Creative Autonomy in a Simple Interactive Music System

Fabio Paolizzo[a,c]* and Colin G. Johnson[b]

[a]Department of Cognitive Sciences, University of California, Irvine, Irvine, USA; [b]School of Computing, University of Kent, Canterbury, UK; [c]Department of Electronic Engineering, University of Rome Tor Vergata, Rome, Italy

*Fabio Paolizzo, PhD

Università degli Studi di Roma "Tor Vergata"

Dipartimento di Ingegneria Elettronica

Via del Politecnico, 1

Roma, IT 00133

+1 949 396 1064

fabio.paolizzo@gmail.com

Colin G. Johnson, PhD

University of Kent, School of Computing, Room S102

Canterbury, UK CT2 7NF

+44 (0)1227 82 7562

Creative Autonomy in a Simple Interactive Music System

Interactive music systems always exhibit an amount of autonomy in the creative process. The capacity to generate material that is primary, contextual and novel to the outcome is proposed here as the bare minimum for creative autonomy in these systems. Assumptions are evaluated using Video Interactive VST Orchestra, a system that generates music through sound processing in interplay with a user. The system accepts audio and video live inputs — a camera and a microphone that capture the interplay of a musician, typically. Mapping of the variance in the musician's physical motion to the sound processing allows identifying salience in the interaction and the system as autonomous. A case study is presented to provide evidence of creative autonomy in this simple, yet highly effective system.

Keywords: interactive music systems; autonomous systems; human-computer interaction; salience; computational creativity; machine improvisation

Introduction

Interactive music systems (IMSs) have both extended existing approaches to music making and introduced entirely new ways for musical creativity. Combining human-computer interaction and machine improvisation, numerous types of systems and frameworks have shaped the literature of the recent past (see Drummond, 2009). A most well-known, yet evergreen paradigm for IMSs suggests that these are defined within a continuum between their capacity to extend the musical creativity of the user — the system behaves as a musical instrument — and the resemblance of capacities which are typical of human players — the system behaves as a player; '[IMSs] are those whose behaviour changes in response to a musical input' (Rowe, 1993, p. 1). More recently, characteristics such as novelty, value and intentionality are identified as determining whether computationally creative agents can be categorized either as tools for creative support, co-creation or as fully autonomous (Ventura, 2017). The concept of musical metacreation is also interesting to this regard, as it frames a notion of authorship in music generation for systems that are 'creative on their own' (Bown et al., 2016). These definitions capture IMSs at a wide range but, interestingly, they all leverage on the concept of creativity as a function of autonomy; any IMS exhibits a capacity to operate at a creative level, which depends on how autonomous the system is.

Not all types of autonomy can allow a machine to interact as a human could do. For example, a machine musician could be producing outcome that is very coherent within its own scheme, but which does not necessarily produce anything meaningful to a human interactor. Vice versa, a machine could not be able to distinguish which information is particularly meaningful to the human agent. Although complexity in computational music creativity varies greatly, autonomy always remains a central property of any IMS. We are interested here in investigating the bare minimum of properties that defines creative autonomy in an IMS. We hypothesise that such systems can be regarded as musically creative also when they do not incorporate any musical knowledge. Instead, we propose to define them as systems which are capable of exhibiting autonomy in a musical task.

In the next section, we investigate further the concept of creative autonomy and provide reference to some of the most well-known approaches enabling this in IMSs. In section 3, we present an IMS that implements creative autonomy at a bare minimum of required features. In section 4, we propose a case study adopting that system and supporting our thesis and definition.

1. Autonomy in IMSs

In order to develop a working definition for IMSs that can also explain the phenomenon of creative autonomy rather than its mere ontology, we shall first note that definitions of autonomy based on observations of the system behaviour focus on the resemblance to the intentionality typical of a human player, as mentioned. Here, resemblance means that the system could 'fool someone into thinking it was human' and/or 'has/suggests a similar level of intentionality as a human'. This concept extends slightly a definition that we presented in the introduction (Rowe, 1992). Specifically, the concept links into debates from the computational creativity literature as to whether computers can be creative in ways that are as creative as humans (Boden, 1998; Colton, 2008), whilst clearly being computers and exercising their creativity in a way that is native to computers (Dartnall, 2013; Kantolaso & Riihiaho, 2018). Although autonomy is typically synonymous with complete independence, here we start investigating from a definition of autonomy as a form of self-determination that is not necessarily free from the influence of some external information fed to the system (Bown & Martin, 2012). From this perspective, a distinction can exist between an autonomous IMS with a seemingly random behaviour, and one that is controllable or influenceable by the user. Notably, the exercise of control does not imply that an agent is aware of being in control. This is a well-known phenomenon in cognitive sciences (Wyer Jr. & Srull, 1994; Tsakiris & De Preester, 2018). For example, in a musical interplay with an IMS, a user may perceive a system that is too autonomous or too predictable as either being an unengaging or over-improvisatory partner, regardless of the actual system properties (Bown & Martin 2012; Ornes, 2019; Yu, 2019). While true autonomy may be difficult to capture through observations,

In the present paper, we suggest that a player can maintain control over a specific set of parameters, while a system exhibits dynamical sonic behaviour; autonomy and control can coexist in a balance. This is a common feature in various IMSs, even for simple interactions ‘altering the relation the system has to itself’ (Sanfilippo, 2012; 2015), which exhibit unpredictability. However, because an IMS operates interactively rather than automatically, we expect the generation in an IMS to allow for the retention of some mutuality to the context in terms of user’s action and perception, as it occurs alongside an interplay or co-invention. The capacity to exchange information within a context and inform the artistic practice is a determining factor in human creativity. Computational creativity can also leverage on a similar capacity through the interaction with a user. In the next section, we reference IMSs that exhibit mutuality in the interplay.

Mutual listening

A well-known approach to the design of computer programs whose behaviour mimics that of a human interplayer in a musical improvisation consists of using algorithms, which monitor the improvisation and use the information gathered to generate new and contextually relevant material. A listener can understand the computer outcome, in terms of a response to a musical gesture from the human player. Early examples of this are GenJam (Biles, 1999) and MusicBlox (Gartland-Jones, 2003), which use interaction and interactive genetic algorithms to define the quality of the contextual fit between the computer-mutated musical fragment and the human performer’s contribution. The formalisation deriving from a definition of the initial population and the use of interaction rules mitigate the capacity for novelty because the (user-dependent) decision-making process is subjective and unilateral.

Multidominance

Other mutual listening works (Chadabe, 1984; Perkis, 1999; Brown & Bischoff, 2002) use the combined behaviour 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 multiple input (musical) gestures are re-interpreted into a complex musical output. Also, this approach denotes a form of shared control where the systems have autonomy in the musical tasks. 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), meaning a form of interaction in which all participants contribute primary material. As such, these systems cannot lead the musical direction of the performance because the primary generator of music material is only the human performer. Multidominance as a system property is also a trait of authenticity and authorship in autonomy, and one of the first systems capable of such style-independent response is Voyager (Lewis, 2000). Voyager carries out sonic behaviour grouping by imitating, opposing, or ignoring the performer's musical dynamic. The system then processes outcomes and reconfigures any algorithm involved in the grouping with 'no built-in hierarchy of human leader/computer follower' (Lewis, 2000, pp. 36).

Modelling knowledge

A computational music model can be achieved by segmenting music sequences in a corpus and analysing those segments for common elements of style. These elements can then be used to recombine the segments into new works (e.g., Cope, 2010; 2016), Similarly, by operating within a machine-learning scheme, music expectation can be modelled (Weng, 2010). For example, OMax learns 'in real-time by listening to an acoustic musician and extracting symbolic units from this stream. It then builds a sequence model on these units constituting an internal knowledge' (Lévy, Bloch and Assayag, 2012, p.1). This type of algorithms can navigate the model and recombine the musician's discourse, who is exposed to a form of stylistic reinjection: the system constantly confronts the player with 'a reinterpreted version of his own playing' (Lévy et al., 2012, p. 1). Other approaches adopt dictionary-based machine-learning models for the imitation of style (Dubnov, 2003; Dubnov and Surges, 2014), also capable of imposing stylistic constraints in the generation process (Pachet, 2016). These systems are highly effective their capacity to contextualize novelty. The approaches confirm the importance of mutuality in music generation also for systems that incorporate knowledge from musical data. However, the complexity of such systems exceeds the bare minimum that we seek in the present article. To this purpose, we shall recall that, simply, '[s]trong interactivity depends on instigation [by the system] and surprise [by the human performer], as well as response' (Blackwell & Young, 2005).

2. Video Interactive VST Orchestra

Video Interactive VST Orchestra (VIVO) (Paolizzo, 2013) is an IMS that was developed concurrently to theoretical research on music and interaction (Paolizzo, 2006). In a typical scenario (Figure 1), the user makes music through a sound source (i.e., a musical instrument), which VIVO receives as an audio signal for sound processing. At the same time, VIVO observes the user's movement by the means of a camera connected to the system and analyses the information to generate music. The system has similarities to VNS (Rokeby, 2010), which also implements an approach for mapping gesture/video to sound. In the present scenario, the audience can hear both the original and unprocessed sound source together with the sound that VIVO generates. In other scenarios, the information that controls the processing could derive from different types of source, such as a video or an external device connected to the system (e.g. haptic, text-based, etc.). In any scenario, the system carries out a simple analysis of the user's interaction and uses that information to control the processing of the audio signal, generating a subsequent musical output. Features differentiating VIVO from most-similar systems are described in the next sub-section.

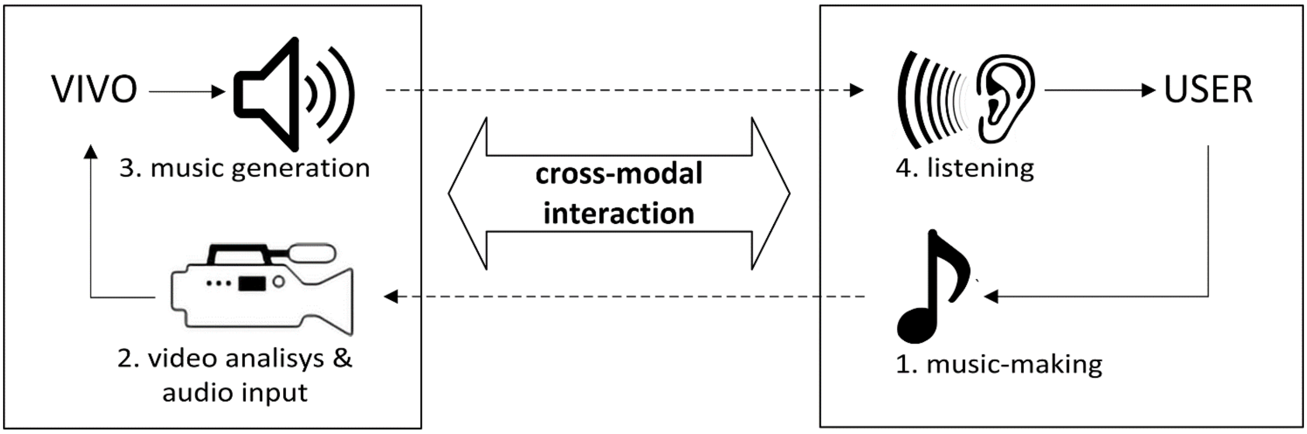

Figure 1. Typical interaction model of a VIVO/user instance.

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. The software requires an audio input for the processing and a video file or a live camera feed for the analysis. Figure 2 shows an overview of the system architecture. The user retains a configurable amount of control and an extended creation capacity that is open-ended because third-parties audio plug-ins can be loaded and mapped into the system. The system is comprised of different software components: (a) a video motion tracking module, (b) a variance-based threshold that detects relevant changes in the interaction, (c) an active-monitoring audio host that monitors and controls the plug-ins loaded into the system to reflect the user’s interaction, (d) an interactive graphical editor for stochastic scores, (e) a single graphical user interface to control the proprietary interface of each plug-in loaded, 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 variance-based threshold on the interplay, through the action of (c) the active-monitoring audio host, as it receives, stores and recalls usage data for the plug-ins loaded.

(a) motion-tracking video module

This module allows for simultaneous detection and mapping of the Quantity of Motion (QoM) — ‘an overall measure of the amount of detected motion, involving velocity and force’ (Camurri et al., 2003). QoM is measured by the number of pixels in the current frame which have changed from the previous frame. QoM is mapped as a scalar to the parameters ranges of the sound plug-ins previously defined in (e), automating the generation for those plug-ins that are currently enabled.

(b) variance-based threshold

This component continuously computes the variance of QoM and monitors its value to exceed a threshold. The threshold adapts to the mean of the variance and can be configured to different scales of sensitivity. When the threshold is exceeded, the component requests a change in the current sequence of active audio plug-ins in (e). The difference between this threshold and the variance of QoM represents the Salience of Action (SoA). Salience is a factor informing the human interplayer about the potential effects of his/her musical actions on the interplay, similarly to a ‘vested interest’ influencing the subject’s self-efficiency (Crano, 1995).

(c) active-monitoring audio host

This virtual host for audio plug-ins automates both the activation of plug-ins loaded by answering requests from (b) and parameters within user-mapped ranges from QoM values in (a). The host subtracts SoM values from an energy variable, *e*, for each plug-in when active in the current audio processing sequence, and maintains a table of the current *e* values of all the plug-ins. Upon request from (b), the plug-in with the highest *e* value is activated and the plug-in with the lowest *e* value is de-activated. *e* represents the energy of the agents in the interaction environment (Impett, 2001), as the capacity of a plug-in to join the sound generation process. Because least used plug-ins are activated and most-used ones deactivated, the approach favours novelty in the generation.

The combined use of (a) the adaptive video tracking module and (b) the audio energy host (e) affords the user with a cross-modal interaction where the movement captured by the camera determines the individual agency of multiple audio plug-ins, each having a separate memory of its overall use. SoA builds on the concept of **salience**, implemented through the simultaneous mapping of changes detected by threshold in (b) and the values assumed by QoM to the audio plug-ins in (e). As shown in our previous research (Bowman et al., 2012), a computational detection of salience within a data stream representing aspects of the interaction process can be used to manifest the potential for an interplayer to act. This occurs also in VIVO, as SoA captures an estimate of the salience in the interaction, allowing the computer outcome to reflect the users' musical interaction and the interplay to retain mutuality. Creative autonomy is sought via salient sonic changes that are synchronic to the user's interaction with the system, which in turn affords a dynamic amount of control to the user. From this perspective, SoA is an indicator that describes the user's intentionality in the action. In terms of salience, the interplay can be described as a form of communication between sign producers, where VIVO generates salient audio cues in response to a user's musical action. Notably, in this process of meaning attribution, the changes that the salience-based generation produces create an expectation for meaning to be found in the outcome, both for the user and for the audience. The automatic sound generation is sign-bearing because of its saliency.

Implications of the system architecture are discussed throughout the rest of the present article.

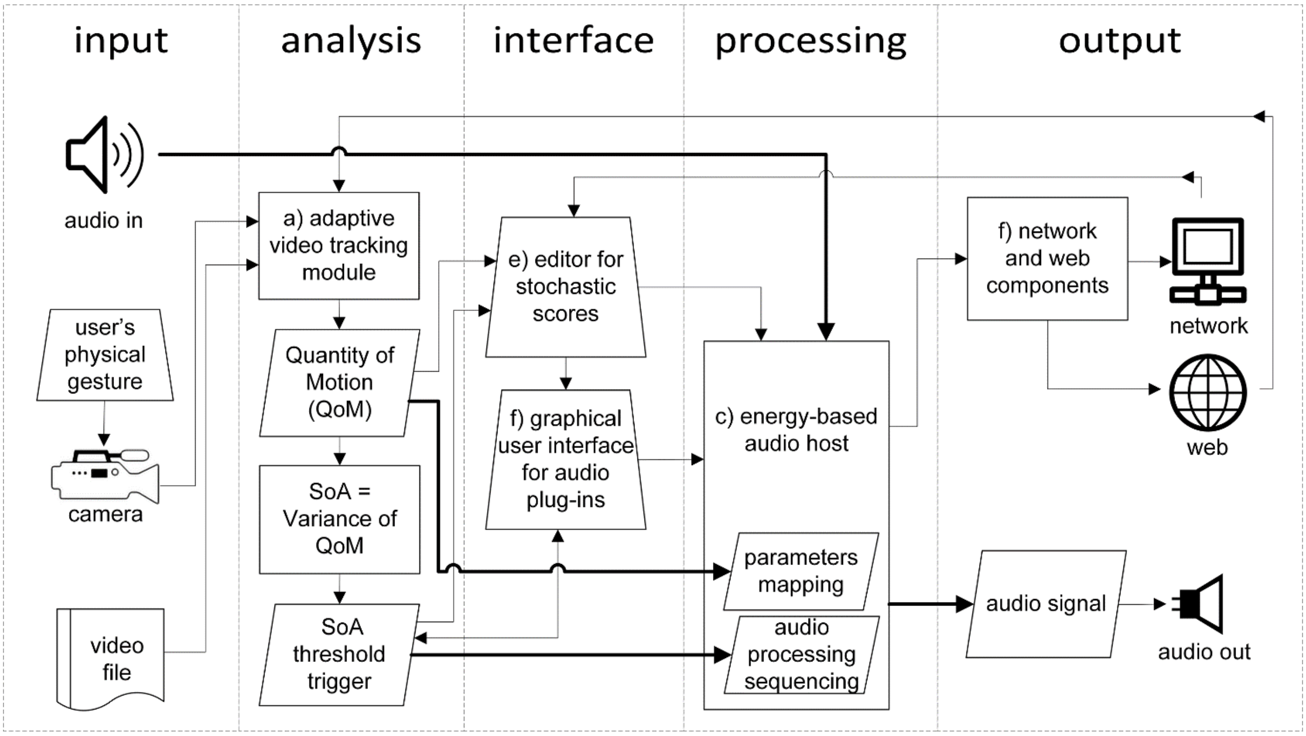

Figure 2. System architecture of VIVO. Connectors in bold provide focus to the present study.

In the typical interaction model of a VIVO/user instance (Figure 3), a user engaging in music-making also enacts gestures with a musical intention (i.e., physical gestures on a musical instrument). VIVO extracts salient information from these gestures in order to generate a sonic outcome. The user is thus caught in an action-reaction loop of self-reflection (Paolizzo, 2010), which stimulates an interpretation of the response in musical terms that includes exploration, encounter and comparison. To this regard, the experience can be assimilated to a reflexive type of interaction (Pachet, 2006). In section 2 of the present article, we have discussed some IMSs relying on mutual listening between player and instrument. Such systems operate in terms of salience, implicitly. By detecting and using salience for sound generation, an IMS can influence the musical conduct much as a human interplayer could. Salience-based systems generally derive the data for the generation from the music played by the human interplayer. Figure 1 presents this approach in a cross-modal interaction. Notably, the explicit use of salience allows drawing effectively from non-musical information which is sequentially structured (i.e., visual sequencing and motor planning but potentially also language), when such extra-musical information retains some coherence to the interplay (i.e., QoM, SoA, *e*). The detection and mapping of variance from non-musical but relevant information introduces novelty in the generation, while also retaining mutuality to the interplay.

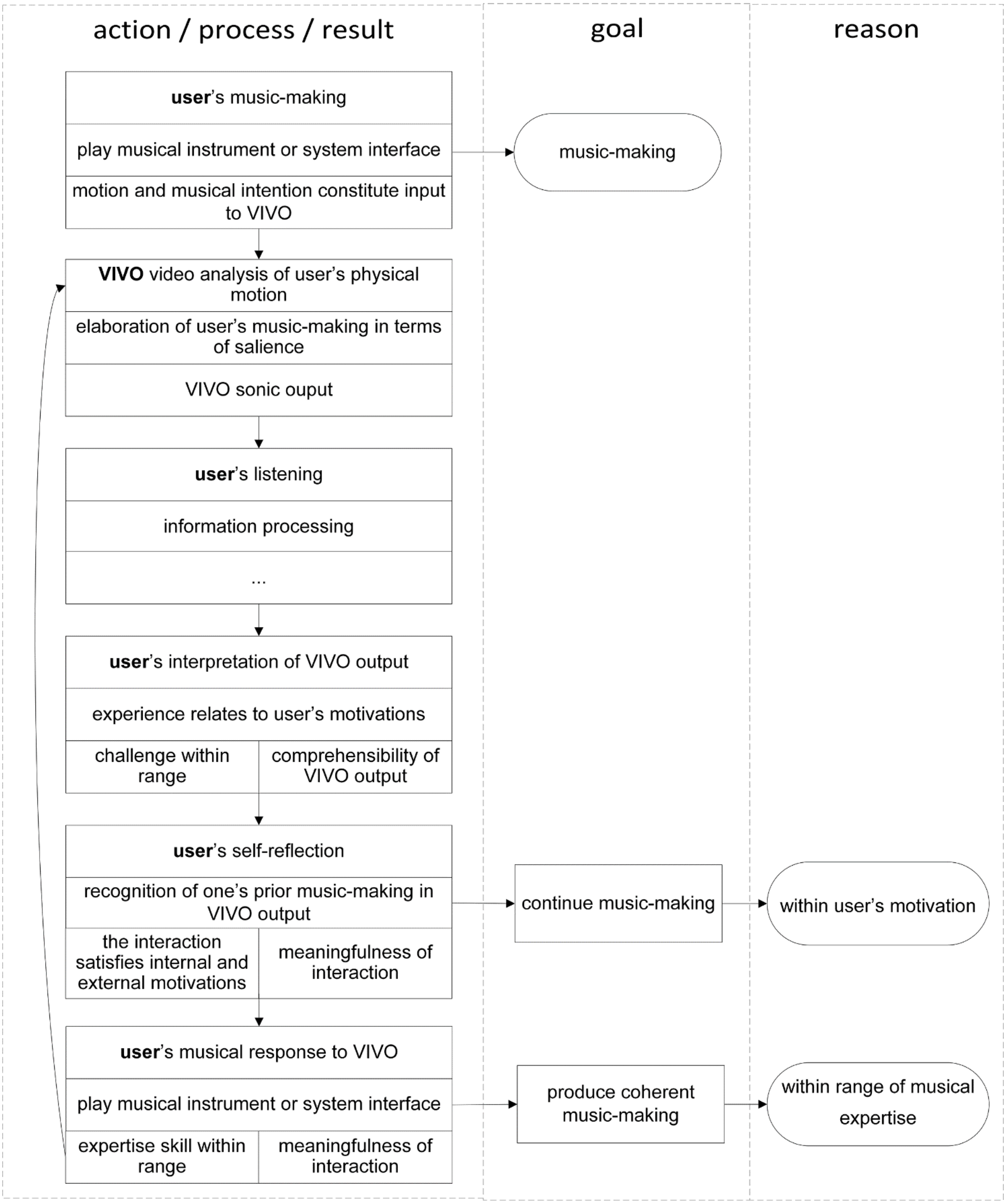
action / process / result
goal
reason
user's music-making
play musical instrument or system interface
motion and musical intention constitute input to VIVO
music-making
VIVO video analysis of user's physical motion
elaboration of user's music-making in terms of salience
VIVO sonic ouput
user's listening
information processing
...
user's interpretation of VIVO output
experience relates to user's motivations
challenge within range
comprehensibility of VIVO output
user's self-reflection
recognition of one's prior music-making in VIVO output
the interaction satisfies internal and external motivations
meaningfulness of interaction
continue music-making
within user's motivation
user's musical response to VIVO
play musical instrument or system interface
expertise skill within range
meaningfulness of interaction
produce coherent music-making
within range of musical expertise

Figure 3. Interaction diagram of a VIVO/user instance. Action, process and result are listed for each stage of the interaction.

3. Overview of pilot studies

The present research has included pilot studies in which VIVO was used for music-making within a variety of scenarios (Table 1). The purpose of these studies was to test the functionality of the system and to highlight implementation strategies that could maximise the perception of creative autonomy within an action/perception feedback loop for both the user and the audience. In order to provide a framework for the case study that we present in section 4, we introduce and discuss some of the theoretical background underlying the pilot studies. This framework incorporates the concepts already discussed in the present article, such as multidominance, mutuality and novelty.

Gestural embedding

In an acoustic instrument, the action-reaction cycle is at the basis of instrumentality and central to playing a musical instrument (Leman, 2008; 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 relation between our actions on it (action) and the sound that we hear (perception) in a process of identification-through-repetition (Emmerson, 2000). In the pilot studies on VIVO, the automatic sound generation exhibits acousmatic properties, as the audio processing forces the sources and causes of sound-making to become as ‘remote or detached from known, directly experienced physical gesture and sounding sources’ (Smalley, 1997, p. 112). Sound generation in VIVO allows designing an action/perception feedback loop for music-making that is bond to a causation mapping — a cause and effect association. In this, a salient cue by the human interplayer is used to mould the automatic generation, which can be perceived as both autonomous and contextual. The cross-modal feedback loop results in a complex and reiterated-but-changing mapping between action and sound. The system affords the user with a connection between physicality and perception, and projects sound generation to a cognitive dimension of musical expectancy. In an IMS, a simple action-reaction mapping can therefore embody a sonification process where the quality of a gesture shapes the music. As mentioned in the previous section, motor knowledge is embedded in VIVO through video-to-sound types of mapping. In this, embodiment constitutes a musical goal-directedness for the human interplayer. For the user, VIVO works as a means to the cultural embedding of gesture, which is typical in what is known as gestural surrogacy — the process of increasing remoteness. Remoteness is a form of uncertainty that can be perceived in the causality between sound sources and sonic events, for example when sources are inferred or imaged (Smalley, 1997).

In the typical scenario of Figure 1, gestural surrogacy occurs through salient, gesture-like generations, which are dependent on the user, who is also stimulated in inferring a causation in the computer-generated sound. The cross-modal nature of the feedback loop is a factor that influences multidominance because a variable amount of unpredictability affects the mapping of video information (action) to sound (generation). The user's and audience's attribution of meaning to the automatically generated sound is dependent on a causal action/perception relation suggested by the system. For the audience, this algorithmic generation is visible in the source from which the QoM is derived and computed (i.e., the video stream capturing a musician playing an instrument). For the user, the system's use of a variance-based threshold trigger allows the sound generation to change in correspondence to salient actions, ultimately increasing the coherence between the sound source and the acousmatic-like sound. Salience informs here the algorithmic generation, thereby preserving musical coherence in the interplay while also introducing gestural surrogacy.

Broadening the action/perception feedback loop

In the pilot studies, gestural surrogacy is established in VIVO when a directly mapped relation is formed between the user's gesture and the perception of the VIVO-generated outcome. Information regarding the action and the perception of non-musical processes which are relevant to the experience are used for the machine improvisation, as discussed. A camera watching the user's body and the surrounding space (as first explored in *Studio1*) or a video file (as in *VIVOtube* and *Invisible Cities*) provided such information. In both cases, non-musical information drove the automatic generation and extended the user's agency in terms of gestural surrogacy. This was achieved in different ways: (i) when instructions were sent to the machine for sound generation (as in the preparation of *VIVOtube*), (ii) 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*), (iii) through gestures on physical interfaces connected to software instruments (as in *Velodrone*), and (iv) 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 all instances, VIVO generated a simultaneous auditory feedback for each input information; QoM and SoA feedbacks referred to a user/VIVO interaction in the physical space, proprioception of users captured by a camera or visual sequencing in a video file.

Enabling self-reflection mechanisms into VIVO

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. Implementations for multidominance through the combined use of a cross-modal action/perception feedback loop and a salience/energy criterion, also afford the user with an experience that may include phenomena of reflexivity and embodiment.

In considering human cognition as embodied, VIVO was designed for facilitating the user’s perception of system autonomy through automatic sound generation recognised as music by the user. In the interacting user’s mind, this also stimulates a subjective capacity for self-reflection. Self-reflection is thus implemented by design by enabling interactions that imply rehearing, reproduction and variation. In the user’s self-reflection, both the perceived self and the perceiving self mirror each other through musical constructs that embody the agent’s activity. The term **reflection** refers here to the recursive nature of the interplay with VIVO; audio plug-ins embody an agency that depends on the user. The interpretation of an object is a process that operates multi-directionally and recursively in a semiotic/semiological feedback loop of meaning and/or sense. The process retrieves new information from new experiences and may potentially continue endlessly. In self-perception, both the perceived self and the perceiving self keep mirroring each other. Enclosed in a recursive loop of self-definition, the **I** is constituent to the same self. However, extending over the boundaries of individual reflection, the I is also the result of an interpretative process which culture incorporates. In the interaction with VIVO, the user’s expectations for meaning and sense to be found in the sound generation leverage on this process of cultural incorporation. At the same time, this leverage is possible because the user’s inner body knowledge provides a basis for the generation.

The present pilot studies suggest that self-reflection may be considered as a status of the network of interaction, which is established between VIVO and the user. Interestingly, some backing to this can be found in recent research on consciousness as a state of matter, rather than as an emerging property (Tegmark, 2015). Similarly, the reflexivity of the interaction and its character of multidominance may constitute a state of the user/system network where the capacity for meaningfulness does not emerge from an evolving process of interaction between human and software agents but rather it is enabled by system properties. We have suggested that a very limited number of properties may be needed for a system to exhibits creative autonomy and have presented the implementations.

Implementation of the dynamic mapping of QoM and SoA aims to establish multidominance (VIVO contributes primary material) through salience-based generation. This is achieved by also enabling an action/perception feedback loop between user and system (provides mutuality and contextuality to the interaction), (c) video-to-sound cross-modality (introduces unpredictability in the generation), and (d) an energy-based activation criterion in the active-monitoring host for audio plug-ins (favours novelty within the generation scenario defined), as discussed. We call reflexive multidominance the state of the system that these implementations manifest for the user.

In the next section, we evaluate our implementations as they enable a contribution of a primary material, through reflexive multidominance, which is contextual and novel to the interplay.

4. Case study: excerpt from Collective

Figure 4 depicts a transcription of the audio recording from *Collective* (Table 1), illustrating a free improvisation between a trombone player and VIVO (also see Supplemental Material for the audio video recording). 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, in order to reflect the actual playing. In the VIVO part, only salient cues are engraved, in contrast to greyed-out parts where timbre is predominant over pitch.

Legend

1) VIVO interplay
2) Threshold trigger
3) VIVO interplay
4) Trombone interplay leads
to threshold trigger

5) Threshold trigger
6) Trombone interplay
7) VIVO interplay

▨ Timbre predominant over pitch

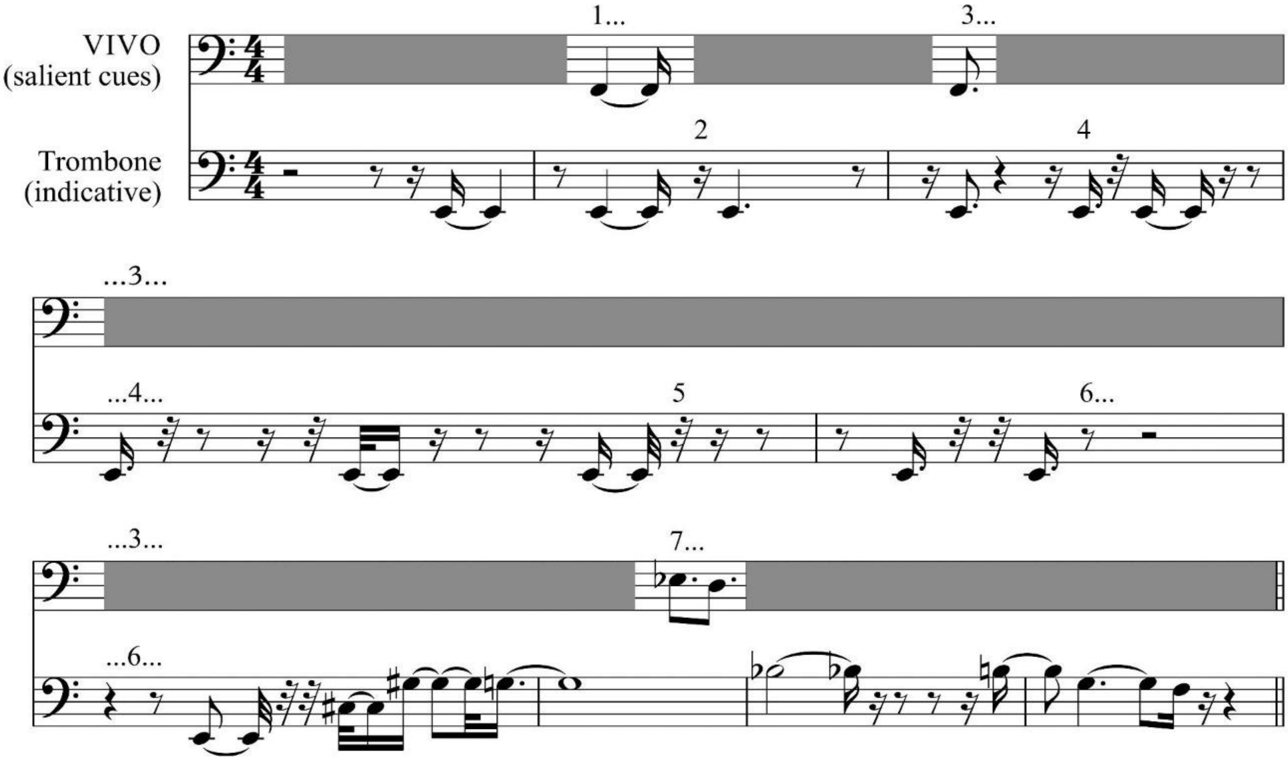

Figure 4. Excerpt of score (automatic transcription) from *Collective*.

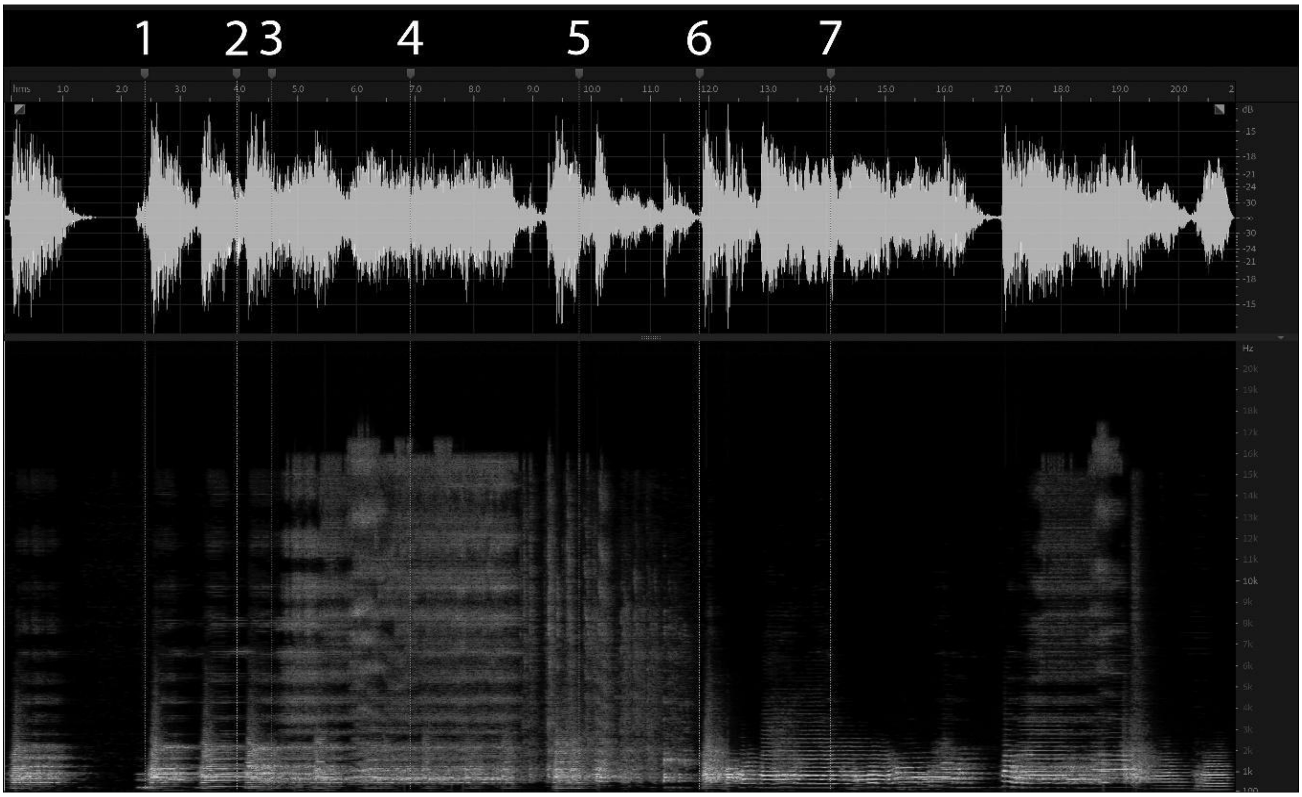

Figure 5. Excerpt of spectrogram (left + right) from *Collective*.

Here, VIVO interplays with a trombone player (Figure 4-1 and Figure 5-1) and the musician’s response results in the activation of the variance-based threshold (4-2 and 5-2). The musician recognises the consequent sound generation as an opportunity for action to achieve a meaningful interplay with the system. Creative autonomy is verified here as the musician listens to VIVO (4-3 and 5-3) and then shapes his own playing accordingly (4-4 and 5-4), thereby activating the threshold again (4-5 and 5-5). Notably, the musician’s achievement of musical phrasing after the first trigger (4-6 and 5-6) confirms the intentionality of this second trigger. VIVO’s interplay initiation depends on the musician’s playing. However, although the system denotes a certain level of autonomy, the interplay remains coherent. 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 interplay with the system, which verifies the effectiveness of **reflexive multidominance** in terms of creative autonomy. Both the musician and VIVO provide primary, novel and contextual material to the interplay, and adapt to each other (Figures 4-6 and 4-7, 5-6 and 5-7).

In the present case study, VIVO works for the musician both as an instrument and as an autonomous player. This mode of operation echoes Rowe's definition of an IMS at both extremes of the continuum proposed in that definition. As an instrument, the system extends the musician's capacity for music-making through an embodiment of the control for sound generation; VIVO is here an extension of the trombone. As a player, the system exhibits autonomy in the interplay through a dynamic mapping of SoA; VIVO is an autonomous player that exhibits creative autonomy.

Conclusions

Autonomy in IMSs is discussed here as a pivotal capacity for self-determination, yet not sufficient for a machine to be autonomously creative as a human agent. Creative autonomy has been investigated as a compound property that at least incorporates the capacity to contribute primary material, introduce novelty in the generation and retain contextuality to the interplay. We have presented VIVO, an IMS for autonomous music generation in real-time, which meets the present criterion for creative autonomy by using a simple detection of motion in a live video signal and the mapping of parameters from this to control a sound processing. In scenarios of interplay with the system, VIVO detects and uses the variance of a musician's quantity of motion and a threshold to determine sound changes that have primary influence in the music. Reflexivity and cross-modality in the experience stimulate an expectation for meaning and sense to be found in the saliency of the generation. We have described this as a form of reflexive multidominance, which the system enables through a mapping of the salience detected in a cross-modal interplay. We have provided details of a case study presenting a musical evidence. In this, the automatic generation denotes contextuality and novelty, and a musician's response that shows awareness of the system's autonomy in providing primary material. The relative simplicity of the system makes a case for reflexive multidominance as a property that enables creative autonomy in IMSs.

In a most-recent research, VIVO was used to generate the sound component of a large multimodal dataset (Paolizzo, 2019) for music emotion recognition and classification (Paolizzo et al., 2019). In future studies, we will use this dataset to investigate further the proposed concept of reflexive multidominance as a property for meaningfulness in music generation.

Funding

The work was supported by the Research Executive Agency, EC, under Grant 659434.

References

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 (AMCMM'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.

Boden, M. A. (1998) Creativity and artificial intelligence. Artificial Intelligence. Volume 103, Issues 1–2, Pages 347-356 (1998).

Bowman, H. et al. (2012). Emotions, Salience Sensitive Control of Human Attention and Computational Modelling. In: *Salience Project*. EPSRC. Retrieved on October 1, 2019 from: http://www.cs.kent.ac.uk/people/staff/hb5/attention.html

Bown, O., Eigenfeldt, A., Pasquier, P., & Dubnov, S. (2016). Special Issue on Musical Metacreation, Part II. Computers in Entertainment (CIE), 14(3), 1.

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 October 1, 2019 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.

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. AAAI spring symposium: creative intelligent systems.

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.

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

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.

Dubnov, S., Assayag, G., Lartillot, O., & Bejerano, G. (2003). Using machine-learning methods for musical style modeling. Computer, 36(10), 73-80.

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

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.

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.

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.

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

Ornes, S. (2019) Science and Culture: Computers take art in new directions, challenging the meaning of creativity. Proceedings of the National Academy of Sciences. https://doi.org/10.1073/pnas.1900883116

Pachet, F., & Sony-CSL, P. (2006). Reflective interactions: from “it’s fun” to flow machines. Acts of pluridisciplinary musical recitations, Lyon, Grame.

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

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

Paolizzo, F. (2010). 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. (2013). VIVO: a Wakeful Instrument for Collective Musical Embodiment. Ph.D. Thesis, University of Kent.

Paolizzo F., Pichierri, N., Casali, D., Giardino, D., Matta M., Costantini G. (2019). Multilabel Automated Recognition of Emotions Induced Through Music. arXiv:1905.12629 [cs.SD].

Paolizzo, F. (2019). Musical-Moods: multimodal dataset. University of Rome Tor Vergata, University of California Irvine. At: https://github.com/fabiopaolizzo/musical-moods

Perkis, T. (1999). The Hub, an Article Written for Electronic Musician Magazine. Retrieved on October 1, 2019 from: http://www.perkis.com/wpc/w_hubem.html

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

Rokeby, D. (2010). Very Nervous System (1986-1990). David Rokeby-Artist. Retrieved on October 1, 2019 from: http://www.davidrokeby.com/vns.html

Rowe, R. (1992). Interactive Music Systems: Machine Listening and Composing. Cambridge, MA: The MIT Press.

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 October 1, 2019 from: http://dariosanfilippo.tumblr.com/

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

Tsakiris, M., De Preester, H. (2018). The Interoceptive Mind: From Homeostasis to Awareness. Oxford University Press

Ventura, D. (2017). How to Build a CC System Paper type: System Description Paper. ICCC.

Wyer, R., Jr., (Ed.), Srull, T. (Ed.). (1994). Handbook of Social Cognition. New York: Psychology Press. https://doi.org/10.4324/9781315807102

Yu, G. (2019) Effects of timing on users’ perceived control when interacting with intelligent systems (Doctoral Thesis). https://doi.org/10.17863/CAM.37907