Gesture cutting through textual complexity: Towards a tool for online gestural analysis and control of complex piano notation processing

Pavlos Antoniadis[1], Frédéric Bevilacqua[2], Dominique Fober[3]
[1]*GREAM, Université de Strasbourg-Ircam,* [2]*STMS-Ircam-CNRS-UPMC, Paris,* [3]*GRAME-Lyon*
Correspondence should be addressed to: info@pavlosantoniadis.com, katapataptwsi@yahoo.gr

***Abstract*: This project introduces a recently developed prototype for real-time processing and control of complex piano notation through the pianist's gesture. The tool materializes an *embodied cognition*-influenced paradigm of interaction of pianists with complex notation (*embodied* or *corporeal navigation*), drawing from latest developments in the computer music fields of musical representation (augmented and interactive musical scores via INScore) and of multimodal interaction (Gesture Follower). Gestural, video, audio and MIDI data are appropriately mapped on the musical score, turning it into a personalized, dynamic, multimodal tablature. This tablature may be used for efficient learning, performance and archiving, with potential applications in pedagogy, composition, improvisation and score following. The underlying metaphor for such a tool is that instrumentalists *touch* or *cut through* notational complexity using performative gestures, as much as they touch their own keyboards. Their action on the instrument forms integral part of their understanding, which can be represented as a gestural processing of the notation. Next to the already mentioned applications, new perspectives in piano performance of post-1945 complex notation and in musicology ('performative turn'), as well as the emerging field of 'embodied and extended cognition', are indispensable for this project.**

1. INTRODUCTION AND STATE OF THE ART

Despite the astonishing heightening of technical standards in musical performance, its premises remain heavily attached to an interpretative model of the past. This model privileges compositional abstract thinking, which is represented as musical notation, to be further „sonified" by the performer. This hierarchy theorizes performance as a transparent channel between the sender-composer and the receiver-listener.

Such a model of musical communication seems to ignore recent developments in aesthetics, cognitive science and computer music technology. Our interdisciplinary research attempts to integrate perspectives from those fields into a revision of interpretation today. In particular, the emphasis on performativity in modern aesthetics, the importance of action in cognition and the field of computer music interaction, form the background of this research.

1.1. Background in performance practice and music technology

Developments in contemporary composition have problematized notation as a transparent interface linking compositional intentionality to performative response. Paradigmatic in this respect is the work of Brian Ferneyhough, which programmatically employs complex notation for inviting multiple interpretational strategies and sonic results, as described in [1]; or the work of Iannis Xenakis, where extremes of physicality function as a performer-specific perspectival point to complex notation, as shown in [2] and [3]. In such cases, the traditional performative paradigm seems to be sabotaged: *Understanding* the notation cannot anymore function as the prerequisite of instrumental *technique* towards an expressive *interpretation*. Our research attempts to offer an embodied, medial and performer-specific alternative to this linear arrangement of *Understanding-Technique-Interpretation*. We refer to it as the UTI paradigm[1].

The UTI aporia echoes a general performative turn in musicology, theatrology and cultural studies. A wholly new set of notions (*event* instead of *work*, *presence* instead of *representation*) and, more importantly, the notions of *embodiment* and *materiality,* become central for a new aesthetic of the performative, as in [5].

The case for a performer-specific theory and praxis finds further defence in the field of *embodied and extended cognition*. This interdisciplinary field has been embraced in recent years by music psychologists, who deal with embodiment, mediation, movement and gesture, as in [6],[7],[8]. The underlying thesis of these studies is that (music) cognition is not reducible to its neural implementation, but is rather distributed among the brain, the body and the environment. This thesis ontologically upgrades gesture and movement into equal components of cognition, potentially resulting in genuine reflection on the UTI aporia. Some basic sources for the field, including J.J. Gibson's "Urtext" *The Ecological Approach to Visual Perception*, will be referenced in detail in 1.2.

The enhanced role of action in music cognition, in combination with the increasing availability of low-cost sensors and interfaces in the turn of the 21st century, become the central parameters in the emerging field of computer music interaction, as documented in [9]. Gestural data can today effectively be captured, analyzed and mapped upon other modalities, paradigmatically sound. This fact opens the way for novel interaction concepts and for the design of interactive multimodal systems and musical robots. The process can be closely tracked down in the context of the NIME (New Interfaces for Musical Expression) conferences since 2001. Those developments remain to be democratized for the larger community of classically trained performers. Complementary to gesture and movement interaction, of special importance for this research is the field of computer music representation, in particular platforms for interactive augmented musical scores. Those platforms provide a link between computer music representation and interaction to be further explored.

1.2. Corporeal Navigation

The concept of *corporeal* (or *embodied*) *navigation* attempts to offer an embodied and medial performer-specific alternative to the UTI paradigm. Instead of a strictly linear arrangement of its formants -understanding notation, then employing purposefully technique and then allowing, in the end, for expressive interpretation-, it proposes the conceptualization of learning and performance as embodied *navigation* in a non-linear notational space of *affordances*[2]: The performer "moves" inside the score in several dimensions and manipulates in real-time the elements of notation *as if* they were physical objects, with the very same gestures that s/he actually performs. This manipulation forms indispensable part of the cognitive processes involved in learning and performing and transforms the notation. This transformation can be represented as a multilayered tablature, as in the following

[1] For a more detailed discussion of the UTI paradigm as manifested in performers' and composers' discourses, from Karl Leimer and Walter

[2] Both terms, *navigation* and *affordance*, are direct references to J. J. Gibson's work, as in [10].

"simple" example of Fig. 1 (simple in the sense that it deals only with the parameters of pitch and texture) :

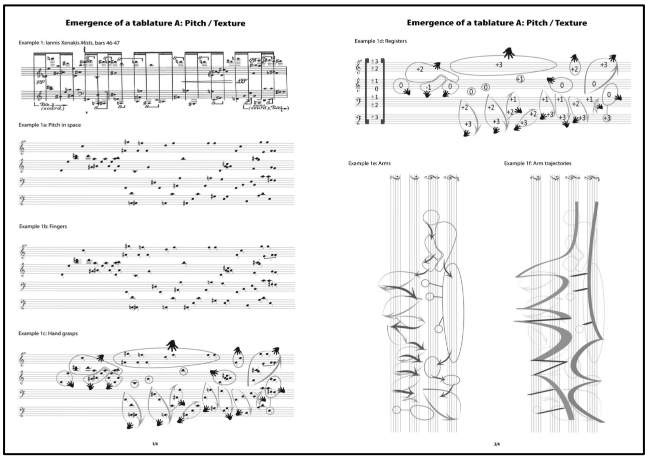


Fig. 1 The embodiment of a Xenakian cloud / fingers-, hand-, and arm-layer in 1b, 1c, 1f respectively

Next to this gestural template, the score-space involves dimensions of continuity and discontinuity according to compositional parameters, as well as the dimension of a singular passage through it: an irreversible, linear, actual performance. An example of coupling with compositional parameters is offered in Fig. 2:

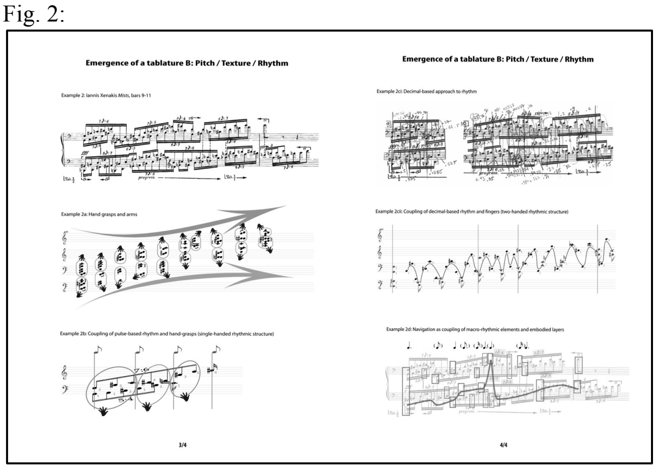


Fig.2 Coupling of gestural template (2a) with complex rhythm in a Xenakian linear random walk. Embodiment of a pulse-based (2b) and decimal-based (2c) approach to rhythm; macrorhythmic prioritizations (2d) and emerging continuities and discontinuities in relation to 2a.

In a nutshell, corporeal navigation signifies the perpetual movement in-between embodied representations of the immobile score-space. This movement produces 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 out of the initial notational image. Interpretation consists in this diachronic movement, rather than in the repetition of a fixed sound-image.

The notion of corporeal navigation draws from developments in the field of embodied and extended cognition (EEC), such as: the notion of the manipulation of external information-bearing structures (here notation), and of action in general, as constitutive of cognition ([11], [12]) ; the notion of self-organized systems and emergent behaviors (the system embodied mind, instrument, notation would be seen as such) from dynamic systems theory[3]; the notions of navigation and affordance from Gibson's ecological psychology as cited above; the notion of conceptualization based on embodied experience from cognitive linguistics, as in [14].

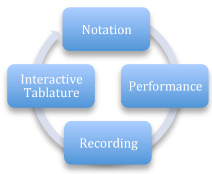
This concept's advantages over the UTI paradigm are the following: a) It is based on individual performative embodied experience, thus it might be a better metaphor for performers (performer-specificity); b) it directly involves notation in "dynamic visuo-gestural formations", unlike most studies of gesture which assume a static notation; c) it is not incompatible with analytical approaches and compositional intentions, which are fed as further dimensions and priorities into the system; d) it can account for the multivalent sound-images of postwar music, but could also be employed for earlier and simpler music as well. Complex post-1945 music serves merely as a point of departure because of the explicit problematization of *understanding* and subsequently of *technique* and of *interpretation*.

2. Gesture Cutting Through Textual Complexity (GesTCom)

2.1. General Description

In the course of the *musical research residency 2013-2014* at the Ircam, we developed a prototype system, called GesTCom. It is based on the performative paradigm of embodied navigation [4], on the INScore platform [15] and on the Gesture Follower [16], [17]. This prototype takes the form of a sensor-based environment for the production and interactive control of personalized multimodal tablatures out of an original score. As in the case of embodied navigation, the tablature consists of embodied representations of the original. The novel part is, that those representations derive from recordings of an actual performance and can be interactively controlled by the player.

The interaction schema takes the following feedback loop form:



More specifically, the input performative gesture produces four types of recorded datasets (gestural signals, audio, MIDI and video), which are subsequently used for the annotation, rewriting and multimodal augmentation of the original score. Those output notations are embodied and extended: They are produced through performative actions, they represent multimodal data, they can be interactively controlled through gesture and they can dynamically generate new varied performances. They can be considered as the visualization and medial extension of the player's navigation in the score-space, creating an interactive feedback loop between learning and performance.

[3] An overview of dynamic systems theory applications to cognition, from Rodney Brooks' subsumption architecture in robotics to the work of E. Thelen, T. van Gelder, R. D. Beer, is offered in [13], pages 114-157.

2.2 Representations

Our tablatures feature three kinds of representations, as demonstrated in figures 3 to 11. They all constitute transformations of the first four bars of Brian Ferneyhough's *Lemma-Icon-Epigram* for piano solo:

1)The first type of representation is based on the original score. It consists of the original image's (Fig. 3) annotations (Fig. 4,5) and multimodal augmentations with videos and gestural signals of a performance (Fig.5). The annotations and augmentations are achieved through the INScore platform, described in more detail in 3.3.
Annotation in this instance (Fig. 4,5) takes the form of a simple graphic segmentation of the original image (shaded rectangles in Fig. 4,5), which has been decided through experiments with the *motionfollower* as described in 3.2. This graphic segmentation can be explicitly coupled with a corresponding time segmentation, in a relation generally described as *time-to-space mapping*.The mapping is expressed in the form:

$$([x_1, x_2[\ [y_1, y_2[)\ (\ [t_1/t_2, t_3/t_4[)$$

whereby pairs of intervals expressing pixels ($[x_1, x_2[\ [y_1, y_2[$) are associated to intervals of musical time expressed in rationals ($[t_1/t_2, t_3/t_4[$), with 1 corresponding to a whole note.
Augmentation (Fig. 5) consists in the synchronization of graphic objects, such as videos and signals, along the designated time-to-space mapping. It takes the form of a master/slave relationship. In Fig. 5, the video and the signal are slaves to a master cursor, moving along the mapping of Fig. 4.
In addition, the annotated Fig. 4 has in Fig. 5 been rotated by 90 degrees clockwise (similarly to the Fig. 1e and 1f, pg.2). The generated perspective of the musical score matches the pianist's perspective of the keyboard: Pitch is distributed on the horizontal axis (lower pitches on the left and higher pitches towards the right as in a keyboard), while time is unfolding vertically, in an inversion of the traditional notational taxonomy (where pitch is represented vertically and time horizontally). Consequently, the video and graphic signal scroll down the notated image, from the right to the left column of Fig.5.

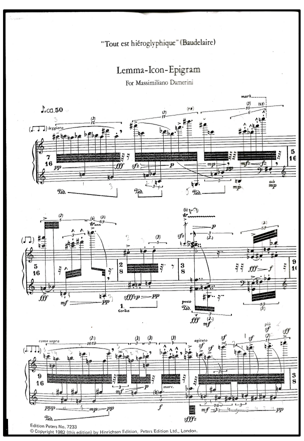


Fig. 3 Original Score

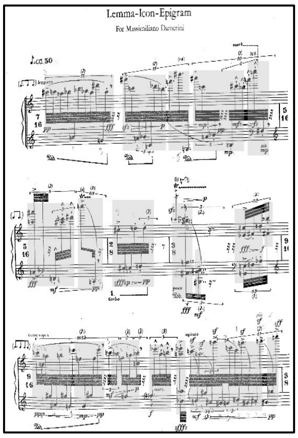


Fig. 4 Annotation: Segmentation and time-to-space mapping

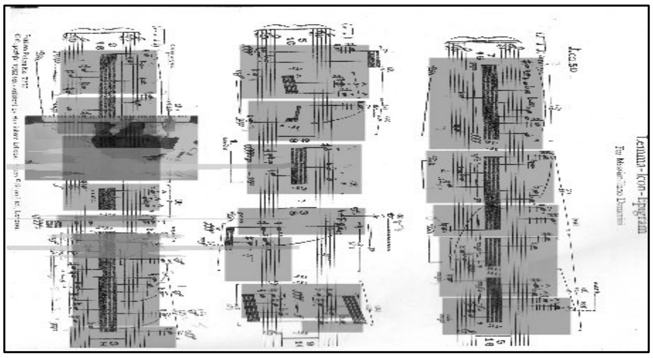


Fig.5 Augmentation: Rotation of Fig. 4 by 90 degrees clockwise and addition of multimodal data: video of a performance plus gestural signal (left column), scrolling down from the right to the left column

2) The second type of representations derives from the midi files of differently prioritized performances, reflecting different embodied layers of the original score according to the embodied navigation paradigm. Brian Ferneyhough's *Lemma-Icon-Epigram*, bars 1-4, serves always as our case-study.
In Fig. 6 a *reduced-proportional* representation, actually derived from a piano-roll of the original, has been generated from the MIDI file of a performance using tools based on the Guido engine[4].This performance reflects a note-to-note (or finger-to-finger) approach to the original notation, corresponding to the so-called „finger-layer" of the embodied navigation model as demonstrated in figures 1a and 1b, pg.2 of the current.
In Fig. 7 and 8, similar representations corresponding to different embodied layers have been used: Figure 7 is based on a transcription of the MIDI file of a performance, which prioritizes the so-called „arm layer" (Fig. 1e, 1f, pg.2). The amount of pitch information in Fig. 6 is now reduced or filtered to mostly the notes played by fingers one and five in both hands. The resulting image retains the contour of Fig. 6 and is much easier to read. The MIDI transcription has been based on MidiSheetMusic 2.6 software.
Similarly, Fig. 8 features the transcription of a performance which prioritizes the so-called „grasp-layer" (as in Fig. 1c, 1d, pg.2): The original note material is now arranged in hand-grasps and transcribed with MidiSheetMusic 2.6.

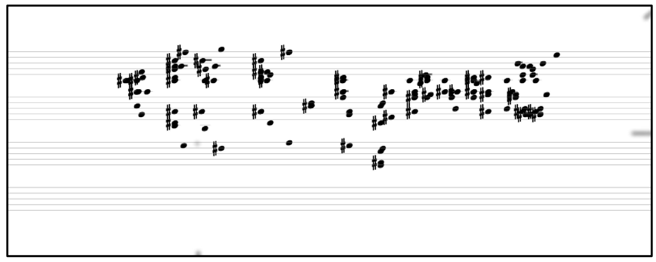


Fig.6 Reduced-proportional representation of the pitch information of the original: Finger-layer

[4] An open source rendering engine dedicated to symbolic music notation, see at http://guidolib.sf.net

Fig. 7 Representation of a performance of the arm-layer:mostly fingers one and five in both hands

Fig. 8 Representation of pitch arranged in hand-grasps: grasp-layer

3) The third type of representations involves again multimodal data: In Fig. 9 we have annotated and mapped the image of the MAX/MSP patch used for our recordings. The image includes MIDI, gestural and audio information, whose graphic representation has been segmented according to the mapping used for Fig. 4 and 5.
In Fig. 10 we have similarly segmented an image from the *motionfollower* MAX patch, depicting the superimposition of the gestural signals of two differentiated performances.

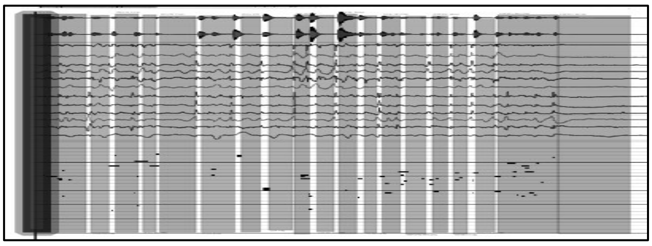


Fig.9 Recording patch image: From bottom up: MIDI information, gestural signals (6 for each hand), audio signals. The segmentation is the same as in Fig. 4, 5.

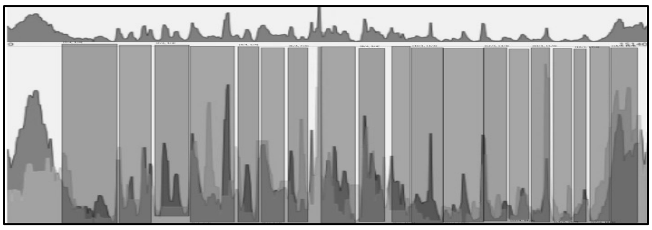


Fig.10 *Motionfollower* patch image:
Superimposed gestural signals of two performances and basic segmentation as in Fig. 4, 5.

Eventually, all above mentioned representations can be freely combined in the graphic space and synchronized through the same time-to-space mapping with INScore (Fig. 11). The resulting tablature is personalized, in that it reflects personal priorities of individual performers; multimodal, in that it enables imaginative combinations of traditional notation, symbolic scores, videos, MIDI, audio and gestural data; malleable, in that it can be substituted from the data of a new recording; interactive, since it can be gesturally controlled, a feature which we will explore more in section 3.4.
In terms of embodied and extended cognition, the player thinks by gesturally navigating several embodied representations. Learning and performing are organized in a perpetual feedback loop and this process is externalized and objectified.

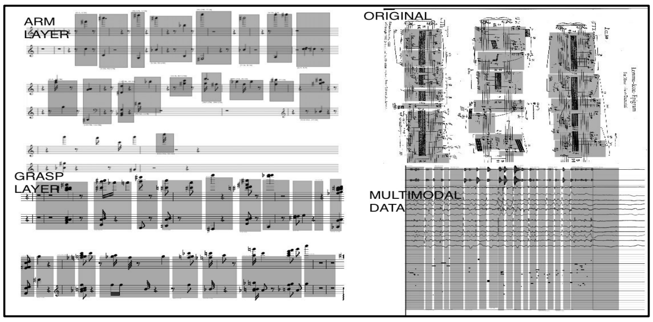


Fig.11: Tablature of combined representions. They can be synchronized with video and audio and interactively controlled. The player navigates between the several representations.

3. GESTCOM METHODOLOGY AND ARCHITECTURE

3.1. Recordings

The personalized interactive multimodal tablature is based on a set of initial recorded data, which is later appropriately mapped on the original score and on its derivative representations. The recording set-up (Fig. 12) was kept fairly simple and lightweight, having in mind performers' needs for mobility. It consisted of a midi upright piano, two microphones for audio recording, a kinekt device for video recordings and a pair of sensors capturing acceleration and angular velocity (gyroscope) data from the performer's wrists. The captured sets of data were synchronized through a recording MAX patch.

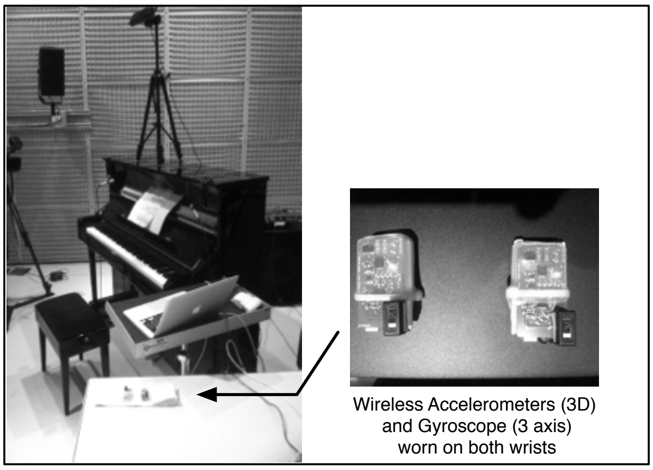


Fig. 12 Recording Set-up

In the course of three months, the first author realized a series of recordings, ranging from explicitly complex piano repertoire after 1950 (works by Iannis Xenakis, Brian Ferneyhough, Jean Barraqué) to mainstream classical repertoire (Johann Sebastian Bach and Ludwig van Beethoven). Those recordings featured several stages of the learning process, ranging from the very first approach of a new score up to the complete performance of selected passages or even complete works. A multitude of prioritization processes as to the approach of the notation, based on the model of corporeal navigation, was employed. The variety of performances of a single notational image is captured as the variation and comparative analysis of the corresponding sets of multimodal data. The question that arised was: how can those data be fed back into the score and transform it.

3.2. motionfollower

At a second stage, a scenario of pianist's interaction with the *motionfollower*, an object in MAX after the Gesture Follower architecture, was implemented.

The Gesture Follower was developed by the ISMM Team at Ircam. Through the refinement of several prototypes in different contexts (music pedagogy, music and dance performances), a general approach for gesture analysis and gesture-to-sound mapping[5] was developed. The "gesture parameters" are assumed to be multi-dimensional and multimodal temporal profiles obtained from movement or sound capture systems. The analysis is based on machine learning techniques, comparing the incoming dataflow with stored templates. The creation of the templates occurs in a so-called learning phase, while the comparison of a varied gesture with the original template is characterized as *following*.

The Gesture Follower was implemented in the so-called "prima vista scenario": This scenario of interaction is based on the assumption that, in the presence of an overwhelming amount of notational information, the performer will rather adopt a top-down approach. S/he will first focus in the global aspects of the musical work before delving into detailed analysis. In that sense, the performer starts the learning trajectory with a quasi sight-reading approach, which prioritizes fluency and forward movement and not necessarily accuracy, and gradually refines detail following personal prioritization paths.

In GesTCom, the prima vista performance is used to train the system (*learning phase*), while the subsequent, varied, prioritized performances are compared to the original (*following phase*). It was empirically found, that given a sufficient degree of fluency of the initial "prima vista" performance, there is a *basic gestural profile* or *segmentation*, which can account for all subsequent interpretational differentiations and refinements, in the sense that the system can successfully follow them. An example of *basic segmentation* has already been cited in Fig. 10. In addition to empirically allowing for the discovery of this segmentation, the use of the *motionfollower* was found to provide useful auditory feedback in the very first stages of the learning process. The motion follower was also employed at the last stage of interaction, as will be described later: In the *following* phase, the system can indicate in real-time the current position in the score, based on the performer's gestural data.

3.3 INScore

At a third stage, the basic gestural segmentation discovered with the use of the *motionfollower* was mapped on the notational and multimodal representations derived from the recording of the performance. Those graphic components were synchronized along this mapping using INScore.

INScore is an open source platform for the design of interactive, augmented, live music scores.

INScore extends the traditional music score to arbitrary heterogeneous graphic objects: symbolic music scores but also images, texts, signals and videos. A simple formalism is used to describe relations between the graphic and time space and to represent the time relations of any score components in the graphic space on a *master/slave* basis. It includes a performance representation system based on signals (audio or gestural signals).

It provides interaction features provided at score component level by the way of *watchable* events. These events are typical UI events (like mouse clicks, mouse move, mouse enter, etc.) extended in the time domain. These interaction features open the door to original uses and designs, transforming a score as a user interface or allowing a score self-modification based on temporal events.

INScore is a message driven system that is based on the Open Sound Control [OSC] protocol. This message-oriented design is turned to remote control and to real-time interaction using any OSC capable application or device (typically Max/MSP, Pure Data, but also programming languages like Python, CSound, Super Collider, etc.)

A textual version of the OSC messages that describe a score constitutes the INScore storage format. This textual version has been extended as a scripting language with the inclusion of variables, extended OSC addresses to control external applications, and support for embedded JavaScript sections.

All these features make INScore particularly suitable to design music scores that need to go beyond traditional music notation and to be dynamically computed.

As already demonstrated in 2.2, the GesTCom methodology takes advantage of the mapping and synchronization aspects of the INScore: Annotations, transcriptions and multimodal representations can be graphically combined and can be synchronized in the time domain. Furthermore, its OSC design allows real-time interaction of INScore and the *motionfollower*, as described in the following section.

3.4 Interaction

At a final stage, we were able to connect the *motionfollower* to the INScore tablature (OSC architecture) and gesturally interact with the tablature in real time. The whole idea is based on the *motion follower*'s learning and following schema, which is used to control the mobile elements of the INScore tablature, for example cursors and videos: At the *learning phase*, the user synchronizes with any element of the tablature, moving along the mapping that we described as basic segmentation (3.1). In the *following phase*, the player can pursue highly differentiated performances and prioritizations and still control the speed of the mobile elements of the tablature through her actual gestural signal. Current position in the score is indicated in real-time. The whole interaction schema could be described and at a later stage even sonified as an *embodied clicktrack*, which relieves the notational complexity and functions for a wide range of interpretational deviations.

A demonstration of the system can be accessed here: https://www.youtube.com/watch?v=KV9nQUhhyuI

3.5 GesTCom Architecture

In summary, the resulting architecture of the GesTCom involves the following components:

- Recording
- Gesture Analysis (*motionfollower*)
- Derivative Representations, Mappings and Synchronizations (INScore)

[5] The term "mapping" here obviously differs from the previously mentioned time-to-space mapping through INScore.

- Personalized Tablature Creation (INScore)
- Interaction (INScore and *motionfollower*)

4. FEATURES AND APPLICATIONS

The system GesTCom offers a novel paradigm for the management of massive amounts of information in the very first stages of the learning process, through a personal, spontaneous performative response. This initial performance segments the score in manageable chunks of information, to be used for the refinement of the performance during the learning process. Each new performance can potentially interactively transform the tablature, thus offering an accurate archive of the learning process and a means of multimodal representation/recording of the performance.

The potential applications of the system are not limited in this specific "prima vista" interaction scenario: In the case of players who favor an analytic approach or do not have the experience or ability to sight-read, we can imagine an explicit mapping of the preferred gestural properties or priorities on the INScore and its use as ground for further learning.

In comparison to other highly developed systems providing augmented feedback to the player, such as the 3D Augmented Mirror-AMIR [18], the novelty of this system lies in the fact that it directly involves notation and its transformations, thus the title „gesture cutting textual complexity".

Next to its obvious applications in pedagogy and musical performance, the system could be thought of as a compositional and improvisational tool (generating notation through gesture), as well as a powerful resource for performance analysis.

Summarizing, the features of the system involve: efficient top-down learning of complex scores through augmented multimodal feedback produced and processed gesturally; easy-to-read reduced representations of the notational information; interaction in the form of an embodied clicktrack ; archiving of learning and performance from the very first step; externalization of the navigation between the annotations, augmentations and transcriptions of the notation; performance analysis.

5. FUTURE DIRECTIONS

Future directions in the design of GesTCom include: accumulating user experience; automating elements of the GesTCom architecture according to performative needs; creating web-resources.

1) User Experience: Testing the tool in selected communitites of performers and in a wide range of repertoires will give us an accurate perspective of performative needs.

2) Architecture: Assuming the reluctance of most performers to program, it will be quintessential to keep the performer as close to the keyboard and to gesture as possible, with developments in:

a) Recording: We wish to implement haptic interactions, through the recording of other forms of gestural data such as piezoelectric, probably in combination with appropriate keyboards as controllers (for example the TouchKeys system).

b) Gesture Analysis: Instead of empirically defining the „basic segmentation" with the *motionfollower*, one could automatically derive it from notational representations employing machine learning.

c) Representations and Mappings, Tablature Creation: Automated time-to-space mapping through gesture, rather than through typical UI events, would considerably make the whole precess of tablature creation more performer-friendly. In this direction one can also predict the incorporation of more user interfaces, such as touchscreens, or controllers, such as the TouchKeys[6].

d) Interaction: The „embodied clicktrack" notion can also be extended, with sonification of the movement along the mapping.

3) Implementation of the GesTCom as an open web resource could enable projects of collaborative learning through the collective creation and sharing of interactive multimodal tablatures.

6. ACKNOWLEDGMENTS

This work was supported through the Musical Research Residency 2013-2014 at Ircam.

7. REFERENCES

[1] B. Ferneyhough: *Aspects of Notational and Compositional Practice*. In J. Boros & R. Toop (eds.) *Brian Ferneyhough-Collected Writings*, pages 2-13. Routledge, London and New York, 1995.
[2] S. Kanach (ed.): *Performing Xenakis*, *The Iannis Xenakis Series vol. 2.* Hillsdale, Pendragon Press, New York, 2010.
[3] P. Antoniadis: *Physicality as a performer-specific perspectival point to I. Xenakis's piano work. Case-study Mists.* In: *Proceedings of the I. Xenakis International Symposium 2011,* Goldsmiths University, London. http://www.gold.ac.uk/media/07.3Pavlos Antoniadis.pdf
[4] P. Antoniadis: *Corporeal Navigation: Embodied and Extended Cognition as a Model for Discourses and Tools for Complex Piano Music After 1945*. In Pedro Alvarez (ed.) *CeReNeM Journal,* Issue 4, pages 6-29, March 2014.
http://issuu.com/cerenem/docs/cerenem_journal_issue_4
[5] E. Fischer-Lichte: *Ästhetik des Performativen*. Suhrkamp Verlag, Frankfurt am Main, 2004.
[6] M. Leman, *Embodied Music Cognition and Mediation Technology*. MIT Press, Cambridge, 2008.
[7]R. I. Godøy & M. Leman (eds.) *Musical Gestures: Sound, Movement and Meaning*. Routledge, New York, 2009.
[8]E. F. Clarke: *Ways of Listening: An Ecological Aproach to the Perception of Musical Meaning*. Oxford University Press, Oxford, 2005.
[9] J. Solis & K. Ng (eds.): *Musical Robots and Interactive Multimodal Systems*. Springer-Verlag, Berlin Heidelberg, 2011.
[10] J. J. Gibson: *An Ecological Approach to Visual Perception.* Psychology Press, London, 1986.
[11] M. Rowlands: *The New Science of the Mind: from extended mind to embodied phenomenology.* MIT Press, Massachusetts, 2010.
[12] A. Clark: *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press, New York, 2011.
[13] L. Shapiro: *Embodied Cognition.* Routledge, London and New York, 2011.
[14] G. Lakoff and M. Johnson: *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. Basic Books, New York, 1999.
[15] D. Fober, Y. Orlarey, S. Letz: *INScore: An Environment for the Design of Live Music Scores*, *Proceedings of the Linux Audio Conference* - LAC 2012.
http://www.grame.fr/ressources/publications/INScore-ID12-2.pdf
[16] F. Bevilacqua, N. Schnell, N. Rasamimanana, B. Zamborlin, F. Guedy: *Online Gesture Analysis and Control of Audio Processing*. In: J. Solis & K. Ng (eds.) *Musical Robots and Interactive Multimodal Systems,* pages 127-142. Springer-Verlag, Berlin Heidelberg, 2011.
[17] F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, F. Guédy, N. Rasamimanana: *Continuous realtime gesture following and recognition*. *In Embodied Communication and Human-Computer Interaction*, volume 5934 *of Lecture Notes in Computer Science*, pages 73–84. Springer Berlin and Heidelberg, 2010.
[18] K. Ng: *Interactive Multimedia for Technology-Enhanced Learning with Multimodal Feedback.*In: J. Solis & K. Ng (eds.): *Musical Robots and Interactive Multimodal Systems,* pages 105-126. Springer-Verlag, Berlin and Heidelberg, 2011.

[6] http://www.eecs.qmul.ac.uk/~andrewm/touchkeys.html