pointed out, "Japanese music is basically transmitted orally, and the scores related to joruri [johruri], in particular, are difficult to express in written form and had not been published much. ASADA's scores are also based mainly on what he noted at the time he himself was practicing or transcribed from records" ("English Abstract")
[9] Tokiwazu-bushi is a type of Johruri, which is a form of Japanese narrative music accompanied by shamisen.

Himmel (sky) that the melody reaches the highest note of the entire song. Also, in example 2.2, the melody develops from the previous pattern and progresses to a lower note at the word oroshi (fall), which is a part of Fuji-oroshi, a noun which denotes a kind of wind that blows downwards from Mt. Fuji.

2.1.1 Pitch Height

This phenomenon has been described by Cox (2001, 2016) as "musical verticality". It shows that the perception of sound involves more sensory modalities than merely hearing, and is a powerful example of the embodied perception of music.

In proposing what he calls "mimetic hypothesis", Cox suggests that music perception involves mimetic comprehension, either overt (mimetic motor action) or covert (mimetic motor imagery). In the 2016 book, he has stated:

> How and why would listening to or thinking about music, apart from planning or recalling one's own performance, have anything to do with embodiment beyond the operations of the auditory system? The answer offered here is that listening to, recalling, or otherwise thinking about music involves one or more kinds of vicarious performance, or imitation (or simulation), and that the role of this imitation in music is a special case of its general role in human perception. (p.11)

And the principles of the hypothesis can be summarised as:

- Part of how we comprehend the behavior of others is by imitating, covertly (MMI) or overtly (MMA), the observed actions of others.
- Part of how we comprehend music is by imitating, covertly or overtly, the observed sound-producing actions of performers. (p.12)

Such a participatory response connects music perception with other embodied experiences, contributing to the association between a sound with particular properties and a vertical position.

In the subsequent discussion in his book, Cox attempts to explain the phenomenon of pitch height, asking exactly “what is high about high notes” (p.85). He tries to answer this by listing a series of bases on which the verticality could have been established, including “literal height”, such as staff notation, resonance spot, propagation, and “metaphoric height”, which involves the “conceptual metaphor” GREATER IS HIGHER.

Conceptual Metaphor Theory is a theory raised by Lakoff and Johnson (1980), in which they tried to explain figurative use of language by the process of conceptually blending, or “mapping”, properties in two domains. Through the process of mapping qualities from a “source domain” that is used as an analogy, to those in a “target domain” that the speaker tries to comprehend. For instance, in the conceptual metaphor LIFE IS A JOURNEY, some qualities of journey from the source domain are mapped to those of life in the target domain, and thus an analogy is made, which is the basis of phrases such as “he is reaching his life goal” or “he has come a long way”. Similarly, Cox believes the conceptual metaphor GREATER IS HIGHER, which maps height to quantity or magnitude, is the underlying process of many mechanisms of embodied cognition that result in describing high notes as high, such as frequency, loudness, exertion in producing the voice, and so on.

I believe, while Cox’ analysis is insightful in pointing out the effects of mimetic participation and related embodied experience, there are certain limitations resulting from the reliance on the notion of conceptual metaphor. Exploring these issues could also extend the discussion of embodied cognition in music perception.

I have always considered Conceptual Metaphor Theory a useful tool for rhetorical analysis, but not always entirely helpful in accounting for what is conceptually happening. One of the metaphor layers for pitch height, in the analysis of Cox, happens to serve as an example here. Cox has listed a total of 10 bases for pitch height, including literal and metaphoric ones, and points out that one could "think of the concept of pitch height as a conceptual blend, with these ten sources as ten input domains" (p.101). However, it is probable that some of these domains are already a result of conceptual blending themselves. For instance, in the metaphor of "effort", GREATER IS HIGHER is said to be the underlying concept, and because higher notes generally require greater effort to produce, one ends up describing these notes as "high". Yet it is obvious that "great effort" itself already has figurative quality itself, and if one were to apply the analysis of conceptual metaphor, it would be based on something like BIGGER MAGNITUDE OF EFFORT IS BIGGER MASS, before it can become a domain to be mapped to another target. While such analysis certainly could stand in semantic or rhetoric terms, it is rather clumsy if it were to be applied in every case, and, moreover, it is definitely difficult to imagine that the lengthy layers of blending actually underlies what seems to be a usage that is so common that it is already incorporated into the default polysemy of the word.

Indeed, another theory that attempts to explain the mechanism of metaphor would focus more on the linguistic aspects themselves, rather than trying to define the conceptual formulae, and I often find it insightful, especially in certain areas overlooked by conceptual metaphor theory. As Carston has pointed out, Relevance Theory is a theory that:

> has developed in tandem with two important and closely connected ideas in cognitive science: (1) that the mind is modular, and (2) that many mental processes are performed by 'fast and frugal' heuristics. Arguments from evolutionary psychology suggest that the mind is massively modular, in the sense that a great many distinct dedicated procedures and processes have evolved to solve specific cognitive problems (Cosmides & Tooby 1994; Sperber 2002). And it seems that, in solving a wide variety of everyday problems, we employ rather

simple, albeit ecologically rational, heuristics rather than foolproof algorithms or explicit reasoning processes (Gigerenzer et al. 1999). (2010, p.154)

And, according to Carston, the theory would consider metaphors as a case of loose language use, "on a continuum with approximations, category extensions and hyperbole". I believe this perspective could better account for some of the sources of assigning height to pitch, and also offer a more extensive view of embodied cognition of music, which will be repeatedly discussed in the following chapters of this thesis.

In elaborating on the perspective of Relevance Theory here, I would like to firstly point out that, in metaphor study, it is sometimes ignored that the English copular verb, as well as predicative and attributive adjectives, the items that people use so naturally as analytic units of metaphoric language, are actually rather polysemous, and their polysemy could correspond to embodied cognition in the broad sense. To begin with, the copular verb is multi-rooted[10]. Subsequently, it still has multiple meanings, such as identificational, predicational, veridical, all represented by the same word, and sometimes the boundary between these meanings may not be entirely clear. Consider, for example, the utterance "his life is a journey". This could be about a person who constantly relocates throughout their life, in which case the statement is closer to being identificational, and "journey" is used in its literal sense. Alternatively, this may be simply highlighting how this person has gone through various different phases in

[10] The verb is actually a product of the integration of three different origins. The first one, conventionally called b-root, had the form of "beon" in Old English and probably originated from Proto-Indo-European *bheu-. It means "to come into being" or "to become". Among the present conjugational variances of the copula, be, been and being are remnants of b-root. The second origin is s-root, which had the form of "eom" in Old English period. It probably originated from Proto-Indo-European *es-, whose meaning remains rather obscure. Stassen (1997) believes it is has locative meaning, while Shields (1978) holds that it could have pronominal origin. Among the present conjugational variances of the copula, am, are and is are remnants of s-root. The third one is w-root , which was "wesan" in Old English and originated from Proto-Indo-European base *wes-. It means "to reside, to dwell". Among the present conjugational variances of the copula, was and were are remnants of s-root. For a detailed discussion, see Stassen (1997)

their life, in which case the statement is closer to being predicational, and “journey” is used figuratively. However, in reality, there might very well be cases where it is not easy to determine if “journey” is entirely literal or figurative, if the person in question does engage in traveling, to a certain extent and ratio. Thus, conceptually, a better way to describe the process would probably be that the speaker simply follows the heuristics of relating selected information included in “journey”, in any domain, to the events being described, depending on the context.

Consider another utterance, “he is curious”. The ambiguity is in a sense not unlike that of the previous one, in that the statement could have an inclination closer to either identificational or predicational. This person in question could be an overall inquisitive mind in general, or particularly interested in a certain topic. Although no figurative use is involved here, the process of relating a selected scope of “curious” to the person in question is similar.

Extended examples of such a process, namely flexibly selecting relevant information contained by a lexical item and associate it with another, can be observed in many cases, where a phrase could be better explained by “the loose use of language”, rather than a specific process of mapping. In Mandarin, for instance, a sharp knife is described as the adjective denoting “fast” or “contented”[11]. There seems to be no conceptual blending here, and indeed it is hardly considered as a figurative use, but simply a polysemy of the adjective, probably related to the speed of the instrument while being used to cut through objects, or the state of the user that is satisfied by its performance. Another example can be found in a German phrase that means “to be drunk”, which literally translates to “to be blue”. The phrase originated from a particular historical background, as Guryanov and colleagues have described:

[11] 快刀 kuài dāo

In the German culture, there are expressions that contain a blue color and denote a pass of work:

• Blau machen – to work blue, skip a working day;

• Blauer Montag machen - walk away from work, do not go to work.

These idioms arose as impressions of the national history of Germany. Initially, the word "der blaue Montag" was the last day of the carnival and nobody worked at that day. The color designation "blau" was explained by the fact that in the following days the church fasts were decorated with blue cloth. The last day on the eve of fasting in Germany passed in a special fun, and many townspeople were in a state of intoxication. In this regard, German language began to emerge idioms in which blue color was used to mean "drunk". These expressions include the following stable phrases:

• Blau sein - to be drunk;

• Blauveilchen - drunkard.

(2017, p.803)

While the phrase is still commonly regarded as metaphoric due to the presence of the copular verb[12], it would be difficult to explain it as a process of blending of domains[13]. Instead, it derived from a series of relevance of its particular historical context, including the calendrical system, ceremonial objects, and social activity.

Similarly, the question of pitch height, which was raised in the beginning of this section, can also be accounted for with different types of relevance, in some cases in absence of a conceptual blending formula. relevant information associated with height would be enough to result in the tendency to describe a particular sound as "high". It could be from the bodily experience of producing the sound, such as the downward torso movement while producing sighing sounds that descends in pitch, or external

---

[12] For instance, in the 2014 article by de Knop: "In German there is also another metaphorical use of blau e.g., Er ist blau vom Alkohol ('He is blue from the alcohol'), which means that he is drunk." (p.81)
[13] Some scholars may define it as a "dead metaphor", as Pawelec has described: "A 'dead' metaphor is a lexical item with a conventional meaning different from its original meaning (or some previous meaning in the chain of semantic change)" (p.118). Lakoff (1987) has objected to the usage of such a category, but his counterarguments are cases where there are original domain structures, which does not seem to exist from the beginning in the case of "blau sein".

experience related to the environment or event containing the sound, such as creatures in higher position compared to the listener making higher sounds.

Thus, pitch height is a crucial phenomenon which, when investigated through the theories of metaphor, exemplifies aspects such as extended embodied cognition and covert mimetic participation, integrating them as different types of relevance in the process of music cognition. My analysis above has shown that the entire picture can only be explained if one integrates the theory of relevance in addition to the conventionally applied theory of conceptual blending. Essentially, this is because the phenomenon is not only embodied conceptually, but also based on various relevance of bodily and external experiences. This serves as a convincing explanation of the emotive effects of music: when audio elements are perceived, very often the perception invokes and associates other experiences through relevance, which eventually gives certain properties to the audio. For instance, as pointed out above, the movement of sighing involves the torso descending slightly in terms of physical height, and as a result it is natural for the pitch to be associated with the movement and perceived as "falling". But this is not based on a formula of conceptual blending, nor is it purely literal. Rather, it is a holistic process that involve a series of relevant association, and the "emotion"[14] of the agent performing the movement might very well be included in the association. More discussion on extended embodiment and emotive effects of music will follow below and in 2.2.

Similar notions can be found in other theoretical frameworks. Shapiro, for instance, has compiled three themes of the principles of embodied cognition study in

[14] Here "emotion" does not refer to a reaction towards a specific stimulus, but rather a more general state. It would perhaps be more appropriate to call it "feeling" if one were to follow the working definitions provided by Juslin and Sloboda: "[feeling] is used to refer to the subjective experience of emotions or moods" (p.10). See also the definition for "mood", from the same page: "This term is used to denote such affective states that are lower in intensity than emotions, that do not have a clear 'object', and that are much longer lasting than emotions (i.e. several hours to days)"

his works, namely conceptualization, replacement, and constitution. As outlined in the 2019 book, the hypothesis of conceptualization holds that the concepts of the organisms are embodied, and that they “consist in activity in the areas of the brain where perceptual, motor, and emotional processing occurs” (p.117). Unique concepts, as well as other cognitive elements, such as attitudes and emotions, often develop based on such an embodiment. On the other hand, the hypothesis of replacement states that the central mechanism of cognition is embodied, rather than computational. It “see[s] embodiment and situatedness, rather than symbol manipulation, as the core explanatory concepts in their new cognitive science. The emphasis on embodiment is intended to draw attention to the role an organism’s body plays in performing actions that influence how the brain responds to the world while at the same time influencing how the world presents itself to the brain” (p.173). Finally, the hypothesis of constitution proposes an extension of cognition consisting of “constituents of cognitive processes extend[ing] beyond the brain” (p.243) while believing that “perceptual experience depends on knowledge of sensorimotor contingencies, where this knowledge, just as standard cognitive science would expect, is represented in the brain” (p.254).

Similarly, Ward & Stapleton (2012) have proposed an embodied cognition model which they label as “4 E’s”, namely embodiment, embedding, extension and enacting of mind. According to them, cognition is the result of bodily structure (embodiment), interaction of individual and environment (embedding), integration of internal and external information (extension) and, eventually, the emergence of meaning during the above three processes (enacting).

Among the notions discussed above, the hypothesis of conceptualization, namely an individual’s concept of something is determined by the embodied cognition, and the hypothesis of constitution, as well as the integrational processes of embedding and extension, are clearly all related to the phenomenon of pitch height. In example 2.1.2, quoted in the beginning of this section, for instance, "sky" is not merely conceptualised as layers of atmosphere, but rather the entire bodily knowledge regarding it, including the fact that it is situated in a relatively higher position over human. It is then natural that a high note would be suitable to represent "Himmel". The flexibility of relevance and extended embodiment allows wider explanatory and analytical possibilities for not only the pitch height problem, but also other effects and musical excerpts. These theories of extended embodiment also leave the possibility to account for other cultures where pitch is not metaphorised in terms of physical height, in that they may be conceptualised in fashions involving other schemata or elements that are particularly prominent in their environment.[15] Again, this also directs the discussion towards listeners' emotive response towards music, since the information from more external levels of embodiment, such as embedding or extension, very often involves emotion by relevance. Section 2.2 will investigate the emotive effects of music in more detail.

2.1.2 Philosophical Influence and Empirical Support

As Gallagher and Zahavi (2007) have pointed out, the embodied approach of cognitive study can be regarded as a viewpoint situated at the intersection of philosophy and cognitive science. Langer, in her 1957 work that discusses the philosophy of symbolism, has argued that semiotically, music is a symbol that is "unconsummated". In her view, music could be regarded as a signifiant without its

[15] Although a claim would be that since the experience regarding physical height is relatively common across different groups of peoples, if pitch is understood in terms of height, the "higher" pitch is generally conceptualised as "high" rather than "low".

signifié[16], because it consists of audio materials that do not readily have semantics. As a result, it allows individual listeners to "fill in" the content of the symbol on their own. This is one of the early theories of music perception that recognises and elaborates on the participatory process of the listeners.

Also, in the discussion of phenomenology, the perceiving agent itself is usually regarded as an inseparable part of the entire epistemological process. Husserl (1913), for instance, has noticed that a subject is deeply correlated with the external world that they are experiencing, and attempted to explore the mechanism involved. In discussing the concept of number, he has explored the idea of "unity", namely oneness, the basic unit of which multiplicity is composed. According to his analysis, there are two types of unities, namely the factual unity of objects and materials, and the eidetic unity of essence and conceived. In the process of experiencing, consciousness connects the two, and it is through such a connection that one properly perceives and construes the external world, with the information from both factual and eidetic realms.

Husserl has elaborated in great lengths and details on how the consciousness and the other two unities operate in correlation, which will not be discussed here. The framework itself, however, seems to also be applicable in music perception, and the result demonstrates the phenomenological view of audio perception: Consider a piece of music played on the piano, for instance. There are single elements that are defined on the basis of physical facts, such as a note, which is a continuous sound in a particular pitch with its fixed range of fundamental frequency. The knowledge of essence, on the

[16] The terms are used by Swiss linguist Saussure in his discussion of semiotics. According to him, a sign or symbol consists of two main components, the signified (*signifié*), which is the concept being expressed, and the signifier (*signifiant*), which is the sound or image that expresses the concept (Saussure, 1916). For instance, if the word "sheep" is to be considered in the Saussurian framework of semiotics, it would be a symbol where the creature sheep is the signified, whereas the sound /ʃiːp/ or the script "sheep" is the signifier.

other hand, could occupy the listener's consciousness in the form of the concept of that particular piece, which is in its entirety, rather than being the sum of a series of notes connected by "and". The unity of the consciousness, however, will emerge from the experience of receiving the music, depending on the phrasing and rhythm, or even dynamics, timber and other features or contextual information that could result in the grouping of sound perceived[17]. Essentially, these elements correspond to audio stimuli, the listener's background knowledge and perceptual process, and the analysis of how they interact would likely lead to a conclusion similar to that of Husserl's, namely that the consciousness of the perceiving agent and the other unities are not only connected, but also to be considered as elements in a holistic process that is not entirely discrete from each other.

This belief gradually leads to an attempt to systematically study how consciousness and experience act out their parts, a motif developed across different branches of phenomenology. For instance, in pursuing the nature of "unity" that is isolated from the subjects experiencing it, Husserl has proposed the notion of noema and noesis, which mean what is thought, and the act of thinking, respectively[18]. Exactly how these two terms should be interpreted philosophically is not a problem without dispute, but in a general sense, they can be regarded as something similar to sensation and perception in cognitive science[19]. Husserl has noted:

> Perception, for instance, has its noema, and at the base of this its perceptual meaning, that is, *the perceived as such*. (p.185)

---

[17] In the original discussion in Ideas, Husserl more often uses the example of spatial objects and perspectives obtained from different standpoints and orientation. For instance, a building is construed, or adumbrated, from what one experiences from different angles and positions related to the building.

[18] From ancient Greek νόημα and νόησις.

[19] Sensation refers to the mere reception of stimuli, whereas perception involves the subsequent interpretation of information.

and gives an example in his discussion:

> Let us suppose that we are looking with pleasure in a garden at a blossoming apple-tree, at the fresh young green of the lawn, and so forth. The perception and the pleasure that accompanies it is obviously not that which at the same time is perceived and gives pleasure. From the natura standpoint the apple-tree is something that exists in the transcendent reality of space, and the perception as well as the pleasure a psychical state which we enjoy as real human beings. (p.186)[20]

Again, Husserl has extended the discussion to not only perception, but also recollection, judgement, and so on, and there are several proposals regarding what precisely these two particular terms mean, but the notion can indeed be regarded as one of the earlier theoretical frameworks that acknowledge and structurally discuss the distinction between sensation and perception. The example above can easily be applied to the pleasure obtained from listening to a piece of music, where the sensation of audio stimulus (noema) causes the pleasant experience (noetic process).

As Gallagher and Zahavi (2007) have noted, such philosophical studies as those of consciousness structure were first regarded as untenable and unverifiable. It seemed to be detached from any meaningful empirical support. But scholars gradually started to include some evidence to substantiate their argument. For instance, phenomenologist Merleau-Ponty (1945) famously has quoted the Müller-Lyer optical illusion to support his argument that consciousness exists and operates in a holistic way. In the discussion, he has examined in great details the theories proposed by psychologists such as Koffka, who in 1935 finished his *Principles of Gestalt Psychology*.

[20] From § 88. The term perception is widely used in the translation and, naturally, does not reflect the sensation/perception distinction noted above, which is from a convention developed after Husserl.

At the same time, cognitive science developed, sometimes seeming to be a thematically related study with a completely different approach to phenomenology. Yet researchers gradually found that in studying something as inherently subjective as consciousness, the introspective model of phenomenology actually had a lot to contribute to the thought processes of argumentation and experimentation. Thus, the two fields eventually found intersections. Varela and Thompson (1991), for instance, have combined the two and propose the theory of embodied cognition, pointing out that cognitive process heavily depends on bodily experience itself. Thus, the external world does not establish itself independent from one's recognition, and one's recognition in turn cannot establish itself without embodied experience.

To date, a substantial body of research in physiology and neurology exists and supports the embodied cognition of music, and in turn the hypothesis of mimetic participation. Cox has briefly suggested mirror neurons as a biological basis of this theory (2016, p.23). Although he does not provide directly obtained empirical evidence, both motor imagery and subvocalisation have been studied and tested among music performers as well as audiences. For instance, it has been reported that famous pianists widely engage in "mental practice", where they use motor imagery to practice (Gieseking & Leimer, 1972; Meister, et al., 2004). Findings even have shown that motor imagery results in change in neurocircuits that in turn affects a performer's actual movement (Mulder, Zijlstra, Zijlstra, & Hochstenbach, 2004), just like actual practice would achieve. Subvocalisation is even more obvious and universal in performers, a lot of whom actually even explicitly vocalise music passages when performing. For a more experimental setting, it has also been found that vocal performers show very similar brain activation patterns when performing actual singing and mental singing (Halpern & Zatorre, 1999; Kebler 2007). These findings support the view that music is processed in a very similar way when it is made and when it is covertly mimicked.

Even among musically naïve individuals who passively receive music, similar evidence has been found. It has been reported that when asked to observe teachers playing an instrument, either followed by imitation of the playing or not, the participants indeed show strong activation in inferior parietal lobule (Buccino 2004; Gerardin et al., 2000), which is a brain region that contains mirror neurons (Buccino, Binkofski, & Riggio, 2004; Manthey, Schubotz, & von Cramon, 2003; Rizzolatti & Luppino, 2001) and it is inferred that mirror neuron in this area are involved when participants observe musicians playing instruments. Also, Koelsch and colleagues (2006) have reported that the audience's motor region would be activated while listening to music. Interestingly, this only happens when music is regarded as pleasant (which, according to their research's definition, is consonant. For a more detailed discussion on this, see section 3.2). This could be regarded as neurobiological evidence of subvocalisation.

The findings above are from either musicians or non-musical participants that are asked to observe or evaluate music. But it may be inferred that similar activity can likely take place in regular audience members who listen to music as well. This may be a useful presumption for my study's practical output, which is composed in an attempt to engage the audience' mimesis and in turn evoke emotion. Also, evidence other than neurobiological ones of music embodiment can be found in Robb (2015), where he analyses different performances of piano music and points out, for instance, how playing against metric gravity is a result of musical embodiment, and how this in turn creates imagined fluidity of notes in listeners. Another study (Godøy et al., 2005) has investigated air playing of instruments and concludes that one's experience of music includes not only audio but also gestures. For discussions on cases outside western classical music, see Clayton and Leante's 2011 research, where north Indian vocal music and Canadian bagpipes music is studied. It has been concluded that different cultures indeed have elements of musical gestures in common, which work in similar ways, such as serving communicative function during performance.

Finally, it may be worthy to note that Fitch (2015) in his article proposing principles of what is called “bio-musicology” has listed four core components of “musicality”. Here what he investigates is not limited to human activity, but something that can be universally considered musical across different species. His four components are song, drumming, social synchronisation and, last but not least, dance. He points out that music has a multimodal nature (which also explains why he lists drumming as one of the core components) and in some languages even, the distinction between dance and music is not made (Nettl 2000; Morley 2014). He has defined that “the expressive movements often made by instrumentalists as they play […] would also be classified as dancing” (p.8), and dancing is in turn a core element of music. As he has continued to draw examples of this in species other than human, this may serve as an example of embodiment of music in a broader biological sense.

2.2 Emotive Effect of Music

> Jump and strike the tifa of maremi wood, let's sing an Erisam!
> Let's shout about the happiness of that land over there.
> Oh, grandparent, too bad you had to put down your arrow, bow, and machete. [x2]
> Let's sing together, let's throng together, let's praise You quickly. [x2]
>
> \- Hosea Mirino, *Woper Pum Sireb*[21]

As discussed above, music perception is highly embodied, rather than the mere process of audio sensation. Through various mechanisms, music becomes deeply associated with movement and emotion. One of the main foci of this thesis is exploring these mechanisms from physiological, cognitive, and philosophical perspectives, attempting to elaborate on how music perception involves paths in several modalities in addition to audio, thereby causing emotion response. As the following discussion will show, the analysis of emotive effect is itself an integral part for the understanding of cognitive process from the embodied point of view, and also touches on one of the core aspects of music.

[21] The original in bahasa Biak:
Woper pum sireb ai marem wawor erisamoma.
Kobuki byaren darori mariseno sup iwa.
Ara apuso fabye rwir ikoi sumber mamun. [x2]
Kowor kofasiar kosuba mu fasau. [x2]

This text is from a piece of Yospan, a type of dance song featuring a repertoire of diverse, constantly alternating steps that accompanies various community feasts and ceremonies in the island of Biak, Indonesia. (Rutherford & Yampolsky, 1996)
Tifa is an hourglass drum of the region, and Erisam refers to a type of wor song, another traditional song that is "deeply rooted in every corner of Biak life" (p.5)

2.2.1 Theoretical Preliminaries

As Shapiro (2019) has encapsulated, human bodies have states that correspond to the occurrence of emotions:

> Heart rate might increase, adrenaline might surge, hair might stand on end. […] Disappointed people slump whereas the proud stand tall. […] A happy person will smile – the result of contraction of the *zygomaticus major* muscle. When disgusted, the *levator labii superioris* tugs the upper lip into curl.

And he has also pointed out that, from the point of view of an embodiment theorist:

> [T]he exercise of emotion concepts will invoke simulations of the very bodily reactions that one experiences when one is feeling emotion. (p.131)

The view is reminiscent of the perspective that information regarding music is deeply embodied, and that same or related physiological reaction can occur when one hears music and when one makes music. Indeed, the discussion of music perception will very often involve the emotive response evoked during the process, and, as the discussion will demonstrate, just as a listener can have covert mimetic participation, they can also experience vicarious emotion response when hearing music. It is not surprising, as both perception and emotion involve aspects that can be comprehended as a bodily experience.

The exact routes through which music connects with emotion will be analysed in frameworks of empirical studies, which shall support the view that both music perception and emotion work in highly embodied fashions. Meanwhile, there are also philosophical theories regarding the issue that are noteworthy, in that they involve the extended embodied aspects of the context in which a listener is situated.

Langer, for instance, in the previously quoted work (1941), has proposed a theory attempting to explain the deep connection between music and emotion. As previously reviewed, she believes that music is a symbol that is decoupled from its signified meaning, a state that she describes as "unconsummated". They are essentially sound materials that do not have predetermined semantics, and it is up to the listeners to connect these symbols with certain contents, thereby filling in the empty signifier. Langer believes emotional satisfaction is obtained by the listener, exactly in the process of consummating the audio symbols.

While what is meant by "emotion" here could still fit in the definition given by Juslin, it is probably less inherent and rather related to mechanisms such as expectancy, imagery, or even aesthetics (these mechanisms will be discussed subsequently in this thesis). Moreover, one may be tempted to point out that some of these views are better applied to particular types of contexts where music has developed to be an abstract and aesthetic form, which is probably also what Langer is more familiar with. In other types of music, there might be less free room for individual listeners to derive associations between a piece of music and unique signified contents of their own choice. For instance, Senegalese poet Senghor (1966) has pointed out in *Négritude* that black African art has triple nature of being "functional, collective and committed". If this could also be applied to black African music, it is clear that it would not be entirely convincing to describe a music that is "functional" and "collective" to be an "unconsummated symbol". However, Langer's view could still provide an explanation for the emotive response that are generated by music, even in cases where the piece of music in question has pre-assumed functions in that specific context, because these functions themselves can already be considered the signified content related to music as a part of extended embodied knowledge that listeners possess, and these functions can also be related to strongly emotional experiences, such as the example quoted in the beginning of this section.

Another theory that attempts to explain the emotive effect of music, as Leman and Maes (2014) have proposed, involves the anticipation of musical progression. A typical example is illustrated in their discussion on music perception and embodied cognition:

> Consider the perception of articulated tones generated on a piano […] Depending on the tones played, the expectation may be strong or weak. For example, in a particular context of the C major key, the chord sequence D7-D7-Dmin7-G7 (one chord per bar) may generate a strong expectation for Cmaj7 on the next bar. Such chords, perhaps enforced by particular rhythms and turns of a melody played on top of them, allow the prediction in time of newly incoming articulated tones. Obviously, the degree of match between the expected tones and the new tones creates tensions and relaxations that may engage our affective and emotional engagement with music. In this case the sequence D7-D7-Dmin7 culminating to G7 creates tension, while the G7 going to Cmaj7 creates a relaxation. In short, perceptive processes evoke anticipations that lead to the emergence of tensions and relaxations, which ultimately lock into practices that give rise to signification, such as mood regulation, bonding and aesthetic experience (p.82).[22]

Leman and Maes have commented that, "over centuries, theories have been developed that capture the essence of how this kind of anticipation works in practice, using a level of description that draws upon rules and attributed functions" (p.82), and have offered an explanation in functional harmony for the given example. Again, this is an explanation specific for one system of music and has its limits, but the mechanism itself, namely prediction, does indeed plays an important role in music perception and the emotive effect involved. In the BRECVEMA model, it is termed as expectancy. It will be discussed in the following chapters as well.

[22] The term "mood" here is mostly consistent with the working definition applied (see footnote 14)

Just like Leman and Maes have pointed out, the earlier theories are inevitably based on pre-established knowledge of the particular systems involved, rather than results obtained from scientific exploration. However, it is now possible to observe and discuss the biological basis underlying these processes, and such explorations have indeed become an important part of cognitive study.

2.2.2 Empirical Studies

Either as an artistic form or social activity, music has been often recognised as something that can evoke emotion response. Even in a religious or ceremonial context, it can be used to associate with emotional elements of the activity, as pointed out above. Indeed, various studies have now applied an empirical approach and conclude that one of the primary purposes of music listening is related to emotion. Juslin and Laukka, for example, have conducted a study where a questionnaire was given to a group of participants with even distribution of age, gender, and music proficiency[23]. The questionnaire was designed to explore the participants' views on several aspects of music, including "Listening context; Musical expressivity; Musical communication; Emotion perception; Emotion induction; Relationship between perception and induction; and Basic motives for listening to music" (p.226). They have concluded that the findings "suggest that emotional states figure prominently in listeners' primary motives for listening to music[24]." (p.232)

Other studies have shown that the emotive effects of music have function in both

[23] In this study, a participant is defined as musically trained if they are able to play one musical instrument. 51% percent of the participants were classified as musically trained, where as the other 49% were not.

[24] the answers to the free response question "Why do you listen to music?" were recorded as follow: "to express, release, and influence emotions" (47%); "to relax and settle down" (33%); "for enjoyment, fun, and pleasure" (22%); "as company and background sound" (16%); "because it makes me feel good" (13%); "because it's a basic need, I can't live without it" (12%); "because I like/love music" (11%); "to get energized" (9%); "to evoke memories" (4%); and "other" (11%). (p.232)

daily and clinical context. As Gupta (2018) has encapsulated:

> Listeners generally use music as a mood regulator. Emotional regulation has often been reported as one of the major reasons for listening to music (Thoma et al., 2012; van Goethem & Sloboda, 2011). Participants listen to music primarily to manage or regulate their moods, and that the reasons for listening to music may change as people grow older (Lonsdale & North, 2011). Mood improvements have also been observed in stroke patients after intensive music listening or music performance training in motor neurorehabilitation protocols (Bradt, Magee, Dileo, Wheeler & McGilloway, 2010; Rodriguez-Fornells, Rojo, Amengual, Ripollés, Altenmüller & Münte, 2012). Music listening perhaps "activates attention, semantic processing, memory, motor functioning and emotional processing, and enhances emotional and cognitive functioning in healthy subjects and in various clinical patient groups" (Särkämö et al., 2008, p. 866). It has been rightly suggested that music is a tool that can be used to "enact micro-practices" that can influence an individual's regulatory functions, such as emotions (Krueger, 2011, p. 2) (p.256).[25]

While qualitative research based on methods such as questionnaires provides evidence directly from the introspection of listeners, quantitative research based on measurement observes the physiological and neurological activity of listeners, which can provide some insights of music perception as a biological process, and the emotion aspects involved. They also offer further details to the viewpoint that emotion concepts are embodied, claimed by scholars such as Shapiro.

A widely applied measurement is functional magnetic resonance imaging (fMRI), which in recent years is capable of showing detailed activation pattern in different brain regions, not only in scenario of emotive response but also in those of many other psychological activities. Indicators that are less pricy to measure are also used. These include, among others, salivary cortisol, heart rate, skin conductance, skin

[25] The term "mood" in this quotation is mostly consistent with the working definition applied (see footnote 14)

temperature or muscle tension in nape of neck (commonly referred to as "chill"), in zygomaticus major ("smile") or in errector pili muscle ("goosebumps"). For measurement of emotive response caused by music, previous studies have shown that skin conductance[26] seems to be the best indicator. Rickard (2004) invited participants to listen to music that they considered to be "emotionally powerful" and measured their skin conductance (SC), nape muscle tension (chill), heart rate (HR) and skin temperature (TEMP) and found that SC was the most statistically significant index of emotive response caused by music. Chill also had significance, although to a lesser extent, while the HR and TEMP failed to indicate any response. Similarly, Craig (2005) monitored listeners' skin conductance (SC), errector pili muscle tension (goosebumps) and skin temperature (TEMP) while they listened to music which they reported as having produced chills. The result also shows that SC (termed "galvanic skin response" in this study) changes significantly when the participants listen to these music pieces, while goosebumps does so to a lesser degree and TEMP does not.

Some neurological bases of SC may explain why it works well as an emotive response indicator. Studies have located the amygdala as the region activated during occurrence of SC (Craig 2005; Laine, Spitler, Mosher & Gothard, 2009). As generally concluded, the amygdala serves "a crucial role in evaluating the emotional significance of stimuli and in transforming the results of this evaluation into appropriate autonomic responses" (Laine, Spitler, Mosher & Gothard, p.1749). Thus, SC response generated with amygdala activation may indeed be very indicative of a listener's emotive response. Moreover, Laine et al. also points out that it is sometimes possible for SC to occur due to only sympathetic arousal. The arousal of sympathetic nervous system

[26] Also referred to as "galvanic skin response" (GSR) or "electrodermal activity" (EDA). While the latter seems to be the current preferred term in neuroscience, this section will stick with SC, which is the most straightforward term.

(SNS) is believed to be related with preparation for "fight or flight" behaviour[27]. It is worthy to note here that, if some SC measured may indeed indicate emotive response and they indeed occur because of sympathetic arousal, it perhaps can support a theory (Juslin 2011, to be elaborated later in this section) that one of the mechanisms through which music induces emotion is brain stem reflex, which Juslin has defined as reaction to extreme aspects of music such as high amplitude and pitch, the characteristics that are associated with threatening environments and set off survival reactions such as flight and fight.

It is to be noted that, as Lang et al. correctly point out in their 1993 study, SC may indeed be a very effective indicator of emotive response, but it is only explanatory in terms of the dimension of emotive arousal, not that of emotive valence; namely only the presence of the emotion can be confirmed, but it is hard to determine the type of the emotion activated[28]. This is perhaps also the case for many other physiological measurements, where only the intensity of emotion can be indexed by the statistical significance, but what kind of emotion it is remains to be explained by other means such as auxiliary report from the participant or certain other measurements that are correlated with emotive valence, such as zygomatic or corrugator muscle activity (which result in smile or frown respectively).

These indexes obtained from measurements generally support the theories discussed previously. As outlined above, Langer's theory has suggested individual cases of how a piece of music is perceived can be highly diverse. Again, while it is not yet possible to really measure a listener's internal dramatisation of a given piece of

[27] as opposed to parasympathetic nervous system (PNS). Although both are autonomic nervous system, SNS is responsible for reactions to threat while PNS regulates the human body during more relaxed situation dubbed "rest and digest".

[28] Categorisation of emotions itself is a widely discussed topic and several models of emotion types have been proposed. Among them the oldest ones include Wundt's three dimensional model (1897) and Ekman's basic emotion types (1972). Later models are usually based on these but further take into consideration factors such as culture, language, and biology.

music or even a given excerpt of a piece, there are neurobiological findings that support the basic ideas of Langer's theory. Experiments (Salimpoor et al., 2011) have found out that if a listener knows and expects a certain climatic moment of music, different parts in their brain release dopamine indeed exactly according to their "journey" of experiencing the music. To be precise, when they are expecting the climax, their caudate is activated; when the climax arrives, the activation moves to nucleus accumbens.

It should perhaps be clarified that these neurobiological findings are highly specialised, focusing on particular moments in music and measuring listeners' physiological reaction to them. It may not directly monitor and display how audiences appreciate and fantasise a particular piece of music on a more advanced level, but it does provide support that music does induce emotion in structural and semiotic ways, which could serve as effective elements in a composition methodology.

Another aspect discussed above is that of anticipation, which is also supported by empirical evidence. For example, Tsai (2014) investigated listeners' emotive response to pop songs in verse-chorus form. Combining structuralistic analyses of volume, texture, rhythm, melody, tempo and harmony of various examples with measurement of SC response, Tsai has concluded that the formal design of verse-chorus form, or, to be more specific, exactly the verse-chorus transition, is what evokes emotions in listeners. He continued to accommodate the findings in the context of schema shift and reward learning, both are cognitive psychology notions related to the mechanism of expectation.

Related studies can also be found from the works of Sloboda, although some of his sampling, for example that from his 1991 work, may not be very universal in that it focuses on a group of listeners where many of them also play music. When asked to report their experience of being moved by music, many of them quoted western classical music (132 out of 165 nominations were classical pieces, with or without vocal) and they also showed high level of music literacy by being able to indicate specific parts of the pieces, or even “precise musical event” (p.119), such as passages containing features like cadence descending through circle of fifths or melodic appogiaturas. However, such studies can still provide reference as to how a specific group of listeners, namely those who are familiar with the production of music, tend to perceive music. His use of questionnaires also provides qualitative accounts that could be useful to viewed side by side with more quantitative results from neurobiological research.

Finally, several attempts have been made to propose a theoretical framework describing how emotion is induced through music. In the research project titled Appraisal in Music and Emotion[29], Juslin (2010) has combined survey, diary, and experiments, engaging in a large-scale project that approaches the question in both qualitative and quantitative ways, with the target of explaining underlying mechanisms while considering individual differences caused by different factors. The BRECVEMA model, which is derived from their analyses, has proposed a series of psychological mechanisms through which emotive effect is achieved by music. The structure of this study mainly follows the framework of BRECVEMA, discussing the emotion-inducing mechanisms by exploring relevant studies, analysing pieces of composition involving these mechanisms, and applying the mechanisms in my own composition output.

According Juslin’s theory, there are in total eight mechanisms proposed, which are Brain stem reflex, Rhythmic entrainment, Evaluative conditioning, Contagion, Visual imagery, Episodic memory, Musical expectancy, and the later added Aesthetic attitude. There are other models that discuss the process through which music induces

[29] acronymised as AMUSE

emotion, and many of them are constructed as alternatives to the BRECVEMA model. For instance, Scherer and Coutinho (2013) have pointed out that the framework lacks the descriptions regarding both the distinction among different levels of functions:

> The “mechanism status” of several entries in the list is unclear if one applies Marr’s (1982) distinction between three levels of function— the computational, the algorithmic, and the implementational levels. For example, contagion seems to operate on the computational level, evaluative conditioning on the algorithmic level, and brain stem reflexes on the implementational level. Thus, the proposed mechanisms seem to be situated on different levels of function.[30] (p.198)

and interactions among the mechanisms:

> [A]s multiple mechanisms are likely to operate in many cases, a unified framework 198 would require a discussion of potential interaction. The complete absence of multilevel appraisal as a central mechanism underlying several of the phenomena discussed, in particular musical expectancies, is also difficult to understand. (p.198)[31]

---

[30] Scherer and Coutinho quoted this from Marr’s description of how an information-processing device carry out its tasks. Marr believes that these three levels are important in comprehending how such devices operate:
At one extreme, the top level, is the abstract computational theory of the device, in which the performance of the device is characterized as a mapping from one kind of information to another, the abstract properties of this mapping are defined precisely, and its appropriateness and adequacy for the task at hand are demonstrated. In the center is the choice of representation for the input and output and the algorithm to be used to transform one into the other. And at the other extreme are the details of how the algorithm and representation are realized physically-the detailed computer architecture, so to speak. (pp. 24-25)
[31] They continue to acknowledge that Juslin did touch on appraisal in previous works, but it is simply not included in the BRECVEMA model:

And it seems unfortunate that the appraisal of goal conduciveness in the framework of music listening motivation (covered by several chapters in Juslin & Sloboda 2010) is completely neglected. (p. 199)

Now that Juslin is definitely aware of the process of appraisal, it is possible that the decision to not list it as a separate mechanism per se is a statement regarding the multilevel and interactive nature of the mechanisms, in that he considers it to be at a different functional level.

Expanding on the notions of "routes" previously proposed by Scherer and Zentner (2001), Scherer and Coutinho (2013) have proposed a structure of five major routes through which music activates emotion response, which are Appraisal, Memory association, Entrainment, Emotional contagion, and Empathy. These routes, by their definition, are systems, or "psychobiological pathways" (p.209) that recruit various mechanisms in combination. This model allows the possibility to organise the perceptual mechanisms into a layered structure, which is helpful especially when one attempts to develop an argument that involves a series of reactions, such as contagion and empathy[32]. In my study, on the other hand, I choose to structurally follow Juslin's model, because the arrangements where the mechanisms are not recruited into integrated pathways would allow me to more easily organise the discussion, especially those involving some methods I attempt to devise in my composition. For instance, procedures related to sudden change of fast tempo in Scherer's and Coutinho's model can be related to novelty (Appraisal), Entrainment, and subsequent effects of Contagion or Empathy, whereas in Juslin's model there is a single section to accommodate it. As stated in the previous chapter, the choice is a practical one, and has more to do with the possibilities to experiment with single mechanisms than with the validity of either framework as a unified account for the issue. Whether either of the models is more explanatory overall as a unified theory describing the process does not affect the effectiveness of this study, which focuses more on individual processes and its practical possibilities.

Likewise, Lennie & Eerola (2022) also have pointed out that Juslin's model lacks a depiction of how the mechanisms interact. They response by proposing

[32] By their definition, the two routes involves functions of different levels, in that contagion can be activated merely by observing cues such as strong motor expressions, which will invite mimicry, whereas empathy involves the capability to have concern and simulate the motivation and appraisal of the observed (pp. 206-208).

Constructivistly-Organized Dimensional-Appraisal model, acronymised as the CODA model, which describes the process in terms of core affect, multidimensional appraisal space, meta-experience, and their interactions with cognitive and behavioural components. According to them, core affects "refers to the neurophysiological state [...] where changes in core affect drive attention through the process of attributing these changes to a stimulus" (p.9). These then feed into the multidimensional appraisal space, or MAS, which is "a key cognitive-affective component in the construction of relevance and meaning for an individual" and "can be seen as a dynamic and ongoing perceptual interaction between a person and their physical and social environment" (p.9). Eventually, the process results in emotional meta-experience, which is "the conscious experience of a specific emotion[33]" (p.13).

Similarly, I chose to apply the framework of Juslin over the CODA model mainly for structural reasons. A mechanism I would like to incorporate in the composition methodology, for instance harmonic contrast, could be associated with either core affects or multidimensional appraisal space, and component interactions involving both. It is simply more convenient to discuss each element in correspondence with the framework's structure.

Moreover, while being an in-depth explanatory model that could be potentially unifying, at the current state the empirical research related to crucial aspects of the CODA model is still developing. There are works of preliminary research that attempts to explore the biological basis for the basic aspects of appraisals from some dimensions, such as Brosch & Sander (2013) and Kafkas & Montaldi (2014). In the former, for instance, Brosch & Sander have discussed in the study the possible neurological mechanisms underlying dimensions of appraisal such as novelty, concern

[33] from Russell (2003), as referenced in the original.

relevance, goal congruence, agency, and so on, but acknowledge that the work is "a blueprint of the 'appraising brain,' with the aim of motivating a stronger integration of appraisal theory and neuroscience" and that "more hypothesis-driven research about the neuroscience of appraisal is needed" (p.166). For instance, attempting to ascribe a neurological mechanism for the dimension of concern-relevance, they point out:

> [M]ost affective neuroscientists agree that the amygdala is a key region of the emotional brain, optimally positioned to rapidly receive sensory information about the environment and to orchestrate emotional responses. Whereas the important hub role of the amygdala in the emotional brain is well established, the computational profile of the amygdala is still debated (see Cunningham & Brosch, 2012; Sander, 2009, for discussion). The amygdala was typically thought to be a module specialized in the detection of fear-relevant information. However, more and more studies accumulated showing amygdala activation to non-fear-relevant negative stimuli or to positive stimuli, and pointed to an important role of the amygdala in reward learning. To account for the wide range of stimuli that activate the amygdala, we advanced the hypothesis that the amygdala detects concern relevance, the relevance of a stimulus for a person given the current needs, goals, and the values of the person (see Sander, Grafman, & Zalla, 2003, for further discussion; see also Cunningham & Brosch, 2012; Sander, 2009, in press). (p.164)

While being detailed and comprehensive, this review provides more of a mapping between the properties associated with amygdala and those associated with what is considered as concern-relevance, rather than a piece of data that supports a hypothesis with clear description and definition, as the authors acknowledge. If an experiment were to be conducted where the participants' neurological activities are measured while they are asked to perform certain tasks in a context with controlled environment and audio stimuli designed to test their relevance threshold[34], it would

[34] In fact, such studies may not be too far from realisation, as Lennie & Eerola themselves suggest a potential design like this in their article. The part where this is included will be quoted in the next chapter.

be a more substantial support for the concepts defined in the model.

Also, certain descriptions regarding the interaction among the components in the model, which is supposed to be the advantage of this model over others, are not entirely easy to test at the moment. For instance, they have argued the following in describing the processing of appraisals:

> The appraisal dimensions of MAS are ongoing and processed simultaneously. The hierarchical nature of appraisal processing (Scherer, 2009b), where certain appraisals are processed before others (novelty, goal-relevance) when a new stimulus is encountered, does not change the fact that all appraisals are ongoing. Thus, changes in one appraisal produces changes in MAS as a whole, even if other appraisals are yet to be processed. We remind the reader how appraisals may constrain each another. Furthermore, re-appraisal can occur as appraisals are processed and feedback from other components is received. (p.13)

This hypothesis, in the 2009 study of Scherer which is referenced in the above quotation, is expressed as follows:

> The proposed mechanism [of the hierarchical nature of appraisal processing[35]] is highly compatible with the assumption of parallel processing as all SECs are expected to be processed simultaneously, starting with relevance detection. It should be noted that the essential criterion for the sequence assumption is not the occurrence of the event and the start of the specific appraisal process but the time at which a particular check achieves preliminary closure, that is yields a result that warrants efferent commands to response modalities, as shown by the descending arrows in Figure 2 (see Scherer, 2004, for details). In this way the assumption of continuously parallel processing is entirely compatible with the notion of a sequence of primary outcomes. The sequence theory postulates that,

[35] In the CODA model proposed by Lennie & Eerola, as quoted above, this happens in the MAS. The term Scherer uses is CPM, which stands for component process model.

> for the reasons outlined earlier, the result of a prior processing step (or check) must be in before the consecutive step (or check) can produce a conclusive result with efferent consequences. (p.1318)

While this model successfully attends to various aspects involved in the process, these aspects are too diverse in terms of their timespans and other properties for some parts of the argument to have a clear definition. The SECs, which stands for stimulus evaluation checks, are a number of criteria that operate at different levels. As Scherer describes:

> All of the appraisal criteria can be processed at different levels of processing such as (a) a low-level neural circuit [...] in which the checking mechanisms are mostly genetically determined and the criteria consist of appropriate templates for pattern matching and similar mechanisms (cf. the notion of "biological preparedness", e.g., for snakes, Öhman, 1986; or baby faces, Brosch, Sander, & Scherer, 2007); (b) a schematic level, based on memory traces from social learning processes and occurring in a fairly automatic, unconscious fashion; (c) an association level, involving various cortical association areas, which may occur automatically and unconsciously or in a deliberate, conscious fashion, and (d) the conceptual level, involving propositional knowledge, and underlying cultural meaning systems, requiring consciousness and effortful calculations in prefrontal cortical areas. (p.1314)

They are obviously of very different timescales: the action potential of a neuron, which is the basis of (a), lasts only around 1 millisecond, and the subsequent refractory period lasts also around the same time[36], whereas the mechanisms in (c) or (d), as well as the subsequent motor or psychological reactions, can be much longer. This could inevitably lead to some confusion, either because of the disparate properties of these mechanisms, or because of the difficulty to measure them in a satisfactory fashion. In

[36] This is relative refractory period. Absolute refractory period can be even shorter.

the empirical work by Grandjean & Scherer, which is also reviewed by Scherer, the authors have pointed out that:

> Full-blown emotions are characterized not only by the underlying appraisal process but also by a series of efferent phenomena like action tendencies, motor expressions, motivational processes, and subjective feeling that can, in turn, influence the appraisal process. One of the tasks for future research will be the assessment of efferent effects of the appraisal results occurring at a specific point in time. […] We expect that it is [during the results of the SECs] that efferent effects are produced, for example, somatovisceral activation and innervations of the striate musculature involved in expression. However, with our present approach we cannot demonstrate this as it requires synchronized recording of EEG signals on the one hand and somatovisceral variables on the other. We are currently working on the technical realization of this complex measurement design. (p.350)37

They also have acknowledged that “Many of [the study’s] limitations are imposed by the fact that our measurement techniques, while representing current state of the art, do not yet allow us to examine some of the issues outlined above” (p.350). Naturally, this involves one of the crucial aspects of music perception discussed in the context of embodied cognition, as motor response, either overt or covert, is a key element in the series of reactions involved. The continuous nature of music as the stimulus also means that these subsequent effects will take place while the stimulus carries on and develops, further complicating the issues identified above. The limitations in the relevant measurement would naturally result in some restriction in the scope of discussion.

Obviously, these issues do not necessarily undermine the validity of the model in

[37] The technical details of the two experiments involved will not be discussed here. The study, basically as Scherer has outlined, “systematically manipulated novelty, goal relevance, intrinsic pleasantness, and goal conduciveness SECs in visual stimuli to test the sequence hypothesis in two experiments with electroencephalographic (EEG) recordings (p.1342)”

question. If anything, these restrictions actually could suggest that the model succeeds in proposing a hypothesis describing the part of the process in perspectives that has not been explored before, hence the needs for further advancement in the neurobiological measurement. However, the ongoing development in empirical studies can require rather technical discussion. Since the technological details of the issue should not be the focus of this study, I choose to apply the structure of the older BRECVEMA model, which has more readily available support in empirical research. Future neurobiology studies may potentially lead to some amendments in the models in terms of how the mechanisms are structured and interacting, but this study will explore each mechanism by means of analysis and composition, which can remain relatively valid even when the overall structure of the models is updated in the future.

Finally, in this section, it would also be important to review certain views that appear to be the antitheses of neurobiological approaches, which will be often cited. One scholar representative of the view is Schiavio, who in his 2014 thesis has criticised that a "scientist" would automatically believe emotion is related to neural state triggered by audio stimuli, and that a there is a missing link in the assumed link between biological and emotive reactions:

> If we ask a scientist about the sense of tension elicited by the cadential chords in the traditional harmonic progression I (in second inversion) V, I, her reply would be in terms of sound waves propagated through the medium of air. Sound travels through disturbances in given medium (air for example) pressure, being consituted by waves of different amplitude and frequency.
>
> But *where* is the tension provoked by this passage if it is not present in the outside world? Where is the *sensation* that makes me feel the need to solve this melodic pattern to the tonic? I believe […] those questions to be wrong questions. […] However, most philosophers and neuroscientists […] suggest that the answer has to be found inside the brain. In particular, a standard answer would be as follows:

i) A musical stimulus reaches the sensory input

ii) The sensory input sends the signal to the auditory cortex

iii) A particular neural state would be set up

Therefore → This neural state corresponds to my subjective feelings (qualia)

As previously discussed, between (iii) and the conclusion of the argument lies the explanatory gap of the 'hard problem of consciousness'. The two solutions usually adopted to explain how (iii) and (iv) relate are dualism (qualia are immaterial entities generated by a physical substance) or reductionism (qualia are to be identified with a particular brain state). In both cases, however, there is a problem. The first approach will always face the ontological issue of causation while the second usually sees the brain as a machine isolated from the rest of the nervous system. (2014, p.31)

As will be reviewed from the discussions in the following chapters, the studies cited that involves neurological basis mostly treat the measurements as only indexes of certain physical states. The correspondence between the physical states and the feelings or emotion, in reality, is usually explained by several other mechanisms that are relevant, rather than automatically assumed based on one solution, whichever that is, for all cases.

I believe some misunderstanding related to the issue may have occurred from two sources, which eventually lead to the impression of dualism and reductionism. The first one is the fuzziness of human language that is used when discussing this topic. Different types of causality can be signified with the same connective in many languages. For instance, as linguists such as Groupe λ-l (1975), Halliday & Hasan (1976) and Van Dijk (1979) have pointed out, the word "because" could refer to physical or epistemological causality, or, some studies put it, causal relation in the *states of affairs level* or *discourse level*[38]. For instance, it makes sense to say either:

[38] Some linguists point out further distinction exist within the discourse level, and propose subgroups of *reasoning* and *speech act performance* therein (Davies, 1979; Sweetser, 1990; Dik et al., 1991). The level of speech act performance is pertinent to the filed of discourse study in linguistics, but is not very relevant here so will not be discussed.

a) The ground is wet, because it rained.

or

b) It has rained, because the ground is wet.

where **a)** assumes a physical causality where the rain wetted the ground, while **b)** refers to epistemological causality where the information that the ground is wet caused the speaker to believe that it has rained[39]. Halliday & Hasan (1976) term the causal relation in **a)** as *external* and that in **b)** as *internal*. Now, imagine a scholar says colloquially the following utterance:

c) The listener had negative emotion while listening to this, because there was obvious activation in amygdala.

They might simply mean that the neurological measurement serves as a signal of the listener's emotion, but if the internal causal relation in discourse level here is misunderstood as an external causal relation in states of affairs level, or, to put it more simply, the epistemological causality is misunderstood as a physical one, the speaker of **c)** would appear like an imaginary "scientist" in Schiavo's scenario that adopts the view of dualism.

Granted, there are indeed radical materialists that believe human emotion as well as consciousness is nothing more than the result of chemical reaction, but that is

[39] I wrote these utterances hypothetically. For actual speech data exhibiting causal relations in these levels from corpus database, see for instance Verstraete et al. (1999) where examples are drawn from COBUILD corpus.

probably not most researchers argue when they incorporate in their studies results from biological measurements, which is supposed to serve only as indexes of "how" emotion are activated, rather than essence of "what" emotions are (which will inevitably touches the hard problem of human consciousness, as Schiavo has pointed out).

The second cause misunderstanding may be misinterpretations of certain literature. Some studies may indeed present a mapping between a brain region and a certain emotion as part of the conclusion, which, if understood in an oversimplified fashion, may appear to be of the view of reductionism. However, the arguments and experiments involved behind such facades are usually much more multi-layered. For instance, amygdala is usually associated with reactions such as fear and anger, but the association is confirmed not only in "subjective feeling", or qualia as Schiavo puts it, but also actual, subconscious response. As Morris and colleagues (1998) have pointed out, amygdala would react to images of angry facial expressions, even when the image has been presented in such a short time that the viewer would not have consciously realised it. This might be due to the fact that such stimuli take another route through the nervous system, but eventually reach the amygdala (LeDoux et al., 1985). Thus, in this case, what a "scientist" would argue is definitely not the simple mapping between the feeling and the brain region based on correlation, as described by Schiavio, but rather the link between the stimuli and activation as observed in different settings, independent of the subjective emotion. The relevance with emotions is only included as a potential result that often arise with such stimuli. Another example would be studies of subvocalisation, which is the mimetic response to a certain music making behaviour on the neurological level, as will be reviewed in the next chapter. In such cases, likewise, no direct correspondence would be assumed between the activation of motor area and the subjective feelings evoked. Rather, the activities in brain region simply serve as indexes of the subvocalisation. What exact emotion is to follow would only be explained by further related mechanisms, such as conditioning or extended embodiment, which are exactly the realms Schiavio would consider to be relevant. Finally, to consider Schiavo's own example of the sensation to resolve a dominant chord, probably very few scientists would attempt to explain the entire phenomenon simply with a correspondence between a neurological state and an emotion. If one were to break down the audio properties involved, there are indeed early theories describing a certain neurobiological status as the underlying mechanism of dissonance[40], which could have something to do with harmonic tension. However, most scientists also accept the evolutionary explanation of the mechanism, namely similar properties are to be identified in animal calls signalling danger, and hence their association with tension. While there is perhaps no definitive way to validate any hypotheses embracing evolutionary views, this is still an argument that relates what is inside the skull to external information and environment, and it is widely accepted today.

Again, there may be scholars who are of the more radical view that emotion, or even consciousness, is something occurring in the brain, while disregarding other components in the cognition such as bodily experience and external information. In a sense, Schiavio has critically assessed such views, and in doing so has offered compelling arguments based on the embodiment theories, while perhaps also demonstrating unresolvable differences in ontological beliefs. However, in the actual context of music perception study, most neurobiological research has the potential to be incorporated with other theories and contextualised to form multifaceted arguments, very often not as incompatible with other perspectives as some would

[40] See 4.3.2

imagine.

As have been pointed out in the previous section, as far as this study is concerned, I primarily explore the cognitive details of emotive responses to music, and how they can offer inspiration for composition methods, but attempt to avoid further issues such as what emotions really are. Under this premise, I will organise the discussion in accordance with the structure of the BRECVEMA model. Among the mechanisms proposed in the framework, Contagion will be the focus of chapter 3, while Visual imagery, Rhythmic entrainment, and Brain stem reflex, will be explored in chapter 4. The three remaining mechanisms, which are Evaluative conditioning, Episodic memory, and Musical expectancy, are more dependent on individual's personal or musical experience. As such, simply from the perspective of the listeners' emotive reaction to excerpts of music, they are more difficult to explore and discuss in a general fashion. Instead, they will be explored in chapter 5 before the conclusion, from the angle of performing musicians and information integration, in the contexts of specific cases where these mechanisms can be observed.

As mentioned, I will discuss the mechanisms firstly by reviewing the relevant theoretical and empirical research, and subsequently by analysing their operation in existing pieces. Finally, I shall integrate my own compositions where I intentionally apply the methodologies derived from these mechanisms into the discussion. In doing so, I will attempt to produce theoretical background and composition output that will complement each other, where the written work establishes some principles to be explored through the music pieces, which in turns interprets and enriches the theories and involved in this research. Through the process, the central thesis of this study, which is that music perception involves not only audio sensation but a series of extendedly embodied processes, will be supported and elaborated in both theoretical and practical ways.

In order to observe and explore the full range of the perceptual process of music, which involves not only audio, but also motor, and cognitive aspects, the flexible structure of Experimental Music Thetare, or EMT, is rather useful. The pieces discussed and composed for this study will be mainly EMTs, and the following section will be a brief introduction and review of it.

2.3 Experimental Music Theatre

> Mamertinus—then Lampridius and Hephæstion—waxed wroth. According to them the subject of a speech was a matter of no moment. To an orator it should be absolutely indifferent whether he undertook to attack or to defend a case. Even meaning had no interest for him. The principal thing was the orchestration of verbal sounds—the melody, the musical assonance of letters—permitting even a barbarian, witless of Greek, to feel the sheer beauty of language.
>
> - Dmitry Merezhkovsky, *The Death of Gods*[41]

Recent theatre studies applying cognitive science seem to have cast a certain doubt on the viability of both the naturalistic belief of creating theatrical illusion and the Brechtian attempt of creating distanced spectating. The notion that performance resides on a continuum between ritual and play has been pointed out by scholars such as Schechner (1976) and later McConachie (2011). If one considers the fact that a role in a ritual is inseparable from the individual that takes the role, while a character in a play is supposed to represent the individual in the diegesis itself, it is not difficult to

[41] The original in Russian:
Мамертин, Лампридий, Гефестион вознегодовали: по их мнению, содержание речи безразлично; оратору должно быть все равно, говорить за или против; не только смысл имеет мало значения, но даже сочетание слов – второстепенное дело, главное – звуки, музыка речи, новые сладкогласные сочетания букв; надо, чтобы и варвар, который ни слова не понимает по-гречески, чувствовал прелесть речи.

find the continuum of performance' nature corresponding to what McConachie (2008) has defined as "conceptual blending" (2008, P.39). Applying the notion of blending proposed by Fauconnier & Turner (referred to McConachie as F and T, as shown in the following quotation), McConachie has argued that:

> audiences generally "blend" the actor and the character together into one image, one concept of identity, to enable their affective immersion in the performance. Like other conceptual universals, identity can work singly or can participate in the more complex cognitive operation of blending. According to F and T, blending is learned in infancy and soon occurs automatically to generate complex cognitive concepts, mostly below the level of consciousness. (p.42)

According to McConachie, this serves as the basis of spectatorship, namely what spectators perceive in a theatre role is a mixture of the actor playing it and the character created thereby. He then goes on to argue that similar processes are rather pervasive: "Because of its flexibility, conceptual blending structures and enables all role-enacted games of make-believe, from playing with dolls to professional tennis" (p.43).

Indeed, as long as an activity involves some kind of performance, the blending of identity seems to occur naturally. McConachie has given an argument that compares the blending in art to that in theatre:

> Interestingly, the Roman critic Horace used the term "blend" (in Latin) to describe the goal of art. "The poet's aim is . . . to blend in one the delightful and the useful," he said. Although Horace knew nothing of the neuroscience and psychology of conceptual blending, he hit on a definition of art that suggests the necessity of blending to enable dramatic impersonation: "useful" actors must combine with "delightful" characters before a blended actor/character can emerge. (p.43)

Moreover, I believe, the requirement can be observed in not only the "delightful" world (art), but also the "useful" (for instance, the speech of an orator). Thus, since the essential ingredient of dramaticality, namely the process of blending, is itself a mechanism that operates in both artistic and realistic and world, the theatre, as a result, could perhaps also be considered a mixture of both in this regard.

If a deconstruction of theatrical illusion guarded by "the fourth wall", a concept popularised by 19th century realism, can indeed be supported, perhaps some parallelism could be drawn between what Lehmann describes as "postdramatic theatre" (2006) and experimental music theatre. As Heile (2016) has pointed out, the traits of "parataxis" and "simultaneity" of postdramatic theatre, as outlined by Lehmann, can be applied to describe experimental music theatre as well. Parataxis refers to a state of theatrical elements being independent of one another rather than forming a clearly structured hierarchy. Similarly, simultaneity dissolves the conventional holistic experience of theatre and allows spectators to obtain individually diverse perceptions where one aspect may overwhelm another. These traits allow an environment where it is possible to separate the aforementioned mechanisms and play with them independently.

One intriguing question, as Rebstock (2012) has raised in the first chapter of his book on experimental music theatre[42], is that although each element is expected to possess equal weight, there are also "compositional principles" that seem to reside in an underlying level and govern other elements. Rebstock has regarded the parataxis of elements and dominance of compositionality as two features of composed theatre between which there exists a "latent tension" (p.7).

Such a tension is observed based on the possibility to pinpoint the effect and boundary of each element. Apparently, when music and theatre are considered two

[42] The term used in the book, as can be observed from the title, is composed theater. Here the word composed is in musical sense as in German *Komposition* is almost always musical.

distinct domains with their respective specialised vehicles, applying musicality in theatre and applying theatricality in music are both identifiable methods that can be analysed as such. For instance, when Meyerhold suggests a character's subtextual mood by the rhythmic flow of a musical piece, one can identify certainly that it is an application of musical element in theatre (Schmidt 2014, Roesner 2016). Similar application is not limited within the performance, but can also be found in training and rehearsal, where for example Chaikin introduces the mindset of "jamming" and "idiomatic improvisation", inspired by jazz musical characteristics, to the actors (Chaikin 1972). Similarly, by the definition from a similar context where the two fields have their respective conventions that are established and well-defined, musical bodies that carry theatrical elements is also a common device widely applied in compositions. Schubert's *Erlkönig,* musical setting of Goethe's ballad (itself with theatrical nature), for instance, has characterizations and dialogues achieved by the singer that plays four different roles with nuanced expression of voice, and by the piano accompaniment that represents the dashing horse. An example with instrumentation in bigger scale, *Ouverture Solennelle 1812* by Tschaikovsky features a range of tunes representing people and conflicting armies. In the middle section the melodies of *La Marseille* and Russian folk music compete against each other, in a way almost like two actors' blocking in the abstract musical space. There are also artilleries, either actual ones or audio representation, that serve as the prop in the narration.

However, there are also plenty of examples where such clear-cut distinction between the two fields is not effectively present or theoretically necessary. The narrative music of Johruri, quoted at the beginning of this chapter, for instance, has musical accompaniment in its performance, but the lyrics itself is highly theatrical, and the delivery of the vocal is in fact termed as "telling" rather than "singing"[43]. While such a genre could theoretically be considered as a storytelling performance that combines both musicality and theatricality by a definition external to its convention, and some tension between the juxtaposed and hierarchical operation of elements, for example, the combination of prosody and melody in the delivery of vocal performance, and the domination of rhythm created from the plucked strings, could be pointed out, in effect, it is merely a setting that carries out its function using the devices available[44], and the elements that could be considered musical or theatrical are in fact results of the practical background.

Because similar possibilities of blurring the roles and boundaries of elements are constantly present in experimental music theatre, it is essentially difficult and not always necessary to define a piece as primarily musical or theatrical. Moreover, it is exactly the liberation of elements that results in such a fuzziness, which eventually makes it ideal for one to play with certain effects in less restricted way, which is intended in this research's composition. Indeed, previous works already show how EMT can provide an ideal background to for the perceptual mechanisms, many of them mentioned in the music cognition models, to function.

?Corporel by Globokar (1982), for instance, is a percussion piece in which no traditional percussion instrument is used. Instead, the performer generates sounds using their body, in percussive manner. The performance note specifies an assortment of percussive techniques and body parts that are combined to produce the expected sound, such as palm hitting a soft torso area or fingertip moving along a limb. Other

[43] 語る kataru, rather then 歌う utau

[44] Johruri can either be performed independently, or accompany other theatrical pieces. Depending on its function, the performers can be placed in different parts of the stage and use different members of lutes of the shamisen family, and the delivery will be adjusted accordingly. For instance, the Johruri for puppet theatre (bunraku) is placed on the side of the stage and uses a low register instrument, which sometimes also simulates certain sound effects.

techniques such as clapping and vocal elements are also used. Overall, the audio elements are displayed and then developed in nonretrogradable rhythm (a pattern proposed by Messiaen). Thus, despite the absence of instruments in conventional sense, the piece is musical in that it is governed by principles not dissimilar to those of musical composition. Meanwhile, however, it is a theatrical piece where the performer produces, for example, an action resembling that of suicide, and even linguistic lines, in one movement. The performer is also required to be in a specific costume, although partly due to concerns related to audio effects. The rhythmic expression in the piece has uniquely evocative effects, as the units carrying the nonretrogradable rhythm are defamiliarized, being disparate from sounds of instruments that are commonly heard from musical pieces. But meanwhile they are very familiar and intimate as they are sounds produced, sometimes unconsciously, in human's everyday bodily experience. The contagion effect would also be a strong, as the sound-generating process heavily involves bodily contact that is visually presented. The simultaneity and accessibility can easily elicit covert mimetic response.

Another example is *Tennei-ji* by Udow (1999). This piece for solo xylophone is inspired and tightly connected with the tradition of Japanese Noh Drama[45]. The composer proposes three role types from the traditional program set of Noh, namely Okina, Katsura-mono and Shura-mono, which can be roughly understood as old man, woman and samurai ghost, as thematic and spiritual reference for different parts of the piece. When performing the piece, the player does not only wear Omote masks used in Noh Drama, but also is expected to adapt the entire body language of Noh, which is minimal in movement and highly condensed in spiritual expression.

[45] Alternatively transliterated as Nō. The character, spelled as 'nou' (のう), is written with Kanji character 能, which means ability or talent.

Performer-researchers also point out that Udow may be deliberately corresponding to other formal elements in Noh in the composition, such as the light-hearted Kyohgen intermezzo represented by a jazz-like section, or the sound of closing and hitting fan, which is one of the few props used in Noh, imitated by the audio effect of Rute Dowel stick (Wu, 2008). Also integrated in the piece is the element of theatre/concert attendants, which Kirby (1972) categorises as a "non-matrixed performance", namely the presence of the audience is regarded as a role, but merely as themselves, rather than residing in any system of acting. Indeed, one of the most achieved authors of Noh, Zeami (1363-1443), has pointed out that a large part of Noh is essentially "the art of the spectator". For performers it should always be taken into consideration to engage the spectators by adjusting the intensity of vocal, bodily, and energetic output according to their state at given moments, so that they also in a way interact with the theatre with optimal degree of concentration (Motokiyo, 1984). This is also explicitly demanded in Udow's piece[46].

In such a setting, the direct integration of Noh drama costume would make it very natural to establish visual imagery for those who have (varying degrees of) knowledge of the genre. Also, the movement incorporated in the performance could potentially invite covert mimetic response and thereby cause contagion effect. As McConachie has described, "[m]uch of the aesthetic purity and force of noh drama comes from its tight focus on the primary emotion embodied by the shite's[47] character and elaborated by other sounds and movements on stage" (p.114).

[46] The ultimate goal of integrating the spectators is to deliver the state of Yuugen, a concept of ephemeral beauty induced by transcendence and nostalgia. The art of emotional interaction between the audience and the performers (also indirectly with those behind the stage) is nowadays also being studied in the context of cognitive science. See the 2008 work of McConachie for further discussion in this topic.

[47] シテ, the leading role in a piece of Noh.

3. Emotional Contagion

This chapter is divided into four sections. In section 3.1 I attempt to establish the difference between contagion and other paths of emotion induction in the BRECVEMA model. I argue that although the difference between emotivism and cognitivism may be over-generalising, it could still have descriptive and practical value if applied in an appropriate context. Section 3.2 explores two different theories of how contagion operates and shows that they may not be in conflict, although sometimes one of them may be positioned as the counterargument of the other. Section 3.3 analyses the mechanism's operation, using *Les Guetteurs de Son*, an EMT piece by Georges Aperghis (1981), as an example. The final section describes the application of emotional contagion in a movement of my own EMT piece, *The Carnival of the Endangered Animals*.

3.1 An Infectious Mechanism

A basic distinction between what Juslin describes as contagion and other mechanisms lies in the nature of emotional induction. Contagion assumes that certain components within music itself are emotional, or at least representative of emotional elements, and they resonate with the audience, thereby transmitting emotion; whereas other mechanisms such as visual imagery or rhythmic entrainment focus more on perceptive information that tend to be emotionally neutral.

Granted, the contrast may not be clearly drawn, not only because visual or rhythmic stimuli could also carry emotional connotation, but moreover, because musical elements such as timbre and tempo may not be perceived separately, but rather are presented as a holistic entity, where they interact and affect each other. Examples of this can be seen in Russo & Thompson's research, where they have asked listeners to estimate the sizes of intervals, which were presented in different ways by various experiment designs. It is reported that the estimations of interval sizes are indeed influenced by factors such as tempo, register, pitch direction (2005), timbre (2005), and, more importantly, singers' facial expression (2010). Especially noteworthy is the 2010 experiment, where the intervals have been judged to be larger than they actually are when the audio is synced to the video where singer has facial expression that is used to produce larger intervals. As Russo & Thompson have commented:

> Facial expressions carry information about pitch relations that can be read by viewers and that influence the perception of music. Even when auditory information was available, visual information still influenced judgments. This finding is intriguing, […] [i]n that pitch relations are fundamental to musical structure and are evaluated early in processing, the findings illustrate that facial expressions are highly relevant to the perception of music. (p.321)

Facial expression is clearly a crucial part of bodily experience and highly contagious; it is also assumed to be the primary locus of emotion expression (Tomkins, 1984; Ekman, 1992). If such a deeply embodied and emotional factor is empirically proven to influence the perception of audio features such as pitch, then indications should be strong that the basic "vehicles" of different mechanisms in BRECVEMA model, or several other models on the topic indeed, may also be interdependent. This shows that in discussing these issues, it would be more ideal to include analyses of instances of music pieces or events, which is based on actual settings where these mechanisms and context interact. Such discussions will be made in 3.3 and 3.4, and also in the following chapters.

Nevertheless, contagion and other mechanisms seem to be residing on the two ends of a spectrum which, if considered in proper context, can establish a certain useful distinction. In theory, it situates the mechanisms in the context of music-emotion theories, providing explanations for several older proposed theories, for instance the affection theory (music evokes emotion) and later romanticism aesthetics (music expresses emotion) in western classical music; in practice, it serves as the basis of different composition methods.

A simplistic solution to establish the distinction would follow the somewhat facile dichotomy based on whether the emotion is "felt" or "perceived". Contagion, for instance, would be regarded as emotions that are felt, as according to Juslin, emotive induction occurs when the listener mimics the emotion in the music, thus, they are supposed to feel it themselves. Emotion evoked through other mechanisms, on the other hand, are examples of perceived emotions where information from musical elements simply expresses such emotions, reminding listeners of them.

Again, later research that has taken a closer look suggests that the contrast is unlikely to be this simple. For instance, reviewed previously in 2.2.2, the work of Zentner & Scherer (2001) has proposed a more specialised model where the "affective state", as they put it, induced by music, exists in several subgroups. They could be related to preferences (e.g., like, dislike etc.), emotions (shorter episodes of response such as angry, joyful etc.), mood ("diffuse affect state" (p.363) such as cheerful, gloomy etc.), interpersonal stances (responses towards another person such as distant, warm etc.), attitudes (longer predispositions such as loving, desiring etc.) and personality traits (nervous, anxious etc.). This is more of a bottom-up approach where they formalise these responses as subgroups according to their respective features, instead of a top-down approach where they presuppose an emotivism-cognitivism (or felt-perceived) distinction and then decide which side does a response fit in. Gabrielsson

(2001), moreover, after reviewing an abundance of previous studies, has found that the emotivist and cognitivist mechanisms themselves actually interact with one another, as well as with other personal and situational factors in particular fashions.

> Albeit there are lots of examples of positive relationship between emotion perception and emotion induction, this relationship is far from general. Examples of negative or no systematic relationship are easily found. The common belief in a strong positive relationship reflects another belief, namely that listeners' responses are exclusively, or at least, predominantly determined by musical factors. As we have seen, this is a simplistic point of view which suggests neglect of personal and situational factors. The music-person-situation interplay can never be disregarded. (p.139)

On reflection, according to the principles of embodied cognition and mimetic response, it is natural that the emotion one feels is affected, at least to some extent, either positively or negatively (Gabrielsson has reviewed evidence of both cases), by the emotion one perceives. Yet it does not necessarily imply that the distinction between the two is invalid. Even in Gabrielsson's study, it is concluded that the emotions resulting from contagion exhibit a higher likelihood to be dependent on personal factors. Also, Zentner and colleagues actually have argued that the felt-perceived distinction exists and find empirical evidence in their previous studies, where they explicitly ask the listeners to report what they "felt" and "perceived". They find that the difference in instructions greatly influences the results and conclude that a major part of emotion categories seems more likely to be perceived than felt, while only some exceptions are more readily felt than perceived. They have also pointed out that in terms of emotional valence, listeners report mostly only positive emotions when asked what they felt, as opposed to both positive and negative emotions when ask what the music expresses (Zentner et al., 2008). These empirical evidence indeed seem to suggest a certain difference in essence. The question then is, are there potentially other ways to define such a difference more effectively?

It might be helpful to take a step back and disregard for now the situation on the listener's end, namely whether they generate internal emotions or accommodate external emotions as a result of music, which, as have been pointed out, could be a distinction less explanatory then it tries to be. Instead, it is worth to consider the source that stimulates emotive reaction in the first place. To proceed in this direction, the emotion model proposed by Ortony, Clore & Collins, acronymsed as OCC model, originally published in 1988, could be an appropriate basis.

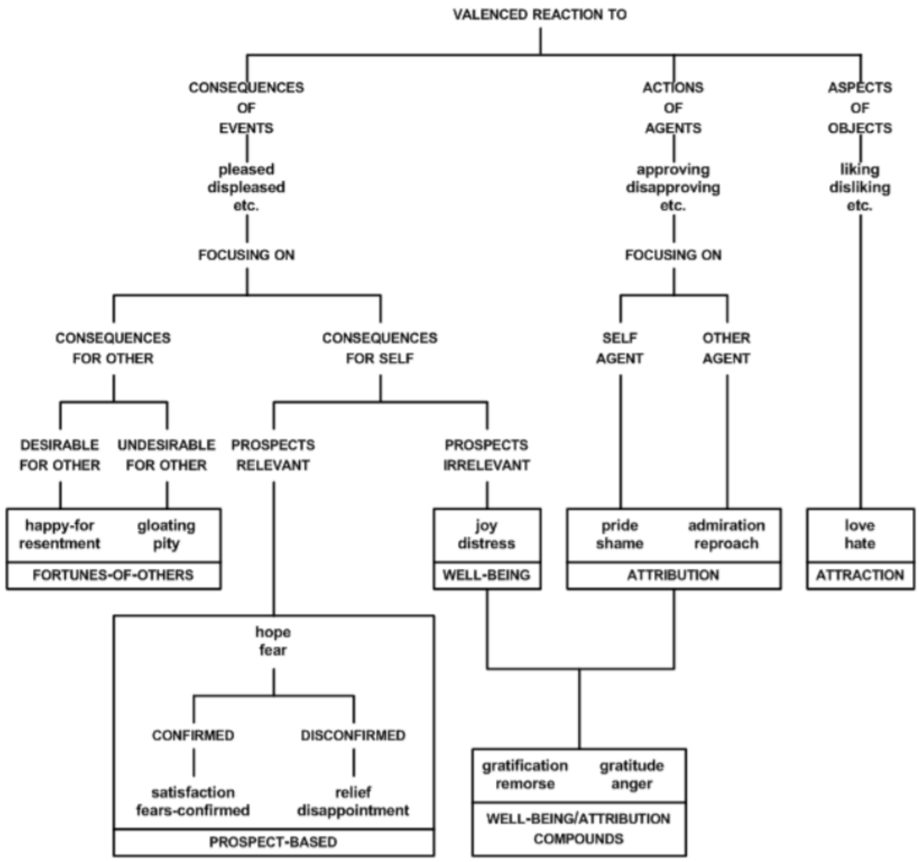


Figure 3.1 OCC Model of Emotion (quoted from Ortony et al., 2005, p.19)

The model is constructed based on an individual's valanced reaction to different aspects that stimulate the emotion. The three aspects in the theory are events, agents, and objects. As Ortony et al. define:

> Our conception of events is very straightforward-events are simply people's construals about things that happen, considered independently of any beliefs they may have about actual or possible causes. Our notion of focusing on objects is also quite simple. Objects are objects viewed qua objects. This leaves us with agents, which are things considered in light of their actual or presumed instrumentality or agency in causing or contributing to events (p.18).

Subsequently the first two have subgroups where the individual's cognitive appraisal focuses on these aspects surrounding self or others. Further subdivisions and bi-valence distinctions at the end of each branch form a framework that accommodates different emotions.

This theory is rather stimulus-oriented, and therefore has the neutrality needed for our argument here. In fact, it is also primarily the stimulus part of the framework that is applied here, as in terms of music perception, what type of emotion is induced at the end result is determined not only by the stimuli but also by other factors such as length of response episodes (as proposed by Zentner & Scherer, reviewed previously) or different types of "routes" that induce affective states[48].

[48] Zentner & Scherer use the term "central route" to describe the fashion through which music induce emotion like other stimuli; and "peripheral route" for alternative cases, which in their analysis appears to be not unlike the scenarios described in James-Lange theory of emotion production. For instance, in discussing the peripheral route, they have proposed:

A solution of the issue focusing on the stimulus-related aspects could be indeed effective, and has been proposed before as well by, for instance, by Lennie & Eerola (2022):

> The distinction between perceived and induced emotions could be as simple as assessing a stimulus' goal-relevance. That is, a subjectively identified emotional episode emerges when a stimuli is seen as relevant enough to an individual to cross a goal-relevance threshold. A perceived assessment can emerge from an externally-directed goal (i.e., directed by the experimenter) to assess stimuli and pigeon-hole them into discrete categories. A participant can evaluate the available evidence: e.g., in a music experiment—acoustic cues, changes in core-affect, physiological changes, appraisals, etc. However, devoid of much goal-relevance, evaluations remain relatively culturally standardised and the activity remains subjectively cognitively cold49 (p.16).

Derived from their discussion on the Constructivistly-Organized Dimensional-Appraisal model, the hypothesis quoted above also suggests that a distinction could be obtained by considering the properties of the stimulus, here "goal-relevance", which is a dimension of appraisal related to affordance, which, according to Lennie & Eerola, "refers to cognitive processes of evaluating the utility of different actions

---

[I]f music could be assumed to systematically influence one of the emotion components, peripheral mechanisms might be invoked to explain a spread to other emotion components, thus in fact producing emotion states that did not exist before. One potential candidate for such influence is rhythm. […] Recent evidence suggests that such coupling of internal rhythms to external drivers, as originally described by Byers (1976), might be present at a very early age (Rochat & Striano 1999). If there is, indeed, a fundamental tendency for synchronization of internal biophysiological oscillators to external auditory rhythms, such coupling may provide a promising explanatory venue for the emotion-inducing effects of music. For example, given the close relationship of respiration and cardiovascular function, a change of respiration through musical rhythm may have an impact on a variety of neurophysiological systems (Boiten et al. 1994), in many ways similar to emotion-induced physiological changes. (pp.371-372)

A more detailed description and related discussion on James-Lange theory of emotion production will be included in 4.2.2.

[49] From the original footnote by Lennie & Eerola:
Cognitively *cold* here used only to refer to the subjective experience. It is not to imply a meaningful distinction between cognitive and emotional mechanisms.

(including mental actions) that different stimuli offer (p.11)”. Thus, it becomes possible to discuss the perceived/induced distinction by analysing the quality of the stimulus, while bypassing the complex quality of the reaction, which could include a mixture of both types of emotions. This approach may in fact also be practically more feasible than what is applied in qualitative methodology, where listeners often have problems making the distinctions themselves. As Gabrielsson have commented in the 2001 work:

> A pertinent methodological problem is that neither all researchers, nor all subjects clearly observe the distinction between emotion perception and emotion induction. In reading certain reports, one feels uncertain concerning what the subjects reported: their perception of emotion in the music, or their own emotional response, or a mixture of both. (p.139)

Similarly, following the framework of OCC model, one could avoid the need to clearly prescribe whether the emotion is felt or perceived, by remaining descriptive of simply “to which” aspect is the individual responding to. Although this may appear to be only a convenient escape, a pivotal point actually exists in Juslin’s theory where stimulus aspects and listener aspect are connected, making the framework compatible with OCC model while providing the basis for contrasting contagion and other mechanisms. According to Juslin, again, listener internally mimics aspects in music, or to be more precise, the features that are similar to the prosody of emotional speech (Juslin and Laukka, 2003), and the correlation is especially pronounced in terms of anger and sadness (p.797). It is then clear that the aspect focused here can only be agents, because when a human mimics something, the target should be another human, or, alternatively, human-related properties observed from other entities. It is noteworthy that the latter accommodates different types of stimuli into the model more flexibly, and the notion is consistent with both the OCC model and the contagion effect of music discussed here. Indeed, as Ortony et al. have explained:

> Agents are not limited to people, even though they are the most usual manifestations. Agents can be nonhuman animate beings, inanimate objects or abstractions, such as institutions, and even situations, provided they are construed as causally efficacious in the particular context. When objects are construed as agents, they are just that-objects construed as agents. So, for example, a person who buys a new car that turns out to be a constant source of trouble might blame the car for his series of misfortunes. In doing so, however, he would be treating the car as though it were an agent, rather than simply as an object. In treating it as an agent, he could disapprove of it. Were he to treat it only as an object, his affective reaction to it would be one of dislike (pp. 18-19).

BRECVEMA model, and indeed various other studies, all have placed emphasis on aspects of music that resemble human vocality when discussing the mechanism of contagion. Even in another path of discussion that takes into consideration social context and sees the contagion as a result of imitation rather than mimicking (see the following section), the focus is still primarily agents instead of events. It is then perhaps safer to categorise that contagion is a mechanism where the valenced reaction occurs primarily to aspects of agents, whereas other paths such as visual imagery or rhythmic entrainment could also focus on events or objects. It still allows a tendency that emotion induced through contagion is more likely to be felt then just perceived because of its mimicking/imitational nature, but there is more flexibility because the definition does not have to rely entirely on such an emotivism-cognitivism distinction, and the listeners' report, which is itself not always reliable.

This new definition inspires composition in different ways. For instance, it seems that by integrating some elements of human musicality, it is possible to induce emotions through contagious effects resulting from the aspect of agents, regardless of a piece's theme, or lyrics and other features. How I intend to work on this in the setting of EMT and specifically this study's composition will be discussed in the last section of this chapter.

3.2 Singing/Meaning

With contagion defined as emotion reaction to the aspects of agents, the next question is then what exactly is mimicked internally from the agents. Two very different reasoning are available for this. Juslin himself, as reviewed above, considers the similarity between music and human vocal prosody to be an important cause. Indeed, in a study reviewed in the previous chapter, for their study of imagined emotion, Kleber and colleagues (2007) have chosen vocal performance as the music stimuli, not only because singing is a more universal and primed form of human music (Molino, 2000), but also it simply involves much more bodily-centered physiological processes:

> Singing and particularly singing of classical music as compared to instrumental music requires more body-core centered motor and anatomic activity involving muscles and internal organs of vital bodily importance such as respiration, salivation, ingestion, speech, and social communication while string instruments and piano playing, investigated in most brain imaging experiments, involves fine control of peripheral muscle groups, (Kleber 2007, p.890).

This universality sets this mechanism apart from other musical structures that induce emotion, for example major/minor distinction in western classical music. Scholarship on the emotion-tonality relationship points out that the source of such distinction may be more than one and contains both nature and nurture factors (Parncutt, 2014). Also in Indian traditional music, different scales are supposed to relate to different emotions and circumstances, but it is an even more codified system, and the context requires specific training to understand. In contrast, Juslin's explanation strictly focuses on only the inherent factor, namely the physical similarity between music and human speech (which Parncutt also actually considers as one of the factors contributing to major/minor tonality difference).

Furthermore, not only literal singing, but also music possessing singing-like quality may be able to activate a mimicking response. Again, when discussing neurobiological basis of subvocalisation in the previous chapter, it was reviewed that Koelsch's empirical study (2006) has pointed out "pleasant" music activates both sensory and motor brain areas. It is now appropriate to further discuss this in the context of this section. Firstly, it is worth remembering that Koelsch has defined pleasant music as being harmonically consonant. In his study, pleasant music stimuli are "joyful instrumental dance tunes" that can all be safely defined as "major-minor tonal music" (p. 241) while the unpleasant, namely dissonant music are the pleasant ones with two transposed versions of themselves, one major second above and another one a tritone above, layered over them. It is then not difficult to understand when Koelsch continued to conclude that, one reason for the harmonious music to activate motor area of brain is simply that they are easier to vocalize.

What is important here is that, the unpleasant stimuli, with has parallelly moving tunes in dissonant intervals, is not only uninviting but also objectively impossible to mimic, thus not possible to be appraised as related to an agent. This is why such stimuli are constantly found to be less likely to activate motor reaction. On the other hand, the adjectives pleasant/unpleasant are rather subjective; indeed, as Scherer & Zentner (2001) have point out, individual evaluative judgements play fundamental rules in musical emotion, and moreover, learned aesthetics may also have effects. Thus, Koelsch's use of words like "pleasant" and "unpleasant" may be misleading here, for the point, at least if one uses such an evidence as support for contagion mechanism, is actually how “human-friendly” these stimuli are, and thus how likely it is to invite imitational behavior, rather than how “enjoyable” they are, which cannot really be defined clearly, as can be concluded from the discussion above.

Naturally, a lot remains to be explored as to to what extent do individuals’ criteria of “pleasant” music vary, but this is another issue and is rather irrelevant here. Rather, it is important to note that both Kleber’s and Koelsch’s findings support that the cognitive reaction to “agents” according to OCC model, which is in practice realised as reaction to human-related feature in music such as vocality or singability, is prone to activate contagion.

The question is then how exactly does bodily engaged processes lead to higher likelihood of emotional response? To answer this, it may help to briefly examine the biological basis of what links subvocalisation to emotion, in support of mimicking as the basis of contagious emotion response. Koelsch (2006) has pointed out that when listening to consonant music stimulus, a circuit in listeners’ brain is activated that consists of Rolandic operculum, insula, and ventral striatum, and that it possibly underlies the motor-emotive response here. First of all, both anterior superior insula and Rolandic operculum are found to be activated both during singing and subvocalisation (Jeffries et al., 2003; Riecker et al., 2000; Wildgruber et al.,1996). Note that in Koelsch’s study the participants were only engaging in covert singing, so the activation here can indeed only be the result of subvocalisation. Subsequently, previous studies also show that left anterior superior insula are activated for the task of auditory planning (Dronkers, 1996) and vocal tract innervation (Ackermann and Riecker, 2004). Finally, the ventral striatum processes positive emotive response, in particular nucleus accumbens, as indicated by studies from Delgado et al. (2000). This is consistent either when listeners are presented with unfamiliar pleasant music (Brown et al., 2004; here pleasant music is defined by participants’ judgement after listening to them) or when they listen to their own favourite music (Blood & Zatorre, 2001).

The studies described above suggest contagion occurs from biological tendency to response to human-like music features in motor ways, echoing Cox' mimetic hypothesis. They usually find support in empirical evidence obtained through MET, fMRI, or other measurements. On the other hand, another interpretation of emotional contagion attempts to proceed from a completely different angle, although its start point can also be identified in Juslin's framework.

The mechanism of contagion, according to Juslin, is based on the brain function's survival value to enhance "group cohesion and social interaction, e.g., between mother and infant" (2001, p.623). Its ontogenetic development occurs as early as the first year of an individual's life cycle. This perspective is related to what is known as an inactive approach of cognition (Gallagher, 2011), where the survival value would be regarded as a functioning mechanism itself, even for adult listeners.

Regarded as a radical branch of embodied cognition theory by Gallagher, the theory has proposed that cognition emerges from the sense-making that occurs during the interaction between an individual and the environment. Similar concept is analysed by Ward & Stapleton (2012) as the interplay between four sectors, dubbed "4 Es", namely embodiment, embedding, extension and enacting of mind, which was discussed in the previous chapter. According to them, cognition is the result of bodily structure (embodiment), interaction of individual and environment (embedding), integration of internal and external information (extension) and the emergence of meaning during the above three processes (enacting).

Thus, meaning is of central importance for scholars that embrace enactivist point of view on human cognition. Delalande and colleagues, for instance, only consider an activity as "musical" when the focus is on the sound itself rather than body movement.

Schiavio (2014), after reviewing Imberty's study (1983) of children music improvisation, which accordingly already requires attention to specific audio properties, proposes the term "teleomusicality" to refer to the music-producing behaviour of infants. It has been argued that as young as six to ten months, infants are already intentionally pursuing a musical goal instead of simply repeating a movement pattern. Quoting Imberty's work, Schiavio has pointed out:

> Michel Imberty's research on children's musical improvisation (1983) showed that their musical production (both freely improvised or based on imitation) has two permanent features, named pivot and colmatage. While "pivot" refers to a stable and defined musical element, "colmatage" specifies the unstable and variable nuances of a musical production. These two sets of musical behaviours would require a particular attention to the sound properties rather than to the sensorimotor behaviours employed in order to obtain the sound. Therefore, pivot and colmatage emerge after (or during) the six-to-ten months' attentive shift. This means that this shift would allow the infant to create musical material, being intentionally directed towards the sounds without being focused on the actions performed. In other terms, it permits the constitution of a first musical context where the infant's goal is intentionally musical. We could name this basic form of intentionality as teleomusicality, considering the fundamental role of the goals of the action for its constitution. (p.103)

In the study of cognitive science, distinctions are sometimes made between mimicking and imitation, where the former refers to the subconscious copying of a behaviour and the latter to the intentional repeat in order to fulfil a specific purpose. It could probably be assumed that scholars with enactivistic views would more or less reject the idea of mimicking as the basis of emotive response to music, because intentionality, being of essential importance to all cognitive activity as the product of 4E's interaction, should be functioning as the underlying mechanism here. Indeed, Schiavio (2014) has argued that the contributing factor of listener's emotive response to music is the meaning brought forth in the specific musical environment, which is co-defined together by the listeners and the performers. The contextual knowledge of concert or any performative process as a social activity determines that listeners react emotionally to a pianist playing a piece, but would not react the same if 1) the music is performed for instance by re-playing data recorded from piano roll or automatic piano and 2) listeners are aware of the fact that no "real" performer is playing it. Naturally, this explanation places heavy emphasis on the "meaning" of the music constructed by the listener and its environment. The listener would still have emotive response if, for instance, they know the piano roll is recorded from a musician that has meaning to them, or listens to CD (which is essentially a re-play) but can relate to the music from the CD, either by recognising the performing musician or recognising recording as a particular activity and derive meaning from it. Thus, emotion can be either the result of the reaction to internal, inherent quality of music that is human-like, or the result of enaction from the external, holistic information that surrounds the musical activity and constructing meaning from it.

Finally, it is also important to note that contagion, just like any other elements in music emotional effect, should not be overemphasised on its own. As pointed out even earlier in the previous chapter, these mechanisms actually interact with each other to induce emotions in listener; and moreover, the reactions of listeners also depend on other factors that are highly individual in nature. They cannot really be analysed without taking into consideration the listeners' active participation in the listening experience. There are cases where one single mechanism may have predominant effect in a certain part of a piece, but generalised impressions of for instance "listeners are more likely to feel than to perceive a certain emotion" are usually results of combination and interaction of different procedures at work.

3.3 Contagion in EMT

Nevertheless, there are pieces which provide very good examples to observe certain effects. *Les Guetteurs de Son*, by Georges Aperghis (1981), for instance, is an EMT piece that demonstrates contagion effect very well.

A video recording of the performance by Dressage Percussion can be found at https://www.youtube.com/watch?v=CGUxZZMKwfw&. The piece, for three soloists, each of whom plays a bass tomtom with hands and bass drum with pedal, consists of human breathing voice, fragmented syllables, lines, and drumming (both movements of drumming, without really hitting the drum, and actual drumming that produces sounds). The especially inspiring parts occur in the solo sections of each musician (at 02:04, 07:38 and 16:16 from the link attached above), which are almost representative of humans' cognitive processes to establish the knowledge of sound-making movements.

The first soloist has movements of dropping hands towards the tomtom drum, but without actually contacting the drum (represented by a minim-like hallowed note in Aperghis' notation). Therefore, no sound is made in the process. However, they also lower their head towards the drum as if expecting a certain sound to come out from the movement. Meanwhile, the sound is made from pedalling the kick drum; however, the pedalling movement occurs behind the bass drum, hidden from the audience.

The two "sound-making" events form a particularly strong contrast- the audience sees the movement without sound and sound without movements, and they are designed to be entirely unsynchronised with each other. This results in a pair of drumming movement and drumming sound that occur independently. The soloist also turns their head and make rapid breathy voice every time when the bass drum hits. This comes across as a reaction to the decoupling of the sound that they "did not make" (from visible movements) and the movement they actually make (but does not result in sound).

Also worth observing is the vocal production in this piece, especially the solo passages again. After a few bars into the first solo, the soloist begins to occasionally hit the tomtom and thereby produce sound. Every time this happens, the drumming is accompanied by an “ah!” spoken by the soloist. This is however not regular, and the soloist continues to turn their head when they do not manage to actually hit the drum. The intermittent successes in drumming of this passage neatly interpret the title of piece. Les Guetteurs de Son, which translates into “the look-outs of sound”, describes a search across two different domains (visual and audio). In this solo, the musician is mostly lost at the detachment of the sound and sound-making movement, and looks for an external source that is suspected to produce sounds. It is however already hinted that sounds can be produced from an internal mechanism, which, although operating with only varied degrees of success here, is reinforced by layering two successful sound-making movements (drumming and vocalising) together.

The doubling of movements is to some extent symbolic of the excess of sound-making behaviour at early stages in infant development. For instance, more phonetic elements than practically needed (that is, of course, in one specific linguistic environment only) usually occur at early infant speech. Yet they are gradually lost at later stage when they are not supported by the “ambient language”, namely the language heard in the infant’s environment (for a more detailed review on this see Vihman & de Boysson-Bardies, 1994).

Admittedly, it is unlikely that any of the audience would remember such experiences from their infancy, before any link of movement/sound was established. The effect here therefore should only be symbolic. Yet it seems clear that Aperghis at least has some intention to illustrate different stages in the development of sound production behaviour, in that in the second solo passage, the percussionist demonstrates entirely different level of mastery on sound making. They play entirely on the tomtom drum, with a great variety of techniques, such as hitting, rubbing, finger trill, pitch altering, all displayed alternately while decorated with the same "ah" from the previous solo, but this time independent of the drumming and sounds at ease and pleasant. Another vocal production from this passage is the mocking of finger trill, accompanied by the gesture in the air and made towards the direction of two other percussionists, appearing to be almost taunting and flaunting. Also, there is no "failed" drumming in this solo; every movement has an audible result.

The last solo passage is equally sophisticated in technique, but compared with the previous solo, the playing is more multi-dimensional as it incorporates two important elements from the first solo, kick drum and silent movement. Same level of virtuosity is displayed- the soloist plays the drum with an assortment of techniques, vocalises occasionally, but they are clearly in a more aloof state of mind. They do not appear to be disturbed by external sound source (the instruments are arranged in the same way for all three musicians, so the playing of kick drum in this case is also hidden; it is designed to appear independent of the musician) as no movement is made in reaction to it; they intentionally make silent movement and are comfortable with it just as they are comfortable with sound; and also, they never direct their attention to the other two musicians. No annoyance (as in the first solo) or ridicule (as in the second solo) is shown; they do not even look away from the drums.

From cumbersome exploration, flamboyant productivity to restrained sophistication, the look-outs of the sound illustrate in their solo excerpts a schema that can be applied to the process of artistic, cognitive or philosophical development. There could be many potential extended interpretations; on the other hand, purely in terms of sound making movements, some emotive effects from these passages can be expected.

Although the entire piece has a rather reduced instrumentation and texture, the bodily experience of drumming and the musical common sense related to it (or the lack thereof) serves as compelling basis on which one can interpret the musicians' expressions as emotive. For instance, it would not be difficult to imagine that the first soloist's anxiousness would be highlighted strongly, because the detachment of movement and sound departs from common sense, or rather, more precisely, from most people's bodily experience, and is therefore very disturbing. In other words, the audio-visual decoupling is essentially "un-musical" and forms a ground for the movements such as head-turning, hushing to be easily perceived as sign of frustration. And, since the un-musicality involves a situation where one does not see what one hears (the "drumming" movement of the kick drum is hidden); and also where one does not hear what one sees (the drumming movements that do not really hit the drum), it presents itself as an uncanny alternate reality where the reliability of one's senses is deprived.

The contagion then takes place, somewhat similar to how one gets agitated when seeing a character in a film gets into danger, although it is clear that one is not in danger oneself. When one's senses are suddenly not dependable, it is also potentially dangerous; and in order to appraise the situation, the audience would vicariously feel the fret and helplessness that they interpret from the soloist' movements.

In a detailed analysis, it is noteworthy how the building up of frustration is achieved by Aperghis' structural design in the solo passages. I compile the table below to visualise the sequencing of sections in this passage.

A		B		C		D		E		F		G
ah?!	ah!	silent	ah!	ah?!	ah!	ah?!	ah!	ah?!	ah!	silent	ah!	ah?!/silent
fh fh fh fh fh fh fh fh fh	h h	fh fh fh fh	h	fh fh	h h	fh fh	h	fh	h h	fh fh fh fh fh fh fh fh fh fh fh	h	fh x29
kick hush turn sigh	turn sigh	turn pant		kick hush	kick hush pant	sigh	kick hush		kick hush	movement cresc.	turn	hush turn sigh pant

Table 3.1 Sections of first solo in *Guetteurs de Son*

The table above lists the movements in the first solo passage in chronological order. In the middle row, symbol "fh" stands for "failed to hit" while "h" stands for "(does) hit". The upper row indicates whether movements are accompanied with vocal production. Symbol "ah?!" stands for the confused and surprised voice when the soloist fails to hit; while "ah!" stands for the voice made when managing to hit the drum. There are also cases where no vocal production is made, and these movements have "silent" in the upper row. The lower row lists other movements or sounds such as kick drum ("kick"), hushing ("hush"), head turning ("turn"), sighing ("sigh"), panting ("pant"). The last column represents the last section of the solo, where the percussionist plays a string of movements which all fail to hit the drum, either with or without vocal production.

When the soloist does hit the drum, the corresponding cells are in black. When they fail to hit the drum, the cells are in grey (when there is vocal production) or in light grey (when there is no vocal production). Each pair of grey/black sequence is marked with a letter that numbers the section (except for G, which only has grey part).

The colour pattern shows clearly how the grey sections, namely the parts where the percussionist fails to hit the drum, get smaller and smaller in the first five times (from letter A to E). Notice that this may not proportionally reflect the duration of each section, as there are different pauses between each movement. Still, it will be clear that the soloist is taking fewer and fewer “failed to hit” (“fh”, grey parts) movements to get to “managed to hit” (“h”, black parts). The building up reaches its peak in letter E, which is the first time where there are fewer grey movement, only one in this case, than black movements, which occurs twice. This is the part where the soloist only fails once and succeeds twice, and between the sections there are also only very short pauses. It is almost as if they were managing to learn how to produce sound from the tomtom. However, right after the two “h” movements in letter E, the passages enters again into a sequence of “failed to hit” movements, represented by the long light grey section in letter F. Especially noteworthy is that the movements of hitting the drum, although always failing, intensify over time almost as if the percussionist were playing with crescendo. The obvious strengthening of movements forms a stark contrast with the audio result, which is silent all through. Although struggling, the soloist fails to produce the sound after all, and the detachment of visual-audio realm reaches its highest manifestation, as does the tension and fret propelled by such a decoupling of movement and sound.

After one last “h” in letter F, the passage enters into the next section, which is the longest sequence of “fh” in the passage, accompanied by various vocal productions and movements. The soloist never really hit the tomtom again and the passages exists with an outro (not represented in the table above), where the other two percussionist join in making “fh” movement altogether. The first soloist plays in a moderate tempo, while the second soloist plays it rather fast. The third soloist, meanwhile, plays it in the slowest speed among the three. This subtly anticipates the three-phase schema represented by the three percussionists, from pursuit, abundance to economy, as described above. Other than that, very few variations occurs as they play the section without much extra expression. As a result, little state of affect[50] can be observed through the movements and the outro serves as a bridge for the emotive effect to fade away, until the piece enters into the next part, signalled by the entrance of kick drum.

In the second solo passage, on the other hand, human musicality and knowledge of musical activity, the two factors reviewed in the previous section (internal musical quality and external holistic information), work together to deliver the complacent and exuberance of the second soloist. The sound they make copies the finger trill pitch contour, and copying/exaggerating audio features is a typical pattern of mockery. Also, the other two soloists are instructed to turn their heads away every time the mocking occurs. With the extended knowledge that musical activity could sometimes be competitive, the audience would be potentially able to interpret the second soloist's movement as taunting and the other two's reaction as annoyance. Again, the movements are neutral, but contagion mechanism such as inherent human vocality and extended social knowledge serve as the grounds for these movements to be interpreted as having emotional valence.

3.4 Composition Output

As described in section 3.1, this thesis views the emotive response induced by the contagion effect as reactions to a particular type of stimulus. This perspective is somewhat along the line of the analysis proposed by Lennie & Eerola, where they consider the goal-relevance of stimuli in determining perceived/induced properties of emotions. However, instead of basing the distinction on the threshold of goal-relevance like Lennie & Eerola, I apply the emotion model proposed by Ortony, Clore

[50] This term is used to mean general affective elements here. As Juslin and Sloboda define, "[Affect] is used as an umbrella term that covers all evaluative—or 'valenced' (positive/negative)—states (e.g., emotion, mood, preference). The term denotes such phenomena in general." (p.10)

and Collins, and regard the effect of contagion as emotive reaction to the stimulus category of “agents”.

In the OCC model, sources of emotion are categorised into actions of agents, aspects of objects and consequences of events. As quoted in 3.1, agents are “not limited to people” and “can be nonhuman animate beings, inanimate objects or abstractions, such as institutions, and even situations” (2005, p.18). In the following, I will introduce a piece from my composition portfolio which is developed combining the principle of emotion contagion, the notion of valenced reaction to an agent, and my own thematic concern that I hope to acknowledge.

The Carnival of the Endangered Animals is an experimental music theatre piece in four movements that combines an acoustic instruments ensemble and electronic recording of sounds from endangered animals. The recordings of the animal sounds are taken from Macaulay Library, a sound recording archive built by Cornell Lab of Ornithology, Cornell University (https://www.macaulaylibrary.org/) and the International Union for Conservation of Nature Red List (IUCN Red List, https://www.iucnredlist.org/) is used as the reference for the conservation statuses of animals.

Macaulay Library is generally regarded as the largest animal sound archive ever compiled, with more than 43 million audio and video recordings, featuring sounds of more than 12,000 species. Among the species whose sound recordings are available in the database, I chose four animals that are assessed as threatened in the IUCN red list. The details of the conservation status group is described in version 4.0 of the Guidelines for Application of IUCN Red List Criteria at Regional and National Levels (https://portals.iucn.org/library/sites/library/files/documents/RL-2012-002.pdf):

Species are assigned to one of eight categories of threat based on whether they meet criteria linked to population trend, population size and structure and geographic range. Species listed as Critically Endangered, Endangered or Vulnerable are collectively described as 'Threatened'. (p.41)

And a chart of all the categories, quoted below, can be found in the resources describing regional red list assessment (https://www.iucnredlist.org/about/regional) and p.14 of Guidelines for Application of IUCN Red List Criteria at Regional and National Levels. A detailed description of the conservation status assessment process can be found at https://www.iucnredlist.org/assessment/process.

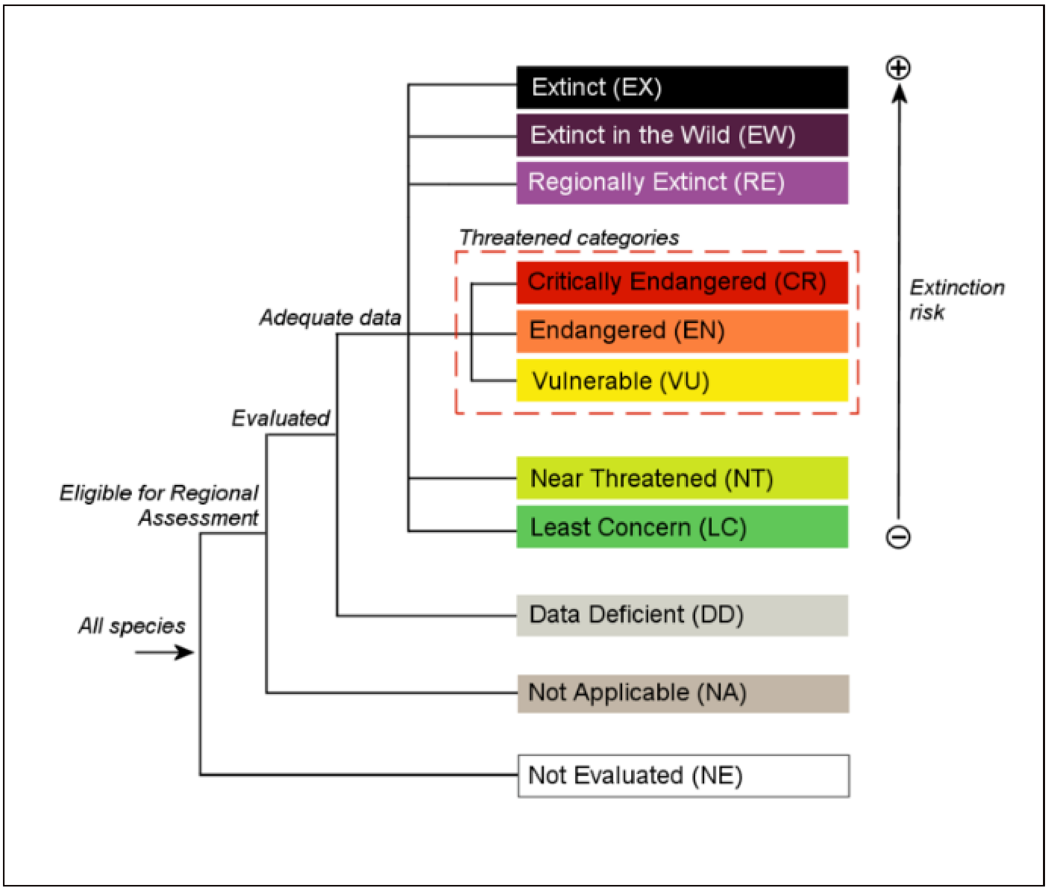


Figure 3.2 Structure of the categories used at the regional level (quoted from p.14 of Guidelines for Application of IUCN Red List Criteria at Regional and National Levels)

The online information from Macaulay Library was retrieved on 1st of November,

2018. The database has continued to be accessible online since then; however, the availability of certain species' audio recording seems to have been constantly updated, and as of 11th of November, 2022, some audio recordings used in this piece are no longer available. The online information from IUCN Red List was retrieved on 11th of November, 2022.

The species whose sound are featured in the composition are listed below:

mvmt.	Species	Conservation Status[51]
1	Philippine eagle (*Pithecophaga jefferyi*)	Critically Endangered
2	Tasmanian devil (*Sarcophilus harrisii*)	Endangered
3	Fin whale (*Balaenoptera physalus*)	Vulnerable
4	Wild Bactrian camel[52] (*Camelus ferus*)	Critically Endangered

Table 3.2 The animals in *The Carnival of the Endangered Animals* and their conservation status

The EMT was originally conceived as a single piece around 7 minutes that includes the sound of all available species listed above. I intended to integrate sounds with different qualities and create a narration which depicts the thriving, and subsequently, danger faced by each species. There were two reasons for this design. Firstly, I wanted to create evolving rhythmic patterns in the piece as it progresses from one section to another, which would strengthen the contrast among each section due to the mechanism of entrainment (this is discussed in detail in 4.2). Secondly, practically, a shorter piece with compact structure would make it more likely to be produced.

In the process of writing, however, I realised the difference in the animals' sound properties would potentially bring forth a much greater variety in terms of rhythm,

[51] Also as of 11.11.2022. The status are to be updated in their respective pages on IUCN Red List:

Philippine eagle	https://www.iucnredlist.org/species/22696012/129595746
Tasmanian devil	https://www.iucnredlist.org/species/40540/10331066
Fin whale	https://www.iucnredlist.org/species/40540/10331066
Wild Bactrian camel	https://www.iucnredlist.org/species/40540/10331066

[52] A different species from the domesticated, not threatened Bactrian camels (Camelus bactrianus) (Mohandesan et al., 2017)

tempo, timbre, and general atmosphere in the music than I expected. It did not seem reasonably feasible to incorporate sounds with such contrasting musicality in one single short movement, and separate sections would also afford more variation in elements such as props and instrumentation. As a result, I decided to compose this piece in the structure where each animal is featured in a movement.

In terms of production, nonetheless, the first movement featuring the sound of Philippine eagle (*Pithecophaga jefferyi*)[53] was nevertheless performed in 2021 at the International Tongyeong Music Festival in Korea as a self-contained piece. The event was however part of a competition and, due to regulation and fairness, some originally designed electronic and other instrumentation had to be edited out. Therefore, the result was a slightly different version, and the score and video recording of this version is not included in the submission.

In each of the movements, the animal's sound is played from a keyboard controller along with the acoustic ensemble. The composition combines the animal sound, the acoustic instruments, and the movements of the musicians, and in doing so attempts to achieve emotive effects that I wish to evoke and associate with the theme of the piece, which is the ecological concern for these species.

Instead of composing materials that are imitating or suggestive of animal sounds, I chose to directly integrate the audio recording into the pieces. The decision is based on both practical and theoretical reasons. In practice, the direct presentation of the animal sound has evocative effects that acoustic instrument cannot achieve. Sounds of birds, for instance, have been having different symbolic meanings and associations across different cultures (Mynott, 2009; N'gweno, 2010; Sault, 2010; Cocker, 2013). Psychologically, as Ratcliffe has pointed out, they commonly have the effect of

[53] This movement is not included in a separate section of the written component of this thesis, but discussion on it can be found in the commentary submitted with the portfolio.

"restoration from attention fatigue and stress", potentially through the mechanisms of "affective appraisals, cognitive appraisals, and relationships with nature" (2015, p.73).

The elements in these sounds that contribute to the effect seem to be difficult to achieve through imitations of acoustic instruments. Ratcliffe has attempted to review the qualities of bird songs in order to explain why they are perceived as restorative. She has analysed the fundamental acoustic properties such as intensity and frequency, however, they do not seem to be the direct factors. On the other hand, the cognitive reasons of associations, and what she has termed as "aesthetic properties", such as familiarity and novelty, pattern, and structure, are found to be related to the effects:

> [F]amiliarity might be a factor in restorative potential due to its role in memory formation, although individuals can, of course, form both positive and negative associations with natural environments and stimuli. (p.84)

> [L]evels of novelty versus familiarity, complexity, and pattern present in a stimulus may contribute to cognitive and affective appraisals and perceptions of restorative potential, particularly due to their potential correlates with concepts from attention restoration theory such as fascination. (p.95)

These views are in line with the perspectives of extended embodied cognition and may explain why integrating actual recorded sounds has unique effects in this context. Meanwhile, theoretically, the inclusion of sounds is justified by the intention to appeal to regard these animals not only as "other species", but also as "fellow creatures".

Ortony, Clore & Collins have pointed out that as long as "they are construed as causally efficacious in the particular context" (p.18), the source "agents" does not necessarily have to be human beings. Indeed, as reviewed above in this chapter, past studies have pointed out that one important aspect of musicality that evokes emotional response is its human-like, or human-related features. If this can be applied more broadly not only to music from acoustic instruments, but also to sound produced by animals, it perhaps can be said that "human-like" features from the animal sounds would potentially have some emotional effect as well.

It is important to note that "human-like" would have different significances in the two cases. Musical instruments are inorganic artefacts produced by humans, while animals are living creatures uncontrolled by humans. "Human-like musicality" from a musical instrument comes from an object, while that from an animal actually comes from another organism, which, in many cases, share considerable genetic similarities in terms of musicality.

One striking instance can be seen in the 2014 study by Pfenning and colleagues. Using methods based on a dynamic programming algorithm and gene expression specialization tree which they developed by themselves, they have analysed brain region gene expression databases of humans and different species of birds and primates and have pointed out that humans and birds share homologous genes related to song-learning:

> Our study indicates that behavioral and neuro-anatomical convergence for the vocal-learning trait is associated with molecular convergence of gene expression changes in the circuits that control the behavior. The findings identify specific molecularly analogous brain regions for song and speech between birds and humans, and further support broader homologous brain regions in which these specialized song and speech regions are located. […]
> Our study has identified molecularly analogous brain regions that function in song and speech in vocal-learning birds and humans. The gene expression similarities we find across species, combined with circuit and functional similarities (40,41), suggest that avian vocal learners can be used as models for speech production at a molecular level. Our experiments provide a candidate set of genes involved in neural connectivity and cell communication functions. (pp. 1343-44)

Human and other species express similar behaviours producing sound, because they share common aspects in terms of genetics and neurology. Thus, while they are distinguished by different levels or types of consciousness[54] in terms of cognition, they are at the same time connected by the same components of musicality in terms of phylogenetics. One speaks of human-like musicality in birds, while there is also bird-like musicality in humans. From this perspective, the recordings of animal sounds in this piece, in addition to providing audio samples of other species that are threatened, can also simply serve as sounds of lives that are bonded with humans.

The second movement of *The Carnival of the Endangered Animals,* for instance, includes the sound of Tasmanian devil (*Sarcophilus harrisii*), which is an endangered marsupial native to the island state of Tasmania, Australia. The sound has a creaking and breathy quality that reminds one of a suffocated individual struggling to breathe. I also include an Aztec whistle, which produces similar sound, and is in the shape of a human face contained in a mammal's head (see figure 3.3 below). While there are several techniques to achieve variation in the sound, such as creating frequency filters by forming an airtight chamber with both hands around the instrument, this instrument is rather straightforward to produce sound and is assigned to the double bassist in this movement, with the additional instruction to emphasise the breathing movement too.

[54] How close animals are to humans in terms of consciousness is still an issue being debated. Various authors have described how certain animals seem to exhibit specific behaviours which can be regarded as the evidence of animal consciousness, such as Linden (2000, 2003). Some scholars believe such behaviours merely develop from "adaptation restricted to a single goal" (Premack, 2007, p.13866), while others believe they show that the animals possess at least simple level of consciousness (Griffin, 2004).

Figure 3.3 The Aztec Whistle used in movement 2

In the beginning, the Tasmanian devil and Aztec whistle engage in an interaction that is similar to call and response, accompanied only by the sparse notes of celeste and cello pizzicato. The musician in charge of the synthesiser is instructed to perform the movement of inhale and exhale along with the sound of the Tasmanian devil. From bar 14 (around 00:45), the woodwinds join in, while the double place the Aztec whistle on the ground in front of the ensemble before changing to the double bass. From bar 20 (around 01:10), the double bass, along with the cello, plays downward portamenti with strong bow pressure, creating low and harsh rumbling sounds. From bar 32 (around 01:57), the woodwinds also stop, the piece returns to the sparse texture, this time with only celeste, cello, and Tasmanian devil, but without the Aztec whistle. It is not until bar 50 (around 03:16) that the whistle joins back, and bar 56 (around 03:48) that the woodwinds also start to play again. The movement enters the final section consisting of repeating chords on the celeste, long notes from the wind trio, cello portamenti with strong bow pressure and the duo of Tasmanian devil and Aztec whistle. All the instruments come to an end at bar 65 (around 04:30), leaving only the Tasmanian devil to cry for the final three times. Below is a table that lists the sections and the parts they play in this movement.

	bar 1 (00:00)	bar 14 (00:45)	bar 20 (01:10)	bar 32 (01:57)	bar 50 (03:16)	bar 56 (03:48)	bar 65 (04:30)
Tasmanian devil	Cry	cry	cry	cry	cry	cry	cry
whistle	whistle				whistle	whistle	
bass			port.				
cello	pizz.	pizz.	port.	pizz.	pizz.	port.	
celeste	chords	chords		chords	chords	repeated chords	
woodwinds		chords	chords			chords	

Table 3.3 Sections in *Tasmanian devil*

I intend to create a fragile, intermittent connection between the marsupial and the whistle, the creature, and the artifact, which are similar in sound qualities but manifested by opposite vehicles. The animal's cry is performed from a piece of machine (keyboard controller) which may not be wholly visible to the audience[55], while the whistle is blown from an instrument that has an animal-shaped façade facing towards the audience. Both musicians will have movements of breathing, which occurs naturally on the whistle and is instructed on the synthesiser. The movements and sounds form a link of breath between the two, which struggles to keeps its continuity and is on the brink of being, or in some part simply is, broken.

The arrangement has two purposes. Firstly, I attempt to evoke a sense suffocation by inviting a mimetic response from the audience. The continuing screeching cries and heavy breathing movements are audio and bodily elements instantly comprehensible to humans, and the reaction usually accompanying these sound and movements, such as discomfort, anxiety, and even affliction, can be reminded and identified. I try to further highlight these reactions by contrasting them with the nonchalant and even slightly jocose background created by the celeste and woodwinds. This is a goal which

[55] This being the second movement, while the first movement preceding it has the sound of Philippine eagle (Pithecophaga jefferyi), it should hopefully be established that the recorded sound coming from the playback system is that of an animal.

I aim to achieve appealing to the basic, subconscious response to the certain stimuli in this piece, and I hope the listeners' reaction will be present even without the knowledge that these animal sounds are from critically endangered species[56].

Secondly, as also argued above, I attempt to blur the distinction between human and other species in this context. In addition to emphasising the aspects which can be seen as common musicality shared by animals and humans, I also arrange the symbolic design where animal sound is played from an electronic artefact while artificial sound is played from an animal-shaped figure. As I have argued, I am trying to present the animals not only as "other species", but also as "fellow creatures". As Doolittle (2008) has pointed out:

> [U]ntil the latter part of the twentieth century, animal songs were kept as "other": composers for the most part believed that animal songs needed human intervention to turn them into music. […] Recent changes in scientific and philosophical thinking about the relationship between humans and other animals, however, ushered in a new willingness to accord the animal voice inherent value. Human composers may of course still interact with animal song -- manipulating it, mixing it with non-animal music, using human musical techniques to explore it more deeply, or figuring out the best ways to present it to a human audience -- but these are only ways of making the animals song more accessible to humans. Whereas earlier composers believed such interventions were necessary to render animal songs musical, many now believe that the music is already there in the animal's song, with or without any interpretation by a human composer. ("V. Conclusion", para. 1)

While Doolittle's argument focuses on the inherent structure of animal sounds that readily contains music, I would like to add that they also contain musicality, which is an important key to access them as music, or as events with musical significance.

[56] Although I clearly have the intention to render the listeners cognizant of the fact that the sounds are from endangered species, hence the title.

Doolittle has given an example of the redshank (*Tringa totanus*) in Scottish Gaelic folklore, which is “a wading bird that spends most of its time between the high- and low-tide marks” and “serves as an intermediary between the worlds of the living and the dead” (“II. Connecting: Animal Songs in Rural and Traditional Cultures”, para.6). When a person is dying, a magical song based on the Gaelic rendering of the call of redshank can be sung to them, helping them to cross between the two worlds. Here, the natural and neutral behaviour of the bird is given a meaning based on the symbolic interpretation by human, and is integrated in the ritual as a guiding figure for the deceased human. Similarly, in this movement of *The Carnival of the Endangered Animals*, the natural and neutral call of Tasmanian devil (whose sound is recorded here without anyone inflicting pain on the animal) is given meaning based on its musicality reminding human with the bodily experience of the respiratory system, and included in the movement as a member of the struggling symbiosis, which is symbolised by the fragmented chain of inhales and exhales.

The two purposes of the design discussed above also serve as a demonstration of the methodological model I try to propose in this thesis: Firstly, the writing process is a dynamic combination of aesthetic decisions of mine and theoretical goals in the framework. The prominent timbre in this piece is chosen for the purpose of emotion contagion, which is theory-oriented, but the exact valence of the targeted emotion is decided based on my thematic concern and overall compositional goal. Subsequently, the result of such as combination, along with the design in other elements such as the props, is an EMT piece that expands the theory of contagion by incorporating animal sounds and their musicality as agents. This is an example of how the aesthetic and thematic goals in practice have multidirectional relations with the theory, in that the practice, while containing elements methodologically constructed based on the theory, also has aspects that confirm and reinvent the theory. This could also potentially provide a model framework for other practitioners who are interested in applying similar theories: if an agent element can be identified or constructed from a specific agent/object/event, there will be potential to explore the emotion contagion effect from new angles.

As a result, this movement serves as one important example of my endeavour to respond to the theoretical framework with practical output. On one hand, the theories of contagion mechanism and mimetic hypothesis provide the basis on which I develop the methodology in this composition; on the other hand, the composition has elements the gives concrete and extended interpretations to the theories, namely the contagion effect evoked by a particular type of stimulus. The end result contains the effect which, combining the audio properties and symbolic meaning of the elements used, appropriately serves the purpose of the thematic concern of the piece. More of this piece will be discussed in the following chapter.

4. Visual, Rhythmic and Audio Elements

The three sections of this chapter focus on music's emotive effects via the mechanisms of visual, rhythmic, and audio structure respectively. They correspond to the three mechanisms in BRECVEMA model: 4.1 on visual imagery, 4.2 on rhythmic entrainment, and 4.3 on brain stem reflex. I will discuss the application of the mechanisms in my composition in the last part of each section.

4.1 Visual Imagery

This section has four parts. 4.1.1 is the discussion on how visual elements work in *Dressur*, an EMT piece by Kagel, to create emotive effects. 4.1.2 explores the basis of visual imagery mechanism, in terms of both music-making and non-music-making elements. Subsequently, 4.1.3 briefly investigates the interaction between visual imagery and other mechanisms, and finally 4.1.4 discusses *Three Singers on Planet M* and *Remember How Apple Smells?*, which are two of my compositions that involve the mechanism of visual imagery.

4.1.1 *Dressur*: An EMT Piece with Intermodal Reference

EMT offers an abundance of opportunities to observe the multisensory interaction between visual and audio, which is helpful in investigating the visual components in music, music-making and music's emotive effect. The flexible, fluid setting that waver between the setting of musical and theatrical piece, a state that Heile (2006) has termed as *metaxis*, particularly allows mechanisms related to visual elements operate in variable fashions. An instance of such in-between-ness can be found in Kagel's 1977 piece *Dressur*. In this piece for 3 percussionists[57], Kagel has constructed a stage on which a xylophone can be both a musical instrument and a part of theatrical scenery; and blocking may serve both acoustic and dramatic purposes.

The arrangement of instruments on the stage instantly reminds one of a dressage arena. Percussionist II, who faces the audience, is behind a xylophone. On each side of the stage there is a desk, on which lay the instruments that percussionist I and III, who are staged near the desks, will use.

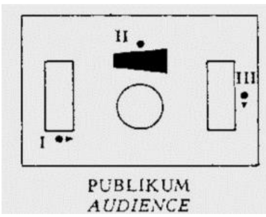


Figure 4.1 The diagram showing the stage arrangement and positions of percussionists in *Dressur* (cited from Kagel, 1977)

The front side cannot have another desk or instrument, which would block audience' sight, but Kagel has put wooden planks there, completing the four sides of an imaginary arena. As pointed above, these setting are both musical and dramatic. Even on the front side, the planks that seem to be the mere complementary side have in fact the duality of acoustic instrument and theatrical scenery, for in later stages of the piece, percussionists will approach them and hit them with other objects (which in turn have the duality of musical instruments/theatrical props), producing specific

[57] Labeled as I, II and III in the score. See figure 4.1 quoted above. In the following I will refer to the players as percussionist I, percussionist II and percussionist III.

sounds.

The piece starts with percussionist II playing the beginning few bars of the famous *Souvenir de Cirque Renz* by Gustav Peter. Although it is only a few seconds before the tune starts to develop and become defamiliarised, the iconic circus music, together with the stage setting and the title, instantly establishes the image of a dressage arena. Notice that it is only in the metaxis of EMT that it is possible to have the visual imagery of *Souvenir de Cirque Renz* extended into, or formalised in the physical world, and strongly combined with what is evoked by the conditioning effect. Otherwise, while the xylophone tune could have evaluative conditioning to evoke the idea of circus[58], it will also evoke other individual-dependent imageries/memories that are entirely beyond the control of the composer. It is of course not to say that EMT is definitely more universal; in fact, dressage is still a culture-specific activity that would elude the understanding of audience who are not familiar with it. However, the metaxis does offer the composers greater access to the musical and theatrical elements, which are deconstructed from the hierarchy of fixed genres, and can be applied at will to direct the available mechanisms towards their purpose.

Equally impressive is how music expectancy is manipulated in this piece. The circus music played by percussionist II is abruptly interrupted by percussionist I, who smashes a chair on the floor. However, they soon resume playing, with more and more variation, while percussionist I tries repeatedly to stop the theme by hitting chair, wooden sticks, sometimes approaching percussionist II. Percussionist III plays castanets for the entire duration, at first in a rather irregular rhythm but then starts to sound like galloping horse, while percussionist I approaches percussionist II holding

[58] Again, the mere idea of circus, just like a character represented by a leitmotif in a Wagnerian opera, would probably require little evaluation in terms of positive or negative emotion, but Juslin’s term evaluative conditioning will be applied for the sake of consistency.

the chair above their head. Eventually percussionist I appears to be about to hit percussionist II on the head with the chair, but ends up gently putting the chair behind percussionist II, who sits down and holds the mallets horizontally in their mouth, like a horse holding bits in the bridle. Again, in the environment of metaxis, the otherwise individually dispersed music expectancy is in some degree shaped by visual and contextual information from the title and setting of the piece. The familiar tune of circus music being interrupted and gradually modified, while accompanied by the galloping rhythm from percussionist III, will be understood as the process of the training (dressage). As the xylophone plays with variations such as retrograde and augmenting leap intervals, the unexpectedness will build up, and because of the context of the piece, listeners will be looking forward to a "resolution", not only harmonically but also dramatically, where the training is complete. Even more unexpectedly, Kagel delivers this resolution not musically, but visually, by having percussionist II representing a bridled horse through holding mallets in their mouth.

There are two mechanisms with intermodal aspect here: the "music expectancy" built up in audio realm is fulfilled by a representation in visual realm, while the "visual imagery" of the horse is completed by both audio (galloping rhythm) and visual (bridling movement) information. As analysed, almost every device involved have the musical/theatrical duality, and it is by using these elements that Kagel produces an ingenious narrative of the training process, which opens the entire piece.

This piece is an ideal example of intersensory interactions in EMT, which involve not only visual imagery, but also its combination with other mechanisms. The following section will examine closer the nature of visual imagery and its interplay with evaluative conditioning and music expectancy.

4.1.2 Visual Imagery: A Bi-layer and Bi-directional Framework

For the discussion of visuality in musical emotion, in order to incorporate different elements and mechanisms as comprehensively as possible, it would probably be beneficial to consider both musicmaking-related and non-musicmaking-related imageries in the framework. Originally, the category mainly discusses the latter, i.e., mental imageries that are related metaphorically to music through schemata derived from bodily experience, such as how beautiful scenery is associated with pleasant music (see Lakoff & Johnson (1980), Juslin (2011), Juslin (2013)). However, from the discussion from the previous chapter where various theoretical and empirical research were reviewed, it should be clear that imageries related to musicmaking should be taken into consideration too, although this category of imagery would have even greater individual differences among different listeners[59]. I base this argument on **a)** the abundance of biological measurement results which point out that neurological circuits of musicmaking are activated while listening to music, and in turn **b)** the theory of enactment that points out the knowledge of musicmaking is an essential part that gives meaning to music.

Both musicmaking and non-musicmaking imageries can potentially work in a bi-directional fashion, which is to say they can both give emotive quality to music or receive emotive quality from music. Consider how the evoked image of *Souvenir de Cirque Renz* in *Dressur* transfers its circus-like atmosphere to the piece and later is given more tension as the piece proceeds musically and theatrically. Indeed, as pointed out in the previous section, EMT, which treats movements and props as independent, active compartments, is an ideal genre where this mechanism can be explored and

[59] While Juslin (2013) already points out that mechanism is highly individual-dependent, further objective factors seem to also play a role when musicmaking movement is involved. For instance, listeners at higher age were reported to be less accurate in multisensory emotion recognition. For a detailed discussion, see Ruffman, Sullivan and Dittrich (2009).

utilised. Its significance to emotive effects of music is not unlike that of computer simulation to universe formation, in that it can re-present the process of how visual-audio association takes shape.

4.1.2.1 Music-Making Movements as Visual Stimulus

Both sections 2.2 and 3.2 have reviewed studies which show how motor circuits can be activated when listeners hear music. The discussion has covered mostly empirical research which outlines the biological basis of this activation, but exactly what these motor reactions do in emotive terms has not been investigated yet. Obviously, as in the cases of most research with neurological measurement as methodology, knowing how the mechanism is triggered is in no way tantamount to knowing what exactly the listener is experiencing, as currently there is no method to actually record the “imagery” of activities such as subvocalisation and (if it occurs) subsequent emotive response. One alternative way to gain some insight on this issue, however, is to investigate music-making performers’ real-life motor information, and see what effect it has on listeners.

Multisensory integration of music perception has gained emphasis over recent years, and visually perceived emotion in music is a subject widely explored, usually with the conclusion that audio-visual presentation of music evokes greater emotive response than audio only presentation does (or visual only, which takes place as control group in studies). Mitchell & MacDonald (2016) investigated listeners’ ability to distinguish the nuanced differences in expression from different musicians’ performances and have found that they tend to be able to correctly match an audio recording of the performance to a video recording of the musician playing, but are less likely to be successful when asked to recognise performances when they are presented only in audio. In an earlier study, Dahl & Friberg (2007) played silent video of music performance to audiences and asked them to rate the emotional intentions" that the musicians were supposed to express. The results have suggested that many emotions in the music performance can be successfully recognised even without audio information, and Dahl & Friberg have concluded that "it is possible for a musician to convey specific emotions using body movements only" (p.448)[60].

Biological measurements have led to similar conclusions. For instance, Chapados & Levitin (2008) have reported greater skin conductance[61] from listeners who were exposed to audio-visual presentation of music than those who received audio only or visual only presentation. From the results they have drawn a conclusion that places emphasis on the nature of perception as a gestalt, a view that can potentially also accommodate other elements in theatre settings:

> The higher levels of electrodermal activity for participants who could both hear and see the performances, as compared to those who could only hear or only see, indicates that the interaction between the two sensory modalities conveyed by musical performances created an emergent property, a holistic perception that was greater than the sum of its parts. This result is consistent with the tension study (Vines et al., 2006), which also identified evidence for emergent properties in their AV condition, using subjective behavioral measurements of tension and other emotions. It appears that there is a nonlinear summation of physiological arousal when both sensory modalities are observed. (p.646)

Visual information even has the ability to "distort" the emotions expressed by

---

[60] Additionally, it is noteworthy that, in the 2007 study, Dahl & Friberg have reviewed another research on dance, which has mentioned how the visual understanding can be influenced by other stage elements, such as lighting:

> Dittrich et al. (1996) reported that when presenting observers with point-light dance performances, Anger was likely to be mistaken for Joy, and vice versa. When the dances were showed in normal lighting, however, the confusion appeared between Joy and Surprise. (p.449)

[61] Again, termed electrodermal activity in the original. Skin conductance will be used in this thesis for consistency. Similarly in several studies to be mentioned in the following, which will not be footnoted repeatedly.

music. Krahé and colleagues recorded performances of Beethoven's Bagatelle (WoO 54), whose two sections were very straightforwardly labeled "lustig" (happy) and "traurig" (sad). They then produced two versions of the video recording, one with the original audio-video matching, the other with the "lustig" part audio matched to the video recording of the "traurig" part and vice versa[62]. The listeners were shown either the original or the incongruent version of the performance recording and asked to evaluate the emotion felt or perceived. Krahé and colleagues have found that those who saw the mismatched version are significantly more likely to report the sad music section as happy. Interestingly, though, in the other section, visual information was actually overridden by audio ones and the happy music section is still perceived to be happy. The different result shows very well the bi-directional nature of multisensory integration, as visual information can both gave emotive meaning to or receive emotive meaning from music[63].

4.1.2.2 Scenery and Item / Scene and Prop

The imagery that is not related to the production of music can also be discussed from the angle of real-time audio-visual pairing to have some idea as to how multisensory integration operates. For instance, Pehrs and colleagues (2014) matched a kissing scene with either happy, sad or no music. Instead of asking the audiences to self-report the emotions after viewing the video, this study showed the video to audiences during fMRI scanning. They found that stimuli with both video and audio triggers the network of fusiform gyrus (FG), amygdala (AMY) and anterior superior temporal gyrus (aSTG), where aSTG modulates the coupling between FG and AMY. In

[62] WoO 54 is originally for piano solo but in this study Krahé and colleagues used clarinet version of the piece for the reason that body movement on this instrument is more salient.

[63] The instructions in the study asked listeners to evaluate the *performance* that they *watch and listen to*, so their answers should apply to both visual and audio stimuli.

a later study, Taruffi and colleagues (2017) used close-up of sad characters as video stimuli, half of which are accompanied by music. The same network is again triggered[64], and the activation strengthens when both music and video are presented.

Again, as Taruffi and colleagues have interpreted in these studies, the effect of audio-visual stimuli on emotive response happen in a way that they modify each other, here exhibited by how the connectivity from FG to AMY is constantly being affected by aSTG when both stimuli are present. Although these studies use visual perception in real life, the key region in charge of multisensory integration, namely aSTG, is reported in another study to be activated also from mental imagery, while FG, the region in charge of processing and sending signals to AMY, also demonstrate reliable overlap in activation from visual perception and mental imagery[65]. Therefore, it would probably be safe to infer that a similar mechanism happens for visual imagery also (for a detailed discussion on visual perception/mental imagery fMRI study, see Ganis et al. (2004)).

4.1.3 Context and Interaction

Discussions and analyses above have shown that visual imagery can be highly individualized, but when functioning in specific contexts or in conjunction with other mechanisms, such as in a piece of EMT designed to activate particular imageries or in an activity involving sense-making based on holistic perception, it has aspects that are more consistent across individuals. This section will investigate its interaction with other mechanisms and attempt to provide an overview of the processes.

The mechanisms of evaluative conditioning and musical expectancy in BRECVEMA model are particularly related to visual imagery. Evaluative conditioning refers to the acquisition of a pairing between musical and another stimulant. It is in this category

[64] With the addition of temporal pole (TP) in this study.
[65] With the right fusiform gyrus being rather inactive from mental imagery.

that Juslin introduces the example of the famous Leitmotiv in Wagner's music, where melodic, timbral, harmonic or other musical elements are used to represent characters, objects, or events in a piece[66]. Subsequently, the intermodal association in both visual imagery and evaluative conditioning may eventually contribute to musical expectancy, where emotion is evoked when the development of music either go against or continue according to listeners' expectation of it (Juslin, 2011). Although it also appears difficult to test, relatively more theoretical and empirical research has been conducted regarding this mechanism, perhaps because it is more suitable for a context where listening is the primary activity, and therefore more clinically tangible. Earliest mentioning can be found in Meyer (1956)[67], while Narmour (1990) further analysed this mechanism through what he calls the implication-realization model, which is an axiomatic approach that sees music as structures composed of basic melodic units, constructing a framework according to which the expectancy works. Lastly, Huron (2008) has proposed a comprehensive theoretical framework in which he distils five distinct system that respond to expectancy. Emotions are evoked when listeners have responses through the five mechanisms, which Huron has further categorised into two pre-outcome responses, imagination, and tension, and three post-outcome responses, prediction, reaction, and appraisal. He has acronymised and termed this as ITPRA theory, of which he has given the following summary:

[T]he five response types are the imagination response (where a foretaste of

[66] Although Juslin specified "positive or negative" stimulants are paired with music in this mechanism (2013, 235-266), judging from this example he gives, it seems also possible that what the musical stimulus evokes may not have a particular emotional valence.

[67] The terminology in this early book, *Emotion and Meaning in Music*, has some consistency with the recent definition given by Juslin and Sloboda. Meyer also makes the distinction between "emotion" and "mood" based on the duration, quoting Weld (1912). The definition of "emotion" is also close to emotive response (except for when the term is used to discuss emotive state denoted by the music, but the distinction is often made explicitly), although it is sometimes used interchangeably with the more general "affect".

> possible future emotions occurs), the tension response (where arousal and attention are tailored to anticipated outcomes), the prediction response (where predictive accuracy is rewarded), the reaction response (where a quick defensive response is initiated), and the appraisal response (where a leisurely assessment of the final state leads to encouraging or discouraging reinforcement). (p.138)

This theory is rather explanatory and can, for instance, fittingly describe the function of the circus tune in the introduction of *Dressur*. Then, the visual imageries evoked in the piece, by the placement of instruments or movements of musicians, can also be integrated into various parts of the framework. In fact, along the line of perception as a holistic process, this theory also has underlying notions in common with the view of extended embodied cognition. I will discuss this further in 5.2.

As early as 1956, Meyer had argued against the measurement of emotional responses in his work. However, later studies have taken the empirical approach and obtained results that could serve as a biological basis of emotive response to music expectancy. Likewise, these studies are designed for audiences that are familiar with cultural-specific musical stimuli. Steinbeis and colleagues (2006) measured the skin conductance and heart rate change of listeners who were exposed to different versions of Bach chorales. In the study the selected modulation sections of chorales are played in the original and alternative harmonisations[68], which either return to the home key instead of modulating (the "more expected" version) or modulates to an even further key on the circle of fifths (the "less expected" version). The results show indeed greater response are detected when the less expected variations are played. Also, in a study by Salimpoor and colleagues (2009), listeners were given pieces of their own choice that they find pleasing. Indicators such as skin conductance, heart rate, body

[68] They did also take into consideration the melodies, which were changed according to the respective re-harmonisation.

temperature and blood volume pulse were measured. Again, emotive arousal was detected, and Salimpoor has attributed this to "individualized top-down processes (e.g., expectancies)" (p.11). In a more recent functional neurological study, Seger and colleagues (2013) presented musical pieces with different types of cadences[69] to the listeners and find strong interaction activated in basal ganglia and cortex region. This structure has been found to be responsible for reward prediction and reward prediction error (Seger 2010; Lee et al., 2012) and is regarded in this study as a central role that processes musical expectancy.

The analyses so far in sections of 4.1 have described how visual imagery function besides other mechanisms, in the context of EMT piece and theoretical frameworks. Based on the discussion, it is possible to conclude an overall pattern of interaction: visual imagery is evoked because of certain evaluative conditioning; and participates in establishing musical expectancy. The bi-directional property of visual imagery can be very well demonstrated by this interaction, in that it receives emotive valence from the evaluation of conditioning but can also serve an emotive function in the process of establishing musical expectancy. As elaborated in section 4.1.1, in EMT pieces with intentional design and control of musical/theatrical components, the interaction can be observed in a relatively predictable form, whereas in more general cases it would occur in much greater variety among individual. Nevertheless, in order to conclude this pattern, it would be helpful to imagine a simple hypothetical scenario. For instance, if a certain piece evokes the image of moonlight, the exact atmosphere of the imagery would be determined largely by the music, and in turn the visual of moonlight would

[69] The piece is in minor key as defined by western classical music; and the study designs four different types of cadences: a "standard cadence" that ends in i; a "deceptive cadence" that ends in VI; a "modulated cadence" that ends in iii (3b) (via German sixth chord); and an "atonal cadence" that ends in a chord which, although being a rather consonant eleventh chord, is far away from the tonic (on minor second above the mediant) and should therefore be unexpected enough to violate music expectancy for listeners familiar with this style of music.

bring other features associated with it to the music, forming certain expectations. These mechanisms will be discussed in the context of my original composition in the next section.

4.1.4 Composition Output

In 4.1.3, I have attempted to describe visual imagery as a bi-directional process through theoretical investigation. Here I will elaborate on the notion and its application through discussing two of my EMT pieces, *Three Singers on Planet M* and *Remember How Apples Smell*? [70]

Three Singers on Planet M is a piece conceived in a fictional setting, the imaginary planet M. The piece includes an introduction of the background:

> Planet M is a tidally locked planet. It does not rotate; hence, one side is constantly bright and the other constantly dark. Most creatures on planet M inhabit the zone between the scorching bright side and the bleak dark side. For the intelligent life there, such an in-betweenness is an essential motif in their existential experience.
>
> Vocal trio is a common music form in many cultures. In these trios, three singers would spread out diagonally across the ground. One singer is close to the bright side, another close to the dark side, and the third singer in the middle. They sing in parallelism, and the pieces usually consist of rather simple phrases with limited variations. The development occurs primarily when the singers transpose their phrases, or move into different positions, which triggers more drastic

[70] This title is quoted from a line in *Metamorphoses* by Mary Zimmerman (1996).

harmonic changes. When there is one singer on each side (bright, middle, dark), the parallelism consists of simple ratio intervals. When there are more singers on any side, greater dissonance occurs and there are more disordered variations in the phrases. This correspondence between simple ratio intervals and evenly distributed voices on the three sides represents Planet M inhabitants' idea that harmony occurs only when equilibrium is kept. Thus, these pieces could be performed, but they also have a non-performative nature. They can be practiced with the three singing members alone, which would be an exercise where they transform their experience of in-betweenness, balance and harmony metaphorically into music.

The imaginary practice is represented as an EMT piece here. During the performance, the downstage, centre stage and upstage respectively have no light, weak light, and strong light to represent the spectrum. In the imaginary situation, the singers can produce these intervals by themselves, but since it would be impractical to ask musicians to learn the practice, scores are provided to prompt the musicians to achieve the desired intervals in each situation. The instrumentation of this piece includes a B-flat clarinet, a cello, and a free-base accordion. It represents the type of trio where the middle singer (cello) remains in the same zone. The other two singers, the clarinet and the accordion, start respectively from the dark zone and bright zone of the stage. Since they will be moving, there are stands provided in all three zones along their position. The placement of musicians and stands are shown in the figure next page.

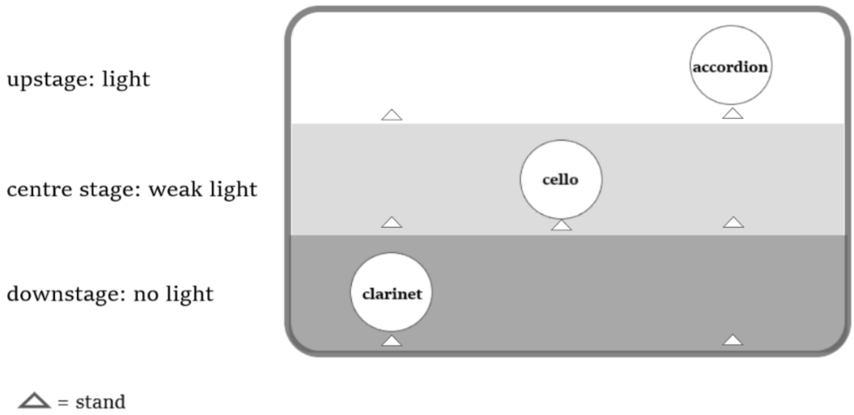


Figure 4.2 The lighting, the initial position of the musicians, and the placement of stands in *Three Singers on Planet M*

The string instrument originally planned for the piece was violin or viola, so that the musician could also move while performing. This was intended so the coordinated footsteps can create some effects of rhythmic entrainment. The final instrumentation was used eventually due to practical reasons, and also the consideration that the piece already is investigating the mechanisms of visual and conditioned elements. As pointed out in the introduction, while I am aware that the process of musical emotions ultimately involves many routes which all interact, I try to focus the controlled aspects in the compositions here for the function of methodological demonstration.

The piece starts with an exposition of themes, repeated three times in parallel fourths (00:00 to 00:39). The same theme is then repeated with singers transposing their parts, recreating the theme in the same intervals but slightly different timbres every time (00:43 to 01:14). Subsequently, the accordion moves to the centre stage, and the same progression of timbre through exposition-transposition is repeated in parallel fifths, but this time the drone on the accordion starts to form transpositions of minor second with parts of other singers in the second and third phrases, gradually creating stronger dissonance (01:15 to 02:35). The piece then progresses into the next part, after the clarinet also moves to the centre stage, and the three musicians congregating in the same zone start to play the phrases with greater dissonance and disorder (02:36 to 05:34). Finally, the clarinet moves to the upstage and the accordion upstage (05:35 to 05:42). The cello plays an accompaniment in single note ostinato, while the clarinet and the accordion plays fragments of the theme phrase, again in fourths (05:43 to 05:53). While the ostinato from the cello continues, the clarinet and the accordion return to their initial positions and play the fragment in fourths one last time (05:54 to 06:21).

I attempt to create a link between the even distribution of the musicians and simple harmony in the music, which corresponds to the notion that exists in the fictional setting of this piece. The visual imagery of musicians congregating into the same zone and increasing dissonance will build up as the piece progresses. Audience could understand the link in accordance with the philosophy introduced in the setting, but they could also derive interpretations of their own, based on the connection between musical and visual development, which itself is a neutral imagery.

The other work, *Remember How Apples Smell?* is an EMT piece with several movements of two sections, which are two variations of the same theme. Between the variations, one of the musicians will always bring out an apple and place it on a desk, which is set in the middle of the stage, and remove it in the end. In the earlier stage of the piece, the movements are based on pastiches of established styles of western classical music, and the two variations are characterised by contrasting moods, achieved by devices such as tonality. The apple serves as a visual prompt for such a contrast. For instance, *Sarabande and Minuet*, which is the first movement of the piece, beings with a slow Sarabande style dance in c minor (00:00 to 00:45). Subsequently, the musician brings out an apple and places it on the desk, and then proceeds to play the Menuet part, which is a light-hearted dance in C major (00:51 to 01:27). Later movements, however, has less readily definable contrast between the two sections. They may be very similar, or simply in the same style, developing in only less obvious properties. The apple, however, continues to be presented in the middle of each movement.[71] The established significance of the apple as a symbol of musical contrast will hopefully persist, tricking the listeners into perceiving certain differences or contrasts in the music.

Overall, the two pieces applies the nature of visual imagery that works in a bi-directional way in order to activate processes that lead to flexible possibilities in the listeners' individual responses. *Three Singers on Planet M* attempts to create evaluative conditioning through a cycle of phases demonstrating the association between musical and visual features. Similarly, *Remember How Apples Smell?* attempts to create the meaning of a visual symbol by repeating a process, which in later stages will bring out musical expectancy and could alter the perception of the music. These are examples that apply the theories related to the visual imagery mechanism in BRECVEMA model, but further demonstrate the creative possibilities by constructing novel imageries in EMT pieces which have limited pre-established meaning in the real world. Thus, the pieces do not aim for predictable associations and emotive responses, but rather attempt to allow more variable interpretations and emotions that are to be potentially established in the process of the music. They serve as examples of how the research interest, which in this case is the application of embodied visual elements, can lead to results that confirm and re-interpret the theories in another scope. The association between visual imagery and musical

[71] The submitted piece here is an incomplete one, for the design requires several movements that pastiche existing styles, which is of limited analytical significance. To demonstrate the intended structure, the submission contains two movements, which are *Sarabande and Minuet* with obvious contrast, and *Elegy* which is more obscure in terms of difference between the two sections. There will be more discussions on *Elegy* in relation to other mechanisms in BRECVEMA model in 4.3.4.

components are present in daily life and commonplace in the practical field I am experienced in, where composers create music for other media. EMT pieces independent of other works such as the compositions discussed above further demonstrate the mechanism in an autonomous context. This creates a setting which is in a sense close to a simulation, where a temporary and virtual association is formed through the mechanism. The result has the significance of illustrating the theory in a fresh context and also provides theory-informed methodology.

4.2 Rhythmic Entrainment

This section is also organised in four parts. 4.2.1 will firstly discuss the biological basis of rhythmic entrainment mechanism, while also exploring its emotion effect. Subsequently in 4.2.2, following the evidence of neurological research, I try to make a distinction between interoception and movement as two different types of entrainment oscillators. In 4.2.3, I will discuss examples of this mechanism in two EMT pieces, *Dressur* by Kagel, from the previous chapter, and *Six Elegies Dancing* by Jennifer Stasack. Finally, I will give an instance of from *The Carnival of the Endangered Animals* where I attempt to apply the effects related to entrainment in 4.2.4.

4.2.1 By-product of Mimetic Mechanism

In section 3.2, two very different theoretical basis of emotional contagion were discussed, namely that of mimicking and that of imitation. The former is defined as spontaneous reaction to human-like audio characteristics in music as a sonic phenomenon, whereas the latter as a posteriori sense-making of all the other external aspects involving music as a social activity. This section will firstly analyse rhythmic entrainment from the perspective of the former, namely the mimicking type of process. Juslin (2011, p.622) has referred to earlier studies that have shown how this mechanism synchronises listeners' heart rate and respiration (Landreth and Landreth, 1974; Harrer and Harrer, 1977), the latter even present in the case of infants (Last & Kneutgen, 1970). Indeed, these activities are all related to ANS (autonomic nervous system), which seem to suggest that the mechanism can take place without listeners' volition and interpretation.

It is possible that the mechanism occurs based on humans' inherent tendency to mimic. This is evidenced by the common biological basis shared by entrainment and mimicking. Quoting Jones' article in *The Oxford Handbook of Music Psychology* (2009), Juslin has briefly mentioned brain neurons as a potential basis for rhythmic entrainment (2011, p.622). It should be noted, however, that the *Oxford Handbook* article cites the research of Will and Berg (2007), which investigates the synchronisation between external audio stimuli and brainwaves, a result of microcosmic neuron oscillation.

On the other hand, studies of a more integrated level of the neural system could indicate more clearly the biological relation between observable output of entrainment and mimicking. Firstly, Lewis & Miall (2003) have identified two types of timing functionalities, which can be considered as the basis of rhythm perception. Perceptions of shorter intervals in milliseconds, termed "automatic timing", involves the cerebellum and motor cortices, whereas identification of longer periodicity, termed as "cognitively controlled timing", engages parietal, prefrontal areas and basal ganglia. As Repp & Su (2013) have outlined, the proposed separation of two distinct timing system is still a matter of debate (Merchant et al., 2008; Shih et al., 2009), but the involvement of the motor circuit in the process, whether subtyped or not, is very well confirmed. Indeed, as Coull, Cheng & Meck (2010) have pointed out, among various areas involving timing tasks, dorsal striatum of basal ganglia (BG), which is the subcortical nuclei primarily involving motor control, seems to be the most important region. There are also studies indicating that BG activates in timing of shorter subsecond intervals, and that it is activated even in tasks that do not ask for motor response (Coslett et al., 2010).

In section 2.2, I reviewed studies showing that audiences' motor region and mirror neurons in inferior parietal lobule are activated when listening to or watching music being performed (Koelsch et al., 2006; Buccino 2004; Gerardin et al., 2000). Similarly, even when one listens to only the audio rhythmic input, an abundance of evidence indicates that motor systems are still engaged. These rhythmic stimuli could be beat patterns consisting of only woodblock sounds (Chen et al., 2008), drumbeat (Bengtsson et al., 2009), or pure tones (Chapin et al., 2010). Again, BG plays an important role in the process, and a further study shows that damage in BG will result in failure to perceive the underlying rhythmic structure (Grahn & Brett, 2009).

Most circuits mentioned above involve regions that are evolutionarily primitive. The function of BG has been retained for more than 500 million years over vertebrate evolution (Grillners et al., 2013; Grillner et al., 2016). The cerebellum is an even older structure commonly present in many species. Thus, associated with motor mimicking, does rhythm perception have an origin that can be traced back equally far? One approach to this question is to examine the mechanism in other species. In a 2012 study, Honing and colleagues played auditory rhythm input to rhesus monkeys (*Macaca mulatta*), omitting certain units in the rhythm, either at irregular points, or regularly at each downbeat. By examining electrical activity (termed as "mismatch negativity component") in the monkeys' brains, Honing and colleagues have found that the monkeys can detect omissions of random units. An earlier study by Ueno and colleagues (2008) suggested that a similar capability is present in chimpanzees (*Pan troglodytes*) as well. However, the monkeys cannot seem to detect regular omission at each downbeat, which is something that human infants are capable of. This seems to show that synchronisation of neural oscillation and auditory input, the basis of motor mimicking, will not necessarily result in the capability to perceive rhythmic structure.

Evidence from other species, however, seems to suggest that birds such as the sulphur-crested cockatoo (*Cacatua galerita eleonora*) and African grey parrot (*Psittacus erithacus*) do perceive rhythm, indicated by their tendency to synchronise their bodily movement to a given rhythm (Patel et al., 2009; Schachner et al., 2009). Patel (2009) has proposed a hypothesis which links vocal learning and rhythmic synchronization, suggesting that only species with vocal learning ability can entrain to rhythm. This seems to confirm and add another layer to Cox' mimetic hypothesis. It should also be noted that in a more recent study, a California sea lion (*Zalophus californianus*), an animal that has been considered non vocal mimic so far, has demonstrated rhythm entrainment after intentional training (Cook et al., 2013).

Thus, in order to gain an overview, it would be important to clarify what exactly is happening in an entraining process. It is obvious that entraining is tightly related to the mechanism of mimicking, but interoceptional reactions, such as heart and respiration rate, and external movement, such as tapping (which is the task used in the above study by Coull, Cheng and Meck (2010), for instance) may have a somewhat different nature. When Juslin wrote entrainment "occurs in some way or another in all animal species" (2011, p.622), he was probably referring to more simple reaction such as brief coordination in movement or swift change in brain wave in certain frequencies, whereas what Patel (2006) has argued in vocal learning and rhythm synchronization hypothesis involves more complex learning mechanism. Perhaps, if no further evidence showing that the California sea lion has vocal learning emerges, the case of sea lion synchronising head movement to rhythm could be considered to be deriving from the former type of entrainment, and the common presence of entrainment in vocal learning birds and non vocal learning sea lion could be a result of human training and subsequent “convergent evolution”[72].

This distinction could be important for the theme of the following section, where I shall discuss how entraining to rhythm results in emotion. Juslin has pointed out that the effect has not been studied systematically; perhaps the distinction could be a starting point. It is naturally beyond this thesis’ scale and capacity to examine if the two types of entrainments have different evolutionary origins, but by considering exactly what activity is entraining to the rhythm, one can still have a simple model where there are two types of paths for the emotive effect of rhythmic entrainment.

4.2.2 Two Oscillator Types

Quoting Clayton and colleagues’ 2005 article, Juslin has defined rhythmic entrainment as the interaction between two “oscillators”, each with its own autonomous rhythmic process. As stated above, this section will explore the emotive effect of this mechanism from the perspectives of two different types of oscillators. One of them involves interoception, and the other movement.

The example Juslin has given in the article seems to be the first type. A brief description of the process has also been provided:

> […] the powerful, external rhythm of the music interacts with an internal body rhythm of the listener such as heart rate, such that the latter rhythm adjusts towards and eventually ‘locks in’ to a common periodicity. The adjusted heart rate may then spread to other components of emotion (e.g., feeling) through proprioceptive feedback, thereby producing increased arousal in the listener. (2011, p.622)

Again, like the “peripheral route” proposed by Zentner & Scherer (2001),

[72] Or convergent “development”, since the result is obviously too small-scale and short-term to be considered as anything close to evolution.

reviewed in the previous chapter in 3.1, this description is similar to the James–Lange style theory of how emotion occurs, as briefly discussed in the first chapter. Proposed by William James and Carl Lange, the hypothesis describes emotion as instigated by biological arousal. According to this theory, if a person sees a bear, their heart rate rises not because they are scared, but rather, their heart rate rises because it needs to propel more blood to the entire body in order to prepare for a fight or flight response required by the situation. The emotion of fear is what comes after such a biological reaction as a result of cognitive appraisal.

This hypothesis is rather evolutionarily oriented. Juslin's description, quoted above, while being less specific regarding the exact heart rate change and type of emotion, does have similar reasoning, especially because he has later quoted Clayton (2009), pointing out that entrainment is "particularly noticeable in activities where rhythmic coordination will make physical work more efficient" (2011, p.622).

The other path through which entraining leads to emotion is through movement. As I have proposed in the beginning of 4.2.1, the mechanism of rhythmic entrainment itself is still spontaneous, based on a process that even many non-human species have. However, after the movement occurs, the subsequent emotive effect takes place due to contextual, external knowledge. Take a march in western classical music for example: it would be near-universal that the salient and powerful rhythm would encourage the listener to produce synchronised bodily movement; however, what the movement reminds the listener of and what the following emotion would depends on one's contextual knowledge regarding that particular march. This set of context information again involves the telos, or purpose and function of a particular piece of music. It is somewhat related to what Schiavo (2014) has termed "teleomusicality", but here it is not only synchronically, but also diachronically telos-specific, as rhythm entrainment very often involves music with particular functionality, such as dance music for a situation where people actually dance. Consider also music accompanying agricultural activities, which may lose its original telos in a different setting of time and space. If an individual possesses necessary knowledge, however, relevant emotions might still arise. Moreover, the effect can also be observed when the purpose of music is simply for general pleasure, as even when listeners listen to music for enjoyment, if entrainment among individuals occur, it has the effect of strengthening social bonding, as Stupacher and colleagues have found in their 2020 study.

4.2.3 Intermodal Entrainment in EMT

The previous sections, in discussing the biological bases of mimetic mechanisms, have reviewed studies that play musical and rhythmic stimuli to listeners and detect activation in the motor region of the brain. These studies have generally defined "rhythmic stimuli" as beat patterns consisting of short audio events, such as woodblock sounds (Zatorre et al., 2008) or drumbeats (Bengtsson et al., 2009). Naturally, in EMT pieces with percussion instruments, such rhythmic elements are found in abundance and have great versatility, often serving as good examples of the entrainment effect. In Kagel's *Dressur*, the piece discussed in section 4.1, for instance, the castanets by percussionist III that takes the rhythm of horses' gallops would have entrainment effects in both routes proposed above- inherently, it is associated with human heartbeats, for the temporal distribution of audio signal in galloping is similar to that of heartbeats, consisting of regular interval of beats and silence. If anything, galloping beats in *Dressur* occur as rapid strings of triplets, more intensive than duplets of heartbeats, therefore potentially more intense. Externally, in the context of the stage's visual imagery, as analysed in 4.1, for those who are familiar with the activity, the beat pattern tends to be interpreted as horse-riding in a dressage arena, and the rhythm would evoke further entrainment and association with, especially for an audience with relevant experience.

It is noteworthy that visual elements play an important role in this example. In addition to the stage arrangement reminiscent of an arena, the galloping pattern recurs through the piece, sometimes played by the percussionist in a crawling posture, reaffirming the equine association. Such a multimodal source of information will likely strengthen the evocative quality of the visual rhythm, provided that, again, the recipient has relevant contextual knowledge that enables them to integrate the quadrupedal posture and sound pattern, associating these information with horse, horse-riding, circus, dressage, or other objects/events that would potentially induce response in emotion or memory, according to their individual experience.

Another example of such intermodality in entrainment can be seen in *Six Elegies Dancing*, a piece for solo marimba by Jennifer Stasack (1987). Corresponding to its title, this piece consists of six movements. It is worth pointing out that, rhythmically, each odd-numbered movement in this piece has phrases in rather irregular and complex rhythmic structures, while each even-numbered movement has phrases consisting of one series, or several groupings, of fast notes of equal length, resulting in at least one layer of periodicity, or tactus, as defined by London (2012). The constant, salient pulse in these movement forms a ground for entrainment of different depths (depending on the number of the periodicity level present). Such an overall arrangement somehow forms intervals of entrainment: a conspicuous rhythmic periodicity creating a certain entrainment effect occurs every other movement.

One difference of perspective to be noted is that, when investigating it as a mechanism that evokes emotion, Juslin has tended to view entrainment as the synchronisation of at least one concrete oscillator exhibited by the listener. On the other hand, London has applied a broader definition, seeing entrainment as "sometimes more and sometimes less than a phase locking of the listener's attentional rhythms with temporal regularities in the musical surface" (2012, p.13). This does not mean London's analysis is vague, however. The loose interpretation of "attentional rhythm" simply facilitates the discussion at the level of metrical framework, separating it from the more complex biological events such as oscillators and neural basis involved (which is indeed still discussed subsequently in chapter 3 of London's 2012 book). Following this definition, I will analyse entrainment mechanisms in Stasack's work broadly in terms of attentional reaction to tactus and other periodicity levels. For the discussion, I will be referring to the rendition of this piece by percussionist and educator, Tim Shuster, available online at https://www.youtube.com/watch?v=B_381RYn5AQ. Rough timecodes of each movement are listed in the table below.

movement and title	timecode (in Shuster's version)
i. adamantly, vigorously	0:00
ii. intensely	1:14
iii. gingerly, very stable	2:15
iv. furiously	4:10
v. with deliberate concentration	5:16
vi. elegiac	8:25

Table 4.1 movements in *Six Elegies Dancing*

As summarised above, movement 2, 4 and 6 have phrases with clear presence of "tactus". If one applies the framework proposed by Jones in her "temporal perspective model" (1987, 1990), it will be possible to further analyse the layers of temporal regularity below or above the pulse level. Based on the numbers and complexity of these levels, it is possible to observe that the entrainment in these movements occurs at different (and increasing, as to be shown) depths.

Movement 2, *intensely*, consists of unbroken streams of demisemiquavers that continue throughout the movement. The aptly titled melody that leaps back and forth covers almost the entire range of marimba, forming an intense contour. Rhythmically, apart from the equal lengths of demisemiquavers, there is no other layers of periodicity present, in that the melody consists of irregular numbers of tuplets leaping notes. The relentless pulse, or tactus, although simple in structure, would evoke basic temporal expectancy, resulting in entrainment of the lowest depth, as London (2012, p.16) has described.

The dramatic effect of the fast pulse is further enhanced by abruptly changing dynamics which keeps breaking music expectancy. This would probably result in brain reaction such as mismatch negativity discussed previously in 4.1.1[73]. The movement ends in an acceleration, before finally reaching an abrupt termination.

Movement 4, titled *furiously*, consists of phrases in fast semiquavers of either alternating dyads or groups of irregular notes. There is a pause before each occurrence of a phrase. Although the phrases are still in tactus of irregular lengths, the groupings of dyads consist of two fifths that alternate, forming a layer of binary subdivision below the pulse level. Thus, this movement has one more level of entrainment. Quoting Jones' temporal perspective model, London describes such a layer of subdivision as "the product of analytic attending, that is, we grasp them as fractions of a larger span" (2012, p.17).

It should also be pointed out that the constant alteration between rapid tactus and silence forms a meta, on-and-off rhythm of the entrainment evoked by the pulse in the repeating phrases, and such meta-rhythm in turns reflects the overall structure

[73] 4.1.1 has reviewed a study of rhesus monkey's mismatch negativity component activated by omission of notes, but this phenomenon is also commonly present in human, which may be aroused by the auditory stimuli's sudden change in frequency or intensity. For a detailed empirical study of this, see Näätänen et al., 1993.

of the piece, which alternates between movements in irregular rhythmic movement (where entrainment is less likely to happen) and those with regular tactus (where entrainment is more likely to happen). It has been pointed out that this piece draws upon the tradition of tai-chi, especially the gestures and motions of the percussionist in movement 5 (Peters, 2015; Kumor, 2016), which will be discussed shortly. But perhaps, in addition to lending some elements for the movements from the shadowboxing martial art per se, tai-chi also inspires the piece as a philosophy. The rhythm entrainment in movement 4 forms meta-rhythm by switching on and off at intra-movement level, and, in turn, the rhythm entrainment of the entire piece again forms a higher layer meta-rhythm by switching on and off at inter-movement level. Such an arrangement where on layer of on-and-off contains another is not dissimilar to the flux and circulation of Yin and Yang, which is the central philosophy for tai-chi and can be observed in the symbol. Although it is iconic and may be familiar to many, the symbol is quoted below.

Figure 4.3 A tai-chi symbol. Reproduced from Jaeger, 2012

Finally, movement 6, *elegiac*, framed by an intro and an outro, consists of a poignant tune and lower voice, played in tremolo on two mallets. Although soft mallets are suggested for both the melody and lower voice, which have pitches sustained for at least a certain duration, the nature of marimba's rolling notes still creates pulse by default, this time in a much darker tone compared to movement 2 and 4. In addition, this movement is in 4/4 metre, having both subdivision level below the tactus and measure level above the tactus. In Jones' model, the presence of these three layers would "provides an attending framework that allows the listener to track rapid, moderate, and relatively slow event onsets; these correspond to subdivisions of the tactus, the tactus level itself, and a higher-level ordering of beats into measures", where the tactus layer "anchors our attentional process and mediates between analytic attending (awareness of local details) and future-oriented attending (awareness of more global processes and goals)" (London, 2012, p.17). It should be pointed out that, although movement 2, 4, and 6 are in ascending order in terms of the number of rhythmic attention layers, it does not mean the former movements are less emotionally engaging. Movement 2, for instance, while having only one basic layer of pulse, constantly breaks musical expectancy by abrupt changes in dynamics, thereby interspersing the entrained attention with other reactions, such as mismatch negativity and brain stem reflex (to be discussed in the next section).

Above I briefly discussed the entrainment in movements 2, 4 and 6, and how the periodicity levels, and in turns, theoretical depths of entrainment, increase. which are central to the entrainment effect. Furthermore, another key aspect of the piece occurs at the end of movement 4 and transition into movement 5. This occurs at 5:12 in the link that has been referred to, when the percussionist plays one last group of semiquavers, but over, instead of on the marimba bar. The playing therefore has only movement, but not sound. Throughout the movement, each phrase is both an audio and visual event, but in this one last instance, it becomes purely visual.

As analysed above, phrases in this movement are groups of fast semiquavers that occur in a constant, straightforward on-and-off meta-rhythm. The clear rhythmic shape, both intra- and inter-phrase, could very likely establish an entrainment, and expectancy thereof, through the course of the movement. After the expectancy is formed, in the last phrase when it suddenly turns visual, it is likely that the listeners would still have analytic attending to the rhythm, due to evaluative conditioning developed so far.

Indeed, previous sections have discussed how entrainment is closely related to mimicking. It is then only natural that the inclusion of visual stimuli in EMT (or, in more general and conventional cases, in any live performance where the audience can see the musicians) would enhance the entrainment, because it offers more information needed for mimicking (either the eventual mimicking is overt or covert). It has also been pointed out that the biological basis of entrainment is largely autonomic. Thus, since it is still something that the mirror neuron could still automatically sense and react, it would be natural to expect that even purely visual stimuli could evoke some mimicking and in turns entrainment effect.

There is some relevant evidence in empirical studies. Synchronisation to visual rhythmic stimuli has been studied by scholars such as Jäncke et al. (2000) and Kawashima et al. (2000) [74]. Such visual synchronisation even tends to be more neurologically stable than audio ones, as shown in the 2007 study by Del Olmo et al., where participants are asked to synchronise with audio and visual rhythm. When the researchers try to disrupt synchronisation by applying repetitive transcranial magnetic stimulation to the participants' brains, more brain regions are activated to maintain the stability of audio synchronisation; however, the disruption does not seem to have

[74] To be expected, as Witt and colleagues point out in their 2008 review study, entrainment to audio and visual stimuli involve common brain area activation, but there also some difference. The basal ganglia mentioned above is mainly related to audio cues, whereas visual synchronisation tend to particularly activate DL-PFC (dorsolateral prefrontal cortex).

the same effect with visual synchronisation.

One thing to be noted is that these studies examine finger tapping, an oscillator that is activated by explicit instruction, unlike interoceptive ones such as heartrate that occur spontaneously. This is probably related to one fundamental difference between visual and audio perception: there is much less selectiveness in the latter, since it is much more difficult to choose to focus one's hearing on a specific source, and not possible to choose to stop hearing whenever one wishes. As a result, since audio stimuli are received more automatically, reaction thereto are also more spontaneous. However, exactly because of this difference, the last phrase of movement 4 becomes a key moment: as described above, it transfers the receptiveness of rhythm from the audio to the visual realm. Synchronisation would otherwise happen less spontaneously once the rhythm turns visual, but through the previous pattern in the movement which may have already established the association between phrases and rhythmic entrainment thereof, the evaluative conditioning would result in the last phrase' capability in still inducing entrainment, even if there is only visual stimuli, namely the movement of playing[75].

After such an attempt to introduce the entrainment in visual mode by an abrupt switch-off of the sound, Stasack proceeds to movement 5, titled *with deliberate concentration*, which incorporates an abundance of movements into the playing. It is likely that a large portion of audience' attention would now be directed to visual elements. Movement 5 can be divided into two sections, which is separated by a middle part where the percussionist turns back and plays as if on an imaginary, vertical marimba on a wall (from 6:44 to 7:04 in the cited video). The section before 6:44

[75] A similar silent phrase is again used in movement 6, but it is at the end of the piece and does not have the role of a transition. Therefore, it is probably more aesthetic/narrative than functional. The phrase also extends longer, with the percussionist rising their arm higher and higher as they continue the playing movement.

features sparse phrases, inserted with several expansive, slow movements of the percussionist's limbs, most likely inspired by movements in tai-chi shadowboxing, which is a practice said to have certain meditative and healing qualities. Although the scientific validity of such claims still appears to be a matter of debate[76], in the specific context of this performance, with the entrainment established in the previous movement and switching to visual elements at the end thereof, audience would be able to "lock in" into not only what they hear but also what they see, and hence covertly synchronise with the bodily experience of the movements, potentially with mimetic response and subsequent effect, either mental or physiological, based on their individual reaction. The section after 7:04 does not have more slow tai-chi movements, but continues to feature air marimba playing, which enables the percussionist to smoothly incorporate the movements of first leaving two of the mallets and then eventually changing to another softer pair, allowing an attacca transition into the next movement and its mournful emotive climax.

Taking into consideration the presence of tactus in even number movements, the increasing number of periodicity levels throughout movement 2, 4, and 6, and how movements (both the overt movements of the musician, and the covert movements evoked in audiences' mimetic reaction) creates visual rhythm, one can then realise the structure of the piece neatly elucidates the piece' title *Six Elegies Dancing*. The odd number movements have titles indicating restraint and repression (*adamantly*,

[76] Studies of health benefits of the practice have indeed been conducted. Li et al. (2012) points out the sport "appears to be effective as a stand-alone behavioral intervention designed to improve postural stability and functional ability in people with Parkinson's disease" (p.519). However, previous review analyses in larger scale, such as Wang et al. (2004), seems to find that most studies generally have room for improvement in subject selection, statistical validity and long-term follow-up. They believe that the sport "appears to have physiological and psychosocial benefits and also appears to be safe and effective in promoting balance control, flexibility, and cardiovascular fitness in older patients with chronic conditions. However, limitations or biases exist in most studies, and it is difficult to draw firm conclusions about the benefits reported. Most indications in which Tai Chi was applied lack a theoretical foundation concerning the mechanism of benefit." (p.493)

gingerly, *deliberate concentration*) whereas the even number movements have titles denoting strong emotion (*intensely*, *furiously*, *elegiac*). It is in movement 6 that the piece eventually reaches its "titular movement", with a sorrowful tune titled *elegiac*. This process illustrates the turmoil state of mourning, where one struggles between constraining and releasing one's grief, and the emotions in even number movements, *intensely*, *furiously*, and *elegiac* also somehow corresponds to the phases of denial, anger and depression, in Kübler-Ross model of stages of grief (Kübler-Ross and Kessler, 2005). Pulse, and subsequent entrainment effect, accompanies every movement where an emotion is present. As a result, the piece can be "dancing" and also "a dance", both because of the alteration between stillness (control, absence of pulse) and movement (emotion, presence of pulse), and because of the visual elements in the later movements.

Overall, both *Dressur* and *Six Elegies Dancing* can have instances of different types of entrainment effects discussed in the previous sections, and both are good examples of how the inclusion of visual elements such as stage design and movements in EMT enhances the diversity and strength that rhythmic entrainment could have in music. The inclusion of visual cues might also reflect a holistic version of musical experience, closer to that of some forms of music in its contextual environments, as briefly raised in 4.2.2.

4.2.4 Composition Output

I have attempted to explore entrainment in the fourth movement of *The Carnival of the Endangered Animals*, which features the sound of wild Bactrian camel (*Camelus ferus*). As has been discussed, the mechanism is a biologically prominent, and rather universal process. It is mostly manifested as the straightforward phenomenon of synchronisation in oscillators. The subsequent emotive response should be more variable, but again, it is individual-dependent. In this movement, on the other hand, I seek to evoke specific emotion through the structural and symbolic designs related to rhythmic entrainment.

The piece consists of a consistent rhythm pattern in 14 (5+4+3+2). The pattern persists approximately from 00:00 to 00:50, building up as more forces join: it starts on the piano, and subsequently is joined by pitched percussions (crotales and xylophone), and then woodwinds (alto flute and contrabassoon). The camel sound does not appear until after 00:52, when one beat in the rhythm pattern is replaced by a knock on the panel or key block of the piano.

This design is inspired by the materials used in the experiment for detecting mismatch negativity component. I attempt to establish a regular tactus first, which is intended to cause entrainment, and subsequently get the listeners' attention through omitting certain beats in the pattern, thereby then highlighting the entrance of the camel sound. The sound plays twice (00:52 to 01:02), each time introduced by a knock that replaces the notes in the rhythmic pattern. The piece then enters a slow middle section (01:03 to 03:05). The same rhythm continues in half speed, and the pattern also continues where each sound of the camel is preceded by a missing beat on the instruments, replaced by a hit using the mallet sticks in this section. From the end of this section, the camel sound is changed to a second version which has a long, grunty tail[77]. It repeats several times, and in the end of the last occurrence, the piano enters again (03:22), playing the pattern in the first section, but this time synchronised to the rhythm of the grunting sound of the camel[78], which is slightly faster. The rest of the ensemble joins from 03:36. The piece continues into the last section, which still has

[77] The recording of this particular sound has some ambience sound mixed in it, so an extra layer of ambience sound taken from the materials in this recording is played from here.
[78] This can be assisted with timecode activated by the controller playing animal sounds.

beats replaced by the knock on the piano, but the camel sound does not appear following that anymore. Finally, the sound occurs at 04:04, when the piece ends in the upward scale on the piano and crescendo in other parts. The tail of the camel sound continues after all the instrument finish, grunting alone in the background of ambience sound.

Although wild Bactrian camel (*Camelus ferus*) used to be considered the same species with the domesticated Bactrian camels (*Camelus bactrianus*), studies by scholars such as Mohandesan and colleagues (2017) have confirmed that they are actually different species, and the former is, indeed, critically endangered. In this piece, I cite the sound of the *Camelus ferus* as a symbol of the ghost of the disappearing animal. The preceding rhythmic pattern's missing beats highlight the camel, both perceptually and symbolically: Perceptually, this would activate electrical activity in the brain, as shown by the experiments reviewed previously. Symbolically, they represent the voice of a distant, feral kind from the past which has almost gone missing in the current scene. Hence, it's audially associated with the missing element in the structure of the tactus. After the instruments, which again represent the artificial and human force, synchronise with the camel, it disappears in the subsequent music, symbolising the domestication process which did not keep the original species. The last occurrence of the camel in the end is a call from the distant wilderness and memory.

Again, while it is perhaps inevitable that I process, or interfere, the sound material in some ways in order to construct a piece of composition, I attempt to acknowledge and attend to, the musicality that is inherently in the sound, adapting the rhythmic element therein as a central feature in the conclusive section. Moreover, while the symbolic intention of the piece may not be instantly delivered if the listeners do not have the background information, certain effects can hopefully still be achieved: The association between the missing beats and the sound would still form, on the basis of the highlighting effect triggered by the missing element in an entrained periodicity, which will attract the listeners' attention. Also, in the last section, the established association between the missing beats and the camel is broken, strengthening the impression of the animal's disappearance.

4.3 Brain stem Reflex

This chapter has four sections. The first two discuss the two basic acoustic properties that trigger the brain stem reflex: 4.3.1 looks at how the factors of volume and context trigger the mechanism, and 4.3.2 discusses the structure and effects of dissonance. The later 2 sections explore the application of the mechanism: 4.3.3 analyses the effects of dynamics in *Sapindales* by Martin Iddon, while 4.3.4 discusses how I attempt to activate the mechanism with multiphonics in in the *Elegy* movement in my piece *Remember How Apples Smell?*

4.3.1 Jump Scare, Loudness War, and Lofi Hip Hop

In the BRECVEMA model, Juslin has defined the brain stem reflex as the reaction to certain characteristics of audio stimuli which are interpreted as related to "a potentially important and urgent event", usually induced by "sounds that are sudden, loud, dissonant, or that feature fast temporal patterns" (2011, p.621), and also "high sound level, quick attack, and sharp timbre" (2015, p.19). To demonstrate this, Juslin conducted a study with colleagues where they play excerpts from pieces with such features[79], and found statistically significant reaction in both skin conductance and facial expression of participants (2015, pp.30-31).

[79] such as Brahms' Symphony No.2, 4th movement where the full orchestra plays suddenly and loudly after a quiet strings section; Mahler's Symphony No.10, 1st movement, where the recapitulation of the theme peaks in a loud, dissonant chord; and, of course, the famous Haydn's Symphony No. 94, 2nd movement with its tutti "surprise" section.

Juslin as argued that this mechanism is rather spontaneous and universal, "reflect[ing] primarily 'hard-wired' reactions to simple features that are not affected much by learning" (2011, p.625). It should be noted that, however, in the 2015 study, it was only western classical music that was played (as described in footnote 26 below, plus an early Stravinsky piece *Firebird*)[80]. As quoted above, Juslin's description of the audio stimuli evoking brain stem reflex is "sudden, loud, dissonant" and with "fast temporal pattern". These are information involving the context, amplitude, harmony, and rhythm of the audio signal in question. While amplitude and rhythm are relatively more straightforward to quantify and define, harmony is less so, and context could apparently interact with all other three aspects.

The effects related to this mechanism is nonetheless a pervasive phenomenon, relatable to various aspects of human's bodily experience. They serve as some of the most prominent and straightforward examples of how musical emotion is associated with embodied cognition. It would be worth discussing what exactly is triggering the reaction through this mechanism, and the nature/nurture elements involved.

First of all, it is probably very common for most humans, and also some animals, to react to a loud audio stimulus that occurs suddenly. Quoting previous studies by Halpern et al. (1986), Foss et al. (1989), and Burt et al. (1995), Juslin and Västfjäll have pointed out in their 2008 article that emotion induced through this mechanism is primarily unpleasant. This is the basis of reactions such as "jump scare", in commercial horror films. As Koelsch & Siebel have pointed out in their 2005 study, such a startle response is a very early procedure in audio perception, so it would quickly activate at the onset of the stimuli. Neurologically, this mechanism is associated with the brain stem's ascending reticular activating system (ARAS) (Joseph, 2000), a formation

[80] The participants were also all students at Uppsala University, but their backgrounds were not mentioned.

consisting of several nuclei and related to vital physiological functions such as alternating between sleep and wakefulness, and entering into the state of high attention.

Based on this property of transition, it can probably be inferred that audio context plays a decisive role in this mechanism. Indeed, while sudden loud audio stimuli can almost universally evoke a strong, instant reaction, should the exposure to loudness continues, the sensitivity to high volume tend to lower gradually. This desensitisation process, known as *temporary threshold shift* or *loudness fatigue*, can be observed when individuals are exposed to either loud music (Sadhra et al., 2002) or high-level impulse noise, such as gunshots (Olszewski et al., 2007), for an extended period of time. The same environment on a smaller scale, namely constantly high loudness within the digital recording of one piece of music, usually resulted from certain post-production practices, leads to a similar process of desensitisation which deprives the piece of its dynamic range, de-emphasising the loudness of the loud parts with overall volume level. Around the last decade, the world of audio engineering started to disfavour such practices and the resulting loudness fatigue, criticising it as "loudness war". A detailed analysis of the background and influences of this can be seen in a 2010 article by Vicker. Quoting other sources such as Levine (2007) and Lawson (2008), Vicker has commented on "loss of excitement and emotion" (p.6) resulting from the loudness war:

> One of the main complaints about hypercompression is that it flattens the dramatic and emotional impact of the music. Levitin stated that "The excitement in music comes from variation in rhythm, timbre, pitch and loudness. If you hold one of those constant, it can seem monotonous." [11]
>
> Lawson used Róisín Murphy's song "Overpowered" as an example: "...as cleverly assembled as 'Overpowered' is, the use of dynamic range compression arguably

> limits its potential for pleasure. The recording is not as aggressively mastered as some, but when the song sounds like it is 'supposed' to hit a peak in volume – 'As science struggles...' – there is no actual increase... In order to be beguiled the listener has to imagine that the volume has increased...." [25] (p.6)

Hence, even though objective measurements of sound pressure levels exist, when discussing the dynamics in terms of the emotive effects it causes, it is crucial to consider the psychological perception of loudness, which is less definite and more context dependent. From long-term daily environments to short-term duration of a song, what is considered loud in any given moment is to be determined by whether it stands out in the dynamic background, which is also the decisive factor as to whether brain stem reflex will be activated.

The extreme cases of such emotive response would be reactions such as "jump scare" as discussed above. Also, there are more subtle instances of such effects to a lesser extent with music with certain particular functions. In my experience, annoyance could be caused by too many occurrences of loud elements, when the music is designed to be a background to other activities which requires concentration, such as important monologue in a theatre piece or difficult tasks in video gaming. In such circumstances, listeners simply do not want to be distracted by the firings of reflexes to sudden strong audio stimuli, which happen involuntarily.

The function and practicality of music is another important perspective in the debate of "loudness war". Recent audio playback devices are advanced enough to represent high definition of audio signal, even quiet parts, with great clarity, which is one of the reasons many audio engineers believe it is time to place more emphasis on preserving the dynamic range of music. However, the mobility and ubiquity of such devices also result in new purpose of music. Because music can be conveniently played from very portable devices, some listeners use them as audio background for general activities such as reading or social gathering. In such cases, high dynamic range becomes not necessarily preferable[81]. A very good example would be Lofi Hip Hop, famously dubbed "beats to relax/study to". Such productions usually apply heavy dynamic range compression. Further discussions of ideal dynamic range in different listening situations and different genres can be seen in Vicker's 2010 article.

Overall, a stimulus needs to be both loud and sudden, as defined by the context, in order to activate a listener's brain stem reflex. The reaction is so fast that it is experienced as instant. Although the mechanism is rather primary and universal, there are different emotion results associated with it, depending on various factors such as audio context, and particular function and aesthetics of the music. It also plays a great part in many aspects of music and music production, especially in today's background, where the technology supports more and more different forms of music listening in different acoustic environments.

4.3.2 Consonance/Dissonance

While it is easily agreed that sudden, loud audio stimuli would startle listeners, harmony, on the other hand, requires more explanation. Current evidence seems to point out that the mechanism of harmony perception has both inherent and acquired properties. Relevant discussions usually apply a general distinction between sensory consonance/dissonance and musical consonance/dissonance, as put forward by Terhardt (1984), where the former is independent of learning or cultural experience,

[81] It is perhaps worth pointing out that in the case of western classical music, dynamic range seems to be generally much bigger than many other modern genres. A production of nocturne by Chopin usually has far greater dynamic range than, for instance, a new age piano piece. This is especially true for orchestral music with larger instrumentation which could feature very loud and quiet parts. For a new a piece of modern orchestral music, before recording and post production, the overall dynamic shape is usually one of the first things an audio engineer would confirm with the production team. Sometimes an engineer would phrase the question conveniently as something like "are we intending to have a more classical-like or modern dynamic range?". Therefore, what is described here may be less applicable to classical music pieces.

while the latter involves acquired features and depends on the listener's musical background.

Sensory analysis points out that the physical basis of the human perception of a dyad as consonant or dissonant is the ratio of frequency. There appears to be a cross-cultural phenomenon to perceive two notes of frequency ratio 2:1, namely an octave, as "alike". Also, intervals with relatively simpler frequency ratio are perceived as more consonant than those with complex frequency ratios (such as 5:4 in a major 3rd and 15:8 in a major 7th). The underlying property under these ratios, as von Helmholtz pointed out in 1863, is the presence of "beats" in dissonant intervals. Originally termed *Schlag* by von Helmholtz, beats refers to the fluctuations in level when two pure sounds close in pitch are played simultaneously. In intervals with simple frequency ratios, due to the coinciding phasing of harmonics, no beat would occur. Intervals with more complex ratios, on the other hand, are more likely to produce beats due to the harmonics being closer to each other in frequency.

At the neurobiological level, when such beats occur, the neural populations in auditory cortex subsequently response by firing in synchronisation with the periodicity of the beats' temporal modulation in amplitude. This causes a sense of "roughness" and result in the interval eventually being perceived as dissonant. The neurological was first pointed out by Meyer in 1898, and later studies have reported further empirical evidence, in both human and other species. For instance, Zentner & Kagan (1998) have found that four-month-old infants tend to stare longer at an audio source when it plays a melody that contains parallel major or minor thirds (according to tonic scale) while exhibiting fret (moving and turning away) when the audio source plays melody in parallel minor seconds. The biological preparedness to not prefer sensory dissonance seem to be present before any learning experience. A later study by Trainor and colleagues (2002) has used only interval on piano as stimuli and has had similar results. Schön and colleagues (2005) conducted an ERP (event-related brain potentials) study on adults using stimuli in categories entirely consistent with the perfect consonance/imperfect consonance/dissonance groups in functional harmony theory in western classical music and have also found result in support of Helmholtz' theory. Also, in their 2001 study, Fishman and colleagues have found that the neuronal ensembles react to dissonant audio stimulus by firing synchronously with the "beats", and this is present in both human and crab-eating macaque (*Macaca fascicularis*), a species of monkey.

On the other hand, behavioural studies on animals also report that other species tend to discriminate between consonant and dissonant sounds. Hulse and colleague (1995), for instance, have pointed out European starlings (*Sturnus vulgaris*) are able to discriminate different chords, and believe they do this by recognising the consonance and dissonance. Also, Watanabe and colleagues (2005) have found reinforcement of perching behaviour is present when Java sparrows (*Lonchura oryzivora*) are presented with consonant stimuli, but not dissonant ones.

As stated in the beginning of this section, these properties are believed to be universal. The topic of sensory dissonance is also what Juslin focuses on when discussing the mechanism of brain stem reflex, which he considers as hard-wired in humans' inherent biology. Outside the range of sensory theory, however, when one examines musical emotions in the bigger scale as a holistic experience felt during a piece of music, there are phenomena that cannot be entirely explained by such theories. To give a few examples, firstly, when a dissonant interval consists of two sequential notes rather than a dyad, supposedly no beats would occur, because the sounds do not occur simultaneously and therefore their harmonic series would not produce the modulation in amplitude. However, studies find such sequences are still judged as dissonant by listeners, infants (Schellenberg & Trehub, 1996) or adults

(Schön et al., 2005)[82] alike. Also, a sensory dissonance could be perceived as musical consonance, by a specific individual who is accustomed to a specific harmonic language. A chord containing augmented 4th could sound "right", given that it is resolved in a particular musical context, for a listener who is familiar with that context. Mechanisms like this is clearly based on the sensory dissonance/consonance (tension/resolution) of individual components, such as the hormonic structure of each chord, but whether if the resulting holistic effect is universal or learned also remains undetermined.

An evolutionary explanation for the consonance/dissonance phenomenon, as suggested by Terhardt (1984), holds that tonal perception is based on being exposed to the vocalisation of animals, particularly humans, of course. These vocalisations are one of the core elements of communication, which is essential for survival. They are therefore pertinent to every individual. One defining characteristic of vocalisation is that they are based on harmonic series, and therefore have tonal quality. The concept behind this argument is that humans understand music, because music is essentially an artifact made as an extension of vocalisation and expression, so it is only natural that emotions are induced by elements such as harmonics and frequency ratios, which are the very mechanism on which vocal cords operate, rather than mere physical properties. Based on the same principle, many basic elements in music and our embodied knowledge of vocalisation are in common. The roughness of sensory dissonance, for instance, is actually present in many animal calls signalling danger (Ploog, 1992). Therefore, negative emotion aroused by similar sounds can be considered as a natural, evolutionary legacy. This perspective helps connecting sensory

[82] Not all these studies view the phenomenon as contradicting to the beats theory. For example, Schön focuses on the fact that melodic intervals are conceived as "less" dissonant than dyads, and think it supports von Helmholtz' theory because non simultaneity reduces roughness. The fact that the roughness is still there, however, could perhaps be something that remains to be explained.

and musical consonance/dissonance, in that we can consider the latter as an extension of the former, and just like some emotive responses are acquired through evolution, there are also responses that can be acquired through accumulated experience of exposure in specific musical contexts. This could offer a potential possibility to see evolution result as nature in small scale, but nurture in bigger scale.

To sum up, similar to what I have attempted to point out in the discussion in rhythmic entrainment, further reviewing the evidence from empirical studies can sometimes refine the categories defined in BRECVEMA model, providing insights in analysing and applying the effects in musical examples. For instance, a detailed description of brain stem reflex would be the brain stem's ARAS is triggered by a loud and sudden sound, and this would occur regardless of the consonance or dissonance contained in the sound. As discussed above, the "jump scare" effect activates the ARAS very quickly. According to Moore (1973), short audio stimuli would not be perceived as tonal, but rather sounds like a click. The quality of "tonality"[83] only emerges as the sound becomes longer. It can therefore be assumed that, since the reflex reacts instantly to the attack of a sound, the pitch, and resulting harmony, would not be the primary factor in the process. If a sudden, dissonant stimulus somehow gives a more unpleasant effect than a sudden, consonant stimulus[84], it would be because afterwards when the pitch and interval is perceived, the roughness in beats locks in with listeners' neural activity in auditory cortex.

[83] Here this means perceiving frequency as pitch, instead of harmonic tonality in western classical musical theory.
[84] This can only be guessed, because it would probably be unethical to test.

4.3.3 Effects of Dynamics in EMT

As the discussion has shown, the activation of brain stem reflex is rather instantaneous and straightforward. On the other hand, in composition, through intentionally designing the features of the soundscape, effects can be achieved that limit, suspend, or sensitise the reaction to the dynamics in the music.

Sapindales, a 2020 piece written for bass clarinet and fixed media by Martin Iddon, was included in the music composition collection of the research network Somatic Music, convened by Björn Heile, of which I had the pleasure of personally being a part. The piece, along with those by other members in the network, was performed as a part of the Sound Festival 2021 in Aberdeen, which I also attended[85]. I have chosen to analyse this piece not only because that it is an ideal example of the mechanism, but also because I had the pleasure of being involved in the rehearsal, performance, and discussion of the piece. As a result, I was able to gain more insights into the piece than I could have with other pieces. This makes the piece an optimal example to be discussed in this section.

A performance of the piece can be found approximately from 48:55 to 1:10:00 at https://www.youtube.com/watch?v=9VAK2H54ptw, which is the recording of the Sound Festival 2021 concert that featured the compositions from the research network[86]. It is performed by Heather Roche, the clarinettist for whom the piece is composed.

The music consists of a field recording by Iddon, taken from Moorlands Nature

---

[85] My EMT piece *Three Singers on Planet M*, discussed in 4.1.4, was also a part of the research network composition and premiered in this event.

[86] Alternatively https://www.youtube.com/watch?v=9VAK2H54ptw&t=2935s, which includes the timestamp that directs viewers to the beginning of Iddon's piece. *Three Singers on Planet M* is also included in this recording, starting from https://www.youtube.com/watch?v=9VAK2H54ptw&t=2120s

Reserve in Yorkshire, and a pre-recorded track of clarinet, and the live bass clarinet. The piece beings with Roche clicking a button on the laptop, starting the playback of field recording. Subsequently, the recorded and live clarinets join, playing mostly long, slow glissandi against the background of environmental sound in the fixed media. Throughout the piece the stage remains dark, and the musician still, creating a contemplative and mysterious visual, which obscures the image of the source that is producing the sound, further shrouding the duo between live and recorded clarinets with vague illusions of audio-visual decoupling[87]. In such a meditative mood, through mechanisms that can be described as "condensation", the dynamics of the piece achieve subtle and sophisticated effects that further contribute to the complexity and symbolism of the piece.

'Condensation' is used as an analogy from both psychoanalysis and physics here. In Freudian psychology, condensation refers to the phenomenon where several separate concepts or identities are represented by one single object in dream (Freud, 1974). In Sapindales, two disparate properties are condensed in the clarinet when one considers its relationship with the fixed media recording:

The elements in the field recording can be categorised into two groups: there is an ambient sound which is constantly present, and there are different kinds of birdsongs and low frequency rumblings which occur more irregularly. The constant ambience ends up serving as a background, whereas the irregular birdsongs and rumblings, being sporadic and varying, cause sensitisation and stronger response from the listeners (Groves & Thompson, 1970), thus become more noticeable. This basic background-foreground structure defines shape of the soundscape throughout.

[87] For further analyses on the audio-visual, doubleness, situatedness, and other themes of the piece, the Somatic Music research network will publish an article commentating this piece, written by several network members other than Iddon. My discussion on the piece from hereon also included texts that will likely be modified and included in the unpublished article.

When the clarinet enters, against the ambient sound, it serves the role of an irregular occurrence, highlighting its presence to the listeners. It possesses a sensitising property in this regard, disrupting the otherwise unvarying ambience. However, in relation to the different types of birdsongs and rumblings which are more scattered, the clarinet sound appears relatively more unmarked and almost becomes a part of the ambience elements. Thus, two dissimilar, almost conflicting properties are condensed in the clarinet sound. This is a duality that emerges from the interplay between the elements with different degrees of regularity, and, since it has the double identity of highlight and backdrop, causing opposite perceptual effects, creates a slight sense of dispute and complexity against the otherwise meditative, unified soundscape.

Inevitably, this structure reminds one of Schafer's definition of "hi-fi environment", where "there is perspective-foreground and background" and "even the slightest disturbance can communicate vital or interesting information" (Schafer, 1993, p.43), although, obviously, the structure here has three elements, with the clarinet as the middle layer, serving as both foreground and background. Such a doubleness results in an unsettling, ever-changing shift of focus between the woody, chanting sound of clarinet that sings in the air, and the twittering birds and rumbling machines that echo among the woods. The subtle yet complex effect is achieved through setting off different layers of reflex to occurrences of stimuli, among the three elements.

It is also noteworthy that, for the listeners who are at the live performance of this piece or watching the video recording of it, the only audio element that is visually represented is the clarinet sound, associated with the human musician on stage. Such an embodiment would mostly ensure that a listener's attention is directed to the clarinet, thus being drawn into the in-betweenness where foreground and background, song and ambience are condensed into one entity.

As a physics term, condensation refers to the change from gas state to liquid state. Essentially, what is changed from gas to liquid is that the range of particle activity is more limited and confined. It is again an analogy that can describe the clarinet in Sapindales, which has no explicit dynamics instructions on the score. The absence of dynamics marking, however, does not imply the absence of dynamics changes. The fluctuations in the volume naturally resulting from the flow of attack, decay, sustain and release is still a salient feature of the clarinet sound. Hence, this is analogous to the particle activity of liquid state, which, although moving in a fluid fashion, is confined in a limited range, compared to that of gas state (which would be the sound with explicit instruction in dynamics and dynamics change by this analogy).

This quality of stillness, yet definitely not lifelessness, corresponds perfectly to the ambience of the environment, which is also still, and meanwhile full of sounds of birds chirping, leaves rustling, and even distant traffic rumbling. Such a correspondence contributes greatly to the aesthetics of Sapindales, and also explains how the clarinet sound blends well into the soundscapes. With the dynamics activity condensed in the range of quiet and natural breathing, the clarinet is almost like a lifeform which is well-integrated as part of the environment. Indeed, it has been observed that animal sounds in a natural environment all fit into unique spectral ranges and rhythmic cycles. Termed by Krause as "niche hypothesis", the theory holds that an equilibrium is achieved when all the creatures vocalise in the same environment, but they can all be heard, because they are situated in such respective niches. Krause has described that:

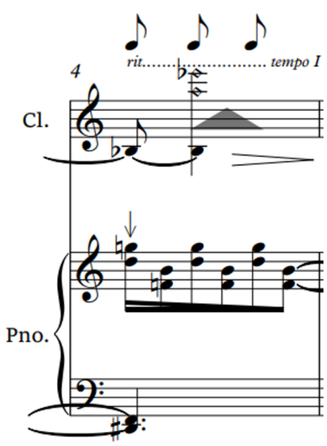
> each creature appears to have its own sonic niche (channel, or space) in the frequency spectrum and/or time slot occupied by no other at that particular moment. Taking a giant leap when considering the habitat as a whole, the sounds of each of these zones are so unique and important to creature life in a given

> location, if one creature stops vocalizing, another immediately joins the chorus to keep that audio bio-spectrum intact. An audio bio-spectrum is an acoustical spectrographic mapping of any particular habitat by frequency (pitch, sometimes tone) and amplitude (loudness) over short periods of time. (1993, pp. 2-3).

Not surprisingly, this is important for species to sustain, as vocalisation usually has important survival functions. Indeed, the accommodation of calls across respective spectral and rhythmic ranges is observed to be necessary for biodiversity to sustain, as habitats with greater biodiversity usually exhibit higher "acoustic space use", namely the ratio of time and frequency ranges occupied by particular vocalisations. This has been observed in various types of habitats, from forests (Aide et al., 2017) to savannas (Ferreira et al., 2018). Such a structure can only be maintained, obviously, when no sound in the environment is overwhelmingly loud, because a signal with overpowering intensity would likely mask the others, even if they occupy different bands of frequencies. The clarinet in Sapindales, while having its natural dynamic expression, is never played with excessive volume that would drown the birdsongs. It also mostly occupies a spectral range that does not collide with those of the birdsongs, achieving an ideal fashion of coexisting with minimum disruption. Such a symbolism is also manifested in the perception of the listeners, as the music unfolds in the absence of intense dynamics peaks which would activate strong reactions of brain stem reflex. It is as if the listeners were transitioned to the location where the soundscape occurs and invited to experience the audio environment which would allow the "virtual symphony of animals", as Krause has put it, to exist.

4.3.4 Composition Output

In addition to dynamics, the mechanisms of harmony in brain stem reflex and its related theory can also be applied in developing methodologies for achieving more complex effects in EMT. For instance, in the *Elegy* movement of *Remember How Apples Smell*, which is the piece from my composition portfolio that has been discussed previously in 4.1.4, I attempt to manipulate the effect of brain stem reflex by using multiphonics on B flat clarinet. The multiphonics have sharp timbre, and I also choose mostly intervals that contain inherent dissonance (see the example below). The resulting sound, shrilling and dissonant, should be particularly aversive, due to its high concentration of energy around the range between 2.4 and 5.5 kHz, which should cause intense reaction in the amygdala, and in turn auditory cortex (Kumar et al., 2008; Kumar et al., 2012). For instance, figure 4.4 shows the energy distribution on the spectrum of one multiphonic example used in bar 18 of the movement, quoted on the top left of the picture.



Example 4.1 A multiphonic note in the *Elegy* movement of *Remember How Apples Smell*, consisting of minor second and augmented fourth.

Figure 4.4 the concentration in the range from 2.4 and 5.5 kHz, captured around the peak of the crescendo.

The multiphonic notes played in this piece, however, have qualities contradicting the effects of brain stem reflex, or strong stimulation in amygdala. As discussed in previous sections, Juslin has pointed out that one feature of the stimuli that cause brain stem reflex is that they must be sudden. The studies in neuroscience have shown similar observation, which is concluded in a more quantitative fashion: Kumar and colleagues have shown that the sound has to have a particular range of temporal modulation frequency, namely the fluctuation in intensity over time, to be detected as unpleasant (2008, p.3813). Indeed, during the fMRI measurement, “[n]o response was observed in the amygdala that correlated with either spectral frequency or temporal modulation frequency alone” but “[t]here was a significant correlation of BOLD[88]

[88] Refers to Blood Oxygenation Level Dependent signal detected during fMRI scan.

response in the amygdala with the interaction between the spectral frequency and the temporal modulation frequency" (2012, p.14186).

While it would be difficult to instruct the exact temporal modulation frequency of the sound to be performed on the clarinet[89], this movement has multiphonic notes around the 2.4 to 5.5 kHz range, which are played in gradual crescendo and diminuendo in different lengths (indicated by the triangle graphs in the score). With the notes with screeching frequency but no sudden intensity, I attempt to achieve a creeping effect which suspends the brain stem reflex until enough intensity is reached in a short enough duration. The purpose of creating such an element that constantly hovers around the edge of harshness fits in the structure of this piece. As described in 4.1.4, this piece employs an apple as the vehicle that carries a sign of transition, which will have been established through the previous movements based on pastiches. When the *Elegy* movement is reached, the listeners would hopefully have some impression and interpretation of the apple. The clarinet multiphonics that constantly lingers between mildness and sharpness will serve as an ambiguous symbol that can potentially be comprehended, or even perceived differently depending on the absence and entrance of the apple. It thus serves the purpose of creating stimuli that tend to be more neutral, here due to their inherent audio property, in the later movements.

Overall, in this example, the theory has helped to accurately shape the audio element needed for the music. The interval, pitch, and dynamics shape of the multiphonics are chosen based on the results of the specific studies reviewed. Meanwhile, the composition output also enriches the theory by extending a straightforward effect into something more subtle and flexible through the use of particular techniques, and also by implementing the mechanism in a setting where it

[89] In Kumar et al. (2008), the sound is perceived as the most unpleasant when the modulation frequency is 1 to 8 Hz.

is involved with other mechanisms, providing a concrete example of interaction. This demonstrates a scenario where the theoretical framework informs the piece of composition, and in order to effectively apply the theory to achieve the precise outcome, technical details in relevant research are also taken into consideration. The result again shows a dynamic relationship between theoretical exploration and practical output, in that separate mechanisms and elements are combined together and holistically deliver the effect. As pointed out in 4.1.4, this piece is conceived from a play's single line, which is used as the piece title. In the scene, the characters are moving in circle and speaking in rotation, while the line suddenly leads to a crucial moment of silence and motionlessness, where all characters stop and sniff in the air in unison. This is an attempt to summon an olfactory sensation that is not actually present by invoking the illusion thereof, probably appealing to the mimetic reaction of the audience. Similarly, this piece attempts to achieve the effect of bestowing a symbolic meaning upon the apple, and it does so by means of conditioning and expectancy. The vehicle to contain it is established on the clarinet multiphonics, which has acoustic ambiguity, namely. This shows that in practice, a particular property described in theory can be realised in a form that is multifunctional and less straightforward.

5. The Remaining Mechanisms and Conclusion

In this thesis I have attempted to incorporate some of the fundamental biological bases, relevant empirical studies, and extended philosophical discussions of 4 of the 7 mechanisms in Juslin BRECVEMA model on how emotions are induced through music. I have also analysed how these mechanisms work in the context of experimental music theatre, which is a unique setting consisting of non-hierarchical elements that allow various intermodal embodied cognition activities to accommodate the mechanisms in particularly flexible and effective ways. The discussion on each mechanism has included a section in the end that talks about how I attempt to incorporate the mechanism and related effects in my own EMT compositions.

As pointed out before, the remaining 3 mechanisms, namely evaluative conditioning, episodic memory, and musical expectancy, are more unpredictable in terms of individual listeners. It would be less easy to have a generalised discussion on, for instance, with what situation or memory the listeners would associate a particular musical element. However, they can actually be explored from the angle of particular groups of musicians, and particular groups of listeners' general stylistic recognition, and offer insights on the operation of embodied cognition in music. Therefore, before I proceed to conclude this thesis, it would be helpful to discuss the following perspectives:

5.1 when such mechanisms work on musicians (evaluative conditioning, episodic memory): as pointed out in 2.1.2, these mechanisms by all means are also obvious in listeners, but they are very individual dependent. For a general discussion, however, some performance traditions have principles that prescribe systematic structures dictating when musicians should perform certain music, and they are connected with these mechanisms, as the discussion will show.

5.2 when such mechanisms do not work on the listeners, because they are "overridden" by reactions occurring at another level (musical expectancy): here I will also extend the discussion of the "sense-making" and "telos" of music and show that the embodied and social information which one possesses will all contribute to musical expectancy, and in turn affect the emotive responses one has. In fact, Juslin has added "aesthetic judgement" in his 2013 work, which is the eighth mechanism in the BRECVEMA model. He argues that this mechanism "rel[ies] more on 'higher' cognitive function, domain-relevant knowledge, and a fluid, individualized process that may change across time and context" (2013, p.247) and that the process of the mechanism "begins with an initial classification of the music as 'art', which will lead the listener to adopt an aesthetic attitude" (same as above). Here I would argue, additionally, that such processes based on a so-called "higher" cognitive function do not only involve an "aesthetic" attitude, but rather could be activated by any socially or culturally relevant contextual information and lead to emergence of meaning and enactment of a specific piece of music, thereby overriding other mechanisms such as basic brain-stem reflex or rhythm entrainment. This could be activated even without the music being regarded as "art". I categorise this interaction broadly as a kind of musical expectancy, and will discuss it in this section.

5.3 Conclusive remark of this thesis.

The topics in 5.1 and 5.2 obviously all require extensive discussion to be fully covered, but due to practical reasons for this thesis, I will briefly explore them, before proceeding to the final section. I will attempt to conclude this thesis by recapitulating that processes related to embodied cognition are fundamental to the perception emotive effect of music, and that they are also meaningful in two important ways: they offer analytical insights to various aspects of music, from underlying phenomenon such as musical verticality (discussed in 2.2), to particular, extended excerpts of musical pieces (as shown in the analyses in 2.3, 3.3, 4.1.1, 4.2.3, and 4.3.3). They are also tightly encoded in humans' cognitive system, involving not only the receiving end but also the producing agent of music (5.1), and interact with other social cognitive processes (5.2). Finally, they can inspire composition methodologies or concepts when the composer deliberately manipulates these mechanisms (as discussed in 3.4, 4.1.4, 4.2.4, and 4.3.4), which is their contribution in practice-based areas of the research.

5.1 Evaluative Conditioning and Episodic Memory of the Performers

The examples of episodic memory and evaluative conditioning given by Juslin involve emotive responses that are evoked by specific pieces or excerpts of music in specific individuals. These phenomena can be rather contextually dependent, but empirical studies on the general presence of their effects have also been conducted. For instance, Sloboda, O'Neill, and Ivaldi (2001) conducted an ESM (experience sampling method) study where they asked non-musician participants to report their activities randomly between 8 AM and 10 PM for a week, by the means of electronic pagers. They have found that music, especially those from the participants' personal choices, are used to accompany more than 40% of the activities and have effects on the listeners by improving focus and alertness. Some participants also mentioned that they realised that they associate music and their emotion, which they were doing subconsciously before. It indeed suggests the emergence of accompanying memories that would form associations with the music, as Schulkind, Hennis, and Rubin (1999) have discovered, a strong correlation is present between the emotion and memory when participants are played excerpts of music. The effect of evaluative conditioning by music has also been explored in empirical research, such as the 2001 study by Bolders, Band, and Stallen, who have successfully established conditioning of valence between music and lexical stimuli in participants, or an earlier study by Blair and Shrimp (1992) who have found correlation between listeners' attitude towards a brand and their experience with a specific piece of music.

As stated above, these mechanisms obviously depend on particular cultural background and personal experience of the listener. However, systematic examples of episodic memory between music and events, or evaluative conditioning between music and performance occasion, can be observed particularly in performers of certain classical tradition. The following discussions will include cases of Formosan Austronesian and Hindi music traditions.

Camake Valaule, for instance, was a musician and music teacher from Paiwan, which is a Formosan Austronesian ethnic group (Taiwanese indigenous people). Having been to college in Taitung, a city in Southeast Taiwan where Austronesian culture is deeply embedded, he had been interested in and actively learning songs/dances from not only his own Paiwan tradition, but also other Formosan Austronesian groups such as Amis, Bunun and Puyuma. After college, he took the job as a teacher in Taiwu Primary School in Pingtung county, where he started to coach student singers, at first one solo singer Lumasan, who won the first prize of National traditional Singing Contest in Taiwan, and subsequently the singing group Taiwu Ancient Ballads Troupe, which he founded and supervised.

The traditional music pieces performed by Taiwu Ancient Ballads Troupe were mostly provided by Valaule himself. He has been collecting traditional songs, particularly from his own Paiwan tradition, which had been only passed down orally before. During the process, he realised that many senior singers did not remember one entire song. He had to collect excerpts from different singers and combine them afterwards. A female singer's response to his request to record love songs offered a possible explanation as to why certain songs tend to be lost: this female singer, once considered as the best love song singer in the entire tribe, refused to demonstrate the love songs that Valaule wished to collect and learn, because she was abiding by the principle that these love songs, along with the emotion and thought carried by the tune, are strictly reserved for premarital days. Now a senior person who already has grandchildren, she considered her singing days to be over because she should not be "goofy and childish" like she used to be in the old days (Kuo, 2014). Interestingly, her grandson, was no other than Camake Valaule himself. This also shows how strong the principles in evaluative conditioning can be for musicians in this convention: even at the request of her own grandchild, for the cause of keeping the musical tradition of her people, she still would not break the rules and sing the tunes outside of its ascribed circumstance.

Another convention that associates elements of music with particular circumstances is the raga (a structure in Hindi music consisting of a particular melodic mode) and rasa (an atmosphere and essence that is supposed to be evoked by a particular raga) in Hindi classical music. Sangita Makaranda, a written work from around 1100 CE by Narada, classifies the ragas into different times of the day (morning, noon, sunset etc.). Another author, Somesvara, also ascribed ragas to different seasons. Granted, this entire performance time system has developed to be very complicated and cannot always be strictly followed, especially in modern days. In radio production, for instance, pieces using ragas of different performance times are actually recorded altogether in one session, and later broadcasted in their assigned times. Somesvara has also added in the end of his work that the seasonal ragas are free to be performed outside the ascribed season for pleasure's sake. However, there are still practitioners that prefer to strictly follow the system like Camake Valaule's grandmother did. Narada warned the musicians and listeners that disasters would befall if the performance times he described are not correctly followed. And indeed, there has been senior Hindi musicians that stick to the system so firmly that they even took issues with the decontextualisation of some genres of music in western classical music, believing that the "Westerners" will soon incur horrible disaster because they have been performing funeral marches when there is no funeral, spring songs when it is not spring, and nocturnes at day[90], all "at wrong occasions and at wrong times" (Kaufman, 1965).

As stated in the beginning of this section, these phenomena serve as examples of concrete systems deriving from the two mechanisms involved. The aspect of the performers is therefore an important part for the framework that should be included in the discussion. Indeed, apart from the two cases briefly explored, there have been and still are many music performance conventions that demonstrate similar characteristics. They are pertinent to the embodied aspects of music mainly because of the following three reasons: 1. Although many such principles are outdated, as are some rules in raga/rasa performance time, for musicians that still do follow them, they form a part of their embodied experience and have significant cognitive effects. In these cases, the production of particular pieces of music requires an external contextual factor, which could be religious restriction, social etiquette or aesthetic

[90] They also mentioned "love song when there is not love", which perfectly corresponds to what Camake Valaule's grandmother said.

principle. The presence of these requisites places particularly great emphasis, if one again refers to the 4 Es process proposed by Ward and Stapleton (2012), on the embedding process, which refers to the interaction between individual and the environment, and it shows how salient and effective the elements in cognitive processes could be: indeed, pieces such as Christmas carol in Europe and Northern America also have sort of their fixed performance time, but conventions in modern days with strictest performance time restrictions, such as Paiwan songs, would not be performed even for the purpose of recording or preserving it, which is an epitome of how certain components in the theoretical framework, such as the environment element in the embedding process, the contextual information in episodic memory and evaluative conditioning, can have overriding effects which dictate the entire process: music does not only evoke memory, but can be confined in the memory with which it is associated. I have shown how certain mechanisms such as visual imagery can be dual-directionally evocative; here these mechanisms are even more tightly connected, because they can be dual-directionally causative.

2. As active principles that regulate the musician's embodied experience, both in terms of performing and emotion, these phenomena also lead to discussions on another side of music's purpose and meaning. When it is dependent on the performer and occasion as to whether or not, or how, something should be performed, the focus of the music's purpose seems to shift from the listeners' side (to be heard) to that of the musician and circumstance (simply to be played). In fact, these cases may constitute a substantial part of the music culture in the world. As Spitzer (2021) has argued:

> A popular idea [of music's origin] is that music arose from the coordinated activity of a group of people, and there was probably no distinction between performers

and audience. Both these aspects were inherent to ancient work music when hominins first banged rocks together […]. They were also present when music broke apart from work and was practised as an activity in itself, in festivals or over an evening drink. The anthropologist Thomas Turino calls this type of music – where the performers and the listeners are the same people – participatory[91].

and believed that:

The norm, historically (and prehistorically) as well as geographically, is to actively do music with other people, not passively consume it on your own. (p.347)

While the definition of "the norm" seems to be not entirely clear in Spitzer's work, such a type of music practice does contribute to important perspectives in the BRECVEMA model. In the singing tradition of Paiwan music, for instance, an important element in the music is the performer's emotion. Individual singers improvise to express their thoughts, and family members sing together tunes that are known to the whole tribe, but with ornaments unique to their own house (Chen, 2010). The listeners of course can still exist in the scenario, and are free to have their own emotive response, but the main focus of the activity here is the episodic memory and evaluative conditioning of the people that are singing the songs. Expressing their own thoughts and emotion, the Paiwan singers also create verse and poetry through improvised singing, sometimes even dialogues[92] (Sun, 2005). Interestingly, other mechanisms on the musicians themselves also occur, such as rhythmic entrainment which results in the singers' dancing (Banai, 2000). Again, musicians from any convention may move along as they play, but here the mechanism is focused and

[91] In his framework, Turino (2008) has classified into 4 types of practices. As Spitzer has quoted:

Turino divides the world's music into four artistic practices, or fields, which he calls the participatory; the presentational; high fidelity; and studio audio art. (p.87)

[92] Although this skill is mostly lost in the younger generation as of 2020 due to much lower proficiency in their native Paiwan language (Lee, 2020).

developed to have its own status as part of the performance.

5.2 Musical Expectancy at Integrated Levels

In the BRECVEME model, the mechanism of musical expectancy originally refers to the induction of emotion which occurs when "a specific feature of the music violates, delays, or confirms the listener's expectations about the continuation of the music" (2011, p.623). Indeed, expectation is a mechanism underlying many aspects of not only music but also emotion and cognition, and there has been other models other than that of Juslin's that deal with this process. In 4.1.3 I briefly mentioned Huron's 2008 book where he proposes a 5-category system, what he calls ITPRA theory, of expectancy in music. It is now appropriate to discuss this further. Huron's core argument is in line with what Meyer has proposed in his 1956 book, which attributes music's emotive effect largely to expectation. As mentioned in 4.1.3, Meyer was not in favour of measurement of biological reactions to music, but his objection of course has at least something to do with the comparatively limited technology back in his time. Many later studies that apply empirical methods actually have been done to support Meyer's work. Huron's work, drawing on various psychological and statistical evidence, has continued Meyer's line of argument and put forward the ITPRA theory of musical expectation, where each of the five letters stand for a type of expectation response in the theoretical framework: imagination, tension, prediction, reaction, and appraisal. These proposed processes all have evolutionary root and can again be accommodated in the context of embodied cognition: imagination and tension, termed as "pre-outcome responses" by Huron, involve the expectation of a potential outcome and preparation for it. These mechanisms lead to behavioural and physiological responses to a potential future outcome. Prediction, reaction, and appraisal, termed as "post-outcome responses" by Huron, occur after the event has taken place. Prediction response occur in accordance with an outcome that confirms or violates the expectation. Reaction and appraisal response occur regardless of the accuracy of the expectation and purely to the outcome itself as an event. Reaction is an instant, subconscious process whereas appraisal occurs after the listener or individual applies intentional assessment of the outcome.

It is easy to see some correspondence between this model and 4e cognition theory: the automatically activated tension and reaction are of somatic nature and part of the "embodiment" mechanism; the imagination and subsequent prediction response involve the "embedding" process where agent and environment, in this case the listener and musical events, interact. The appraisal response is related to the "extension" mechanism where one integrates internal and contextual information. Eventually, meaning and emotion emerge through "enacting". It should be noted that Huron often supports his theory with results from statistic and experimental studies, such as analyses of melody tendencies (chapter 8) and experiment in melody length expectation (pp.229-230). The evidence usually demonstrates clearly how the proposed responses operate in musical environment and enhances the validity of the model, and also in turns the embodiment implication of the mechanism.

Also, already from Huron's own description of the processes, one can tell that the expectation responses work in correlation with many other mechanisms discussed in previous chapters. Indeed, Huron has clearly pointed out that anticipation is not the only emotion inducing mechanism, and Juslin also has stated that the mechanisms in the BRECVEMA model interact with each other. As mentioned in previous sections, processes such as visual imagery (4.1.1) and rhythm entrainment (4.2.3) constantly interrelates with expectation, or, in Juslin's term, musical expectancy. To recapitulate, these analyses involve contextual information and formal feature of the pieces discussed. In Kagel's Dressur, expectancy arises when (or if) the arena-like stage design reminds the audience of circus or dressage environment. In Stasack's Six Elegies Dancing, expectancy forms with the structural design that alternating between movements with and without salient rhythmic groupings. To add to the discussions, here is another analysis from the perspective of biological basis and shows how music expectancy interact with features involving brain stem reflex: Chen and Tsai (2016) tested the respiration rate and heart rate of participants listening to heavy metal music excerpts with sudden change from loud to soft passages, and found that their heart rates in the soft passages are significantly lower than when they listen to the same soft passage preceded by soft sounds of sea. This shows that music expectancy (in this case, the contrast formed by violating the expectancy for the tension to continue) have significant effects on listeners' emotion response. Specifically, the sound quality that would usually be associated with aggression in loud passages of heavy metals actually helps to enhance the relaxing quality in the soft passage by violating the expectation.

In addition, expectancy can be observed not only at the scale of local components such as dynamics or rhythm, but also global, integrated level such as stylistic information. Familiarisation with musical components, be it harmony, rhythm, instrumentation, and others, or rather the synthetics of all the above, would activate a specific listening mode that affects the perception of audio elements. As stated in the beginning of this chapter, this could have overriding effects on other mechanisms. This is similar to what Juslin later proposes and terms as "aesthetic attitude", but it can actually be triggered by identification of an audio stimulus as other potential symbols, not only as an aesthetic object, and it apparently also varies between individuals, depending on different levels of knowledge of contextual information they retrieve from the music. Koelsch, Schmidt & Kansok (2002) played chord sequences to groups of musicians and non-musicians and have found that ERAN (early right anterior negativity)[93] responses in musicians are significantly stronger at "out of context" [94] Neapolitan chords, showing that listeners familiar in a style, context, or "musical syntax" [95] , do have stronger tendencies to develop certain expectancy in the progression of musical elements. In another study by Steinbeis, Koelsch & Sloboda (2006), researchers played to the participants the original and recomposed versions of J.S. Bach chorales where the two recomposed versions have a chord replaced with another one that is either "more expected" or "less expected". The listeners' brain electric responses to the "less expected" version were significantly larger. It is particularly important to point out that in some examples, the replaced chords have the same type, for example a major third chord. This clearly indicates that even in separation the two chords should have the same harmonic property[96], in the specific context where listeners have certain expectations in harmonic progression, they are perceived very differently. This is again in the musical syntax of western classical music harmony, and trained musicians with greater familiarity with the context again have significantly stronger reaction.

Finally, Huron has also associated the concept of expectation with a synchronic group of musicians' attitudes in regard to the music they create. Analysing the musical properties and historical contexts of composers such as Wagner, Schoenberg, and Stravinsky, Huron has concluded that there is a collected mentality associated with the so-called musical modernism which can be regarded as the deliberate violation of expectation:

---

[93] A type of brain electric activation that is shown to be responsive to musical irregularities (Koelsch et al., 2001)

[94] According to western classical music harmony

[95] A term used in previous studies such as Bharucha & Krumhansl (1983), Swain (1997), Patel and colleagues (1998), Maess and colleagues (2001).

[96] They are played in a midi piano sound with equal temperament.

> There are characteristic features of modern music whose most straightforward interpretation is the goal of psychological irritation or unease. Many of the musical devices used by Wagner, Schoenberg, Stravinsky, and others make sense only when viewed as serving the goal of psychological disruption through thwarted expectation. (p.350)

Although this statement may be overgeneralised, from the perspective of composition, it could still provide some relevant narratives. Long after composers such as Wagner and Schoenberg, an unexpected absence of cadence or combination of interval might no longer be so "disruptive", but the effect of musical expectancy (or lack thereof) still plays decisive roles in how audio stimuli are received. As reviewed above, previous studies have shown how an existing expectancy alters the perception of certain musical elements. It is then possible to have one explanation as to why certain music are "interesting" and "novel" in the context after modernism era: when a piece of composition does not invite the listeners to recognise a certain pattern or context, less expectancy related to the "'higher' cognitive function, domain-relevant knowledge" would be evoked, and it is more likely for the audio elements to have effects in their primitive form, such as rhythmic and harmonic characteristics. It is of course not to say that music summoning expectancy are less worthy in any way, but simply that in this case, the music detached from the higher-level meaning-making of enactment tend to deliver effects in the most basic level of embodiment cognition in a more controllable fashion.

5.3 Conclusive Remarks

This thesis started with a Merleau-Ponty quote which states that music is nothing more than "a series of audio sensation". Through the thesis, I have attempted to argued that for audio sensation to be perceived as music, much more aspects are involved in the process. I have chosen to argue this following the structure of the BRECVEMA model of musical emotion proposed by Juslin, because the theoretical framework is the most appropriate for my approach, which explores the issue in three perspectives: theoretical, analytical, and practical.

The theoretical approach often combines fields of study such as aesthetics, phenomenology, linguistics, cognitive science, neuroscience, and biology. It further examines each category in the BRECVEMA framework from the point of view of embodied cognition, providing details that are applicable in the analytical and practical sections.

The analytical approach investigates pieces of EMT, which is a particularly flexible setting that enables many methods involving the mechanisms in the framework to be applied in the composition. For the mechanisms of emotional contagion, visual imagery, rhythmic entrainment, and brain stem reflex, I discussed particular pieces where the processes described in the BRECVEMA model and embodied cognition theories proved to be idea bases for analysis. For the mechanisms of evaluative conditioning, episodic memory and musical expectancy, I also analysed some cases of general music performance and perception which exemplified the operation of the mechanisms.

The practical approach included my own EMT composition which intentionally applies the mechanisms in the BRECVEMA models. I have submitted three pieces which include instances of methodology developed on the basis of theories related to emotional contagion, visual imagery, rhythmic entrainment, and brain stem reflex, and discussed the relevant parts in the end of each mechanism's section.

Overall, this thesis is an attempt to demonstrate the contribution of a research methodology that combines theoretical and practice-based approaches in the context of embodied perception of music. As have been argued in 3.4, 4.1.4, 4.2.4, and 4.3.4, the combination results in a mutually beneficial relation between the theory and the practice. The theory helps constructing the analytical tools, the methodologies, and the materials for the compositions, while the practice output enriches the theories by demonstrating extended interpretations and emergent interactions of the mechanisms. These two paths of relation serve as principles for a methodological model, which in the future can be applied for the investigation of relevant topics. Since musical emotion emerges from processes in physical, cognitive, and phenomenological realms and involve all participants in the experience, the combination of theory, analysis and composition can usefully offer concrete, clear examples that depict different perspectives of the big picture. The discussion in this study, for instance, has effectively elaborated on the notion that music involves much more than merely audio sensation, including the many aspects denied by Merleau-Ponty: the pieces analysed have shown how music can evoke, for instance, one's memory of horses in a circus, feeling of being situated in a forest, knowledge of a tradition that summons essences with melodies, and dream of a distant planet.

Bibliography

Ackermann, H., & Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: a review and a hypothesis. *Brain and language*, *89*(2), 320-328.

Aide, T. M., Hernández-Serna, A., Campos-Cerqueira, M., Acevedo-Charry, O., & Deichmann, J. L. (2017). Species richness (of insects) drives the use of acoustic space in the tropics. *Remote Sensing*, *9*(11), 1096.

Banai, M. (2000). The Beauty of Indigenous Music. *The singing and appreciation of native music*, edited by G. Lin, 199–204. Yilan County: National Center for Traditional Arts.

Bengtsson, S. L., Ullen, F., Ehrsson, H. H., Hashimoto, T., Kito, T., Naito, E., Forssberg, H. & Sadato, N. (2009). Listening to rhythms activates motor and premotor cortices. *cortex*, *45*(1), 62-71.

Blair, M. E., & Shimp, T. A. (1992). Consequences of an unpleasant experience with music: A second-order negative conditioning perspective. *Journal of Advertising*, *21*(1), 35-43.

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the national academy of sciences*, *98*(20), 11818-11823.

Bolders, A. C., Band, G. P., & Stallen, P. J. (2012). Evaluative conditioning induces changes in sound valence. *Frontiers in Psychology*, *3*, 106.

Boiten, F. A., Frijda, N. H., & Wientjes, C. J. E. (1994). Emotions and respiratory patterns: Review and critical analysis. *International Journal of Psychophysiology*, 17, 103-28.

Bradt, J., Magee, W. L., Dileo, C., Wheeler, B. L., & McGilloway, E. (2010). Music therapy for acquired brain injury. *Cochrane Database of Systematic Reviews*, (7).

Jarvis, M. S., & Doupe, A. J. (2013). Translating birdsong: songbirds as a model for basic and applied medical research. *Annual review of neuroscience*, *36*, 489.

Brosch, T., Sander, D., & Scherer, K. R. (2007). That baby caught my eye... attention capture by infant faces.

Brosch, T., & Sander, D. (2013). Comment: the appraising brain: towards a neuro-cognitive model of appraisal processes in emotion. *Emotion Review*, *5*(2), 163-168.

Brown, S., Martinez, M. J., & Parsons, L. M. (2004). Passive music listening spontaneously engages limbic and paralimbic systems. *Neuroreport*, *15*(13), 2033-2037.

Buccino, G., Binkofski, F., & Riggio, L. (2004). The mirror neuron system and action recognition. *Brain Lang*, 89(2), 370-376.

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J., et al. (2004). Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. *Neuron*, 42(2), 323-334.

Burt, J. L., Bartolome, D. S., Burdette, D. W., & Comstock Jr, J. R. (1995). A psychophysiological evaluation of the perceived urgency of auditory warning signals. *Ergonomics*, *38*(11), 2327-2340.

Byers, P. (1976). Biological rhythms as information channels in interpersonal communication behavior. In *Perspectives in ethology*, (Vol. 2) (ed. P. P. G. Bateson & P. H. Klopfer) pp. 135-64. New York: Plenum Press.

Carston, R. (2010). Lexical pragmatics, ad hoc concepts and metaphor: from a relevance theory perspective. *Italian Journal of Linguistics*, *22*(1), 153-180

Chapados, C., & Levitin, D. J. (2008). Cross-modal interactions in the experience of musical performances: Physiological correlates. *Cognition*, *108*(3), 639-651.

Chapin, H. L., Zanto, T., Jantzen, K. J., Kelso, S., Steinberg, F., & Large, E. W. (2010). Neural responses to complex auditory rhythms: the role of attending. *Frontiers in psychology*, *1*, 224.

Chen, S. (2010). *The heritage and development of the musical culture of Paiwan tribe: illustrated with Tai-wu township of Pintung county*. (Unpublished master's thesis). Tainan National University of the Art, Taiwan.

Clayton, M. (2009). The social and personal functions of music in cross-cultural perspective. *The Oxford handbook of music psychology*, 35-44.

Clayton, M., & Leante, L. (2011). Imagery, melody and gesture in cross-cultural perspective. *New perspectives on music and gesture*, *203*.

Chaikin, J. (1972). The Presence of the Actor. *New York: Theatre Connmlnications Group*.

Cheng, T. H., & Tsai, C. G. (2016). Female listeners' autonomic responses to dramatic shifts between loud and soft music/sound passages: a study of heavy metal songs. *Frontiers in psychology*, *7*, 182.

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

Cocker, M. (2013). *Birds and People*. London: Vintage Publishing.

Cook, P., Rouse, A., Wilson, M., & Reichmuth, C. (2013). A California sea lion (Zalophus californianus) can keep the beat: motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *Journal of Comparative Psychology*, *127*(4), 412.

Coslett, H. B., Wiener, M., & Chatterjee, A. (2010). Dissociable neural systems for timing: evidence from subjects with basal ganglia lesions. *PLoS One*, *5*(4), e10324.

Coull, J. T., Cheng, R. K., & Meck, W. H. (2011). Neuroanatomical and neurochemical substrates of timing. *Neuropsychopharmacology*, *36*(1), 3-25.

Cox, A. W. (1999). The metaphoric logic of musical motion and space. PhD diss., University of Oregon.

Cox, A. (2001). The mimetic hypothesis and embodied musical meaning. *Musicae scientiae*, *5*(2), 195-212.

Cox, A. (2016). Music and embodied cognition: Listening, moving, feeling, and thinking. Indiana University Press.

Craig, D. G. (2005). An exploratory study of physiological changes during "chills" induced by music. *Musicae scientiae*, *9*(2), 273-287.

Critchley, H. D. (2002). Electrodermal responses: what happens in the brain. *The Neuroscientist*, *8*(2), 132-142.

Cunningham, W. A., & Brosch, T. (2012). Motivational salience: Amygdala tuning from traits, needs, values, and goals. *Current Directions in Psychological Science*, 21, 54–59.

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception*, *24*(5), 433-454.

Davies, P. J. (1979). When an Order for Specific Performance Fails. *Mod. L. Rev.*, *42*, 696.

Del Olmo, M. F., Cheeran, B., Koch, G., & Rothwell, J. C. (2007). Role of the cerebellum in externally paced rhythmic finger movements. *Journal of neurophysiology*, *98*(1), 145-152.

Delalande, F., Addessi, A. R., & Young, S. (2009). Analysing the first spontaneous musical behaviour: A pedagogical and anthropological objective. *Proceedings of the European Network of Music Educators and Researchers of Young Children*, 529-536.

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of neurophysiology*, *84*(6), 3072-3077.

Dik, S. C., & Hengeveld, K. (1991). The hierarchical structure of the clause and the typology of perception-verb complements. *Linguistics*.

Dittrich, W. H., Troscianko, T., Lea, S. E., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, 25(6), 727-738.

Doolittle, E. (2008). Crickets in the concert hall: A history of animals in western music. *Trans. Revista Transcultural de Música*, (12).

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, *384*(6605), 159-161.

Ekman, P. (1992). Facial expressions of emotion: an old controversy and new findings. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *335*(1273), 63-69.

Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krauser, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M. & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal* *of personality and social psychology*, *53*(4), 712.

Fauconnier, G., & Turner, M. (2008). *The way we think: Conceptual blending and the mind's hidden complexities*. Basic books.

Flowerdew, J. (1990). Problems of speech act theory from an applied perspective. *Language Learning*, *40*(1), 79-105.

Fishman, Y. I., Volkov, I. O., Noh, M. D., Garell, P. C., Bakken, H., Arezzo, J. C., Howard, M. A., Steinschneider, M. (2001). Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. *Journal of Neurophysiology*.

Fitch, W. T. (2015). Four principles of bio-musicology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140091.

Foss, J. A., Ison, J. R., Torre Jr, J. P., & Wansack, S. (1989). The acoustic startle response and disruption of aiming: II. Modulation by forewarning and preliminary stimuli. *Human factors*, *31*(3), 319-333.

Ferreira, L. M., Oliveira, E. G., Lopes, L. C., Brito, M. R., Baumgarten, J., Rodrigues, F. H., & Sousa-Lima, R. S. (2018). What do insects, anurans, birds, and mammals have to say about soundscape indices in a tropical savanna. *Journal of Ecoacoustics*, *2*(1), 1-1.

Freud, S. (1977). *Introductory lectures on psychoanalysis*. WW Norton & Company.

Gabrielsson, A. (2001). Emotion perceived and emotion felt: Same or different?. *Musicae scientiae*, *5*(1_suppl), 123-147.

Gallagher, S. (2011). Embodiment and phenomenal qualities: An enactive interpretation. *Philosophical Topics*, *39*(1), 1-14.

Gallagher, S., & Zahavi, D. (2007). *The phenomenological mind: An introduction to philosophy of mind and cognitive science*. Routledge.

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cognitive Brain Research*, *20*(2), 226-241.

Gerardin, E., Sirigu, A., Lehericy, S., Poline, J. B., Gaymard, B., Marsault, C., et al. (2000).

Partially overlapping neural networks for real and imagined hand movements. *Cereb Cortex*, 10(11), 1093-1104.

Gieseking, W., & Leimer, K. (1972). *Piano technique*. New York: Dover

Godøy, R. I., Haga, E., & Jensenius, A. R. (2005). Playing “air instruments”: mimicry of sound-producing gestures by novices and experts. In *International Gesture Workshop* (pp. 256-267). Springer, Berlin, Heidelberg.

Grandjean, D., & Scherer, K. R. (2008). Unpacking the cognitive architecture of emotion processes. *Emotion*, *8*(3), 341.

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

Grahn, J. A., & Brett, M. (2009). Impairment of beat-based rhythm discrimination in Parkinson's disease. *Cortex*, *45*(1), 54-61.

Griffin, D. R., & Speck, G. B. (2004). New evidence of animal consciousness. *Animal cognition*, *7*(1), 5-18.

Grillner, S., Robertson, B., & Stephenson-Jones, M. (2013). The evolutionary origin of the vertebrate basal ganglia and its role in action selection. *The Journal of physiology*, *591*(22), 5425-5431.

Grillner, S., & Robertson, B. (2016). The basal ganglia over 500 million years. *Current Biology*, *26*(20), R1088-R1100.

Groves, P. M., & Thompson, R. F. (1970). Habituation: a dual-process theory. *Psychological review*, *77*(5), 419.

Groupe λ-l. (1975). Car, parce que, puisque. *Revue romane*, *10*(2), 248-280.

Gupta, U. (2018). Personality, Gender and Motives for Listening to Music. *Journal of Psychosocial Research*, *13*(2).

Guryanov, I. O., Rakhimova, A. E., & Rudnick, A. (2017). Socio-cultural aspect of coloristic components of idioms in German discourse. *QUID: Investigación, Ciencia y Tecnología*, (1), 800-805.

Hallam, S., Cross, I., & Thaut, M. (Eds.). (2011). *Oxford handbook of music psychology*. Oxford University Press.

Halliday, M. A. K., & Hasan, R. Cohesion in English 1976 London. *England Longman*.

Halliwell, S. (Ed.). (1987). *The Poetics of Aristotle: translation and commentary*. UNC Press Books.

Halpern, D. L., Blake, R., & Hillenbrand, J. (1986). Psychoacoustics of a chilling sound. *Perception & Psychophysics*, *39*(2), 77-80.

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

Harrer, G., & Harrer, H. (1977). Music, emotion and autonomic function. In *Music and the Brain* (pp. 202-216). Butterworth-Heinemann.

Hanslick, E. (1854). On the musically beautiful. Trans*. Geoffrey Payzant. Indianapolis, In: Hackett*.

Von Helmholtz, H. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (A.J. Ellis, Trans.) Brunswick; Eng. Trans Courier Corporation.

Heile, B. (2016). Towards a theory of experimental music theatre: 'showing doing', 'non-matrixed performance' and 'metaxis'.

Honing, H., Merchant, H., Haden, G. P., Prado, L., & Bartolo, R. (2012). Rhesus monkeys (Macaca mulatta) detect rhythmic groups in music, but not the beat. *PloS one*, *7*(12), e51369.

Hoshino, A. (2011). Scores of Japanese Music Transcribed by ASADA Masayuki. *Research and Reports on Intangible Cultural Heritage.* (5), 77-107.

Hulse, S. H., Bernard, D. J., & Braaten, R. F. (1995). Auditory discrimination of chord-based spectral structures by European starlings (Sturnus vulgaris). *Journal of Experimental Psychology: General*, *124*(4), 409.

Huron, D. (2008). *Sweet anticipation: Music and the psychology of expectation*. MIT press.

Husserl, E. (1913). *Ideas: General introduction to pure phenomenology*. (B. Gibson, Trans.) Routledge.

Imberty, M. (1983). *Les écritures du temps. Sèmantique Psychologique de la Musique*. Bordas, Paris.

Jaeger, S. (2012). A geomedical approach to chinese medicine: the origin of the Yin-Yang symbol. *Recent Advances in Theories and Practice of Chinese Medicine*, 29-44.

James, W. (1894). Discussion: The physical basis of emotion. *Psychological review*, *1*(5), 516.

Jäncke, L., Loose, R., Lutz, K., Specht, K., & Shah, N. J. (2000). Cortical activations during paced finger-tapping applying visual and auditory pacing stimuli. *Cognitive Brain Research*, *10*(1-2), 51-66.

Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, *1016*(1), 749-777.

Jeffries, K. J., Fritz, J. B., & Braun, A. R. (2003). Words in melody: an H215O PET study of brain activation during singing and speaking. *Neuroreport*, *14*(5), 749-754.

Jones, M. R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception & psychophysics*, *41*(6), 621-634.

Jones, M. R. (1990). Musical events and models of musical time. *Cognitive models of psychological time*, 207-240.

Jones, M. R. (2009). Musical time. *The handbook of music psychology*, 81-92.

Joseph, R. (2000). Neuropsychiatry, Neuropsychology. *Clinical Neuroscience*, *3*.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code?. *Psychological bulletin*, *129*(5), 770.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and brain sciences*, *31*(5), 559.

Juslin, P. N., Liljeström, S., Västfjäll, D., & Lundqvist, L. O. (2010). How does music evoke emotions? Exploring the underlying mechanisms.

Juslin P. N. (2011) How Does Music Evoke Emotions? In *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 605-642). Oxford University Press.

Juslin, P. N., & Sloboda, J. (2011). *Handbook of music and emotion: Theory, research, applications*. Oxford University Press.

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of life reviews*, *10*(3), 235-266.

Juslin, P. N., Harmat, L., & Eerola, T. (2014). What makes music emotionally significant? Exploring the underlying mechanisms. *Psychology of Music*, *42*(4), 599-623.

Juslin, P. N., Barradas, G., & Eerola, T. (2015). From sound to significance: Exploring the mechanisms underlying emotional reactions to music. *The American journal of psychology*, *128*(3), 281-304.

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

Kafkas, A., & Montaldi, D. (2014). Two separate, but interacting, neural systems for familiarity and novelty detection: A dual-route mechanism. *Hippocampus*, *24*(5), 516-527.

Kaufmann, W. (1965). Rasa, raga-mala and performance times in North Indian râgas. *Ethnomusicology*, *9*(3), 272-291.

Kawabata, Y. (2011). *The Sound of the Mountain* (E. Seidensticker, trans.). London: Penguin Classics. (Original work published 1949–1954)

Kawashima, R., Okuda, J., Umetsu, A., Sugiura, M., Inoue, K., Suzuki, K., Tabuhi, M., Tsukiura, T, Narayan, S. L., Nagasaka, T., Yanagawa, I., Fujii, T., Takahashi, S., Fukuda, H. & Yamadori, A. (2000). Human cerebellum plays an important role in memory-timed finger movement: an fMRI study. *Journal of neurophysiology*, *83*(2), 1079-1087.

Kirby, M. (1972). On acting and not-acting. *The drama review: TDR*, 3-15.

Kivy, P. (1989). *Sound sentiment: An essay on the musical emotions, including the complete text of the corded shell*. Temple University Press.

Kivy, P. (1990). *Music alone: Philosophical reflections on the purely musical experience*. Cornell University Press.

Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T., & Lotze, M. (2007). Overt and imagined singing of an Italian aria. *Neuroimage*, 36(3), 889-900.

de Knop, S. (2014). A contrastive study of colour terms in French and German causal constructions. *Multilingual cognition and language use: Processing and typological perspectives*, 73-96.

Koelsch, S., Gunter, T. C., SchroÈger, E., Tervaniemi, M., Sammler, D., & Friederici, A. D. (2001). Differentiating ERAN and MMN: an ERP study. *NeuroReport*, *12*(7), 1385-1389.

Koelsch, S., Schmidt, B. H., & Kansok, J. (2002). Effects of musical expertise on the early right anterior negativity: an event-related brain potential study. *Psychophysiology*, *39*(5), 657-663.

Koelsch, S., Fritz, T., DY, V. C., Muller, K., & Friederici, A. D. (2006). Investigating emotion with music: an fMRI study. *Hum Brain Mapp*, 27(3), 239-250.

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in cognitive sciences*, *9*(12), 578-584.

Koffka, K. (1935). *Principles of Gestalt psychology*. International library of psychology, philosophy and scientific method. New York: Harcourt, Brace and Company.

Konečni, V. J. (2008). Does music induce emotion? A theoretical and methodological analysis. *Psychology of Aesthetics, Creativity, and the Arts*, 2(2), 115.

Krahé, C., Hahn, U., & Whitney, K. (2015). Is seeing (musical) believing? The eye versus the ear in emotional responses to music. *Psychology of Music*, *43*(1), 140-148.

Krause, B. L. (1993). The niche hypothesis: a virtual symphony of animal sounds, the origins of musical expression and the health of habitats. *The Soundscape Newsletter*, *6*, 6-10.

Krueger, J. W. (2011). Doing things with music. *Phenomenology and the cognitive sciences*, *10*(1), 1-22.

Kübler-Ross, E., & Kessler, D. (2005). *On grief and grieving: Finding the meaning of grief through the five stages of loss*. Simon and Schuster.

Kumar, S., Forster, H. M., Bailey, P., & Griffiths, T. D. (2008). Mapping unpleasantness of sounds to their auditory representation. *The Journal of the Acoustical Society of America*, *124*(6), 3810-3817.

Kumar, S., von Kriegstein, K., Friston, K., & Griffiths, T. D. (2012). Features versus feelings: dissociable representations of the acoustic features and valence of aversive sounds. *Journal of Neuroscience*, *32*(41), 14184-14192.

Laine, C. M., Spitler, K. M., Mosher, C. P., & Gothard, K. M. (2009). Behavioral triggers of skin conductance responses and their neural correlates in the primate amygdala. *Journal of neurophysiology*, *101*(4), 1749-1754.

Lakoff, G., & Johnson, M. (1980). Conceptual metaphor in everyday language. *The journal of Philosophy*, *77*(8), 453-486

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive science*, *4*(2), 195-208.

Lakoff, G. (1987). The death of dead metaphor. *Metaphor and symbol*, *2*(2), 143-147.

Landels, J. G. (2002). *Music in ancient Greece and Rome*. Routledge.

Landreth, J. E., & Landreth, H. F. (1974). Effects of music on physiological response. *Journal of research in music education*, *22*(1), 4-12.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261-273.

Lange, C. G. (1887). *Über Gemütsbewegungen: eine psycho-physiologische Studie*.

Langer, S. K. (1957). *Philosophy in a new key: A study in the symbolism of reason, rite, and art*.

Harvard University Press.

Last, G., & Kneutgen, J. (1970). Music for sleep. *Munchener Medizinische Wochenschrift (1950)*, *112*(44), 2011-2016.

Lawson, J. (2008) *The Compression and Expansion of Musical Experience in the Digital Age*. [Master's thesis, University of Vermont]

Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *American psychologist*, *46*(8), 819.

Lee, D., Seo, H., & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. *Annual review of neuroscience*, 35, 287-308.

Lee, P. N. (2020). Cultural inheritance on indigenous music education: a Paiwanese music teacher's teaching. *Music Education Research*, *22*(2), 159-172.

Lennie, T. M., & Eerola, T. (2022). The CODA model: a review and skeptical extension of the constructionist model of emotional episodes induced by music. *Frontiers in psychology*, 1389.

Levine, R. (2007). The Death of High Fidelity: In the age of MP3s, sound quality is worse than ever. *Rolling Stone*, 2007 December 27.

Lewis, P. A., & Miall, R. C. (2003). Brain activation patterns during measurement of sub-and supra-second intervals. *Neuropsychologia*, *41*(12), 1583-1592.

Li, F., Harmer, P., Fitzgerald, K., Eckstrom, E., Stock, R., Galver, J., Maddalozzo, G., & Batya, S. S. (2012). Tai chi and postural stability in patients with Parkinson's disease. *New England Journal of Medicine*, *366*(6), 511-519.

Liljeström, S. (2011). *Emotional reactions to music: prevalence and contributing factors* (Doctoral dissertation, Acta Universitatis Upsaliensis).

Linden, E. (2000). *The Parrot's Lament: And Other True Tales of Animal Intrigue, Intelligence, and Ingenuity*. New York: Plume.

Linden, E. (2003). Octopus and the Orangutan: More True Tales of Animal Intrigue, Intelligence, and Ingenuity. New York: Plume.

Lipscomb, S. D., & Hodges, D. (1996). Hearing and music perception. In *Handbook of music psychology* (pp. 83-132). Institute for Music Research.

London, J. (2012). *Hearing in time: Psychological aspects of musical meter*. Oxford University Press.

Lonsdale, A. J., & North, A. C. (2011). Why do we listen to music? A uses and gratifications analysis. *British journal of psychology*, *102*(1), 108-134.

Manthey, S., Schubotz, R. I., & von Cramon, D.Y. (2003). Premotor cortex in observing erroneous action: an fMRI study. *Brain Res Cogn Brain Res*, 15(3), 296-307.

Marr, D. (1982): *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman and Company

McConachie, B. (2008). *Engaging audiences: A cognitive approach to spectating in the theatre*. Springer.*

McConachie, B. (2011). An evolutionary perspective on play, performance, and ritual. *TDR/The Drama Review*, *55*(4), 33-50.

Meister, I. G., Krings, T., Foltys, H., Boroojerdi, B., Muller, M., Topper, R., et al. (2004). Playing piano in the mind--an fMRI study on music imagery and performance in pianists. *Brain Res Cogn Brain Res*, 19(3), 219-228.

Merchant, H., Zarco, W., & Prado, L. (2008). Do we have a common mechanism for measuring time in the hundreds of millisecond range? Evidence from multiple-interval timing tasks. *Journal of Neurophysiology*, *99*(2), 939-949.

Merezhkovsky, D. S. (1901). *The Death of the Gods* (H. Trench, trans.). London: Archibald Constable. (Original work published 1895)

Merleau-Ponty, M. (2004). *The world of perception* (O. Davis, trans.). Routledge. (Original work published 1948)

Merleau-Ponty, M. (1964). *The primacy of perception and its philosophical consequences.* Northwestern University Press. (J. M. Edie, trans.) (Original work published 1964)

Meyer, M. (1898). Zur Theorie der Differenztöne und der Gehörsempfindungen überhaubt, *Beiträge zur Akustik und Musikwissenschaft*, 2

Meyer, L. (1956). *Emotion and meaning in music* (p. 307). Chicago: University of Chicago Press.

Mitchell, H. F., & MacDonald, R. A. (2016). What you see is what you hear: The importance of visual priming in music performer identification. *Psychology of Music*, *44*(6), 1361-1371.

Mohandesan, E., Fitak, R. R., Corander, J., Yadamsuren, A., Chuluunbat, B., Abdelhadi, O., Raziq, A., Nagy, P., Stalder, G., Walzer, C., Faye, B. & Burger, P. A. (2017). Mitogenome sequencing in the genus Camelus reveals evidence for purifying selection and long-term divergence between wild and domestic Bactrian camels. *Scientific reports*, *7*(1), 1-12.

Molino, J. (2000). *Toward an evolutionary theory of music and language*. The MIT Press.

Monroy, M., & Keltner, D. (2023). Awe as a pathway to mental and physical health*. Perspectives on psychological science*, 18(2), 309-320.

Moore, B. C. (1973). Frequency difference limens for short-duration tones. *The Journal of the Acoustical Society of America*, *54*(3), 610-619.

Morley, I. R. M. (2014). A multi-disciplinary approach to the origins of music: perspectives from anthropology, archaeology, cognition and behaviour. *Journal of Anthropological Sciences*, *92*, 147-177.

Morris, J. S., Öhman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature*, *393*(6684), 467-470.

Motokiyo, Z. (1984). On the art of the No drama: The major treatises of Zeami. *Trans. J. Thomas Rimer and Yamazaki Masakazu. Princeton, NJ: Princeton UP*.

Mulder, T., Zijlstra, S., Zijlstra, W., & Hochstenbach, J. (2004). The role of motor imagery in learning a totally novel movement. *Exp Brain Res*, 154(2), 211-217

Mynott, J. (2009). *Birdscapes: Birds in our imagination and experience*. New Jersey: Princeton University Press.

Näätänen, R., Paavilainen, P., Titinen, H., Jiang, D., & Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology*, *30*(5), 436-450.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. University of Chicago Press.

Nettl, B. (2000). An ethnomusicologist contemplates universals in musical sound and musical culture. *The origins of music*, 463-472.

N'gweno, F. (2010). Sound, sight, stories and science: Avoiding pitfalls in ethnoornithological research, with examples from Kenya. In S. Tidemann & A. Gosler (Eds.), *Ethno-ornithology: Birds, indigenous peoples, culture and society* (pp. 103-113). London: Earthscan

Öhman, A. (1986). Face the beast and fear the face: Animal and social fears as prototypes for evolutionary analyses of emotion. *Psychophysiology*, *23*(2), 123-145.

Olszewski, J., Miłoński, J., Olszewski, S., & Majak, J. (2007). Hearing threshold shift measured by otoacoustic emissions after shooting noise exposure in soldiers using hearing protectors. *Otolaryngology—Head and Neck Surgery*, *136*(1), 78-81.

Ortony, A., Clore, G. L., & Collins, A. (2015). *The cognitive structure of emotions*. Cambridge university press. (Original work published 1988)

Parncutt, R. (2014). The emotional connotations of major versus minor tonality: One or more origins?. *Musicae Scientiae*, *18*(3), 324-353.

Patel, A., Iversen, J., Bregman, M., & Schulz, I. (2009). Studying synchronization to a musical beat in nonhuman animals. *Annals of the New York Academy of Sciences*, *1169*(1), 459-469.

Pawelec, A. (2006). The death of metaphor. *Studia Linguistica Universitatis Iagellonicae Cracoviensis*, (123).

Pehrs, C., Deserno, L., Bakels, J. H., Schlochtermeier, L. H., Kappelhoff, H., Jacobs, A. M., Fritz, T. H., Koelsch, S. & Kuchinke, L. (2014). How music alters a kiss: superior temporal gyrus controls fusiform–amygdalar effective connectivity. *Social cognitive and affective neuroscience*, *9*(11), 1770-1778.

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., Howard, J. T., Wirthlin, M., Lovell, P. V., Ganapathy, G., Mountcastle, J., Moseley, M. A., Thompson, J. W., Soderblom, E. J., Iriki, A., Kato, M., Gilbert, M. T. P., Zhang, G., Bakken, T., Bongaarts, A., Bernard, A., Lein, E., Mello, C. V., Hartemink, A. J., & Jarvis, E. D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*, *346*(6215), 1256846.

Ploog, D. W. (1992). The evolution of vocal communication. *Nonverbal vocal communication: Comparative and developmental approaches*, *6*.

Premack, D. (2007). Human and animal cognition: Continuity and discontinuity. *Proceedings of the national academy of sciences*, *104*(35), 13861-13867.

Ratcliffe, E. (2015). *Restorative perceptions and outcomes associated with listening to birds*. [Doctoral thesis, University of Surrey]

Rebstock, M., & Roesner, D. (Eds.). (2012). *Composed Theatre: Aesthetics, Practices, Processes*. Intellect Books.

Repp, B. H., & Su, Y. H. (2013). Sensorimotor synchronization: a review of recent research (2006–2012). *Psychonomic bulletin & review*, *20*(3), 403-452.

Rickard, N. S. (2004). Intense emotional responses to music: a test of the physiological arousal hypothesis. *Psychology of music*, *32*(4), 371-388.

Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., & Grodd, W. (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *Neuroreport*, *11*(9), 1997-2000.

Rizzolatti, G., Fadiga, L., Gallese,V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Res Cogn Brain Res*, 3(2), 131-141.

Robb, H. J. (2015). Imagined, Supplemental Sound in Nineteenth-Century Piano Music: Towards a Fuller Understanding of Musical Embodiment. *Music Theory Online*, 21(3).

Rochat, P. & Striano, T. (1999). Emerging self-exploration by 2-month-old infants. *Developmental Science*, 2, 206-18.

Rodriguez-Fornells, A., Rojo, N., Amengual, J. L., Ripollés, P., Altenmüller, E., & Münte, T. F.

(2012). The involvement of audio–motor coupling in the music-supported therapy applied to stroke patients. *Annals of the New York Academy of Sciences*, *1252*(1), 282-293.

Roesner, D. (2016). *Musicality in theatre: Music as model, method and metaphor in theatre-making*. Routledge.

Rolls, E. T. (2007). Emotion elicited by primary reinforcers and following stimulus-reinforcement association learning. *The handbook of emotion elicitation and assessment*, 137-157.

Ruffman, T., Sullivan, S., & Dittrich, W. (2009). Older adults' recognition of bodily and auditory expressions of emotion. *Psychology and Aging*, *24*(3), 614.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychol. Rev*. 110, 145–172.

Russo, F. A., & Thompson, W. F. (2005). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception & psychophysics*, *67*(4), 559-568.

Russo, F. A., & Thompson, W. F. (2005). The subjective size of melodic intervals over a two-octave range. *Psychonomic bulletin & review*, *12*(6), 1068-1075.

Rutherford, D., and Yampolsky, P. (1996). CD liner note. *Music of Indonesia, Vol. 10: Music of Biak, Irian Jaya*. Smithsonian Folkways.

Thompson, W. F., Russo, F. A., & Livingstone, S. R. (2010). Facial expressions of singers influence perceived pitch relations. *Psychonomic bulletin & review*, *17*(3), 317-322.

Tomkins, S. S. (1984). Affect theory. *Approaches to emotion*, *163*(163-195).

Sadhra, S., Jackson, C. A., Ryder, T., & Brown, M. J. (2002). Noise exposure and hearing loss among student employees working in university entertainment venues. *Annals of Occupational Hygiene*, *46*(5), 455-463.

Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., & Zatorre, R. J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PloS one*, *4*(10), e7487.

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature neuroscience*, *14*(2), 257.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences*, 14, 303–316.

Sander, D. (2009). The amygdala. In D. Sander & K. R. Scherer (Eds.), *The Oxford companion to emotion and the affective sciences* (pp. 28–32). New York, NY: Oxford University Press.

Sander, D. (in press). Models of emotion: The affective neuroscience approach. In J. L. Armony & P. Vuilleumier (Eds.), *Handbook of human affective neuroscience*. Cambridge, UK: Cambridge University Press.

Särkämö, T., Tervaniemi, M., Laitinen, S., Forsblom, A., Soinila, S., Mikkonen, M., ... & Hietanen, M. (2008). Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. *Brain*, *131*(3), 866-876.

Sault, N. (2010). Bird messengers for all seasons: Landscapes of knowledge among the Bribri of Costa Rica. In S. Tidemann & A. Gosler (Eds.), *Ethno-ornithology: Birds, indigenous peoples, culture and society* (pp. 291-300). London: Earthscan.

De Saussure, F. (2011). *Course in general linguistics*. Columbia University Press.

Schachner, A., Brady, T. F., Pepperberg, I. M., & Hauser, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology*, *19*(10), 831-836.

Schafer, R. M. (1993). *The soundscape: Our sonic environment and the tuning of the world*. Simon and Schuster.

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual review of neuroscience*, *33*, 203-219.

Seger, C. A., Spiering, B. J., Sares, A. G., Quraini, S. I., Alpeter, C., David, J., & Thaut, M. H. (2013). Corticostriatal contributions to musical expectancy perception. *Journal of cognitive neuroscience*, *25*(7), 1062-1077.

Schechner, R., & Schuman, M. (Eds.). (1976). *Ritual, play, and performance: readings in the* *social sciences/theatre*. Seabury Press.

Schellenberg, E. G., & Trehub, S. E. (1996). Natural musical intervals: Evidence from infant listeners. *Psychological science*, *7*(5), 272-277.

Scherer, K. R., & Zentner, M. R. (2001). Emotional effects of music: Production rules. *Music and emotion: Theory and research*, 361(2001), 392.

Scherer, K. R. (2009b). The dynamic architecture of emotion: evidence for the component process model. *Cogn. Emotion* 23, 1307–1351.

Scherer, K. R., & Coutinho, E. (2013). How music creates emotion: a multifactorial process approach. In Cochrane, T., Fantini, B., & Scherer, K. R. (Eds.), *The emotional power of music: Multidisciplinary perspectives on musical arousal, expression, and social control*, 183-214. Oxford University Press.

Schiavio, A. (2014). *Music in (en) action. Sense-making and Neurophenomenology of Musical Experience* [Doctoral thesis, University of Sheffield].

Schmidt, P. (2014). *Meyerhold at work*. University of Texas Press.

Schön, D., Regnault, P., Ystad, S., & Besson, M. (2005). Sensory Consonance:: An ERP Study. *Music Perception*, *23*(2), 105-118.

Senghor, Leopold Sedar. "Negritude: A humanism of the 20th century." *Optima* 16, no. 1 (1966): 1-8.

Shapiro, L. A. (Ed.). (2014). The Routledge handbook of embodied cognition.

Shapiro, L. (2019). Embodied cognition. Routledge.

Shields, K. (1997). On the Origin of Dialectal Ablaut Patterns of the Present Active Indicative of IE* es-'To Be'. *Historische Sprachforschung/Historical Linguistics*, *110*(2. H), 176-180.

Shih, L. Y., Kuo, W. J., Yeh, T. C., Tzeng, O. J., & Hsieh, J. C. (2009). Common neural mechanisms for explicit timing in the sub-second range. *Neuroreport*, *20*(10), 897-901.

Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of music*, *19*(2), 110-120.

Sloboda, J. A., O'Neill, S. A., & Ivaldi, A. (2001). Functions of music in everyday life: An exploratory study using the Experience Sampling Method. *Musicae scientiae*, *5*(1), 9-32.

Spitzer, M. (2021). *The Musical Human: A History of Life on Earth.* Bloomsbury Publishing.

Stassen, L. (1997). *Intransitive predication*. Oxford University Press.

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of cognitive neuroscience*, *18*(8), 1380-1393.

Stupacher, J., Witek, M. A., Vuoskoski, J. K., & Vuust, P. (2020). Cultural familiarity and individual musical taste differently affect social bonding when moving to music. *Scientific reports*, *10*(1), 1-12.

Sun, D. (2005). Singing with a Pen: The background and present condition and future trends of Taiwan aboriginal literature. *Journal of Taiwan Literary Studies*, *1*, 195–221.

Taruffi, L., Pehrs, C., Skouras, S., & Koelsch, S. (2017). Effects of sad and happy music on mind-wandering and the default mode network. *Scientific reports*, *7*(1), 1-10.

Terhardt, E. (1984). The concept of musical consonance: A link between music and psychoacoustics. *Music perception*, *1*(3), 276-295.

Thoma, M. V., Ryf, S., Mohiyeddini, C., Ehlert, U., & Nater, U. M. (2012). Emotion regulation through listening to music in everyday situations. *Cognition & emotion*, *26*(3), 550-560.

Trainor, L. J., Tsang, C. D., & Cheung, V. H. (2002). Preference for sensory consonance in 2- and 4-month-old infants. *Music Perception*, *20*(2), 187-194.

Tsai, C. J. (2014). On Verse-Chorus Form: Shema Shift and Reward Learning in Pop Songs. *Applied Psychology* (61), 239-286.1

Turino, T. (2008) *Music as Social Life: The Politics of Participation*. Chicago: The University of Chicago Press

Van Dijk, T. A. (1979). Pragmatic connectives. *Journal of pragmatics*, *3*(5), 447-456.

Varela, F., & Thompson, E. (1991). The embodied mind: cognitive science and human experience.

Verstraete, J. C. (1999). The distinction between epistemic and speech act conjunction. *Belgian essays in language and literature 1999*, 119-130.

Vickers, E. (2010, November). The loudness war: Background, speculation, and recommendations. In *Audio Engineering Society Convention 129*. Audio Engineering Society.

Vihman, M. M., & de Boysson-Bardies, B. (1994). The nature and origins of ambient language influence on infant vocal production and early words. *Phonetica*, 51(1-3), 159-169.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, 101(1), 80-113.

Wang, C., Collet, J. P., & Lau, J. (2004). The effect of Tai Chi on health outcomes in patients with chronic conditions: a systematic review. *Archives of internal medicine*, *164*(5), 493-501.

Ward, D., & Stapleton, M. (2012). Es are good. *Consciousness in interaction: The role of the natural and social context in shaping consciousness*, 86-89.

Watanabe, S., Uozumi, M., & Tanaka, N. (2005). Discrimination of consonance and dissonance in Java sparrows. *Behavioural processes*, *70*(2), 203-208.

Weld, H. P. (1912). An experimental study of musical enjoyment. *The American Journal of Psychology*, 23(2), 245-308.

Wildgruber, D., Ackermann, H., Klose, U., Kardatzki, B., & Grodd, W. (1996). Functional lateralization of speech production at primary motor cortex: a fMRI study. *Neuroreport*, *7*(15-17), 2791-2795.

Will, U., & Berg, E. (2007). Brain wave synchronization and entrainment to periodic acoustic stimuli. *Neuroscience letters*, *424*(1), 55-60.

Witt, S. T., Laird, A. R., & Meyerand, M. E. (2008). Functional neuroimaging correlates of finger-tapping task variations: an ALE meta-analysis. *Neuroimage*, *42*(1), 343-356.

Wu, S. H. (2008). The Application of Theatrical Elements in Percussion Works. PhD Thesis, Taipei National University of the Arts.

Wundt, W. M., & Judd, C. H. (1897). *Outlines of psychology* (Vol. 1). Scholarly Press.

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion*, *8*(4), 494.

Zentner, M. R., & Kagan, J. (1998). Infants' perception of consonance and dissonance in music. *Infant Behavior and Development*, *21*(3), 483-492.

Zimmerman, M., & Slavitt, D. R. (2002). *Metamorphoses: A play*. Northwestern University Press.

Macaulay Library. Retrieved 1.10.2018
https://www.macaulaylibrary.org/

The IUCN Red List of Threatened Species. Retrieved 11.11.2022
https://www.iucnredlist.org/

Georges Aperghis (1981), *Les Guetteurs de Son*. Performance by Dressage Percussion. Retrieved 21.9.2019
https://www.youtube.com/watch?v=CGUxZZMKwfw&

Martin Iddon (2020), *Sapindales*. Performed by Heather Roche
Retrieved 11.11.2022
https://www.youtube.com/watch?v=9VAK2H54ptw&t=2935s

Oli Jan (2021), *Three Singers on Planet M*. Performed by Ensemble Thing
Retrieved 11.11.2022
https://www.youtube.com/watch?v=9VAK2H54ptw&t=2120s

Jennifer Stasack (1987), *Six Elegies Dancing*. Performed by Tim Shuster
Retrieved 25.10.2020
https://www.youtube.com/watch?v=B_381RYn5AQ