

Hyperinstruments as interactive systems of music composition

NICOLA BARONI

Istituto di alta formazione musicale Conservatorio Claudio Monteverdi Bolzano

Abstract. Form shaping has been a principal focus of music composition since the mid XX Century, when the classical musical structures and listening practices started to be questioned by the avant-gardes. An advanced and pioneering system of composition was represented by Xenakis's use of computers for elaborating scores from stochastic processes inspired by physical laws and complex mathematical behaviours; through non-linear mass distributions analogous forms were produced in macro-dimensions (orchestra) and micro-sounds (electronics). Starting from the 80s, a further development of this approach was based on the mapping of interactive actual human gestures onto pitched and synthetic sound contours [1]. More recently Godoy and Jensenius have been exploring cognitive and computational correspondences between human gestures and music, rooting their concept of music on the traditional electroacoustic idea of the sound-object as a primary building block [2]. A sound object is a gestural form-bearing perceptual unit, a fragment of concrete sound typically in the range of a few seconds, which can be seen as a structural counterpart of the more traditional element called "musical note". The notion of gesture as a sensitive metaphor for the interpretation and analysis of music forms has become a consolidated topic over the last decades, blurring boundaries between score-based and electroacoustic composition. The current development of sensing systems, such as sound analysis in real-time and motion tracking, are supplying factual means for researching into the field of performance-based interactive music. Since their origin, the interactive behaviours of Hyper-instruments has been implemented as a means of empowering performers to intentionally influence the electroacoustic outputs of score-based compositions through their performance gestures on-stage [3]. Starting from the notion of interactivity we consider the potentials of the current sensing systems to be part of complex, digitally formalized, compositional networks and processes, in the light of the current emergence of embodied cognition frameworks. This paper explores topics bridging the meanings of music composition and music gesture, presenting as a conclusion some hypotheses which support innovative systems of performance based real-time digital composition implemented by the writer.

Keywords. Scores as instruments, gesture-based composition, physical computing

1. Composition and scores

The concept of a contribution by the performer to the compositional music process is an ancient topic, since traditionally music scores allow degrees of freedom to individual and even on-the-fly performance choices. Pre-classical scores mostly delineate frameworks organised as pitch/duration note-based "discourses", expanding in metric/harmonic sequences within defined macro-forms. The score, as a designed representation, need to be completed by live ornamental and polyphonic contributions from the performer, who shares and knows the relative compositional technique.

[Allemande bass]



Figure 1. Example of XVIII Century Figured Bass notation. Numbers as a guide for extemporary harmonisation and polyphony invention. F.E. Niedt, *The music guide*, Clarendon press, Oxford, 1989

It should be noted that extra-European traditional practices mostly neglect to consider composition and performance as distinct roles, and in the case of written music documents we most often find collections of tunes, patterns, lexicons, symbolic associations, congruous behaviours to be mastered and “composed” by the “performer”, who elaborates original expressions from sets of principles. The *Werktraue* idea of a score as an ideally whole connotative entity started to emerge within the Western Classical and Romantic era, shifting the role of the performer to the more constrained responsibility of a subjective accomplishment of the Work, taking into account a subset of implicit meanings, whose sonic realisation denotes an art of interpretation. In this way music scores can be seen as a full symbolic representation of the sounds required by the composer, in other words as the Text of the composition.

1.1 Recording technologies

The development of recording technologies during the past century appears to have caused a dual process: on the one hand dramatically increasing the need for a perfect and objective accordance of live performances with the written Classical score, but on the other hand offering the status of a corollary textuality to multiple recorded performances, often quite different from one another. Through recording we can objectively analyse different performance renderings of a same score, we can also extract and examine features of non-written compositions, and even textually evaluate free improvisations, since they are recorded on a support [4].

This situation led to the rethinking of some terms of the debate about what composition is. In addition, the recording technologies allowed to analyse and formalise the most subtle sound morphologies allowing timbre to be considered as a principal object of knowledge and compositional treatment. Timbre parameters started to be structurally relevant and no longer confined to the standardised and ancillary roles traditionally given by the Western tradition. In this context John Cage’s claim about the impossibility of an exhaustive textuality of the music score, and the following deduction that every score has clear degrees of indeterminacy, appears significant: the choice of which parameters should be more precisely defined inside a score is therefore a social habit, or an individual decision [5].

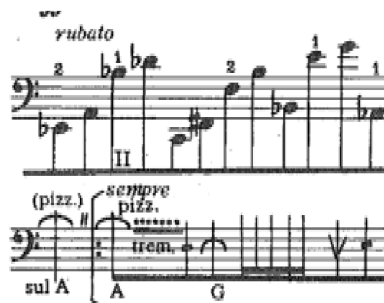


Figure 2. Mix of common and action notations K. Penderecki , *Capriccio per Siegfried Palm* Schott, Mainz, 1968

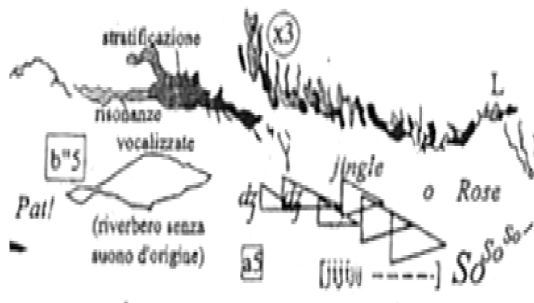


Figure 3. Graphic and electroacoustic score L. Berio, *Thema (Omaggio a Joyce)*, Suvini Zerboni, Milano, 1958

1.2 Unconventional scores

The mid XX Century witnesses wide-ranging experimentations of new scores, abandoning traditional note-oriented approaches and developing non-connotative features such as action notations (defining which instrumental gesture is to be performed irrespective of the resulting sound), free-graphic approaches, timbre and process-oriented notations, verbal instructions, combinatorial systems and circuits.

1.3 Interactivity

The persistence of traditional notation strategies now offers a multilayered landscape which, in the case of software composition, currently allows to produce programs intrinsically intended both as representational and operative, in other words acting as scores and instruments at the same time [6]. This assertion can be considered as a kernel of interactivity in composition. The radical thrust of conceiving composition as a combined action of textual machines treated as instruments was pioneered at the advent of electroacoustic music through physical manipulations of recording tapes, variable voltage controls of mathematical rules actually synthesising sounds, and algorithmic systems of note composition through rule-based or data-driven combinatorial processes.

Interactivity is underlined by Horacio Vaggione's action approach to composition, escaping linear formalisations towards multi-syntactical strategies borrowed from object-oriented programming methods. In this perspective algorithms are not seen as abstractions allocating mechanisms towards a result to be straightly listened to, but rather as processing tools producing their own rules and incapsulating the listening action of the composer as part of the operation [7]. In fact the potentials to exploit computation for analysis, symbolic representations (such as scores and rules) and sound synthesis even in one single environment currently allow networking, contextual and semantic behaviours previously unpredictable in terms of complexity. In this direction we could quote, among others, productions and researches oriented to multi-agent ecosystem methods of real-time composition inscribing human choices and environmental conditions as part of AI procedures [8].

2.Sound and gesture

Electroacoustic music is characterised by the direct manipulation of sound on supports (recording tools, editors or softwares). In this way the so called sound-based composition potentially allows to bypass the presence of a traditional performer and a symbolic representation (score), embedding sound synthesis, transformations, organisation, storage and diffusion inside a group of machines: sound can thus be directly shaped without any intermediate layer, by means of a chosen studio-machine acting as an instrument-support tool. In this way becomes natural to create music derived by real-life sounds, thereby extending the concept of music timbre.

Traditional music theories were grounded on the concept of music notes, discrete chunks of "ideally pure" sounds, sharing a scalar space of frequencies (pitch) and durations, functionally organised through standardised or innovative macro-forms often relating to dance, poetry, mathematics, architecture, or rhetoric figures. In the last

century, the further extension of the notion of music sound to all the possible audible phenomena, of which traditional instrumental sounds are a special family, produced new contrasting theories mostly developing Schaeffer's concept of *Musique Concrète*.

2.1 Sound objects and morphologies

Schaeffer's phenomenological approach to music explored the perceptual qualities of real-world sounds, creating an idea of composition based on sound fragments that exist in reality, considered as discrete and complete "sound objects", aiming to remove music from the idea of structured "sound abstractions" [9]. The "sound object" is a fragment of recorded tape, or a continuous sound repetition through a closed groove (the so-called *sillon fermée*). Through repetition or de-contextualising manipulation the "sound object" is abstracted from its reality becoming an object of music contemplation. In the age of analog technologies in the mid XX century, this extraction of sound objects was a physical action/gesture of composition through slice/paste strategies acting upon the actual recording support. The length of the sound-object, broadly modelled on the archetype of the "note", shares with the note the potential to be treated in a phonetic fashion. Forcing the linguistic comparison we might argue that the note is open to be seen as an arbitrary sound potentially part of a pseudo-logical music organisation, and in this sense many older theories and pedagogic approaches stress language-based metaphors describing music forms as phrases, periods and macro-structural abstractions, generally intended as devoid of arbitrary meanings. Differently, a sound-object retains its concrete overall shape: it is a small perceptual pattern, a unit of an audible gesture, in a sense a "timbre" block. The last work of Schaeffer represents a systematic effort to organise a lexicon of typo-morphologies of sound objects, a *Solfège* based on the perceptual surface characters of these catalogued sound units: in other words on their action/perceptual content. Principal categories of the inventory relate to iteration, continuity, grain, impact, saturation, allure, profile and internal dynamic [10].















TIMBRE DYNAMIQUE		1	2	3	4	5	6	7
GENRES D'ATTQUES	Tracé bathygraphique							
	Nature d'attaque Symbole conventionnel	ABRUPTE ou explos.  (choc ou plétre) sans résonance appréciable.	RAIDE  (marteau feutré) avec forte résonance liée	MOLLE  (pizz ou mailloche douce) avec résonateur	PLATE  (pseudo attaque) ou mordant	DOUCE  son posé sans ataq. apparente	SFORZANDO ou appui  ou crescendo rapide	NULLE ou très progressif.  perception du profil
PRÉDÉTERMINATION DU PROFIL en fonction du genre d'attaque	Profil dynamique	pointe dynamique (choc)	régulier dégressif	renforcem. du résonateur	nul, sauf la pseudo-attaque	profil nul	Profil caractérist. sons en général courts	seul cas de seuil émergence du profil
	Profil harmonique	son double (2 timbres)	appauvrissement	réponse du résonateur	nul dans les instruments tels que l'orgue varié en musique électr. ou cordes	Profils souvent progressifs	Timbrage caractérist.	Profils le plus souvent liés ou artificiellement indépendants

Figure 4. Typo-morphologies of sound objects [10]

Among the multiple productions and theories developed after Schaeffer, *Spectromorphology* is currently considered as a main electroacoustic compositional frame and subject of reflection. The accent is placed on the time and spatial features of sound in relation to the macro-evolution and dynamic consistencies of the composed sound, not confining the analysis to object typologies, but showing an event-based constitution of the virtual-sound world of electroacoustic music. Framed by the main categories of gesture and texture, sound movements are catalogued in terms of the rooted/floating qualities, trajectories, propagations, multi-dimensional and behavioural aspects of sound organisation [11].

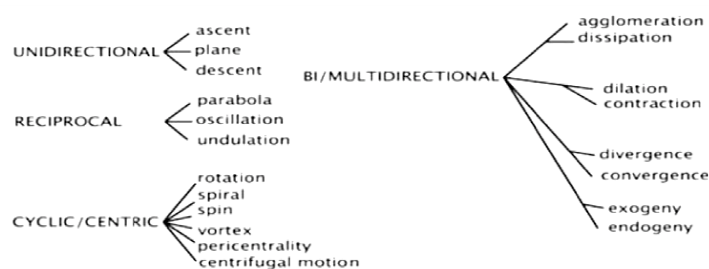


Figure 5. Gestural dynamics of spectromorphology events [11]

2.2. Time Scales of Music

On the other hand, starting from Stockhausen's pioneering research, and taking into account the developments of sound science and digital sound processing research, music sound categories can be unified in terms of time-perception inside the so-called theory of the Time Scales of Music [12]. In this sense Macro, Meso and Sound Object time scales appear to be falling within a time range consciously detectable and analysable by humans and traditionally scored and represented. Sound objects share a similar time scale (a few seconds) with respect to the traditional music notes (approximately from 200 milliseconds until 3-4 seconds), while macro and meso levels can be easily reabsorbed into the terms of traditional macro and intermediate music forms. Micro time scales can instead describe and compute events and manipulations difficult to be logically managed prior to the advent of digital means.

The fastest events perceivable and producible by humans cannot be below a threshold of 100 milliseconds ca., and the human spontaneous tendency is to group them in patterns when they are very quick. Below this threshold we find a blurring zone of roughness and reverberation extremely important to detect the character of sound attacks and dynamics linked to a global unconscious identification of timbre and emotional qualities of the sound. The time scale roughly between 1 and 20 milliseconds pertains to the perception of pitch (from 50 to 1000 Hz.). A faster timescale, from less than 1 millisecond until a few milliseconds, relates to filtering, digital effects, and interestingly to the real perception of timbre qualities through unconscious auditory fusion [13].

2.3. Digital Composition

The software potential to declare, compute and process heterogeneous functions proceeding through diversified time scales obviously represents a huge advantage. Musical programming, formal and/or graphic representations help to empower complex kinds of analysis and to frame consistent music structures, which need to be “performed” by the system (automatically or by human actions) in order to generate a composition. For this specific purpose, it seems unimportant whether the “compositional performance” happens in real-time (on stage) as opposed to off-line and step-by-step (in studio), or if the result is intended for producing a notated score rather than to directly shape sounds.

The relevant fact is that every kind of Computer Aided Composition involves softwares to enact processes implied by a final composition, generally too complex to be fully controlled by a human mind, and requiring a human response (or evaluation/choice) in front of non-deterministic outputs resulting from the initial conditions set by the composer: obviously algorithms are a huge collection of tools, not the composition. The focus on processes and interactive design whose output cannot be fully foreseen show a non-classical attitude to viewing the essence of the composition as the living dialectic between diverse entities and agents [14]: composers can be interested in showing the autonomous results of the composed pre-conditions, or be part of the system in order to live-constrain the system, maybe adding further layers. If the result has instead to be a fixed score, in any case composers can operate a choice for the most successful final work from among different outputs generated by non-deterministic systems, or exploit computers only for local problem solving.

2.4. Notions of Gesture

If traditional scores depend on the performance gesture (at least imagined in the case of an expert) in order to be realised, and are probably the final fixed result of previous instrumental/conceptual gestures, new technologies appear to have more intimately embedded gestural approaches to composition, as previously mentioned while discussing on sound-objects and spectromorphology. If gesture appears as a native rationale in the field of sound-based composition, since a “concrete sound” is intrinsically a gesture, we notice a growing trend to deploy the category of gesture also in score-based, even traditional, music. Bierwisch defined music as a gestural form because of its iconic and combinatorial status, dynamically oriented to shape surfaces, contours and irregularities, navigating through structures, in opposition to language which is essentially a logic form [15].

Gestures denote non-verbal transfers of information through body movements not necessarily conveying conventional meaning, and often emphasising emotion and expression. An interesting isomorphism linking gesture and music regards the joining of physical motions with human intentions, by a rhetoric attitude calling for a feedback [16]. Sound and gesture share a physical, dynamic, spatial and semiotic attitude, and in case of sound producing (instrumental) gestures they manifest a joint intention, semiosis and embodiment. In this sense the action upon a controller cannot be defined as a gesture. But the trajectories of notes on a score, just as the direct sounds on a support, are indeed considered as gestures relying on their physical, semiotic or perceptual consistencies.

3. Interactive Music

Interactive music needs a sensing input coming from the real world and its factual status relates to digital processes. Sensing is a kind of physical computing, which exploits audio input (microphones or pickups) and/or motion tracking mainly in the form of optical and inertial systems, and can also be integrated by force detectors and potentially any other means of body and environment monitor systems. What happens in the world flows as a vector of data acting as a collection of variables in real-time, depending on the quality of the analysis, the kind of features and trajectories chosen to be extracted, and the types of interaction wanted by the composer. In other words the complexity of this hermeneutic step relies on a transparent transformation of low-level physical quantities into mid/high levels of meaningful features.

In the case of audio input treated as a data collector, interactive artists exploit objects of analysis relating to acoustic knowledge and music theories. Motion tracking often involves algebra, geometry and kinaesthetic descriptors, taking into account the current consolidating tendency towards a search for corporeal high-level features often relevant to embodied cognition theories. In this sense the body is seen as a mediator between matter and mind and the search moves to defining the relations between corporeal articulations (countable patterns of movement) and subjective intentions like non-verbal messages, socially shared techniques of movement, functional cues and behavioural resonances [17]. A subset of analysis linking the Schaefferian sound typomorphologies to the functional segmentation of music-related-actions such as sound-producing, excitatory, modulatory or sound-accompanying actions can be found in the field of the so-called Music Retrieval Ontologies [18]. Machine learning systems are sometimes applied for the detection of complex gestures such as bow-movements [19].

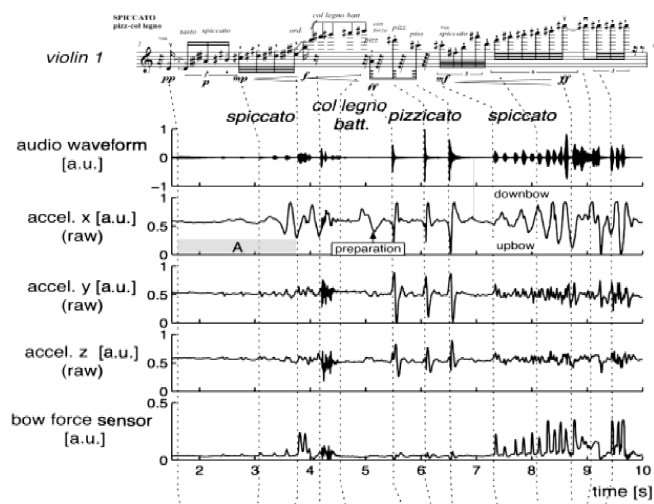


Figure 6. Plot of gesture segmentations [19]

Music Information Retrieval is mostly concerned with the implementation of objects able to extract information from the raw audio signal processing its spectrum, the iterative patterns of amplitude or brightness contours, in order to return significant perceptual features through complex reverse engineering, giving rise to high-level music descriptors.

3.1 Composition and instrument

Interactivity allows a live dialogue between performance on stage and electronics, allowing a consistency partially lost when the Live Electronics are controlled by off-stage machines, and even by on-stage controllers.

If the performer turns a switch on the electric guitar we will hear the sound effects changing, but if the sound effects are variably dependent on the kinds of patterns, timbres or intensities currently played by the guitar we notice an increase in complexity and expectancy. It is self-evident that interactive systems hybridise the concepts of instrument, performance and composition. Since performance influences the electronic sound, playing an instrument involves playing also the electronics and the final relation between both: in other words to “live-compose” a multilayered structure. In this case software composition must be procedural, modular and reactive (in a sense “performative”). Software interactive design shows overlapping aspects among the categories of instrument and composition.

Therefore interactive music is often inscribed in a pre-composed score, in this way the spread of the interconnections becomes local and is absorbed by the planning responsibility of the composer; the composition can also leave small or large windows of free exploration to the performer offering more elastic results. Many systems are instead based on improvisation, opening a broad HMI dialogue whose responsibility is shared by the performer and the composer-programmer. Radical experimental approaches involve one single performer prototyping his/her interactive languages and exploring new music boundaries [20]. It is well known that interactive systems can also allow the audience to gain channels of influence upon a live performance. A taxonomy of interactivity can be built on the continuum among the range of complexity of the systems. When just a few linearly shaped parameters drive the variable machine response the system is defined as instrument-like, while greater complexity relates to a more compositional response [21]. Originally complexity was linked to an idea of unpredictability sometimes useful for increasing human creativity by enhancing the sense that a machine interacts instead of simply reacting; the improvement in sensing tools and high-level descriptors obviously contributes to the perceptiveness of any systems. We further note the possibility to discriminate between note-based and sound-based approaches, the latter approach being more involved in timbre and spatial electronic treatments. Note-based interactive systems, originally built upon the MIDI protocol, were able to manage in real-time traditional note-oriented “languages”, allowing to implement HMI systems dialoguing in terms of music symbols and structures. Currently softwares easily mix and swap both approaches.

Hyperinstruments (also called Digitally Augmented Instruments) are a special family of interactive systems implementing an acoustic-digital unity focused on the typical performance actions of the traditional instruments. Through features extracted by sound analysis and/or motion tracking upon the sound-producing gestures, and a net of digital mappings, they follow a “chamber music” ideal continuity from performance to digital composition [3].

4. Gestural systems of real-time composition

Hyperinstruments, since not physically modifying traditional instruments, rely on acknowledged techniques and expressive rhetoric patterns. The idea of navigating within virtual worlds is currently quite common, often at the cost of losing continuity with the real world. Augmented Reality, as a true world filled of data, needs gestures (non-verbal transfers of meaning), rather than controllers. The goal of my systems is an intimate sound re-appropriation of symbolic score-machine flows.

My reference software is the interactive music environment MAX/Msp [22], through which I recollect networks of sensing data coming from a minimal equipment of audio pickups and/or inertial motion units [23], whose resulting features are analysed through specialised libraries. Compositions are themes (narratives) upon which the performer is requested to operate a search, to make choices, to explore the sounds coming from the electronics, elaborating individual strategies. External verbal scores tell the performers how to influence the overall sound result and to how to guide the system, which variably develops in part automatically and in part as a consequence of the performance. The laptop screen acts as a variable animated score proposing and responding (sometimes generating interactively common notation as a result of the performance gestures that have to be sight performed in a loop). In the case of an ensemble the performers send reciprocal messages, interactive scores and elaborate on-the-fly pre-determined collective goals. The performers can gain a detailed knowledge of the interaction through rehearsal, but they can also interact loosely, intuitively and discovering step-by-step. The verbal scores inform the performers which are the means to interact with, and they can monitor the composition behaviour by listening and through the visual screen. Depending on each single composition the performers can communicate and interact through note-intervals (onset/pitch detection), instrumental timbres, rhythmic patterns, contrasting music sequences (in this case recognised by the system through machine learning), or pitch ranges. In the case of motion tracking the best results have been obtained by bowing styles recognition and sound accompanying gestures. Performers learn how to expand their gestures in order to integrate their acoustic result with the system's behaviour and sound as a single consistency. Each interaction is a special software instance focusing on specific techniques, sound/event search, performance problem solving according to the benchmark "fiction" trajectory [24].

The systems are intended as gesture-based compositions since the non-linear nodes of the local mappings are constrained by input gestures which are physical signals, intentions, and performance techniques mediated by software symbolic actuators. The input gestures (timbres, note contours, sound patterns) are intimately complex and the performer has to understand how the machine selects their features and modulates the "socially" goal-oriented tasks. The semiosis between human and system (and through humans in the case of an ensemble) operates through scores and representations. Scores are generated as gestural resonances, local messages and autonomy/heteronomy negotiations displaying the specific narrative. In this sense performer and pre-programmed system are treated as agents of a single environment for shared strategies of composition. Improvisation is allowed as an emergent strategy of contextual adaptivity, but performers need to predetermine fixed individual strategies not in order to gain control (since the system is self-regulating) but in order to gain a maximum of meaning.

References

- [1] I.Xenakis, *Formalized Music*, Pendragon Press, New York, 1992
- [2] R.I.Godøy & M.Leman, *Musical Gestures: Sound, Movements and Meaning*, Routledge, New York, 2009
- [3] T.Machover, Hyperinstruments. A progress report 1987-1991, *MIT Media Laboratory* (1992), <http://opera.media.mit.edu/publications/> (last accessed 7/17)
- [4] V.Caporaletti, *I processi improvvisativi nella musica*, LMI, Lucca, 2005
- [5] J.Cage, *Silence*, Wesleyan University Press Paperback, Middletown, 1961
- [6] N.Schnell & M.Battier, Introducing Composed Instruments, Technical and Musicological Implications, *Proceedings of the 2002 Conference on New Instruments for Musical Expression* (2002)
- [7] H.Vaggione, Some ontological remarks about music composition processes, *Computer Music Journal* **25:1** (2001), 54-61
- [8] A.Eigenfeldt, Real-time Composition as Performance Ecosystem, *Organised sound* **16:2** (2011), 143-153
- [9] P.Schaeffer, *A la recherche d'une musique concrète*, Éditions du Seuil, Paris, 1952
- [10] P.Schaeffer *Traité des objets musicaux*, Éditions du Seuil, Paris, 1966
- [11] D.Smallley, Spectromorphology: explaining sound-shapes, *Organised sound*, **2:2** (1997), 107-126
- [12] C.Roads, *Microsound*, MIT Press, Cambridge, Mass., 2004
- [13] T.Whishart, *Audible design*, Orpheus The Pantomime Ltd., York, 1994
- [14] A.Di Scipio, A Constructivist Gesture of Deconstruction. Sound as a Cognitive Medium, *Contemporary Music Review*, **33:1**, 87-102
- [15] M.Bierwisch, *Musik und Sprache: überlegungen zu ihrer Struktur und Funktionsweise*, Peters, Leipzig, 1979
- [16] C.Cadoz & M.W.Wanderley, Music-gesture, in *Trends in gestural control of music*, eds. M.Battier & M.W.Wanderley, Ircam Centre Pompidou, Paris, 2000
- [17] M.Leman, *Embodied Music Cognition and Mediation Technology*, MIT Press, Cambridge, Mass., 2007
- [18] R.I.Godøy et al., Classifying Music-Related Actions, *Proceedings of 12th International Conference on Music Perception and Cognition*, (2012)
- [19] F.Bevilacqua et.al, The Augmented String Quartet: Experiments and Gesture Following, *Journal of new Music research* **41:1** (2012), 103-119
- [20] G.Lewis, Too Many Notes: Computers, Complexity and Culture in Voyager, *Leonardo Music Journal* **10** (2000), 33-39
- [21] R.Rowe, *Interactive Music Systems*, MIT Press, Cambridge, Mass., 1993.
- [22] <http://cycling74.com/>
- [23] <https://sites.google.com/site/speckledcomputing/cello2>
- [24] <https://nicolabaroni.com/artworks>