Playing without mental representations: embodied navigation and the *GesTCom* as a case study for radical embodied cognition in piano performance[1]

Pavlos Antoniadis[1] and Anthony Chemero[2]

[1] EUR ArTeC, Université Paris 8 & IRCAM
[2] Department of Philosophy and Psychology, University of Cincinnati

Background in Performance, Musicology and Music Interaction. Performance of notated piano music has remained attached to a past paradigm of interpretation, which is based on the mental representation of musical scores. This paradigm is challenged by developments in contemporary music, musicology and technology. The complexity of the musical act problematizes its encapsulation in a text and musicology has acknowledged the fact, through its performative and embodied cognitive turns. The primacy of multimodal interactions in complex musical systems, the central role of embodiment in performance and the mediation of the musical act through technology crack music notation open and invite us to reflect on the very essence of mental representability.

Background in Cognitive Science. Anthony Chemero's research is both philosophical and empirical; typically, it tries to be both at the same time. The research is focused on questions related to nonlinear dynamical modeling, ecological psychology, complex systems, phenomenology, and artificial life. His 'radical embodied cognition' problematizes the role of mental representations and offers concepts and tools for the development of a theory of embodied interaction with the musical score.

Aims. To argue that the role of mental representability in learning complex notated piano music can be outsourced on the embodied interaction between symbolic and environmental information; to do so against both the traditional textual interpretation paradigm and predictive processing in embodied cognition; to present the *GesTCom* interactive system and argue that it makes a convincing case for radical embodied cognition in piano performance; to explore reciprocal relationships between radical embodied cognition and complex music performance.

Main contribution. We propose a novel paradigm of pianists' interaction with complex music notation defined as *embodied navigation* (Antoniadis 2018). Its novelty lies in rethinking the classic notion of textual interpretation as embodied interaction, and musical performance itself as a dynamic system. In the radical version of the embodied navigation model, and in line with the radical embodied cognition (Chemero 2009), the processing of the musical text can be explained even without the need for mental representations, as dynamic interaction between the elements of the system: body, mind, instrument, notation and interactive systems. This embodied navigation paradigm is materialized in the *GesTCom* (Gesture Cutting through Textual Complexity) (Antoniadis 2018), a dedicated interactive system for learning notated music. It is a modular, sensor-based environment for the analysis, processing and real-time control of complex piano notation through multimodal recordings. The system optimizes the performer's learning experience through longitudinal multimodal documentation, real-time activity monitoring with augmented feedback, and adaptation of notation's complexity to the user's developing skills.

Implications for musicological interdisciplinarity.The contribution above is inscribed in what today constitutes a paradigm-shifting web of knowledge around musical performance. The relevant fields of the project include both the humanities and the sciences, as well as artistic research. In humanities, traditional approaches stemming from historic & systematic musicology and music pedagogy are complemented by the performative turn in musicology, the wider field of performance studies, and aspects of complexity in post-1950 compositional and performative aesthetics. In sciences, the role of embodiment in cognitive processes (embodied cognition/cognitive psychology), the study of physical movement through interactive technologies (Human-Computer Interaction) and the creation of new interfaces for musical expression are combined with computational approaches in musicology and dynamic systems theory. In terms of interdisciplinarity between musicology and cognitive science, both the theory and the interactive system presented here are inspired by radical embodied cognition. Inversely, they can serve as a case study for many of the non-musical claims of this theory, in a mutual and reciprocal gesture between the two fields.

Keywords: piano, complexity, notation, interaction, embodiment, cognition, representation

---

[1] This is the preprint version of an article accepted for publication in the peer-reviewed *Journal of Interdisciplinary Music Studies*, special issue "Embodiment in Music" following CIM19 conference in Graz, Austria (eds. Andrea Schiavio and Nikki Moran): https://musicstudies.org/ (accessed 06.06.2020)

Introduction

The current article questions the role of mental representations in the performance of notated piano music. In the traditional view, mental representations are considered a prerequisite for the musical act. The pianist presumably internalizes the musical text and performing technique, before being able to play and offer a personal interpretation of the text. In our hypothesis, this foundational function of mental representations can be outsourced to embodied interaction with music notation. Instead of internalizing and interpreting a static text, the pianist is assumed to constantly reshape notation, as if it were plastic matter. This dynamic reshaping is mediated by the pianist's embodiment and by the notation's materiality. This transformation feeds back into the performer's responses, in a constant interactive schema. Even if mental representations, say of a certain snapshot of the constant notational transformation, were to emerge, their explanatory power would be insignificant in virtue of the interaction. And since notation is not the malleable matter posited in our hypothesis, we have developed technological means for its dynamic adaptation to the pianist's actions.

Thus, the epistemological (as opposed to an ontological[2]) claim of this paper is that:

Piano performance can be better explained as embodied interaction with symbolic notation, not as textual interpretation. In this form of interaction, the role of mental representations is only contingent, not a sine qua non for music performance.

In considering only a contingent role for mental representations, our approach aligns with the branch of cognitive science known as *radical embodied cognition*. Its main feature is the rejection of mental representations and mental computations as explanatory tools. In their place, radical embodied cognition employs tools from *ecological psychology*, describing the interactions of organisms and their environment, and *dynamic systems theory*, describing the way systems are changing over time. Thus, the reformulation of musical interpretation as embodied interaction takes a more specific form: We show how the pianist interacts with the musical score, as part of a self-regulated environment, and how this interaction changes over time. We propose a theory of *embodied navigation* of this notation-environment and we develop a technological system called the *GesTCom (Gesture Cutting Through Textual Complexity),* in order to track and simulate the dynamics of the process. Both the theory and the interactive system presented here are inspired by radical embodied cognition. Inversely, they can serve as a case study for many of the non-musical claims of this theory, in a mutual and reciprocal gesture between the two fields.

The focus of the current paper will be on extremely complex contemporary piano music. This focus does not indicate an aesthetic, ideological or repertoire preference of sorts. We strategically employ the most complex piano music of our time, because it systematically problematizes representability, computability and predictability. In that way, it is clearly revealing the inadequacy of traditional approaches based on mental representations. At the same time, this repertoire allows us to rethink the very notion of musical complexity in new terms. Beyond the traditional conception of graphic complexity in quantitative and qualitative terms, we are talking about

---

[2] An epistemological claim differs from an ontological in that it refers to optimal forms of humans' access to knowledge, without making truth claims. We are making a claim about how performance is to be better understood, not a claim about what performance is.

complexity of dynamic interactions between parts of the performative system and about textual complexity that is grounded in a multimodal reality. In that revised sense, complexity of interactions may well extend into graphically simpler repertoire, for example the mainstream classical piano repertoire, as well as into notated piano music of other musical genres. However, for reasons of clarity and scope, this paper will focus on musical examples from the complex contemporary genre.

Next to radical embodied cognition, such a broad conception of complexity aligns with current trends in the field of Human-Computer Interaction, for example, the study of "extreme users in extreme situations"[3] rather than the study of average user behavior. Our claim is that our understanding of the most complex tasks can illuminate our understanding of simpler tasks as well. And indeed: The reformulation of interpretation into interaction prompts us to rethink a number of phenomena which seem to evade mental representability in the form of memorization. Traditional annotation practices, sight-reading and chamber music or ensemble performance, whereby a constant presence of the musical score is the case, would be some of them. Again, the scope of this paper will not allow us to expand beyond solo piano music. From the point of view of radical embodied cognition, the focus on extremely complex tasks and systems has an additional advantage: If we can show that complex music performance supports basic claims of radical embodied cognition, then we will have done so using a higher-order, 'representation-hungry' example of cognition with a motor component, rather than some minimally cognitive behavior.

One might still wonder: But why would we need a new theory and a new technology transforming notation to evade mental representations, given the fact that performers still do use notation, no matter how complex, for learning and performing? Our initial response to such question is similar to: Why do we need electronic mail in the place of physical mail? Or why did we ever need writing in the first place, if we can remember things? We did, because the very nature of visuographic media is to store information that would otherwise be hard or impossible to mentally represent. The interaction with such media has always been technologically mediated. In a certain way, neither the theory of embodied navigation nor the technology of the *GesTCom* are entirely new. Both are founded on the very evolution of music notation parallel to the evolution of musical practice. Both reflect the complexity not only of the codification but also of the deciphering, and the ensuing changes in the balance between mental representation and external mediation. Current music practices that appear to be hybrid, employing both mental representations and external media, could actually become fully externalized in the future, in the same way that the use of the internet challenged our ways of using information rather than internalizing it. The reshaping of the notation that we suggest here corresponds then to historically ubiquitous annotational practices of performing musicians. Our approach absorbs those practices into the very fabric of the *GesTCom* interactive system and substitutes the internalization of learning with immediate responses to a changing score-environment. This approach can have significant consequences not only for learning, pedagogy and real-time performance, but also for composition, improvisation and musical communication, as will be shown in the last section.

It is important to underline that this paper does not negate altogether the possibility of mental representability in musical performance. Mental representations may well be existent and desired, standing in for several parts of the system, as when one performs

---

[3] "Extreme users in extreme situations" is, for example, the main field of study of the laboratory *ex)situ*, at the Université Paris-Saclay https://ex-situ.lri.fr/about accessed 06.06.2020.

from memory[4]. What this paper does negate is the power of mental representations as explanatory tools for what is happening in the interaction between a pianist and the musical score, and especially so in the case of extremely complex music.

Our argument is structured as follows:

In the first section, we make an introduction to radical embodied cognition and mental representations. We review the modeling of mental representations in terms of coupled oscillators and we propose a transformation of the musical communicative chain based on this modeling.

In the second section, we look at the evolution of musical complexity and how it invites a paradigm shift from interpretation to interaction.

The third section presents our theory of embodied navigation of complex notation and how it relates to theories of performance that posit mental representations, both the classic and the latest ones.

In the fourth section, we present the *GesTCom* interactive system and we explain how it instantiates the embodied navigation theory.

In the end, we point to future directions and correspondences between radical embodied cognition and embodied navigation.

---

[4] Even in this case, the dynamics of the performer's interaction with the environmental structures such as the instrument, the acoustics and the audience could be claimed to evade mental representability. The music notation itself could be claimed to have been transformed entirely into movement and sound, without the need to mentally represent it. But we will leave this for another paper.

FIRST SECTION

Notes on mental representations

a. Radical embodied cognitive science

The theoretical foundation for the work described in this paper is what is called ‘radical embodied cognitive science’ (Chemero 2009). Radical embodied cognitive science differs from other kinds of embodied cognitive science primarily in its rejection of mental representations and mental computations as explanatory tools. In most varieties of embodied cognitive science, bodily motions and environmental resources are taken as supplementing, or even transforming, the mental representations and computations that are presumed to constitute cognition (e.g., Clark 1997). In radical embodied cognitive science, mental representations and computations are not taken to constitute cognition. The explanatory work done by mental representation and computation in embodied cognition is replaced by Gibsonian ecological psychology (Gibson 1979) and nonlinear dynamical systems theory (Port and van Gelder 1996). In radical embodied cognitive science, the object of study is coupled animal-environment systems. As such, radical embodied cognitive science is radical in two senses: first, it is a stronger rejection of standard cognitive science; second, relying on the original meaning of ‘radical’, it is embodied at its foundation, in contrast to other forms of embodied cognitive science that take features of embodiment as supplementing a computational and representational approach.

Because radical embodied cognitive science differs from embodied primarily over the status of mental representations, we spend the next section outlining some definitions and distinctions among proposed kinds of mental representations.

b. Distinctions and definitions

In this, we will give some definitions of the notion of mental representation, and make some distinctions that will be important later. Such definition and further discussion is important, so that we can substantiate the claim that complex piano performance of notated music is actually not only possible, but better effectuated and explained through an outsourcing of these traditional mental representations to embodied interaction with symbols mediated by technology. That is, complex piano performance is best explained by radical embodied cognitive science.

One of the main criteria differentiating traditional theories of representation from more recent theories is the decoupleability of the mental representation from the representational target, the thing or state being represented. To make this distinction clear, it is easiest to bring in a further distinction, between ‘presentation’ and ‘re-presentation’ (Grush 1997). Suppose that you are seated on a piano bench, looking at a score. We could potentially identify a cluster of neural activity that is responsible for your ability to see the score. That cluster of neural activity would be a presentation of the score. Suppose you feel frustrated at the difficulty of the score, and go out for a walk, but as you walk, you remember a part of the score. We could potentially identify a cluster of neural activity that makes it possible for you to remember the score, it might even be the same cluster of neural activity. This cluster would then re-present the score in its absence. Any view of mental representation would acknowledge that the re-presentation is a representation. The main issue concerning the decoupleability of representations is whether the presentation of the score also counts as a representation. That is, the issue is whether a state of an agent could be a representation of a feature of the world, even if that state could never do its job when not in the presence of that feature of the world. For example, for a visual representation of a book on the table to do its job (guiding reaching for the book), it must be working in the presence of the book. That is, the representation can only work when it is coupled to the target. We will call a view that allows both presentations and re-presentations to be representations ‘the traditional view’(Millikan 1984, Markman and Dietrich 2000). In contrast, according to what we will call ‘the decoupleable view’, presentations themselves are not representations; only re-presentations are. To be a representation according to the decoupleable view, a state of the agent must be able to do its job when the target is not present. This view takes it that mere presentation is not sufficient for representation (Haugeland 1991, Smith 1996). Thus the cluster of neural activity that enabled you to see the score is not a representation, but the cluster of neural activity that enabled you to remember it is. The decoupleable view makes far fewer things count as representations; most cognitive scientists and philosophers of mind believe the traditional view. That is, most philosophers of mind and cognitive scientists, including most embodied cognitive scientists, believe that even basic bodily activity with things that you are currently perceiving involves mental representations.

We can understand these varieties of representation in terms of coupled oscillators. Neurons are oscillators, as are brain areas, so it is natural to imagine mental representations in terms of an oscillator in the brain coupling to an oscillating target in the environment. Here we will discuss two kinds of oscillator that could serve as the model of a neural oscillator acting as a mental representation. Several models of neuron action potential (Fitzhugh 1961; Nagumo, Arimoto, and Yoshigawa 1962) come in the form of relaxation oscillators, named that because they slowly accrue voltage and then suddenly fire, relaxing or releasing their energy. These oscillators synchronize readily with rhythmic input when that input matches their preferred frequency of oscillation. When presented with an input pattern consisting of pulses at the appropriate period, a relaxation oscillator will synchronize its firing with the pulses. If these pulses are rhythmic, the oscillator synchronizes and “beats along” by emitting its own pulses in tandem. But when the driving stimulus is removed from a network, the oscillators decouple immediately and return to a quiescent state. In this way the oscillators are unable to couple with a target that is absent. A neural relaxation oscillator could be a mental representation according to the traditional view, but not the decoupleable view. It takes a much more complex oscillator to satisfy the decoupleable view. For example, adaptive oscillators (McCauley 1996) are hybrid oscillators, with some of the properties of relaxation oscillators and some from other types of oscillators. In particular, adaptive oscillators can change their phase to match a signal that arrives at an inopportune time. They can also change their preferred frequency of oscillation to match that of an incoming signal. Adaptive oscillators can beat along in real time to rhythmic stimuli, a task akin to tapping one’s foot along with music, even when the rhythm is noisy, as in the case with a human drummer. Because they can adapt their phase and preferred frequency, adaptive oscillators can continue to beat along with a rhythm to which they are no longer coupled. So although very simple relaxation oscillators can serve as traditional representations, it takes the much more sophisticated adaptive oscillator to be a decoupleable representation. The distinctions among these theories of representation and the oscillators that can instantiate them will be important later in the paper.

There are several further distinctions that will be important later in the paper. First is the distinction between mental representation and other external, non-mental representations. Oscillators, or images, or sentences in language of thought inside an agent might serve as mental representations, but there are other representations all around us; things like photographs, maps, sentences in articles, and, crucially for this paper, musical scores are representations, but not mental representations (Morgan 2013). Another crucial distinction, as far as external representations are concerned, is between static and dynamic representations. Sentences, images, maps, and the like are static in that they are, within some relevant period of time, unchanging; dynamic representations, like films, are dynamic in that, within some relevant period of time, they are changing. In the context at hand, that is music notation, we associate static representations with either traditional scores or even static representations of the interaction with the scores (as in the tablatures for embodied navigation theory, see section 3). This is opposed to mobile scores or to adaptive representations of how the representational content changes through the embodied interaction of the score and the user (as is the case in the *GesTCom*, section 4). Finally, we will need to distinguish between symbolic or amodal representations and grounded or modal representations. Theories that address how symbols mean something outside them point towards an ecological definition of symbols, firmly grounded in multimodal reality (Glenberg 1999, Barsalou 1999). These theories differ from standard cognitive science, which regards symbols as amodal, abstract, formal tokens to be computed (Fodor 1975). Music notation comprises both modal and amodal characteristics. The notes of a musical scale, for example, can be symbolized with numbers or letters and transposed or subjected to other sorts of abstract operations. Those tokens are amodal. At the same time, the notes of the scale can be represented graphically in space, refer to a certain characteristic of sound (frequency), refer to a certain placement on the piano keyboard, and they are abstracted from a reality of sound which is much more complex than just a frequency. This real sound is essentially multimodal, depending on the sound source, the acoustics of the space, or the way it has been produced, to name just a few of the notation's abstracted qualities. In that sense, notes can also be modal symbols, acquiring meaning only in meshing with reality.

c. Representations and musical performance

In this paper, we focus on performance from a musical score, as opposed to performance from memory. In coupled oscillator terms, the musical score is constantly coupled to the pianist. For the traditional view discussed in the previous section, this constant coupling involves mental representations on the part of the performer, as opposed to the decoupleable view, which does not consider this coupling as mediated by mental representations. Thus, the view which we will be questioning here is the traditional view of mental representations, as opposed to the decoupleable view, whose questioning would only be relevant in a scenario of playing by heart.

We use the following notation to distinguish between classes of mental representations and their expression through classes of coupled oscillators:

$$T \rightarrow MR$$

for mental representations MR constantly coupled to their targets T.

$$(T) \rightarrow MR$$

for mental representations MR of a potentially absent target (T).

$$\{\mathrm{T} \rightarrow \mathrm{MR}\} \rightarrow \mathrm{rosc} = \mathrm{XR}$$

for the relationship {..} between a present target T and its mental representation MR modeled as a relaxation oscillator rosc. This oscillator can be an external representation XR, onto which the initial content of the mental representation MR is outsourced.

$$\{(\mathrm{T}) \rightarrow \mathrm{MR}\} \rightarrow \mathrm{aosc} = \mathrm{XR}$$

for the relationship {..} between a potentially absent target (T) and its mental representation MR modeled as an adaptive oscillator aosc that can function as an external representation XR.

On the basis of these equations, the musical communicative chain between composer and performer through a musical score in classical music:

$$(\text{music}) \rightarrow \text{composer} \rightarrow \text{score} \rightarrow \text{performer} \rightarrow \text{music}$$

that is, an absent in material terms music, conceived by a composer, codified in a score, transmitted to a performer, who will produce the present in material terms, movement and sound, interpretation of the music, can be represented as follows:

$$\{(\mathrm{T}) \rightarrow \mathrm{MR1}\} \rightarrow \mathrm{XR} \rightarrow \{\mathrm{MR2} \rightarrow \mathrm{T}^*\}$$

Note the double coupling of the score as external representation XR to the mental representations of both the composer MR1 and the performer MR2. The asterisk in the final T* signifies that the produced target of the performer is the interpretative variation of the composer's intended target.

In our proposed formulation, the mental representations of the performer MR2 become contingent, and thus appear in parenthesis (MR2), if and only if the external representation of the score XR is functioning as a relaxation oscillator that couples directly to T*. The left-right arrow ↔ indicates the interactive relationship between the score and the target.

$$\{(\mathrm{T}) \rightarrow \mathrm{MR1}\} \rightarrow \mathrm{XR} = \mathrm{rosc} \leftrightarrow (\mathrm{MR2}) \leftrightarrow \mathrm{T}^*$$

The red part in the equation then refers to the constant reshaping of the score during the performer's embodied interaction with it. This reshaping traditionally takes the form of score annotation, but will take on a more literal meaning through the introduction of notation processing via the *GesTCom* system, as presented in the last section.

On the basis of the distinctions we outlined in the previous subsection, such score will be defined as: a causally coupled, external, dynamic and grounded or modal representation. This representation can drive the performer's action without mental representations, through constant adaptation in order to act as a relaxation oscillator.

SECOND SECTION

Complexity

The role of this second section is to show why and how the increasing complexity of music notation invites the paradigm shift from interpretation to interaction. After an anthology of historical references, we will outline the emerging distinction between a complexity that is static, symbolic and representable, and a complexity that is dynamic, grounded and interactive. This distinction echoes the categories of representations elaborated in the previous section.

a. The shifting meanings of complexity

The notion of complexity in music is far from clear, as opposed to other domains, such as physics, mathematics or economics. We are still lacking a unifying theory of musical complexity (Toop 2010). The term ‘complexity’ has nevertheless generously been used in historical and systematic musicology, contemporary music composition, music analysis and recently in music psychology.

In traditional musicology, notational complexity has invariably been associated with composition. The very history of composition is often described in terms of a mostly linear evolution from simple to complex. This trajectory leads from abstract to specific forms of representation of musical parameters, with pitch and rhythm having always been hierarchically dominant. In this reductionist and modernist view (Pace 2007), music history is limited to a singular technical progression, from medieval neumes to *New Complexity*’s nested rhythms and live interactive music’s patching environments.

The use of the term in contemporary composition is mostly in line with this historical and technical concept of complexity. It is defined in what several commentators indicate as ‘quantitative and qualitative terms’: an explosion of both the amount of notational information (quantitative aspect) and of their inter-relations (qualitative aspect). As Richard Toop puts it, complexity is a state, in which "there are not necessarily many things, yet in which I sense many levels of relationships between the few or many things" (Toop 1993).

The very distinction between qualitative and quantitative complexity is crucial for performance. The British composer Brian Ferneyhough invites the performer’s exploration of multiple prioritization and interpretation paths through his notational mazes. He seeks to develop “a notation which demands of the performer the formulation of a conscious selection-procedure of [..] the information [..] and a determination of the combination of elements (strata) which are to be assigned preferential status at any given stage of the realization process"(Ferneyhough 1995, p. 4). The US American cellist and composer Franklin Cox suggests that complex music has brought about a fundamental paradigm shift away from the traditional model of interpretation. The reason is its purely quantitative characteristics, namely: “[..] extreme degrees of both density and fine detail, and [..] coalescence of highly rationalized materials, notated challenges and organization with an extreme physicality and almost irrationality of results” (Cox 2002, p.70). The paradigm shift consists in the transformation of the communicative chain between composer-performer-listener into “[..] an overlapping series of volatile conflicts between incompatibles. Thus, notation is treated as an essentially opaque medium, (to paraphrase Derrida, notation is always already 'writing', with all its historical sedimentations) and such notation demands less reading than decipherment" (Cox 2002, p. 76). Interestingly enough, Cox's response to the open-endedness of musical meanings, in the sense of a Derridean *différance*[5], suggests a heightened role for embodied intelligence in radical complexity (Cox 2002).

The French composer Pierre Boulez is in many ways one of the forefathers of such ideas. In an early text (Boulez 1986), he is asserting that incompatibilities between composition and performance result from the intentional impossibility of one-to-one realization of all information and of all relationships between them. Thus, the augmented quantitative and qualitative complexity is functioning as a constraint for human abilities, but also as a source of refreshing creativity and indeterminacy in seeking to address those incompatibilities. In other cases, such as the music of the Greek-French composer Iannis Xenakis, such impossibilities are partly aiming at sheer physical pressure and transcendence of the performer's limits, a form of expressive athleticism (Varga 1996). Franklin Cox has described such attitudes under the rubric of *energetic striving* (Cox 2002).

Elsewhere, extended instrumental techniques[6] problematize the priority of pitch and rhythm and explore the materiality of the instrument and the sound. Such is the case in the German composer's Helmut Lachenmann *musique concrète instrumentale*[7] (Lachenmann 2004), but equally in tablature notation of decoupled actions with indeterminate sound results, as in the case of the German composer Klaus Karl Hübler (Hübler 2002). Such approaches rethink the notion of traditional technique by deconstructing it and are certainly not uncommon in many forms of free improvisation.

The dynamicity of learning processes and of the performers' input to the creative process, the embodiment of the performer as a factor shaping composition and the materiality of the traditional instrument already point towards a dynamic, grounded and interactive complexity. It is exactly those elements that problematize both the transparency of the communicative chain between composer and performer, as well as the traditional role of mental representations.

Another front against the reductionist view of notational complexity emerges in the so-called *performative turn* in musicology. Since the 1980s, an increasing focus on performance rather than composition broadens the meaning of complexity beyond the musical score (Cook 2013). Live performances, oral traditions, improvisations and recordings enter the realm of analyzable music (Arbo and Ruta 2014). New methodologies, often grounded on computational techniques and sophisticated technological tools, are being developed. Exemplary is in this respect Philippe Lalitte's analysis of the *Ten pieces for Wind Quintet* by György Ligeti (Lalitte 2015). Finally, new conceptions of even the most complex notations are offered, for example, Nicholas Cook's notion of *complexity as social script* (Cook 2013).

---

[5] *Différance* is a French term coined by the philosopher Jacques Derrida to indicate the simultaneous difference and temporal deferral of meaning, as a multiple semantic space opened up by his process of textual deconstruction.

[6] Extended techniques are all unidiomatic by traditional standards uses of the instruments as sound sources, for the production of noise and mixes of noise and sound.

[7] Lachenmann defines *musique concrète instrumentale* as follows: "With it we signify a music, whereby the sound events are chosen and organized in such a way, that the mode of their production becomes at least as important as the resulting acoustic qualities in themselves. Those qualities, such as timbre, amplitude etc. are sounding, so to speak, not for themselves, but rather describe or signal the concrete situation: We hear in them, under which circumstances, with which materials, which energies and against which resistances is a certain sonic or noise-producing action performed." (2004, p. 381).

Finally, there are alternative views on contemporary music historiography that offer non-reductionist approaches to musical complexity. For example, the musicologist Makis Solomos has described the emergence of sound in the music cultures of 20th and 21st centuries (Solomos 2013). In his framework, music is rethought in an ecological sense, not only in terms of its sonic contents but also as an emergent network of social interactions between composer, performer and listener.

> “The phenomenological explanation remains valid for numerous musics. But the refocusing on sound itself cannot be always thought of as ‘reductionist’. More generally, it presents itself as ‘emergence’. In its strong interpretation, which is developed in several tendencies in the cognitive sciences, physics and biology, this word (emergence) designates an evolution, which after a critical threshold of complexity generates new properties. For example, passing from a music which is centered on tones to a music which is of sound does not necessarily mean that the latter simply replaces the former: this may result by the very fact that, for example, the work on pitch, through its complexification, results in objects which do not end up being perceived as chords, but rather as composite sounds, as timbre (as is the case with spectral music). In this music, more often than not, the word emergence will assume a weaker interpretation, designating on the one hand that the refocusing on sound is generated by a progressive evolution in the inside of music, and that this evolution proceeds by complexification.” (Solomos 2013, p.9, our translation[8])

Interestingly enough, such ecologies do not *exclude* the musical score but rather include it in the sense of a ‘non-human agent’[9]. Such approach though differs significantly from Cox's “incompatibilities and conflicts in the communication chain” (Cox 2002) and similar ideologies of *complexism* (*Komplexismus*), as in (Mahnkopf 2002), which remain attached to a musical score paradigm.

b. *Intra-complexity* and *inter-complexity*

From the non-exhaustive anthology cited above it becomes clear, we hope, that the classic conception of notational complexity is imploding under its own density. Composition, performance and musicology have all come to acknowledge that notation cannot any more function as a self-enclosed symbolic cocoon of the musical work. At the same time, music notation in its most complex forms remains the *lingua franca* of composer-performer communication and one of the pinnacles of human culture. How could we crack notation open to dynamic transformation, multimodal grounding and social interaction, without abolishing the musical score altogether? Our distinction between *intra-complexity* and *inter-complexity* is a first step towards this direction.

Intra-complexity has been the main trope for the historical development of notation. It features the abstraction and codification of embodied actions and sounds into some form of symbolic language. This trope is not limited to Western art music but refers to most sophisticated non-Western notational systems as well (for example, Byzantine notation (Floros 1980)). This abstraction remains always partial, or ‘metonymic’, as Philippe Manoury (Manoury 1998) puts it, in that it substitutes the symbolic part for

---

[8] “L’explication phénoménologique reste valable pour nombre de musiques. Mais le recentrement sur le son ne peut pas toujours être pensé comme « réduction ». Plus généralement, il se présente comme émergence. Dans son sens fort tel qu’il est développé dans certains courants des sciences cognitives, de la physique, de la biologie…, ce mot désigne une évolution qui, à partir d’un seuil critique de complexité, génère de nouvelles propriétés. Par exemple, passer d’une musique centrée sur le ton à une musique du son ne signifie pas nécessairement que le second se substitue simplement au premier : cela peut se produire également lorsque – comme il en va avec la musique spectrale – le travail sur les hauteurs, par sa complexification, produit des objets qui finissent par ne plus être perçus comme des accords, mais comme des sons composés. Dans ce livre, le plus souvent, le mot « émergence » sera pris dans un sens plus faible, désignant, d’une part, le fait que le recentrement sur le son se produit selon une évolution progressive et intérieure (à la musique), et que, d’autre part, cette évolution procède par complexification.”

[9] Di Scipio 2015, pp. 286-287

the multimodal whole, which depends on performance. Intra-complexity is the defining stylistic feature of a large range of divergent musical languages throughout music history, including ideological overtones in recent history, in the sense of the composer's Claus-Steffen Mahnkopf's notion of *Komplexismus* (for a good review see Mahnkopf 2012, pp. 54-64).

Acknowledgment of the fact that intra-complexity is only part of a complex dynamic system in performance brings us into the realm of *inter-complexity*. Inter-complexity describes the interaction of notation with the other parts of the system, namely: with the performer, with the instrument and with the sound ecology, and certainly with other musicians and human agents. Such definition can be extended to include interactions in technologically advanced constellations (multimedia) and interactive multimodal systems, that is interactions with non-human agents. Dynamic scores, augmented feedback, multimedia and score-following would be some of the domains of application for such conception of inter-complexity.

The current article aspires to present a synthesis of these two views of complexity: It deals with intrinsically and interactively complex piano music, which is further re-inscribed into augmented, dynamic, interactive & multimodal musical scores. Please note that intra-complexity is not necessarily resulting in inter-complexity. In that sense, even a simple score in terms of intra-complexity may be generating complex interactions, and, vice versa, a very dense score might be part of a very simple network of interactions in systemic terms.

c. Rethinking interpretation as interaction

We have described performative responses to this shift in complexity as the transformation from a UTI into a TUI performance paradigm:

UTI stands for 'Understanding-Technique-Interpretation'. It refers to the dominant paradigm of interaction with a score, as transmitted in traditional musical education and musical praxis. According to this paradigm, the basis of music-making is the mental representability of the information contained in a musical score (first stage). This mental representability allows for the employment of instrumental or vocal technique as realization tool of the composer's vision (second stage). At a third stage, mastering both the mental and the physical aspects of the work eventually allows for an indeterminate degree of freedom of interpretative expression, which depends on the style of the music being performed.

TUI stands for 'Tangible User Interface'[10]. It refers to the proposed paradigm of performer-specific embodied interaction with complex notation. It assumes that, due to its intrinsic and extrinsic complexities, notational surface information and relations can never be mentally representable before having been fully realized. What happens in reality is that most performers process notation in an embodied way, through their active interaction with the score and the instrument. Performers 'touch' notation and transform it, in the course of learning, into embodied structures, which contrary to the symbolic notation are dynamic, malleable, and interactive.

In terms of mental representation:

If the UTI was represented by a transparent and symmetrical relationship between targets, mental and external representations on the part of composer and performer, thus:

$$T \rightarrow MR \rightarrow XR \rightarrow MR \rightarrow T$$

---

[10] The use of the term here is metaphorical and should not be confused, with, say, the work of Hiroshi Ishii on tangible user interfaces (Ishii 1997).

then the TUI conception of notation introduces the incompatibility of targets and representations, tending to the elimination of mental representation altogether in favor of a common interface-oscillator, the musical score:

$$\{(T) \rightarrow MR1\} \rightarrow XR = rosc \Leftrightarrow (MR2) \Leftrightarrow T^*$$

The search for an alternative to the UTI paradigm, one which acknowledges the primacy of embodied experience rather than symbolic cognitive processing of a text, leads us then smoothly into the domain of the philosophy of mind and of cognitive science, and namely to radical embodied cognitive science as defined in section 1. In the following section, we will be developing this TUI paradigm into an embodied navigation theory based on radical embodied cognition.

THIRD SECTION

Embodied Navigation

a. Theory of embodied navigation

In previous work (Antoniadis 2018) we have presented the theory of embodied navigation of complex notation. Similarly to radical embodied cognitive science, this paradigm is based on ecological psychology and dynamic systems theory.

As far as ecological psychology is concerned, the musical score is conceived as an environment to be navigated by the performer. This environment consists of *affordances*, that is of relations, or *couplings*, between the features of the musical score and the abilities of the pianist. The perceived features of the score are invitations to action, given proper abilities. The perception of these action-oriented features is posited to be *direct*, evading mental computations and mental representations.

Here is a toy example, which could facilitate the understanding of an ecological definition of music notation: Imagine a notated middle C on the piano, the only parameter being represented here being pitch. In terms of mental representation, this notation indicates an abstract frequency of approximately 261.6 Hz. In terms of ecological psychology, and given a piano and a pianist, this single C affords certain actions. Its perception and localization on the keyboard may be direct in the case of a trained pianist, without necessarily needing to mentally represent the intended frequency. The pianist may play it with any of the two hands, any finger, with a variety of touches, or even in non-idiomatic ways, for example with her knuckle or with her palm. The sound result, which will always be a C, assuming that the pianist hits the right note, can be extremely varied. This single notated C *affords* a great variety of actions and sound results.

Now let's assume that the composer adds notational information as to some other parameter of this C. It might be its duration, dynamic, articulation, touch, timbre, to mention just a few possibilities. This information changes the perceived affordances of the environment, as well as the potential actions and sonic outcomes for the pianist. The pianist might still imagine the action and the sound, mentally represent the notational image or a realization of it, given the low amount of information. But the pianist could as well physically explore, or navigate, the space of possibilities afforded by the notation.

To conclude this first part of the example, imagine now this middle C with all its parameters in the context of a complex composition, say a Ludwig van Beethoven piano sonata. Isn't it true that the affordances of this C will be dramatically reduced due to its further contextualization inside the score? It might be the inner voice of a passing c minor chord, but it could as well be the culmination point of the whole work. Whatever the case might be, there still is a great variety of interpretations of this C by virtue of its contextualization. We posit that its interpretation is not necessarily mentally representable, or even intentional, but is calibrated by the environment to be navigated (a whole piano sonata) and the embodied couplings of the performer to it.

We leave for the moment this toy example concerning ecological psychology aside, in order to resume our presentation of the embodied navigation theory with its second foundation: dynamic systems theory. Dynamic systems theory is the theory of systems that change over time. Such systems comprise multiple parameters, or *variables*, which are non-linearly coupled to each other. The behavior of such systems is often very complex and its mathematical modeling results in complete maps of all possible states of the system, called *state-spaces*. These state-spaces can predict the behavior of the system, given certain variables. For an example of a state-space, refer to (Beer 2003, p. 222).

The concept of a state-space is very tempting as a tool for modeling the affordances of a musical score and their development over time. If we consider the system of a pianist coupled to a score, then embodied navigation can be modeled as a processing of a layout of affordances through embodied action. Assuming that we could map all the possible realizations of the score in action and sound, this mapping would constitute the state-space of interpretation as interaction.

The reality is more convoluted: The interaction between a pianist and a score comprises too many variables to be able to be modeled mathematically; and the prediction of all possible states of this interaction would be debatable on artistic grounds (more on that in the "predictive processing" section).

Still: The idea of a state-space is crucial for modeling the dynamics of affordances in the musical score environment. Our state-space is represented as a series of annotations and transformations of the score that we call *tablatures*. The term tablature is used here somewhat loosely in relation to its strict historical meaning, in order to indicate hybrids between action-oriented and symbolic representations.

At a first stage, those tablatures depict the relation between pitch localization and physical *coarticulation* as defined by the musicologist Rolf Inge Godøy:

> "Coarticulation means the subsumption of otherwise distinct actions and sounds into more superordinate actions and sounds, entailing a contextual smearing of otherwise distinct actions and sounds, e.g. rapid playing of scales and arpeggios on the piano will necessitate finger movements included in superordinate action trajectories of the wrists, elbows, shoulders, and even whole torso, as well as entail a contextual smearing of the singular tones into superordinate contours of the scales or arpeggios.(…) One essential element of coarticulation is that it concerns both the production and the perception of sound, hence that it clearly unites

sound and action into units, into what we prefer to call sound-action chunks in music."(Godøy 2011, p. 2)

In the same way that the pianist's performing mechanism is coarticulated in interdependent and functionally coupled nested parts (fingers, hands, wrists, arms, shoulders, torso..), we posit that pitch notation affords different groupings, or *embodied layers*. We distinguish between a *finger-layer*, a *hand-grasp layer* and an *arm-layer* of the music notation.

Consider the previous toy example: In the case of the middle C, the notation affords ten different instances of fingering, as far as the finger-layer is concerned. In the case of this one note, only the finger-layer of the notation is so to speak activated. Attention: In order to physically activate this finger, the pianist will activate also other parts of the performing mechanism. Activation of the finger-layer refers not to the physical activation of the muscles to move the finger, but to the affordance of the notation.

Consider now this C as a part of a c minor four-part chord played by the right hand. This chord can be considered as the group of four fingers, but the relation of the fingers is not arbitrary. It is dictated by physiology and by a certain ordering, which avoids fingers crossing over each other. The fingers belong to a hand grasp and the notes will be played with say the fingering 1, 2, 3, 5. But the chord does not afford to be played with the fingering 3, 1, 2, 5… In that sense, in the case of this chord both the finger-layer and the grasp-layer are activated and coupled.

Consider finally this C as part of the previous chord which is transposed an octave up. The displacement of the hand necessitates the activation of the upper arm physiologically and the corresponding activation of the arm-layer of the notation. The finger playing the C couples to the hand grasp playing the first chord and is part of the arm trajectory necessary for it to play its corresponding note in the second chord.

In that sense, coarticulation creates constraints and reshapes notation. This reshaping changes the affordance layout during learning and performance.

Next to the tablatures of embodied layers, another representation originating in the idea of a state-space and extending the coupling to variables other than pitch is the *network of multi-parametric intentionality nodes*. The term 'intentionality' here refers to the composer's intentionality as codified in the musical score. The term 'nodes' makes reference to the fact that each note is the crossing point of multiple strata of notational information. This multiplicity contributes, among other factors, to the possibility of divergent sound images, or what would traditionally be called interpretation. The term 'nodes' implies also the possibility of a connectionist architecture, which could represent the prioritization processes that constitute interpretation: the special 'weight' of each note as a node, where multiple streams of information converge, is decided by the performer.

In the toy example above, imagine then that the c minor chord is arpeggiated, the composer requiring a different duration, dynamic and touch for each note. Each note can be represented as a node, where three strata of variables intersect and interact with the finger-layer and the grasp-layer. Those strata change the action and the sound outcome and the prioritization of parameters can be represented as a special weight for each note-node.

b. Complex examples of embodied navigation

Let us now consider some really complex examples from contemporary piano music.

The following example by Iannis Xenakis *Mists* (Figure 1, 'emergence of a tablature A') features heightened intra-complexity in the domain of pitch. Pitch is here generated according to several probability laws detailed by Xenakis in (Xenakis 1980) and resulting in the so-called *stochastic* constellations of Figure 1. These textures are compared by Xenakis to the Brownian movements of gas molecules. Our hypothesis is that the sheer quantity of notes and the lack of readily available patterns or structural relationships renders learning strategies based on mental representation and memorization too difficult or impossible. We suggest that, in this and similar situations, learning involves the gradual chunking and transformation of the pitch material according to the performer's embodiment, and more particularly according to the physical coarticulation of the performing mechanism.

We have then represented learning as the emergence of a multilayered tablature based on an organization of pitch information according to physical coarticulation: Fingers or finger-layer, grasps or grasp-layer and arm movements or arm-layer. You may clearly see the gradual transformation of the notation, passing from a grouping based on individual fingers (Figures 1a, 1b) to one based on hand grasps (Figure 1c) and one based on arm movements (Figures 1e, 1f). In more detail: Figure 1a consists in a transcription of the original Xenakis passage in reduced proportional notation. Since we focus on pitch, we have removed the sixteenth note stems that articulate time in the original (*time-space notation*) and we have used four staves to represent the pitch space, instead of the original two. In this way, we avoid octave doublings and we have a clear representation of pitch in space. Figure 1b represents the finger-layer, with blue numbers corresponding to right hand digits and orange numbers corresponding to left hand digits. Figure 1c shows the organization of these fingers in hand grasps, either physically graspable as a chord, or 'quasi-grasps' that are to be arpeggiated. The grasps are indicated with ellipses, quasi-grasps with ellipses whereby one side is an arrow. This is the grasp-layer of the passage. Figure 1d has removed pitch and kept only the grasp-layer defined through register placement. Figure 1e features the intermediate arm movements that interconnect the different grasps. Figure 1f has merged those intermediate movements with the 'quasi-grasps' into two linear trajectories and it has abstracted the grasp-layer. This is the arm-layer of the notation. In these last two representations (examples 1e, 1f), the musical staves have been rotated by ninety degrees clockwise, so that the representation matches the view of the keyboard by the pianist: The right hand trajectory in blue is on the right side, and the left hand trajectory in orange is on the left side of the pianist. The straight lines in purple that stick out of the trajectories indicate rapid leaps, or 'edges'. Interestingly enough, the initial pointillistic image of the notation has been transformed in two mostly linear trajectories, only interrupted by 'edges'. The hand-grasp groups of Figure 1c are nested in these trajectories.

Please note that the embodied layers represented above equal the alleged state-space of affordances throughout the learning process. Although presented here in linear fashion, those static representations do not constitute linear stages of learning but changing possibilities of action, that the performer non-linearly navigates.

In that sense, the transformations of the initial notated image are aspects of the inter-complexity of the passage: the multiplicity of relations between physical coarticulation and intra-complexity. These relations resist a definitive mental representability due to their constant transformation in performance.

The following example (Figure 1, 'emergence of a tablature B') explores the parameter of complex pitch combined with complex rhythm. Using a different passage from the same piece, *Mists*, we show an embodied navigation model for dealing with extreme rhythmic tasks. According to (Cox 2002) and (Schick 2006) among others, when confronted with tasks, such as the exemplary mosaic of non-coinciding polyrhythms in four parts (example 2), performers resort to *mediation techniques*. The core idea is that, instead of computing and performing the irrational rhythms *per se* in the mind, performers externalize the task. They are using beats, pulses, decimal positioning inside the beat or tempo transformations, in order to make the original impossible task manageable. Our own contribution here is that those mediation techniques are essentially embodied and coupled to pitch. Thus, different embodied layers correspond to different hierarchical layers of rhythm.

In the first mediation example 2a, we show how the whole passage's pitch content is chunked in hand-grasps and arm movements, exactly as was the case in example 1. In example 2b, we show how those grasps, indicated with blue ellipses for the right hand, are coupled to eighth-note beats, thus facilitating the performance of the complex rhythm 7:6 against 5:4. There are no exact computations, but a fitting or distribution of the two streams of notes, so that they can fill the space between two successive beats. In example 2c, we show how bimanual execution of the passage is facilitated through the computation of the decimal position of each note inside the beat (2ci) and the corresponding ordering of the attacks between the hands as a single line (2cii). This mediation technique couples with the finger-layer rather than the grasp-layer, although bimanually. In the last example (2d), we present the coupling of the arm-layer with several macro-rhythmic characteristics and phrasing of the passage: Red boxes indicate the introduction of a different tempo, blue boxes indicate the entry of the scale material in canon, yellow highlights indicate rhythms that are overall slower than the pulse, the brown line indicates a steady flow of sixteenth notes traversing the passage diagonally throughout the voices. In short, those are different paths, to invoke Ferneyhough's terminology, or prioritization nodes, throughout the passage. Please note that none of these mediation techniques is in itself sufficient as a representation, mental or external, of the original notated rhythm. Each mediation technique illuminates and makes possible different aspects of the rhythm and the final performance depends on an oscillation or navigation between these techniques.

The inter-complexity, as the relation between physical coarticulation and nested rhythmical layers, problematizes then the notion of mental representability of rhythm as mental calculation. Just another word on why the rhythms in the example above cannot be mentally represented and subsequently reproduced. Ignoring all issues of multiple hierarchical layering, the simplest explanation for this would be the problem of coordination of the tuplets. In the case of a simple 3 against 2 tuplet, the physical coordination is easily mentally representable and then achievable through the least common multiple of 6: If we imagine an underlying sextuplet, the notes of the original triplet fall in the positions 1, 3, and 5 of the sextuplet, the notes of the duplet fall in the positions 1 and 4 of the sextuplet, and thus one can very well visualize and then exactly perform the rhythm. On the contrary, in the case of an irrational rhythm, such an approach cannot work due to extremely big least common multiples. For example, in the first over-layering of tuplets in figure 2, example b, Xenakis asks the performer to play a 7:6 against a 5:4 with her right hand. This is a non-coinciding polyrhythm, which would coincide in a cycle of 12 sixteenths (the least common multiple of the denominators 6 and 4), in which case the polyrhythm equals a 14:12

against a 15:12, or a 14:15. Now, in order to play a 14:15, one would need a least common multiple of 14*15=210, which practically means that one would need a 210-tuplet in order to pinpoint each of the 14-tuplet and 15-tuplet notes in the same way one did with the 3:2 polyrhythm. This is obviously impossible. What one actually does, is practically navigate between the different embodied couplings as described in the second example, exploring several possible realizations of the rhythm, and using simple heuristics such as "note x comes before the beat and note y comes after", etc. In that sense, the rhythm is not a priori representable, but rather the result of an interaction between the symbolic information and the nested embodiment.

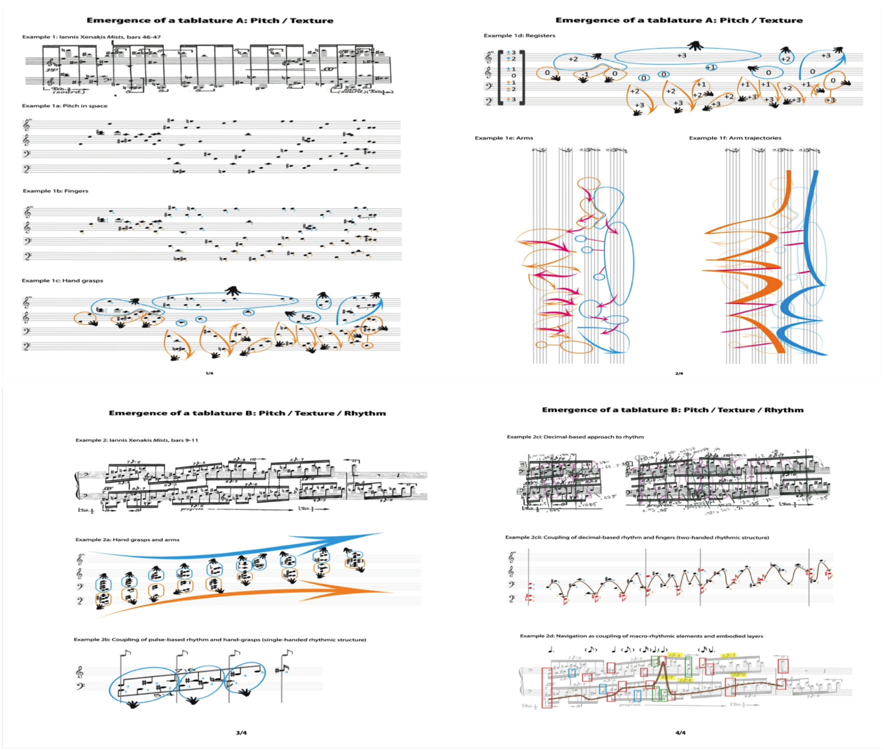

Figure 1. Emergence of a multilayered tablature based on physical coarticulation. The upper part of the image (A, examples 1a to 1f)) shows the coupling between pitch / texture and embodied layers. The lower part of the image (B, examples 2a to 2d) shows the coupling between complex pitch/texture (2a), complex rhythm (2b-2d) and embodied layers.

Finally, similarly to the extension of the model from pitch (example 1) to pitch and rhythm (example 2), the layout of affordances can be extended to include any notational or musical / analytical parameter. In this way, we are creating multidimensional spaces to be navigated in an embodied way. Alluding to connectionist architectures, we have defined such spaces as networks of intentionality nodes. In the following example (Figure 2, left side), each note represents a node, whereby multiple strata of notational information intersect. These strata are represented with different symbols over-layered on each note, in particular symbols for: simultaneity of attack, beats, hand-grasps, decimal positioning, change of speed, change of articulation, change of dynamics, rests (see legend in the bottom of the left side panel of figure 2). The more strata converge through a note, the more important is its ‘weight’ in the resulting network of nodes. Thus, we can model interpretative decisions in the form of complex interactions between embodiment and any notational or analytical parameter.

Thus far, we have dealt with the feature of inter-complexity which we defined as dynamicity. This dynamic processing of notational information during the learning process has been neglected in theories of notation, and so have the ubiquitous annotational practices of performing musicians. But the most compelling perspectives for a truly anti-representational and anti-computational approach open up through the focus on what we defined as the second feature of inter-complexity: the multimodal grounding of the notation, as shown in (Figure 2, right side). Notation refers to real physical energies, and these energies can today be registered through multiple capture systems. These modalities include sound, image, video, acceleration through inertial sensors, MIDI information and finger positioning through capacitive sensors on the surface of the keys, among many others available in the market. The combination of such energies, in the form of multimodal data, with the musical score, as constantly dynamically processed by the performer, creates the true possibility of outsourcing the learning process onto the environment and making it fully communicable. The registration of multimodal grounding will be the object of section four on the *GesTCom*.

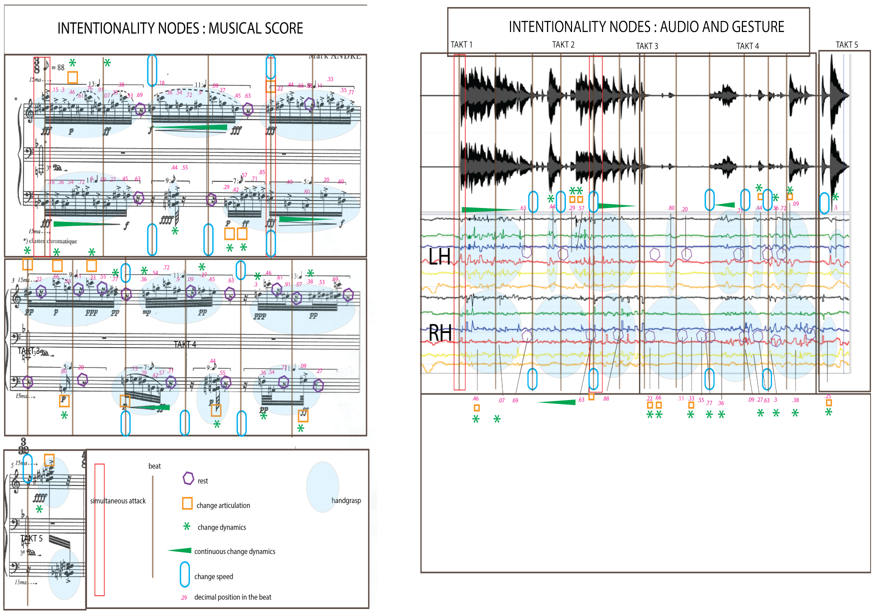

Figure 2. Intentionality nodes in Mark Andre’s *Contrapunctus,* p. 1. In the left side, each note carries several symbols, which represent different parameters, including articulation and dynamics. In the right side, those parameters have been mapped into a representation of multimodal data, audio (on top) and gestural signals (3D accelerometers and 3-axis gyroscopes, bottom part). The blue ellipses in both parts represent the grasp-layer of the notation.

For the moment, please compare the intentionality nodes in an example by the German-French composer Mark Andre (Figure 2) in notation form (left part) and in multimodal form (right part). The network of multi-parametrical nodes, including the grasp-layer (blue ellipses), is mapped on both the annotation and the multimodal data (from top to bottom: stereo sound, 12D gestural signals, six dimensions for each hand). Each change in the weight of each individual note is reflected and registered in the multimodal data. Thus, we can combine first-person, subjective symbolic annotations, with third-person, objective and communicable multimodal registrations.

c. Conclusion

To sum up: Embodied navigation is conceived as an extension of physical movement. It is a higher-order, diachronic movement, in a score-space, or score-environment. This environment constitutes a state-space of affordances: of relations between the abilities of the performer and the possibilities of action enabled by the fixed score. Those relations change over time, producing a new and infinitely malleable space. The movement functions between learning and performance, between detailed and global aspects and between the continuity of performance and the resistance of decoding. The qualities of this navigation (its directionality, its speed, its viscosity etc.) define what can sound from the initial incomprehensible and/or unplayable image. Interpretation is this diachronic movement, in place of the reproduction of a static sound-image.

In that sense, it is important to stress that embodied navigation is not a *metaphor* for virtual movement inside a musical score, but rather the temporal extension of physical movement in the physical space of the instrument as articulated by the affordances of the musical score. Similarly to navigation defined as non-metaphorical movement, the score-space is not a *metaphorical* space where virtual movement would take place, but rather the dynamic, in-time articulation of the fixed physical spaces of instruments and musical scores by physical movement. Navigation takes place in multiple time-scales as ontogenesis of the performance, similar to the development of a living organism, from the first stages of learning up to the multiple performances, which are traditionally termed ‘interpretations’.

In such a dynamic process, mental representations become contingent. The performer is instead constantly coupled to an oscillating external representation, the ever-changing musical score, and the system performer – score can be thought of as a coupled oscillator. Embodied navigation embraces a radical anti-representationalist dynamical stance, according to which cognition can be described as a dynamical system without the need for mental representations. It could however embrace softer versions of embodied cognition, such as 4E cognition (enacted, embedded, embodied and extended cognition, Rowlands 2010). A ‘hard’ (rather than radical) version would accentuate the dynamic interplay between internalization and action. Such position would not necessarily reject mental representations, but would posit *action-oriented* (Clark 1997) or *pushmi-pullyu* (Millikan 1984) mental representations. Eventually, there is the possibility of a ‘soft’ version, which would simply acknowledge the situated nature of the tasks at hand and the ergonomic / heuristic use of scores and instruments, for what is essentially a mental cognitive task (*embedded cognition*, Rowlands 2010).

In the following two subsections, we compare the theory of embodied navigation to two alternatives based on mental representations.

The first is a very good example of the UTI model of interpretation, a renowned ancient piano method from the 1930s, still very much influential today: The *Leimer-Gieseking method of modern piano playing*. The second is a state-of-the-art model of *predictive processing* in music performance, proposed by Marc Leman (Leman 2016). Although predictive processing is the latest wave in embodied cognition, the positing of mental representations as the basis of embodied cognition is strikingly closer to the UTI than to embodied navigation.

d. Embodied navigation and the *Leimer-Gieseking method*

This subsection is dedicated to the presentation of the Leimer-Gieseking method of piano playing[11] as a radical example of the UTI model. As will be shown shortly, the Leimer-Gieseking method assigns a leading role to conscious memorization and mental representability during learning and performance. In that way, it is a very good measure, against which the ecological qualities of the embodied navigation model can be gauged.

Throughout their book, Leimer and Gieseking (from now on abbreviated to LG) passionately advocate the memorization of the score as the cornerstone of any subsequent learning process. This memorization is to be achieved through systematic internal representation away from the instrument:

> “it is essential, before beginning with the practice of the piece, to visualize the same, whereupon, if this has been done thoroughly, we shall be able to play it correctly from memory. To be capable of doing that in short time, the memory must be trained by means of reflection (systematic logical thinking).”(1972:11)

In other words, the experience of performing on an instrument begins with a completely disembodied mental activity, which includes visualization, reflection and memory training. In terms of our definition of the body, the instrument and the score as the three fundamental environmental structures at the performers' disposal, in the context of a dynamic system of performance, we realize that the LG strategy essentially does away with all of them. It disposes of the first two (the body and the instrument) and offloads the third (the score) onto the performer’s mind.

The process of visualization, reflection and memory training is thoroughly described in several examples throughout this book. The authors offer analyses of the mental working-out of pieces of the standard repertoire. The analysis is highly detailed, but in a prioritized way: pitch and duration as the barebones of composition are exhaustively examined, while the more 'physical' variables, that is articulation and dynamics, are not initially entering the mental frame. Structural observations are active in a relatively loose way, not with a rigid intention to grasp an overarching formal schema or reflect on the process. In other words, there is a moment to moment, or one note at a time, memorization process, which brings to mind the notion of *episodic memory* (Rowlands 1999).

Episodic memory is the memory of discrete, isolated events in time, as opposed to semantic memory, which relies on the relations between seemingly isolated events, and procedural memory, the memory of sequences of actions. In the LG memorization process, the balance between its episodic and semantic properties is considerably

---

[11] The book was originally published in 1931 in two individual volumes, *Modernes Klavierspiel* and *Rhythmik, Dynamik, Pedal*. It has had a solid reputation and influence in piano pedagogy and celebrated its 30th edition by Schott in 2011: https://en.schott-music.com/shop/modernes-klavierspiel-no39286.html accessed 06.06.2020. Here we will be referring to the translation in English, published by Dover (1972).

leaning towards the first. The highly localized, measure-by-measure, hand-by-hand, nature of the description of the musical text, and the simplicity of the syntactic relationships observed, point toward an episodic experience of 'being in one place at a time', only that this happens mentally, not in real time and space.

A metaphor employed by the authors is very telling about the nature of mental representations employed:

> "When a part of a composition has been played for the first time, a picture of the same becomes imprinted on the brain. This picture varies in clearance according to the mental constitution of the pupil. In general, a very faint impression is left on the memory, similar to a photograph which is not clear or has been under-exposed. Through constant repetition the picture becomes more and more distinct and finally resembles a clear, sharp photograph." (1972:47)

But it is not only abstract understanding of the music, which is governed by mental representation. Technique is also claimed to be fully representable in the mind:

> "By further development of the idea, one acquires the ability even to prepare the technical execution through visualization, so that, without studying on the instrument itself, the piece can be perfectly performed and this in a most astonishingly short time."(1972:11)

The absolute banishment of the instrument from the learning process is later praised as an achievement, and even as a sign of superiority:

> "Only a very few of the elect are born with the talent of immediately and intuitively grasping the meaning of a composition; and they alone have the capability of reaching to so high a degree of mental and manual ability that they can mentally comprehend and correctly render a composition, by means of the fingers, practically without further practice". (1972:33)

Our final note considers the persistence of Leimer and Gieseking's internalism in the very foundation of piano technique. As opposed to other modern piano schools, which ecologically stress the interdependence of muscles and gravity as energy sources for piano playing (Sándor 1981, Breithaupt 1905), LG focus on muscular control. Whereas their foundations call for 'natural' playing, employing the least possible physical strain, the way to achieve this often becomes mystified. For LG, a notion of consciously controlled relaxation as a complement to the conscious exertion of the muscles is of key importance, but in the expense of every notion of gravity or weight. Thus their rejection of all ancillary movements:

> "I contrive to raise a feeling of relaxation from within, as it were. This is generally attempted by the aid of visible movements. All movements are injurious." (1972:12)

While acknowledging the fact that coordination and muscular interdependence is indispensable, LG also advocate muscular strengthening. Fixation of joints is constantly mentioned during the description of the individual modes of touch, and the contradiction of these remarks to the idea of relaxed playing is to be acknowledged later:

> "A strong fixation is unavoidable in forte and fortissimo playing. But one should always think of relaxing the muscles whenever the opportunity arises, so that the fixation will be interrupted and lessened. As we have already stated, the relaxation must ensue from within, minus any noticeable movement."(1972:111)

The points we would like to keep from this brief presentation are: the conscious control of muscular relaxation and exertion; the muscular strengthening; the fixation of joints and avoidance of movement. These three technical features, plus an exclusively mental practicing and the memorization of the score consider the pianist as an entirely self-contained system, dedicated to a perpetual quasi-biological development, both mental and physical. The reluctance to ecologically employ any resources other than the mind in the course of learning and performing is absolute-in fact considered as a sign of weakness, almost inferiority, if it does happen. The mental representability of every and each element of piano playing becomes the absolute prerequisite for musical performance.

It should be clear that such conception of pianistic development is rather alien to the model of embodied navigation. In this model, pianistic development is associated with the increasing ability to meaningfully interact with the environmental structures at the pianist's disposal: embodiment, gravity, the instrument, acoustical spaces, and quintessentially for this paper, the musical score. This distributed interactivity allows for the outsourcing of all elements internalized in the case of LG.

e. Embodied navigation and sensorimotor prediction

Having juxtaposed embodied navigation to a radical instance of internalism in traditional piano playing, we suggest now a second cross-reference to the opposite direction: predictive processing as the latest wave in embodied cognition. We will show that the epistemological claim for embodied navigation, namely that "musical performance is better explained as a complex dynamic system, which does not necessarily involve mental representations", seems to be at odds with a particular description of sensorimotor prediction by Marc Leman. In this sense, and despite its firm positioning in the field of embodied cognition, this particular theory of sensorimotor prediction shares a common core with the LG model of musical performance presented in section d. above: Mental representability of the musical score as the foundation of performative actions. This common core is the exact point of divergence between radical embodied cognition in the form of embodied navigation and both theories presented in d. and e.:Leimer & Gieseking's internalist pianism and Marc Leman's model of sensorimotor prediction for musical performance.

i) Predictive models

In chapter 6 of his latest book *The Expressive Moment: How Interaction (with Music) Shapes Human Empowerment* (Leman 2016), Marc Leman argues that sensorimotor prediction is one of the cornerstones of what he calls the dynamics of expressive interaction (the other two being *entrainment* and *expressive alignment*, to which chapters 5 and 7 respectively of his book are dedicated). In the context of the embodied navigation paradigm, while entrainment and expressive alignment fit well with notions of expressive (de)coupling of parametrical layers, the notion of sensorimotor prediction raises questions. We address those, being fully aware of a fundamental difference: That the model of embodied navigation addresses higher-level and longitudinal processes, in comparison to sensorimotor prediction that addresses interaction at a 'molecular', instantaneous level.

The theory of expressive interaction, according to Leman, "is based on the idea that the human brain is a predictive machine that forms beliefs through sensory interaction with the environment" (p. 124).

In its basic form, the prediction model posits that motor commands by a performing subject are inextricably linked to sensory expectations, namely expectations about the kinesthetic / proprioceptive outcome of the movement (internal expectation), the sensory outcome of the action, and the potential perception of other external or environment stimuli (external expectations). Those expectations seem to work part and parcel with the commands in both directions, so that either a motor command generates the expectation of a desired output (in so-called *forward models*) and / or, vice versa, the expectation generates the relevant command (in so-called *inverse predictive models*).

If the first component of this prediction-based modeling is the close relation between motor commands and expectations, then the second component addresses the comparison between the actual perceived outcome, or external simulation, to the expected one. Thus, a trumpeter who has used her lips and fingers to produce an expected pitch might for different reasons have failed to do so. The perception of the difference between the expected and perceived output is termed *prediction error*. This error allows then for a third stage in the sensorimotor prediction loop, namely the updating of the system of expectations and motor commands, so as to minimize the prediction error and fit it to the real outcomes.

ii) An example of predictive model for musical performance

A very simple example for the application of the expressive prediction theory to musical performance is provided in (Leman 2016), p. 134, summarized by the author as in the quote:

> “The procedure works as follows: First, the notes are re-conceptualized as a scale using a chunking process. Second, the scale chunk is handled as an object concept. Third, the object concept activates a motor command involving an inverse predictive model. This sets the readiness for action, and initiates the action at the time of the action execution. Fourth, the motor command is executed using a forward predictive model that simulates the action to produce expected sensory outcomes. Fifth, the produced sensory outcomes are checked whenever that is possible. The prediction error may be used as sensory feedback to adjust the ongoing predictions. Note that predictions allow for a parallel processing of action and perception. Thus, at the start, the musician will imagine the D major scale. This imagery prepares for a performance of the first four notes. Then, while playing these notes, the musician imagines the E-flat major scale, which prepares for the performance of the first four notes of the E-flat scale just as the D major sequence is completed, and so on. The thought experiment suggests that interaction with music can indeed be broken down into a range of cognitive processes, including chunking, imagery, action intention, action execution, and the use of auditory and kinesthetic feedback. In short, the theory of predictive processing offers an understanding of processes in music playing, and it may be instructive as a cognitive model for music education.” (Leman 2016, p. 139)

Here are some observations about Leman’s suggested model in the context of embodied navigation exposed before:

Firstly, the simplicity of the musical material being chunked here, namely parts of major diatonic scales, does allow for the rapid conceptualization – objectification – internalization and subsequent imaging of the desired action and sensory output. But what would happen in the case of more complex musical material, such as the example of Xenakis’ *Mists* reviewed above, whereby multiple strata of parametrical information need to be organized in connectionist intentionality nodes networks? The sheer complexity, both qualitative and quantitative of the material, makes the application of predictive models hard from a computational point of view.

Second, interactive aesthetics in many strands of post-1950 Western art music composition and other forms of music-making, for example free improvisation, are programmatically aiming at *unpredictability*. The best example would be Klaus Karl Hübler's (Hübler 2002) layers of independent actions, with the intentionality of an indeterminate sound result. How would predictive models be applied in such cases?

Third, the process is being presented in strict algorithmic fashion, with the only exception being the feedback loops allowed in terms of prediction error. Such linearity seems to be at odds with the claim for musical performance as a non-linear system of interactions between the embodied mind, the notation and the instrument. The non-linearity, we claim, aims at explaining the emergence of intentionality, *before* it is fully at work in predictive models triggering actions; or, to use Merleau-Ponty's terms: it aims to "slacken the intentional threads which attach us to the world and thus bring them to our notice." (Merleau-Ponty 1962, p. xiii)

Fourth, at the core of this model stands the brain's ability to imagine and simulate actions and outcomes, based on previous learning. The compatibility of computational and embodied variants of cognition, such as also advocated by (Clark 1997) and (Rowlands 2010), allows for a smoother integration of predictive models in embodied cognition debates, as in Clark (2016). However, the radical embodied stance criticizes such models as lapses towards a representationalist view of cognition. The question being raised is: Would prediction be possible without mental representations? A positive answer to this question would allow for a smooth integration of predictive models in the program of radical embodied cognition. Positive answers to this question have actually been proposed. An example is the question around the *optic flow* and its sufficiency to guide action, as opposed to reliance on extra-retinal information or efference copies, which equal mental representations. For an overview, refer to (Chemero 2009, p. 126).

Fifth, such models are claimed to allow for generalization at the level of expression:

> "Note also that what is said here about note chunks can also be applied in more detail at the level of expression. We are then speaking about articulations of notes, coarticulations and about the articulation of expression arcs that control the timing, dynamics and the articulation of larger sequences of notes". (Leman 2016, p. 139)

However, the passage from a cognitively manageable amount of information to the complex nested information that constitute expression is not further analyzed. How is this passage effected? Linearly or non-linearly? The level of expression has already been shown, so is our thesis, as an emergent property out of interactions, rather than an independent layer on top of some 'basic' sensorimotor learning.

It is exactly music of staggering notational complexity and navigational aesthetic intent that questions predictive models. At the same time, such music invites alternatives based on the notion of direct perception and embodied cognition.

Even in the case of music not that complex, a main line of criticism from the point of view of radical embodied cognition is expressed by the following passage by Thelen, questioning particularly Piaget's *object-concept* cited by Leman:

> "We propose here a radical departure from current cognitive theory. Although behavior and development appear structured, there are no structures. Although behavior and development appear rule-driven, there are no rules. There is complexity. There is a multiple, parallel, and continuously dynamic interplay of perception and action, and a system that, by its thermodynamic nature, seeks certain stable solutions. These solutions emerge from relations, not from design. When the elements of such complex systems cooperate, they give rise to behavior with a unitary character, and thus to the illusion of structure. But the order is always

executory, rather than rule-driven, allowing for the enormous sensitivity and flexibility of behavior to organize and regroup around task and context." (Thelen 1994, xix)

Another model for the assessment of why predictive models are not fit to describe complex dynamic processes, such as the performance of a complex musical score, has been provided by Van Gelder's critique of computational approaches to the *Watt Governor*.

Remember a supposed computational approach to the Governor, which very much resembles Leman's account in that it chunks the process into algorithmic steps:

"1 Measure the speed of the flywheel.
2 Compare the actual speed against the desired speed.
3 If there is no discrepancy, return to step 1. Otherwise,
a measure the current steam pressure;
b calculate the desired alteration in steam pressure;
c calculate the necessary throttle valve adjustment.
4 Make the throttle valve adjustment.
Return to step 1."
(Van Gelder 1995: 348)

Van Gelder proposes a comprehensive dynamic description, which could essentially start at any given step of the algorithm, or rather avoid altogether such a diachronic sequence of steps, since the speed of the flywall, the height of the balls and the opening of the throttle are constantly coupled and interdependent, given the steal flow. In the same vein, given the first sparkle of intentionality or motivation of the performer to engage with a complex musical notation, all linearly arranged elements of the system that Leman provides us with, that is chunking, imagery, action, intention, action execution, and the use of auditory and kinesthetic feedback, would be deconstructed into a self-organized performing system. In such system, motor commands would, for example, appear to alter or even shape the very chunking and object concepts. And this notwithstanding that, in the hard model of embodied navigation, such object-concepts a.k.a mental representations are judged as contingent in the first place.

iii) Are predictive models incompatible with the notion of embodied navigation?

Let's review some of the ideas that both models, predictive and radical embodied cognition, can share.

First, the idea of chunking. While the chunking process in Leman's predictive model is supposed to be an objectification / conceptualization of musical material on the basis of existent patterns (scales), the embodied navigation model suggests multiple chunkings at multiple hierarchical levels of movement coarticulation, based on notational affordances. In that sense, chunking of action-oriented descriptors appears to be a useful heuristic for the management of cognitive overload.

Second, the idea of an algorithmic sequence with environmental feedback loops seems counter-intuitive to the idea of navigation along and between several dimensions or strata of chunks. However, the scenario of learning as refinement towards perfection, can indeed profit from a strong predictive model. Still, such a model aims clearly at an understanding of interpretation as reproduction rather than as navigation through a state-space longitudinally shaped.

Third, the idea of intentionality cannot be evaded through a navigational model, since motor commands at the lowest physiological level do have incontestable intentional beginnings (Libet et al. 1983, Alexander et al. 2016). The question is rather: Which out of multiple intentions codified in a complex score is driving those motor commands and how is the choice to be made? Or whether (as we claim) embodiment navigates and assembles these fragmented intentions, without necessarily unifying them into a singular concept driving performance. There has been enough evidence to show that, the type of intentionality associated with music notation is often misleading, since most often contemporary scores codify both sensory outputs and motor commands alike in a symbolic way. In that sense, symbolic notation allows for multiple possible entry points and fosters exploration rather than prediction, at least in some stages of the learning process.

Fourth, the very basis of the *GesTCom* system materializing the embodied navigation paradigm is effectively prediction, in the form of probabilistic architectures (*Hidden Markov Models*), as will be shown in the next section. The question here is rather, to what extent are the states and layers of such models equivalent to object-concepts or mental representations. At the user level, those systems are clearly based on interaction dynamics of entrainment and alignment, and such dynamics have been shown to be able to evade representation.

Concluding: This section remains open-ended as to the compatibility of a model of embodied navigation with sensorimotor predictions. Computational limitations, interactive aesthetics bordering on indeterminacy, arguments from phenomenology and radical embodied cognition, and musical expression, all raise questions concerning the nature of intentionality, mental representations, algorithmic structuring and linearity, as well as expression, in predictive models. At the same time, there are notions that indicate potential compatibilities: The notion of action-oriented chunking; performance scenarios that tend towards a reproduction rather than a navigation model; the distinction between the intentionality of motor commands and intentionality in higher-order processes of learning and performance; and eventually, the Bayesian architecture of interactive systems themselves.

FOURTH SECTION

The *GesTCom*

In this section, we will overview an interactive system instantiating the embodied navigation model. The system highlights the two features of notational inter-complexity: the dynamicity and the multimodal grounding of notation during learning and performance. Its final aim is the constant adaptation of notation through the performer's physical movement. This adaptation would enable the coupling of performer and notation as oscillators, and would thus minimize the need for mental representations.

At its current state of development, the *GesTCom* is a sensor-based environment for the analysis, processing and real-time gestural control of complex piano notation through multimodal recordings. The acronym *GesTCom* stands for "Gesture Cutting through Textual Complexity". It is a project initiated at IRCAM (Institut de recherche et coordination acoustique / musique), Paris, in the context of the first author's 2014 Musical Research Residency in collaboration with the team *interaction-son-musique-*

mouvement. It continued at LabEx GREAM (Groupe de Recherches Expérimentales sur l'Acte Musical), Université de Strasbourg, in the context of his PhD, always in co-direction with the IRCAM.

a. Concept

The system's initial goal was set in the realm of *representation*:
How to address the limitations of prior means, such as the multilayered tablatures and intentionality nodes networks presented in the embodied navigation section. These tablatures and networks remain static, symbolic and subjective, while attempting to capture the dynamicity of embodied navigation. The *GesTCom* allows for the creation of malleable and objective, albeit personalized representations, through the inscription of recorded multimodal data, including sound, in the symbolic notation. A derivative goal was the representation of the longitudinal learning trajectory and the archiving of embodied navigation over time.

The system's further goals were set in the domain of *interaction*:
First, how to effectuate, rather than merely represent, the transformation of complex notation through physical movement. This transformation followed two directions: a) either simplification through information reduction, seeking efficiency in the first stages of the learning process; b) or proliferation of information, through complex tablatures offering augmented multimodal feedback, including sound.

A second requirement in the domain of interaction was the real-time control of these new notations through physical movement. In a nutshell, the *GesTCom* generates and controls multimodal interactive tablatures by means of inertial sensors non-invasively placed on the wrists of the performer. These interactive features effectively merge notation and instrument into an organic whole. Such interface exhibits Veitl's criteria of global instrumentality: materiality, visibility, readability, performativity, systemicity and causality (Veitl 2006).

Crucially for this paper, the radical vision of embodied navigation, as dynamic interaction of the elements of the performance system, with only contingent use for mental representations, is actively pursued: The performer externalizes the gestural processing of the notation, which becomes reproducible and communicable. Complex notation turns into a live signal to be directly perceived, through the updating of traditional annotational practices. The interactive dynamics of entrainment, alignment and sensorimotor learning (Leman 2016) become palpable in the relationship with the musical score, which turns into an oscillator coupled to the performer.

b. Methodology

The *GesTCom* methodology features the following steps:

a. Multimodal recordings of the performance of complex piano notation

b. Quantitative and qualitative analysis of the data and correlation to the original notation

c. Offline processing of the original notation on the basis of the data

d. Online interaction with the new output notation through inertial sensors

This methodology will be clarified later through dedicated examples, but we could describe it briefly as follows:

In the first step, the performer generates multimodal data through a recording of a performance of the original notation. In the second step, the quantitative and qualitative analysis of the data, in comparison to the qualitative analysis of the notation, results in a shared segmentation. On the basis of this segmentation, the original notation is processed and reshaped into a multilayered tablature in the third step. Eventually, again by virtue of this common segmentation and machine learning techniques, the multilayered tablature can be trained to follow the performer in variations of the initial performance. The performer, that is, interactively controls the tablature through gesture and expressively navigates networks of intentionality nodes in real-time.

Please note that the new output notation can be fed back into this loop, generating new performances, recordings and tablatures. Thus ensues the following interaction schema, which instantiates longitudinally the embodied navigation paradigm:

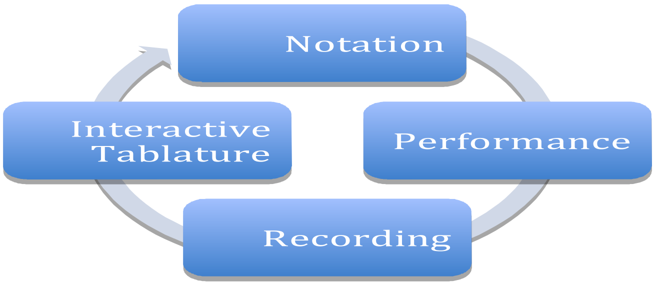

Figure 3. The *GesTCom* interaction schema

c. Architecture

The *GesTCom* is a modular system. In terms of hardware, it comprises systems for the capture of movement, audio, *kinect* video, MIDI and capacitive data from sensors on the piano keys. In terms of software, it is equipped with modules for the capture, analysis and control of the multimodal data, as well as modules for the augmentation and interactive control of music notation. Each of these systems functions both as stand-alone and integrated in the above methodology.

The first two steps, recording and analysis, are based on the library of Max / MSP[12] objects for multimodal analysis of sound and motion, interactive sound synthesis and machine learning known as MuBu[13] (multi-buffer) and developed at IRCAM by the ISMM[14](*interaction-son-musique-mouvement*) team.

---

[12] https://cycling74.com/ accessed 06.06.2020

[13] http://ismm.ircam.fr/mubu/ accessed 06.06.2020

[14] https://ismm.ircam.fr/ accessed 06.06.2020

The third step, notation processing, features INScore[15], a platform for multiple graphic representations, as well as customized command-line tools based on the Guido Engine Library[16]. Both are developed by Dominique Fober at GRAME (Centre national de création musicale), Lyon.

The last step, interaction, features:

a) the concept of gesture following, implemented in a customized *motionfollower* patch after the Gesture Follower project[17];

b) the connection of the motionfollower to the INScore representations through a dedicated Max / MSP patch. The outcome of this patch is an interactive multimodal tablature controlled through movement. Both the motionfollower and the *GesTCom* tablature have also been developed at IRCAM by the ISMM and the first author.

For a more technically detailed analysis of the systems and methodologies in question, including the relevant patches and the Hidden Markov Model probabilistic architecture of the motionfollower, please refer to (Antoniadis 2018, pp. 357-401), (Antoniadis & Bevilacqua 2016), (Antoniadis 2018b).

d. Selected examples

In the following selected examples, you may see some *GesTCom* applications for two complex piano works: Iannis Xenakis' *Mists* (videos 1 & 2) and Brian Ferneyhough's *Lemma-Icon-Epigram* (videos 3 & 4).

Video 1: https://www.youtube.com/watch?v=io9iGpVUAkI accessed 06.06.2020

Video 2: https://www.youtube.com/watch?v=Rql732JUm5M accessed 06.06.2020

Video 3: https://www.youtube.com/watch?v=YLB7uayipd4 accessed 06.06.2020

Video 4: https://www.youtube.com/watch?v=4rtRgaARiSU accessed 06.06.2020

i) Video 1

In the first video, the first two steps of the methodology are demonstrated (recording & analysis). You may watch the reproduction of multimodal data, recorded during the learning process of Iannis Xenakis *Mists*, p.1. In Figure 4 you may see an annotated snapshot of the Max/MSP patch that enables the reproduction and annotation of the data.

From top to bottom in the blue screen of the right side of the figure, there is a visual representation of the following datasets:

---

[15] http://inscore.sourceforge.net/ accessed 06.06.2020

[16] https://guido.grame.fr/ accessed 06.06.2020

[17] http://ismm.ircam.fr/gesture-follower/ accessed 06.06.2020

- stereophonic audio
- twelve gestural signals from inertial sensors on the player’s wrists. 3D acceleration data are shown with black, green and blue signals, for both the left hand (‘LH ACCEL’) and the right hand (‘RH ACCEL’); 3-axis gyroscopic data are shown with red, yellow and orange signals for the left hand (‘LH GYRO’) and the right hand (‘RH GYRO’).
- MIDI information from the keys and the two piano pedals. Color coding indicates velocity.
- Capacitive data from TouchKeys[18] sensors on the keys of the piano. Color coding indicates the position of the finger on the key, clusters are traces of hand-grasps.

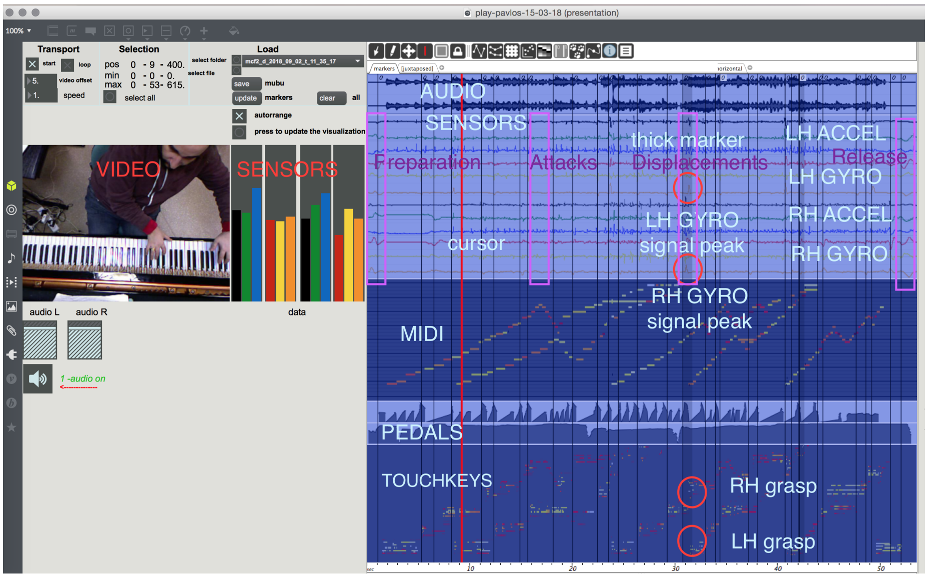

Figure 4. Max/MSP patch for the reproduction and analysis of recorded multimodal data (audio, movement, video, MIDI, TouchKeys)

The black markers superimposed over these datasets indicate a segmentation which is defined from both quantitative and qualitative data.

The quantitative data used here are the orange gyroscopic signals for both hands and the TouchKeys clusters. The orange gyroscopic signal peaks (shown with red circles in Figure 4) indicate two of many hand position changes, or hand displacements, which are visible in the video at the moment the red cursor crosses over the thick marker (00:29-00:30 in the video). This marker is thick because both hands are displaced towards the bottom of the keyboard, as opposed to thin markers, whereby only one hand is displaced.

The qualitative data used are performer’s annotations, that is multimodal tablatures as in Figure 2a. In this particular instance, it is the hand-grasp layer, indicated with blue and orange ellipses for the right and left hand respectively in Figure 2a. There is also

---

[18] http://touchkeys.co.uk/ accessed 06.06.2020

a qualitative comparison of the gestural data to the *kinect* video (left side of Figure 4), for visually confirming the displacement of the hands.

This pattern-matching between gyroscopic signals, TouchKeys data and annotated hand-grasps is indicative of the fact that subjective annotations can have an objective expression in multimodal data. In (Antoniadis & Bevilacqua 2016) we have in fact described a syntax of piano movement derived from multimodal data. This syntax takes the form of ‘movement envelopes’ consisting in movement Preparation, Attack, Displacement and Release phases. We called them ‘PADR envelopes’. In Figure 4, you may see the relevant purple boxes indicating some instances of those types of movement. Practically, attacks are accelerometer activations between the gyroscope activations that indicate displacements. So, in this particular example, attacks take place in-between the markers of displacements and the preparation and release gestures are visible before and after the sound (Figure 4).

On the basis of this segmentation, the traditional forms of annotation employed by performing musicians in symbolic form could be effectuated automatically through performance, in the form of multimodal data. Applying machine learning, one can recognize each of the four phases of the movement envelope. This application is currently in development.

Beyond the comparative analysis of movement and annotations, another function of this module is the monitoring of the learning process, including: the recording of practicing sessions, its use as an ‘augmented mirror’ in teaching, as well as for the longitudinal archiving of the process.

Finally, this multimodal analysis of performance can be used comparatively with score-based music analysis techniques, to offer insights about the relations between deeper compositional structures and performative responses. For example, in (Antoniadis 2018, pp. 405-462, “Embodying Algorithms”) we have offered a comparative analysis of types of texture and form compared to types of physical movement in Xenakis’ *Mists*.

ii) Video 2

In the second video**,** the two subsequent steps of the methodology are demonstrated (notation processing & interaction).

The segmentation discussed in the first video becomes the basis for the creation of a score-following system, which is sensitive to all sorts of deviations (tempo, articulation, dynamics, false notes, as explained in the video). This system is based on a probabilistic motion-following methodology employing Hidden Markov Models (Bevilacqua 2010) and on the syntax of piano-specific movement we showed above (Antoniadis & Bevilacqua 2016).

The novelty here is that the segmentation is defined by the performer, rather than a composer-defined symbolic score[19]. The performer defines the gestural units to be

---

[19] As is the case in state-of-the-art score following systems, such as *Antescofo* https://www.antescofo.com/ accessed 06.06.2020

followed during a ‘recording phase’ (not to be confused with the recording presented in the previous video!), and the system couples with a great range of the player’s variations in the ‘following phase’. In particular, the system is probabilistically comparing the incoming varied gestures during ‘following’, to the originally recorded gestures during ‘recording’. This module may generate simpler scores for real-time performance, or even be used for the control of live electronics, page turning and other interactive applications.

This simplification of the notation in the form of a reduced proportional representation (Figure 5) has substituted the initial complex notation by Xenakis. This feature might be useful not only for real-time performance, but also in the first stages of the learning process.

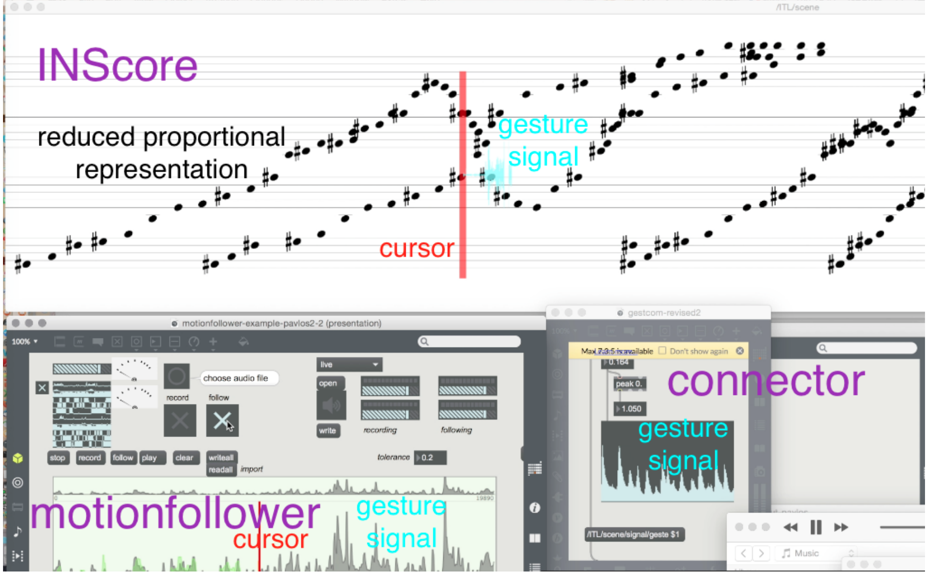

Figure 5. Augmented interactive score controlled through sensors and connected to the motionfollower through a connector Max/MSP patch.

Let’s take now a closer look at this module, since it will be important in the next video example as well. It is consisting of one INScore representation, connected interactively to the motionfollower Max/MSP patch through another Max/MSP patch (indicated in purple as ‘INScore’, ‘motionfollower’, ‘connector’, Figure 5).

An INScore script generates this augmented interactive score, which consists of the following graphic objects: the reduced proportional score representation, a cursor and a signal, as shown in Figure 5. The reduced proportional notation has been generated automatically by the MIDI data shown in video 1/ Figure 4, using command-line tools based on the GUIDO Engine and developed by Dominique Fober. The cursor in red and the signal in blue are controlled through the inertial sensors in the wrists of the pianist. The signal in blue comes from the motionfollower.

The score’s interactive possibilities are based on the motionfollower, an object in Max / MSP and a customized patch shown in Figure 6. In a first phase, a gesture is recorded (as in the video, 00:00 – 00:25). This gesture is represented by the grey signal in Figure 6. This signal is the sum of the twelve signals we saw in video 1/ Figure 4, plus audio energy. In a second phase (00:25 – 01:24), this gesture is compared probabilistically to a new, incoming gesture, represented by the green signal in Figure 6, which is superimposed over the grey signal. The system essentially predicts the probability of the new gesture being similar enough to the recorded gesture. If this is the case, the system follows the player, and this following is indicated by the smooth movement of the cursor. If not, the cursor is moving with a certain viscosity, gets stuck, jumps abruptly forward or is waiting for the performer and so on. Such examples of failed followings will be shown in the third video example.

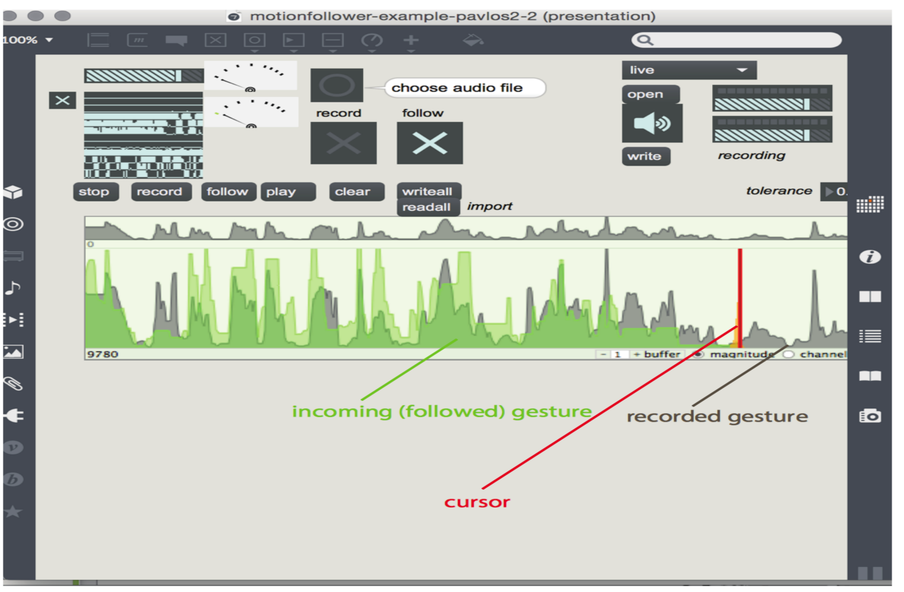

Figure 6. Max/MSP patch for motion following.

On top of the visual feedback, the motionfollower may offer sonic feedback, as the initial recording is of both movement and sound.

The third component is a Max/MSP patch which functions as a connector, sending the incoming, new signal of the motionfollower, to the INScore tablature in the form of Open Sound Control[20] messages.

The crucial element, which allows for the motion-following to be reflected in the notation and thus become score-following, is that both the gesture and the notation are sharing the same basic segmentation.

In the recording phase (00:00 - 00:25), the user follows any mobile element of the INScore, in our case the red cursor, which is set to move at a desired speed, like a classic metronome would do. The musical score has already been graphically segmented, according to the movement segmentation defined in video 1, and assigned

---

[20] http://opensoundcontrol.org/introduction-osc accessed 06.06.2020

a duration according to a specific INScore space-time formalism (explicit mapping). In this recording phase, the motionfollower learns, so to speak, the mapping from the performer (implicit mapping), who follows the mapping of the INScore (explicit mapping). In the ‘following’ phase, the performer can pursue highly varied performances: a faster performance (00:25 – 00:45 in video 2), a faster performance with softer dynamics (00:45 – 01:02), a performance with different (staccato) articulation and even mistaken notes (01:03 – 01:24). This time, it is not the performer that follows the system, but rather the system that follows the performer, given that the segmentation is correct and common in all these varied performances. Thus, the performer may control the mobile elements of the INScore tablature. The feedback of the follower has been extended to score compound representations. The gesture-following has been turned into score -following.

iii) Video 3

In the third video, we have a more sophisticated variation of the previous example, featuring notation processing & interaction. This example features the expressive navigation of embodied layers and networks of intentionality nodes, effectively simulating the dynamics of the embodied navigation theory.

The *GesTCom* system is trained on the basis of the notation on the right side of the screen, as shown in Figure 7. It is a reduced version of the first page of Ferneyhough's *Lemma-Icon-Epigram,* representing the grasp-layer of the original notation, as shown in Figure 8. This version is reduced, in the sense that only a small fraction of the original note information is included. Red notes indicate fingers five and blue notes indicate thumbs, so that the grasp-layer is defined as segments between fingers one and five, omitting the intermediate information (for more information on this reduction process, please refer to (Antoniadis & Bevilacqua 2016).

Having trained the system on this plain example, one can fill in the complex multi-parametrical information, in the same way that one would learn the piece in top-down fashion (Ferneyhough 1982): starting with global gestures and adding later the notational details.

The system can couple to varied performances, not only in the sense of expressive deviations as in the second video, but also in the broader sense of navigation of different embodied layers (grasps or fingers), loops and distortions of the material as shown in the video:

00:00 – 00:28 present an unsuccessful attempt to follow the player performing the grasp-layer, which is shown in the reduced notation on the right-hand side of the video and in Figure 7. The cursor and the sound are momentarily stuck in 00:18 and 00:26 and then they don’t follow until the end. This indicates that the new gesture (green signal) was not computer as similar enough to the recorded gesture (grey signal. The recording of this signal is not shown in the video).

00:29 – 00:39 demonstrate a successful attempt to follow the performance of the grasp-layer in the reduced notation (Figure 7). The movement of the cursor and sound are smooth and follow until the end. The green and grey signals are characteristically similar.

00:40 – 00:54 show a successful attempt to follow the performance of the original non-reduced notation by Ferneyhough (Figure 8), always on the basis of the grasp-layer (Figure 7): What one hears in the video is a performance of the score in Figure 8, the system however follows the segmentation which is shown in the right-hand side of the video and in Figure 7. The fact that the system follows indicates that the gesture to play the full score in Figure 8 and the gesture to play the reduction in Figure 7 (the recorded gesture not shown in the video) share a common segmentation, as in video 1.

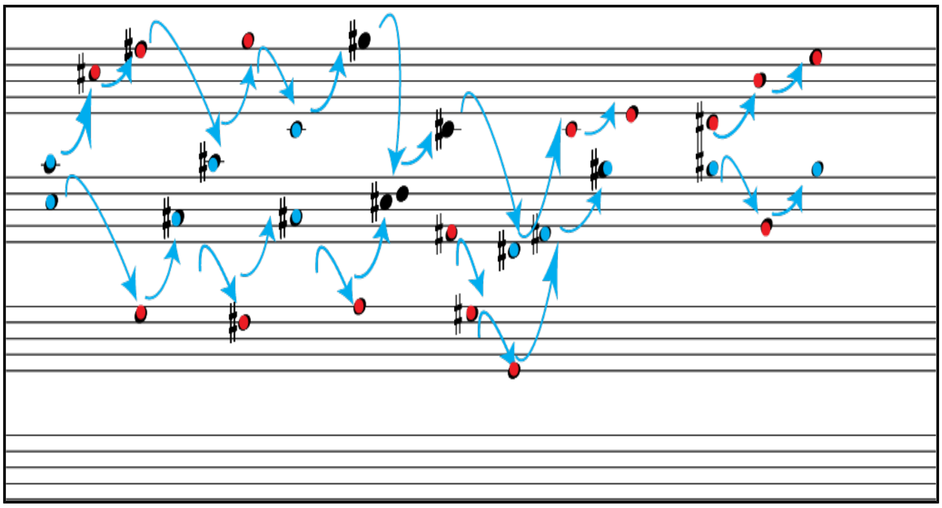

Figure 7. Reduced version of Ferneyhough's notation in *Lemma-Icon-Epigram*, p.1. Red notes indicate the fifth fingers and blue notes the thumbs for both hands.

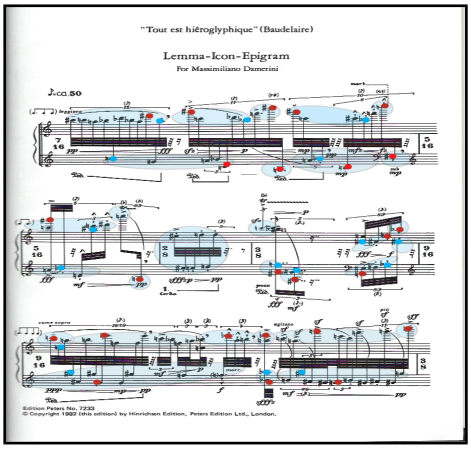

Figure 8. Annotation indicating the hand grasp layer on Ferneyhough's original score of *Lemma-Icon-Epigram*, p.1. Reproduced with kind permission by Peters Edition.

The degree of variation here is already much greater than sheer expressive deviations as in video 2.

00:57 – 01:17 show a successful attempt to follow a performance of the original notation, but much slower. The beginning is unsuccessful because the follower does not recognize the preparation gesture, but upon restarting the performance, the system follows.

01:18 – 01:42 shows a following of a performance which is even more varied. The original is played slower with staccato articulation and there is an intentional looping towards the end (01:35), whereby the cursor stops and waits for the performer to proceed correctly.

01:43 – 02:01 features the most varied performance of all, a literally distorted version of the original, whereby tempi, pitch, articulation dynamics are varied, even new material is added. However, since the basic segmentation of the hand-grasp layer remains intact, the system can still follow.

Next to the visual information in the form of representations, signals and cursors, the system offers audio feedback of the first recording.

The third module, then, is a simulation of the embodied navigation. It can be used for top-down learning on the basis of simplified information.

iv) Video 4

In the fourth video, a different instance of multimodal augmentation (as opposed to the reductions above) of the notation is presented (Figure 9). It features the synchronization of several videos, signals, symbolic notations, audio through the INScore module of the system and its connection to the play-patch of video 1 (Figure 4). The reduced notation presented in Figure 7 is used both in the video on top of Figure 9 (element a.), in the form of a scrolling notation rotated by ninety degrees clockwise (element b.), as well as in the compound representation below (element d.), in comparison to the annotated gestural signals (element c.) and to the original annotated score (element e.). The red numbers on top of this reduced notation indicate the number of hand displacements in each bar of the original notation. Finally, there is a live gestural signal in red (element f.), which allows for the live interactive control of this complex tablature (please note that this signal appears blue in the video). Such representation could be too complex for real-time performance, it does indicate though the possibilities of the system as far as the multiplicity of external representations in performance is concerned. It may function as a valuable source of information for performance analysis and even as an artifact in itself. And it can certainly be connected to the motionfollower as in videos 2 & 3 and interactively controlled.

Summarizing: After a recording phase of multimodal data, as in the first video, several strategies of either reduction (second and third) or augmentation of the notation (fourth video) are combined with machine learning techniques in the form of this motion following architecture. The result is a system which aims at simulating the dynamics of human learning as embodied navigation. The system learns from the performer and at the same time aids the performer to learn, in a mutual adaptation process. In other words, the system notation – performer turns into a system resembling to systems of coupled oscillators. As shown in the first section, the behavior of such systems can function as an epistemological replacement for mental representations, in the sense that the system can be explained and understood without mental representations.

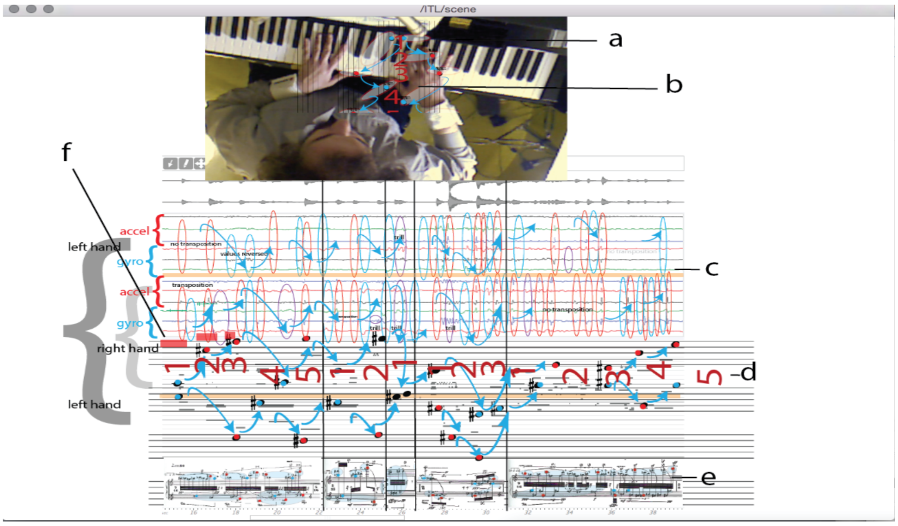

Figure 9. Compound interactive representation featuring a. video, b. video annotation presenting the reduced notation of Figure 7 scrolling inside the first video, c. annotated gestural signals, d. reduced notation of Figure 7 with numbers of hand displacements per bar, e. original notation, f. live gestural signal in red

e. Features for an anti-representationalist and anti-computationalist approach to musical performance

In relation to the embodied navigation paradigm, we may summarize the *GesTCom* technical features and conceptual contributions as follows:

In the realm of representation:
a. Registration of real physical energies as opposed to mere symbolic representation.
b. Monitoring of the learning process.
c. Longitudinal archiving in both physical (data) and symbolic (notation) form.
d. Dynamic as opposed to static representation of the embodied navigation paradigm.
e. Externalization of the internal mental processes associated with learning.
f. Extension of the traditional annotation practices through technology.
g. Reproducibility and communicability of the learning processes.

In the realm of interaction:
a. Simplification of the complex notation through performance data.

b. Proliferation of the representations and augmented multimodal feedback.
c. Real-time gestural control of the symbolic and multimodal notation.
d. Effective transformation of the notation into an interface.
e. Radical embodied paradigm with contingent use for mental representations.
f. Direct perception of notation as signal.
g. Entrainment, alignment and sensorimotor learning in the interaction with symbolic information.

Through those features, the notation is turned into an oscillator to be coupled in real time with the performer. As indicated in the notation of chapter 1, {(T) → MR1} → XR = rosc ↔ (MR2) ↔ T*, the usual role of mental representations on the part of the performer is thus outsourced to the adaptive notation itself.

f. Applications

The applications envisaged for the *GesTCom system* include a broad range of musical activities. Beyond the applications in embodied interactive learning, performance analysis and score-following demonstrated in the videos, the *GesTCom* has already been used as a tool for musical composition and in live improvisation contexts[21]. Currently (June-December 2020), the *GesTCom* is employed for the creation of an interactive multimodal recording of the work for solo piano *Evryali* by Iannis Xenakis, integrating full body motion capture data and creating a virtual score-space to be navigated by the listener/viewer (post-doc project at ArTeC / Université Paris 8 / IRCAM, in collaboration with Jean-François Jégo and Makis Solomos).

As far as user testing is concerned, the prototype system has for the moment mainly been the object of the first author’s artistic research. Its development has been based on recordings made by the first author himself over the course of six years. The system has however successfully been accessible to other pianist/users in several public settings, including conferences, pedagogical workshops and human-machine interaction workshops[22]. In these contexts, pedagogical applications proved very promising.[23] User testing following strict Human-Computer Interaction and New Interfaces for Musical Expression protocols is planned as a post-doc project at the Technische Universität Berlin, Abteilung Audiokommunikation with funding from the Humboldt Stiftung (2021-2022).

Future work

Further correspondences between embodied navigation / the *GesTCom* and radical embodied cognition are envisaged:

There have been decades of experimental work and rigorous dynamical modeling of human rhythmic behavior that treats the behavior in terms of coupled oscillators (Haken, Kelso and , Bunz 1985). This research treats abilities to produce particular rhythmic as intertwined with the ability to produce other patterns. Thus, learning to drum in 9:4 is not just adding an ability, but also transforms our abilities to produce other rhythmic patterns, making some previously patterns more difficult to maintain

---

[21] Antoniadis 2018, pp. 568-570 http://panosghikas.com/unrealtime/#PAVLOS_MENU, accessed 06.06.2020

[22] Antoniadis 2018, pp. 570-582

[23] Antoniadis 2018, pp. 565-567

(Amazeen, Sternad, and Turvey 1996). Amazeen et al. (1996) characterize this interconnected of abilities in terms of the *deformation of the space of the possible rhythmic patterns* an individual can perform. Given the renewed notion of the score as interface in the *GesTCom* methodology, its transformation effectively into an oscillator that couples with the performer, one could envisage modeling musical learning in the same terms. That is, learning a new score is not just an addition to a performer's repertoire, but is instead a transformation of the entire space of playing abilities. Such a transformation should be detectable in experimental settings and should be able to be modeled using the same dynamical models that other are used in rhythmic behavior generally.

Similarly, Chemero's dynamic theory of *affordances 2.0* (Chemero 2009, p. 153), which describes the interaction of affordances and sensorimotor abilities over time, could further contribute to an understanding of how learning operates on different scales: the learning of a single work, how this contributes to the longitudinal evolution of a single pianist (ontogenesis) and even how learning evolves between generations and historically (phylogenesis). The addition of an interactive loop with processes of neural plasticity to the interaction between the pianist and the score, under the notion of Varela's (Varela, Thompson and Rosch 1991) *autopoiesis*, could provide us with a very complete theory of musical learning.

Eventually, the rich work on perception by dynamic touch (Gibson 1962; Shockley, Carello, and Turvey 2004) could offer a genuinely anti-visual and anti-representational further direction as to musical learning: It could be shown that the pianist's perception of the affordance *moveability*, measured through an inertial tensor and in relation to the grasp-layer in the embodied navigation methodology, is actually correlated with the complexity of the musical score, which is literally felt and 'computed' in the wrist of the pianist. 'Understanding the score' could then be theorized as a form of dynamic touch of the notation, materializing the thus far only metaphorical notion of the notation as TUI – Tangible User Interface.

Conclusions

In the previous four chapters, we have shown how a renewed notion of notational complexity invites a theoretical model of embodied navigation inspired by radical embodied cognitive science. Subsequently, this theoretical model was materialized in a dedicated interactive system, the *GesTCom*, which allows for playing and learning in a way that does not privilege internalization and mental representation.

In the first section, we showed how the instantiation of mental representations into systems of coupled oscillators provided the theoretical springboard for a transformation of the musical communicative chain, whereby a musical score-oscillator would couple with the performer, rendering her mental representations contingent.

In the second section, we showed why musical complexity in its purely graphical form problematizes mental representations and invites a new ecological approach to performance.

In the third section, we presented this approach as 'embodied navigation of the notation'. In this approach, notational complexity is considered from the point of view of ecological psychology and dynamic systems theory as a *state-space of affordances* to be navigated in an embodied way. We juxtaposed this approach to both traditional memorization strategies and theories of predictive processing in music making.

In the fourth section, we addressed issues of representation and interactivity in the embodied navigation model through the development of the *GesTCom* interactive system for live control of the notation. This system effectuates the concept of a notation-oscillator that couples with the performer.

While music has always been very fit as a case-study for radical embodied cognition, it was rather audio-based phenomena, like beat perception, which were foregrounded. Performance of notated music has been underrepresented, although it offers a higher-order task, combining auditory perception with bodily movement and symbolic communication through musical scores. The conception of musical scores as oscillators that couple with performers can open up new perspectives for both music-making and radical embodied cognitive science.

Acknowledgments

We wish to express our gratitude to Frédéric Bevilacqua, Riccardo Borghesi and Dominique Fober for their ongoing contributions to the concept, design and development of the *GesTCom* system. This paper is dedicated to the memory of the Greek musicologist Katy Romanou (1939-2020), who passed away on May 22, 2020.

References

Alexander, P., Schlegel, A., Sinnott-Armstrong, W., Roskies, A. L., Wheatley, T., & Tse, P.U. (2016). "Readiness potentials driven by non-motoric processes. Consciousness and Cognition", 39, 38–47. https://doi.org/10.1016/j.concog.2015.11.011

Amazeen, E. L., D. Sternad, and M. T. Turvey (1996). "Predicting the nonlinear shift of stable equilibria in interlimb rhythmic coordination". *Human Movement Science*, 15, 521–542.

Antoniadis, P. and Bevilacqua, F. (2016)."Processing of symbolic music notation via multimodal performance data: Ferneyhough's *Lemma-Icon-Epigram* for solo piano, phase 1" in *Proceedings of the TENOR 2016 conference*, p. 127-136. Cambridge: Anglia Ruskin University 2016.
http://tenor2016.tenor-conference.org/TENOR2016-Proceedings.pdf, accessed 06.06.2020

Antoniadis, P. (2018). *Embodied Navigation of Complex Piano Notation: Rethinking Musical Interaction From A Performer's Perspective*, PhD thesis. Strasbourg :Université de Strasbourg – IRCAM, 2018.
http://theses.unistra.fr/ori-oai-search/notice/view/2018STRAC007?height=500&width=900, accessed 06.06.2020

Antoniadis, P. (2018b). "GesTCom: A sensor-based environment for the analysis, processing and real-time control of complex piano notation through multimodal recordings", invited talk at *Séminaires Recherche et Technologie*, IRCAM, 15.10.2018.
https://medias.ircam.fr/x2253e1 accessed 06.06.2020

Arbo, A., Ruta, M.(eds.) (2014). *Ontologie Musicale: Perspectives et débats*. Paris: Hermann

Barsalou, L. (1999). “Perceptual Symbol Systems” in *Behavioral and Brain Sciences,* 22: pp. 577–609.

Beer, R. (2003). “The Dynamics of Active Categorical Perception in an Evolved Model Agent”. *Adaptive Behavior*, 11, 209–43.

Bevilacqua, F., Schnell N., Rasamimanana N., Zamborlin B., Guedy F. (2011). “Online Gesture Analysis and Control of Audio Processing” in: Solis, J. & Ng. K. (Eds.): *Musical Robots and Interactive Multimodal Systems*. Berlin Heidelberg: Springer-Verlag, pp. 127-142.

Boulez, P. (1986).“Time, notation and coding” in Nattiez, J.J. (ed.), *Orientations*. London: Faber & Faber, pp. 84-89.

Breithaupt, R.M. (1905). *Die natürliche Klaviertechnik.* Leipzig : C. F. Kahnt Nachfolger

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge Massachusetts: MIT Press.

Clark, A. (1997). *Being There*. Cambridge, Mass.: MIT Press.

Clark, A. (2016). *Surfing Uncertainty. Prediction, Action and the Embodied Mind*. New York: Oxford University Press.

Cook, N. (2013). *Beyond the Score: Music as Performance*. Oxford: Oxford University Press.

Cox, F. (2002). “Notes Toward a Performance Practice for Complex Music” in Mahnkopf, C.-S., Cox, F. and Schurig, W. (Eds.): *Polyphony and Complexity, New Music and Aesthetics in the 21st Century*, i. Hofheim: Wolke Verlag, pp. 70-132.

Di Scipio, A. (2015). “The Politics of Sound and the Biopolitics of Music: Weaving together sound-making, irreducible listening, and the physical and cultural environment” in *Organised Sound* 20(3): 278–289. Cambridge: Cambridge University Press.

Ferneyhough, B. (1982). *Lemma-Icon-Epigram for solo piano*. Edition Peters, No. 7233.

Ferneyhough, B. (1995). “Aspects of Notational and Compositional Practice” in Boros, J. and Toop, R. (Eds.): *Brian Ferneyhough: Collected Writings*. London: Routledge, pp. 2-13.

Fitzhugh, R. (1961). “Impulses and physiological states in theoretical models of nerve membrane” in *Biophysical Journal*, 1, 445.

Floros, C. (1980). *Einführung in die Neumenkunde*. Noetzel Verlag.

Fodor, J. (1975). *The Language of Thought*. London: Thomas Crowell.

Gibson, J. (1962). “Observations on active touch”. *Psychological Review*, 69, 477–490.

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

Glenberg, A. M. and Robertson D. A.(1999). “Indexical understanding of instructions” in *Discourse Processes*, 28:1, pp. 1-26.

Godøy, R.I. (2011).“Sound-Action Chunks in Music”. In: Solis, S. & Ng, K. (eds.) *Musical Robots and Interactive Multimodal Systems*. Berlin-Heidelberg: Springer

Grush, R. (1997). “The architecture of representation”. Philosophical Psychology, 10, 5–

24.

Haken, H., J. A. S. Kelso, and H. Bunz (1985). “A theoretical model of phase transitions in human hand movements” in *Biological Cybernetics*, 51, 347–356.

Haugeland John (1991). “Representational genera” in *Philosophy and Connectionist Theory*, ed. Ramsey,W., Stich, S. and Rumelhart, D. (Eds.). Hillsdale, N.J.: Erlbaum.

Hübler, K. K. (2002). “Expanding String Technique”, in Mahnkopf, C.-S., Cox, F. and Schurig, W. (Eds.): *Polyphony and Complexity, New Music and Aesthetics in the 21st Century*, i. Hofheim: Wolke Verlag, pp. 233-244.

Ishii, H. (1997) “Tangible bits: towards seamless interfaces between people, bits and atoms” in: Proceedings of the *ACM SIGCHI Conference on Human factors in computing systems CHI '97*, 234-241.

Lachenmann, H. (2004). *Musik als Existentielle Erfahrung, Schriften 1966-1995*. ed. Josef Häusler. Wiesbaden: Breitkopf und Härtel.

Lalitte, P. (2015). *Analyser l’interprétation de la musique du XXe siècle. Une analyse d’interpretations enregistées des Dix pièces pour quintette à vent de György Ligeti*. Paris: Hermann.

Leimer, K. and Gieseking, W. (1972/1931). *Piano Technique*. New York: Dover.

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. Cambridge*:* MIT Press.

Leman, M. (2016). *The Expressive Moment. How Interaction (with Music) Shapes Human Empowerment*. Cambridge: MIT Press.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). “Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (readiness-Potential)”. Brain, 106(3), 623–642. https://doi.org/10.1093/brain/106.3.623

Mahnkopf, C. S. (2002). “Complex music: an attempt at a definition” n Mahnkopf, C.-S., Cox, F. and Schurig, W. (Eds.): *Polyphony and Complexity, New Music and Aesthetics in the 21st Century*, i. Hofheim: Wolke Verlag, 54-64.

Manoury, P. (1998). *La note et le son : écrits et entretiens (1981-1998)*. Paris: Harmattan

Markman, A. B., and E. Dietrich (2000a). “In defense of representation” in *Cognitive Psychology*, 40, 138–171.

Markman, A. B., and E. Dietrich (2000b). “Extending the classical view of representation” in *Trends in Cognitive Sciences*, 4, 70–75.

McAuley, J. D. (1996). “On the nature of timing mechanisms in cognition” in *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. Hillsdale, N.J.: Erlbaum.

Merleau-Ponty, M. (1962). *Phenomenology of Perception*. London: Routledge & Paul.

Millikan R.G. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, Mass.: MIT Press.

Morgan, Alex (2013). “Representations Gone Mental”, in *Synthese, An International Journal for Epistemology, Methodology and Philosophy of Science*, Berlin Heidelberg: Springer

Nagumo, J., S. Arimoto, and S. Yoshizawa (1962). “An active pulse transmission line simulating nerve axon" in *Proceedings of the Institute for Radio Engineers*, 50, 2061–2070.

Pace, I. (2007). "Complexity as Imaginative Stimulant: Issues of Rubato, Barring, Grouping, Accentuation and Articulation in Contemporary Music, with Examples from Boulez, Carter, Feldman, Kagel, Finnissy". Paper given at conference *Tempo, Meter, Rhythm. Time in Music after 1950*, Orpheus Institute, Ghent, April 11-14. https://openaccess.city.ac.uk/id/eprint/6472/ accessed 06.06.2020

Port, R., and T. van Gelder (1995). *Mind as Motion*. Cambridge, Mass.: MIT Press.

Ramsey, W. (2007). *Representation reconsidered*. Cambridge: Cambridge University Press.

Rowlands, M. (1999). *The Body In Mind: Understanding Cognitive Processes*. Cambridge: Cambridge University Press.

Rowlands, M. (2010). *The New Science of the Mind. From Extended Mind to Embodied Phenomenology*. Cambridge, Massachusetts: MIT Press.

Sándor, G. (1981). *On Piano Playing. Motion, Sound and Expression*. New York: Schirmer.

Schick, S. (2006). *The Percussionist's Art. Same Bed, Different Dreams*. New York: University of Rochester Press.

Shockley, K., C. Carello, and M. T. Turvey (2004). "Metamers in the haptic perception of heaviness and moveableness". *Perception and Psychophysics*, 66, 731–742.

Smith, B. C. (1996). *On the Origin of Objects*. Cambridge, Mass.: MIT Press.

Solomos, M. (2013). *De la musique au son. L'émergence du son dans la musique des XXe-XXIe siècles*. Rennes: Presses universitaire de Rennes. Aesthetica, Pierre-Henry Frangne, Roger Pouivet, 978-2-7535-2638-9. hal-00769893. https://hal.archives-ouvertes.fr/hal-00769893/document accessed 06.06.2020

Thelen, E., Smith, L. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Toop, R. (1993). "On Complexity" in *Perspectives of New Music,* 31, 1993, pp. 42-57.

Toop, R. (2010). "Against a Theory of Musical (New) Complexity" in Paddison, M., Deliese, I., *Contemporary Music, Theoretical and Philosophical Perspectives*, Surrey: Ashgate.

Van Gelder, T. (1995). "What Might Cognition Be, If Not Computation" in *Journal of Philosophy* , 92, 1995, pp. 345–381.

Varela, F., E. Thompson, and E. Rosch (1991). *The Embodied Mind*. Cambridge, Mass.: MIT Press.

Varga, B. (1996). *Conversations with Iannis Xenakis*. London: Faber and Faber.

Veitl, A. (2006). "Musique, causalité et écriture: Mathews, Risset, Cadoz et les recherches en synthèse numérique des sons"in Kanach (ed.), *Musicologie, informatique et nouvelles technologies*. Paris-Sorbonne : Observatoire Musical Français, pp. 121-128.

Xenakis, I. (1980). *Mists pour piano*. Editions Salabert

Biographies

Pavlos Antoniadis (PhD in musicology, University of Strasbourg-IRCAM; MA in piano performance, University of California, San Diego; MA in musicology, University of Athens) is a pianist, musicologist and technologist from Korydallos, Athens, Greece, currently based in France. He performs complex contemporary and experimental music, studies embodied cognition and develops tools for technology-enhanced learning & performance. He is postdoctoral researcher at EUR ArTeC, Paris 8 & IRCAM, and following up at the Berlin Institute of Technology (TU-Audiokommunikation) as a Humboldt Stiftung scholar.

Tony Chemero is Distinguished Research Professor of Philosophy and Psychology and a core member of the Center for Cognition, Action, and Perception at the University of Cincinnati. He got his Ph.D. in Philosophy and Cognitive Science from Indiana University in 1999. After that, he taught at Franklin & Marshall College, where he was Professor of Psychology. In 2012, he became Professor of Philosophy and Psychology at the University of Cincinnati. Tony's research is both philosophical and empirical. It is focused on questions related to dynamical modeling, mental representation, ecological psychology, artificial life and complex systems. He is author of more than 60 articles and the book Radical Embodied Cognitive Science (2009), which is part of the Bradford Books series from MIT Press. Tony lives in Cincinnati, OH, USA.