VIMOs: Enabling Expressive Mediation and Generation of Embodied Musical Interactions

Hugo Scurto, Frédéric Bevilacqua
IRCAM
Paris, France

ABSTRACT

This paper presents a recently-started doctoral project towards a new framework, called Virtual Intelligent Musical Objects (VIMOs). VIMOs aim at combining features of interactive machine learning tools, autonomous agents, and collaborative media into the creation of motion-based, user adaptable, shareable interactive music systems. We propose two models for stylistic motion learning and generation under a "design through performance" interactive workflow. We discuss further human and computer-related research challenges involving expressiveness rendering and social musical interaction, as well as novel artistic and educational applications to be led within the scope of VIMOs.

CCS CONCEPTS

•Human-centered computing → Gestural input; •Applied computing → Sound and music computing;

KEYWORDS

Expressive movement; Music; Machine learning; Interaction design; Embodied cognition; Social computing.

ACM Reference format:
–. 2017. VIMOs: Enabling Expressive Mediation and Generation of Embodied Musical Interactions. In *Proceedings of International Conference on Movement Computing, London, UK, June 2017 (MOCO'17),* 4 pages.
DOI: 10.475/123_4

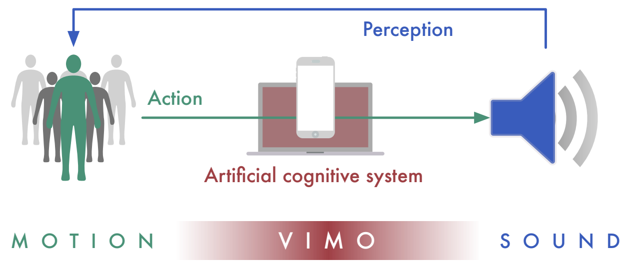

Figure 1: VIMOs' workflow. The cognition level enables real-time adaptation to one or several users' motion.

1 INTRODUCTION

At the intersection of human-computer interaction and machine learning research, interactive approaches to machine learning have been successfully applied for designing interactive music systems. Promising applications range from performing arts and sound design [4] to health research [6]. For now, interactive machine learning frameworks have essentially focused on individual users designing custom motion-sound relationships — a use case for which it has proven effective [11]. Current challenges include the three following points. First, in motion-based interactive systems, creating gestures can sometimes be a major obstacle, notably for non-expert users [20]. Second, current systems propose "mechanical" and fixed interaction behaviors that poorly account for phenomena found in human-human interaction [16]. Third, as musical communication often involves more than one performer and/or listener, interactive systems should fully support collective and social uses.

This doctoral work will develop a framework, called Virtual Intelligent Musical Objects (VIMOs, see Fig. 1), that addresses the forementioned issues by combining three aspects. First, it should contain a tool enabling incremental and adaptive learning from a user's motion style. Second, it should implement a motion-based autonomous sonic agent for creative human-machine co-operation, or improvization. Third, it should behave as a shared motion-based interactive medium where synchronization and entrainment would drive artistic expression learning [17]. We believe that VIMOs may constitute a gateway to individual and collective embodied musical interactions. Aiming for VIMOs' framework requires adopting a holistic approach, drawing from embodied music cognition to the design and evaluation of machine learning models for expressive interaction.

2 BACKGROUND

2.1 Embodiment and Music Interaction

Embodied interaction has been proposed in the last decade by Human-Computer Interaction (HCI) researchers as a common basis to areas such as tangible and social computing, where users can express collectively and acquire skills directly through interactive media [10]. In this spirit, effort has been devoted to design interactive tools that feels more "natural" to users, aiming for fluid, custom interaction flows between human ideas and computer actions [3]. Body movements have constituted one of the most investigated input for such embodied interactive systems, notably leading to the creation of virtual agents that can convey affective expressions to users [14].

Interestingly, music and performing arts constitute exemplary case studies for previously-cited HCI research. Leman proposed

DOI: 10.475/123_4

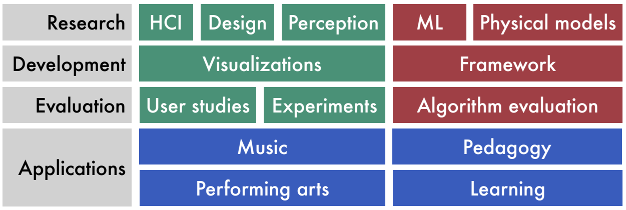

Figure 2: Research themes and potential contributions. In green: Human-centred aspects. In red: Computer-centred aspects. In blue: Application domains.

that corporeal engagement with music may be at the core of intentionality attribution between performers and listeners. He insisted on several attributes for designing motion-based interactive music systems — such as the creation of autonomous social agents creating an illusion of non-mediation [16]. Moreover, he recently argued that such expressiveness transfer and entrainment in musical interaction is essentially realized through continuous flows involving motion patterns and dynamics [17] — thus matching dance theorists' views on so-called "movement qualities" [1].

2.2 Expressive Motion and Music Learning

Machine learning (ML) has been investigated by researchers to support expressive motion-sound mapping in interactive systems [8]. In the last decade, interactive approaches to machine learning have made such technology more accessible by adopting a mapping-by-demonstration approach [12]. With such tools, users can build custom motion-sound relationships through a two-step, iterative design process: first, demonstrating gesture examples; second, experimenting with the newly-created model; then, alternating between such design and performance steps until converging to a satisfying system. Several user studies have shown this approach very useful when the goal is to parametrize a motion space [11, 12], or to track temporal profiles of gestures [4, 7, 12]. However, while users emphasized the potential advantages of a continuous corporeal interaction flow [20], interaction design remains separated in distinct steps. Also, agent-like imitative motion generation is not directly supported.

Alternatively, other interactive music systems aimed for such agent-like, continuous interaction flows. The OMax/SOMax systems respectively allow for improvization generation by recombining stylistically-consistent elements of a memory [2] and real-time improvization guidance [5, 9]; Pachet's Flow Machines also allow for such real-time reactive improvization using different computational approaches [18]. However, these three frameworks are based on discrete representations of pitch-based musical content, and might not be able to extrapolate a given style — that is, generation is only made possible by a recombination process.

Interestingly, physical models have been used as a fully generative approach for real-time continuous trajectory creation in motion-sound mapping [19], even allowing a quality-based encoding and visualization of motion in dance technology [1]. Yet, such generation and encoding remain hard for users to customize. Overall, very few attempts have been made to turn these systems into medium-like, shared systems.

3 RESEARCH THEMES

This doctoral project aims at developing new interactive music systems, called Virtual Intelligent Musical Objects (VIMOs), that would combine features of interactive media, customizable tools, and expressive agents into the creation of shareable *moving sonic forms* (as defined in [16]). It adopts an interdisciplinary approach with potential contributions in different fields as shown in Figure 2. For clarity purposes, research questions are divided in two groups: human-centered, and computer-centred.

3.1 Human-Centred Questions

We identified three features that we deem essential to the design of interactions at stake in VIMOs:

(1) Embodied interaction: Users should interact with motion only to maintain real-time interactive flows.
(2) Expressive/stylistic interaction: Users should be able to express fine stylistic features.
(3) Collective interaction: Users should be able to interact collectively and share expressive social cues.

We motivate these three points with previously-cited works in embodied cognition and human-computer interaction that stress the creative potential of fluidly interacting with a virtual expressive agent. Importantly, interactive visualizations will be implemented to facilitate user perception of the system's behaviour and interaction in a mixed reality environment.

3.2 Computer-Centred Questions

We identified three features that we deem essential to the design of models at stake in VIMOs:

(1) User adaptation: Models should support online incremental adaptation to users' motion space and style.
(2) Autonomous behaviour: Models should be able to generate moving forms that reflects users' style.
(3) Expressiveness rendering: Models should encode movement qualities, imitating and extrapolating users' style.

We motivate these three points with previously-cited works in machine learning and style modelling that highlight creative advantages of each. Combining these three features is in itself challenging: it would require designing flexible models that support real-time learning of individual or multiple users motion.

4 CURRENT WORK

4.1 Interaction Design

We propose a new interactive workflow, named "design through performance" (see Fig. 3), that aims at supporting the three features described in section 3.1.

The workflow aims at merging interactive machine learning systems' design and performance steps into one continuous, fully interactive experience. Instead of alternating between customization and expression tasks, users would be allowed to do both by only moving, thus supporting both embodied and stylistic interaction. We hypothesize that such continuous user adaptation could also mediate collective interaction: as the system's reaction to user input would be direct, different users could interact together while

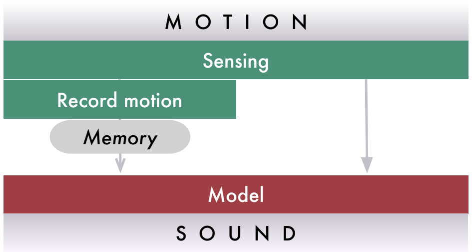

Figure 3: A design through performance workflow.

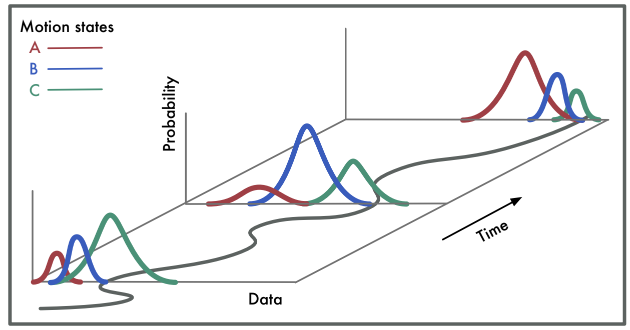

Figure 4: Interactive mapping at stake in VIMOs. Here, three gaussians adapt in real-time to users' motion.

performing. Overall, our wish is to propose a new "computer-in-the-loop" interaction paradigm, where humans would perceive the computer as an expressive tool, a creative partner, and an intelligent medium at once [3].

4.2 Computational Models

So far, we have investigated two models that respectively aim at the two first features described in section 3.2 — namely user adaptation and autonomous behaviour. Importantly, both models' data-driven implementation could support a shared use; yet, we only tested them in an individual setup for now.

4.2.1 User Adaptation. This work acts as a follow-up to recent works in motion-sound mapping. Our goal is to adapt these mapping-by-demonstration approaches to our design through performance workflow. We recently submitted a paper that details the model's implementation and workflow.

We propose an online implementation of a gaussian mixture model (GMM, see Fig. 4) for continuous learning and adaptation to user motion input. GMMs are very general and versatile probabilistic models for designing motion-sound relationships [12]. Here, we use them to perform online clustering, which is identifying groups of similarity in user motion input in real-time. The evolving GMM-based modelling of user-specific motion space "internal structure" can then drive sound synthesis in several expressive ways, all of them being characterized by direct, corporeal interaction with sound and personalized exploration of motion in relation to sound. An improvizational use case involving various controllers and sonic environments showed advantages compared to previous systems to directly, physically explore and reflect users' relevant motion states — be they positions, postures, or dynamical attributes.

4.2.2 Autonomous Behaviour. While our first model focuses on learning motion states, this second model focuses on motion patterns. It builds on research done on human-computer musical improvization and pitch-based style modelling [5, 9]. A crucial aim relies on adapting such approaches to motion data.

We propose to investigate a reactive factor oracle (FO) for continuous generation of, and adaptation to user-specific motion patterns. A factor oracle is an agent that models a sequence of states by creating links between states that share a common context. Here, we use it to learn online motion pattern sequences (where motion-sound states would be defined using one previously-described GMM representation), and generate new stylistically-consistent moving forms by recombining learned sequences of states in reaction to users' performance (where moving sonic improvizations would emerge through human-computer co-adaptation). Preliminary observational experimentations showed promising abilities for generating novel embodied musical interactions in a design through performance workflow, such as listening and physically imitating a moving sonic form.

5 PLANNED WORK

Designing VIMOs requires completing our current work with several studies encompassing different research fields (see Fig. 5). For clarity purposes, we divided them in three groups: human-centered, computer-centred, and real-world applications.

5.1 Human-Centred Studies

We plan to lead a twofold human evaluation of VIMOs: (a) studying user interaction with VIMOs from a HCI perspective, and (b) testing VIMOs' perception from a cognitive perspective. The first part will respectively investigate individual and collective interaction with VIMOs through longitudinal user studies and participatory workshops. Various user groups will be involved to better understand what creative uses novice and expert musicians would potentially praise or discover. The second part will study how people perceive VIMOs' stylistic encoding of motion through controlled experiments. This would be a first step towards studying expressive style learning between two or more persons, where a teacher would first have a VIMO learn his or her movement style that could be transferred to other users through performance.

5.2 Computer-Centred Challenges

The computer-centred challenges of this thesis are threefold: (a) investigating other models while continuing working on our two current models, (b) evaluating them from a computational perspective, and (c) implementing them in a global framework equipped with interactive visualizations. First, in the online GMM current implementation, the number of clusters at stake is established in advance. We plan to optimize the number of clusters to enable adaptive motion space partitioning, in which states of motion variations could be taken into account as users perform with the system. Also, we tested the reactive FO for the learning of only one motion representation at once. Combining several motion representations could allow for more relevant recombinations of user-specific motion

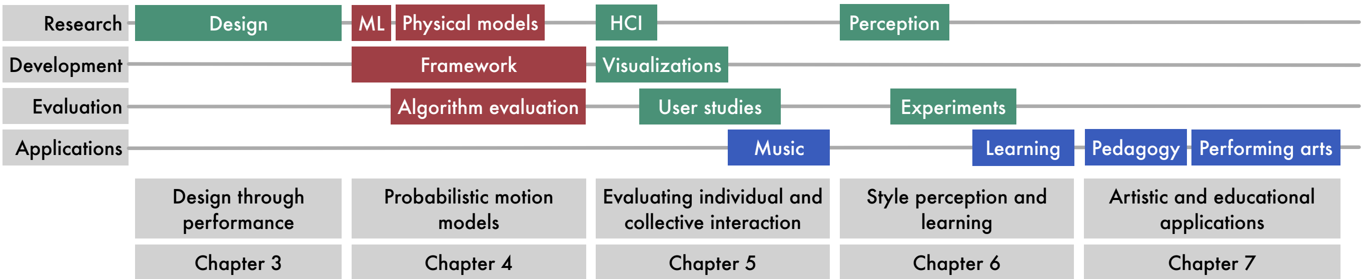

Figure 5: Targeted timeline.

patterns. We also plan to address the issue (3) of expressiveness rendering by investigating physical models, following work in dance technology [1] and robotics [13, 15]: a first step may be to use them to continuously interpolate between the agent's state sequence in accordance to users' motion style. Finally, computational evaluation will be led by comparing models' performance to other up-to-date systems using synthetic databases.

5.3 Real-World Applications

An important goal of this doctoral work will be to use VIMOs in real-world artistic and educational contexts. First, applications in music, interactive, and performing arts will be covered. Collective uses of VIMOs may be tested in an interactive concert setting, using recent platforms for mobile-based collaborative performances[1]. Also, pedagogical applications in art mediation, music, and dance will be investigated to evaluate individual and collective uses of VIMOs by users with different levels of expertise. Based on recent works [20], we especially hypothesize that VIMOs' design through performance workflow might mediate novel forms of musical communication between music therapists and disabled users. Lastly and on a long-term view, VIMOs may be used in a rehabilitation setting, where the embodiment of expressive qualities would drive movement learning through music.

ACKNOWLEDGMENTS

Omitted for anonymous review.

REFERENCES

[1] Sarah Fdili Alaoui. 2012. *Analyse du geste dansé et retours visuels par modèles physiques: apport des qualités de mouvement à l'interaction avec le corps entier.* Ph.D. Dissertation. Université Paris Sud-Paris XI.
[2] Gérard Assayag, Georges Bloch, Marc Chemillier, Arshia Cont, and Shlomo Dubnov. 2006. Omax brothers: a dynamic yopology of agents for improvization learning. In *Proceedings of the 1st ACM workshop on Audio and music computing multimedia.* ACM, 125–132.
[3] Michel Beaudouin-Lafon. 2004. Designing interaction, not interfaces. In *Proceedings of the working conference on Advanced visual interfaces.* ACM, 15–22.
[4] Frédéric Bevilacqua, Bruno Zamborlin, Anthony Sypniewski, Norbert Schnell, Fabrice Guédy, and Nicolas Rasamimanana. 2009. Continuous realtime gesture following and recognition. In *International Gesture Workshop*. Springer, 73–84.
[5] Laurent Bonnasse-Gahot. 2014. An update on the SOMax project. *Ircam-STMS, Internal report ANR project Sample Orchestrator* 2 (2014).
[6] Eric Olivier Boyer, Bénédicte Maria Babayan, Frédéric Bevilacqua, Markus Noisternig, Olivier Warusfel, Agnes Roby-Brami, Sylvain Hanneton, and Isabelle Viaud-Delmon. 2013. From ear to hand: the role of the auditory-motor loop in pointing to an auditory source. *Frontiers in computational neuroscience* 7 (2013), 26.
[7] Baptiste Caramiaux, Nicola Montecchio, Atau Tanaka, and Frédéric Bevilacqua. 2015. Adaptive gesture recognition with variation estimation for interactive systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 4, 4 (2015), 18.
[8] Baptiste Caramiaux and Atau Tanaka. 2013. Machine Learning of Musical Gestures.. In *NIME*. 513–518.
[9] Axel Chemla. 2015. *Guidages de l'improvisation musicale homme-machine.* Master's thesis. UPMC.
[10] Paul Dourish. 2004. *Where the action is: the foundations of embodied interaction.* MIT press.
[11] Rebecca Anne Fiebrink and Perry R Cook. 2011. *Real-time human interaction with supervised learning algorithms for music composition and performance.* Citeseer.
[12] Jules Françoise. 2015. *Motion-Sound Mapping by Demonstration.* Ph.D. Dissertation. UPMC.
[13] Auke Jan Ijspeert, Jun Nakanishi, Heiko Hoffmann, Peter Pastor, and Stefan Schaal. 2013. Dynamical movement primitives: learning attractor models for motor behaviors. *Neural computation* 25, 2 (2013), 328–373.
[14] Michelle Karg, Ali-Akbar Samadani, Rob Gorbet, Kolja Kühnlenz, Jesse Hoey, and Dana Kulić. 2013. Body movements for affective expression: A survey of automatic recognition and generation. *IEEE Transactions on Affective Computing* 4, 4 (2013), 341–359.
[15] Dana Kulic, Wataru Takano, and Yoshihiko Nakamura. 2007. Incremental learning of full body motions via adaptive Factorial Hidden Markov Models. In *Proc. of the International Conference on Epigenetic Robotics.*
[16] Marc Leman. 2008. *Embodied music cognition and mediation technology.* Mit Press.
[17] Marc Leman. 2016. *The Expressive Moment: How Interaction (With Music) Shapes Human Empowerment.* MIT Press.
[18] François Pachet, Pierre Roy, and Fiammetta Ghedini. 2013. Creativity through style manipulation: the flow machines project. In *2013 Marconi Institute for Creativity Conference, Proc.(Bologna, Italy*, Vol. 80.
[19] Jan C Schacher, Daniel Bisig, and Philippe Kocher. 2014. The map and the flock: Emergence in mapping with swarm algorithms. *Computer Music Journal* (2014).
[20] Hugo Scurto and Rebecca Fiebrink. 2016. Grab-and-play mapping: Creative machien learning approaches for musical inclusion and exploration.. In *Proceedings of the International Conference of Computer Music.*

[1] Link to the related research project, omitted for anonymous review.