Embodied music cognition

Exam No. 7043547

MSc in Philosophy

The University of Edinburgh

2010

“I have read and understood The University of Edinburgh guidelines on Plagiarism and declare that this written dissertation is all my own work except where I indicate otherwise by proper use of quotes and references.”

Table of contents

Introduction

The aim of this thesis is to argue in favour of the embodied music cognition paradigm (hereafter: EMC), as opposed to traditional (computational) theories of musical mind. The thesis consists of three chapters. The goal of the first chapter is to (1) examine computational (disembodied) music cognition, focusing on the main problem of this approach (namely: the *symbol–grounding problem*). In the second chapter, (2) I will present and discuss Marc Leman's EMC, that may serve as a response to the problems of the computational view of the musical mind. Although this framework is interesting, it is unclear in several places. For that reason, I intend to enrich it with the references to the recent works on human mirror neuron system (hereafter: MNS) and enactivist views on music cognition. Given that, in the last chapter, I will (3) answer the question, whether mental representations are necessary in music cognition.

The motivation for that thesis is the belief that the EMC is an immensely interesting proposal for the explanation of our interaction with music, based on the accurate recognition of the body's *active* involvement in music cognition. Thereby, this fresh and promising approach gives justice to the phenomenology of musical experience (neglected by traditional research), and can serve as an argument for the embodied paradigm in cognitive science in general.

This introductory chapter will consist of remarks intended to provide a necessary background for the discussion. Firstly, I am planning to (i) draw the frameworks of the traditional (disembodied, computational) and the embodied approaches to music cognition. The second part will briefly sketch the (ii) levels of musical involvement and the description of music.

- **Disembodied and embodied music cognition.**

What is music cognition? In general, this area of study can be described as the interdisciplinary research that is intended to understand the cognitive processing that occurs in the production and perception of music. As stated in the title of this section, it can be claimed that music cognition comes in two juxtaposed varieties: (a) traditional (computational) and the (b) embodied kind.

Traditional music cognition can be accused of negligence of the role of the non – neural body (and its active contribution) in cognitive processing. It is focused, for instance, on musical attention and memory. The idea behind this approach is that music cognition is done *all in the head*. Thus, these disembodied studies are focused on various mental capacities that one must possess in order to perform and perceive (understand) musical structures. In other words, it can be claimed that this kind of music cognition is occupied with the study of passive, input – output processing of symbolic (or sub – symbolic) representations of heard music, all of which will be described in the first chapter of this thesis. In this vein, scholars like Fred Lerdahl and Ray Jackendoff (1983) theorized about the cultural constraints that shape our understanding of musical structures and the acquisition of so-called *musical idioms* that are unique to any given culture.[1]

The embodied model of music cognition is, on the other hand, built on the aspiration to provide a more holistic view on human – music interaction. It aims to give justice to the phenomenology of music experience (that is: recognising the ways in which our interaction with music is grounded in movement, feeling and expression).

Embodiment (as well as the enactivism), in general, rose from feelings of dissatisfaction with computational research on mind. Works in the embodied (Gibbs, 2006; Clark, 2008; Robbins & Aydede, 2009a) and the enactive (e.g. Varela, et. al., 1991; Noë, 2004, 2008) paradigms did, quite significantly, influenced research on music cognition (e.g. Leman, 2008, Godøy & Leman, 2010a). Built on the assumption of the tight coupling between the mind, the body and its environment (whereby action and perception are not separate) embodiment and enactivism (in general and applied into musical research) draw perhaps the most stimulating and thought – provoking perspectives in contemporary cognitive science.

[1] I will return to their works on the (possible) innate capacity to understand music (that is, to construct the principles of musical grammar) later in this thesis (in chapter 1). Their works serve as a fine example of the application of the computational view of mind to music cognition. For now, note that their idea (of tonal grammar) is based on the language – like processing of music.

What is enactivism? In general, this approach is based on the claim that various aspects of our cognition are dependent on our sensorimotor abilities and skills that allow us to directly retrieve information from the environment. Underlining this kind of active and skilful involvement, it can be claimed that cognition is *what we do*, rather than what passively happens to us. This directly implicates the body in cognitive processing, since it is obvious that without the brain's cooperation with the non – neural body no sensory inputs (and no motor outputs) are possible. Yet, while both embodiment and enactivism are grounded in recognition of the body as the active component of cognition, in the following thesis they are to be treated as separate accounts. The reason for that lies in the assumption that the EMC is based on the representational approach to the science of mind, whereas the enactive approach tends to diminish the necessity of mental representations in cognition. The latter is rather focused on direct perception of the information (so – called *ecological paradigm,* to be introduced later on).

For now, it is likewise important to underline that while computational representations (on which traditional (music) cognition is based) are amodal, the representational embodied view endorses *modality – specific representations*, that allow organisms to cope more effectively with their environments. For the sake of these introductory remarks concerning embodiment, it is of great importance to indicate the distinction between *on – line* and *off – line cognition* (Anderson, 2003).

On – line cognition refers to dependence on dynamical interactions between sensorimotor area in the brain and the relevant parts of the non – neural body. Here, the role of the body in cognition is directly indicated. Off – line cognition on the other hand denotes the dependency of the cognitive function on sensorimotor areas in the brain even in the absence of sensory input and motor output. Having recognized that, the role of the body is implicated indirectly – not all forms of embodiment involve bodily dependence in a strict sense; yet, and this is important for the upcoming argumentation, numerous studies indicate at least some kind of causal dependency between the sensorimotor body and the brain (Robbins & Aydede, 2009b).

By what means may this be connected to musical research? Raymond W. Gibbs (2006) argues as follows. Music perception and production are always tightly associated with embodied movements (such as: rhythmic gaits, gestures, as well as the gestural control of musical instruments (in order to produce sound) or breathing). In this thesis I will mainly focus on the various phenomena of perception (and understanding) of music explained in terms of the embodied music cognition framework. How can music cognition be embodied? How does it involve movement? In short, according to an immensely interesting book entitled *Ways of listening: an ecological approach to the perception of musical meaning* Eric F. Clarke (2005) explained three possible connections between movement and sound. These are:

> (1) sound […] is a direct consequence and indicator of motion, as in the clatter of stones down a hillside, the distinctive pitch and dynamic profile of passing motorbike, or the creak of a branch swaying in the wind; (2) the relation between the auditory perception and human action is encapsulated in the functions of *mirror neurons* or the most general *sensorimotor contingencies* […]; (3) the production of musical sound inevitably involves movement, from the relatively discrete movements of vocal tract in singing, to the more manifest movements of hand-clapping and instrumental performance.

The so-called *mirror neurons* indicated by Clarke are one of the main themes of the following chapters. As will be seen, the mirror – neuron system in the brain is proposed to be a mechanism that allows an individual to understand the meaning and intention of communicative signals by evoking a representation of that signal in perceiver's own brain (Molnar – Szakacs & Overy, 2006, 2009). Moreover, and which is more strictly important for this thesis, recent research in neuro-imaging has proven that musical processing (in terms of action) is supported by this system. MNS can be also conceived as a necessary prerequisite for any type of communication, and the link between action and perception, in the way introduced above (Gallese, 2003). How does it work? Recent research suggests that parietal mirror neurons have the special property of coding the motor acts as belonging to an action sequence, and predicting the intended goal of a complex action. What is more, subsets of premotor mirror neurons reveal audiovisual properties that are able to represent the actions independently of whether they are performed, seen – or, interestingly – heard (Molnar – Szakacs, 2006; Kohler et al., 2002; Fogassi et al., 2005). In the second chapter of this thesis, it will be shown how (based on MNS) a comprehensive model of musical communication (called: SAME) was derived, which can support Marc Leman's framework of embodied music cognition.

Sensorimotor contingencies (e.g. Hurley and Noë; 2003) are especially important for the enactive approach. They should be understood as a set of rules of interdependence between stimulation and movement (Gibbs, 2006). By their means perceivers learn how to master the ways in which they retrieve information from the environment (for example: change as a function of movement with respect to the environment). Thus, perception of music performed live is an *activity* involving a set of skilful activities, such as: movements of the head and the rest of the body.

Given that short introduction, in the following thesis I intend to take as my starting point Marc Leman's (e.g. Leman, 2008) framework of embodied music cognition, already mentioned. Leman claims that the human body ought to be considered as the *natural mediator* between the mind and the physical (in this case: musical) environment. This mediator is deployed in time and space by means of (musical) gestures. Musical gestures are to be understood as the *meaningful combination of movements and music.* Following Leman's argument (e.g. Godøy & Leman, 2010b), when we hear (or see) something, we are able to make sense out of it by relating it to our body movements (or the image of the body movements in our mind). This is consistent with on – line and off – line cognition as described above.

According to Leman (2008), the embodied approach to music cognition broadens the focus of musical research in general (by providing new perspectives on multi-modal music perception, kinaesthesia and affective involvement with music), as examined in the following thesis. Yet I intend to argue that some parts of his framework are unclear and need deeper consideration. Before we proceed to the first chapter (dealing with computational music cognition and its gaps), I intend to describe briefly the levels of musical involvement and musical description, as they provide background for the following argumentation.

- **Levels of musical involvement and the description of music.**

We can distinguish two levels of involvement with music, namely direct and indirect (Leman, 2008). *Direct* involvement with music can be roughly defined as: *a matter of corporeal immersion in sound energy, which is a direct way of feeling musical reality*. In other words, one is directly involved with music if this interaction causes subjective, purely personal qualities, like a feeling of self – reward, or (simply) happiness. Leman describes this type of involvement (that can occur while being part of the audience at a concert or of an assembly playing music), as less concerned with cognitive reflection, description or interpretation. What is (probably) meant here is that one's interaction with music is based on kind of pure contemplation of musical values, in which *only here and now matters* to the person that is experiencing it. The vagueness of such description suggests that despite best efforts, this kind of musical involvement is hard to be explained in words, and calls for a provision of a different (than linguistic) form of music description. The distinction between linguistic and non – linguistic descriptions will become more vivid in the paragraph below. *Indirect* involvement with music is to be understood by means of the mediation, for instance, done by linguistic description of music or the audio player. In this type of involvement, there is no direct *immersion in sound energy* / *musical reality*, mainly because of the existence of the mediator between human and musical source.

There are three levels of musical description, based on first, second and third – person perspectives. The *first – person* description is based upon the interpretation of intentions that are attributed to music. For instance (Leman, 2008), I can attribute the description of *inner victory of the spirit* to a passage from Beethoven's *Piano Sonata in A.* In other words, this is my *interpretation* of this passage. Interpretation, in general, is aimed at finding the source of the attributed intentions. This search is often made by the lens of personal history and experience that enters the domain of linguistics based descriptions of the surrounding world. For that reason, musical descriptions have the status of a meaningful yet personal narration that can later be understood by people sharing similar interpretations of environment. As such, a first – person description is built on both intentions and a linguistic basis.

On the contrary, the *third – person description* is based on repeatable measurements of the given phenomena that can be obtained by a person, a machine or a person with the help of a machine. Putting knowledge of human information processing and the user's convention into a machine, it is possible to achieve the measurement of the high – level structural and semantic properties (pitch, loudness, tempo or particular affects such as *happy* or *sad)* directly from the physical properties of the musical piece. The agent's subjective involvement with music can be observed from the third – person point of view. For instance, the brain activity causing the movement of arms and legs in response to music can be recorded and then analysed from the third – person perspective.

Things become more complicated with the *second – person description* that serves as the starting point of Leman's (2008) framework of the embodied music cognition. This level of musical description is based on the question as to whether it is possible to have a form of description that involves intentionality but is less linked with interpretation. Leman proposes that it may be based on an accurate recognition of corporeal articulations. The main difference between the first - person and second - person descriptions lies in the distinction between experience, articulation and experience as being articulated. As it will be described in greater details in the second chapter of this thesis, Leman (2010a) proposes the idea of embodied music semantics. As such, a gestural based second - person description serves as an interesting alternative to linguistic descriptions of musical experience. We may now proceed to the first chapter of this thesis that is intended to discuss the traditional (computational) view of music cognition, as later opposed by the embodied and enactive approaches.

Chapter 1

The computational approach to the science of (musical) mind that will be examined in this chapter comes in two basic varieties: cognitivism and connectionism. Firstly, I intend to provide (1) a brief overview of these models, with the examples of their application into musical research. At the end of this chapter, (2) I will conclude by demonstrating the ways in which the computational approach to musical mind is explanatorily insufficient (e.g. because of its negligence of the phenomenology of musical experience) and faces the *symbol – grounding problem*.

- **Computational mind and music.**

i. Cognitivism.

With a necessary simplification, the cognitivist approach can be defined in terms of the *mind as a digital computer* metaphor. The underlying assumptions of this model are: (1) the brain is the information – processing machine (hardware, or (in Minsky's words) *meat machine*), while (2) the mind serves as its software, on the basis of which the brain operates. As seems to be obvious, the proposed resemblance between mind and the digital computer is stated in terms of their *functions*, that both of them are supposed to perform (Clark, 2001). This view is grounded in the developments of formal logic (in other words: the set of symbols and the ways of joining them into complex propositions, with the set of rules that specify the derivation of the new complexes of symbols from the old ones). Noteworthy, in this case, is that formal logic parallels the grammatical structure of language, as will be illustrated on the following pages of this chapter.

One may ask: how is it the above paragraph is related to human thought? It is assumed that mind is a sort of device that operates on the symbols that are *representing* the features of the external world. Thus, cognition is thought to consist of acting on such symbolic (mental) representations that are realized in the form of symbolic code in the brain - in the same manner in which they are realized in the digital computer (Varela et al., 1999). Given that, the more far – flung assumption was made: with the resemblance to the mind, the digital computers (physical symbol systems) have necessary and sufficient properties for intelligent behaviour (for instance: copying, creating and writing performed on symbolic representations). Anything that has such abilities can be conceived as being able to behave in intelligent way. Cognitivism (which recognizes the brain as operating as an input – output device) takes the sensory data (input) as encoded in symbolic representations and manipulated (by the central processor) in the mind (output). How can this be related to musical research?

- **Application of cognitivism to musical research.**

One of the most interesting applications of cognitivist theories to musical research can be found in the book entitled *A Generative Theory of Tonal Music* (1983) written by Fred Lerdahl and Ray Jackendoff. In short, with the application of the (cognitivism influenced) linguistics of the time, these authors provided a formal theory of tonal music grammar. Their approach was built on the idea that the perception and processing of language and music are parallel; both of them involve unconscious construction of abstract structures, of which the events of the (musical or spoken language) surface are the only audible parts. In the case of music, these abstract structures are constructed as accounting for one's musical understanding and musical anticipation.

Thus, the grammar of tonal music is to be understood as the set of principles that: (i) define the abstract structures of tonal music available to the listener experienced in the tonal (musical) idiom, and (ii) relate these structures to surfaces of musical piece in the idiom (Jackendoff, 1995).

As briefly stated above, *musical idioms* are the musical styles that are unique to a given culture. The grammar presented in this model is intended as an account of the experienced listener's *final state* understanding of a musical piece; the structures that the listener can attain, given the full familiarity with the idiom and with the piece and no limitations of short-term memory or attention. Their study is focused on the four types of structural relations that the listener perceives while listening to the musical piece. These are: (i) grouping structure and metrical structure of musical rhythm, (ii) hierarchies in the relative structural importance of tones, (iii) patterning tension and relaxation over time (Patel, 2008). Jackendoff and Lerdahl's theory of musical processing was intended to show how the principles of the listener's internalized musical grammar can be deployed in real time, in order to build musical representations (Jackendoff, 1995). It is noteworthy that, unlike linguistic theories of generative grammar, their theory was not about *generating* pieces of music (or marking them as *grammatical* or *ungrammatical*), but rather should be understood as a set of principles that matched pieces with their proper structures (Lerdahl & Jackendoff, 1995).

Given the existence of the "processor" in the listener's mind, Lerdahl and Jackendoff put forward the models of musical processing, which are worth explaining briefly. These models are: (a) *serial single – choice (processor),* (b) *serial indeterministic model,* and (c) the *parallel multiple – analysis model.* I am going to provide a quick overview of these models, in order to illustrate the application of cognitivist computation in music cognition.

Interestingly, the first model of musical processing described by Jackendoff (1995) faces the problem that will later on feature in criticism of the computational theory of mind. The argument explains the inadequacy of the computational account of the mind, as follows. Jackendoff described the *single – choice* model as a processor that computes only one analysis at a time for a musical piece and that when confronted with the indeterminacy among multiple analyses chooses the most likely alternative as the one to continue with. As was seen above, the cognitivist approach to the mind treated the brain as working on principles similar to the digital computer that works in sequences, applying one rule at one time. This computational limitation is known as the *von Neumann Bottleneck.* Moreover, as already mentioned, contemporary research on the work of the brain, shows it is not true that it works in such a sequentially limited way.

The second, *serial indeterministic model* is analogous to the previous model. It computes only a single analysis at a time for a piece of music. The difference is, however, that the serial indeterministic model bides its time and collects evidence without making a decision until such time as a unique analysis can be settled on. According to Jackendoff, the main idea behind this model appears in the case of the linguistic parsers (i.e. devices performing syntactic analysis). Parsers compute preliminary analyses of local parts of a sentence, but do not connect them up into a global analysis until a single correct structure can be determined for the whole.

The third, *parallel multiple – analysis model* is based on the idea that when the processor encounters a choice point among competing analyses, processing splits into simultaneous branches, each computing the analysis for one possibility. When a particular branch drops below some threshold of plausibility, it is abandoned. The branch that remains at the end of the piece contains viable structures for the piece as a whole. This model is highly unprecedented in works on language perception.

Another interesting application of the cognitivist tools to musical research can be found in Jackendoff's (1995) hypothesis of the so – called *implication – realization approach*. Implication – realization approach is an explanatory tool that captures in formal terms listeners' expectations of what is to come in the musical fragment that he is listening to. When these kinds of expectations are fulfilled by the following music, the listener feels a positive affect (satisfaction); if not he experiences negative affect (e.g. surprise or disappointment). Traditionally, this was based on the premise that in particular melodic elements set up expectations or implications about the melody's completion. Jackendoff takes the idea further. He argues in favour of what he calls "*prospective hearing*" - i.e. the theory that listeners are using the principles of musical grammar in order to assign multiple possible analyses to the fragment of music heard thus far. Let us now focus on the second variety of computational mind.

ii. Connectionism.

Perhaps the best way to describe the second variety of the computational view of the (musical) mind is to start with the way in which it differs from cognitivism. The above description of the cognitivist approach stressed the linear, serial – processing of symbol systems that operate on the forms that parallel those of language (grammar) and formal logic. Connectionism (also known as: parallel distributed processing or artificial neural networks) on the other hand operates often on non-linear functions. What is more, it bears some distant relationship to the architecture and the workings of the biological brain.

An artificial neural network is composed of many simple processors that are linked in parallel by a mass of wiring and connectivity (that is where the name of the view comes from). In the biological brain the "simple processors" are neurons (in fact, the neurons are much more complicated than connectionist units) and the connections are axons and synapses. Both neurons and connectionist units are generally sensitive only to local influences (that is: each element takes its input from a small group of its neighbours and passes outputs to another small group). Following Andy Clark's (2001) summary of this view, I will concentrate on the main differences between cognitivist symbol – processing computation and connectionism.

I have mentioned the non – linear functions on which the connectivist networks often operate. Non – linearity in this case means that the numerical value of the output is not proportional to the sum of the inputs. I take as a model of an artificial neural network "NETtalk" as provided by Sejnowski and Rosenberg in late 80's, whose task was to take written input and turn it into coding for speech (Clark, 2001). What is impressive about this network is the fact that it consists of learning algorithms (*back-propagation learning*) that allow it to begin with a set of randomly selected connection weights (e.g. information about the number of units, fixed by the designer). Next, it is exposed to large number of input patterns, for which each initially incorrect output is produced. The network consists of an automatic supervisory system that monitors this output and compares it to the correct output. After a number of attempts, the network learns an assignment of the weights that are sufficient for the effective solving of the problem. In that way, the connectionist network *encodes* the problem – solving information.

Following Varela et al. (1999) I have stated that according to the idea behind symbol – based systems, the human mind is a device that operates on symbols that *represent* the features of the external world. In that case, cognition is thought to consist of acting on these symbolic representations that are being realized in the form of symbolic code in the brain. Connectionism on the other hand operates on what Clark (2001) calls *distributed representations* (i.e., patterns that involve distributed and superpositional coding schemes). The idea behind them is that they do not consist of simple facts (e.g. about external world) alone, but also encode information about subtle similarities and differences. For example, distributed representations may contain the presence of a cat in the visual field, with small variations of the cat's orientation. Moreover, the notion of superpositional coding may be expressed in terms of pattern activation in the situation where a black panther is in the visual field (as it shares some structures of the cat activation pattern), whereas it would not activate while perceiving a white fox. Distributed superpositional coding allows the forcing of more information into the encoding system, mostly because of its ability to exploit more highly structured syntactic vehicles than words. It is also worth noticing that while the cognitivist approach is operating on symbolic representations, connectionism can be depicted as a *sub-symbolic* paradigm. Following research done by Smolensky, the essential idea behind this depiction is that whereas basic (cognitivist) physical symbol systems displayed a kind of semantic transparency (that is: familiar words and ideas were rendered as simple inner symbols) connectionist approaches introduced a greater distant between daily talk and the contents manipulated by the computational system. In that sense, parallel processing networks have an ability to operate also on *dimension shifted representations* that display fine – grained context sensitivity (Clark, 2001). It is also important to mention the recent, third generation of connectionism, the so-called *dynamical connectionism*. The difference between the previous generations of connectionism (sketched above), and the dynamical approach lies in the shift from still “static” inner symbols to a much wider range of time - involving and dynamic properties. Inspired by works of Wheeler or Port and Van Gelder, dynamical systems introduce various (more) neurobiologically realistic properties of sub-symbolic systems, such as: (a) special purpose units (whose activation function is tailored to task or domain), (b) even more complex connectivity or (c) computationally salient time delays in the processing cycles (Clark, 2001).

- **Connectionism and its application to musical research.**

Many applications of the connectionist approach in musical research can be found in the collection of papers by Peter M. Todd & Gareth D. Loy, entitled simply *Music and Connectionism* (1991). It is not necessary to deeply analyse the arguments of each article in this collection, but the short descriptions of the themes of the research described in some of them will sufficiently give us the spirit.

For instance, in their *A Neural Network Model for Pitch Perception* (1991), Sano & Jenkins examined the human ear, in order to propose a connectionist model that was able to examine the sensitivity discrepancy in pitch perception. The model was intended to concentrate on the pre-processing of the auditory stimulus, and emphasize the neural representation of pitch perception. Stages of pre-processing were utilized to reduce the level of information input to the network, and were based on back-propagation algorithms as described in the previous subsection. To provide another example, Robert O. Gjerdingen in his article *Using Connectionist Models to Explore Complex Musical Patterns* (1991) describes a connectionist model that stands for someone who (without possession of the necessary formal training) does not understand how music *works.*

Giving yet another example of connectionism applied to musical research, Niall Griffith (1995) described the use of an artificial network in the context of the model that represented (and classified) pitch use in melodies. Griffith used the Kohonen Feature Map (KFM) that developed the representations that had high visual information content (namely, it was used to visualize the key and degrees of musical scales that were derived from abstractions of musical sounds identified with pitch and interval). The representations developed with the KFM form a map that can be seen to correspond directly with the images used by musicians to represent key relations. Thus, recognising the similarities between aspects of language and music (and the linguistic and the visual analogies used by musicians), made the visual representations of musical structure provided in Griffin's paper quite interesting. Yet, there are many worries that can be raised against both varieties of the computational approach to (musical) mind. For that reason, I will conclude this chapter with a critique of these disembodied views.

(2) Conclusions.

There is a variety of problems for the computational approach to the musical mind. For instance, one of the main problems for the cognitivist approach may be recognized in the case of the limitation of the (sequences of) operations that mind can perform. As in (early) digital computers, cognitivism claims that mind is performing the operations by applying one rule at the time. This case is known as *von Neumann Bottleneck* (Varela, et. al, 1991). Yet given our contemporary knowledge about the performance of the brain, it is widely accepted that this postulation is false. The brain does not work that way.

Moreover cognitivism (as well as connectionism) can be accused of the neglect of the non – neural body (that is assumed to be *passive* (i.e., play no causal role) in (music) cognition).

The third possible problem for the cognitivist approach is concerned with the listener's musical memory. If the perceiver knows the musical piece very well, there should be no place for shock, disappointment, but only continuous satisfaction. On the contrary, in the second chapter of this thesis I will argue in favour of the alternative model of musical anticipation, as based on the outcomes of research on mirror neurons.

In general, the main problem for both varieties of computational approach to the (musical) mind lies in the question of how symbolic (or sub – symbolic) representations acquire their meanings. According to Anderson (2003), the relation between these symbols and their meaning (*sign* and the *signifier*) is rather arbitrary and thus unclear, and thereby it *enforces a kind of distance between the inner arena of symbol processing and the external world of meaning and action.* As I intend to show in the following chapter, scholars working on the embodied view have suggested that the answer to this problem lies in the reference to our non – neural bodies and their sensorimotor activity.

Furthermore, the computational approach to the musical mind does not do justice to the phenomenology of musical experience (that is: the recognition of the ways in which our interaction with music is grounded in movement, feeling and expression). Therefore, the computational research may be viewed as being quite artificial and too abstract to provide knowledge about the actual relation between humans and the world of music.

Chapter 2

In the previous chapter, I described briefly the computational approach to the musical mind. This chapter will consist of three main subsections. Firstly, (1) I introduce the enactive view on music perception that serves as a non – representational alternative to the traditional approach and is the paradigm of embodied music cognition. I will analyze prediction and anticipation in music cognition, as well as emphatic and emotional involvement with music. In the second subsection of this chapter (2) I examine Marc Leman's (representational) framework for embodied music cognition. Next, with reference to the different layers of mental representations (corresponding to Leman's view on musical semantics), I will also (3) suggest how the *symbol – grounding problem* can be solved.

- **Enactivism and music cognition.**

The enactive approach to (music) cognition, in general, argues against the notion of mental representations. The main reason for that lies in the claim that representationally based cognition has *virtually nothing to say about what it means to be human in everyday, lived situations* (Varela et. al, 1991). It is consistent with my argument against the computational approach to music cognition, accusing it of not giving justice to the phenomenology of musical experience. Before we proceed, it is important to understand the assumptions of the so – called *ecological paradigm*, on which the enactive approach is based.

The ecological paradigm is inspired by the works of J. J. Gibson (e.g. 1979), who proposed the theory of direct – perception, which holds that the duality between the organism and its environment can be overcome by the fact that the organism perceives the action – relevant properties (*affordances)* of this environment. The idea of direct – perception of music can be briefly stated by the following claims (Leman, 2008):

- Perception serves to disambiguate the stimuli in our environment. It involves prediction, and in that it is based on past experiences, rather than on logical analysis of present stimulus. Moreover,

- Perception is an emerging effect of resonance communication by means of which natural constraints and cultural constraints interact with each other. Furthermore, perception of our own actions generates internal representations of intentionality (goal – directed actions) which are associated with valence attributions.

- Perception of action of others proceeds in terms of our own ontology of intentions, and hence in terms of prediction and anticipation.

How can action – relevant *affordances* be related to music? In order to understand this, we need to refer to Joel W. Krueger's (2010) view on *musical affordances*, as follows. Musical affordances should be understood as the properties of the listener – music relation, because:

> …they are realized within the relation between a feature of the environment (e.g., particular structural qualities of a piece of music) on one hand, and a perceiver-side ability or skill (e.g. motor capacity, perceptual and affective sensitivity) enabling the pickup and appropriation of this structural feature, on the other. The relevant perceiver-side skills are what make music affordances show up as available for engagement and appropriation. But *musical affordances* are simultaneously dependent upon qualities of the music, too – hence their relational (or *interactional*) nature.

Given the relational nature of musical affordances, a musical piece *affords* engagement and appropriation. Thus – claims Krueger - musical engagement is caused by the listener's recognition of musical affordances; thus, they emerge through the dynamic interaction between the active listener and the musical piece.

For now, it is important to flag that Marc Leman's framework of embodied music cognition is also based on a (quite narrow understanding of) musical affordances. Leman's approach to the ecological paradigm (Leman, 2010a) is based on the idea that we ought to consider the acting body (rather than the mind or disembodied brain) as the centre of the mediation process. Thereby, we implicitly acknowledge the action – oriented bias of humans as a driving force behind meaning formation. As such – according to him - it allows the subjects to interpret the energetic forms in the physical environment (affordances). Given that, we may now return to the description of the enactive approach.

There are (at least) two major varieties of enactivism (Krueger, 2009): (i) *cognitive enactivism* and (ii) *perceptual enactivism.* The first aspires to provide a broad account of mind and cognition (e.g. Varela et al., 1991), while the latter focuses on questions about perceptual consciousness and subjectivity. In the following overview of the enactive approach to music cognition, I will focus on perceptual enactivism.

In his book entitled *Action in Perception* (2004) Alva Noë defined the enactive approach to perception in the following way: *"… perceiving is a way of acting. Perception is not something that happens to us, or in us. It is something we do* (…). *Perception is a kind of skilful bodily activity*". One may ask: how can cognition *work* without reference to the notion of mental representations? The idea here is that (in the vein of the ecological paradigm) cognition depends on immediate world directed interaction.

The idea of enactive perceptual experience implicates three things: (a) our bodies (or, rather, sensorimotor systems), (b) an implicit understanding of the abilities of our bodies (e.g. moving around, reaching for things and so on), as well as (c) the way in which these bodily activities (in Noë's terminology these are called: *sensorimotor contingencies)* alter our perceptual access to the world (Krueger, 2009). To provide an example, when I perceive a red tomato, I do not have a visual experience merely of the front side of it. Rather, as Noë forces us to believe, I am experiencing the *whole tomato* (including the backside of it). What is more, the qualities (colour, shape, location in space, etc.) are never wholly given in the visual experience, but – as Noë argues – they are *amodally* present in our experience (that is: it is in some *there* in my experience). How can the back of tomato be given to me without my direct perception? Noë calls this *the problem of perceptual presence* (Noë, 2008). The solution to this problem can found in our sensorimotor contingencies, namely: the backside of tomato is always given as something that can be touched or manipulated; it is directly accessible. As such, it is a matter of (musical) *affordances;* our bodily engagement shapes the character and content of consciousness, and the experience is a matter of situated, skilful coping with the environment.

- **Prediction and anticipation.**

It is fairly unclear how – without appealing to the notion of mental representations – the enactive approach can explain prediction and anticipation in the perception of music. Krueger (2010) discusses the case of an infant's *sonic expectations* (i.e., anticipation that the dissonant interval will be resolved by a consonant interval), which seems to be a most visceral and primitive aspect of musical experience of all ages. Yet, I argue, it is still vague why it shouldn't be considered in the terms of musical representations. I argue that musical prediction and anticipation should be explained in terms of research on the human mirror neuron system. Therefore, I will describe the SAME model.

Both Leman's idea (2008) and the SAME model (Overy & Molnar – Szakacs, 2006, 2009) are based on the coupling of action and perception, based on the mirror neuron system (MNS). The human MNS provides and operates on the mental representations that enable prediction. How does it work? The SAME model is based on evidence that the MNS shows sensitivity to auditory stimuli related to actions (Aziz – Zadeh et al., 2004; Buccino et al., 2005). Furthermore, given the evidence (Molnar – Szakacs et al., 2005; Zatorre and McGill, 2005) that the area of mirror neuron system in the brain may be the actual source of predictive models of upcoming events in sequential processing (a feature common to music and action), the SAME model suggests that this system can serve as a pre – condition for embodied communication and more general cognitive abilities (Overy & Szakacs, 2006).

Thus, the hypothesis behind the *Shared Affective Motor Experience* is that humans may comprehend all communicative signals (visual, auditory, linguistic or musical) in terms of their understanding of the motor action behind that signal and the intentional and emotional state of the executor. Thus, a musical phrase can be used to express the individual's semantic intention or emotional state, and a listener can understand that intended expression via perception of the *motion* of the signal (Overy & Szakacs, 2006). Parietal mirror neurons have the special property of coding motor acts as belonging to an action sequence, and predicting the intended goal of a complex action. What is more, subsets of pre-motor mirror neurons reveal audiovisual properties that are able to represent actions independently of whether they are performed, seen – or, interestingly – heard (Molnar – Szakacs, 2006; Kohler et al., 2002; Fogassi et al., 2005).

Given the above deliberations, I argue that the explanation of prediction and anticipation should be explained in terms of mental representation. Consider the following example. Marc Leman (2008) provides a study where two listeners were asked to move a joystick, while listening to fragments of *guqin*[2] music. The first listener knew the musical fragment well, but was not an *expert* in Chinese music. The second listener was entirely new to that type and the particular piece of guqin music presented to him. They were requested to move the joystick in accordance with the music. Meanwhile, their corporeal articulations were monitored.

The observations made by Leman in this case study suggest that listeners are able to attune their elementary movements to characteristics of the sound energy (e.g. onset or pitch). In the first (experienced) listener, the corporeal resonance behaviour had a more pronounced predictive character – the onset of the listener's elementary movement often agreed with the onset of the guqin player's elementary movement.[3] Moreover, the speed of the movement can be related to the speed of the pitch change; whether the pitch went down or up affected the decrease or increase of the listener's movement velocity. Leman claims that these data provide evidence for his hypothesis, that the listener's perception is predictive in the short term. As such, he claims, listeners are capable of grasping music as an intended moving form. In doing this, the movements take the character of an action (goal – directed movement) whose intentionality is in turn projected onto the sound energy, and, by extension, onto the movements of the listener. Thus, claims Leman (2008), player and listener are able to establish a relationship of mutual information exchange at the level of action. The above data also suggest that music perception involves motor attuning component, which helps the listener to *read the minds* of composer and the performer. Thus, Wilson and Knoblich (2005) argue that the predictive behaviour that is made possible by motor attuning facilitates perception. This predictive character of corporeal attuning is also understood to be the expression of fundamental and automated patterns of communication (Bargh and Chartrand, 1999) and the essence of human intelligence (Hawkins and Blakeslee, 2004).

[2] The guqin is a Chinese plucked string instrument that is played by plucking the string with the right hand and producing different pitches by manipulating the string with the left hand. It differs from a guitar - for instance -because it does not have frets.

[3] Term „elementary movement" is to be understood as: the movement of a part of the body between two points in space (Leman, 2008)

- **Musical empathy and joint – sense making.**

Enactive ideas on musical empathy and joint – sense making are not only thought – provoking, but may serve as a possible way of explaining prediction and anticipation. I intend to present the embodied view on empathy first, and then compare it with enactive ideas.

Leman (2008) claims that embodied empathy with music draws on shared and distinct representations of perception and action. Moreover, empathy involves a regulatory mechanism by which the subject keeps track of the self in relation to music. Such regulatory mechanisms are assumed to play an important role in social contexts where the subject tends to adapt his or her expressive behaviour to the expressive behaviour of other social agents. Thus, there are strong indications that music can be conceived as a "virtual" social agent, and that listening to music can be seen as a socializing activity in the sense that it may train the listener in social attuning and empathic relationships.

This view is consistent with Joel W. Krueger's (forthcoming) view on musical empathy. Krueger claims that music dynamically motivates even the earliest infant – caregiver interactions. What is more, various studies quoted by Krueger seem to show that infants respond to music in a way that suggests they are best viewed as *active perceivers.* Thus, consistent with Leman's idea described above, it is suggested that music can play a crucial role in guiding early communicative exchanges (that are important in the social context of sharing musical experiences).

However, the main difference between Leman's and Krueger's propositions lies in terms of representations. While Leman is committed to the ways in which this kind of involvement with music draws on representations, Krueger argues that music serves as an immediate (non – representational *direct perception*) and non – inferential way that allows emphatic involvement. In the later, enactivist view, the character of shared musical experience can be defined in terms of dynamic processes of (Krueger, forthcoming):

- joint *sense – making*, enacted via temporally – extended patterns of
- skilful engagement with music that are
- synchronically and diachronically *scaffolded* by the surrounding environment.

The first point is especially necessary for my argument, as the notion of *joint sense – making* seems to be consistent with the ideas that underlie Leman's *collaborative semantics* as described below. Krueger understands this notion as a *cooperative process of meaning construction and appreciation.* It is noteworthy that Krueger is committed to a non – representational view, where sense making is understood as a *relation and affect – laden process by which an organism actively generates a meaningful world* (e.g. Varela et al, 1991).

Yet, both the emphatic involvement with music and the joint – sense making can be described in a better way by the (representational) SAME model. According to Overy & Szakacs (2006), emotions and action are intertwined on several levels, and this affective coupling may provide the neural basis of empathy (Carr et al., 2003; Leslie et. al., 2004). In other words, empathy (especially the aspect that requires no intermediary cognitive process but automatic and immediate *motor identification* (Gallese, 2003)), it is suggested, can be based on the mirror neuron system that is involved in the production of emotional representations that bind us together with others. Furthermore, Overy & Szakacs (2006) provide also the following evidence:

> Two regions, including the posterior inferior frontal gyrus and the anterior insula are commonly activated during musically evoked emotional states (Koelsch et al., 2005, 2006; Menon and Levitin, 2005). These two structures may hold the key to understanding how the brain uses a simulation mechanism to represent emotional states evoked by musical experience. As described above, the posterior inferior frontal gyrus (BA 44) is the frontal component of the human fronto-parietal mirror neuron system. With its ability to link perceptual and behavioral representations of a stimulus during the perception of emotionally arousing music, the mirror neuron system may simulate an emotional state in the listener (Gridley and Hoff, 2006).

The evidence for the functions of the mirror neuron system that are related to the topic of this thesis may suggest that Leman (while being committed to the representational view) should take into greater consideration the brain-side of musical interactions. Yet, the reference to the idea of the SAME model (and the evidence behind it) seems to explain ideas of empathy and joint sense – making more clearly and convincingly than the enactive view. Let us now consider the case of emotional involvement with music in a greater detail.

- **Emotions and music.**

Given the fact that the proper recognition of *musical affordances* enables us to theorize about emotional response to music, Kruger (2010) argues for a form of listening (based on the historical and cultural assumptions attached to musical pieces) that can be described as being a source of emotional power and expressivity. As such, we immerse ourselves within the immediacy of the phenomenal richness of music as a basis for the emotional content of a musical piece (Davies, 2001). In the following deliberations, I will refer to this way of listening to music as *transparent listening.* The *transparency* of it lies in the fact that while experiencing music we are not, in general, considering the various acts that led to the production of musical sound; as such then, experience of music is *detached from the circumstances of its production* (Hamilton, 2007). The emergence of emotional response to (affordances in) music can be found in studies on infants that, despite the lack of conceptual musical education, have preferences for consonant musical intervals, rather than on dissonant ones (Kruger, 2010; Trainor and Heinmiller, 1998). On the basis of such research outcomes, Krueger (2010) hypothesizes that – as infants are engaged with music, and tend to reveal emotional responses – music allows emotional self – organization. In other words: music is perceived as an affordance – laden structure that furnishes the possibility of emotion construction and regulation. Yet, for this to happen, infants must be engaged with music in a sensorimotor way, which leads us back to the assumptions of perceptual enactivism.

On the other hand, the fundamental relation between embodiment, bodily actions and emotions can be captured in the idea that to *be moved* refers to the feeling as if one is in a different position in regard to one's situation (Gibbs, 2006). Starting from the classic example, the metaphorical ideas that *good is up* and *bad is down* (Lakoff & Johnson, 1980, 1999), people usually experience their emotions as movements toward something in themselves. Whether we are *moved by a song* or, for example, *feeling like being on top of the world*, we tend to experience our emotions as movements towards something in ourselves. It is noteworthy that we need to recognize the relationship between emotional experience and bodily changes (beyond those conveyed by the face). Research in neurophysiology focused attention on the anatomical structures underlying emotions. The results of various studies (Gibbs, 2006) are consistent with the idea that people do believe that different emotions are associated with different localized bodily symptoms. Moreover, some researches have suggested that a particular body movement may induce specific emotions. To provide an example: empirical research investigated people's recognition of emotion from gait (Montpare, Goldstein & Clausen, 1987). Research outcomes suggested that emotions such as happiness, sadness, anger and pride in fact be revealed in gait. Styns et. al. (2007) did an interesting study into the intricate relationship between music and bodily movement, particularly on how music influences the way humans walk. Other research has shown that expressive music can induce subliminal facial expressions in listeners (Witvliet and Vrana, 1996).

The representational, embodied view is thus consistent with the SAME model. All of the outcomes of the research presented in this subsection are consistent with the suggestion that the perception of emotion in music may arise at least in part from its relation to one's physical posture and gesture (Jackendoff and Lerdahl, 2006; Davies, 1994). Let us now focus on the framework of embodied music cognition provided by Marc Leman.

(2) Marc Leman's framework of embodied music cognition.

The assumption behind Marc Leman's model (2008) is that humans interact with music on the similar basis on which they pursue social interactions, namely: they tend to understand each other's intentional actions. In that vein, these intentional actions are encoded in *moving sonic forms* (i.e., patterns in physical energy), that allow subjects to decode them in terms of interaction with music and the understanding of musical intentionality. The underlying process that allows this understanding is so-called *behavioural resonance*. Leman distinguishes a variety of these resonances: from a purely passive response to suddenly heard loud music, to tapping the beat or empathic involvement with the music (i.e., behaving as if feelings engendered by the music are shared).

Roughly, behavioural resonance can be defined as: measurable corporeal expressions and articulation of subjective feelings, moods, flow experiences (e.g. Csikszentmihalyi, 1990) and feelings emerging in response to perceived music. As I intend to argue below, Leman's idea of behavioural resonance can be understood in terms of a deep connection between mirror neurons (as provided in the SAME model).

Leman (2008) argues that his framework offers new perspectives for musical research. For the purposes of this chapter, I will concentrate on three aspects: (i) the subjective aspect of musical experience and musical gestures, as well as on his EMC model in terms of (ii) musical perception (with the account on Leman's ecological perspective) and (iii) musical communication.

- **Subjective experience of music and musical gestures.**

As Leman (2008) argues, one of the most important consequences of taking EMC into account in musical research is the suggestion that empirical experiments should consider their subjects as an *active* contributor. Thus, empirical experiments in the field of music cognition should be designed in a way that does not prevent active involvement with the music. In the previous chapter, I argued that embodied music cognition is based on the second – person descriptions (that is: a form of description that involves intentionality but is less linked with interpretation). As such, the EMC assumes that subjective experiences of music are expressed directly by means of measurable corporeal articulation (gestures).

What are musical gestures and how they can be studied? In general, *musical gestures* should be understood in terms of a meaningful combination of movement and music. In this vein, in the article *Musical gestures, concepts and methods in research* Alexander Refsum Jensenius et. al. (2010), there is a discussion about three different viewpoints from which musical gestures can be studied, that is (a) the subjective phenomenological level, (b) the objective, and (c) the communicative (phenomenological, biomechanical and functionalist) viewpoint. In short, the first of these viewpoints focuses on the descriptive aspects of gestures – such as describing a gesture in terms of its speed, the amount of space that it uses, or the frequency range. The second, objective view point focuses on the conditions for gesture generation, such as various biomechanical and motor control constraints. The last point of view is focused on the purpose of movement or action in a certain context (e.g. whether it is sound – producing, or sound – modifying). For now, this brief description of musical gestures is sufficient.[4] In the following section[5], I will examine Leman's view of the role and the way in which gestures work in a greater detail.

▪ Embodied music perception and the ecological paradigm.

Leman's (2008, 2010a) framework of embodied music cognition is based on the so – called *ecological paradigm*. I have already described the idea of direct – perception above. Yet, the notion of *intentionality* is also important in music perception. Leman understands the term (embodied / corporeal) *intentionality* in the following way. Corporeal intentionality is an emerging effect of action / perception couplings; its underlying engine can be defined in terms of the sensorimotor system. This system turns the physical energies of music into an imaginary world of objects, that have qualities, balances and intentions and vice versa. As such, corporeal articulations are to be understood as the expressions of these processes, and, in consequence, as turning physical energy into an action – relevant and action – intended ontology (Leman, 2008). Thus, *corporeal intentionality* differs from the traditional, cerebral idea of intentionality.

Cerebral intentionality is based on the exploration of speculative pursuit of possible interpretations – in other words, it is interpreting the source of intentions attributed to music. On the other hand, the essence of corporeal intentionality lies in the articulation of moving sonic forms, with a particular emphasis on movement in relation to behavioural resonances of the human body.

The idea is that we perceive music in terms of our own actions, so that perception is claimed to induce a transition from musical audio streams to the trajectories of auditory sensing in the brain, to motor trajectories our bodies. Musical movements (gestures/articulations) are therefore executed through movements of the head, hands, legs or the whole body.[6] In that vein, when the music is violent and aggressive we spontaneously tend to perform movements that go along with violent and aggressive sounds.

Leman's framework also offers an interesting account of the aesthetic pleasure that we gain by listening to music. According to sensorimotor theory, claims Leman (2008), the

[4] More detailed description of various aspects of musical gestures can be found in Godøy & Leman's book *Musical Gestures* (2010).
[5] See the subsections of this chapter focused on embodied music perception and embodied music production.
[6] See the quote from Eric F. Clarke in the introductory chapter.

senses provide the information to create action – related models of our environments and these models allow us to anticipate the consequences of our actions. An interesting example of the application to music can be found in Repp and Knoblich's (2004) paper on how pianists are able to recognize their own performances out of a number of others. It is suggested that the better we know music (or, the better we are acquainted with *musical idioms*), the more we are inclined to predict on the basis of movement. The overlapping of performed and perceived actions may thus be the origin of aesthetic pleasure.

It is also important to notice that perception of music is a multisensory activity, namely, involving the multisensory integration. The integration of senses (serving as the basis for sensorimotor theories) makes it possible to experience the music not only as pure sound, but also an association of visual, tactile and motor experiences. Given that, we also need to focus on the so-called *corporeal / embodied attuning* and listeners' *empathy with music.*

Singing along with music can be pointed out as an example of corporeal attuning. Attuning brings the human body into accordance with music, and may be considered as a form of synchronization in the sense that it aims at being as much as possible in harmony with features of the moving sonic forms of music; it is a kind of *playing together with music*. As such, attuning aims at melody, harmony, rhythm and timbre, or patterns related to expressiveness, and feelings. It is based on the idea that the world is perceived in terms of cues that are relevant to the subject's action – oriented ontology (Leman, 2008). Let us consider the example of vocal attuning, a technique based on low-voice singing and humming along with perceived music. Pilot research on that topic has shown that a structural feature of music as complex as tonality can be captured by means of vocal attuning. Vocal attuning involves the subject in a pleasurable state of behavioural resonance, invoking signification based on corporeal articulations.[7] That outcomes lead us to the, so-called, *embodied empathy with music.*

Moreover, in Leman's (2008) pilot study (that is simplified for the sake of argumentation) listeners were asked to express perceived music through corporeal articulations. The hypothesis was that listeners would be able to improve their performance as a result of implicit learning. They were asked to move their arms along with three short pieces of Chinese guqin music, with which they were unfamiliar, and the music didn't have a real beat. The results of the research suggested that embodied listening is prone to implicit learning.

[7] For more, see: Chapter 5 in: Leman, 2008.

- **Embodied music production.**

Leman's framework is also interesting from the music production point of view. In this subsection, I intend to describe Leman's view of interaction with acoustic instruments, as well as the role of gestures in music production. A nice example of human interaction with acoustical instruments can be found in the case of *guqin*-playing (Leman, 2008). The reason for picking this acoustic instrument is the fact that the player's subtle gestural control is direct. In contrast with guitar or violin, guqin involves no *mediating device* (such as: frets or bow). The movements that contribute to the formation of tone can be described as constituent *tone gestures.*

As these tone gestures can be related to the corporeal attunement of listeners, we need to focus on them in greater detail. First, *elementary movements* in the following study can be defined as the movement of a body part between two points in space. In Leman's (2008) study, markers were attached to particular parts of subject's bodies (e.g. finger, wrist, elbow and head), and were monitored with the camera system. Given the fact that playing music is a continuous activity, the camera recorded a continuous displacement from which a continuous velocity pattern on the sonograph was derived. A main conclusion of his study was that guqin tones were generated by a complex interplay of very efficient corporeal articulations, consisting of preparatory (counter)movements and effective (control) movements which together defined an action that lead to the production of tone. As such, the characterization of music as moving sonic forms captures the idea that the sound structure encodes aspects of the player's actions. The sound structure reveals the encoding of biomechanical energy as sound energy, behind which are movements aimed at producing tones.

(3) Musical semantics, layers of (musical) mental representations and the symbol-grounding problem.

As frequently outlined above the traditional (computational) view of (music) cognition relies on the idea of (symbolic / sub – symbolic) representations of the external world which are the basis of (input – output) computational processing. The problem of how mental representations acquire their meaning (*symbol – grounding problem)* faced by such an approach concerns also musical processing. At the end of chapter 1 we have recognized how this problem can be solved with reference to the body; but how exactly does it relate to music? What is the basis for musical semantics? To answer these questions I will refer to Marc Leman's paper entitled *An embodied approach to music semantics* (2010a). In this paper, mental representations (in the process of the formation of musical meaning) are analysed from two points of view: as focusing on (a) sharing expressions (joint activity that is driven by music), and (b) movement based perception.

Given that, Leman distinguishes four layers of mental representations that correspond to: (1) *Representational* musical semantics, (2) *Referential* musical semantics, (3) *Corporeal* and (4) *Collaborative* musical semantics. Let us now examine these levels in greater details.

- **Representational musical semantics.**

The idea of musical grammar presented in the preceding chapter serves as a good example of representational music semantics. As Leman (2010a) nicely describes the representational semantics approach, it can be characterized as:

> (i) Focused on the representation of structures in music (…). (ii) Focused on the mechanisms of pattern-processing, either computational or as brain activity (…). (iii) Focused on causal explanations and predictions, in the sense that the pattern-processing is assumed to engender changes in systems that are linked with meaning (...).

Given that, the neglect of the non – neural body and its activity outlined in the previous chapter becomes clear. Let us consider the idea of mental representations in representational semantics in relation to the notion of anticipation (Leman, 2010a). Mental representations are seen as carriers of structure and deviations from expectancies based on these representations as sources of meaning. Thus, the actual meanings of mental representations result from their structures (Leman, 2010a).

Once mental representations are acquired, they function as models for the anticipation of new information. Depending on what is expected, this may lead to tension, surprise and various semantic responses (Huron, 2006). Recent brain research suggests that tonal – harmonic sequences may engage the brain in a tension – resolution process, which can be interpreted as the semantic processing of music (e.g. Patel, 2003).

While the traditional view of (musical) mental representations seems to be facing the *symbol – grounding problem* explained in the previous chapter, Leman suggests that representations are *calibrated* by body movement.[8] He claims that motor patterns have to be followed for effective perception of musical structure. In that sense, motor patterns should be conceived as executing structures that regulate mental representations, or form schemata in terms of executive or corporeal representations. Leman moreover suggests that musical semantics can be rooted in the dynamics of the human body and in activities that involve social communication.

- **Referential musical semantics.**

The second approach to musical semantics is based on the following idea: sonic patterns function as pointers to musical meaning. In language, sound and meaning are closely related to each other, but this relationship in sonic patterns and meaning is more loosely defined. For that reason, Leman (2010a) provides the following distinction between (a) *extra – musical meaning in music* and (b) *intra – musical meaning in music.*

The first (*extra – musical* meaning) can be defined as meaning that lies outside music. In other words, as the sonic pattern – meaning relation is loosely defined, inexperienced

8 See: the third chapter of this thesis (section: *How the body calibrates representations*?).

listeners can be enabled to recognize or form its associations with the real world. To provide an example: an inexperienced listener may have problems recognising the extra – musical reference to the *expression of passion* as can be found in Berlioz' *Symphonie Fantastique,* and so on.

On the other hand, *intra – musical* meaning can be conceived as a kind of association between several structural components due to their commonality. It can be observed while thinking of a particular melodic or rhythmic theme (that is repeated and slightly varied at different moments during the musical piece). Another example is the appearance of *musical quotes* (fragments of somebody else's musical work) in a particular musical piece. While the identification of intra – musical references may not be the main concern of most listeners, it is important from the viewpoint of musical analysis and education.

Given that, *referential* semantics can be either hermeneutical (extra – musical meaning) or structural (intra – musical meaning). The hermeneutic approach can be characterized as a search for metaphors. I have described in brief Lakoff & Johnson's (1980, 1999) ideas of embodied metaphors, this – combined with musical cognition – may have influenced Leman's idea of corporeal / collaborative semantics, as follows.

- **Corporeal musical semantics.**

The last type of musical semantics, and perhaps the most important for the following deliberations, can be described as looking at meanings as being generated through the mediating activities of the human body (Leman, 2010a). As such, it is consistent with the enactive approach to musical meaning and the research on mirror – neurons (especially the SAME model) as described below. Corporeal music semantics is based on the idea that musical stimuli can be perceived and become *calibrated* (acquire meaning) through bodily movements (gestures), rather than through mental construction and imagination. Consider an example: a discotheque dancer. Leman claims that the dancer uses music as an *activator* to move, and through this movement *music acquires its meaning.* This meaning cannot be obtained just by having a mental representation of dancing. The second claim made here is that our motor capacities constraint our perception. Moreover, embodied engagement with music *calibrates* the content of mental representations of music. So Leman argues that the traditional view of mental representations is insufficient to provide musical meaning, unless they will involve the deployment of the human body. In Leman's words (2010a), the human body, understood as the mediator:

> (i) may adjust the anticipations that are provided by mental representations, (ii) may engage in corporeal presentations (unfolding of structural corporeal articulations in space and in time) that derive their meaning from the sole fact that they are effectuated. The deployment of the mediator can be perceived by the subject (the feeling of what happens), and it addresses communication channels that can be opened up to other subjects in a social setting. It thus forms an essential part in understanding how mental representations function, and how meaning formation may work.

How then is the meaning of the mental representations acquired? Leman (2010a) claims that the human body is a determining fact in meaning formation processing, due to the fact that all of the meanings derive from an action – oriented bias that is grounded in the deployment of the mediator. Thus, he claims, that it is justified to speak about corporeal semantics of music, where meaning formation processes are linked with music and corporeal articulations. Moreover, corporeal meanings are significations that result from the spatial and temporal deployment of the body in response to music; in that sense the meaning is determined by the body activity.

- **Collaborative musical semantics.**

Collaborative semantics covers meaning that emerges from musical practices in a social context. It should be understood as an extension of the notion of corporeal semantics towards the social domain. It draws on the idea that corporeal semantics involves a deployment in space and time through articulations that have their origin in the social domain.

It is noteworthy that embodied expressions – more than mental representations (of the representational semantics) can be seen as genuine vehicles for interaction – based meaning formation. Music serves as an excellent opportunity to apply personal expressions to moving forms, both for senders (musicians) who encode their expression through a corporeal articulation into the musical moving forms, and for receivers (listeners) who perceive these as expressive patterns. Moreover, according to Leman (2010a), the latter can afford behavioural resonances that underlay social bonding. Thus, these patterns can be shared with other people and collaborative semantics is exactly about how meaning may emerge *naturally* from this interaction, by just doing it. As such, Leman's model of corporeal semantics is complementary to the SAME model, as it will be argued below.

For now, it is also important to outline that in corporeal semantics, the meaning can be seen as the emerging effect of social musical interaction. Using Leman's (2010a) description, this type of musical semantics can be briefly characterized as being:

> (i) multi-modal, because it involves audio, visual fields, movement and other modalities, (ii) mediated, because it focuses on the mechanisms that transfer energetic forms to experiences, or experiences to energetic forms, for multiple subjects, (iii) combined objective/subjective, involving the subjective involvement in joint attentional frames, joint intentionality, common ground, expression, gesture, social bonding aspect in collaboration, and exchanges of energetic patterns via multiple channels.

- **The symbol-grounding problem.**

To conclude this section, let us reintroduce the layers of mental representations that can be found in the foregoing models of musical semantics. In the first type of semantics (*representational*), we have seen the question of how mental representations acquire their meaning. It forces us to recall the *symbol – grounding problem* introduced in the previous chapter.

The solution to this problem can be found in reference to the non - neural body, as in the arguments by Gibbs (2006) or Leman (2010a), which I take to be complementary. The second approach to musical semantics, *referential* was characterized as the search for metaphors. Given that fact, it can be suggested that *extra – musical meaning* is to be acquired in reference to associations with real – world meaning, one can think about musical pieces that can be interpreted as, for instance, sound illustration of running or crying and so on. The third (*corporeal*) and the fourth (*collaborative*) musical semantics are of special importance to the following arguments. Representations in these two models are seen to acquire their meaning by direct relation to bodily activities, and social bonding based on the perception of one's corporeal articulations.

Chapter 3

In this chapter I intend to answer the question as to whether a commitment to mental representations is necessary in music cognition. In the previous chapters I described the representational, computational view of music cognition. The problem with this approach was the *symbol – grounding problem,* which – as I argued – can be solved by reference to the non – neural body. Yet, in the second chapter of this thesis I presented briefly the enactivist challenge to representational (embodied) views on music cognition. Thus, we are led to the question of whether reference to the notion of mental representations is necessary in music cognition.

In this chapter I attempt to answer this question in the following steps. Firstly, I intend to (1) examine Marc Leman's understanding of the notion of mental representations. I will show how – in musical semantics – mental representations acquire their meaning. Secondly, we will see (2) what makes the case of mental representations, and what are the differences in understanding of the notion of *musical affordances* in the embodied (Leman) and the enactive (Krueger) views of music cognition. Next, (3) I intend to show how – in Leman's view – mental representations are *calibrated.* Finally, I will (4) conclude this chapter by answering the question about the necessity of mental representations.

- **Marc Leman's understanding of the notion of mental representations.**

The way in which Leman understands the notion of mental representations is fairly unclear in his main work - *Embodied Music Cognition and Mediation Technology* (2008). Luckily, he straightens it up in his later article entitled *An embodied approach to music semantics* (2010a). In this article, Leman argues in favour of maintaining the notion of mental representations in cases when the body interacts with music. As such, it is opposed to the enactive ideas presented in the previous chapter of this thesis. Yet, Leman claims that these kinds of representations are not sufficient for his idea of the embodied approach to (corporeal) musical semantics. Thus, he argues (Leman, 2010a) that if we are to preserve mental representations in music cognition, they should involve the role of body as mediator, that:

> (i) may adjust the anticipations that are provided by mental representations, (ii) may engage in corporeal presentations (unfolding of structural corporeal articulations in space and in time) that derive their meaning from the sole fact that they are effectuated. The deployment of the mediator can be perceived by the subject (the feeling of what happens), and it addresses communication channels that can be opened up to other subjects in a social setting. It thus forms an essential part in understanding how mental representations function, and how meaning formation may work.

Given a proper understanding of the connection between the human body (mediator) and mental representations, their meaning can be *calculated.* Mental representations (understood as information – bearing structures) on their own are insufficient for the process of computation of meaning. Thus, according to Leman, the traditional ideas of mental representations (introduced in the first chapter of this thesis), ought to be broadened by an accurate understanding of the meaning – formation contribution of the non – neural body. In other words – claims Leman (2010a), corporeal meaning is a signification that results from the spatial and temporal deployment of the body in response to music. Thus, meaning is determined by bodily activities. Given that, let us examine the case of embodied mental representations in greater detail.

- **What makes the case for embodied mental representations?**

Leman's idea of embodied mental representations is analogous to works by Raymond W. Gibbs (2006) that I will describe shortly in order to make this case more clear. Gibbs view, although not necessary concerned with music cognition, is based on the idea that mental representations acquire their meaning in relation to the non – neural body and its activities.

Traditionally (in the disembodied sense), concepts are to be understood as: stored mental representations that enable people to identify the objects and events in their environment. As such, concepts are basic rules / categories that are not only defined by their relation to an external object, but also enable representations to acquire their meanings. In other words, the concepts are assumed to be mentally represented in terms of *prototypes,* with degree of category membership determined by similarity to the prototype. Raymond W. Gibbs provides the following example: sparrows are closer to the prototypical bird for Americans than are penguin or ostriches, and this makes it easier for people to verify statements such as “A sparrow is a bird” than those such as “A penguin is a bird” (Gibbs, 2006). As can be easily seen, this description matches the idea of amodal symbolic representations in cognitivism. Moreover, it is important to add, these symbols are language – independent, context – independent and disembodied. Following Gibbs (2006), there is a variety of problems with the traditional view of concepts. For instance, evidence shows that people represent different properties quite differently in different contexts. These findings suggest that the prototypes are closely tied to individual contexts and are not necessarily abstract representations that emerge from specific instances of a given concept.

On the other hand, cognitive linguistics embraces the embodied view on the conceptual basis of mental representations. Assuming the language – mind – body linkage, in various works it is claimed that that our concepts are grounded in (and structured by) various patterns of our perceptual interactions, *bodily actions* and manipulations of objects (e.g. Lakoff & Johnson, 1999). Also noteworthy, interesting accounts of specific (physical and embodied) dynamic forces that underpin our abstract concepts (e.g. Talmy, 2000) can be found. We tend to understand them in terms of our bodily experiences (for instance: “I have been *forced / pushed by* (abstract) X to do Y”). These patterns (called *image schemas*) emerge throughout the sensorimotor activity of our bodies (Gibbs, 2006). These remarks serve as a basis for the later idea of the corporeal musical semantics, as examined in the previous chapter. For now, it is important to state once again that the embodied view of mental representations assumes that they are grounded in (and structured by) various patterns of our perceptual interactions, *bodily actions* and manipulations of objects.

A possible worry for Leman’s (2008, 2010a) ecological view is that his understanding of musical affordances (as accounting for bodily actions / expressions of musical interaction) is too narrow. Thus, I argue that Krueger’s (2010) enactive view deals with this notion in a better and more solid way. Krueger assumes that affordances are realised in the relation between (a) features of the environment (in the case of musical *structures* that are built in a way that allows the perceiver to pick up the information and appropriate it in order to behave in a certain way (e.g. dance to it) in response), and the (b) perceiver. Leman, on the other hand, seems to overlook the *structural* feature of the affordances, and recognize only the fact that they are to accounting for gestures. Furthermore, his account can be accused of not recognizing the affordances on the side of musical performer (feedback) that allows engagement in musical production.

- **How does the body calibrate mental representations?**

The basic idea behind Leman's and Gibbs' argument is that mental representations ground their meanings in the sensorimotor activities of our bodies. Thus, Gibbs' idea of mental representations as being grounded in the body is consistent with Leman's idea of (body as) mediator. To illustrate that fact, in the following subsection I intend to explain Leman's idea of mental representations as being *calibrated* by body. As we have seen in the previous chapter, the *corporeal* (and therefore also the *collaborative*) musical semantics originates from the idea that (musical) meaning is not "in the head", rather, it emerges through (bodily) mediated relationship between mind and the energetic forms of music (so – called *moving sonic forms,* that allow subjects to decode them in terms of interaction with music). Thus, meanings cannot be calculated (computed) from structures, but rather, from bodily actions (e.g. sensory – motor interactions with physical objects). Moreover, body movement may be largely driven by sensory feedback, and therefore the *feeling of what happens* (Damasio, 2000) occurs via continuous *calibration* (sensory motor feedback).

- **Are mental representations necessary?**

As seen above, both Krueger and Leman use notions of direct – perception. The main difference between their views lies in the question whether to preserve the notion of mental representations of music. Krueger – in his enactive view – argues against the necessity of mental representations in music cognition, while Leman, on the other hand, claims that this notion should not only be preserved, but also expanded by the acknowledgment of the non – neural body (mediator) and its *active* contribution to the process of musical interaction. This begs the question of whether *throwing* mental representations out of music cognition is not too radical a move?

In the previous chapter, we saw that most of the ideas of the enactive view of music cognition can be easily accommodated by the representational research on mirror neurons (SAME model). Furthermore, Leman's idea of embodied mental representations of music seems to solve the problems of traditional, computational (musical) mind. Given that, I view the enactive view on music cognition as being too radical, and giving no predominating reasons for rejecting the notion of mental representations of music and its explanatory value. The notion of (embodied) mental representations of music should be preserved.

Conclusions

The aim of this thesis was to argue in favour of the embodied music cognition paradigm, that was opposed to traditional (disembodied) theories of musical mind. I have, correspondingly: (1) examined computational music cognition, while focusing on its *symbol-grounding problem*, (2) presented and discussed Marc Leman's framework of embodied music cognition that serves as a response to the problems of computational musical mind, (3) answered the question, whether mental representations are necessary in music cognition. I have argued that they should be preserved, and that the enactive ideas (of throwing mental representations out of music cognition) are too radical.

Bibliography

Anderson, M. L. (2003). Embodied cognition: A field guide., in: Artificial Intelligence 149 (1): 91-103.

Aziz-Zadeh, L., Iacoboni, M., Zaidel, E., Wilson, S., Mazziotta, J. (2004). *Left hemisphere motor facilitation in response to ma manual action sounds*, in: The European Journal of Neuroscience, 19,2609–12.

Bargh, J. A. & Chartrand, T. L. (1999). *The unbearable automaticy of being*, in: American Psychologist, 54 (7), 462 – 479.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V.,Rizzolatti, G. (2005*). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioural study*, in: Brain research. Cognitive brain research, 24, 355–63.

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*, in: Neuron, 42, 323-334.

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). *Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas*, in: Proceedings of the National Academy of Sciences, 100, 5497-5502.

Clark, A. (2001). *Mindware: an introduction to the philosophy of cognitive science*, Oxford University Press.

Clark, A. (2008). *Supersizing the mind: embodiment, action, and cognitive extension*, Oxford University Press.

Clarke, E. F. (2005). *Ways of listening: an ecological approach to the perception of musical meaning*, Oxford University Press.

Csikszentmihalyi, M. (1990). *Flow: the psychology of optimal experience*. New York: Harper & Row.

Damasio, A. R. (2000). *The feeling of what happens: body and emotion in the making of consciousness*.

Davies, S. (1994). *Musical meaning and expression*. Ithaca and London:Cornell University Press.

Davies, S. (2001). *Philosophical perspectives on music's expressiveness.* in P. N. Juslin & J. A. Sloboda (Eds.), Music and emotion: Theory and research (pp. 23–44). Oxford: Oxford University Press.

Fogassi, L., Ferrari, PF., Gesierich, B., Rozzi, S., Chersi, F., Rizzolatti, G.(2005). *Parietal lobe: from action organization to intention understanding*, in: Science, 308, 662–7.

Gallese, V. (2003). *The manifold nature of interpersonal relations: The quest for a common mechanism*. Philosophical Transactions of the Royal Society B: Biological Sciences, 358, 517-528.

Gibbs, R. W. (2006). *Embodiment and cognitive science*, New York: Cambridge University Press.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gjerdingen, R. (1991). *Using Connectionist Models to Explore Complex Musical Patterns*, in: Todd, P. M. & Loy, G. D. Music and Connectionism, MIT Press.

Godø*y, R. I. & Leman, M. (2010a). Musical gestures: sound, movement, and meaning*, Routledge.

Godø*y, R. I. & Leman, M. (2010b). Editorial Preface, in:* Musical gestures: sound, movement, and meaning, Routledge.

Gridley, M.C., Hoff, R. (2006). *Do mirror neurons explain misattribution of emotions in music?*, in: Perceptual and motor skills, 102, 600–2.

Griffith, N. (1995).*Connectionist visualisation of tonal structure*, in: Artificial Intelligence Review 8: 394 – 408, 1994 – 5.

Hamilton, A. (2007). *Aesthetics and music*. New York: Continuum.

Hawkins. J. & Blakeslee. S. (2004), *On intelligence.* New York: Times Books.

Hurley, S. & Noë, A. (2003). *Neural Plasticity and Consciousness*, in: Biology and Philosophy 18: 131 – 168.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*, MIT Press.

Jackendoff, R. (1995). *Languages of the mind: essays on mental representation*, MIT Press.

Jackendoff, R., Lerdahl, F. (2006). *The capacity for music: what is it, and what's special about it*?, in: Cognition, 100, 33–72.

Jensenius, R. J. et al. (2010). *Musical gestures, concepts and methods in research,* in: Godøy, R. I. & *Leman, M. (*2010*).* Musical gestures: sound, movement, and meaning, Routledge.

Koelsch, S., Fritz, T., von Cramon, D.Y., Muller, K., Friederici, A.D. (2006). *Investigating emotion with music: an fMRI study*, in: Human Brain Mapping, 27, 239–50.

Koelsch, S., Siebel, W.A. (2005). *Towards a neural basis of music perception*, in: Trends in Cognitive Science, 9, 578–84.

Kohler, E., Keysers, C., Umilta, M.A., Fogassi, L., Gallese, V., Rizzolatti, G. (2002). *Hearing sounds, understanding actions: action representation in mirror neurons*, in: Science, 297, 846–8.

Krueger, J. W. (2009) - *Enacting Musical Experience*, Journal of Consciousness Studies, 16(2-3), 98-123.

Krueger, J. W. (2010) – *Doing things with music*, in: *Phenomenology and the Cognitive Sciences.*

Krueger, J. W. (forthcoming), *Empathy, enaction, and shared musical experience*. in: The Emotional Power of Music: Multidisciplinary Perspectives on Musical Expression, Arousal and Social Control, Oxford University Press.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Lakoff, G. & Johnson, M. (1999), *Philosophy in the flesh.* Cambridge University Press.

Leman, M. (2008). *Embodied music cognition and mediation technology* ,MIT Press.

Leman, M. (2010a), *An Embodied Approach to Music Semantics*, in: Musicae Scientiae, discussion forum 5: 43 – 67.

Lerdahl, F. & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*, MIT Press.

Leslie, K. R., Johnson-Frey, S. H., & Grafton, S. T. (2004). *Functional imaging of face and hand imitation: Towards a motor theory of empathy*, in: Neuroimage, 21, 601-607.
London: W. Heinemann.

Menon, V., & Levitin, D. J. (2005*). The rewards of music listening: Response and physiological connectivity of the mesolimbic system*, in: *Neuroimage, 28*, 175-184.

Molnar-Szakacs I & Overy K. (2006*). Music and mirror neurons: from motion to 'e' motion*, in: Social Cognitive and Affective Neuroscience, vol 1 (3):235-241.

Molnar-Szakacs, I. & Overy K. (2009). *Being Together in Time: Musical Experience and the Mirror Neuron* System, in: Music Perception , vol. 26 (5), 489-504.

Molnar-Szakacs, I., Iacoboni, M., Koski, L., & Mazziotta,J. C. (2005). *Functional segregation within pars opercularis of the inferior frontal gyrus: Evidence from fMRI studies of imitation and action observation*, in: Cerebral Cortex, 15, 986-994.

Molnar-Szakacs, I., Kaplan, J., Greenfield, P. M., & Iacoboni, M. (2006). *Observing complex action sequences: The role of the fronto-parietal mirror neuron system*, in: NeuroImage, *33*, 923-935.

Montpare, J., Goldstein, S. & Clausen, A. (1987). *The identification of emotions from gait information,* in: Journal of Nonverbal Behaviour, 1, 33-42.

Noë, A. (2004). *Action in perception,* MIT Press.

Noë, A. (2008). *Precis of Action in Perception,* in: Philosophical and Phenomenological Research, 76(3) :660–665.

Patel, A. D. (2008). *Music, language and the brain*, Oxford University Press.

Patel, A.D. (2003). *Language, music, syntax and the brain*, in: Natural Neuroscience, 6, 674–81.

Repp, B. H. & Knoblich, G. (2004). *Perceiving action identity: How pianists recognize their own performances,* in: Psychological Science, 15(9), 604 – 609.

Robbins, P. & Aydede, M. (2009a). *The Cambridge handbook of situated cognition,* Cambridge University Press.

Robbins, P. & Aydede, M. (2009b). *Introduction,* in: *The Cambridge handbook of situated cognition,* Cambridge University Press.

Sano, H. & Jenkins, B. K. (1991). *A Neural Network Model for Pitch Perception,* in: Todd, P. M. & Loy, G. D. Music and Connectionism, MIT Press.

Styns, F., Van Noorden, L. Moelants, D., Leman, M. (2007). *Walking on music,* in: Human Movement Science.

Talmy, L. (2000). *Towards a cognitive semantics*, MIT Press.

Todd, P. M. & Loy, G. D. (1991) *Music and Connectionism*, MIT Press.

Trainor, L. J., & Heinmiller, B. M. (1998). *The development of evaluative responses to music: Infants prefer to listen to consonance over dissonance*, in: Infant Behavior & Development, 21(1), 77–88.

Varela, F.J., Thompson, E., Rosch, E,. (1991). *The embodied mind: cognitive science and human experience*. MIT Press.

Wilson, M. & Knoblich, G. (2005). *The case for motor involvement in perceiving conspecifics*, in: Psychological Bulletin, 131(3), 460 – 473.

Witvliet, C., Vrana, S. (1996). *The emotional impact of instrumental musicon affect ratings, facial EMG, autonomic response, and the startle reflex: Effects of valence and arousal,* in: Psychophysiology, Supplement, 91.

Zatorre, R., McGill, J. (2005). *Music, the food of neuroscience?,* in: Nature, 434, 312–5.