Abstract

Interactive Music Systems (IMS) have introduced a new world of music-making modalities. But can we really say that they create music, as in true autonomous creation? Here we discuss Video Interactive VST Orchestra (VIVO) — an IMS that considers extra-musical information by adopting a simple salience-based model of user-system interaction when simulating intentionality in automatic music generation. Key features of the theoretical framework, a brief overview of pilot research, and a case study providing validation of the model are presented. This research demonstrates that a meaningful user–system interplay is established — what we define as reflexive multidominance.

Keywords: interactive music systems, computational creativity, salience, reflexivity, multidominance.

Tridimensional continuum: instrument, player, and score

Rowe (1993) defines an IMS in terms of a continuum between instrument and artificial player. Here, we propose to extend this definition to include the score as an interdependent dimension of a three-dimensional space. This perspective mandates that an IMS (a) always affords an amount of control to the user via interactions with the instrument, (b) conveys system autonomy (i.e., suggests an autonomous player) to the user, and (c) constrains the invention scenario by its system architecture or score. We demonstrate the importance of this three-dimensional space by looking for commonalities among different systems for each specific dimension. Notably, by introducing the score as an explicit dimension of the continuum, we are able to investigate user interaction and system autonomy — a current limitation in IMSs. We hypothesize that, for an IMS, system autonomy exists within this space, and that it is possible to use extra-musical information (e.g., the quality of a gesture) in the generation of a sound in order for a user to identify the sound as being intentionally organized, which is a crucial step in interpreting music as meaningful.
The system architecture (A.K.A score)

The score, or system architecture, is the set of processes and instructions that dictate music generation either directly or indirectly. Starting from early systems (Vercoe, 1984; Dannenburg, 1984) to more recent ones (Arshia, 2008; Ritter et al., 2013; Echeveste et al., 2013; Rodriguez-Serrano et al., 2016), a score-follower is an example of a score-focused IMS. In any system implementing user-mirroring, algorithm autonomy and content meaningfulness depend on a balance between implemented features and the user’s interaction with such system functionalities. In terms of computational creativity, we may describe this as computer invention. Indeed, this balance also frames computer generation, which is not open-ended as we would expect, because every IMS embeds at least some implicit knowledge of the predefined compositional structure in the system architecture.

Although ‘scores’ can be responsible for variable generations not foreseen by design, they are akin to musical scores relying on indeterminacy or chance in some form. The computational score that is embedded in an IMS since its design is similar to the score that a composer may conceive in aleatoric music. As in an aleatoric composition, an IMS may embed more than one of the following techniques, separately or in combination (Griffiths, 2001): (i) interactive processes adopting randomness for generating fixed compositions, (ii) mobile or polyvalent generations whose sequencing is partially or completely determined by the user, (iii) methods of indeterminate notation or feedback providing freedom of interpretation to the interacting user. Similarly to an aleatoric composition, an IMS maintains an underlying musical identity.

When indeterminacies in score generation occur, it is in relation to the other dimensions of the space — the instrument and the player. Drawing on Rowe’s continuum, we can state that an IMS is not only an instrument and a player, but also a score in the broadest sense. VIVO (Paolizzo, 2004; 2014) embeds these different techniques in the system architecture in order to either interact with the user to determine the amount of indeterminacy in the generation, or force the system to closely follow the user-defined composition.

The instrument: interacting with the user

All IMSs are musical instruments in that they extend the creative capacity of the user. This extension can range from entirely deterministic to non-deterministic. A deterministic system can have a physical body that permits the user to generate sound in new ways (e.g., using sensing technologies on an acoustic instrument to generate or transform the sound, as in Machover & Chung, 1989). In VIVO, a camera connected to the system and used to control the sound can sense physical gestures on a musical instrument. A system can also surpass physicality by allowing the user to control music generation in a virtual environment, creatively. For example, the developer can implement an interaction mapping between a system’s complexity and its intrinsic limitations. The interaction design of VIVO can accommodate for technical limits such as sampling rate, quantization and latency (e.g., Tarabella, 2004) in order to control the relations between the gestures necessary to produce sounds on a specific instrument and the intradiegetic and extradiegetic elements of a performance (Paolizzo, 2013). In both cases, the sound production exceeds that of traditional instruments, and the user still has reliable control of the system output.
Non-deterministic systems vary in the control that they afford the user. That is, the system can allow a user’s performance to influence a musical outcome without controlling it entirely. Similarly, VIVO can be set to provide an illusion of control for the users, while actually only reinterpreting their real-world input (e.g., Schiemer, 1999) by operating at the edge of control (Drummond, 2009) or exhibiting a notion of ‘discontrol’ (Wijnans, 2010).

**The player: establishing an autonomous system**

A system may be considered autonomous when its behavior resembles the intentionality typical of a human player. “Resemblance” means here that the system could ‘fool someone into thinking it was human’ and/or ‘has/suggests a similar level of intentionality as a human’. This links into debates in the computational creativity literature as to whether computers can be creative in ways that are as creative as humans, whilst clearly being computers and exercising their creativity in a way that is native to computers. Although typically synonymous with complete independence, here we define *autonomy* as a self-determination that is not necessarily free from the influence of some external information fed to the system (Bown & Martin, 2012). Within this framework, a distinction can exist between an autonomous IMS with seemingly random behavior, and one that is controllable by the user. Along this spectrum, a user may perceive a system that is too autonomous or too predictable as being an unengaging or over-improvisatory partner, respectively.

In VIVO, the user maintains control over a specific set of parameters, but the system can also be set to exhibit dynamical and unpredictable sonic behavior. This is a common feature of many IMSs, even for simple interactions “altering the relation the system has to itself” (Sanfilippo, 2012; 2015). We expect the level of autonomy in computer generation to allow for the retention of some mutuality to the context, as it occurs alongside human interplay/co-invention.

**Listening**

A well-known approach to the design of computer programs whose behavior mimics that of a human interplayer in a musical improvisation consists of using algorithms that monitor the improvisation and use the information gathered to generate new, contextually relevant material. A listener understands the computer outcome in terms of a response. Early examples of this are GenJam (Biles, 1999) and MusicBlox (Gartland-Jones, 2003), which use interaction and interactive genetic algorithms to define the goodness of contextual fit between the computer-mutated musical fragment and the human performer’s contribution. The formalization implicit in the definition of the initial population and of the interaction rules — a score-focused approach — is mitigated in terms of openness by the subjectivity of the (unilateral) user decision-making process. The approach is also implemented in VIVO, and as for any user-defined stochastic score, its automatic unfolding depends on the musical choices made during the interplay.

**Multidominance**

Examples of mutual listening works (Chadabe, 1984; Perkis, 1999; Brown & Bischoff, 2002) use the combined behavior of software and human agents to determine overall system complexity. In terms of system autonomy, this is an improvement over unilateral human-to-
machine interactions, as input gestures are re-interpreted into complex musical outputs. Such systems denote a form of shared control. However, only response-response interactions can be determined, as the software agents do not exhibit a capacity for *multidominance*, a term borrowed from Douglas (1991). As such, these systems cannot lead the musical direction of the performance since the primary generator of novel material is the human performer. Multidominance as a system property is also a trait of authenticity in autonomy (as similarly introduced by Lewis, 2000). One of the first systems capable of such style-independent response is Voyager (Lewis, 2000). It carries out sonic behavior grouping by imitating, opposing, or ignoring the performer’s musical dynamic. The system then processes outcomes and reconfigures any algorithm involved in the sonic behavior grouping with “no built-in hierarchy of human leader/computer follower” (Lewis, 2000, pp. 36).

**Modelling knowledge with stylistic reinjection**

A computational model of music can be achieved by segmenting music sequences in a corpus and analyzing those segments for common style elements. This information is then used to recombine these segments into new works (e.g., Cope, 2010; 2016). Although knowledge-based systems using hard-coded rules are extremely score-focused, using the commonalities in a large corpus allows these models to avoid plagiarism while yet providing musical coherence and producing novel variations of existing music. Generation for these systems is however most efficient in non-real time environments and depends on a model of “creativity that is coming from the programmer and not the program” (Wiggins, 2008).

A successful approach to real-time generation is that of implementing Markov chains with constraints for modelling musical styles (Pachet at al., 2001; Pachet, 2002). The approach was later improved with the use of combinatorial design games for creating active lexicons from knowledge models of the user’s interaction (Pachet, 2008), and more recently by the use of Flow Machines for music and text (Pachet, 2016). Active lexicons are a type of grounded machine learning approach that has some similarities to the OMax system, which listens to an acoustic musician and plays along interactively by constructing real-time symbolic models in order to recombine the player’s discourse into new material. OMax constantly confronts the player with “a reinterpreted version of his own playing” (Lévy et al., 2012, p. 1). In *stylistic reinjection* (Assayag et al., 2006), rather than opposing the user, the system provides a measured amount of challenge. As previously suggested, “[s]trong interactivity depends on instigation [by the system] and surprise [by the human performer], as well as response” (Blackwell & Young, 2005).

**Adaptive models and self-evaluating systems**

A rule- or script-based approach for controlling the generation parameters of an improvisation session can be found in IMSs such as PyOracle, which builds predictive oracles to study decision problems adaptively on arbitrary feature data without quantization. Features derived from input data define graph data structures for the machine improvisation. This improves on systems that construct the oracle by converting audio data into a symbolic form taken from a fixed alphabet in order to “to guide the machine improvisation output in ways that are meaningfully matched to the live musical input” (Surges & Dubnov, 2013). A more recent theoretical investigation of meaningfulness (Surges, 2015) also considers components such as
listeners’ expectations and boredom, as well as machine self-evaluation mechanisms. This approach, as well as that of multidominance implemented as reflexive interactions between the user and a mirror image of themselves (Pachet, 2008) in an interactive, reflexive musical system (Ferrari & Addessi, 2014), are critical to consider in the study of computational meaningfulness.

**VIVO and the simulation of intentionality**

A final approach is exemplified by VIVO, a wakeful system (Paolizzo, 2013). Here, the term wakeful incorporates the philosophy that computers and human beings can interact via an evolving language, which is expressed by VIVO in its simulation of intentionality. Ultimately, the evolutionary and open-ended nature of this framework provides for an improvement in terms of autonomy.

VIVO implements theories from previous research on music and interaction (Paolizzo, 2006), which investigate the relationship between the human capacity for interpreting and interacting with music. In a typical scenario (Figure 1), the user controls a sound source (i.e., a musical instrument), that sends an audio signal to VIVO for sound processing. In turn, and within a user-defined computational score, the system analyzes the interaction and uses that information to control the processing of the audio signal.

![Figure 1. Typical interaction model for a VIVO/user instance.](image)

**Design of VIVO**

VIVO is an open source computer program developed in MAX/MSP (Cycling 74, 2017), which is capable of real-time audio processing and sound synthesis by loading and using external audio plug-ins (VST, VSTi, DirectX, AU) in the program. Figure 2 shows an overview of the system architecture. The system is comprised of different software components: (a) an adaptive video tracking module, (b) a salience detection algorithm, (c) an editor for stochastic scores, (d) a single graphical user interface for every loaded audio plug-in, (e) a dynamic host for automating the audio plug-ins, and (f) network and web components to send and receive external data for extended configurations.

In the present study, we focus on the influence and implications of (b) the salience-based detection algorithm on the interplay when receiving data from (a) the video tracking module. These two software components allow the simultaneous detection and mapping both of the Quantity of Motion (QoM) — the amount of movement in a video stream (Camurri et al., 2003; Paolizzo, 2013) — and of the Salience of Action (SoA) — the dispersion either of the QoM or of the manual control on the graphical user interface. The SoA is computed by (b) the salience detection algorithm. Multidominance is implemented by (a) the adaptive video tracking module, (c) the dynamic host and (f) the network and web components.
The meaningfulness of the system generation is sought in VIVO via salient sonic changes that are synchronic to the user’s interaction with the system, which affords control to the user. These changes create an expectation for meaning to be found in the outcome for both the user and the audience, and provides both parties with a basis for identifying the automatic sound generation as novel music via a reflexive interaction, as the present study will show. This ultimately allows the user to identify the sonic generation as a meaningful simulation. The generation framework offers additional open-ended creation capacity to the user via the external audio plug-ins loaded. Implications of the system architecture are discussed throughout this article, while the salience-based detection algorithm is discussed in the Methodology.

**Figure 2.** System architecture of VIVO.

**Autonomy and computational creativity in VIVO**

Earlier, we discussed autonomy in the present context as indicative of self-determination. However, in considering our tridimensional continuum the term ‘configuration’ is seemingly more appropriate than ‘determination’ in describing computer invention in an IMS. While an IMS can automatically generate novel music, it does so under the purview of the generative framework indicated by its score. While imperceptible to the user, the autonomy of such a framework is naturally limited. So, although the capacity of a sound algorithm for self-configuration is relevant for computer invention, autonomy as a property is not sufficient for a machine to interact as a human could do in terms of creativity. For example, a machine musician could be producing outcome that is very coherent within its own schema, but which does not necessarily produce anything meaningful to a human interactor.

**Building a model of musical meaningfulness**

Coherence and novelty are not sufficient criteria for creativity. Creativity is instead a complex feature of intelligence whose literature is somehow well established (e.g., Wallas,
The computational modeling of creativity can be approached as a historical or psychological problem (Bolden, 1998), through music (Wiggins, 2001), and can extend beyond the use of symbols to favor instead features commonly found in connectionism, geneticism and artificial life (Dartnall, 2013). On the other hand, computational creativity can come from a simpler perspective. Indeed, “skillful, appreciative and imaginative behaviours” have been proposed to be “the bare minimum required to support the perception of creativity” (Colton, 2008).

The perception of creativity need not imply an understanding of its source. For example, the perception of music relies on one’s capacity to have meaningful expectations surrounding the listening experience (Delalande, 1993). The study of musical expectations has a long tradition (i.e., Narmour, 1992; Huron, 2006; Meyer, 2008; Deliège et al., 2011), and recent advances show that expectations lead our perceptual, cognitive and emotional reactions to music because they permeate all brain functions (Trainor & Zatorre, 2015). Computational learning models applied to the modeling of musical expectations are adequate in reflecting behavior changes and cultural differences or accounting for relevant non-musical information that is sequentially structured such as language, visual sequencing and motor planning (Pearce & Wiggins, 2012). However, such models are not the only alternative to rule-based modeling.

The growing interest in using dynamic systems theory and tools to study and model cognition lies on the claim that cognition is a dynamic phenomenon (see Wilson et al., 2017). However, the very existence and nature of cognitive representations (e.g., static/dynamic, in/variant, non/context-dependent, and dis/embodied) is still uncertain, and dynamic systems theory and perceptual symbol theory have been proposed as complementary to each other in their so-called embodied representations (Dietrich & Markman, 2014). Such a perspective provides further conceptual backing to the present tridimensional continuum for IMSs. We define here the score of an IMS as a symbolic system dynamically embodied by both the user and the computational musical instrument.

Human-computer interaction in an IMS can be seen as a communication problem determined by expectations, in which interactions are most meaningful when the computational outcome is ascribable to an environment of communication feasible to the user’s mind (Paolizzo, 2011). Much as a human musician might not understand the algorithmic origins of computer-generated music, a computer is absolutely not able to distinguish why any information is particularly meaningful to a human agent — an important point in the scope of next-generation IMS design. Communicative commonality between the user and the system can and should be integrated into the system in order to improve computer-human understanding. We want IMSs to introduce novelty and not just be inferior imitations of human improvisers. However, we want them to do so in a way that is not too different from what humans may recognize as music.

The extra-musical informing the musical

While we may be limited to creating computer systems that are only as inventive as we are, humans recognize and understand music by drawing from a wide variety of contexts, most of which are extra-musical. For example, flow is a mental state wherein a person’s abilities are closely matched to the demands of a task(s), thereby leading to an optimal state of cognitive functioning and immersion (Steels, 2004). Similarly, the concept of a sense of coherence (SoC)
Antonovsky, 1987; 1993) describes a subjective experience that contributes to the experiential knowledge base that fuels human interpretation. Therefore, a key element for an optimal musical experience lies in the balance between the complexity of a musical task, the user's level of expertise in the task, and their perceived engagement in the task. This assumption is rooted in studies on flow theory that specifically investigate IMSs (Pachet & Addessi, 2004; Pachet, 2008). Flow and SoC are collinear concepts (Lutz, 2009), and the first empirical application of their integration specifically aimed at IMSs (Paolizzo, 2013) is further validated in the present study.

Extra-musical information is understood here as a type of extradiegetic element that affects the interplay between the user and the IMS. It can operate similarly to extrinsic motivations (i.e., external rewards), which are a fundamental component of flow and that can hinder or foster intrinsic motivations (Engeser & Rheinberg, 2008). Extra-musical information contributes to determining the extrinsic motivations of the interplay, which also draw from implicit (unconscious) motives such as achievement, affiliation-intimacy, and power — the so-called ‘big three’ motives (McClelland, 1987; cf. Heckhausen & Heckhausen, 2008). These unconscious motives orient the musical attention, and select and energize the musical conduct towards those specific classes of musical object-goal directedness identified by the user as providing task-intrinsic incentives/rewards. Explicit (conscious) motives, such as “cognitively-based verbal learning of rules, demands, and expectations that reflect people’s self-attributed view” (Schiepe-Tiska & Engeser, 2012), also influence the decision for or against a musical action/opportunity. From this perspective, we can describe the opportunity for a musical action in an interplay in terms of salience.

In music, salience is a factor of semiotic importance and prominence that and contributes to, and concurs with, the perception of causation (Ockelford, 2004). Salience is thus a factor informing the human interplayer about the potential effects of a musical act on the interplay, similarly to a ‘vested interest’ influencing the subject’s self-efficiency (Crano, 1995). A computational detection of salience within a data stream representing aspects of the interplay can be used to manifest that potential for the interplayer to act.

In the first section of this article, we discussed a number of IMSs that rely on mutual listening. Such systems implicitly operate in terms of salience or could easily be enabled to do so. By detecting and using salience for the sound generation, an IMS can influence the musical conduct much as a human interplayer could. Such systems generally derive the data for the computation from the music played by the human interplayer. It is important to note that the explicit use of salience allows drawing effectively from non-musical information which is sequentially structured (i.e., language, visual sequencing, motor planning), as long as it is inherent to the interplay.

In the typical interaction model of a VIVO/user instance (Figure 3), a user aiming at music-making produces extra-musical actions (i.e., physical gestures on a musical instrument) carried out with a musical intention. VIVO uses salient information from these actions in order to generate a sonic outcome. Congruently to the concepts of flow and SoC, the user is thus caught in an action-reaction loop of self-reflection (Paolizzo, 2010a).
Current IMSs operate mainly, if not exclusively, at the sonic/musical level, and because of this limitation have very little autonomy as the algorithms dictating sonic generation are
greatly constrained by the score and the user’s control. Recent studies suggest that listeners are able to change their semantic relations with the sonic world through functional adaptation at the level of sensing, acting and coordinating between action and perception, in biological, psychological, and cultural terms that involve motor, kinesthetic, haptic and visual, besides the purely auditory components (Reybrouck, 2014). Drawing from Hans-Georg Gadamer (Malpas, 2009), because human interpretation of musicality is constantly evolving with new musical and non-musical experiences we might expect the state of the art in IMS research to be delivering algorithms that work on a multitude of levels using both musical and extra-musical components, such as culture and psychology. For example, research on next-generation systems of interactive music information retrieval (MIR) — like the Musical-Moods project (Paolizzo, 2017) — is focused on enabling systems to mold automatic music generation according to cross-modal classification, as well as by its recognition and prediction of features from both musical and extra-musical information (i.e., emotional states).

The present research describes pilot studies in which VIVO has been used for music-making. The purpose of these studies was to test the system’s functionality and highlight implementation strategies that would maximize meaningfulness for both the user and audience of an open-ended computer generation under different conditions of interaction.

**Methodology**

In the present pilots (Table 1), feedback regarding the quality of the experience was solicited from every participant. Following the concept of SoC, quality was assessed in terms of comprehensibility of the musical relations, manageability of the available resources, and whether the demands posed by the interplay were capable of inducing meaningful expectations. In music semiology, a verbalization regarding a referent can be compared with those from others participating in the experience (Stefani, 1998). Similarly, musical coherence was defined here in terms of the user’s verbalized experience — their subjective perception of the quality of the musical interaction and outcome.

| VIVOTube | An audio/video installation and a dedicated internet page hosting the installation simultaneously. It allows on-site and internet users to search and select videos from YouTube and load it in VIVO for sound processing. The generated soundtrack resulting from the mapping between the sound processing by GRM Tools audio plug-ins is mapped to the motion of the images detected in the video. |
| --- | --- |
| Studio1 | A musical piece for guitar and VIVO that represents a typical standard example of pilot study. It includes a stochastic and user-defined computational score and a sound processing scenario, progressively refined through rehearsals. A camera detects the musician’s physical motion in order to control the sound processing. |
| Interactive Music Group | A group of performers and researchers established at the Centre for Music Technology in the University of Kent, in order to conduct research in the area of interactive music. The music works composed and conducted for the group (Paolizzo, 2010b) defined and adopted rules for improvisation aiming at influencing the performers’ musical conduct in aiming at musical coherence. |
| Velodrone | A cycling and musical event. The project aimed to extend the conventional concept of a competition and of a concert. Participants created the music by cycling. Physical motion related indirectly to sound generation (extradiegetic mapping). Users experienced flow in the experience but did not perceive a sense of coherence in the interplay. |
| Invisible Cities | A short movie for prepared piano and VIVO that fits under the umbrella of ‘remix cinema’. Only motion in the video is mapped to the sound processing. Commonality of interaction between the musician and VIVO is achieved by monitoring changes in the motion threshold, as cues for the interplay (both intradiegetic and extradiegetic). |
In the pilots, either a camera or a video file player were connected to the adaptive video tracking module. The QoM was mapped in the dynamic host to specific sub-ranges of multiple parameters (one-to-many / QoM-to-parameters) of the audio plug-ins (one-to-many / QoM-to-plug-ins) loaded into the system (INA GRM, 2004), in order to process the input audio signal. Furthermore, the implementation of a salience detection algorithm allowed using the QoM as an indicator of notable changes in the user interaction.

The computation of salience, SoA, draws from that of the standard deviation, $\sigma$, (Figure 4) of the QoM. In comparison to standard deviation, SoA allows (a) adjusting the computation to the context as based on a time window, (b) compensating for low frequency of execution time (i.e., for a low computation cost or when syncing to a video framerate) — and (c) improving the response time in the mapping as a result of the shorter ramp times allowed by the algorithm. Computation QoM in terms of SoA, allowed setting and experimenting with threshold values for the automatic enabling/disabling of each plug-in loaded. When the SoA triggers the threshold, the system changes the plug-ins generating the sound, in order to reflect the change detected in the QoM. The present approach draws from and extends that previously implemented for non-real-time simulation environments (Impett, 2001).

In the pilots, computer invention was modelled by tuning the system to both afford musical meaningfulness during automatic generation to each user based on their personal extra-musical background, and incorporate musical comprehensibility and manageability of sonic stimuli. To achieve this end, feedback indicating each listener’s semiotic process was compared, and future iterations of the system were adjusted accordingly (Paolizzo, 2013). For example, when listeners could not establish connections between the internal elements of a piece, they also described difficulties in comprehension, management, and attribution of meaningfulness to the sonic structures during the listening experience (i.e., as in VIVOtube, Studio1 or in the Interactive Music Group). Consequently, specific improvements addressing these difficulties were implemented in later pilots (e.g., removing user-identified problematic plug-ins from subsequent versions of a score, fine-tuning system parameters or delivering pilots differently designed). By modeling the sound generation in terms of salience of action, there existed an indeterminacy in the design that resulted in an openness of the music paradigm that did not necessarily lack in musical coherence. This approach provided grounds for the user(s) and the system to achieve musical coherence in a non-deterministic interplay, in order to foster the system autonomy.
Implementing musical meaningfulness

In regards to musical coherence, it is important to focus on a technique that could elicit a sense of coherence in every user rather than from a limited subset of users. For a user to manage and comprehend sonic stimuli as music, those stimuli must exhibit a structure that can be feasibly interpreted by said user. There is a general agreement that human beings expect to find a meaning in every outcome that appears to be intentionally organized (e.g., as in the theory of mind, Premack & Woodruff, 1978). This principle refers to the inherent structure of how the human mind works in terms of expectations, and regardless that believed-in classes of mental states actually exist in reality (Dennett, 2003). As previously discussed, expectations of meaningful structural organization in music, as in any other language — regardless of the
extent to which we actually consider music to be a language — is a central factor in human interpretation (Delalande, 1993) and in a listener’s attribution of meaning to computer-generated music. In the context of the active user/musician, musically-related interactions are consequently intentional actions, carried out with a musical goal-directedness embodied as motor knowledge (Menin & Schiavi, 2012). Such a sensorimotor integration is also defined as the action-reaction cycle (Leman, 2008). In this cycle, “motor actions, auditory signals and visual information become coupled and lead to a strengthening of the connections between the associated regions of the brain” (Keebler et al., 2014).

In regards to an IMS, our tridimensional continuum describes a computational model of creativity where previous extra-musical experience, for example in the form of analogies, patterns, or language, could be coupling with and thereby influencing our musically-related experiences and interaction with the system. Given that previous experiences allow a user to anticipate what is likely to happen next, it is important to consider the role of the IMS in supporting — or breaking — these predictions.

**Gestural embedding**

In an acoustic instrument, the action-reaction cycle is the basis of instrumentality and central to playing a traditional musical instrument (Maes et al., 2014). Similarly, the principle of action/perception holds that when we excite the physical body of an acoustic instrument, we can see the direct relationship between our actions on it and the sound that we hear in a process of identification-through-repetition (e.g., Emmerson, 2000). Conversely, in an IMS, the physical binding is irrelevant to the kind of sound that can be produced. The same computational substrate can produce many different sounds without physically constituting the source of the sound. In such a system, the action-reaction mapping is embodied in a different medium — the sonification process — during which the quality of a gesture shapes the music. For example, in VNS (Rokeby, 2010) video-to-sound mapping has some similarities to VIVO: for both systems, the users can find themselves in an action-reaction cycle from physical motion to sound. In the present research, electroacoustic music, wherein sound algorithms process one or more sound sources, was used (as defined by Landy, 2006), rather than sound synthesis. Hence, while the system architectures may appear similar in regard to some of the components (i.e., the adaptive video tracking module mapped to the dynamic host), they are different substantially.

Electroacoustic music is not merely a matter of sound generation but rather a sonic abstraction, wherein “the sources and causes of sound-making become remote or detached from known, directly experienced physical gesture and sounding sources” (Smalley, 1997, p. 112). Electroacoustic music listening is intrinsically acousmatic, as it is always perceivable — to a variable degree — in terms of sound-objects (Shaffeur, 1977) maintaining non-dependent relationships to their origins (Wishart, 1986). Although the study of human cognition in electroacoustic music is mostly unpaved, the literature offers diverse perspectives (e.g., Di Scipio, 1995; Smalley, 1996; Eaglestone et al., 2008; Van Nort, 2009; Emmerson, 2013) letting us trace back to such a phenomenon of remoteness. The phenomenon is particularly relevant in an IMS, as a piece of information technology mediates the musical interaction, and consequently the sonic outcome and its source. This results in a complex and reiterated-but-changing mapping between action and sound. The scope of such a mediation surpasses physicality and perception as it projects, for the user, sound generation to a cognitive dimension of musical
expectations. The musical goal-directedness constituting and embodied by the musician as motor knowledge is, as mentioned, embedded in VIVO through mapping. VIVO works for the user as a means for the cultural embedding of gesture, which is typical in what is known as *gestural surrogacy* — the process of increasing remoteness (Smalley, 1997). Notably, while remoteness can foster an evolving interpretation for the listener, it can also reduce the perception of causality between action and perception.

In VIVO, the user’s and audience’s attribution of meaning to the automatically generated sound is dependent on causal action-perception relationships suggested by the system. For the audience, this algorithmic generation is visible in the source of the computed QoM (i.e., a musician playing an instrument). For the active user, the system’s use of a salient threshold allows sound generation to change in correspondence with salient actions, ultimately increasing the coherence between the sound source and the electronic sound. These changes are thus structural musical cues, which concur to a form of intentional synesthesia through technology (Suslick, 2012). Salient actions, although extra-musical, act here as intradiegetic by informing algorithmic generation, thereby preserving musical coherence while also increasing remoteness.

**Broadening the action/perception feedback loop**

As explored in the pilots (Paolizzo, 2013), gestural surrogacy can occur in VIVO when a user interfaces with a software instrument and a directly mapped relationship is formed between that action and the perception of a sonic outcome. This action/perception feedback loop is a critical component in musical meaningfulness. In the present study, action/perception was considered as non-musical, experiential information for determining computer invention. A camera watching the user’s body and surrounding space (as first explored in *Studio1*) or a video file (as in *VIVOtube* and *Invisible Cities*), provided this information. In both cases, non-musical information drove automatic generation and extended the user’s agency in terms of gestural surrogacy and sense of coherence. This was achieved in different ways: (1) when instructions were sent to the machine for sound generation (as in the preparation of *VIVOtube*), (2) through the processing of sound resulting from physical gestures on a musical instrument (as in all the pilots, with the exception of *VIVOtube* and *Velodrone*), (3) through gestures on physical interfaces connected to software instruments (as in *Velodrone*), and (4) through any gesture (e.g., dancing, as in *Collective*) or multimedia providing motion dynamics that could be mapped to a software instrument (as in *VIVOtube, Invisible Cities* and *Collective*). It should also be noted that there were instances wherein a performative gesture could not be mapped, for example when using a video file (as in *VIVOtube* and *Invisible Cities*), or when the interface was a physical device (as the bicycles in *Velodrone*).

In every pilot, VIVO generated for each input information a simultaneous auditory feedback. This feedback referred to visual sequencing in watching a video file determining the QoM or a user-VIVO interaction in the physical space, or to users’ proprioception captured by a camera. We will now discuss further and verify the nature of this reflective and embodied experience for the user.
Enabling self-reflection mechanisms into VIVO

A feature of any acoustic musical instrument is to afford an experience characterized by a unified vision of body, mind and environment as a form of embodied cognition. Drawing from Cowart’s definition of cognition as embodied, we propose that a cognitive process develops for the user “when a tightly coupled system emerges from real-time, goal-directed interactions” (2004) between that user and the autonomous system. In terms of our tridimensional continuum, each dimension of an IMS can exhibit embodiment as a property. In the continuum, the coupling between user and system takes place as the instrumentality of the IMS — the control interactively retained by the user (Figure 3). Embodiment is thus a factor depending on the use of the system, which is however implicit in the design, as discussed. The score is the cognitive and computing dimension of that physical coupling. This abstract dimension comprises both the musical and the extra-musical knowledge embedded into the system, and extends into the user’s cognitive processes developing from the interaction. Finally, system autonomy is the dimension where that physical/cognitive coupling emerges and develops as a result of the user’s interaction with a system responding like an environment that is musically defined; the system responds to the user as not only an autonomous player, but also one that is capable of a musical interplay.

As discussed, the user’s perception of causality in an IMS reinforces that of musical coherence regarding the system generation and, consequently, of its meaningfulness for that user — musical expectations receives a grounding from perceiving causality. In electroacoustic music, as in IMSs lacking embodiment in any of the three dimensions of the continuum, the separation between action and perception that occurs while increasing remoteness also precludes this embodiment experience for the user. When using VIVO, users are able to achieve gestural surrogacy while still being afforded an anchor, in the form of salient musical cues, on which to ground their meaningful expectations. Grounded cognition theories postulate that the brain intrinsically ties sensory information to the perceptual modality in which that information is perceived (Barsalou, 2008; Pezzulo et al., 2013). According to such a view, both acoustic instruments and VIVO allow multimodal information to shift dynamically for the user, “in reaction to the instrument and one’s interaction with it” (Keebler et al., 2014). However, in contrast to acoustic instruments, VIVO is a piece of information technology that mediates (processes), and reflects (re-presents), the user’s interactions.

In considering human cognition as embodied, VIVO was designed not only for facilitating the user’s perception of system autonomy through stochastically-generated sonic constructs recognized as music by the user, but also for stimulating a subjective capacity for self-reflection. Self-reflection is implemented by design through interactions that imply rehearing, reproduction, and variation. In such a self-reflection, both the perceived self and the perceiving self mirror each other through musical constructs that embody the agent’s activity. VIVO affords the user a complementary way to extend their inner body knowledge in order to experience self-reflection in music. A mapping between motion tracking algorithms, sound generation and the user interface is established. The pilot research suggested that self-reflection may be the cognitive process that emerges and develops in an IMS where score, instrument and player dimensions are embodied by both the user and the system. We will now further evaluate this working hypothesis.
Case study

Figure 5 (also see Supplemental Material) depicts a transcription of the audio recording from the Collective 4 pilot (Table 1), illustrating a free improvisation between a trombone player and VIVO. The transcription was generated automatically from the recording via the automatic music transcription software Melodyne 4 (Celemony Software, 2017) using standard settings for polyphonic music. The transcription was adjusted manually in the engraving process for both the trombone and the VIVO parts of the score to reflect the actual playing. In the VIVO part, only salient cues are engraved, in contrast to also engraving those cues where timbre is considered predominant over pitch.

**Figure 5.** Excerpt of score (automatic transcription) from Collective4.
Here, VIVO interplays with the musician (Figure 5.1) and the musician’s response results in the triggering of the salience threshold of motion (5.2). The musician recognizes this trigger as an indication that they are achieving a meaningful interplay with the system. The musician also recognizes that only partial control can be retained over the partially autonomous system. This is verified as the musician listens to VIVO (5.3), and then shapes his own playing accordingly (5.4) thereby triggering the threshold once again (5.5). Notably, the musician’s achievement of musical phrasing after the first trigger (5.6) confirms the intentionality of this second triggering. VIVO’s interplay depends on the musician’s playing but because the system indicates a certain level of autonomy, this stochastic control generates a coherent musical outcome. Furthermore, the musician does not attain musical coherence casually, for example, by independently adding his own playing to the computer generation. Instead, the musician achieves musical phrasing in what we describe as reflexive multidominance by establishing an ongoing interplay with VIVO (Figure 5.6 – 5.7) and proactively elaborating on its outcome. The combination of these user-system properties facilitates the user’s recognition of the outcome as musical.

In the case study, VIVO works for the musician both as an instrument and as an autonomous player. As an instrument, the system extends the musician’s capacity for music-making through embodiment; the system is here an extension of the acoustic instrument controlled through physical gestures. The score embeds VSTs loaded in the dynamic host (musical knowledge) and uses QoM and SoA from the musician’s gestures on the acoustic instrument (extra-musical knowledge) into the system. Because the musician engages in reflexive multidominance through the interplay, the study also suggests that system autonomy is here a property shared with the user: the score extends to the user’s cognitive processes as a result of their perception of that autonomy.

**Conclusion**

The capacity for an autonomous interactive music system depends on how musical intelligence is implemented such that the user can access and control it. The present...
tridimensional continuum makes the role of musical knowledge explicit in the system architecture — the score — and allows for the implementation of computer invention in terms of what listening affords to the user's perception and cognition, rather than only in terms of acoustic qualities. Notably, a system may appear autonomous to the user while it is actually only capable of simulating intentionality during automatic sound generation. VIVO is a system of this kind. The system architecture of VIVO is modelled on the premise of collinearity between mental flow and sense of coherence, thereby enabling a pro-active user to fully engage in music-making with the system and maximize musical coherence. The system analyzes and re-uses extra-musical information derived from the user's music-making process to inform the automatic sound generation.

In VIVO, a video component monitors the user's physical gestures while producing music and triggers a relevant change in the sound generation for when thresholds for salient motion have been met. Because the user's musical intention determines this extra-musical information, the user enters an interplay with the system but does so while actually engaging in a process of self-reflection driven by goal-directed interactions and expectations. As a result, the automatic sound generation is perceived as meaningful regardless of the indeterminacy employed by the system. This is an improvement over current systems, which have no real 'embeddedness' beyond what they immediately do. Through different pilot projects (Table 1), the study related the system outcome to the participants' interaction with the system and their verbalized experiences in order to address the subjective perception of the quality of the musical interaction and outcome and to refine the system architecture accordingly. Validation for this approach is evidenced by the musician's recognition that a meaningful interplay — what we have described here as reflexive multidomiance — was in fact established with VIVO. The study also suggests that the interaction with an IMS may extend to the user's cognitive processes when resulting from (i) instrumentality extending the user's music-making capacity through embodiment, and (ii) the user's perception of autonomy in the interaction as a property that is shared with the system.

VIVO is currently being used in the Musical-Moods project as a basis for next-generation systems of artificial music intelligence to draw from both extradiegetic and intradiegetic multimodal information to model computational creativity. Future works will apply the present validation approach to the current multimodal integration in the project. The approach will also include an analysis of the interplay between multiple users/VIVO instances, as well as a comparison to other systems in terms of quantitative analysis.

**Supplemental material**

*Collective4,*
audio/video documentation:
https://archive.org/details/collective04

VIVO,
audio/video tutorial:
References
Antonovsky, A. (1987). *Unraveling the Mystery of Health: How People Manage Stress and Stay Well*. San Francisco: Jossey-Bass.

Antonovsky, A. (1993). The structure and properties of the sense of coherence scale. *Social Science & Medicine, 36*(6):725–733.

Assayag, et al. (2006). OMax Brothers: a Dynamic Topology of Agents for Improvisation Learning. In *Proceedings of the First Workshop on Audio and Music Computing for Multimedia (AMCM’06)*, Santa Barbara, CA.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*:617–645.

Biles, J. A. (1999). Life with GenJam: Interacting with a musical IGA. In *Proceedings of the 1999 IEEE Conference on Systems, Man and Cybernetics*, pp. 652–656. Tokyo, Japan.

Blackwell, T., & Young, M. (2005). Live algorithms. *AISB Quarterly, 122*:7–9.

Bown, O., & Martin, A. (2012). Autonomy in Music-Generating Systems. In *1st International Workshop on Musical Metacreation*. Palo Alto, CA.

Brown, C., & Bischoff, J. (2002). *Indigenous to the Net: Early Network Music Bands in the San Francisco Bay Area*. Retrieved on June 1, 2017 from: http://crossfade.walkerart.org/brownbischoff/IndigenoustotheNetPrint.html

Camurri, A., Mazzarino, B., & Volpe, G. (2003). Analysis of Expressive Gesture: The EyesWeb Expressive Gesture Processing Library. *Gesture Workshop, 2*(915):460–467.

Celemony Software GmbH. *Melodyne Editor* (Version 4.0) [Software]. Retrieved on June 1, 2017 from: http://www.celemony.com

S Chadabe, J. (1984). Interactive Composing: An Overview. *Computer Music Journal, 8*(1), 22–27. doi:10.2307/3679894

Colton, S. (2008). Creativity Versus the Perception of Creativity in Computational Systems. In *AAAI Spring Symposium: Creative Intelligent Systems*, pp. 14–20.

Cowart, M. (2004). Embodied Cognition. *Internet Encyclopedia of Philosophy*. Retrieved on June 1, 2017 from: http://www.iep.utm.edu/embodcog/

Cope, D. (2010). Recombinant Music Composition Algorithm and Method of Using the Same. *Google Patents*.

Cope, D. (2016). Public lecture/personal communication. *Gassmann Electronic Music Series*. University of California Irvine. January, 2016.

Dannenburg, R. B. (1984). An On-Line Algorithm for Real-Time Accompaniment. In *Proceedings of the 1984 International Computer Music Conference (ICMC–84)*, pp. 193–198. Paris.

Dartnall, T. (Ed.). (2013). *Artificial intelligence and creativity: An interdisciplinary approach* (Volume 17). Springer Science & Business Media.

Delalande, F. (1993). *Le condotte musicali. Comportamenti e motivazioni del fare musica e ascoltare musica*. Bologna: Editrice Clueb.

Deliège, I., Davidson, J., & Sloboda, J. A. (2011). *Music and the mind: Essays in honour of John Sloboda*. Oxford University Press.
Dietrich, E., & Markman, A. B. (2014). *Cognitive dynamics: Conceptual and representational change in humans and machines*. Psychology Press.

Di Scipio, A. (1995). Inseparable models of materials and of musical design in electroacoustic and computer music. *Journal of new music research*, 24(1), pp. 34–50.

Douglas, R. L. (1991). Formalizing an african-american aesthetic. *New Art Examiner*, June 1991. pp. 18–24.

Drummond, J. (2009). Understanding Interactive Systems. *Organised Sound*, 14(2):124–133.

Emmerson, S. (2000). Losing touch?: the human performer and electronics. In S. Emmerson (Ed.), *Music, Electronic Media and Culture*. Aldershot, UK: Ashgate Publishing Limited.

Echeveste, J., Cont, A., Giavitto, J.L., & Jacquemard, F. (2013). Operational semantics of a domain specific language for real time musician-computer interaction. *Discrete Event Dynamic Systems*, 23(4), pp. 343–383.

Eaglestone, B., Ford, N., Holdridge, P., & Carter, J. (2008). Are Cognitive Styles an Important Factor in Design of Electroacoustic Music Software? *Journal of New Music Research*, 37(1), pp. 77–85.

Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, 32(3), pp.158-172.

Ferrari, L. & Addessi, A.R. (2014). A New Way to Play Music Together: The Continuator in the Classroom. *International Journal of Music Education*, 32(2), 171–184.

Gartland-Jones, A. (2003). MusicBlox: A real-time algorithmic composition system incorporating a distributed interactive genetic algorithm. In S. Cagnoni et al. (Eds.), *Applications of Evolutionary Computing*, Vol. 2,611 of Lecture Notes in Computer Science, pp. 145–155. Berlin/Heidelberg: Springer.

Griffiths, P. (2001). Aleatory. In S. Sadie & J. Tyrrell (Eds.), *The New Grove Dictionary of Music and Musicians* (2nd Edition). London: Macmillan Publishers.

Huron, D. B. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT press.

Impett, J. (2001). Interaction, Simulation and Invention: a Model for interactive music. In E. Bilotta, et al. (Eds.), *Artificial Life Models for Music Applications*, pp. 108–119.

Keebler, J.R., Wiltshire, T.J., Smith, D.C., Fiore, S.M., & Bedwell, J.S. (2014). Shifting the paradigm of music instruction: implications of embodiment stemming from an augmented reality guitar learning system. *Frontiers in Psychology*, 5.

INA-GRM (2004). GRM tools (Version 2) [Software], Windows and OSX. Retrieved on June 1, 2017 from: http://www.inagrm.com/categories/anciennes-versions-older-versions/

Kaufman, J. C., & Sternberg, R. J. (Eds.). (2010). *The Cambridge handbook of creativity*. Cambridge University Press.

Landy, L., (2006). Electroacoustic music studies and accepted terminology: you can’t have one without the other. In *Electroacoustic music studies conference*. Beijing.

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

Lévy, B., Bloch, G., & Assayag, G. (2012). OMaxist Dialectics: Capturing, Visualizing and Expanding Improvisations. In *Proceedings of the 2012 Conference on New Interfaces for Musical Expression* (NIME 2012). Ann Arbor: University of Michigan.
Lewis, G. E. (2000). Too Many Notes: Complexity and Culture in Voyager. *Leonardo Music Journal, 10*:33–39. Cambridge, MA: The MIT Press.

Machover, T., & Chung, J. (1989). Hyperinstruments: Musically Intelligent and Interactive Performance and Creativity Systems. In *Proceedings of the International Computer Music Conference (ICMC)*, pp. 186–190.

Maes, P. J., et al. (2014). Action-based effects on music perception. *Frontiers in Psychology, 4*(1,008):1–14.

Malpas, J. (2009). Hans-George Gadamer. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Stanford: The Metaphysics Research Lab. Retrieved on June 1, 2017 from: http://plato.stanford.edu/archives/sum2009/entries/gadamer/

Menin D., & Schiaivo, A. (2012). *Rethinking musical affordances*. AVANT, 3(2):202–215.

Meyer, L. B. (2008). *Emotion and meaning in music*. University of Chicago Press.

Narmour, E. (1992). *The analysis and cognition of melodic complexity: The implication-realization model*. University of Chicago Press.

Ockelford, A. (2004). On similarity, derivation and the cognition of musical structure. *Psychology of Music, 32*(1), 23-74.

Pachet, F. (2002). The Continuator: Musical Interaction with Style. In *Proceedings of ICMC*, Göteborg, Sweden, September: ICMA, pp. 211–218.

Pachet, F. (2008). The future of content is in ourselves. *ACM Computers in Entertainment (CIE), 6*(3):31.

Pachet, F. (2016). A joyful ode to automatic orchestration. *ACM Transactions on Intelligent Systems and Technology (TIST), 8*(2), p.18.

Pachet, F., & Addessi, A. R. (2004). When Children Reflect on Their Playing Style: The Continuator. *ACM Computers in Entertainment, 1*(2):14.

Pachet, F., Roy, P., Barbieri, G. & Paris, S.C. (2001). Finite-length Markov processes with constraints. *Transition, 6*(1/3).

Paolizzo, F. (2004). VIVO (Video Interactive VST Orchestra), Version 1. Rome, Italy.

Paolizzo, F. (2006). *Musica e Interazione*. Master Thesis, University of Rome “Tor Vergata”.

Paolizzo, F. (2010a). VIVO (Video Interactive VST Orchestra) and the Aesthetics of Interaction. In M. Wolf & A. Hill (Eds.), *Proceedings of Sound, Sight, Space and Play 2010 (SSSP2010)*, June. Leicester, UK: De Montfort University.

Paolizzo, F. (2010b). VIVO (Video Interactive VST Orchestra): An Interactive and Adaptive Musical Instrument for Self-reflection in Music. In: *The International Journal of the Arts in Society*. Vol. 4(6), pp. 149–159.

Paolizzo, F. (2011). Wakeful software and wakeful musical instruments: a theoretical approach to the implementation. In: *Proceedings of the 17th International Symposium on Electronic Art (ISEA 2011)*, September. Istanbul, Turkey. Retrieved on June 1, 2017 from: http://isea2011.sabanciuniv.edu/~paolizzo/

Paolizzo, F. (2013). *VIVO: a Wakeful Instrument for Collective Musical Embodiment*. Ph.D. Thesis, University of Kent.

Paolizzo, F. (2014). VIVO (Video Interactive VST Orchestra), Version 2. GPLv3. Canterbury, UK: University of Kent. Retrieved on June 1, 2017 from: http://fabiopaolizzo.com/pub/

Paolizzo, F. (2017). Musical-Moods. Retrieved on June 1, 2017 from: http://www.musicalmoods2020.org/
Pearce, M. T., & Wiggins, G. A. (2012). Auditory expectation: The information dynamics of music perception and cognition. *Topics in cognitive science*, 4(4), 625-652.

Perkis, T. (1999). *The Hub, an Article Written for Electronic Musician Magazine*. Retrieved on June 1, 2017 from: http://www.perkis.com/wpc/w_hubem.html

Reybrouck, M. (2014). Music and its inductive power: a psychobiological and evolutionary approach to musical meaning. Proceedings of *ICMPC-APSCOM 2014. Joint Conference*, pp. 100–104. College of Music, University of Yonsei.

Pezzulo, G., et al. (2013). Computational grounded cognition: a new alliance between grounded cognition and computational modeling. *Frontiers in Psychology*. 3(612):1–11.

Ritter, M., Hamel, K. & Pritchard, B. (2013). Integrated multimodal score-following environment. In *ICMC*.

Rokeby, D. (2010). *Very Nervous System* (1986-1990). David Rokeby-Artist. Retrieved on June 1, 2017 from: http://www.davidrokeby.com/vns.html

Sanfilippo, D. (2012). LIES (distance/incidence) 1.0: a human-machine interaction performance. In *Proceedings of the 19th Colloquium of Musical Informatics (XIX CIM)*. Triest, Italy, pp. 21–24.

Sanfilippo, D. (2015). Dario Sanfilippo – Projects. Retrieved on June 1, 2017 from: http://dariosanfilippo.tumblr.com/

Schiemer, G. (1999). Improvising Machines: Spectral Dance and Token Objects. In: *Leonardo Music Journal*. Vol. 9(1), pp. 107–14. Cambridge, MA: The MIT Press.

Shaffeur, P. (2017). *Treatise on Musical Objects: An Essay across Disciplines*. Trans. Christine North and John Dack, University of California Press.

Smalley, D. (1996). The listening imagination: listening in the electroacoustic era. *Contemporary Music Review*, 13(2), pp. 77–107.

Smalley, D. (1997). Spectromorphology: explaining sound-shapes. *Organised sound*, 2(2):107–126.

Steels, L. (2004). The autotelic principle. In Embodied Artificial Intelligence, Lecture Notes in Artificial Intelligence (Volume 3139). In Fumiya et al. (Eds.), Springer Verlag, Berlin. 231–242.

Stefani, G. (1998). *Musica. Dall’esperienza alla teoria*. Milan, Italy: Ricordi.

Surges, G., & Dubnov, S. (2013). Feature selection and live query for audio oracle based machine improvisation. In *Proceedings of the 14th International Society for Music Information Retrieval Conference*. Curitiba, Brazil.

Surges, G. (2015). Generative Audio Systems: Musical Applications of Time-Varying Feedback Networks and Computational Aesthetics. UC San Diego: Music. Retrieved on June 1, 2017 from: http://escholarship.org/uc/item/9fk810hq

Suslick, K. S. (2012). Synesthesia in science and technology: more than making the unseen visible. *Current opinion in chemical biology*, 16(5), 557-563.
Tarabella, L. (2004). Handel, a Free-Hands Gesture Recognition System. In: Proceedings of the 2004 Second International Symposium Computer Music Modeling and Retrieval (CMMR 2004). Esbjerg, Denmark: Springer Berlin/Heidelberg, pp. 139–148.

Trainor, L. J., & Zatorre, R. J. (2015). The neurobiology of musical expectations from perception to emotion. The Oxford handbook of music psychology, 285-306.

Van Nort, D. (2009). Instrumental listening: sonic gesture as design principle. Organised sound, 14(02), pp. 177–187.

Vercoe, B. (1984). The Synthetic Performer in the Context of Live Performance. In Proceedings of the 1984 International Computer Music Conference (ICMC-84). Paris: International Computer Music Association, pp. 199–200.

Wallas, G. (1926). The Art of Thought. New York: Harcourt Brace.

Wiggins, G. A. (2001). Towards a more precise characterisation of creativity in AI. In Case-based reasoning: Papers from the workshop programme at ICCBR, Vol. 1, pp. 113-120.

Wiggins, G. A. (2008). Computer models of musical creativity: A review of computer models of musical creativity by David Cope. Literary and Linguistic Computing, 23(1), 109-116.

Wijnans, S. (2010). The Moving Body as a Spatial Sound Generating Instrument: Defining the Three Dimensional Data Interpreting Methodology (3DIM). PhD Thesis, Bath Spa University. Retrieved on June 1, 2017 from: http://www.mudanx.nl/PhD/2.2.5.html