ERC Advanced Grant 2019

Part B2

REACH: Raising co-creativity in cyber-human musicianship

Section a. State-of-the-art and objectives

Presentation

Digital societies will see the development of deep interweaving between human creativity and autonomous information-processing capabilities of virtual and physical objects, that turn joint human-machine activity into a mixed reality with symbiotic interactions. In the cultural, artistic, educative or economical domains, co-creativity between humans and machines will bring about the emergence of distributed information structures of a new kind, as in networked ambient intelligence or in artistic performances of mixed artificial and human agents, that will disrupt known cultural orders and profoundly impact human development. With technologies allowing one to extract semantic features from physical (e.g. acoustic, light) and human (e.g. physiological, gesture, language) signals, combined with novel generative learning of symbolic representations, we are beginning to unveil the increasing complexity of cooperation, synergy or conflicts inherent to cyber-human networks. These coming hybrid micro-societies need renewed conceptual and technological frames in order to be fully understood and managed. To this end ***the REACH project aims at understanding, modeling, and developing musical co-creativity between humans and machines through improvised interactions, allowing musicians of any level of training to develop their skills and expand their individual and social creative potential.*** Indeed, improvisation is at the very heart of all human interactions, and music is a fertile ground for developing models and tools of creativity that can be generalized to other activities, as in music the constraints are among the strongest to conduct cooperative behaviors that come together into highly integrated courses of actions. *REACH will study shared musicianship occurring at the intersection of the physical, human and digital spheres as an archetype of distributed intelligence, and will produce powerful models and tools as vehicles to better understand and foster human creativity in a context where it becomes more and more intertwined with computation.*

Positioning

The psychologist Margaret Boden has given much attention to the many relations between creativity and machines [Boden, 1997]. For her, creativity is the ability to find new, surprising and socially valuable ideas or artifacts, and can occur in three main ways: it can be combinatorial (new configurations of known materials), exploratory (discovering new paths in conceptual / mental spaces) or transformative (when the space itself is disrupted giving way to ideas that were properly inconceivable before). But what is the situation when part of the creativity is delegated to machines, when manifestations of co-creativity emerge from symbolic interactions between human and artificial agents? In addition to the novelty / effectiveness criteria, cyber-human co-creativity is strongly felt when two features of improvisation linked to emergence [Canonne & Garnier 2011] and non-linear dynamics [Mouawad & Dubnov 2017] are identified: (1) *emergence of cohesive behaviors that are not reducible to, nor explainable by the mere individual processes of agents*; (2) *apparition of non-linear regimes of structure formation, leading to rich musical co-evolution of forms*. Our assumption is that such surging phenomena result from *cross learning processes between agents involving complex feed-backs loops and reinforcement mechanisms,* and we will strive in our project to create the optimal conditions for them to occur. As a major consequence, the states and behavior of participating agents are in return modified continuously, making them evolve in terms of knowledge and skills. This vision of computational co-creativity foreseen by REACH is a pioneering one with respect to state-of-the-art work, including recent heavy-weight research program such as Google Magenta, Sony Flow Machines, MIT Media-Lab Opera of the future, or Spotify Creator Technology Lab (see e.g. [Roberts & al 2018], [Hadjeres & al 2017]), who are generally centered on style learning and do not address directly the co-action and emergence problems.

A major originality of this project is to constitute the first attempt to build a scientific and technological mesh of knowledges and tools around computational co-creativity and musical improvisation through a highly interdisciplinary approach combining computational sciences of music, artificial intelligence, anthropology of improvised knowledges and practices, and mixed reality. It will be the first time that the notion of improvisation is brought as a general scheme of cyber-human communication targeting creative behaviors, and that methods for augmenting human capacities and for enhancing computational creativity are brought together to model symbiotic interaction. The REACH project will provide theoretical foundations, technologies, software suites and experimental data and make them available in open-access to all the research community, and to other communities as well (e.g. artists, instrument makers, game developer, medias).

REACH co-creative instruments will bring about the long-awaited convergence between physicality, information processing and creativity, in a manner that will be maneuverable and enjoyable by all, at all levels of skills and expertise. Such progress will be likely to disrupt artistic and social routines, as well as amateur and professional music practices, eventually impacting music industry in a game-changing way. As an illustration, in 2014, amateur and professional guitarists spent more than $4.25 billion on buying musical gear. However, it has become very difficult to retain new buyers (90% of guitar buyers stop playing in the first year, according to Fender). By adapting to individual skills, levels and musical tastes, tools resulting from REACH research will increase the intensity and pleasure of creative use, contributing to human development and attaching users to artistic creativity in a constantly renewed esthetic experience.

REACH will explore and foster cyber-human co-creative strategies, based on the integration of artificial listening, interactive learning, structure discovery, mixed reality and social science assessment. It will provide a new ground for understanding symbiotic interaction between the digital, the physical and the human world, with deep repercussions on the unfolding of individual and collective creativity.

State of the art

Formal tools

As a precursor to most recent studies on computational creativity, interactive musical generativity has been addressed through such topics as algorithms and computer architectures [Blackwell 2012], dynamics of interactions seen as a complex system [Canonne & al 2011] or synchronization and matching of audio descriptors between live play and stored model [Moreira 2013]. As a particular case of generative systems, improvisation systems do need for their part an intelligent listening component. Form analysis, a major topic in Music Information Retrieval (MIR) [Doras & al 2019], has proved useful for that matter and has been studied with methods from different fields. [Giraud et al. 2015] used dynamic programming to perform fugue analysis by detecting the occurrences of each motif. A similar method was used for sonata form analysis by [Bigo et al. 2017]. [Bimbot et al. 2016] introduced the System & Contrast model where the general structure of a musical segment is divided recursively in contrasting morphological elements living on a geometrical model of polytope. Most of these approaches, however, are not causal and do not cope with the necessity of discovering structures on-line and in real-time, as expert improvisers do. ***In order to attain improvised co-creativity in live interaction, REACH will focus on on-line discovery of musical structures and interactive learning processes involving cross feed-back loops between artificial and human agents.***

Sequence models

Following Feder's seminal work on optimal coding and universal predictors [Feder 1996] the PI has shown in collaboration with S. Dubnov that context models (that assign probabilities to events under the condition of past event sequences) can become adaptive and support Variable Markovian Order by using an optimal compressor to encode sequences [Assayag & al 1999], [Dubnov & al 2003]. Unfortunately, this optimality is only asymptotical with increasing sequence length. The PI has then discovered the musical properties of a model imported from formal languages theory and bioinformatics, called the Factor Oracle (FO) [Crochemore & al 2007]. The FO is an on-line algorithm, linear in time and space, delivering a time-oriented graph with a "cartography" of arbitrary long Markovian contexts, overcoming the optimal compressor method limitation and well adapted to on-line learning and generation of new sequences [Assayag & al 2004, Assayag & al 2010, Maniatakos, Assayag & al 2012]. More recently, The PI has supervised a series of PhD research works, combining FO with a Bayesian probabilistic framework and graphical models [Deguernel, Vincent, Assayag 2018, 2019], or with scenario planification and pulse-based improvisation [Nika & al 2017], where his students were awarded the "Young Researcher Prize in Science and Music" in 2015 and the "French Association of Computer Music Young Researcher Prize" 2016. Since the PI's pioneering work, studies on the modelling of formal sequences for the apprehension of musical styles and improvised creative interaction has developed significantly in the SMC (Sound and Music Computing) communities. Realistic situations of interaction (stage, studio, video game) involving severe constraints of reactive performance have directed research towards certain types of problems and solutions. Statistical sequence models of music have shown their ability to capture surface features of musical styles. [Conklin 2003] has summarized in a general survey the techniques for generating music from statistical models including various computationally efficient Markovian systems such as n-grams, that can easily be learned from examples. A fusion of statistical and formal models (under the form of logical constraints) has been attempted by [Pachet & Roy 2011] and pursued in the ERC Flow Machine project [Hadjeres & al 2017].

The PI's theoretical work on formal sequences models has led him to the invention of the OMax Improvisation software, with long time collaborators M. Chemillier, S. Dubnov, and G. Bloch [Assayag & al 2004, 2006]. OMax (with its siblings SoMax, ImproteK and Djazz, see [Nika & al 2017]) is now an internationally recognized reference in the field of cyber-human improvisation [Blackwell 2012] with a wide dissemination, including more than 100 concerts, workshops and sessions all around the world, involving world-class musicians such as Steve Lehman (#1 Jazz Album of the year by The New York Times) or Mike Garson (historical Pianist of David Bowie since the *Aladdin Sane* LP). OMax interaction architecture and its learning and generation scheme [Assayag, Bloch 2007] enable on-the-fly acquisition during performance time without a priori knowledge [Lévy, Bloch, Assayag 2012]. OMax allows as well adaptive synchronization between musician and digital agents with respect to the melodic, harmonic, textural and rhythmic dimensions using chroma descriptors and pulse phase alignment [Bonnasse-Gahot 2014]. OMax has been a game changer and is now a familiar paradigm in machine improvisation by promoting the *stylistic reinjection* aesthetic model [Assayag & al 2006], and even coining a new expression "omaxiser" (French for "to omaxize"). However, despite its powerful improvising skills, OMax "listens" to the humans only to learn, *not to decide what to play,* and thus cannot achieve co-creativity in the sense that we have explained above. ***REACH will implement co-creativity as a dynamic adaptation process by parallelizing and fully orchestrating the listening, learning, and generating processes through all phases and at all levels of the musical development.***

Music Information Dynamics

Music Information Dynamics (MID) [Dubnov 2010], [Potter, Wiggins and Pearce 2007], [Abdallah and Plumbley 2009] studies evolution in information contents of music assumed to be related to structures captured by cognitive processes such as formation, validation and violation of musical expectations [Meyer 1956]. In particular, a MID measure called Information Rate (IR) was proposed by Dubnov in relation to human judgements of emotional force and familiarity [Dubnov, McAdams and Reynolds 2006]. IR measures the reduction in uncertainty about a signal when past information is taken into account. It is formally defined as the information-theoretic mutual information between past and present of a signal. In [Dubnov, Assayag, and Cont 2011] a new formulation of IR, using Audio Oracle [Dubnov, Assayag, Cont 2007], was developed using the difference in compressed coding length obtained by knowing the past. This allows the automatic selection of the best audio descriptors (feature selection) and of their distance thresholds so as to maximize the information in the FO and to drastically improve its performance in music generation [Surges and Dubnov 2013]. The PI and colleagues were able to propose a framework for anticipatory machine improvisation and style imitation using an early version of the Audio Oracle with three main modules: memory, guides and learning [Cont, Dubnov & Assayag 2007]. The guides are reinforcement signals from the sound environment to the system, or from previous instances of the system onto itself, feeding the reinforcement learning [Sutton & Barto 2018] module. This was the first work applying active learning with reinforcement in the music domain. The difficulty, however, was in the interpretation of the reinforcement signals, a usual hurdle in music modeling, linked to the difficulty in defining meaningful cost functions (what is a "good" aesthetic decision or artistic behavior?)

A new MID structure designed by Dubnov, the VMO (Variable Markov Oracle) has since brought interesting results. It was shown that motif detection using VMO corresponds well with human expert identification of shapes in music [Wang and Dubnov 2015] and outperforms existing algorithms for musical pattern detection. Wang, Dubnov and Mouawad have been successful in combining VMO with methods from nonlinear dynamics in order to reduce a multivariate time series down to a flat symbolic sequence while detecting recurrences of sub-sequences of variable length. Such recurrences were then depicted by a symbolic recurrence plot to obtain recurrence quantification analysis (RQA) measures, thus clearing the path to high-level musical feature discovery (including affects) in complex multivariate signals [Mouawad and Dubnov, 2016, 2017].

On a historical note, as a precursor of MID, G.D. Birkhoff had already formalized the notion of aesthetic in terms of the relations between order and complexity [Birkhoff 1933], later formulated in terms of information theory by A. Moles [Moles 1958] and M. Bense [Bense 1969], where complexity is measured in terms of entropy, and order is the difference between uncompressed and compressed representations of the data. In the above depiction of the PI and colleagues' works in MID, the past is used for compression, and so the method resembles to the computational aesthetics approach in that it provides a graph of time-varying relative entropy, thus allowing the use of IR as a "local aesthetics" measure for the choice of future materials in composition and improvisation [Dubnov, Assayag 2012]. MID thus simulates the cognitive dynamics at work in the process of listening, enjoying, and anticipating music, *albeit for a single listener*. As a precursory work, Rabinovitch & al have analyzed *joint cognition* of human-robot pairs by modeling sequential transient dynamics of several modalities, including working and episodic memory, emotion and attentional inhibitory control [Rabinovich and Varona, 2016]. One of their results indicate that the Kolmogorov-Sinai entropy is a useful feature for characterizing the level of joint human-robot initiative. ***In REACH, we will be the first to address the question of joint cognitive dynamics of co-creative improvisation unfolding in collectives of human and artificial and using mixed reality instruments.***

Creative physicality

On the physical embodiment side, the "Smart Instrument" resulting from the Ircam IMAREV research project (PI: A. Mamou-Mani), commercialized by the HyVibe sound tech company, was the first of its kind to bring embedded computing, mechatronics and physical modelling straight to the fingers of the musician [Meurisse & al 2014], [Benacchio & al 2012]. To this effect, this technology combines sensors, actuators, embedded algorithms, control interface, and a printed circuit board designed to be the fastest sound card in the world with a one milli-second reactivity. The super low latency system permits to control feedback and transform the guitar's natural vibrations from the soundboard via the bridge where the instrument performs best [Benacchio & al 2016]. Active acoustics experiments have indeed been carried out in several labs such as the ExCITe Center at Drexel University [McPherson 2010] or the EPFL acoustics lab [Lissek & al 2018], and of course at Ircam in the IMAREV project (with experiments on augmented violin, clarinet, string quartet). However, these results stop short of uncovering the deep relation between generative creativity and physical embodiment now permitted by active control. ***By bridging the gap between Smart Instruments with active acoustic control on the physical side, and digital intelligence for co-creative interaction on the logical side, REACH will make possible the advent of*** **Creative Instruments*****: physical instruments loaded both with active control and creative agents.***

Social science assessment

It is impossible to understand the dynamics of creativity in improvisation, whether the latter be considered as a general marker of freedom and will in human behavior, or as a technical and aesthetical issue in music, without summoning the anthropologist's gaze on emerging cyber-human musical cultures. On the social science side, the types of improvisation considered here belong to one of the two categories defined in [Bailey 1999] as idiomatic and non-idiomatic, which roughly correspond to regular pulse-based improvisation and free improvisation. The term "idiom" used by Bailey highlights the existence of a cultural context and a community of people sharing this idiom as they could share a language. This raises the question of considering a particular social and cultural context in the evaluation of improvisation agents, as well as evaluating the integration of digital or mixed reality devices such as the Creative Instrument in live performances — possibly modifying the artist's and audience's relationships and disrupting their perceptual and emotional patterns. The study of technology in musical performances lies at the crossroad between performance studies [Auslander 1999] and sound studies [Pinch & Trocco 2002] as in Butler's in-depth work on DJ and laptop live improvisation practices [Butler 2014] mainly devoted to the specific context of electronic dance music although this style of music serves as an exemplary instance of trends that are present to varying degrees in almost all music today.

On a more conventional music instrument side, improvisation has been studied in the reference works on jazz interaction [Monson 1996, Berliner 1994]. Concerning pulse-based improvisation, a more specific problem arose from the studies on micro-timing, and there exists an active research field dealing with the study of rhythmic interaction between musicians in the context of jazz improvisation. The notion of participatory discrepancies introduced in [Keil 1995] has paved the way to a study of timing in jazz at a micro level. The pioneering analysis developed in [Cholakis 1995] has been followed by researches involving measurement of timing in musical performances [Doffman 2008] leading to the introduction of a model of entrainment that describes rhythmic interaction between musicians as the coupling of autonomous oscillators [Clayton 2012], an idea also explored in the field of neurosciences by [Large 1994, 2002]. Rhythm and meter detection has raised important works in human and even animal cognition [Patel & al 2008], [Fitch 2013]. This subject is also connected to technological issues through the particular domain of automatic score following [Echeveste et al. 2013] which is directly related to the PI's work on the improvisation software family OMax. All these studies tend to configure a contemporary and urban "anthropology of knowledge" [Chemillier 2018] applied to improvisational practices, specially when they introduce companionship with machines. ***The REACH project will be the first of its kind to join interdisciplinary researches in computational music science with anthropological assessment going both ways: by providing human data to feed the computational models, and by inseminating the human field with computational systems in order to observe cyber-human interactions using social science methods.***

In order to approach music improvisation strategies, it is also useful to look at what has been done in the realm of conversational interactions modeling. Undertaken by linguistics researchers as early as in the 1970s, a number of studies have focused on linguistic and paralinguistic information conveyed in the improvised interaction of agents, whether it consists of speaking [Thórisson 2002], transmission of para-verbal cues and their contextualization [Allwood 1976] or simple signals such as head nods during a conversation [Ishi & al 2014]. These signals follow a certain dynamic evolution [Thórisson 2002] and are very much dependent on the context and on previous interactions in order to be decoded correctly or even only perceived. An example often used is that of irony because in this case information is transmitted through prosody that invert the meaning transmitted by words [Fujie & al 2004]. Non-verbal behaviors have been modeled using artificial agents [Kopp 2010] showing that a greater richness of interaction with an artificial system leads to better engagement and therefore greater user efficiency, which translates into higher success rate for the interaction and lengthening of its duration [Novielli & al 2010]. Interaction is no longer seen as a mere exchange of signals, but rather as an adaptive construction, using many modalities over different time scales [Knapp 2013]. This is for example the case with the Talking Heads system [Steels 2003], where agents manage to construct a representation of the world through a game of language (that could be assimilated to free improvisation) by identifying and naming the objects in their field of vision. ***The REACH project will design creative agents who build up their own representation of the world and continuously adapt their pace to their perceptions and beliefs, in dialog with other agents, through multiple interacting modalities and musical dimensions, rather than fulfilling arbitrary rules.***

Freedom and Idiom

In this non-verbal communication between agents, improvisation intrinsically carries the notions of spontaneity and reactivity. It is not, however, naturally antagonistic to those of planning and organization, specially in music, as long highlighted in the cognitive studies on this subject [Pressing 1984]. This is the case for jazz grids, or the written bass in baroque improvisation, or Indian raga prescriptions that constitute formalized sequences of constraints. This conditioning is found in some interactive music systems with a structure organizing improvisation over the long term. In this category, the concept of "controlled improvisation" [Donze & al 2013] proposes a process for generating a sequence of control events guided by a reference sequence and satisfying a formal specification through a graph representation. [Surges & Dubnov 2013] allow one to write scripts determining the evolution of the audio descriptors to be taken into account in the analysis of the input signal, and the evolution of the generation parameters as well. After works on the automatization of jazz harmonic substitutions in blues progressions [Steedman 1984, Chemillier 2004], and in the continuity of their research on navigation inside a FO cartography capturing the sequential logic of a live improviser [Assayag 2004, 2007], the PI and his team have introduced both formal long-term "scenarios" and reactive conditioning from live environment, in order to provide richer guides to improvisation [Nika & al 2012, 2017a, 2017b], [Bonnasse-Gahot 2014].

In REACH, co-creative dynamics of improvised interaction will draw on non-verbal communication sttrategies, temporal adaptation of interaction, high sensitivity and reactivity to the live environment, and formal structures of scenarios articulating short-term and long-term anticipation.

Artificial Intelligence

Deep Learning methods such as restricted Boltzmann machines (RBM) [Hinton 2007], auto-encoders and variational auto-encoders (VAE) [Kingma 2014, 2019], convolutional networks [Humphrey13], recurrent networks [Bengio & al 2012], have disrupted research practices in music data mining, classification and information retrieval and opened new perspectives in generative algorithms. In computational music science, there was already a considerable background in symbolic AI and optimization, such as constraint programming, local search, genetic algorithms, pattern matching, operational research, geometrical and algebraic models of music spaces, see [Andreatta 2013], [Herremans, Chuan & Chew 2017], [Carpentier, Assayag, Saint-James 2009], [Assayag, Truchet 2011]. Progresses in deep learning of music structures are now coming step after step. The DeepMind WaveNet architecture [van der Oord et al. 2016] generates raw audio waveforms through a convolutional feedforward network, with applications mainly in speech generation, although it has been tested on piano pieces. MidiNet is a variant working on symbolic music data (MIDI) and featuring a generative adversarial network (GAN) [Yang & al 2017]. In the PI's team, the composer Daniele Ghisi has used sampleRNN method [Mehri & al, 2016] to create a totally autonomous "ghost" female voice for "La fabrique des monstres", a 5 voices vocal piece inspired by Mary Shelley's Frankenstein and used in Jean François Peyret's eponymous play created on major French stages [Crestel & al 2018]. SampleRNN features auto-regressive multi-layer perceptrons and recurrent neural networks in a hierarchical structure able to capture small to medium scale variations in the underlying sound data from sound samples to notes and motives. DeepBach [Hadjeres & al 2017] uses two recurrent networks (Long Short-Term Memory), one summing up past information and another summing up information coming from the future. Magenta's MusicVAE [Roberts & al 2018] is a hierarchical variational autoencoder encapsulating recurrent networks for learning latent spaces of musical scores, showing interesting sampling, interpolation, and reconstruction performance.

A new task force in the PI's Team has been recently created (ACIDS: Artificial Creative Intelligence and Data Science, lead researcher P. Esling) that has developed several first-of-kind methods for interactive, user-centric computational creativity based on state-of-the art machine learning. For example, the live orchestral piano system [Crestel & Esling 2017] generates on the fly complex orchestrations from a live keyboard play. It is based on a conditional RBM trained on a corpus of piano reductions of emblematic symphonic repertoire. The Variational Timbre Space is a VAE on audio instrument samples, providing for the first time a latent space conditioned on timbre perception data and allowing the creation and control of novel instrument sounds [Esling, Chemla, Bitton 2018]. One important problem highlighted by these recent researches in machine learning is that different modalities of a same musical object lead to different data representations, but still carry a common semantic content that has to be discovered. One of the central aims in REACH will thus be to uncover semantic dimensions where different modalities co-exist, by computing lower-dimension embedding spaces that directly provide interactive transformation and generation.

However, most of these state-of the-art models lack understandability, user-centric view, and incrementability as would be needed for creative interaction. ***REACH research will address signals and structures that appear only through co-creative interaction and will offer an unprecedented framework to study the interoperation between acoustic signal, gestural embodiment, instrument mechatronics, symbolic musical information and user-centric perceptual and anthropological data, through deep AI methods.***

Objectives

The research objective of this inter-disciplinary project is to model and enhance co-creativity as it arises in improvised musical interactions between human and artificial agents, in a spectrum of practices spanning from interacting with software agents to mixed reality involving instrumental physicality and embodiment. Such creative interaction strongly involves co-improvisation, as a mixture of more or less predictable events, reactive and planned behaviors, discovery and action phases, states of volition or idleness. Improvisation is thus at the core of this project and indeed a fundamental constituent of co-creative musicianship, as well as a fascinating anthropological lever to human interactions in general.

The outline of the project unfolds as follows:

- **Understanding**, modelling, implementing music generativity and improvised interaction as a general template for *symbiotic interaction between humans and digital systems* (cyber-human systems)
- **Creating** the scientific and technological conditions for *mixed reality musical systems*, based on the interrelation of creative agents and active control in physical systems.
- **Achieving** distributed co-creativity through complex *temporal adaptation of creative agents* in live cyber-human systems, articulated to field experiment in *musical social sciences*.

These three directions are already innovative and ambitious in their own right. Setting up a synergy between them, in order to foster understanding of co-creativity between natural and artificial beings on a socially significant scale, is an entirely novel initiative. It is the first organized interdisciplinary attempt at deciphering and shaping cyber-human co-creativity, an emerging field of study obviously called to considerable developments in the digital era. Revealing the mechanisms of co-creativity in music will also be a remarkable projector on creativity in general, as music is one of the most highly organized, interactive and complex human activity — at the same time an abstract, sensitive and physical one, yet profoundly shaped by communication and improvisation, and as such a powerful metaphor of human creative interactions.

We will put in place a research ecosystem allowing us to answer two central questions:

- *How to* ***augment the digital*** *agents capacities* with enhanced computational creativity as well as with cyber-physical extensions, so *they can develop convincing interactions with humans*
- How to ***augment the human*** *capacities* by expanding their individual and social creative potential, through novel collaborative strategies and mixed reality, so *they can naturally immerse themselves in complex schemes involving digital intelligence.*

In REACH, co-creative interactions between humans and machines will be studied dynamically from different perspectives (music sciences, social sciences, computer science) in order to highlight the combined conditions under which these interactions occur, their temporal behavior, adaptive dynamics and control strategies, so as to drastically enhance their creative potential.

We emphasize the fact that shared *creativity is an emerging process*, resulting from complex interactions and multimodal, cross feed-back loops between natural and artificial actors, a process that cannot be reduced to an agent's production considered in isolation, or to a single layer of technical skill. *This conception neutralizes the endless philosophical question of whether artificial entities can be qualified as "creative"* and safely shifts the REACH research interest towards the positive exploration of the best possible models and tools to allow for co-creative cycles to bootstrap and settle.

We also believe that "computational creativity" [Wiggins 2006, Colton and Wiggins 2012] will socially take root in a decisive way when it eventually creates the conditions of emergence of cyber-human co-creativity, and when it eventually clings to an augmented physical reality allowing humans to experience a rewarding embodied relation. To this effect *REACH will work towards bridging the gap between abstract computational logics and mixed reality experience anchored in creative musical instruments.* In such configurations of physical *interreality* (a mixed reality where the physical world is actively modified by human action), the human subjects participating to co-creative scenarios will be directly engaged in the digital, the physical and the social dimensions of the experience. Such a seamless integration of cyber to physical to human will be the next frontier in Creative Artificial Intelligence [Boden 1998, Briot & al 2019] and REACH aims at being a powerful instrument towards such progress. ***REACH will implement interreality through Creative Instruments: music instruments that will embed an unprecedented concentration of digital musicianship, making them full partners of human in co-creative scenarios.***

Section b. Methodology

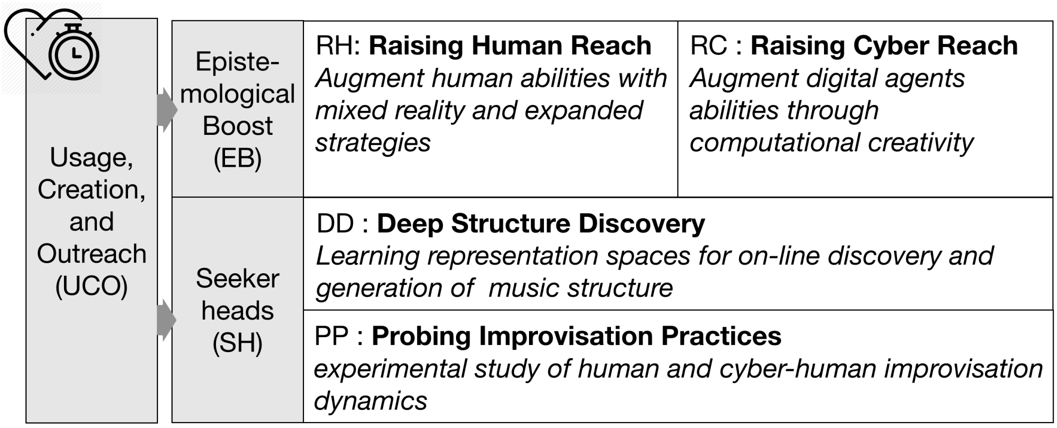
Research Strategy

The fundamental working hypothesis behind REACH is that one has to operate in two symmetrical and converging directions in order to *simultaneously raise* (thus the name of the project) the capacity of human and cyber agents, so that they meet at the point where *synergistic effects may happen and make way for co-creative behavior*. The two central tasks (1) and (2) of the project will follow this scheme:

1. **Raising Cyber Reach** (RC): Augment the digital agents' abilities with enhanced computational creativity, autonomy and multimodal sensitivity to context, as well as with cyber-physical extensions in mixed reality instruments, so they can enter into a convincing expert interplay with humans.
2. **Raising Human Reach** (RH): Augment the human abilities, by interfacing them into cross-modal, embodied mixed reality so they can interact in a natural and creative way with artificial agents, and by expanding their reach through novel collaborative strategies infused by social sciences field experiments.

In order to provide solid foundations and tools to sustain this *double motion from Cyber to Human and from Human to Cyber* (1) and (2), we will surround it with two domains of action (3) and (4):

3. **Deep Structure Discovery** (DD): Machine Learning of complex multi-variate signals involving *structure discovery and knowledge construction* in musical signals
4. **Probing Improvisation Practices** (PP): Field research providing human (or combined cyber-human) data, experimental assessment and heuristic guides for model parametrization through an *anthropology of improvisation practices, knowledges and processes*.



RH and RC form the **Epistemological Boost** (EB) component of our research strategy: by symmetrically augmenting humans and artificial agents until they come together at their co-creative nexus by producing emergent information structures, we will achieve an epistemological leap [Bachelard 1934] beyond the blocking idea that artificial systems cannot be creative by themselves. This turn things around in stating that creativity is not a state anyway, but rather a dynamical effect of interaction in a complex system, likely to show that *radical novelty* already identified as a characteristic of Emergence [Corning 2002]. *By building on this epistemological boost, REACH will be able model deep interactions that in turn will trigger co-creative behaviors.*

DD and PP form the **Seeker Heads** (SH): a cluster of exploratory researches in music structure discovery and improvisation shaping, that will sustain and feed the EB components by providing general learning models, creation heuristics, behavioral examples, and experimental protocols and data.

These two component EB and SH will be articulated by a **Usage, Creation, and Outreach** (UCO) package acting as a *project clock*, or a heart-beat pulsating the timing of a series of iteration between modelling, prototyping, experimenting, and creation / outreach. Going through fast research and development cycles favoring interaction between all work packages, with regular milestones and deliverables, it will smooth-out the interdisciplinary cooperation and prevent unforeseen problems that could potentially seize up the organization of the project. UCO will clock at several nested periodicities (bimonthly, semester, year, biennial) to schedule coordination meetings, delivery checking, scientific seminaries, local and international workshops and dissemination events detailed in the work plan below. *Although REACH is high-risk high-gain, this heart-beat of animation, stimulation and coordination will provide a high level of confidence and security in smoothing the pace of research and production interaction, inside the PI's team and with external collaborators as well.*

Inside this work program, we will use methods drawn mostly from (1) Interactive computational creativity, (2) mixed reality augmentation of music instruments (3) machine learning and musical information dynamics, and (4) social sciences and anthropology of improvised practices. *Such a high level of inter-disciplinarity in the REACH project is realistic since the PI has a unique background in heading for six years a world-renowned laboratory encompassing all sciences of music and sound, where he has successfully organized wide-ranging, multi-domain research strategies.* The PI will nevertheless surround his team with a network of expert researchers or economical actors, from which he will seek one-off expertise in well-defined perimeters, at various stages of the project, notably in the fields of Musical Information Dynamics and cognition, Social Science / Anthropology, Augmented reality and Acoustics. These experts, with whom the PI already has a rich and thorough collaboration history, are listed in the detailed research plan below for each action where they will be associated.

Interactive Computational creativity and mixed reality will be at the core of the Epistemological Boost (EB). Improvised interaction between humans and digital agents [Blackwell 2012] is dependent on mastering artificial creativity, which brings together several key research issues: knowledge, perception and intention modeling, real-time interactive learning — featuring cross feed-backs between human and digital entities that continuously modify the context (moving target)—, time structure formalization. Mixed reality will be instantiated by Creative Instruments reassembling active acoustics and creative agents. ***By going beyond state-of-the-art cyber-physical systems (that create a continuity between the digital and the physical worlds), REACH will promote cyber-human systems that create a continuity between digital and human creativity.***

Machine Learning and Musical Information Dynamics of complex multi-variate signals such as polyphonic musical signals, involving structure discovery and knowledge assessment, will be essential to understand and process (often in real-time) the variety of signals exchanged by humans and creative agents. These signals bear implicit multi-dimensional and multi-scale semantic structures that are hard to tract computationally, but that are nonetheless necessary for machine perception and learning to work properly. Representation learning of these structures, through artificial intelligence methods such as statistical learning and deep learning will be at the core of the Seeker Heads (SH). Musical Information Dynamics (MID) characterize semantically important changes in information contents of audio signals [Abdallah & Plumbey 2009]. REACH will extend state-of-the-art MID towards longer term correlations and attention control, in order to address the question of intention and decision over larger scale structures and bigger time spans. ***This will bring a better understanding of joint cognition and creativity and will feed the design of creative agents. It will also help us to assess that co-creativity is actually boosted by (EB), by checking information measures correlated to variations in surprise, interest, flow, affects, and analyzing their tendencies over different types and scales of interactions.***

Social Science, anthropological approaches will be used to observe, analyze, archive and understand improvisational practices and knowledges between humans, as well as between humans and machines. Data from these studies will help guide the conception of the perceptual, generative and interaction models in EB and DD. This work at the interface of technology, artistic practices and urban anthropology, will proceed on live performance involving different modes of human-machine interaction, in diverse cultural background. It will use the methods of the field survey in the usual sense of ethnomusicology: performance data collecting, audio-visual and gestural capture of live context of musical sessions, interviews with participants (musicians, audience), sessions of re-listening, analysis of the musical contents archived by the softwares. A series of questions will be examined: How do the musicians react to the computer actions? Does it influence the way human musicians play themselves? Do they perceive computer productions as intentional, on-purpose? Do they perceive computer interaction as realistic in the sense that human could produce it, or rather as un-human, thus potentially bending towards totally unheard artistic forms and aesthetics? ***This will bring understanding of joint cognition and action from the human science angle and will complement and correlate the "objective" MID assessment of co-creativity emergence, by "subjective" survey of human behavior and discourses in creative situations.***

Team, host and collaborative network

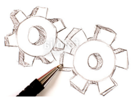
Host Institution (HI) IRCAM, founded by famous composer and conductor Pierre Boulez, is the largest institution in the world gathering scientific research, technology engineering, contemporary music production, concert season and teaching. It offers absolutely unique conditions for REACH researches that seat at the crossing of several scientific, aesthetic, social and technological domains. The hosting Lab STMS (Sciences and technologies of music and sound, a joint operation by IRCAM, CNRS and Sorbonne University) inside IRCAM is world leader in Sound and Music Computing. The hosting team in STMS, called Music Representation, is headed by the PI and internationally renowned in Musical Informatics and Creative Interaction. The REACH team itself will be formed by the PI (40% of FTE) plus three postdocs (PoD), three PhD students and a research assistant. The couples (PoD1, PhD1) will focus on the EB cluster, (PoD2, PhD2) on the PP task, and (PoD3, PhD3) on the DD task (see details in section C on the Web form, and in the schedule below).

The hosting lab will provide a great amount of punctual expertise as needed in the different tasks of the project (specified below in the task breakdown). Pr. Philippe Esling is an expert in Creative Artificial Intelligence and the founder of the ACIDS research group (Artificial Creative Intelligence and Data Science). The S3AM team (acoustics, mechatronics, signal processing, and physical modelling, head: T. Hélie), ISMM team (embodiment, gesture, interaction, head: F. Bevilacqua), and PDS (Cognition, Perception, Computational Neurosciences, head: P. Susini) are recognized leaders in their domain and will provide invaluable punctual expertise in these fields. The IRCAM Music Production and Teaching Department, used to manage international seasons of artistic productions, will greatly support the PI for experimental sessions and artistic outreach.

Regular international collaborators of the PI in Europe, USA and Japan, will provide punctual expertise or service delivery where needed: the award winning Hyvibe high tech sound company, on prototyping and experimenting the Creative Instruments; UCSD Lab CREL (Center for Research in Entertainment and Learning, (Dir. Shlomo Dubnov, a recognized leader in Musical Information Dynamics) on machine learning, artificial listening and edutainment applications ; EPFL Meta Media Center / Montreux Jazz Heritage Lab (Resp. Alain Dufaux) will give access to huge amounts of annotated live musical data through the Heritage Montreux Jazz Festival Archive, a registered UNESCO Immaterial Memory of the World Heritage; EHESS CAMS Lab (Ecole des hautes Etudes en Sciences Sociales, Centre d'Analyse et de Mathématique Sociales, PI: Marc Chemillier) will provide us with social science / anthropological expertise for the PP task. The PI also maintain continuous collaboration and PhD students exchange with several AI labs at Tokyo University (Pr. Harada Cyber Physical Systems Lab, Pr. Hirose Cyber Laboratory) and Tokyo Metropolitan University (Pr. Ikei Virtual Reality and Ultra Reality Lab) who will provide expertise on cyber-physical systems, mixed reality and deep learning theory and practice.

REACH is a disruptive proposal as it addresses the future shape of creative human and digital intelligence workship. However, it is underpinned by a large amount of leading research and results from the PI and his team, thus limiting the risk and laying foundations for conceptual advances, new technologies, societal changes and mentoring and transmitting to students. This work will also be greatly facilitated by the fact that the PI has long been used to collaborate with professional and amateur musicians, to participate with them to artistic performances, workshops, master-classes or studio experimental sessions. All the experiments will be based on volunteering and exchange, and the musicians will have the opportunity to use for themselves freely and indefinitely all the software tools and instruments resulting from the project. When particular situations involving copyrights arise (such as using performance captures for a documentary, or a record) specific agreements will be signed among parties.

Detailed Work Plan



Task **Raising Human Reach** (RH). Augment human abilities for efficient improvised behavior through multimodal & affective interfaces, embodied interaction in mixed reality and novel collaboration strategies.

RH.1 ***From individuals to collective***

We will develop an adaptive, distributed system of creative co-improvised interaction between collectives of human and artificial agents. This involves modeling, identifying and controlling short-term occurrences (reactions, synchronization) as well as longer-term ones (vocabulary, patterns, structures), referring to various aspects of expressivity and style, to a variety of cognitive states and processes related to memory, listening, recognition, intention, initiative and learning. Current models of computational creativity focus on the ability to generate novel content and do not generally consider the rich interplays and collaborations that emerge in live collective setups. Instead, we will aim at actual co-creative partnerships, focusing on shared improvisation dynamics [Assayag, 2016], by evaluating a variety of collaboration strategies such as dialogue, enrichment, convergence and divergence [Sawyer 2000]. To evaluate the success of these novel strategies, we will rely on the analysis of musical information dynamics (MID) by exploiting predictive relationships between cognitive interpretations of musical structures and statistical properties of audio-musical signals [Wang & Dubnov 2015]. In a distributed co-creative system, multimodal cross feedbacks affect collective behavior and increase player engagement through emotion and motivation. This in turn feeds critical information to the learning systems and alter in some way the learning targets [Ikei & al 2016]. We will study the effect of emotional and motivational dynamics on collaborative creativity in order to design multimodal / sympathetic interaction. These advances will help creative agents to enhance their collective coordination by modeling their beliefs, emotions and motivations on other agents' behavior.

RH.2 ***Creative Instruments for Human Augmentation***

Recent studies in embodied interaction revealed that body scheme modification via virtual and augmented reality affects our perception and behavior, as an augmented body tends to overcome usual physical limitations of real bodies [Ogawa & al 2016], and it enhances our perception, performance and creative skills as well [Ban & al 2013]. We will realize augmented interaction in mixed musical reality by the inception of Creative Instruments, augmented instruments loaded with creative agents from task RC. Creative Instruments will be designed on top of Smart Instruments based on active acoustics control. Multiple sensors (e.g. hexaphonic for the guitar, optic for winds or piano) [Benacchio & al 2016], will be added in order to finely capture human expression. New actuators will be added in order to augment the fusion between the natural acoustic regime and the augmented reality productions. A wireless connection will enable the transfer of signals and data between the instrument and remote servers for data collection. The software and hardware integration of creative agents from RC will be started by using generic single boards computers (e.g. Raspberry Pi) to validate the coupling between low-level digital vibration processing, digital audio algorithms and creative AI algorithms, inserted in the physical instrument with control interface.

The Creative Instrument will load the musical instrument with *Machine Musicianship* [Rowe 2004]: a digital intelligence that will make the instrument become autonomous and play along freely with the user, through its authentic acoustic properties and physical interaction, indeed installing a mixed reality where the users' body and mind will feel augmented by the instrument unlimited capacity to expand their intentions in the harmonic, melodic, rhythmic or orchestration dimensions. The addition of creative algorithms (co-improviser, harmonizer, orchestrator, creative looper etc.) will fundamentally alter the musician's perception of his instrument and of his own body interacting with it, thus switching-on the feeling of human augmentation and turning on the co-creative experience. This task will be in constant interaction with task PP in order to study the human side of this cyber-human symbiosis and to collect experimental data on the original performance situations initiated by these technologies.

RH Punctual expert collaborations: *A. Mamou Mani (Hyvibe CEO), Pr. Y. Ikei (Head Virtual Reality and Ultra Reality Lab at Tokyo Metropolitan University) will provide expertise and feed-back on mixed reality and human augmentation. IRCAM teams ISMM, S31M and PDS will provide expertise on embodied interfaces, active acoustics and music cognition.*

Task **Raising Cyber Reach** (RC). Augment computational creativity of AI agents in order to raise their awareness, their autonomy and their inventiveness

RC.1 ***Reaching creativity in perception, decision, and agency***

Listening to music and generating improvised responses require to deal with prediction uncertainty and decision making at multiple time scales. We will build models for enhancing agents' anticipation and decision skills through attentional and motivational dynamics. We will build on our previous results on musical information dynamics and information geometry as initiated by [Cont, Dubnov & Assayag 2011], operating directly on audio signals through expectation statistics and Bayesian graphical models for belief propagation in multi-agent joint decision making [Deguernel & al 2018]. We will extend our works notably by enlarging the temporal and structural horizon of musical expectation and decision to multivariable time-series using the Variable Markov Oracle [Wang &Dubnov 2015] and by exploring computational models of self, belief and intentionality [Assayag 2016].

Musical improvisation requires an ability to listen and quickly apprehend the musical structure correctly. These cognitive processes involve long-term memorization and structure discovery on multiple time scales. We will apprehend these structures by modeling the global temporal shapes and boundaries using Information Rate [Dubnov & Assayag 2011], based on accurate understanding of acoustic events [Tokozume & al 2018], and relating these to perceptual properties in deep embedding spaces as proposed by [Esling & al 2018]. Modeling improvised interactions also involves multiple perception modalities, with expectancy playing a central role. However, these concepts are difficult to tract with machine learning models. We will address this problem by turning improvisation principles into optimization objectives. To do so, we will gather models of improvisation behaviors in social sciences from task PP and explore their relations to statistical properties and information dynamics analysis [Assayag 2010]. Then, we will consider these principles as optimization objectives in order to train an adaptive and anticipatory system of interaction as an encoder of the uncertainty of multiple futures in human behaviour, as proposed by [Kanehira & al 2019] in the video domain.

These tools will help creative artificial agents to equip themselves with minimal "cognitive" structures in order to be able to listen, learn and interact with other human or artificial agents in highly noisy environment. In particular, these agents will be able to show a certain level of reflexivity on their own behavior and display different forms of creative memory at different scales and type of activations (procedural, reflex, episodic, semantic) so as to simulate the qualities of awakening (curiosity triggered by a stimulus), attention (listening to other agents), motivation (whether or not you want to learn) and initiative (decide whether to play or not).

RC.2 ***Temporal Adaptation in advanced interaction strategy***

In interaction contexts involving a variety of agents, such as developed in RH.1, cyber agents should estimate the capacity and intentionality of humans (or other agents) through schemes of temporal adaptation of collective interaction, in order to match their expectations and contribute to the emergence of a cohesive behavior over time [Sanlaville, Assayag, & al 2015]. We will extend to collective improvisation the methods from nonlinear dynamics such as proposed for affects and auditory scenes by [Mouawad and Dubnov, 2016, 2017], by using a variable Markov oracle (VMO) that reduces a multivariate time series down to a symbolic sequence while detecting recurrences of sub-sequences of variable length [Wang and Dubnov, 2015] and evaluating symbolic recurrence quantification analysis (RQA). We will explore computational models of emergent synchrony in social interactions through discrete network models involving empathy/dominance dualities [Knapp 2013]. Then, we will define general mechanisms for dynamic adaptation and synchronization of agent behaviors in co-creative interaction [Nika & al 2017]. Another approach will be to learn joint embedding spaces linking symbolic, acoustic, performance and perceptual sources of musical information in highly adaptive interactive contexts by defining adequate information distances to minimize between semantically similar vectors in latent embedding spaces jointly learned with variational inference as pioneered by [Esling et al. 2018], developing in our case new hierarchical temporal models that allow to ensure that the organization of these latent spaces follow a temporally-coherent organization.

RC Punctual expert collaborations: *long-time collaborator Pr Shlomo Dubnov (Head CREL Lab, UCSD) is a world leading expert in Music Information Dynamics and Machine Listening ; he will shed light on information theoretic issues.*

Task **Probing Improvisation Practices** (PP). Anthropological, social and usage study of human and cyber-human improvisation dynamics.

PP.1 ***Human impact of technological artifacts in live performances***

This task deals with the general problem of live performances integrating digital artifacts, now taking unprecedented proportions with the presence on stage of increasingly sophisticated devices (collaborative robotics, holograms, virtual reality, network performances, etc.) competing or cooperating with humans according to different modes of human-machine interaction [Chemillier 2016]. The contribution of this transformed reality to the performing arts modifies the artist's and audience's relationships and disrupts the observer's perceptual and emotional patterns (the extreme case of fans' attachment to the hologram singer Hatsune Miku). Musicians are challenged by the confrontation with such devices that push the limits of imagination and virtuosity and in turn stimulates their creativity, opening the way to the exploration of new frontiers of musical idioms in which improvisation is taking place. It is a major anthropological and aesthetic challenge to question the presence of these machineries on stage, understand how the human cope with them, and discover new humanistic ways of using them in order to enrich the experience of improvisation in diverse cultural backgrounds. An important part of this task will be to carry out: urban-ethnology enquiries in situations where this human-machine co-creativity is likely to happen; anthropological studies on knowledges and practices relative to improvisation seen as confrontation to the unknown; field surveys on the evolution of performing arts codes and manners brought about by technological artifacts. Some music such as jazz and oral tradition music are more likely than others to develop a strong participative relationship with the public, based on the ritual context in which they were born, and new form of music performances (installation, immersion, public interaction) are emerging. A particular attention will be paid to improvised interactions that transgress the boundaries between artist and audience. Micro-sociological analysis such as symbolic interactionism [Gibson 2010] applied to the relations between musicians and sociological survey of the response of the public to the creativity of the machine will be pursued for that matter.

PP.2 ***Rhythmic synchronization of humans and machines***

Adjusting the tempo of the machine with that of the musicians is a difficult task requiring thorough experimentation in order to collect empirical data, as well as development of measurement and interaction tools featuring innovative ergonomics. Research conducted in music cognition on group interaction in collective improvisation [Canonne & Garnier 2012] shows that there is no single tempo of reference, but as many tempos as musicians who are subject to some sort of negotiation and adaptive resynchronization. In the long term, the challenge is to obtain fair acceptability of the machine's rhythmic productions in relation to the requirements of phrase, flexibility and temporal precision (that are characteristic e.g. of the subtle art of swing developed by jazzmen or by the Banda or Aka Pygmies in their polyphonies and polyrhythms [Arom 1991]). This task combines empirical studies on these synchronization processes with research on formalization and ergonomics solutions to integrate such strategies in relation with RH and RC. For example, the synchronization with human partners can be done by combining tempo detection algorithms such as proposed by Edward Large [Large 2002] along captured manual pulse for correction. Adjusting the tempo of the machine with that of the other musicians is then like a car clutch that adjusts two different speeds of rotation, and it is essential, from the point of view of augmented improvisation, that the person controlling the computer may experience the rhythm feeling the same way the other partners do, through proper ergonomics. In case of problem, experienced musicians know how to use strategies to play specific patterns when there is a flutter in the tempo (drum roll to temporize, note held on the bass to interrupt the walking bass, etc.) in order to achieve a resynchronization of the band. Similar strategies will be developed for computational agents.

PP.3 ***Scenario based improvisation induced from large corpuses***

One aspect of improvisation environments is to condition improvised musical phrases, recombined from bits of a corpus that has been learned off-line, by imposed harmonic chord progression, or otherwise strict compositional scenario [Nika 2017]. The a priori musical knowledge on the corpus allows for efficient preprocessing and indexing methods to help match patterns with extreme efficiency between the scenario structure (e.g. harmonic grid) and the improvisation space (e.g. melodic phrase), thus permitting real-time computational creativity that still respects the compositional constraints. For example, there are different strategies in the way jazzmen follow harmony (respecting harmonic contour or favoring melodic logic at the risk of generating dissonances with the underlying chords). Field surveys on the strategies employed by human experts of different cultures, in traditional as well as in cyber-human concert situations, will be carried in order to understand how musicians of diverse cultures negotiate this dialectic of freedom and rule. This task will help collect data, formalize improvisation rules, and propose guidelines and specifications for creative agents design in tasks RC and DD.

RC Punctual expert collaborations: *long-time collaborator Pr Marc Chemillier (CAMS Lab, EHESS) is a recognized anthropologist of oral knowledges with a focus on improvised practices including machines. He will provide expertise on field experiments and behaviors in idiomatic and non-idiomatic improvisation.*

Task **Deep Structure Discovery** (DD): Learning representation spaces for on-line discovery and generation of deep music structure

DD.1 ***Deep Structure Explorer***

It is recognized that structures in music exist in or emerge from compositional apparatus, from live performance developments and from audience perception, and that comprehending them fully would involve coordinated formal and cognitive studies of objective productions and states of mind in these three instances [Chew 2017] [Smith, Shankler & Chew 2014]. In REACH we concentrate on live co-creative situation of free or structured improvisation, for which well such adapted tools for music structure exploration are rare. In order to make things tractable, we will draw a compromise between long term structure discovery and short-term musical surface recognition. We will search for constantly moving short-term and medium-term predictions of structure development, that will serve as a guide to improvising agents. This task will account for cognitive processes involved in extraction, memorization and creation of phrase structures involved in global and local shaping of improvisation. We will have to address the estimation of temporal shapes emerging outside known idiomatic contexts (e.g. jazz, pop), or constituting variations or deformations of known idioms in uncertain performance context. To this effect, information measures such as information rate or correlation density [Dubnov, Assayag 2011] will be exploited to find structural boundaries in variable Markov models. We will assess on-line learning methods such as long term / short term memory models, in order to introduce temporal sensibility to the past in predicting the next structural units, in combination with language models, such as word sequence selection [Deguernel, Vincent, Assayag 2019], to account for structural segmentation boundaries potentially linked to changes in harmony, melody, timbre, rhythm or energy [Smith & al 2014], and for emergent grammar systems organizing the sequences. We will use supervised / unsupervised machine learning tools such as cross-modal variational autoencoders [Chemla-Romeu-Santos & al 2019], recurrent networks [Briot & al 2019] and thematic memory parallel activation [Bonnasse-Gahot 2014] in order to learn joint signal/symbolic representation spaces where structures can be tracked or generated dynamically. Results of this task will help in the design of creative agents in task CR.

DD.2 ***Deep Style Shaping***

Previous works of the PI and colleagues summarized in a book of reference on the Structure of Style [Argamon & al 2010] have shown the importance of style modelling in the creation of coherent sequential music material and in the shaping of musical form. However, the problem with this approach is the relative lack of originality of generating agents, as their criteria are mostly based on conformity. Jean-Pierre Briot in his thorough study on deep learning for music has evoked the idea that original style creation should be accounted for as much as coherence and conformity in style models [Briot & al 2019]. Up to now, deep learning algorithms are still fragile in determining long-term structure and capturing higher-level content such as form, emotions and style in music [Herremans, Chuan and Chew2017]. Or they restrict themselves to exploring creativity on a single dimension e.g. melody [Yang & al 2017]. We will renew these approaches by extending the idea of Generative Adversarial Networks (GAN) [Goodfellow & al 2014], which have already proven creative and originals in the image domain, in two directions. First, style will be evaluated by the discriminator network by looking at conformism, ambiguity or disruption in the productions of the generator. This can be compared to what has been experienced in visual arts with Creative Adversarial Networks (CAN) by maximizing deviation from established styles [El Gammal & al 2017]. On a historical note, this generator / discriminator strategy had been prophesized by Abraham Moles in his 1971 book *Art and the Computer* [Moles 1971] years before the popularization of neural networks. Second, a novel concept of Multidimensional On-line Adversarial Network (MOAN) will be developed in order to account simultaneously for the recurrent and interactive construction of Music —by conditioning generation both on past elements and on simultaneous input from other agents, and for the multi-dimensionality of music — by training the network on realistic timed, polyphonic and harmonically labeled data. This task will be strongly articulated to tasks RH, for evaluating style emergence in collective coordination of agent, to RC for loading operational models of creative agents with style capacities, and PP for anthropological assessment of stylistic behaviors.

DD.3 ***Deep Creative Strategies***

The limited availability of high-quality labeled audio dataset could limit the scope of machine learning algorithms. However, this drawback can be overcome by data expansion, between-classes learning and semi-supervised generative techniques capable of learning unsupervised latent representations of data. Such progress has been realized in the visual medias, notably at the Harada Lab in Tokyo University [Tokozume & al 2018], one of the main nodes the PI's network of collaborations. We will explore these techniques' potential for sound and musical interaction and adapt them accordingly. We will study dimension raising inspired by automatic 2D to 3D image mapping [Kato & al 2018], by renewing drastically the usual musicological hypotheses on music structure. We will consider that structure in music is an n-dimensional complex ($n > 2$) that is living in a hidden space behind the 1-dimensional audio signal, just as 3D depth and hidden faces are implicitly hidden behind the 2D image. This amounts to say that sound structures of higher dimensions and richer information content have to be induced first from lower dimension audio signal, an upheaval from the classical view according to which the signal is an information-wise expansion of the structure (or correlatively, the structure or semantic information is a compression of the signal) [Dubnov, Assayag 2012]. This innovative approach will unleash generative creativity as there is an infinity of structures that can map to a particular signal, and an infinity of transformations that can easily be performed on them. By creating embedding spaces in which the structures are represented, we will also be able to directly generate articulate improvisations by simple geometric operations such as translations, rotations, deformations etc., or, in the case of stochastic manifolds such as VAE latent spaces, by moving along geodesics [Chemla-Romeu-Santos & al, 2019]. This task will feed task RC with fundamental advances in generative methods for creative agents.

<u>RC Punctual expert collaborations:</u> Pr. T. Harada (Harada Lab, Tokyo University) is a leading researcher in Machine Learning for Multimedia. We will consult him on deep learning methods for sound signal to music structure induction.

Task **Usage, Creation, and Outreach** (UCO): control iterations between research, creation, and outreach

UCO is the heart-beat of REACH, defining the rhythm of iteration between modelling, prototyping experimentation, creation and outreach actions. UCO will smooth out interactions between tasks, stimulate production of deliverables at a steady rate, and provide continuous control and sensible decision handles to the PI and his team in order to process any emerging issue at a very early stage. UCO will be conducted in constant collaboration with expert musicians, including world class performers, thanks to HI IRCAM Music Creation and Production Dpt. support, and to the EPFL Metamedia department who keeps the Montreux Jazz Festival audiovisual archive and organizes yearly workshops with the musicians at the festival. They provide us with an invaluable opportunity to meet musicians and disseminate through master-classes and workshops. The models delivered by research packages will be tested, refined and validated during experimentation with these musicians in studio and public concerts, using audio recording and video capture, instrument sensing, critical re-listening sessions, musician interviews and debriefing.

UCO will be the center task for organizing *<u>an annual workshop in IRCAM</u>* where all the international collaborators and experts consulted will meet and share views on the evolution of the project, with an open day to share results and prospects with the research community at large; *<u>every two years a large scale international event</u>* will be organized in the style of the Improtech Workshop – Festival co-created by the PI (see ikparisathina.ircam.fr), where scientific and artistic ideas will confront each other. An internal coordination and scientific seminary will be held with the team every two months, and every semester a check will be performed on delivery schedule with re-coordination meetings as needed.

<u>Participation of the PI's team to major conference</u> in the domain (including SMC, ICMC, ISMIR, MCM, ICMPC, TENOR, CMMR, ACM Multimedia, AES, ICCC, NIME, MOCO, CHI, ICSC, PSN, SNT) will be organized. *<u>The PI and his team will regularly publish in journals</u>* including the Computer Music Journal, J. of New Music Research, J. of Mathematics and Music, J. of Music and Science, Int. J. of Music Science, Technology and Art, Int. J. of Music and Performing Arts, Critical Studies in Improvisation, ACM/IEEE Trans. on Multimedia, ACM Computers in entertainment, J. of the AES, Ethnomusicology, L'Homme, Anthropologie et Sociétés, Terrain, Psychology of Aesthetics, Creativity, and the Arts, Music Perception, Digital Creativity, Artificial Intelligence, Minds and Machines, IEEE Trans. on Pattern Analysis and Machine Intelligence, IEEE Computational Intelligence Magazine, ACM Trans. on Intelligent Systems and Technology. Milestone papers will be published in open-access journals s.a. PLOS One, Cultural Anthropology, J. of Artificial Intelligence Research, Digital Humanities Quarterly or Music Theory Online. *<u>At least one international journal special issues</u>* will be co-edited by the PI and the postdocs; at the end of the five years, *<u>a book will be proposed</u>* synthesizing the results of the project.

Work Schedule

R: report due; P: algorithm version / prototype; U: UCO delivery check pulse; M: annual workshops; I: biennial international workshops

Work Plan	Year 1		Year 2		Year 3		Year 4		Year 5		Post-doc			PhD		
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	1	2	3	1	2	3
Raising Human Reach (RH)																
RH.1 *From individuals to collective*		R			R		R				X					
RH.2 *Creative instruments for human augmentation*				R		P		P			X					
Raising Cyber Reach (RC).																
RC.1 *Reaching creativity in perception, decision, and agency*				R		P		P		R	X		X	X		
RC.2 *Temporal Adaptation in advanced interaction strategy*		R		R		P		R			X			X		
Probing Improvisation Practices (PP).																
PP.1 *Human impact of technological artifacts in live performances*		R		R		R		R		R		X			X	
PP.2 *Rhythmic synchronization of humans and machines*		R		R		P						X				
PP.3 *Scenario based improvisation induced from large corpus*												X				
Deep Structure Discovery (DD):																
DD.1 *Deep Structure Explorer*		R		R		P	R						X			
DD.2 *Deep Style Shaping*				R		P		R					X			
DD.3 *Deep Creative Strategies*						R		P		R			X			X
Usage, Creation, and Outreach (UCO)																
Experimentation, validation, user feed-back		U	U	U	U	U	U	U	U	U	X	X		X	X	
Workshops (M)eeting, (I)nternational		M	I	M		M	I	M		M	X	X	X	X	X	X

References

Abdallah S.A. and M. Plumbley, D., Information dynamics: patterns of expectation and surprise in the perception of music, Connection Science, Vol. 21, No. 2. (2009), pp. 89-117.

Allombert, A., Desainte-Catherine, M., **Assayag**, G., « Towards an hybrid temporal paradigm for musical composition and performance », Computer Music Journal, 2012, vol. 4, n° 36

Allombert, A., Desainte-Catherine, M., and **Assayag**, G. Iscore: a system for writing interaction. Proc.3rd international Conference on Digital interactive Media in Entertainment and Arts, DIMEA '08, vol. 349, 360-367. ACM, New York, 2008

Allombert A., Desainte-Catherine M., Larralde J. and **Assayag** G., «A system of Interactive Scores based on qualitative and quantitative temporal constraints», Proc. of the 4th International Conference on Digital Arts (Artech 2008), Porto, 2008.

Allwood, J., Linguistic communication as action and cooperation. Gothenburg monographs in linguistics, 1976, vol. 2, p. 637-663.

Agon, C., Bresson, J., **Assayag**, G., OpenMusic: Design and Implementation Aspects of a Visual Programming Language », 1st European Lisp Symposium - ELS'08, Bordeaux, 2008

Allombert, A., **Assayag**, G., Desainte-Catherine, M., « A system of Interactive Scores based on Petri Nets », SMC'07 - 4th Sound and Music Computing Conference, Lefkada, pp. 158-165, 2007

Andreatta M., Musique algorithmique, In N. Donin et L. Feneyrou (eds), Théorie de la composition musicale au XXe siècle, Symétrie, Paris, 2013

Arom, S., African Polyphony and Polyrhythm: Musical Structure and Methodology, Cambridge University Press, 1991 - 668 pages.

Argamon, S., Burns ,K., and Dubnov, S., (eds), "The Structure of Style: Algorithmic Approaches to Understanding Manner and Meaning", Springer SBM, 2010

Assayag, G., Improvising in Creative Symbolic Interaction. In Smith; Chew; Assayag (Eds), Mathematical Conversations: Mathematics and Computation in Music Performance and Composition, World Scientific; Imperial College Press, pp.61 - 74, 2016

Assayag,, G. Creative Symbolic Interaction, 40th Intl. Comp. Mus. Conf. and 11th Sound and Music Comp. Conf. (ICMC / SMC joint conf.), Athenes, Greece, pp 1-6, 2014.

Assayag, G., Truchet, C. (Eds) « Constraint Programming in Music », Constraint Programming in Music, ISTE Ltd and John Wiley & Sons Inc, 256 p., 2011

Assayag, G., Bloch, G., Cont, A., Dubnov, S., Interaction with Machine Improvisation, in S.Argamon, S. Dubnov and K.Burns (Eds) The Structure of Style: Algorithmic Approaches to Understanding Manner and Meaning, Springer, pp.219-245, 2010

Assayag, G., Gerzso, A., « New Computational Paradigms for Computer Music », New Computational Paradigms for Computer Music, ed. Gérard Assayag et Andrew Gerzso. (Ircam/Delatour), 2009

Assayag, G., Bloch, G., Navigating the Oracle: a Heuristic Approach, International Computer Music Conference '07, Copenhagen, 2007, pp. 405-412

Assayag, G., Bloch, G., Chemillier, M., OMax-Ofon, Sound and Music Computing SMC'06, Marseille, 2006

Assayag, G., Bloch, G., Chemillier, M., Improvisation et réinjection stylistiques. Le feed-back dans la création musicale contemporaine - Rencontres musicales pluri-disciplinaires, Mar 2006, Lyon, France.

Assayag, G., Bloch, G., Chemillier, M., Cont, A., Dubnov, S., Omax Brothers: a Dynamic Topology of Agents for Improvisation Learning, Audio and Music Computing for Multimedia, ACM Multimedia '06, Santa Barbara, 2006

Assayag, G., Dubnov, S., « Improvisation Planning and Jam Session Design using concepts of Sequence Variation and Flow Experience », Sound and Music Computing 2005, Salerno, 2005

Assayag, G., Dubnov, S., Using Factor Oracles for machine Improvisation, Soft Computing, Springer Verlag, 2004, 8 (9), pp.604-610, 2004.

Assayag, G., Dubnov, S., Delerue, O., Guessing the Composer's Mind: applying Universal Prediction to Musical Style, ICMC: International Computer Music Conference, Beijing, 1999

Assayag, G., Agon, C., Fineberg, J., Hanappe, P., An Object Oriented Visual Environment for Musical Composition, ICMC: International Computer Music Conference, Thessaloniki, 1997, pp. 364-367

Assayag G., Agon, C., OpenMusic Architecture. International Computer Music Conference ICMC'96, Hong

Kong 1996.

Assayag, G., Visual Programming in Music, International Computer Music Conference ICMC'95, Banff, 1995, pp. 73-76

Arias, T., Desainte-Catherine M., and Dubnov, S., Automatic Construction of Interactive Machine Improvisation Scenarios from Audio Recordings, 4th Intl Workshop on Musical Metacreation, International Conference on Computational Creativity, Paris, 2016

Auslander, Ph. Liveness: Performance in a mediatized culture. New York: Routledge, 1999.

Ayad, L., Chemillier, M., Pissis, S., , Creation Improvisation on Chord Progressions using Suffix Trees, Journal of Mathematics and Music, vol 12 (3), 2018.

Bachelard, G., La formation de l'esprit scientifique, Paris, Alcan, 1934 (reprint. Paris PUF coll. « Quadrige », 2013)

Bamberger, J., Discovering the musical mind: A view of creativity as learning. New York, NY: Oxford University Press, 2013.

Ban, Y.,, T. Narumi, T. Fujii, S. Sakurai, J. Imura, T. Tanikawa and M. Hirose: Augmented Endurance: Controlling Fatigue while Handling Objects by Affecting Weight Perception using Augmented Reality, CHI2013, pp.69-78. 2013.

Boden, M., Computers models of creativity. AI Magazine 30(3): 23–34. 2009

Boden M.A., Creativity and Artificial Intelligence. Artificial Intelligence, 103:347-356, 1998.

Boden, M.A., Dimensions of Creativity, M. A. Boden (ed.), The. MIT Press, Cambridge, Mass., 1996

Boden, M.A., The Creative Mind: Myths and Mechanisms. New York: Basic Books. 1991

Bailey, Derek, L'improvisation. Sa nature et sa pratique dans la musique, trad. Isabelle Leymarie, Paris, Outre Mesure, 1999.

Berliner, Paul. Thinking in Jazz: The Infinite Art of Improvisation. Chicago: University of Chicago Press, 1994.

Benacchio, S., Mamou-Mani, A., Chomette, B., Finel, V. Active control and sound synthesis—two different ways to investigate the influence of the modal parameters of a guitar on its sound. Journal of the Acoustical Society of America, 139 (3), pp.1411, 2016

Benacchio, S., Mamou-Mani, A., Chomette, B., Causse, R. Active control applied to string instruments. Société Française d'Acoustique. Acoustics 2012, Apr 2012, Nantes, France. 2012.

Bengio, Y., Boulanger-Lewandowski, N., & Pascanu, R. (2012). Advances in optimizing recurrent networks. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, (2012): 8624-8628.

Bimbot, F., E. Deruty, G. Sargent, and E. Vincent. System & Contrast: a polymorphous model of the inner organization of structural segments within music pieces. Music Perception 33(5):631–661. 2016

Bigo, L., M. Giraud, R. Groult, N. Guiomard-Kagan, and F. Levé. "Sketching sonata form structure in selected classical string quartets." In Proceedings of the International Society for Music Information Retrieval Conference. pp. 752–759. 2017.

Birkhoff, G.D., Aesthetic Measure, Harvard Univ. Press, 1933.

Blackwell, T., Oliver Bown & Michael Young. Live Algorithms: Towards Autonomous Computer Improvisers. In Jon McCormack & Mark d'Inverno, editeurs, Computers and Creativity, pages 147–174. Springer Berlin Heidelberg, 2012.

Bloch, G., Dubnov, S., **Assayag**, G., « Introducing Video Features and Spectral Descriptors in the OMax Improvisation System », International Computer Music Conference ICMC'08, Belfast, 2008

Butler, M. Playing with Something That Runs. Technology, Improvisation, and Composition in DJ and Laptop Performance, Oxford University Press, 2014.

Bonnasse-Gahot, L., An update on the SOMax project, IRCAM Report, Soma v0.1.3 (2014)

Boyer, E., Arthur Portron, Frédéric Bevilacqua, Jean Lorenceau. Continuous Auditory Feedback of Eye Movements: An Exploratory Study toward Improving Oculomotor Control. Frontiers in Neuroscience, 2017, 11

Briot, J.-P. , Hadjeres G., and Pachet, F.D., Deep Learning Techniques for Music Generation, Computational Synthesis and Creative Systems, Springer, 2019

Bresson, J., Guédy, F., **Assayag**, G., Musique Lab 2: From computer-aided composition to music education, Journal of Music, Technology and Education, Intellect, 2013, 5 (3), pp.273-291.

Bresson J., Agon C., **Assayag** G., Visual Lisp/CLOS Programming in OpenMusic. Higher-Order and Symbolic Computation. Mars 2009, vol. 22, n° 1, p. 81-111

Canonne, C., Garnier, N., "A Model for Collective Free Improvisation", Mathematics and Computation in Music. Third International Conference MCM 2011, IRCAM, Paris, France, June 15-17, 2011. Proceedings, Springer, 2011.

Canonne, C., Garnier, N., Cognition and Segmentation In Collective Free Improvisation: An Exploratory Study. International Conference on Music Perception and Cognition, 2012, Thessaloniki, Greece.

Carpentier, G., **Assayag**, G., Saint-James, E. « Solving the Musical Orchestration Problem using Multiobjective Constrained Optimization with a Genetic Local Search Approach », Journal of Heuristics, 2009.

Carpentier, G., Tardieu, D., **Assayag**, G., Rodet, X., Saint-James, E., (2007) « An Evolutionary Approach to Computer-Aided Orchestration », EvoMUSART, vol. LNCS 4448, Valence, 2007, pp. 488-497

Carsault, T. Nika, J. & Esling, P. Using musical relationships between chord labels in Automatic Chord Extraction tasks. 19th ISMIR Conference (ISMIR2018), Paris, France, 2018

Clayton, Martin. What is entrainment? Definition and applications in musical research. Empir. Musicol. Rev. 7, (2012) 49–56.

Chemillier, M. De la simulation dans l'approche anthropologique des savoirs relevant de l'oralité: le cas de la musique traité avec le logiciel Djazz et le cas de la divination, Transposition, Hors-série n°1, Musique, histoire, société, 2018

Chemillier, M. Nika Jérome, « Étrangement musical »: les jugements de gout de Bernard Lubat à propos du logiciel d'improvisation ImproteK, Cahiers d'ethnomusicologie, n° 28, 2016, pp. 61-80.

Chemillier, M. Jazz et... musiques électroniques, P. Carles, A. Pierrepont (éds.), Polyfree. La jazzosphère (et ailleurs): une histoire récente (1970-2015), Paris, Outre Mesure, ch. 3, pp. 43-54, 2016.

Chemillier, M., Ruse et combinatoire tsigane. De la modélisation informatique dans les répertoiresmusicaux traditionnels (à propos du livre de Victor A. Stoichita: Fabricants d'émotion. Musique et malice dans un village tsigane de Roumanie), L'Homme, n° 211, 2014, pp. 117-128.

Chemillier, M., Fieldwork in Ethnomathematics, Nick Thieberger (ed.), The Oxford Handbook of Linguistic Fieldwork, Oxford University Press, 2011, chapter 14, pp. 317-344.

Chemillier, M. L'improvisation musicale et l'ordinateur. Transcrire la musique à l'ère de l'image animée, Terrain, n° 53 « Voir la musique », 2009, pp. 67-83.

Chemillier, M., Toward a formal study of jazz chord sequences generated by Steedman's grammar. Soft Computing, Vol. 8, No. 9, pp. 617–622, 2004.

Chemla-Romeu-Santos, A., Ntalampiras, S., Esling, P., Haus G., and **Assayag**, G., Cross-Modal Variational Inference for Bijective Signal-Symbol Translation, 22nd International Conference on Digital Audio Effects DAFx 2019

Chew, E.. From sound to structure: Synchronising prosodic and structural information to reveal the thinking behind performance decisions. In C. Mackie (ed.): New Thoughts on Piano Performance, selected papers from the 2013 and 2015 London International Piano Symposium. Kindle Amazon, 2017

Chew, E. Mathematical and Computational Modeling of Tonality: Theory and Applications. International Series in Operations Research and Management Science, New York, NY: Springer, 2014

Cholakis, Ernest, Sound Analysis of Swing in Jazz Drummers: An Analysis of Swing Characteristics of 16 well known Jazz Drummers, 1995, http://www.numericalsound.com/sound-analysis.html

Collins, D. A synthesis process model of creative thinking in music composition. Psychology of Music, 33(2): 193-216, 2005.

Cook, N., Beyond the Score: Music as Performance. Oxford, UK: Oxford University Press, 2013

Conklin, D., Multiple viewpoint systems for music classification. Journal of New Music Research, Volume 42, Number 1, pp. 19-26, March 2013

Conklin, D. Music Generation from Statistical Models , Proceedings of the AISB 2003 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences, Aberystwyth, Wales, 30– 35, 2003.

Crestel, L. Esling, P. Ghisi, D. Meier, R. Generating orchestral music by conditioning SampleRNN, Timbre 2018: Timbre is a many-splendored thing. Montreal, 2018.

Crestel, L., Esling. P., Live Orchestral Piano, a system for real-time orchestral music generation. 14th Sound and Music Computing Conference 2017, Jul 2017, Espoo, Finland. pp.434, 2017

Crochemore, M., Ilie L., Seid-Hilmi, E., The Structure of Factor Oracles, Int. J. Found. Comput. Sci. 18(4), 781–797 (2007).

Colton, S. and Wiggins, G. A.. Computational creativity: The final frontier? In Proceedings of 20th European Conference on Artificial Intelligence (ECAI), pages 21-26, 2012

Cont, A., Dubnov, S., **Assayag**, G., On the Information Geometry of Audio Streams with Applications to Similarity Computing, IEEE Trans. on Audio, Speech and Language Processing, 19 (4), pp.837-846, 2011

Cont, A., Dubnov, S., **Assayag**, G., GUIDAGE: A Fast Audio Query Guided Assemblage, International Computer Music Conference, ICMC'07, Copenhagen, 2007

Cont, A., Dubnov, S., **Assayag**, G., Anticipatory Model of Musical Style Imitation using Collaborative and Competitive Reinforcement Learning , in Butz, M., Sigaud O., and Baldassarre G. (eds.), Anticipatory Behavior in Adaptive Learning Systems, Lecture Notes in Computer Science, Springer Verlag, Berlin, vol. 4520, pp. 285-306, 2007

Cont, A., Dubnov S., **Assayag**, G., A Framework for Anticipatory Machine Improvisation and Style Imitation", Adaptive Behavior in Anticipatory Learning Systems / Simulation of Adaptive Behavior (SAB 2006) Rome, Italy, 2006.

Corning, P. A., The Re-Emergence of "Emergence": A Venerable Concept in Search of a Theory, Complexity, 7 (6): 18–30, 2002

Daikoku, T., Musical Creativity and Depth of Implicit Knowledge: Spectral and Temporal Individualities in Improvisation. Frontiers in Computational Neuroscience vol.12. 2018.

Déguernel, K., Vincent, E., **Assayag**, G., Learning of Hierarchical Temporal Structures for Guided Improvisation. Computer Music Journal, accepted 2019 .

Déguernel, K., Vincent, E., **Assayag**, G., Probabilistic Factor Oracles for Multidimensional Machine Improvisation. Computer Music Journal, 42:2, article 4, Summer 2018

Deguernel, K., Nika, J., Vincent, E., **Assayag**, G. Generating Equivalent Chord Progressions to Enrich Guided Improvisation: Application to Rhythm Changes. SMC 2017 - 14th Sound and Music Computing Conference, Espoo, Finland. pp.8, 2017

Deguernel, K., Nika, J., Vincent, E., **Assayag**, G., Using Multidimensional Sequences For Improvisation In The OMax Paradigm. 13th Sound and Music Computing Conference, Hamburg, Germany, 2016.

Doffman, Mark, Groove! Its Production, Perception, and Meaning in Jazz, M.A. Thesis, U. of Sheffield, 2005.

Doras, G., Esling, P., & Peeters, G. On the use of u-net for dominant melody estimation in polyphonic music. IEEE International Workshop on Multilayer Music Representation, Milano, Italy, pp. 66-70 (2019)

Donze, A., S. Libkind, S.A. Seshia, D. Wessel, Control improvisation with application to music. Technical report No. UCB/EECS-2013-183. EECS Department, University of California, Berkeley, 2013.

Dubnov, S., Characterizing Time Series Variability and Predictability from Information Geometry Dynamic, vol. 8085 of the Lecture Notes in Computer Science series, pp. 658-668, Springer, 2013

Dubnov,S., Musical Information Dynamics as Models of Auditory Anticipation, in Machine Audition: Principles, Algorithms and Systems, ed. W. Weng, IGI Global publication, 2010

Dubnov, S., Unified View of Prediction and Repetition Structure in Audio Signals with Application to Interest Point Detection, IEEE Transactions on Audio, Speech and Language Processing, Vol. 16, No. 2, pp. 327-337, 2008

Dubnov, S., **Assayag**, G., Music Design with Audio Oracle using Information Rate. First International Workshop On Musical Metacreation (Mume 2012), Oct 2012, Palo Alto, United States. pp.1-1, 2012.

Dubnov, S., **Assayag**, G., Cont, A., Audio Oracle Analysis of Musical Information Rate, Proceedings of IEEE Semantic Computing Conference (ICSC2011), United States. pp.567—571, 2011

Dubnov, S., **Assayag**, G., Universal Prediction Applied to Stylistic Music Generation, in Assayag, G., Feichtinger, H.G., Rodrigues, J.F. (eds.), Mathematics and Music, Springer-Verlag, Berlin, pp. 147-158, 2002

Dubnov, S., **Assayag**, G., Cont, A., Audio Oracle: A New Algorithm for Fast Learning of Audio Structures, International Computer Music Conference, Copenhaguen, 2007

Dubnov, S., **Assayag**, G., Lartillot, O., Bejerano G., Using Machine-Learning Methods for Musical Style Modeling, IEEE Computer, vol. 10, n° 38, Octobre, 2003

Dubnov, S., Burns, K.,Kiyoki, Y., Cross-Cultural Multimedia Computing - Semantic and Aesthetic Modeling. Springer Briefs in Computer Science, Springer 2016

Dubnov, S., Cont, A., **Assayag**, G., Audio Oracle: A New Algorithm for Fast Learning of Audio Structures », International Computer Music Conference, Copenhaguen, 2007

Dubnov, S., McAdams, S., and Reynolds, R., Structural and Affective Aspects of Music from Statistical Audio Signal Analysis, Journal of the American Society for Information Science and Technology, 57(11):1526–1536, 2006

Dubnov S., and Surges, G., Delegating Creativity: Use of Musical Algorithms in Machine Listening and Composition, in Digital da Vinci, pp. 127-157, Springer, 2013

Dubos, A., Frédéric Bevilacqua, Joseph Larralde, Joël Chevrier, Jean-François Jego. Designing Gestures for Interactive Systems: Towards Multicultural Perspectives. 16th IFIP Conference on Human-Computer Interaction (INTERACT), Sep 2017, Bombay, India. pp.524-526.

Echeveste, J., Cont, A., Giavitto, J.-L., Jaquemard, F. Operational semantics of a domain specific language for real time musician–computer interaction. J. of Discrete Event Dynamic Systems, 2013, vol. 23, n°4, pp 343-383.

Elgammal, A., Liu, B., Elhoseiny, M., Mazzone, M.: CAN: Creative adversarial networks generating “art” by learning about styles and deviating from style norms (2017). ArXiv:1706.07068v1

Esling, P., Chemla-Romeu-Santos, A., & Bitton, A. “Generative timbre spaces: regularizing variational auto-encoders with perceptual metrics”. Proceedings of International DaFX conference, pp.23-27, Portugal (2018)

Esling, P. Chemla, A. Bitton, A. “Bridging audio analysis, perception and synthesis with perceptually-regularized variational timbre spaces”, Proc. 19th International Society for Music Information Retrieval, vol. 1, Paris, France. AR: 23,2% Best paper award. (2018)

Fujie, S., Yagi, D., Matsusaka, Y., et al. Spoken dialogue system using prosody as para-linguistic information. In: Speech Prosody 2004, International Conference. 2004.

Fitch WT. Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. Front Syst Neurosci. 2013;7:68.

François, A. R. J., Schankler, I., and Chew E., Mimi4x: An interactive audio-visual installation for high-level structural improvisation. International Journal of Arts and Technology, 6(2): 138–151, 2013.

François, A.R.J., E. Chew, D. Thurmond (2008). Visual Feedback in Performer-machine Interaction for Musical Improvisation. In Proc NIME, 277-280.

Gibson, W., The group ethic in the improvising jazz ensemble: A symbolic interactionist analysis of music, identity, and social context. Studies in Symbolic Interaction. 2010.

Gilhooly, K. J., Thinking: Directed, Undirected and Creative. London: Academic Press, 1996.

Giraud, M., R. Groult, E. Leguy, and F. Levé. 2015. “Computational fugue analysis.” Computer Music Journal 39(2):77–96.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozairy, S., Courville, A., Bengio, Y., Generative adversarial nets (2014). ArXiv:1406.2661v1

Hadjeres, G., Pachet, F., Nielsen, F.: DeepBach: a steerable model for Bach chorales generation (2017). ArXiv:1612.01010v2

Herremans, D., Chuan, C.H., Chew, E.: A functional taxonomy of music generation systems. ACM Computing Surveys (CSUR) 50(5) (2017)

Hinton, G.E.: Boltzmann machine. Scholarpedia 2(5), 1668 (2007)

Kato, H., Ushiku, Y., Harada. T. Neural 3D Mesh Renderer. The 31th IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

Ikei, Y, Hirota, K., Amemiya T., Kitazaki, M., Five Senses Theatre: A multisensory display for the bodily Ultra Reality, Emotional Engineering Vol.4, (Ed.) Shuichi Fukuda, Springer, 2016

Ishi, C. T., Ishiguro, H., et Hagita, N.. Analysis of relationship between head motion events and speech in dialogue conversations. Speech Communication, 2014, vol. 57, p. 233-243.

Kanehira, A., Takemoto, K., Inayoshi, S., Harada, T., Multimodal Explanations by Predicting Counterfactuality in Videos. The 32nd IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

Keil, Charles. The Theory of Participatory Discrepancies: a progress report. Ethnomusicology, 39(1), (1995) 1-19.

Kingma, D.P., Welling, M., An Introduction to Variational Autoencoders, arXiv:1906.02691, 2019

Kingma, D.P., Welling, M.: Auto-encoding variational Bayes (2014). ArXiv:1312.6114v10

Knapp, M., Hall, J., Nonverbal Communication in human interaction, Cengage Learning, 2013.

Kopp, S., Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. Speech Communication, 2010, vol. 52, no 6, p. 587-597.

Large, Edward W., and John F. Kolen. 1994. Resonance and the perception of musical meter. Connection Science 6(2–3):177–208.

Large, Edward W., and Caroline Palmer. 2002. Perceiving temporal regularity in music. Cognitive Science 26:1–37

Lartillot, O., Dubnov, S., **Assayag**, G., Bejerano, G., Automatic Modeling of Musical Style, International Computer Music Conference, La Havane, 2001, pp. 447-454

Lévy, B., Bloch, G., **Assayag**, G., « OMaxist Dialectics: capturing, Visualizing and Expanding Improvisations », NIME12, Ann Arbor, 2012, pp. 137-140

Lissek, H., Rivet, E., Laurence, T., Fleury, R., Toward Wideband Steerable Acoustic Metasurfaces with Arrays of Active Electroacoustic Resonators; Journal of Applied Physics. 2018.

McPherson, A.P. Kim, Y., Augmenting the Acoustic Piano with Electromagnetic String Actuation and Continuous Key Position Sensing. Proc. Intl. Conference on New Interfaces for Musical Expression NIME'10, 2010

Maniatakos, F., **Assayag**, G., Bevilacqua, F., Agon, C., On architecture and formalisms for computer assisted improvisation , Proc. Sound and Music Computing conference, SMC'10, Barcelona, 2010

Mehri, S., Kumar, K., Gulrajani, I., Kumar, R., Jain, S., Sotelo, J., Courville, A. & Bengio, Y. (2016). SampleRNN: An Unconditional End-to-End Neural Audio Generation Model (cite arxiv:1612.07837)

Meurisse, T., Mamou-Mani, A., Causse, R., Chomette, B., Sharp, D. Simulations of modal active control applied to the self-sustained oscillations of the clarinet. Acta Acustica, Hirzel Verlag, 2014, 100 (6),pp.1149-1161.

Meyer, L.B., Emotion and Meaning in Music. Chicago: Chicago University Press, 1956.

Moles, A., Théorie de l'information et perception esthétique, Flammarion, Paris, 1958

Moles, A., Art et ordinateur, Pris Casterman, 1971.

Monson, Ingrid. Saying something: jazz improvisation and interaction. Chicago: University of Chicago Press, 1996.

Moreira, J., Roy, P., Pachet, F. VirtualBand: Interacting with Stylistically Consistent Agents.ISMIR, pages 341-346, Curitiba (Brazil), 2013

Mouawad, P., Dubnov, S., On Modeling Affect in Audio with Non-Linear Symbolic Dynamics, Advances in Science, Technology and Engineering Systems Journal, Vol. 2, No. 3, 1727-1740 2017

Mouawad, P., Dubnov, S., On Symbolic Dynamics and Recurrence Quantification Analysis for Affect Recognition invoice and Music, Perspectives in Nonlinear Dynamics 2016, Berlin

Nika, J., Chemillier, M., ImproteK, integrating harmonic controls into improvisation in the filiation of OMax, Proceedings of the International Computer Music Conference, pp. 180–187, 2012.

Nika, J., Chemillier, M., **Assayag**, G., ImproteK: introducing scenarios into human-computer music improvisation, ACM Computers in Entertainment, 14: 2, article 4, 2017.

Nika, J., Deguernel, K., Chemla--Romeu-Santos, A., Vincent, E., **Assayag**, G., DYCI2 agents: merging the "free", "reactive", and "scenario-based" music generation paradigms, Proc. ICMC, Oct 2017, Shangai, China. 2017

Novielli, N., De Rosis, F., and Mazzotta, I., User attitude towards an embodied conversational agent: Effects of the interaction mode. Journal of Pragmatics, 2010, vol. 42, no 9, p. 2385- 2397.

Ogawa, N., Y. Ban, S. Sakurai, T. Narumi, T. Tanikawa and M. Hirose: Metamorphosis Hand: Dynamically Transforming Hands, ACM Augmented Human 2016, 2016.

Pachet, F. and Roy, P. Markov constraints: steerable generation of Markov sequences. Constraints, 16(2):148-172, March 2011

Pachet, F., Musical virtuosity and creativity. Computers & Creativity, 2012.

Patel, Aniruddh D., John R. Iversen, Micah R. Bregman, Irena Schulz, and Charles Schulz. 2008.

Investigating the human-specificity of synchronization to music. Proc.10th Intl. Conf. on Music Perception and Cognition 1–5, 2008.

Pinch, T., Trocco, F. Analog days: The invention and impact of the Moog synthesizer. Cambridge, MA: Harvard University Press, 2002.

Potter, K., Wiggins, G., and Pearce, M., Towards greater objectivity in music theory: Information-dynamic analysis of minimalist music. In Musicae Scientiae 11(2), 295-322, 2007

Pressing, J., Cognitive processes in improvisation. Advances in Psychology, Vol. 19, pp. 345–363, 1984.

Rabinovich, M.I, and Varona, P., (2016), Consciousness and joint Human-Robot Creativity; Chaotic Chunking, Frontiers Neuroscience 2016

Rabinovich, M., Trisan I., and Dubnov, S., Nonlinear dynamics of human creativity, IEEE International Conference on Systems, Man and Cybernetics, 2014

Roberts, A., Engel, J., Raffel, C., Hawthorne, C., Eck, D.: A hierarchical latent vector model for learning long-term structure in music. In: Proceedings of the 35th International Conference on Machine Learning (ICML 2018). ACM, Montreal, PQ, Canada (2018)

Rowe, R., Machine Musicianship. MIT Press, Cambridge, MA, USA, 2004.

Rueda, C., **Assayag**, G., Dubnov, S., A Concurrent Constraints Factor Oracle Model for Music Improvisation », XXXII Conferencia Latinoamericana de Informática CLEI 2006, Santiago, 2006

Sanlaville, K., Assayag, G., Bevilacqua F., Pelachaud, C., Emergence of synchrony in an Adaptive Interaction Model, Intelligent Virtual Agents 2015 Doctoral Consortium, 2015, Delft,

Sawyer, R.K., Improvisational Cultures: Collaborative Emergence and Creativity in Improvisation, Mind, Culture, and Activity, 7:3, 2000

Schankler, I., Chew, E., and François A. R. J., Improvising with digital auto-scaffolding: How Mimi changes and enhances the creative process. In Digital Da Vinci, ed. Newton Lee, pp. 99–125. Berlin: Springer Verlag. 2014.

Scurto, H., Frédéric Bevilacqua. VIMOs: Enabling Expressive Mediation and Generation of Embodied Musical Interactions. 4th International Conference on Movement Computing, Jun 2017, London.

Smith, J., Chew, E., **Assayag** G., (Eds), Mathemusical Conversations: Mathematics and Computation in Music Performance and Composition, Feb 2015, Singapore. World Scientific; Imperial College Press, 2016.

Smith, J. B. L., Schankler, I., Chew, E., Listening as a Creative Act: Meaningful Differences in Structural Annotations of Improvised Performances. Music Theory Online, 20(3). 2014. www.mtosmt.org/issues/mto.14.20.3/mto.14.20.3.smith_schankler_chew.php

Steels, L. Evolving grounded communication for robots. Trends in cognitive sciences, 2003, vol. 7, no 7, p. 308-312.

Smith, J. B. L., Chuan, C., Chew, E., Audio Properties of Perceived Boundaries in Music. IEEE Transactions on Multimedia Special Issue on Music Data Mining, 16(5): 1219-1228, 2014.

Steedman M.J., A Generative Grammar for Jazz Chord Sequences, Music Perception, vol. 2, n°1, 1984, 52-77.

Surges, G. and Dubnov, S. Feature Selection and Composition using PyOracle. Workshop on Musical Metacreation, Ninth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment. Northeastern University, Boston, MA. October 14-15, 2013.

Sutton, R. S., Barto, A. G., Reinforcement Learning: An Introduction. The MIT Press, 2018.

Thórisson, K. R. Natural turn-taking needs no manual: Computational theory and model, from perception to action. In: Multimodality in language and speech systems. Springer Netherlands, 2002. p. 173- 207.

Tokushige, H., T. Narumi, S. Ono, Y. Fuwamoto, T. Tanikawa, and M. Hirose: Trust lengthens decision time on unexpected recommendations in human-agent interaction, HAI2017, 2017.

Tokozume, Y., Ushiku, Y., Harada, T., Learning from Between-class Examples for Deep Sound Recognition. The 6th International Conference on Learning Representations (ICLR 2018), 2018

Tokozume, Y., Harada, T., Learning environmental sounds with end-to-end convolutional neural network. The 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017), 2017

Toro M., Rueda, C., Agon, C., **Assayag**, G., GELISP: A Framework to represent musical constraint satisfaction problems and search strategies , Journal of Theoretical and Applied Information Technology, JATIT, 2016, 86 (2), pp.327-331.

Toro M., Rueda, C., Agon, C., **Assayag**, G., NTCCRT: A Concurrent Constraint Framework for Soft Real-Time Music Interaction, Journal of Theoretical and Applied Information Technology, JATIT, 2015, 82 (1), pp.184-193.

van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., Kavukcuoglu, K.: WaveNet: A generative model for raw audio (2016). ArXiv:1609.03499v2

Wang, C., and Dubnov, S., Pattern Discovery from Audio Recordings by Variable Markov Oracle: A Music Information Dynamics Approach, Proc. of 40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, 2015

Wang C. and Dubnov, S., The Variable Markov Oracle: Algorithms for Human Gesture Applications, IEEE Multimedia, vol. 22, no. 4, pp. 52 – 67, 2015

Wang C., Hsu, J., and Dubnov, S., Machine Improvisation with Variable Markov Oracle: Toward Guided and Structured Improvisation, ACM Computers in Entertainment Journal, Second Special Issue on Music Meta-creation, 2016

Wiggins, G.A., Searching for computational creativity. New Generation Computing, 24(3):209–222, 2006.

Yang, L.C., Chou, S.Y., Yang, Y.H.: MidiNet: A convolutional generative adversarial network for symbolic-domain music generation. In: Proc. of the 18th International Society for Music Information Retrieval Conference (ISMIR 2017). Suzhou, China (2017)