1. Excellence

1.1 Quality and pertinence of the project's research and innovation objectives (and the extent to which they are ambitious, and go beyond the state of the art) #@QUA-LIT-QL@#

Introduction:

Based on a simple principle consisting of juxtaposing short sounds sharing acoustic characteristics, the domain of Corpus-Based Concatenative Synthesis (CBCS) emerged from commercial voice resynthesis at IRCAM at the turn of the millennium. Pioneered by Diemo Schwarz (2000), the realm has enjoyed steady growth ever since, making Creative Music Information Retrieval (CMIR) techniques to become today accessible to musicians. A consensual critique of early uses of Corpus-Based Concatenative Synthesis accused its lack of smoothness due to abruptly chopped sounds, such as could be heard in several musical works or interactive scenarios produced since the early days of CataRT, but the recent spread of the practice has helped overcome those issues, with the ERC-funded FluCoMa[1] (2021) and REACH[2] projects, as well as with higher-level softwares such as AudioStellar or lfo-lab's Mosaique. Today's compelling capacities of CBCS urges composers, creative coders and sound designers to master those technique for music creation and design of interactive scenarios. Such computing technology challenges today the notion of human creativity as it is able to accurately simulate and sometimes surpass the behaviour of a human, a problem aptly summed up by Thor Magnusson in the field of music AI: "If a computer can do this, what are you going to do?".

Most probably driven by the hardware sells of tech giants, but also possibly more decentralised as envisioned in Web 3.0, some form of cyberspace is expected to play a growing role in how we live, work, and interact with each other in the digital world, but how may music making be part of this. The so-called metaverse and the realm of Networked Music Performance (NMP) both flourished during the Covid years, and praised new forms of telecommunication as a palliative to physical distancing. NPM research typically correlates the sense of presence and fluid interaction with a threshold in low latency audio, while technologies of virtual immersion primarily convey a sense of presence of other participants through a perceived adequation of visuals with binaural sound (Serafin 2018). With this remote social interaction in mind, the project consists of using the aforementioned CMIR tools for musical Sound Interaction in Virtual Environments (SIVE).

Overview:

The division between the voice (God-made) and instruments (man-made) in the Middle Ages, somehow relates to contemporary communities gathered around classical instrumental music (funded by patrons or the state) and electronic music (linked to pop culture and boosted by industry). In such context, as already witnessed in the visual arts domain, an AI earning an award by generating the audio of an instrumental piece evokes an act of profanation, for it presents a threat to (hitherto godly) creativity, reserved to humans, with machines only supposed to realise delegated tasks. Issues of authorship involved here pose unprecedented challenges, nevertheless Fears relative to advances in AI remind of many other previous disruptive technological advances in which artists have always found their way. Furthermore, Artists often operate on a longer time scale than the ever evolving-technological industry, and therefore scarcely feel concerned about the latest warnings of the press's headlines. In art music, one attitude – still persisting - consists in operating without technology: composers such as M. Feldman, and H. Lachenmann have avoided electronic music throughout their career, which, in hindsight, appears strategic when considering both artistic and technologic obsolescence in the early days of computer music. Nevertheless, Co-creation, as put forward by REACH, appears a more coercive approach today, particularly because of the ability AI has to model styles and thus approach unprecedented aesthetical questions: award-winning AI-generated artworks best hide the glitches characteristic of their time by the evocation of remote artistic styles (baroque painting for Jason Allen, black and white photograph for Boris Eldagsen). Observing discrepancies between music presented in concert halls and research experiments circumscribed to laboratories, the current proposal investigates whether this emerging stylistic exploration, afforded by AI in general and CBCS in particular, could forge the ground for an authentic aesthetic enquiry, in an environment (XR) predominantly targeting the entertainment industry and monetization. How disruptive it may seem from an artistic standpoint, through its use of 3d avatars and still discontinuous browsing of sonic data, the project is looking for a Historical continuity by focusing on instrumental contemporary practice, or on the creative use, since 1945, of instruments whose evolution mostly froze in the 19th Century[3]. Through a practice-

---

[1] https://www.flucoma.org/

[2] http://repmus.ircam.fr/_media/reach/reach_erc_adg_2019_b1.pdf

[3] https://youtu.be/p-cIRxFiEDI?list=PLc_WX6wY4JtnwPpysSL2F3C_pfg1xxPXs

based enquiry, the proposal addresses two major research questions: Firstly, how to bring the old chamber music making paradigms to worlds opened by immersive technologies and AI? Secondly, is small/custom data AI capable of *producing* (in its etymological sense of "a coming into being") aesthetically valid sonic output, and how this compares with deep-learning's biased tendency to re-produce normative output.

The global metaverse market is expected to grow by 41.6% every year from 2023 to 2030, and yet it remains today a challenge to foresee how the musical domain may join this movement in a meaningful way, without undergoing oversimplification imposed by the entertainment industry. The design of instruments (VRMIs) and virtual agents seems the most plausible strategy for experimenting with Sonic interaction in Virtual Environments (SIVE) in a telematic musical performance. In such a context, how to design meaningful interactive instruments for the metaverse? Within this emerging domain, a few projects are already surpassing the mere proof of concept: in academia Rob Hamilton's physical modeling approach explores various elaborate forms of interaction with virtual strings, bows and plectrums, albeit still limited when compared to the dexterity of a trained instrumentalist. In the industry, a few softwares have taken to VR the metaphors of the modular synthesiser and DJ setups but rarely allow exploration outside of the EDM style. In contrast, with CBCS, the E.R. proposes a bias towards pre-recorded audio and machine learning-aided corpus manipulation, judged musically richer and stylistically more versatile than the aforementioned approaches. In comparison with physical models in which parametric gestures can closely imitate the real world (e.g. by shortening a string to make it sound higher), CBCS imposes more abstract mappings in the gestural interaction. Which types of virtual representation may best convey the user a sense of control over pre-existing musical structures? How gestural interaction with virtual objects may best represent the action of browsing a sound corpus? Which virtual objects, which haptic and visual feedback will best convey the metaphor of the instrumental gesture? The curation and segmentation of the corpus (a database of samples), as well as the extraction of a model (sound map) will play a crucial role in the search of plausible instrumental simulation. Which sonic descriptors, which machine learning algorithms to use to reveal similarities? How long should each sample be and how can they be chained after another to produce plausible sonic results? Bearing in mind that the metaverse is primarily expected to foster new forms of social interactions, and that the gaming industry has already provided robust solutions tackling issues with potential conflicts between different states of a distributed internet application, how may those VRMIs be played collectively? Finally, the resemblance between human-controlled avatars and computer-controlled agents (most commonly pre-captured motion sequences) makes an apt case to study the notion of agency in music. How should a virtual agent react or respond to the actions of the user to simulate the co-creation happening in collective improvisations or chamber music?

Objectives overview:

The main focus of the research consists in using FluCoMa to build models for Virtual Reality Musical Instruments (VRMIs) presented to the user in the form of collective experiences simulating chamber music, combing humans or virtual agents. Objectives (O) are outlined as follows: **O1/** Specification of interfaces for input gestures, together with multimodal feedback. O1 aligns with the field of NIME (New Interfaces for Musical Expression, see state of the art), Work Packages 1 & 2, and *activators* in section 1.2.2. **O2/** Specification of the sound corpus: the range of possible acoustic responses of the system and its organisation in the form of a sound map. O2 consists in improving CBCS resynthesis to obtain plausible acoustic instrument simulations. **O3/** Specification of how agents (human or virtual) may interact in NMPs and the extent to which such performance relates to the concert experience from an audience perspective. Putting emphasis remote interaction and considering the agent a half-way entity between human and computer, O3 shifts the focus from Human-Computer Interactions (HCI) towards the emergence of computer-mediated interactions between agents/performers.

State of the art:

The project's objectives reveal its interdisciplinary nature and allow us to outline three distinct areas for shaping COSMIX's state of the art.

1/ NIME and XR: The field of NIME focuses on innovative interfaces and technologies, and gathers researchers, artists, musicians, and technologists sharing their knowledge on new musical interface design. With 8 publications at the NIME conference dedicated to virtual, augmented and mixed reality between 2011 and 2019, the realm reveals a steep growth with 17 new proposals between 2020 and 2022. Beyond the NIME community, Turchet, Hamilton[4] and Çamci (2021) have undertaken one the most comprehensive overarching studies of the realm in "music in

---

[4] Hamilton being one of the rare oft-cited authors in the field to achieve artistically convincing VR concert pieces/improvisations

https://youtu.be/TYmuIZqY7BE

extended realities", listing 260 publications and 200 works, and yet acknowledging their still fragile artistic output: "For Musical XR to gain a stronger foothold in the artistic landscape, composers and performers will need to envision long-standing practices in this domain". The pioneer research of Stefania Serafin in Sonic Interaction in Virtual Environments (SIVE) will be examined in section 1.2.3. Several ERC-funded projects such as *Meta-Gesture Music: Social, Embodied, Interactive Musical Instruments* (PI Tanaka, Goldsmith, London) and *RapidMix: Realtime Adaptive Prototyping for Industrial Design of Multimodal Interactive eXpressive technology* for sound gesture interaction and user-centred design will also provide elements of response. The industry, finally, exposes a few musical games on VR commercial platforms (Steam and the Occulus store): Virtuoso, Exa the infinite instrument, and Electronauts propose customisable emulation of DJ setups in VR for Electronic Dance Music (EDM) production. Maestro invites the user to simulate the gestures of a conductor. Patch XR and SynthVR, finally, share the metaphor of a modular synthesiser which the user can connect in various ways. The originality of PatchXR Creative coding Environments (CCE) lies in its ability to expose high-level immersive experiences to the general public as well as enable mid and low-level prototyping for advanced users, following the syntax of the visual programming environments Max and Pure Data, but expanding the scope to sound, movement, and 3D graphics.

2/ Corpus-Based Concatenative Synthesis and Machine learning applied to audio Among the principal CCE libraries for CBCS and machine learning (ML), the following taxonomy can be drawn:

A/ In Max/MSP: the Max ecosystem, developed by Miller Puckette in the 1980s at IRCAM, remains today an efficient solution for rapidly connecting Graphical user interfaces and experimenting with CBCS and/or machine learning. 1/ mubu: IRCAM's ftm and its successor mubu expose extensive possibilities for ML implementation but are most specifically targeting gesture recognition. Diemo6 Schwarz's CataRT, widely acknowledged as the pioneer tool for CBCS is now fully implemented in mubu. Despite the clear porosity between ML and such data-visualisation-inspired methods, CataRT has only sporadically implemented ML algorithms. 2/ Ml.*, by Benjamin D. Smith, is situated at the other hand of the spectrum as it exposes a minimal machine learning library, with nevertheless elegant implementations of Markov Chains, Hidden Markov Models, MultiLayer Perceptron (MLP), and unsupervised methods such as Self-Organising Maps (SOMs). 3/ dada. Ghisi's library for computer-aided composition allows for manipulation of databases, 2d plotting and distance calculation circumscribed to the symbolic domain (musical scores). 4/ FluCoMa: The ERC-funded project led by P.A. Tremblay presents similarities to the aforementioned libraries, with a focus on signal decomposition and machine learning. An accessible yet versatile and powerful package for creative MIR, FluCoMa's efforts in outreach and community building through detailed pedagogical resources may help bring together a community which has partly failed emerging around MuBu at IRCAM. 5/ SOMAX2: An environment for human-computer co-improvisation, SOMAX2, unlike its predecessor Omax, is corpus-based, and thus shares similarity with FluCoMa, with a focus on machine improvisation. G. Assayag, P.I. of the REACH ERC-funded project, describes it as follows: "A metaphor I like to use is that of the map and the territory. The idea is this: the musical style is like a territory"

B/ Outside Max: 1/Mosaique: An emerging project at the University of Montreal, Mosaique is built on the top of Max/MSP, OpenGL/Jitter and FluCoMa. The FluCoMa dependency together with its focus on 3D makes it a project strikingly close to COSMIX. 2/AudioStellar ressembles CataRT and FluCoMa but targets a higher-level user experience. Its processing pipeline (MFCC analysis, t-SNE dimension reduction) relates to those most commonly used in FluCoMa. 3/Audioguide: The software by B. Hackbarth differs from other CBCS softwares by its absence of Graphical User Interface and its original focus on non-realtime morphology 4/SuperCollider: A few libraries in SuperCollider attest of the desire in the community to make a creative use of MIR techniques: SCMIR, by Nick Collins (2011), and MIRLC (A. Xambo), Artemi-Maria Gioti, a renown researcher/composer involved in the MusAI ERC funded project, also uses supercollider in most of her work 5/ Wekinator: Based on the idea that "small data is beautiful data", Wekinator has inspired many artists and encouraged the use of ML to control digital synthesis, most commonly mapping the x dimensions of an input gesture to the y dimensions of a synthesiser. It has no direct to CBCS.

3/ Networked Music Performance: Foreseen by Don Foresta, who exhibited in 1986 at the Venice Biennial one of the first computer networks between artists, networking technologies were soon applied to the realm of computer music by The Hub in 1997 in the bay area. Meanwhile in research/academia the field enjoyed sustained interest, with as leading figure Chris Chafe, appointed director at CCRMA in 1996, who famously developed JackTrip and published extensive research on low latency audio and real-time internet collaborations (also coined telematic performances). Belfast's SARC (F. Schroeder) also produced significant research output in the field. The genre kept going with a few enthusiasts, and saw an obvious renewed interest during the covid years, culminating in the quarantine sessions between, among others, Orpheus and CCRMA. The E.R.'s research and artistic output has long focused on the notion of networked music performance, in local settings (over Wi-Fi), foreseeing performances in the cyberspace when the technology would allow it.

Extent to which the objectives are ambitious and go beyond the state of the art: Reaching XR with FluCoMa:

FluCoMa dedicates a large part of its scope to the building of sound maps or "intelligent models", a concise choice responsible for its robustness and relatively accessibility to musicians with no priori knowledge of data science, but whose utility will remain abstract for users not familiar with CBCS and NIMEs. The E.R sees a potential in exposing this knowledge to a wider community practice-based music researcher through user-level user experience in XR. **A shared interest in 3D among the FluCoMa community:** FluCoMa exposes both 2D and 3D map representations. Harvard Professor Hans Tutschku, freelance composer and researcher Balint Laczko, and researcher Dominique Thibault (Mosaique) have proposed prototypes of 3d visualisation within the Max jitter environment. VR outperforms Max in terms of 3d rendering and interaction. PatchXR, therefore, opens a prolific research potential. **Academia:** Computer music Kadenze MOOC such as those delivered by Perry Cook, Rebeka Fiebrink, or Matt Wright, to name a few, aptly exemplifies how music and computer science are today increasingly taught in interdisciplinary curricula. With the rise of ubiquitous AI, the FluCoMa shows high potential in that regard: drawing on XR workshops initiated at his home institution, the ER seeks to encourage music technology students to dive deeper in creative coding with ML.

Originality and innovative aspects of the research, advancements within the field

The ER aims to demonstrate that creating instrumental models that closely mimic human behaviour surpasses the current capabilities of Virtual Musical Instruments (VMIs) typically embedded in digital audio workstations. Common VMIs available to the general public today can convincingly simulate reverberated "strings", but will unmistakably sound artificial when attempting to simulate a solo violin in a dry space. Emerging generations of musicians are getting accustomed to the poor quality of their expression, which spurs to pursue the path opened by projects like REACH and FluCoMa, where CBCS helps simulate incomparably more human expression. The use of CBCS will prevent the caveat faced by physical modelling VR-implementations (e.g. Hamilton's double bass) in which HCI in VR favours a linear sound-to-gesture mapping (e.g. by plucking and modifying the string's length), which arguably leads to simplistic gestural performance and constrains artistic expression. In comparison, the sound palette of CBCS is richer, but will demand more strategy in mapping to allow plausible control through gesture.

1.2 Soundness of the proposed methodology

1.2.1 Overall methodology:

The overall methodology outlines a model paradigm that will help devise four self-contained work packages (WP1-4 entitled "TEAL" for Tangible, Embodied, Automat, and Listener) related to O1 and O3. Two additional packages (WP5-6) deal more specifically with O2 and CBCS.

Low/High interruption tolerance paradigm: The research methodology of articulates four types of "instruments" classified across a continuum inspired by Malloch and Wanderley measuring their autonomy towards human control, thus distinguishing three types of Human-Computer Interaction (HCI): 1/ "low interruption tolerance paradigm" in which, as in classical instrumental playing, the gesture's energy is directly translated into sound, and 2/ "high interruption tolerance paradigm" where the model generates algorithmic music on its own, or can even listen and react to another agent :

Low interruption -High interruption

Tangible - - - - - - - - - - Embodied - - - - - - - - Automaton - - - - - - - - - Listener

Work Packages 1-4: TEAL NIMES: Tangible, Embodied, Automat, Listener

WP 1 Tangible: sound map query via tangible interface Three principal modes of interaction will be investigated: 1/ collision of objects with the controller (buttons, pads, or more evocative 3d meshes, with each item representing either a sample/data point, or a cluster) 2/ displacement of a grabbable object, affected or not by gravity. 3/ friction: how rubbing/scratching can help be evocated in VR Albeit impalpable in VR, the combined perception of visuals 3d meshes with haptic feedback creates a plausible illusion of tangibility. Section 1.2.2. will explain that excitation methods are corpus independant. The 3D textured mesh responsible for the excitation of a flute corpus is therefore more likely to be represented by bellows or a tangible surface rather than by a flute.

WP 2 Embodied Inspired by Tanaka's 30-year-long exploration of the myogram, Laeticia Sonami's artist keynote at NIME 2014 at Goldsmith, and the research undertaken by R. Fiebrink in Interactive Machine Learning (IML) in the same university, embodied interaction will focus on gestural tracking of hands and head without the medium of a tangible object. Previous experiments with *theremins* (see footnote 5, embodied interaction) have raised important issues relative to feedback, as the user cannot rely on visual cues to predict the sounds he/she will trigger, and performs, so to speak, in the dark, "touching sounds" through oral and haptic stimuli. Further investigatation

embodied interactions by tracking gyroscope information (pitch yaw roll, or x y z rotation) yielded by hand-controllers, which prompts from the user a fully embodied exploration of wrist muscles activation. By extension, and informed by the research of R. Fiebrink on interactive machine learning, more dimensions will be added to the input vector (for instance, head position, head rotation, distance between head and left and right controllers…) in order to train an mlp regressor (or classifier) exposed in Flucoma.

WP 3 Automata: An important distinction with instruments designed in WP1 & 2 is that an automaton continues playing after the parametric controls have been tuned. A straightforward application of a FluCoMa sound map to VR can be achieved by chaining with the help of Markov processes the k samples corresponding to the k nearest data points selected via the x y z coordinates of a 3D theremin. From there more complex compositional rules can be implemented such as preset management and interpolation. The mention of Markov processes draws attention to one specificity of REACH-related softwares, which concerns the ability to model sequences of data, which is required when building autonomous agents. SOMAX2 and DICY2 inherit from Omax in that regard music pattern analysis and generation based on Factor Oracles, which will inspire higher-order Markov Chains in the ER's implementation to capture dependencies and patterns in sequential data for predictive capacities.

WP 4: Fully autonomous Listener: Whether embodied by an avatar or merely triggered by control messages, an instrument (e.g. based on a flute corpus) reacting to another (e.g. a piano piece corpus) will constitute an elaborate cyber-human type of interaction. REACH-related software implementations will be important references here. WP4 will imply the mapping of non-overlapping timbral spaces (e.g. when a clarinet corpus needs to react to a piano corpus). Amplitude and pitch tracking will convey a sense of immediate reactivity of the listening module, but musical change happening over longer periods of time will be realised via K Means clustering, thus defining a simple compositional rule according to which both instruments will change clusters at the same time.

Work Packages 5-8: Corpus Manipulation

WP5 Mining: Data mining applied to musical material requires thorough knowledge and "know-how". Based on the accessible and informative documentation of the FluCoMa project, the ER's workflow consists so far of using umap dimension reduction over MFCC analysis. This standard pipeline, extensively tested by the E.R., has proven general purpose for the distinction of timbral categories, but a more refined study of the toolkit will allow a more custom use of segmentation, descriptors, analysis and dimension reduction, tailored to the type of audio the model is fed with. One major tension has been identified between 1/ short samples revealing accurate timbral categories and 2/ long samples performing more realistic resynthesis.

WP6 Retrieval: The exploitation of machine-learning empowered analysis described in WP5 requires the implementation of optimization solutions for nearest neighbour search queries in PatchXR. An ICMC publication (Bell 2013) described a simple solution based on Euclidian distances providing the nearest point, but the kd-tree object exposed in FluCoMa would allow better search performances as well as the ability to retrieve K neighbours (Bell 2013, figure 6), crucial to all TEALs with the exception of first tangible category.

WP7 Resynthesis: Undesired artefacts inherent to CBCS can be minimised through the implementation of cross-fades and polyphony. Such methods will depend on the type of sound material, leading to convincing results with resonating instruments such as strings and percussions, but will scarcely apply to monophonic instruments (e.g. saxophone). Both REACH and FluCoMa software implementation strategies will be examined in that regard. All TEALS but the first (tangible) deal with sequential data, addressing the need for Markov Chain implementation in PatchXR, in which only Bernoulli processes are exposed so far.

WP8 Velocity: based on findings gathered from WP1, 2, 5 and 6, the ER will seek interactions carrying the embodiment of velocity, missing in former implementations.[5] Combining dimension-reduction-enabled timbral control with velocity control holds a contradiction in the sense that loudness (velocity) has an incidence on timbre (the louder the brighter). An ideal implementation would allow velocity to have incidence on retrieval (WP5), but using a low-pass filter in resynthesis (WP6) might be more realistic.

1.2.2 Integration of methods and disciplines to pursue the objectives

From the iterative implementation of the eight work packages will gradually emerge an important milestone: the final implementation specification of the COSMIX toolkit in the form of three deliverables, each responding to one of the challenges expressed in the objectives. Deliverables will not be instruments, but building blocks for instrument and

---

[5] Early FluCoMa uses in XR fell in the categories of WP1/ tangible buttons https://youtu.be/L-MiKmsIzjM WP2/ embodied interaction/theremin https://youtu.be/cM-utIxv2Ww and WP3/ automata https://youtu.be/DC_BL_HGPLA

interaction prototyping. The three classes of high-level user interfaces will follow the paradigm of visual programming (3d blocks to be connected to one another).

Deliverable 1, Tangible and embodied Activators:

All Activators (input) are meant to be connected to a sound map (a corpus) in order to produce any sound. **Tangible objects:** Informed by the research on displacement, collision and friction of WP1, tangible objects will take various visual shapes according to the type of expected interaction, 1/ Strings may be struck or bowed at a certain velocity. 2/ Sticks: inspired by the work of Malloch and Wanderley (t-stick), sticks will implement gravity, so that the object can roll on a surface, be held in hand, thrown in the air, or played like a vahila harp, with tuneable strings around it. 3/ Sabres: optimised for short samples tightly matching the input gesture, sabres are best suited to accelerometer data, for large dynamic palette 4/ Continuous interaction: PatchXR exposes several high-level objects which prompt the user for diverse gesture-based interaction such as pressing, scratching, pooling, and circular motion to name a few. **Tangible-visualisers** Based on the plotter paradigm inherited from CataRT and FluCoMa, those activators are literal visualisation of sound maps. Their haptic and visual feedback provides clarity in the sample selection process, but lack the variety allowed by higher-level processing of information of tangible activators. **Embodied-visualisers** are inspired by latest versions of CataRT and Balint Laczko's *The Hum*, in which lines are drawn between the query point and K nearest neighbours: the user by moving his hand, dynamically change the pool of k samples at his disposal, which he can then trigger via wrist rotation for instance. **Embodied-activators** situate in the continuity of theremins or worlds in which e.g. left hand position is directly mapped a point in the sound map. The right hand can then control triggering envelope. **Embodied-MLPs** will be exposed to the user as a mirror, in front of which the avatar records a few postures then used for classification or regression.

Deliverable 2, SoundMaps: SoundMaps load a given corpus (an audio file) as well as metadata related to timetags, timbral similarity information (1D, 2D and 3D umap), descriptor-based information (pitch, loudness, centroid) as well as labels (class 1, class 2…). SoundMaps contain considerable information about each segment of sound, but do not represent them graphically unless connected to a "tangible cloud" visualiser. Visually, SoundMaps will be blocks taking the form of a 3d mesh best representing the corpus of the acoustic instrument being analysed (e.g. clarinet).

Deliverable 3, Agents: Objective 3 has addressed the crucial question of agency, latent in FluCoMa, and most explicitly developed in the REACH project.

Automata: An automaton has the ability to continue playing after the user has set its parameters, and differs in this sense activators. It operates in time via Markov processes, predicting on its own or with user's input the sequence of samples to be played.

Full-agent: Full agents extend automata by their ability to listen to *influencers* (a term used in REACH/Somax2), to which they respond by activating another sound map, given specific constraints or thresholds set by the user. Their behaviour implies to listen to: control data from the user (activator), an incoming audio stream (sound map), and higher-level symbolic information (e.g. which cluster the *influencer* is currently playing, also provided by the sound map), in order to respond or stop accordingly.

1.2.3 Gender dimension and other diversity aspects

The state of the arts has only briefly mentioned Stefania Serafin, whose pioneer works and authority in the field will be an invaluable source of information. Trained at IRCAM, Serafin holds a PhD from Stanford/CCRMA on physical modelling of bowed strings instruments, supervised by Julius Smith. Pioneer in the field of virtual instrument simulation in VR, Serafin foresaw in 2016 that: "instruments that include a simulated visual component delivered via a head-mounted display or other forms of immersive visualisation have not yet received much attention". With over 5000 citations, Serafin worked with Eduardo Fouilloux at Aalborg university on the conception of Mux, which later became PatchXR. In Sonic Interaction in Virtual Reality (2023), Serafin et al. focus on binaural sound rendering for immersive environments, an important feature implemented in PatchXR in 2020. In the book Sonic Interaction in Virtual Environments (SIVE) co-edited with Geronazzo (Geronazzo, 2023), Serafin presents results from seven editions of IEEE VR workshops she organised on the topic.

1.2.4 Open Science Practices

Based on former experience and research in web technologies, the ER will build a free-access website (Work Package 8) gathering sound maps from FluCoMa users, as well as sound bank from volunteer musicians/artists, with a mention of the instrumentalist, the author of the piece and of the copyright agreement. The initiative will make 3d representations of sound corpora available to the general public. Prototyped in node.js, the website will use plotly.js for display and howler.js for sound. Like the fluid.plotter object in FluCoMa displays 2-dimensional data, the website

will display sound corpora in 3d, with on optional export to the .patch format, for direct export to the PatchXR web portal (commonly used by users to load assets) and .fbx for lower-level export to Unity.[6]

1.2.5 Research data management and management of other research output

The elaboration of the aforementioned website will expose 3-dimensional sound maps, focusing on acoustic instrumental material, most commonly based on pre-existing solo pieces by living composers. An important source for data collection here will be the Babelscore database, which contains around 1200 scores for solo instruments, often with a recording. WP 9 will consists of obtaining copyright agreements from the people involved in the production of the sound corpus (composer, instrumentalist, and FluCoMa corpus extractor). As proved by publications at ICMC23, TENOR23, and JIM23, the research findings will be disseminated via academic publications at major international conferences, in th general field of computer music (ICMC, SMC, CMMR), as well as more specialised ones (TENOR, ISMIR, NIME, SIVE), themselves preparatory to a submission to the Leonardo Journal.

1. 3 Quality of the supervision, training and of the two-way transfer of knowledge between the researcher and the host

1.3.1 Describe the qualifications and experience of the supervisor.

1.3.2 Planned training activities for the researcher…

The training activities of the ER will primarily concern the mastery of the FluCoMa toolkit, working closely with the supervision Pr. P.A. Tremblay, essentially on WP5 6 and 7 to extent the possibilities of Corpus-Based Concatenative Synthesis. Other training activities include expert feedback from FluCoMa developers (Gerard Roma, Owen Green), PatchXR developers (Edo Fouilloux, Victor Beaupuy), SIVE expert Stefania Serafin, aesthetic assessment (Bryn Harrisson), in form of one-to-one meetings.

Learning through frequent engagement with the core members of the FluCoMa team, via one-to-one conversations but also through FluCoMa's discussion forum, the ER will participate in an active maintenance of the toolkit, required now that the FluCoMa funding has come to an end and some of its developers work outside of the University of Huddersfield. Through punctual interventions in Harker and Tremblay's classes, the ER will have the opportunity to present his research to music and music technology students within the university of Huddersfield. Testing early prototypes through remote collaborations, the students will discover an original iteration of the FluCoMa project with cutting edge VR ongoing developments, giving the E.R. invaluable feedback on the design of his instruments, corpora, and toolbox. With his long-term vision to further develop the COSMIX project within the AMU/PRISM-CNRS environment after the fellowship, the E.R. will also seek to foster Marseille-Huddersfield remote collaborations, rehearsals and performances.[7]

The research on Cyber-Human co-creativity, or notion of agency, at the core of the REACH project, will inform scenarios in which an automat (whether or not embodied in an avatar) has the capacity to respond in real-time to the actions of the user. A secondment of a duration of four months is therefore planned at IRCAM in the Représentation Musicales team in order to implement forms of agency in PatchXR. The RepMus team at IRCAM is led by Gérard Assayag, who is also PI of the REACH project. A 3-month non-academic placement is also envisaged at PatchXR, starting after the end of the scholarship. The purpose will be to work on the native implementation of machine learning algorithms such as KD Tree and other strategies for nearest neighbour query (WP6).

1.4 Quality and appropriateness of the researcher's professional experience and skills

Rooted in the ER's solid classical training in some of Europe's best conservatoires (Paris CNSM and London Guildhall School), the ER's background as a fellow artist member of the French Academy of the Arts makes a compelling case for developing meaningful collaborative music experiences. His combination of artistic and technical expertise as well as research records constitutes a solid ground for an interdisciplinary approach. The ER studied at IRCAM in 2014-2016, where he was exposed to many ideas that influenced the current proposal. During a collaboration with the ISMM (Interaction Son Musique Mouvement) team, he could work and interact with Norbert Schnell, Diemo Schwarz and Frédéric Bevilaqua whose scientific contribution and software developments have substantially influenced the FluCoMa project.

2. Impact

---

[6] Node.js prototype: https://youtu.be/-YiP7I-pf-k; Unity prototype https://amubox.univ-amu.fr/s/JYxk7LsPk8AHgtQ?

[7] A first remote performance (Paris, Marseille, Berlin) was presented at JIM23 https://youtu.be/npyfwqN02qE

2.1 Credibility of the measures to enhance the career perspectives and employability of the researcher and contribution to his/her skills development

Born in Aix, the ER began working at Aix-Marseille University and the PRISM-CNRS[8] laboratory in 2017, where he progressively built a local reputation, getting involved in the organisation of CMMR 2019, teaching at PRISM's dedicated Acoustique et Musicologie Masters, culminating in the attraction and organisation of the international TENOR 22 conference (Technology for Musical Notation and Representation), after having published in the same conference every year since 2017. Working from 2017 with Risset's team has had significant impact on future career prospects of the applicant, and from there his artistic career and research output converged toward the investigation of innovative ways to combine music, science and technology. While presenting a paper at ICMC in New-York (2019), the E.R. witnessed the presentation by P.A. Tremblay of ongoing the FluCoMA project which convinced him of the necessity for a creative use of MIR and machine learning in the audio domain. The residency in 2022 in Marseille of Diemo Schwarz, inventor of corpus-based concatenative synthesis, acted as a catalyst for the current proposal, the concert at the *Planetarium, Observatoire de Marseille* being an important source of inspiration for the COMSMIX acronym. After following Schwarz's expert technical guidance on the use of the FluCoMa toolkit, it soon became clear that the unprecedented artistic affordances and intelligence of such systems should be taught in the Master Acoustique et Musicologie. The extension of this domain to XR, with its incidence on multimodal perception, is also perfectly in line with PRISM's vision: Kronland and Serafin worked together at Stanford CCRMA in the early 2000s, and have often collaborated since, Serafin being regularly invited for PRISM's doctoral students' defences. The funding of the COSMIX project would, therefore, secure a permanent position there.

2.2 Suitability and quality of the measures to maximise expected outcomes and impacts, as set out in the dissemination and exploitation plan, including communication activities

The E.R. seeks to maximise and sustain the impact of the COSMIX project through an intertwined synergy of research publications, technological development, teaching and artistic output, during the return phase at PRISM-CNRS. The workshop[9] realised in 2023 with the acoustique et musicologie master students showed indeed vivid interest among students, but also presented the potential caveat of falling into mere consumerism. This is where the COSMIX project endeavours to design and subsequently teach a more sophisticated approach to the environment, introducing the aforementioned techniques of automatic signal decomposition, corpus manipulation and audio-tailored machine learning algorithms, and developing scripts to fluidly automatise the transfer of information from an environment (FluCoMa) to the other (PatchXR). As demonstrated in publications and performances at Tenor 23, JIM 23, and ICMC 23, the "musical metaverse" is an intriguing medium to convey artistic research exploring today's disruptive possibilities of Musical Information Retrieval, machine learning and HCI. Early public performances with the system displayed musicians performing remotely (from home). This lets us foresee complex scenarios in which different tasks can be distributed among the performers, potentially competing with what can be achieved on a real stage. Dissemination and communication activities will therefore prioritise dissemination through the paper+demo/performance format, widely spread across today's major international computer music conferences. Frequent communications, most specifically among the SIVE community, will ensure the adequate management of intellectual property. Yet for dissemination of the toolkit through the community, the strategy for dissemination will be that of *open source* user-generated content, as opposed to the NFT monetisation of instrument currently envisaged by PatchXR.

2.3 The magnitude and importance of the project's contribution to the expected scientific, societal and economic impact

Envisaged as a landmark experience in the E.R.'s development, the COSMIX project has chances, if accepted to grow as a significant part of PRISM's identity[10]. An inquiry encompassing the study of timbre, data visualization/machine learning, HCI and immersive technology driven by the poetic analogy between cosmology and

---

[8] The PRISM laboratory (Perception, Representation Image Sound Music) can be conceived as Jean-Claude Risset's last achievement. An assistant of Max Matthews at Bell labs in the 1960s, and widely recognised as one of fathers of computer music along with his close friends and colleagues John Chowning and David Wessel, Risset famously contributed to the understanding of timbre perception.

[9] https://vimeo.com/832201597

[10] The young interdisciplinary lab, in no small part modeled on IRCAM, brings together researchers in "hard sciences" (physics and computer science) and humanities (music, cinema, visual arts), to reflect on a common object: music technology, with a focus on digital sound synthesis, special immersion, and multimodal human perception.

sound maps has a potential to interest the Leonardo Journal[11] thus contributing to the laboratory's scientific impact. For the social and economic counterpart, the privileged exposure PatchXR has gained through support and endorsement by *meta*[12], in line with MSCA's recent focus on industry, is likely help the COSMIX project reach large audiences.

3. Quality and efficiency of the implementation

3.1 Quality and effectiveness of the work plan, assessment of risk and appropriateness of the effort assigned to work packages

Following Thor Magnusson's humanist ambivalence towards progress perceived in his ERC-funded Intelligent Instruments project, the research is driven by the search of meaningful musical interaction going beyond the proof of concept of latest technological achievements. Yet, the project has a strong technological component and the general specification and syntax of the COSMIX toolkit defines a major milestone. Its modularity is likely to take the three-part architecture discussed in 1.2.2 (Activators, Sound maps, agents), but the paradoxes it contains may ask for adjustments, and needs to take place during the last third of project (Sept. 2025): 1/ the TEAL packages described in the methodology outline a classification which contradicts tri-partition: are automata and listeners considered activators or agents? 2/ for optimization purposes (the display of too many data points makes the app crash), sound maps by default do not represent data points unless connected to "tangible-visualiser", a specification also subject to amendment. 3/ Is an instrument defined as the coupling between an activator and a map?

	May 24		Sept.		Jan.			May 25		Sept.		Jan.	
Month:	1 2 3 4	5 6 7 8	9 10 11 12	13 14 15 16	17 18 19 20	21 22 23 24
WP1-4 TEAL	| WP1 Tangible | WP2 Embodied | WP3 Agent | WP4 Listenner |
WP 5-9	| WP5 Mining	| WP2 Retrieval	| WP7 Re-synthesis | WP8 Velocity
Milestones	|JIM|	| Final specification |
Deliverables	1/ Software:	| Activators | Sound Maps | Agents |
	2/ Publications:	|ISMIR24	|SMC25	|CMMR |submission to
	|NIME24	|ICMC24	|NIME25	|Leonardo Journal

3.2 Quality and capacity of the host institutions and participating organisations, including host arrangements

As part of the training, and given the strength of contemporary music at Huddersfield (the city holds the Huddersfield Contemporary Music Festival, HCMF, one of the world-leading festivals in the domain), the E.R. will presents his ongoing research project to Graham McKenzie (HCMF chief executive[13]) and Professor Bryn Harrison from the music composition department in order to advice on the aesthetic dimension of the project. The music of Professor Bryn Harrison has long drawn the attention of the E.R. and his prominent use of repetition, in soft dynamics and on large temporal scale makes an apt model for computer-simulation with FluCoMa. Harrison examined the E.R.'s composition doctorate in 2016, where he praised its artistic value: "There is no doubting the high quality of this music, which is often hauntingly beautiful and colourful" and provided him with thorough feedback on the music of Morton Feldman and its unfolding in time, which profoundly nourished the E.R.'s development both as a composer and as researcher.

References

[11] Published by the MIT Press, The Leonardo Journal covers the application of contemporary science and technology to the arts and music. Its executive editor, Roger Malina is former director of the Observatoire Astronomique de Marseille Provence.

[12] https://patchxr.com/blog/best-of-2022/

[13] https://hcmf.co.uk/university-of-huddersfield-awards-honorary-doctorate-to-graham-mckenzie/

Michele Geronazzo, Stefania Serafin (2023) "Sonic Interactions in Virtual Environments", Human–Computer Interaction Series (HCIS), Springer, open access.

S. Serafin, M. Geronazzo, C. Erkut, N. C. Nilsson and R. Nordahl (2018) "Sonic Interactions in Virtual Reality: State of the Art, Current Challenges, and Future Directions," in *IEEE Computer Graphics and Applications*, vol. 38, no. 2, pp. 31-43, Mar./Apr.

Tremblay, P.A. (2021) Pierre Alexandre Tremblay, Gerard Roma, Owen Green; Enabling Programmatic Data Mining as Musicking: The Fluid Corpus Manipulation Toolkit. *Computer Music Journal* 2021;

L. Turchet, R. Hamilton and A. Çamci (2021) "Music in Extended Realities," in *IEEE Access*, vol. 9

Schwarz, D. (2000) A System for Data-Driven Concatenative Sound Synthesis. *Digital Audio Effects (DAFx)*, Dec 2000, Verona, Italy. pp.97-102.