A Meta-interactive Compositional Approach that Fosters Musical Emergence through Ludic Expressivity

MÁRIO ESCARCE JUNIOR*, Lancaster University, UK
GEORGIA ROSSMANN MARTINS*, Phersu Interactive, Brazil
LEANDRO SORIANO MARCOLINO, Lancaster University, UK
ELISA RUBEGNI, Lancaster University, UK

The concept of gamified interactive models and its novel extensions, such as playification, has been widely approached in order to engage users in many fields. In fields such as HCI and AI, however, these approaches were not yet employed for supporting users to create different forms of artworks, like a musical corpus. While allowing novel forms of interactivity with partially-autonomous systems, these techniques could also foster the emergence of artworks not limited to experts. Hence, in this paper we introduce the concept of meta-interactivity for compositional interfaces, which extends an individual's capabilities by the translation of an effort into a proficiency. We present how this approach can be effective through a novel system that enables non-experts to compose coherent musical pieces through the use of imagetic elements in a virtual environment. We conduct experiments with a population of musical experts and non-experts, showing that non-experts were able to learn and create high-quality musical productions through our interactive approach.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**; • **Applied computing** → *Arts and humanities*; • **Computing methodologies** → Artificial intelligence.

Additional Key Words and Phrases: interactive musical systems, emergent art, procedural content generation, meta-interactivity

ACM Reference Format:
Mário Escarce Junior, Georgia Rossmann Martins, Leandro Soriano Marcolino, and Elisa Rubegni. 2021. A Meta-interactive Compositional Approach that Fosters Musical Emergence through Ludic Expressivity. (September 2021), 32 pages.

1 INTRODUCTION

Producing a coherent musical corpus in an emergent fashion is a very challenging effort for interactive musical interfaces and AI systems to overcome. Since music is a form of expression inherently associated with feeling and sentiment, it is essential for developers to conceive devices capable of supporting and transforming an expressive motivation of an individual into a sound structure that the general public might acknowledge as a human-made musical piece. In addition, partially autonomous systems that propose to support the creation of music usually demand artists and sound designers a previous knowledge in determined AI techniques, such as Machine Learning, in order for the model to be trained according to their own goals, as we can observe in Roberts et al. [55]. As for the interactive musical

*Both authors contributed equally to this research.

Authors' addresses: Mário Escarce Junior, m.escarce@lancaster.ac.uk, Lancaster University, Lancaster, Lancashire, UK; Georgia Rossmann Martins, georgia@phersu.com.br, Phersu Interactive, Belo Horizonte, Minas Gerais, Brazil; Leandro Soriano Marcolino, l.marcolino@lancaster.ac.uk, Lancaster University, Lancaster, Lancashire, UK; Elisa Rubegni, e.rubegni@lancaster.ac.uk, Lancaster University, Lancaster, Lancashire, UK.

interfaces, aside from the entertainment-based ludic approaches from video games [16, 66], sometimes they require a good understanding of musical theory (and sometimes practice) for the user to co-create pieces that can be recognized as "music". Or yet, the interactive model reduces the compositional experience to mini-games [26, 62], in which the user's creative expressiveness does not produce a result, restricting the experience to an instant entertainment, where the usage of sounds only serves as a scoring system element. That is, the user is not really creating a new music, but trying to mimic/play an existing one according to a scoring system. Thus, in many situations, non-expert users might feel discouraged and unmotivated to have a quality creative experience with an interactive musical system.

Another obstacle in the development of emergent content approaches is finding a perfect balance between the user's freedom to express himself/herself musically and the constraints of the co-creation algorithm, like the one we are going to propose in this work, in the attempt to foster the musical product of the interactive system to be coherent and homogeneous according to a given number of variables (e.g., tone, rhythm, tempo, etc). In our case, since it is intended that a non-expert also have a satisfactory experience with our system, the efforts must be centered on providing algorithms that work underneath the player's experience layer, trying to accurately fit the user's input in a *temporal musical structure* in a discreet fashion, improving the quality of the musical outcome while preserving the user's expressivity and original intent. That is, the algorithm should be able, regardless of how the notation system will be presented (whether it is visual, audible or tactile), to allow the user to transform his/her ideas with a good accuracy through the interactive model.

Many researchers study the interaction of human agents with partially autonomous systems for emerging artworks. For instance, Jacob and Magerko [32] propose an approach where a human and an agent collaborate to produce movement-based performance pieces. Davis et al. [11] describe a system where a user takes turns with an AI system when drawing in a canvas. Similarly, many contemporary systems produce musical pieces by collaboration between humans agents and AI agents. In all previous works, however, human agents need expertise in order to produce high-quality outcomes.

Thus, in this work we introduce the concept of meta-interactivity, and how it can establish a relation with different expressive endeavors in order to homogeneously produce a desired outcome. This approach can be seen as a powerful gamified tool that can be used to explore a user's creative proficiencies, such as drawing, and extend its potential to other expressive fields, such as music production. To explore this concept, we present a Virtual Reality Musical Instrument (VRMI) called Bubble Sounds, a novel interactive art installation that enables non-experts in musical theory and performance to produce an original and pleasant ambient music piece through a friendly interface that allows him/her to use visual elements, such as colors, as tools to sculpt a coherent musical corpus.

Bubble Sounds differs from other approaches in the field in the way we propose the interactive model in which users produce music, as well as in the way it offers a guideline for novice users to produce a coherent musical corpus. In this way, we are motivated not only to propose a novel interactive concept but to also present a Virtual Reality Musical Instrument as a proof of concept in order to verify the efficiency of meta-interactivity. Some of the questions we try to answer in this work are:

- Is meta-interactivity efficient to improve the user experience and support the generation of music?
- Does it successfully accomplish its goal to extend users' proficiencies?
- Considering this approach being employed in Virtual Reality Musical Instruments (VRMIs), are the outcomes generated by this approach acknowledged as high quality music by the general public?

- In meta-interactivity, can visual interfaces combined with sound interfaces create powerful experiences in terms of improving the quality of outcomes generated by partially-autonomous systems?
- Can art practices, such as music composition in the context of VRMIs, be dissociated from the human factor?

In order to tackle these questions, we evaluate our system on its compositional capacity with human subjects, as well as the music produced through it. We show that non-expert users in musical theory and practice were capable to learn and create music as good or even better than experts.

2 RELATED WORK

This is an interdisciplinary work, that benefits from, collaborates and dialogues with different fields. For instance, HCI, AI, game studies, procedural content generation and fine arts are some that provide us fertile soil to discuss the many elements that orbits the meta-interactive approach. However, we highlight that the main contribution of this work, especially regarding the system we will present, resides in the following fields: (i) Virtual Reality Musical Instruments (VRMIs) and Immersive Virtual Musical Instruments (IVMIs), as discussed by Serafin et al. and Berthaut et al. [4, 59]; (ii) studies in interfaces capable to foster efficient devices for the collaboration between human and synthetic agents. In the following, we discuss in detail the related work for each relevant field.

Video game industry. More specifically in the music video games genre, we can find works that explore different forms of interactivity through the collaboration between musical systems and a human agent. E.g., Proteus [36] allows the user to explore a procedurally generated world, where its flora and fauna emit its own musical signature, whose combination generates dynamic changes in the audio output. Vib Ribbon [44] offers a gameplay mechanic where the player interacts with rhythmic obstacles generated by the music that is being played. Similarly, Patapon [38] proposes the user to execute rhythmic sequences on the joystick in order for the characters to proceed across levels. In these examples, however, the space for the user expressiveness is very restricted, since, in the first, the player does not control the music that is being generated. In the other two examples the only action allowed is usually to press determined buttons at predetermined moments.

Tedpole Treble [5] is an adventure game where the player should escape from enemies, and by doing so the character interacts with elements of the level design (presented as a musical sheet) producing its own voice track. The game features the possibility for users to create their own stages in composition mode. Although it allows new musical corpus to emerge, this mode focuses on the replayability fostered by the level editor, not in the musical diversity itself. That is, the obstacles in the level are fully customizable by the user, but the music is not created in real-time (although it presents variations by the repositioning of the game assets). In Child of Eden [50], successor of Rez [49], players interact with the virtual environment through a shooting up gameplay mechanic, where the targets produce melodic lines for a composition upon its destruction. Similarly, Beat Saber [67] is a virtual reality rhythm game set in a futuristic world where the player slash rhythmic enemies as they fly towards the camera. Fract OSC [63] is a music-based puzzle game that allows players to explore an abstract landscape that includes puzzles using platforming and music-based game elements. By completing these puzzles, players can compose an ambient music that continues to play in the environment and unlocks extra content in the game. In these examples, we can observe a huge convergence of the usual gameplay mechanics approaches applied in musical video games, but with a focus on challenges and rule-based interactive models over the expressive creation.

On the other hand, some works rely more on the compositional interface than in the music creation mechanic itself. For instance, Sentris [35] is a musical performance game where users can create their own music by stacking "Sound

Blocks" into a spinning loop, resembling some current audio interfaces for the creation of both music and sound effects. Similarly, many works also foster musical emergence through the collaboration between users and interactive systems in non-game experiences. These generally provide their own form of musical notation for the user to interact with virtual environments, allowing individuals to create musical structures using pre-defined samples. E.g., MusicLab [24] is a web platform that offers a variety of musical experiences, including a sound experience based on the work of the plastic artist Kandinsky. Electroplankton [31] is an interactive sound experience based on a compilation of several mini-games. The user can manipulate creatures called "planktons" in a loop of interaction with the virtual environment in order to produce sounds.

Previous works also proposed interactive mechanics inspired by video game-like input patterns, with a stronger narrative background [43] and a more present rule-based mechanic for emergent music generation given the players actions [34]. In this work, however, we experiment with a different type of environment, fully centered on the musical creation, and thus adopting a more minimalistic gameplay mechanic.

HCI approaches. Interactive design approaches aims to bring daily life experiences into the virtual environment. For instance, the Drift Table [21] is an electronic coffee table that displays slowly moving aerial photography controlled by the distribution of weight on its surface. It was designed to instigate reflection on the ways technology can support ludic activities, as well as improve our general understanding of ludic design. SoundSelf: A Technodelic [18] provides the user a guided meditation experience. Through the usage of voice input, it is possible to explore a sound-based virtual world where modulations of the user's voice creates the feeling of presence in an abstract environment. Previous works also explore possibilities for new musical practices to emerge through novel mechanic design. For instance, Koray et al. [64] use the notion of "cultural probe" – a technique used to gather inspirational ideas from individuals' lives – as input to digital musical instruments. Regarding the production of interactive interfaces for the emergence of music, Electronauts [61] gives the users the opportunity to become a DJ in a VR experience, where it is possible to jam with other users and even professional musicians through the remixing of pre-existing tracks. In music and interactivity, Holland et al. [27] introduce recent research in Music and Human Computer Interaction, also known as Music Interaction, and discusses its role in HCI. According to the authors, current musical activities include a wide range of expressive forms, besides performance and composition; it also includes collaborative music making and human and machine improvisation. From this perspective, specific issues may arise, including: (i) whether Music Interaction should be easy for newcomers; (ii) what can be learned from the experience from those "in the groove", that is, users that already find themselves in a flow of the experience, according to Mihaly Csikszentmihalyi's definition of immersion [8]; and (iii) what can be learned from the commitment of non-experts in musical theory and practice with the system mechanics and interface. In performance, Astaire [70] proposes a collaborative hybrid VR dance game for two players sharing the same environment.

Many HCI works also seek a better understanding of how music impacts players behavior in their experience while playing games. For instance, Rogers et al. [57] investigate whether it is possible to influence video game players on risk-taking behavior through the presence of elements in the background music. The authors concluded that music plays a major role in the immersion of players, as well as in their learning process of the gameplay mechanics. Other works also assess how much music in games contributes to the player experience, more specifically considering VR approaches [58]. In their assessment, authors observed that players tend to perceive time passing significantly quicker when music was being played in the proposed VR experience. This is a very meaningful outcome, given the usual reduced role of audio in comparison to the visual elements in video games (and the audiovisual production). For example, it is common, nowadays, the usage of headsets for team communication in online versus games, reducing the role of music and its impacts in the experience to a second plan.

Additionally, Lucero et al. [39] investigate if playful interactive models based in game elements, such as gameplay mechanics, can support the creation of better engaging experiences. According to the authors, although playful mechanics for interactive experiences are important elements to foster engagement, it is often neglected. In other words, playfulness should be more explored as a feature to improve the quality of many products and processes that goes beyond entertainment-based outcomes.

VRMIs and IVMIs. Works have been exploring different forms in which users interact with virtual musical instruments. For instance, Mäki-Patola et al. [40] introduce and analyze four gesture-controlled musical instruments designed to allow for rapid experimentation of new interfaces and control mappings. Cabral et al. [6] present a novel 3D virtual instrument with a particular mapping of touchable virtual spheres to notes and chords of a given musical scale. The objects are metaphors for note keys organized in multiple lines, forming a playable spatial instrument where the player can perform certain sequences of notes and chords across the scale employing specific gestures. Mitchusson [47] uses the audio module and the Virtual Reality capabilities from the Unity engine to generate, through a dice-rolling mechanic, sample effects and audio mixing to generate experimental music outcomes. Gibson and Polfreman [23] present a framework that supports the development and evaluation of graphical interpolated mapping for sound design. The approach is capable to compare the functionalities of previously developed systems leading to a better understanding of the outcomes these systems are capable to generate.

Innovative devices that work like digital musical instruments were also proposed in the past. For instance, the world-renowned reacTable [33] is a novel multi-user electro-acoustic music instrument with a tabletop tangible user interface. It proposes new engaging ways in which musicians and sound designers can perform and compose music on the fly. Different from many audio interfaces, the application does not use regular input devices, such as a mouse or a keyboard, and presents a more flexible way in which individuals can interact with the installation.

AI approaches. A common approach is to use Deep Learning models to support the generation of music and new composition methods. E.g., Roberts et al. [55] enable the integration of generative models through the use of the Google Magenta Interface to arouse musicians creativity. This system was experimented on a live jazz performance with a piano and drums duet. Also, Huang and Raymond [29] create coherent music with melodic and harmonic structures that can be acknowledged as pieces composed by a human agent. Castro [7] trained Deep Learning models to perform coherent improvisations. Agarwala et al. [1] use generative Recurrent Neural Networks to create musical sheets without predefining compositional rules to the models. Rocksmith [66] presents a mode where it is possible to explore the cooperation between the user and an AI system through a conventional musical instrument, such as an electric guitar. The user can jam with the system, which provides a non-adaptive background track (i.e., it is the user who must adapt to the tempo of the song).

Still AI related, but more specifically in the Generative Art field, many researchers develop AI systems where humans and algorithms cooperate in art emergence. We can observe very inspiring works that foster different forms of interaction such as performance and improvisation through the interaction with partially autonomous systems [14, 51, 52]. For instance, Davis et al. [11] propose an experience where a user takes turn with a computer AI system when drawing in the same canvas. In Jacob and Magerko [32], a human and an agent collaborate to produce movement-based performance pieces using a co-creative agent. These works, however, happen apart from the electronic games development efforts, thus are not focused on providing friendly imagetic interfaces that guides non-experts in any kind of content generation process, demanding them to be properly trained in order for the desired outcome to emerge.

Procedural systems have an enormous power for extending the life spam of an interactive experience since they present new elements to an audience every time they are executed. In this sense, Hoover et al. [28] propose the use of Functional Scaffolding, a developed method that modularizes the structure of a MIDI song and generates autonomously a follow-up based on harmonization to interpret in world matrix of the early stages of Super Mario Bros [48] the tiles assets of the game, as if they were a musical score.

Fine arts. More specifically regarding the music industry and performance, the work of Björk in Biophilia [68] is also a very interesting example of the convergence between music and technology, since it was the first studio album in app format. It contains a wide range of musical genres, such as electronic music and experimental music, and was made on an iPad. It offers a series of applications that Björk described as a collection that encompasses music, applications, the internet, installations and live performance.

Simon Penny [53] analyses common post-cognitive paradigms and makes a parallel of these with general discussions of art and other cultural manifestations. The author argues that computational cognitive rhetoric cannot account for the "quintessentially performative qualities of arts practices", emphasising the way in which modern computing techniques, such as AI, facilitate the emergence of generative and interactive artistic practices. He points out that conventional aesthetics of the plastic arts cannot efficiently address the convergence of artistic manifestations with state of the art technologies, and argues for a new "performative aesthetics" capable to comprehend this new trend.

In general, many works discussed in this section explore a wide variety of approaches from HCI to collaboration between AI systems and human agents. They do, however, demand the user proper training in order for the generative outcome to be "coherent". Also, some examples only allow users to edit pre-existing tracks, not actually allowing users to compose music from scratch. In our work, however, we propose an approach where even non-expert users can compose a music piece that can be acknowledged by the general public as something created by a trained musician.

3 MUSICAL BACKGROUND

In this section, we briefly present some concepts from music theory that are being approached in this work. It is important to emphasize, however, that users do not need to learn and understand these concepts to compose music through our approach. Readers that are familiar with basic music theory concepts can skip this section.

According to some definitions [2, 37, 41, 56], the main concepts that will be discussed further in this work are:

- Note: a note is a single musical sound (e.g. C, C#, D, etc). They can vary in pitch and duration, and its a fundamental building block of music.
- Tone: a constant sound most frequently characterized by its pitch, such as "C" or "D", but which also comprises sound quality (i.e. sound texture), duration and even intensity. In many forms of music, different tones are changed by modulation or vibrato (fluctuations in height and frequency).
- Pitch: refers to how the human ear perceives the fundamental frequency of sounds. Low frequencies are perceived as low tones and the highest as high tones. Adult humans can detect sounds in a frequency range from 20 Hz to 20 kHz.
- Melody: a linear succession of musical notes that can be perceived as a single entity, that when combined generates variations in pitch and rhythm.
- Harmony: a simultaneous combination of musical notes that evolves across time, producing a pleasant effect among listeners. It can be perceived as a base structure of a music, in which the melody comes upon. Along with melody, it is one of the building blocks of the western music.

- Chord: multiple notes played simultaneously, the most commons ate triads (3 notes being played at the same time) and tetrads (4 notes being played at the same time).
- Accident: a quality of a musical note that increases or decreases it in semitones. It allows the same note to sound slightly different, varying from bass to treble (e.g. C#, D#, F#, etc).
- Octave: an octave is the distance between a given note, like C, and the next instance of the same note that is either higher or lower in a scale. In this way, we have a full 12 cycle of notes, such as C, C#, D, D#, E, F, F#, G, G#, A, A#, B, and when we reach the C again, although it is the same note, it is going to be one octave above (or below, in case it is an ascending scale).
- Rhythmic figure: symbols that represent the duration of notes. It is not an absolute measure, since it can vary depending on the beat value and the tempo.
- Scale: a scale is any set of musical notes ordered by a fundamental frequency or pitch. A scale ordered by an increasing pitch is an ascending scale, and a scale ordered by a decreasing pitch is a descending scale. There are many types of scales, but the most common ones that will also be mentioned in this work are the *chromatic scale*, that presents all the 12 notes in the scale, considering its accidents (C, C#, D, D#, E, F, F#, G, G#, A, A#, B), the *pentatonic scale*, that presents 5 notes per octave, and the *diatonic scale*, that is a sequence of 7 successive natural notes. It includes five whole steps (i.e. whole tones) and two half steps (i.e. semitones) in each octave, in which the two half steps are separated from each other by either two or three whole steps, according to their position in the scale (e.g., if we determine F as a fundamental, we will them have the sequence F—C—G—D—A—E—B).
- Fundamental: a reference note chosen among all the 12 possible notes in a chromatic scale. Starting from a fundamental, it is possible to build musical scales.
- Comma: is the smallest interval the human ear is able to perceive. Between semitones intervals, such as C - C#, for example, we have these microtonal variations.

4 META-INTERACTIVITY

In this work we introduce the concept of meta-interactivity as a novel approach to be explored in gamified experiences and AI systems. Gamification consists of using particular game development techniques and strategies in other contexts that are not necessarily associated with gameplay mechanics themselves [12, 25]. It is a reframing of Game Design elements in order for it to be applied in different fields or domains, especially through the design of gratification-wise mechanics for improving both experiences in the concrete reality and all kinds of interactivity we can have within cyberspaces. Bringing such elements to experiences other than games, as we can observe examples for educational purposes [13], for staff training [22], and many others, fosters a huge impact in points such as engagement, productivity and focus, thereby making tasks in any context tangible and more pleasant to achieve. According to Lucero et al. [39], features that make games and play engaging can also make other kinds of products more enjoyable and meaningful, increasing the quality of the overall user experience. Playfulness, in other words, can be a positive feature in products that goes beyond pure entertainment. Similarly, the concept of playification [42] has recently been proposed as an extension to gamification. Differently from gamification, that uses specific game elements to engage individuals (e.g. score and trophy system), playification aims on employing a gameplay mechanic by itself (that is, using the actual game's interactive model instead of external rewarding elements of the experience) thus substantially approximating any experience to the one of a video game. This approach was employed in order to resignify sessions of physiotherapy for elderly inpatients, aiming for more efficiency in treatments.

With meta-interactivity, we aim on addressing the same goal and effect: engage users to accomplish complex tasks, such as music composition, through a ludic effort, that not only makes the tasks easier to be accomplished but also more engaging and fun. The way we propose such effect is rather different, however. Our effort is centered on algorithms that establish bridges capable to translate different artistic endeavors, that are mainly based on expressive efforts, into new artistic instances. From this novel form of user-system relation, we expect coherent forms of artworks emerging from interfaces designed with ludic elements, that evokes an aptitude of the user and converts it into a new instance. It can be comprehended as a way to translate different creative motivations, like “drawing” to “musical composition”, in order to propose different relations between the system and the users, such as “painting a music”, “playing” (in the musical sense) on a canvas. Ultimately, it consists of designing minimalist, very intuitive, and easy-to-use interactive models that generate complex outcomes.

This kind of subverting behavior has been observed in classical art, more specifically in the abstractionism movement, generally understood as a form of expression that does not represent objects on its proper form, according to our perception of the concrete reality. For instance, Kandinsky connected painting with musical composition, drawing forms that refer to motion and that can be perceived as a musical notation system [9]. Bringing a similar translation model to digital environments establishes very particular challenges in order for the approach to be effectively implemented. For instance, Hunt and Kirk [30] examined many different strategies for mapping human gestures onto interactive systems in order to improve performers expressivity in live performances, allowing them to compose music in real-time.

Other works explored new forms of collaboration between the user and a system that also subverts the interaction, as we can observe in the concept of implicit collaboration [34, 43], that consists on the translation of the player’s actions, such as exploring a virtual environment, into a new instance, such as a musical corpus. In these works, however, the user was not necessarily aware that he/she was actually contributing with the quality of the emergent music. Our approach, on the other hand, proposes that the user’s focus should be on the musical production, as if they were manipulating a musical instrument. Also, the relationship we intend to establish between the user and the system should allow, through a playful interface that does not stipulate excessive restrictions on the creative side, the appearance of complex and coherent music.

Other similar translation approaches have also been explored in recent works. For instance, Duckworth et al. [15] describe the design and development of Resonance, an interactive tabletop artwork that targets upper-limb movement rehabilitation for patients with an acquired brain injury. The artwork consists of several interactive game environments, which enable artistic expression, exploration and play. Each environment aims to encourage cooperative and competitive modes of interaction for small groups of participants in the same location. This is an example of subverted game mechanics usage as tools to achieve different goals, such as people rehabilitation. The potential of similar approaches employed as a way to extend individual’s capabilities are, however, yet to be explored.

Meta-interactivity allows us, furthermore, to reflect in more depth about the role of games and play as a cultural phenomenon. Miguel Sicart [60] argues that play evokes a sense of presence in the world, it is a way to understand our surroundings. It is also a form to connect individuals, fostering interactivity. It goes beyond the game itself; it is a mode of “being human”. According to the author, a theory of play does not derive from a particular object or activity, it is a tool capable of bringing complex interactions between people as an extension of their daily life activities. Thus, it is not separated from reality, but part of it. Through this perspective, we can glimpse that, once the proper bridges capable to establish and turn an individuals’ engagement into a proficiency are built, it is also possible to turn dull activities from our daily lives into a playful experience, that allows us to succeed in many different endeavors without the necessity of time consuming – and sometimes boring – training.

On the other hand, time invested in an aptitude can be very rewarding; it is great to get through a learning process and perceive an evolution. However, in cases such as the music theory and practice learning, it also leads to withdrawal, since often a satisfactory evolution is attributed to "talent" [46], and this is a factor that distance people from developing basic musical skills. We believe the meta-interactive approach can help to deal with the overwhelming steps involved in the learning process of determined tools and techniques, such as music practice and composition, giving individuals more resources to explore tasks in a more intuitive, powerful and playful way.

In the next section, we will introduce a system developed through a meta-interactive approach, where a coherent musical compositional emerge out of a playful interaction of a user in a virtual environment. Although the user is engaged on the proposed activity (that is, the user is aware he/she is composing a music), it is done through a very minimalistic interface that proposes the user to "play" (in a ludic sense) with a virtual world by bursting bubbles under the sea, and this action is translated to a composition effort for interesting music to emerge. That user can learn and appropriate from interactions with the environment and use it in his/her favor for a complex behavior (i.e. a new musical instance) to be created.

4.1 Bubble Sounds

Bubble Sounds is an interactive musical system developed as an installation for artistic environments. It offers a game-like experience, but its proposal is rather different from the common entertainment-based approach focused on reward-wise mechanics. It enables the emergence of music through the interaction of human agents with the system, but not limiting the coherent production to experts in musical theory and performance. In fact, the idea is that anyone, regardless of cultural background and expertise, should be able to create music through its novel notation system. Hence, this approach enables non-experts to compose interesting musical pieces through the usage of imagetic elements in a virtual environment, that turns the system in a kind of ludic musical instrument that explores the user's perception of visual elements and translates it into music.

It was originally developed for virtual reality devices, more specifically for the Oculus Rift, and it uses the device's accelerometer to allow a 360 degrees visualization and interaction with the 3D environment. It also uses a microphone to capture sound input, being these the possible agency for the user over the compositional system. This project was developed using the Unity 3D engine, and the game assets were created using Blender 3D.

The aesthetics of the environment is based on an underwater theme, where bubbles are emerging from the ground all the time. It presents a peculiar form of arranging musical notes, composed by a stack of concentric circles that works like a vortex, going all the way up from the virtual environment's floor to its top, as shown in Figure 1. The user is located at the center of this vortex, which can be comprehended as a stack of tracks that orbits the user's position in the 3D environment, capable to receive musical notes. It is a tridimensional representation of a musical score, which is circularly interconnected merging the edges and forming a loop.

A differential of this system is how the interactive model was conceived to foster new ways in which the user dialogues with the system. As mentioned earlier, the system uses a microphone to capture the user's input, and this is how the notes are activated in the vortex. As the user directly stares at a determined bubble, he/she can clap hands, and the variation in the audio frequency will release the notes inside the bubbles, which will then start orbiting a vortex track, producing a musical loop. The tempo in which each of these notes orbit the vortex also follows a procedure that allows the emergent musical corpus to be coherent in a matter of rhythm, and it never generates cacophony (i.e. unpleasant and chaotic musical structures), as we will explain later. In this way, since the system arranges the notes in terms of tempo, allowing for coherent rhythmic structures to emerge, the clapping hands also produces a meaningful

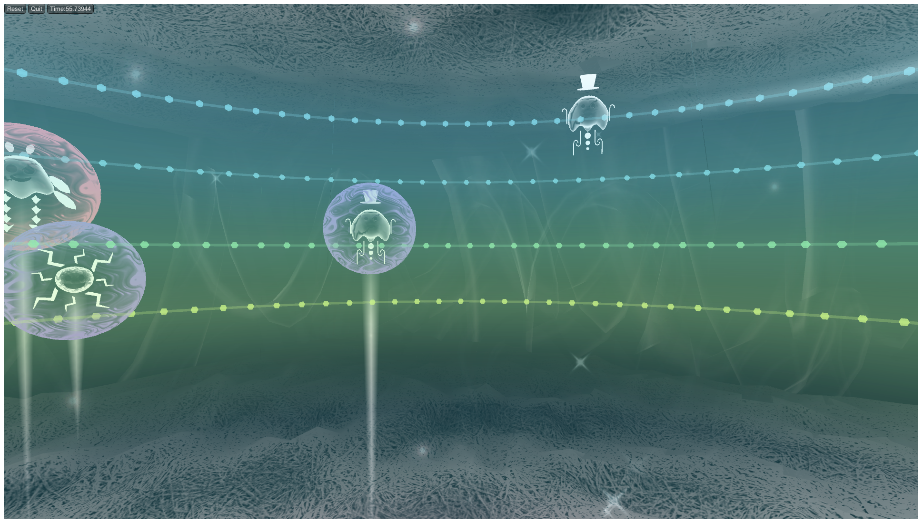


Fig. 1. Bubble Sound's interface.

rhythmic for anyone outside the experience observing the user, generating 2 layers of interactivity that produces coherent musical instances. Before going into more depth about how our approach achieves this, we will present the system overview in more detail.

In the underwater environment, the user sees 12 types of bubble colors randomly emerging from the ground all the time, each one corresponding to a musical note that matches with a colored circular track in the vortex. Each bubble contains a microorganism, from a total of 6, that presents variation in their speeds and octave, as shown in Figure 2. It is up to the user to choose, according to a color procedure that matches note qualities (i.e. tones) with colors, when to release these notes from the bubbles in order to trigger its sonorities.

In other words, the colors carry an important guideline for musical composition: it establishes for the microorganism inside the bubble a range from bass to treble sounds following a visual procedure, done by relating the pitch of each note in the bubble to the color spectrum (Figure 2). The user can interact with the bubbles to release the 6 different types of microorganisms placing them on the vortex of tracks according to his/her own desire.

In the same way, users can also remove notes that are already orbiting according to their on will and on the fly. Thus, it is possible to test if a determined note suits the composition, and if not, it can can be easily removed through a clap of hands (or a mouse click) after having this note in the focal point, and then the microorganism will fade out and disappear from the vortex.

The fastest microorganisms represent high pitch notes, and the slowest ones the low pitch notes. Hence, besides learning the microorganism tone and speed, users can also easily identify musical notes even before releasing them from bubbles by analysing the speed in which bubbles are dislocating vertically. Thus, even with microorganisms idle inside the bubbles, the user might be able to understand, only by observing the visual elements (i.e. bubble colors meaning tones) and its motions (bubble speeds meaning the octaves, which is the same for the microorganism within), the sound quality of the notes that are appearing on the screen. They can then judge whether the notes characteristics fits or not the music being composed.

NOTE	COLOR
F	#A955A9
E	#3A6BA0
D#	#0499CB
D	#38B09E
C#	#56AE77
C	#8CC84F
B	#FDFD05
A#	#FFA42A
A	#FF7940
G#	#FE4A58
G	#FD3667
F#	#E13D9F

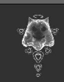
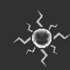
MICROORGANISM	SPEED	OCTAVE
	(slowest) 5 (00:00:16.0)	(lowest pitch) 1
	10 (00:00:08.0)	2
	20 (00:00:04.0)	3
	40 (00:00:02.0)	4
	80 (00:00:01.0)	5
	(fastest) 160 (00:00:00.5)	(highest pitch) 6

Fig. 2. Microorganisms speeds and note-to-color match system.

Like the bubbles, there are also 12 tracks in the stack (vortex), representing all the possible notes in a scale, be it a chromatic, pentatonic or diatonic, considering its accidents (C, C#, D, D#, E, F, F#, G, G#, A, A#, B), as shown in Figure 3. Only the commas, that are intervals smaller than a half tone, are not being approached; which gives us a good range of tonal variations from the western musical system. As previously mentioned, the vortex works like an interactive tridimensional representation of a musical score, merged like a cylinder and with a stylized score sheet projected in its interior faces, where the user has a 360 degrees view of the whole cluster of notes orbiting him/her.

The vortex structure works as follows: the lowest notes in the same octave (i.e., in the same progression of notes, where we have the first one, the *fundamental*, and the following 7 notes, until we reach the next octave from the fundamental) orbit at the lowest part of the screen (i.e., lower tracks), and the highest notes at the top (i.e., upper circle tracks), also shown in Figure 3. The lowest note in the stack of tracks on the vortex is F# (our fundamental), and the highest note is F (also forming a loop vertically), that gets more sonorously high pitched as it goes up. These notes are associated with a color, as mentioned earlier. E.g., F# is pink and F is lilac, as shown in Figure 2. Thus, besides the tonality that each note emits, which might be enough for expert users to use in his/her creative process, the color system will also support non-expert users to quickly recognize and determine the best notes to use in their composition. This translation approach of tones to colors does not only favors non-experts, as experts might also benefit from it for quick notes recognition and activation.

When the bubbles are burst by the user, the microorganisms inside them start to orbit around the player in its own vortex track and in a pre-determined speed according to the microorganism's type (as shown in Figure 2), considering the color match of the bubble and the vortex track. Only then it starts to emit its sound (while inside the bubbles, the microorganisms are idle, not producing their sonorities).

There is a 3D asset/model in the experience we call "pillar", that are faint shapes that hold the whole structure of tracks (these objects are being identified on Figure 3 and 6, along with other interactive key-elements of the system's

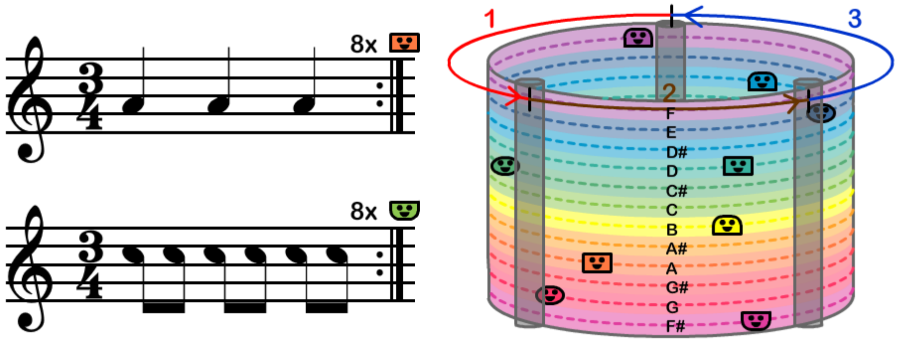


Fig. 3. Representation of a 360º musical score, projected inside a cylindrical structure. Note that microorganisms have different speeds, as shown on the rhythmic figures presented on the compass on the left.

mechanics). When a microorganism is released and starts orbiting the vortex, it collides with these pillars, triggering the note it carries. There are 3 pillars around the cylindrical musical score structure, so, in a 360º turn, a microorganism will have its sonority triggered 3 times, as shown in Figure 3. The duration of each musical note between cycles of interaction with the pillars varies according to the quality of each microorganism and its respective speeds. For example, in Figure 3, we identified the rhythmic figures corresponding to the note A (orange microorganisms) and C (green microorganisms), showing that the A sustain (i.e. the time its audio signal will endure across time) will resonate for a longer period than C's.

The user has full control of the tone and pitch of the microorganisms triggered to orbit the vortex. Colors attributed to notes are fixed in the infrared to ultraviolet spectrum (e.g. *F* will always be lilac and *B* will always be yellow, as shown in Figure 2), so it helps creating a pattern for users to manipulate these colors in their creative process. Similarly, there are 6 octaves of the same note for users to use in their composition, that are determined by the microorganism speeds. In this way, it is possible to have many organisms orbiting the same vortex track presenting audio variation (from bass to treble), according to the 6 types of microorganisms that presents their own speeds and octaves, as shown in Figure 4. Microorganisms orbiting the same vortex track have the same tone (e.g. only *C*, or *C#*, or *D*, etc) but they can vary in their height channels from low to high pitch. This is visually expressed by the slow and fast orbit motion of the microorganisms in the same vortex track according to their octaves (i.e., their correspondent speeds).

Hence, in short, bubble colors point to note quality, from F to F# (a complete 12 note loop) and the microorganisms point to speed (and, as such, by the octave of that note). In this way, the system not only provides a very good range of notes for the user to choose (similarly to most pianos, that offers 7 octaves, the system is offering 6) but it also offers an organized way to visualize the cluster of notes orbiting each track, clearly separated by colors. This feature, as a compositional resource, also resembles to octave pedals, used in musical instruments such as guitar and bass for more tonal diversity, since commonly string instruments are restricted to 22 or 24 frets on their necks.

It is important to emphasize that the user does not need to burst the bubble to understand the speed of the microorganism within, since the actual speed is presented even when they are idle inside bubbles, by the speed they are emerging from the ground (it is the same speed in which it is going to orbit its correspondent vortex track). Also important to note that all these elements from music theory, such as tone, pitch, sustain, scale, melodic and harmonic lines, etc., were completely abstracted to a purely visual system based in colors and movements, with a clean interface,

Fig. 4. All notes possibilities (represented by bubble colors), varying from F to F#, and all the microorganisms octaves the bubbles can carry.

that presents no buttons or complex instructions. The users just look around the virtual environment and clap hands to trigger the sonorities carried by the microorganisms.

The release command is made through audio recognition, in which all the time a determined bubble is on the line of sight of the user identified through the Ray Trace, that is, the vector that represents the distance between the player's location (i.e. where the user is looking at) from the location of the actual bubble, the audio frequency variation will cause the bubble to burst releasing the microorganism within. As mentioned earlier, the recommended audio signal input for this action is to clap hands, as it also produces an interesting new layer of music interactivity for an external listener.

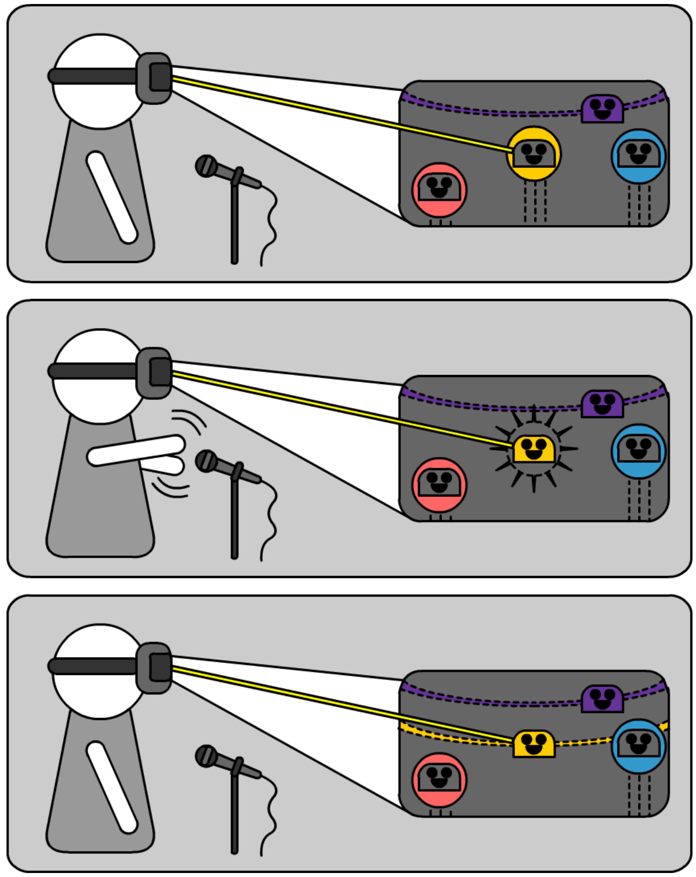


Fig. 5. Interactive flow of Bubble Sounds.

In Figure 5, we identify the interactive flow for a musical note to be triggered during the experience. The sequence shown is:

- In the first frame, the user spots any emerging colorful bubbles with idle notes moving on the y-axis of the screen. We can observe that the Ray Trace is aiming at a determined bubble (yellow line coming from the user to the spotted bubble) as it is in the user focal point.
- In the second frame, with a bubble in the focal point (whose color determine the note), the user can release the microorganism within (which defines the octave according to its speed) by clapping hands on the microphone.
- In the third frame, after the bubble is burst, the microorganism follows a linear path to the corresponding vortex track, and starts orbiting the player, producing its sonority.

As mentioned before, in the same way users can trigger idle notes, they can also remove orbiting notes at will as it might not be wanted to a given composition or changes are required due to a natural evolution in complexity and meaning of a musical piece across time. This can be done through a similar procedure described above (i.e. the user spot an unwanted microorganism and clap hands or left click in the mouse button). Thus, yet minimalistic, the system's interface allow for an easy customization in the emergent musical corpus.

At the exact moment a bubble is burst, the microorganism inside follows a linear interpolation from the point it is released (considering bubbles only moves in the y-axis) to the appropriate track, represented by the match of the bubble color with a corresponding circular layer in the infrared/ultraviolet spectrum. Note that the quick motion of the microorganism from the point it is released to its actual track will not affect the sound; it will only trigger the note it carries when it is actually orbiting its corresponding track, colliding with pillars. It is also important to emphasize that the encapsulated notes (the ones inside the bubbles) do not produce sounds; they only start to react when they are properly released.

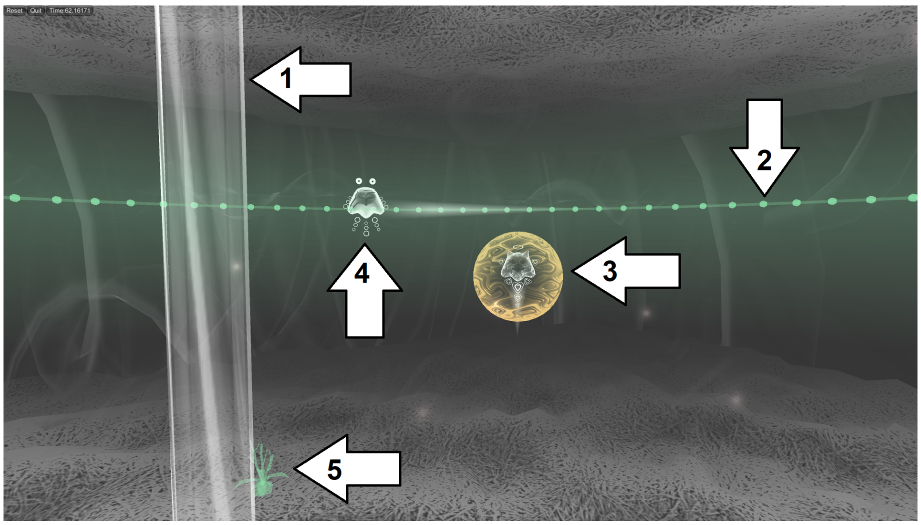


Fig. 6. Through the arrows you can identify the following elements of the system: 1. Pillar; 2. Vortex track; 3. Bubble; 4. Microorganism; 5. Dynamically generated asset.

A feature of the system that is not yet being fully explored in all its potential in the current version are the dynamically generated 3D assets in the environment's background, as shown in Figure 6. They are presented in a wide range of possible 3D assets, like algae, corals, shipwrecks, treasure chests, etc. It is intended that these assets help creating soundscapes that matches our actual underwater theme, also having a sound signature attached to them. However, the sound output they provide are different in length – they tend to last longer, as a subtle background harmonization provided by the system. Since the system was designed as an interactive installation for artistic environments, it is intended for these sounds to always keep the system active, producing subtle harmonization, thus never leaving it completely silent in case no one is interacting with it. In this way, the system can still provide a pleasant audiovisual experience for those just watching or passing by it. The intent behind this module is also to dynamically create micro-narratives associated with the underwater theme. We will discuss more about this module in the Discussion section.

As previously mentioned, our system focuses on allowing non-expert users to produce a coherent musical corpus in a matter of *temporal structure*, enabling the system to work like a smart musical instrument, capable to foster creative freedom for the user to musically express himself/herself without limiting the compositional process (thus allowing the output to result in a pleasant musical piece). This is done by a simple procedure: we make each microorganism orbit the vortex in its own speed, that follows a predetermined geometric progression (5, 10, 20, 40, 80, 160), as shown in

Figure 2. This approach guarantees there will not be notes being played out of the tempo of the music, thus guiding non-experts to generate rhythmically coherent musical structures.

For the interested reader, a video demonstrating our system, as well as some music generated through it by experts, novices and a random algorithm (also used in our assessment sessions) are available at: https://phersu.com.br/bubblesounds.

5 MUSIC EVALUATION

Before we describe our results, we briefly discuss some musical definitions that align with our system, not only to situate the path this work develops towards, but also because it helps us to identify some of the challenges we are trying to overcome. This work aligns with a music definition attributed to the modernist composer Edgard Varèse, in which music can be comprehended as "masses of organized sounds that move against each other" [45], as well as the definition proposed by the music theorist David Temperley, in which he points out to the listening phenomena as a key element in the emergence of meaning when interpreting sounds [65]. Both definitions bring important considerations on how to properly assess interactive musical systems and their desired outcomes (i.e. musical products that can be acknowledged as human-made pieces), especially regarding the effort of finding non-subjective musical classification criteria of a given musical corpus. That is, the subjectivity around whether a given music sounds "good" or "bad" among a group of individuals may vary according to a lot of factors (e.g. cultural background, age, etc), so it is complex to find the correct parameters to determine quality. It is also difficult to find the right questions to ask when it is desired to evaluate musical quality among a group of people.

One way to deal with this problem is to try to identify "figurative" elements in the generated music, making it approach something "real" that is already known and comfortable to us. This can be done by identifying repetitions in the music, structures that can be recovered over time in order to emphasize a musical identity. Thus, the analysis is based, at first, on fundamental elements of music, such as harmony and rhythm, since these are the elements that produces loop patterns that gives the music a signature (considering it is constantly repeated across time). It is, however, very challenging to find a balanced way to evaluate such systems, and it seems to be a recurrent problem in many works related to procedural music generation, since it is always necessary to cope with subjectivity, especially regarding the feedbacks from human subjects, whose perception of quality varies according to several factors (e.g. cultural background). In this regard, Applebaum [3] presented a more simplistic way to determined whether or not a determined sound mass should or should not be classified as "music". He proposes to switch the common recurring question "is it music?" to "is it interesting?", instead. Based on this change of perspective when questioning a sound mass, we built an assessment method we will discuss in details in the next section.

6 EXPERIMENTS

We ran experiments in 2 different stages. The first stage consisted of 42 human subjects, in which 29 (69%) were non-experts in music theory, whose experience was mostly based on listening, and 13 (31%) were experts, with high knowledge in both music theory and practice. We trained the users during the experiment sessions for 3 minutes, providing them basic instructions about the system functionalities, while they freely interacted with the virtual environment in order to have a general feel of it. Our goal at this stage was to evaluate Bubble Sounds and its meta-interactive mechanic.

The second stage consisted of 111 human subjects, who answered an online form that randomized three music samples from the 3 different databases captured on the previous assessment sessions, from the compositions of both experts and non-experts along with randomly generated music.

That is, we randomly picked 10 samples created by the experts, 10 created by the non-experts and also added other 10 generated by a random script to be included in the database. These random samples were generated by an algorithm that randomly triggers some notes and puts them into orbit around the vortex. At the same time, it also has a “filter” that arbitrarily removes notes and makes them inactive, simulating the agency of a user in the system across time; as they can place and remove notes at will.

Besides simulating human agency, this resource aims on keeping the random generated musical samples less polluted in terms of quantity of orbiting notes. If the random algorithm simply kept triggering notes endlessly, we would not only experience performance-related issues but it would also facilitate listeners to easily perceive and acknowledge these productions as something made under artificial methods. Our goal was to ensure, to our best, that the random creations do not sound very dissonant to the human productions.

Each sample on the form (corresponding to random, non-expert and expert) had their orders randomized. And for each category, we randomly choose one out of the 10 samples to display to the user. Our goal at this stage was to evaluate the musical outcome of our system. The users did not have any contact with the interface of Bubble Sounds at this stage.

In this section, note that we say ***expert/non-expert samples*** to refer to the musical samples produced by experts/non-experts, respectively; and ***expert/non-expert evaluators*** to refer to the expert/non-experts users who evaluated those productions. This study was approved by the ethics committee from Lancaster University.

6.1 Stage 1 - System evaluation

The human subjects from this session were mostly identified as game developers from FUMEC University, graduate students from IFMG, computer science students from UFMG and USP and game enthusiasts and aspirant musicians nominated by participating individuals, all from Brazil. Before starting the evaluation session, each user agreed and marked a consent form, with clear explanation of the whole process. No sensitive data from any of the users was recorded.

The version of the system used for the assessment was adapted to work without both the VR device and the microphone input, since the sessions were conducted online. It was used a regular monitor with mouse controller for manipulating the virtual environment view, which turned the interaction into a more “video game-like” experience. After freely experimenting with the system, the human subjects were asked to compose a 1 minute sound piece, which was recorded to be used in the second stage of the assessment. We queried the users who experimented with the system the questions shown in Table 1.

In Q1, we identified the user by its proficiency in music theory and practice, which enabled us to evaluate both the general scenario, considering the 42 feedbacks, as well as the individual scenarios from both experts (29 users) and non-experts (13 users). We considered experts all individuals who have mastered the practice of at least one instrument, have had any solo or group performance experiences, have basic knowledge in music theory (e.g. capable to read musical scores) or who had previous compositional experience.

In Q2, we queried the users about their overall experience with the system. We obtained an excellent evaluation in this part, considering the general scenario, with a $\bar{x}$ = 8.09 out of 10, showing that users had a pleasant experience during the session, as shown in Figure 7 (a). In the independent scenario, considering just experts (13 users), we observed a $\bar{x}$ = 8. As for the non-experts (29 users) we observed a $\bar{x}$ = 8.12, showing that, regardless of the musical expertise of the users, all had satisfactory experiences.

Q1.	How do you classify your skills as a musician? ( ) Novice. ( ) Expert.
Q2.	From 1 to 10, how do you classify your experience with the system? Uninteresting - 1 ( ) 2 ( ) 3 ( ) 4 ( ) 5 ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 ( ) - Interesting
Q3.	How do you rate your own composition? Bad - 1 ( ) 2 ( ) 3 ( ) 4 ( ) 5 ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 ( ) - Excellent
Q4.	How much fun you had during the experience? Little - 1 ( ) 2 ( ) 3 ( ) 4 ( ) 5 ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 ( ) - Plenty
Q5.	Establish a relationship between the generated music with a feeling (E.g.: calm, anguish, tension, etc).
Q6.	How do you classify the compositional interface of the system? Bad - 1 ( ) 2 ( ) 3 ( ) 4 ( ) 5 ( ) 6 ( ) 7 ( ) 8 ( ) 9 ( ) 10 ( ) - Excellent
Q7.	Which of the compositional elements best supported your music creation process? ( ) The relationship between colors and tones for partial identification of musical notes. ( ) The relationship between movement and tempo for rhythmic definition. ( ) The imagetic representation of low and high pitch tones in the vortex tracks. ( ) None of the above.
Q8.	Identify the main focus of your attention during the experience with the system: ( ) My focus was on sound production. ( ) My focus was on the imagery elements, such as the objects that appeared in the background. ( ) Arbitrary, I just explored the interactive possibilities provided by the system. ( ) My focus was on learning the interface and its usability. ( ) None of the above.
Q9.	Would you like to share any additional details about your experience? If so, write it below.

Table 1. System evaluation questionnaire.

In Q3, we queried the users about their perception over their own composition with the system. We obtained a good result in the general scenario, with a $\bar{x} = 7.28$, as shown in details in Figure 7 (b). Considering the independent scenarios, we observed a $\bar{x} = 7.3$ regarding the perception of experts over their own creation, and $\bar{x} = 7.22$ for non-experts in the same question. We believe the means from both groups represents an interesting result, considering that users only had 3 minutes of training before the sessions, but still showed to be satisfied with their own composition. Thus, this result shows that our approach is promising on enabling non-experts to produce interesting musical pieces with our ludic interface, in a very spontaneous fashion (i.e. not requiring much training to do so). This way, we succeed in our attempt to offer a system that not only is easy to learn, but also supports interesting musical pieces to emerge even without much training.

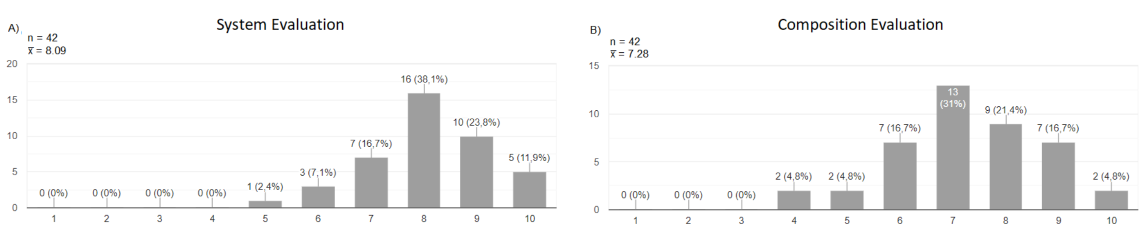


Fig. 7. General perception over the experience with Bubble Sounds.

In Q4, we queried the users about how much "fun" they had during their experience. We understand this is a very subjective question, but our intent behind it was to also verify how well our minimalistic interface went according to the user's perception of Bubble Sounds, considering its proposal of generating complex music through a very simple yet intuitive interface. In the general scenario, we observed a $\bar{x}$ = 7.92, as shown in details in Figure 8 (a). The result here is also promising, especially considering that the version used during the sessions was not using the VR device nor the microphone input, which we consider important elements to engage the user in this experience. Thus, even with simplified interactive mechanics, the system still offered a fairly good level of entertainment. Considering the independent scenarios, we observed a $\bar{x}$ = 7.92 considering just the expert users, and $\bar{x}$ = 7.9 considering non-experts. Considering the results obtained in Q3, where users rated their composition, with the means obtained in Q4, we have a very positive glimpse of how well our approach is addressing the targeting problem (i.e. allowing non-experts to produce coherent music).

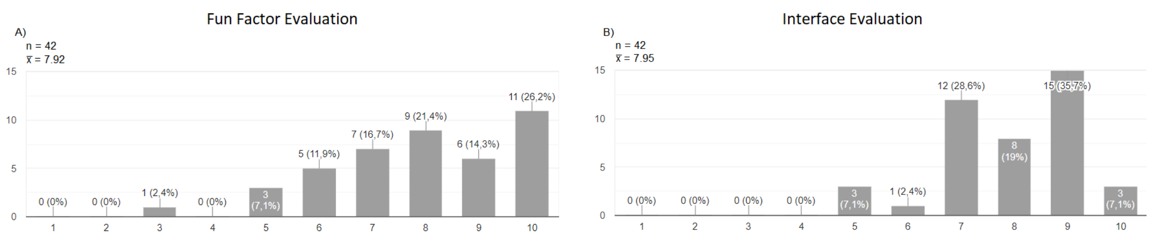


Fig. 8. General perception over the compositions generated by Bubble Sounds.

In Q5, we queried the users if they could identify and assign a feeling to their own composition. Since the question allowed users to freely express anything they wanted, we observed a wide range of adjectives, with a preponderance of the feeling "relaxation", through synonyms like "calm", "zen" and "relaxation" itself, with a portion of 50% of the whole share. Following this, we had "tension", through synonyms like "anguish", "dark" and "tension" itself, representing other 30%. This contrast between "Calm" to "Dark" in the responses evokes the music comprehension as a *listening phenomena*, as defined by Temperley [65], where the cultural background of each individual plays a major influence on how the perception over a determined musical corpus is conceived. We strongly believe this perception may vary according to the population's cultural background, and even other parameters such as age, for example. Other 20% consisted of feelings determined by nouns, adjectives and verbs such as "Dawn", "Outerspace", "Beautiful" and "Chase".

In Q6, we queried the users about the system's interface, in order to assess how well our approach supports the user in the compositional process. Ultimately, by evaluating the system's interface, we were also expecting feedbacks regarding our constraints, that is, how much did our algorithm actually empowered the user on their expressivity or limited it. We observed a $\bar{x}$ = 7.95 in the general scenario, as shown in Figure 8 (b). In the same way it happened in Q4, this was a very positive result, since the interface showed to be efficient in supporting user's expressivity even though it was presented in a simplified form (i.e. without VR and microphone input). Thus, even with the adaptation for the mouse input, the users acknowledged that the interface provided the proper resources for their creation and were able to learn from it. Considering the independent scenarios, we observed a $\bar{x}$ = 7.84 for the experts and $\bar{x}$ = 7.93 for non-experts. A higher mean for non-experts was expected, given that non-expert users likely rely more on the interface and its imagetic devices for fostering musical composition than experts. We believe that, at some point, due to their background training in musical listening and perception, experts were able to manipulate sounds directly, as if they were manipulating a synthesizer, rather than only the imagetic elements in the interface, as non-experts. We also believe that the overall perception of users around the interface, specially in regard to the constraints, was also positive. Considering also the results obtained in Q2 (users overall experience with the system) and Q3 (users rating their own compositions), the system seems to have provided an interesting level of freedom for users to create music, not constraining the production in a way to nullify expressivity. That is, by providing a color-to-tone translation approach the system does restrict at some level, since the duration of the notes is also fixed according to the octave. However, this might be fine as long as it still does allow users to employ their ideas and convert them to music. At some extent, we can argue that any musical instrument constrains, whether by the way they arrange notes in a scale, the way they present their octaves or even in the amount of fretboards (in case of string instruments), for example.

In Q7 and Q8, we asked the users to identify, respectively, the system elements that may have supported him/her in the music creation process and also what was their focus while interacting with the 3D environment. Our goal was to identify what users considered the most important translation mechanic for ludic interaction, based on the visual elements created to support musical composition.

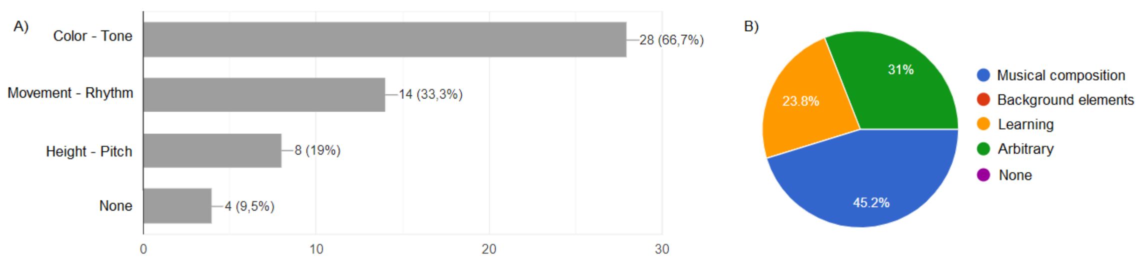
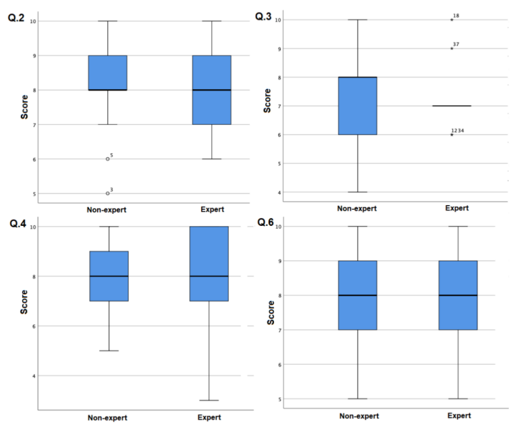


Fig. 9. a) Users' perception over the co-creation mechanic and b) Their focus during the experience.

In Q7, considering the general scenario, two important features of Bubble Sound's mechanic were acknowledged by the users as supportive compositional elements, and they were **the relationship between colors and tones for partial identification of musical notes** (66.7%) and **the relationship between movement and pulse for rhythmic definition** (33.3%), as shown in Figure 9 (a). The relation between tone (low and high pitch) and the vertical representation of the vortex of tracks was acknowledged by 19% of the users, while 9.5% pointed that none of the listed compositional resources were used in their own advantage. Considering the individual scenario, we observed that 70.9% of the experts pointed that the color/tone relation was the feature that best supported their creation, followed by 32.2% that pointed the movement/rhythm relation, also pointed as an important element. A small portion of the expert's pool (19.3%) also pointed that the height/pitch relation supported their composition, while 9.6% did not mark an option. This result shows an interesting trend of our imagetic approach to support not only non-experts, but also experts. At first, we believed that experts would appropriate more of the pure sounding elements for their creation (e.g. the representation of height/pitch), leaving the color mechanic a little aside, but that was not the case in this assessment. As for non-experts, we observed that 66.7% of users pointed the color/tone relation as the main feature to support their creation, followed by 23.7% pointing the movement/rhythm relation as an important resource for production. Other 9.6% pointed that the height/pitch relation was an important resource on their compositional process.

In Q8, considering the general scenario, the majority of the users (45.2%) identified their focus while experimenting with the system was on the **sound production** (i.e. they were actually engaged in the task of composing a music through the imagetic interface). Considering the means obtained for both experts and non-experts in Q3 ($\bar{x}$ = 7.92 and

$\bar{x}$ = 7.9, respectively), where we asked them to rate their own composition, that was an interesting achievement for our approach, since almost half of the whole pool of users were able to produce quality music, according to their own evaluation, being actually aware of what they were doing (i.e. their primary goal was to compose a song, not play around with the interface, and they were satisfied with the result), as shown in Figure 9 (b). Many users (31%) also pointed that their experience were arbitrary, that is, they did not have creative focus during the experience and were just "playing" with the interface. This result highlights an important effectiveness of our approach in generating a coherent musical corpus; that is, even those who were not necessarily engaged in music creation did enjoy the musical product of the system, and thus our approach shows to be promising in its task of promoting the dynamic emergence of music in a coherent fashion. On the other hand, 23.8% of the users pointed that their focus was in learning the interface and its usability. Thus, during their time of experimentation, their focus was in uncovering what the system could actually offer in their advantage. No users identified the dynamic generated assets in the background as relevant compositional guideline.



T- test	Musical knowledge		
Variables	Non-expert (N=29)	Expert (N=13)	p-value
Q2 – Experience score			
Mean	8,13	8,00	0,740
Median	8,00	8,00	
SD	1,15	1,23	
CI	(-0,651 – 0,909)		
Q3 – Composition score			
Mean	7,23	7,23	0,992
Median	8,00	7,00	
SD	1,59	1,09	
CI	(-0,980 – 0,970)		
Q4 – Fun score			
Mean	7,90	7,92	0,973
Median	8,00	8,00	
SD	1,56	2,18	
CI	(-1,191 – 1,152)		
Q6- Interface score			
Mean	7,95	7,85	0,837
Median	8,00	8,00	
SD	1,18	1,57	
CI	(-0781 – 0,960)		

Fig. 10. Non-expert and Expert means comparison for Q2, Q3, Q4 and Q6 through a t-test.

We performed a t-test in our numerical variables. Our goal was to have an in-depth perspective around the means for both experts and non-experts in **Q2**, **Q3**, **Q4** and **Q6**, and then verify if the proficiency level played a major influence on their perception. As shown in Figure 10, a t-test revealed that the high $p - values$ ($p > 0.1$) in the performed question points to a non-significant variance in the experts and non-experts perception regarding, respectively, their overall experience, composition score, how much "fun" they had and also their evaluation over the system interface. We could then conclude that proficiency, the dividing element of our population, did not play a major influence in the perception of users, and both provided positive feedback about their experience with the system.

6.2 Stage 2 - Music evaluation

During this stage, human subjects were asked to answer an online form containing 3 audio samples captured from the first stage of the assessment. Thus, the participants in this stage did not experiment with the system, they just listened to music produced through it. Before starting the evaluation session, each user agreed and marked a consent form, with clear explanation of the whole process. No sensitive data from any of the users was recorded.

Q1.	How do you classify your skills as a musician? ( ) Novice. ( ) Expert.
Q2.	Which music was the best? ( ) Sample 1. ( ) Sample 2. ( ) Sample 3.
Q3.	Identify the relation between samples. ( ) Sample 1 resembles Sample 2. ( ) Sample 1 resembles Sample 3. ( ) Sample 2 resembles Sample 3. ( ) The 3 samples sound diverse. ( ) The 3 samples sound similar.
Q4.	Which track sounded more professional, presenting a higher level of sophistication? ( ) Sample 1. ( ) Sample 2. ( ) Sample 3.

Table 2. User perception questionnaire.

The samples and their order were presented in a randomized fashion on the form, that was fully developed by the authors, available at https://phersu.com.br/bubblesounds. Thus, the users did not know whether they were listening to a random, novice or expert sample. We queried the users who experimented with the system the questions shown in Table 2.

In Q1, we queried the users about their knowledge in music theory and practice. From the 111 feedbacks received in this stage, we had 41 coming from experts (36.9%) and 70 from non-experts (63.1%).

In Q2, we asked the users to identify which sample out of the 3 in the form they liked the most. As mentioned earlier, both the order of presentation and the sample itself were randomized at the beginning, so the user was not aware if a given sample represented a production from random, non-experts or experts. As shown in Figure 11 (a), the musical production from experts was acknowledged as better music for the majority of the human subjects, with 39.6% of preference, followed by the non-experts, with 36.9%. The random production was the least recognized as better music, with 23.4% of the whole share. We run a chi-square test for our categorical variables, and considering the difference between experts and non-experts, we found a $p > 0.1$, an expected value given the proximity between the averages (36.9% and 39.6%, respectively) showing that the difference between experts and non-experts production was not statistically significant. This is a very good result, since it reinforces that non-experts were able to produce music not much different and comparable to the expert's production.

In fact, we also analyzed a scenario considering all human production (i.e. experts and non-experts samples) against computer-only executions (i.e. random samples), as shown in Figure 11 (b). Considering the human vs. computer scenario, we observed an impressive 76.6% for the user's production against 23.4% of random. According to the chi-sqaure test, we found a $p \leq 0.09$, showing statistical significance, giving us 91% confidence that human production is likely to be acknowledged as better music than the random generated ones. This result helps us to draw a clear distinction of our partially-autonomous approach if compared to other works, such as Roberts et al. [55] and Hang and Raymond [29], which employ a more autonomous fashion, not necessarily considering the human expressivity on their productive flow. In our case, our goal is to promote an effective human-machine collaboration, where the first may have creative freedom and the second may provide creative support in a non-limiting way.

These results show an interesting trend of our approach in allowing non-experts to produce music almost as good as the experts. In addition, it also shows that our system is effective on fostering coherent musical corpus even in the chaotic random executions, since a considerable share of the pool acknowledged the random production as good music. It did not, however, dissociate the necessity of the human factor from the relation with the system (i.e. our algorithm showed to be strong as a device for emergent music, however, it is not independent from human feeling/sentiment). We can then assume that the musical feeling associate with sentiment is still important in the creative process of Bubble Sounds, as any form of art, if we analyze it through the perception of a ludic musical instrument.

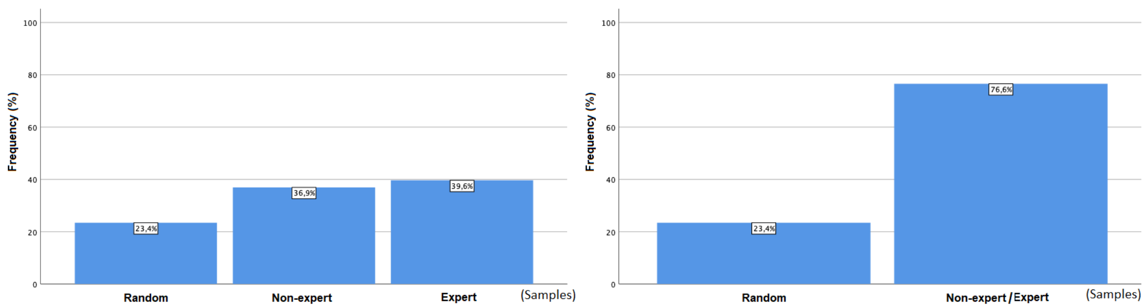


Fig. 11. a) The best voted samples according to the preference of both experts and non-experts. b) All samples (non-experts and experts) against random executions.

In Q3 we queried the users about what kind of relationship they could establish between the random, expert and non-expert samples. As shown in Figure 12, 31.8% of non-experts evaluators observed a resemblance between the non-experts and experts production. That was a very interesting result for our approach, since it shows that non-expert users were able to learn and produce music with similar quality as the ones produced by experts. As for the experts, it was pointed by 26.9% of expert evaluators that non-expert and expert samples sounded similar, which is very good result, given their refined listening skills.

It was pointed by 30.6% of expert evaluators that the 3 samples actually sounded similar, and a higher rate (42.3%) for non-experts was also observed. This was an expected result, given that the current stage of the system's development only presents a limited quantity of timbres and sound textures, which makes them sound similar according to some users. It can be compared to the same guitar with the same effect pedals, like a distortion or a reverb; it sound similar in timbre, but are differentiated by the melodic or harmonic structure they produce through an individual's expressivity. However, our objective at this stage did not go through the necessity of a greater amount of sounds in the system (but it will certainly be a focus on the next iteration of this work).

In Q4 we queried the users about how "professional" the samples sounded. As shown in Figure 13 (a), an important achievement was observed for our system, since according to the experts evaluators, the non-experts samples were acknowledged as better music than the experts samples themselves. This outcome was particularly interesting considering how refined the perception of expert musicians are; they are likely to be able to decompose a musical corpus to judge, and thus their analyses happen under a more meticulous criteria.

This result was also reinforced on in Figure 13 (b), where it is shown the individual scenarios of the preference of expert and non-expert evaluators over the human production (i.e. expert and non-expert samples) vs. machine

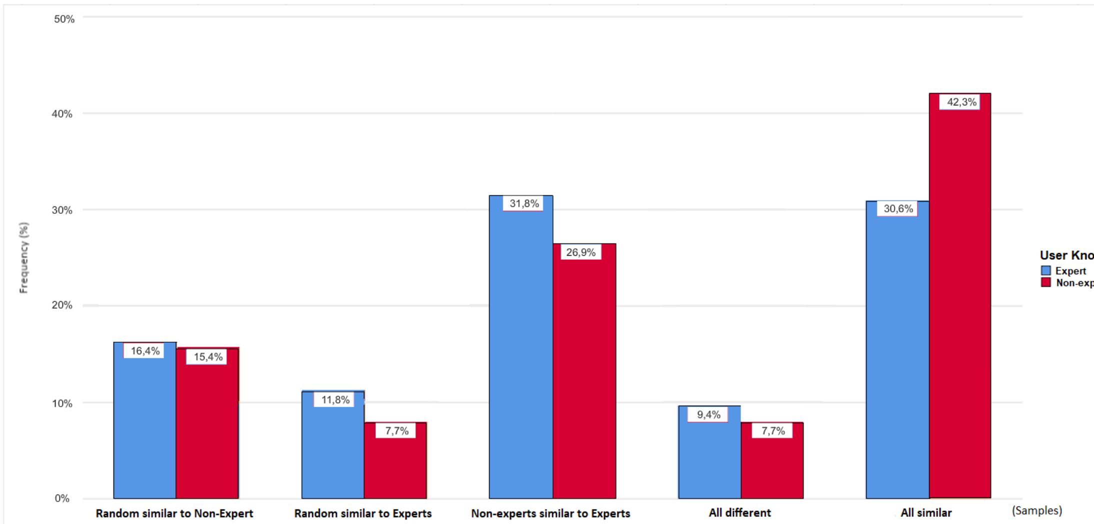


Fig. 12. The relation between the samples according to the user evaluator's perception.

production (i.e. random samples). Considering the expert evaluator perspective, a scenario of individuals acknowledging human samples as better music over the random samples with a larger margin was observed, with a portion of 85.4% for human production against 14.6% of random. Similarly, considering the non-expert evaluator perspective, human samples was once again acknowledged as better music over the random samples with a larger margin of 71.4% for human production against 28.6% of random.

We also observed that users who liked the samples produced by human agents also presented a high probability to rate the expert samples as more "professional" (with a $p \leq 0.01$ according to a chi-square test). Thus, a clear distinction between the music produced with the human factor (e.g. feeling, sentiment) was observed over the production of a random algorithm. This accounts to our effort on establishing an efficient collaboration between a human agent and the system, also showing that, while not fully-autonomous, the system empowers the users on their creative efforts. This way, we achieved our goal on proposing a system that is not self-sufficient (i.e. for complex musical outcomes to emerge, it needs the human-factor), and that could improve the user's compositional capabilities.

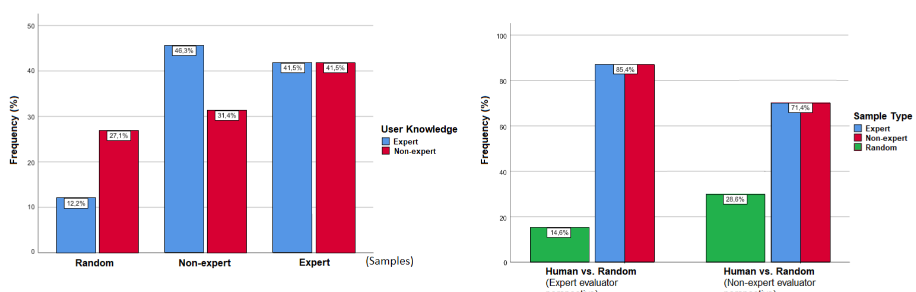


Fig. 13. a) Individual user's perception regarding how professional each sample sounded. b) All human samples (non-experts and experts) against random executions.

7 DISCUSSION AND FUTURE WORK

Given the current stage of development of our system, we believe our approach showed an immense potential to allow non-expert users to create good music that might actually sound relevant to an audience. For instance, the results show that Bubble Sounds color system helped non-experts to create high quality productions, also without constraining significantly the experts expressiveness, finding a good balance between the conceived creative freedom and supportive guideline. The equalization of these two elements in an interactive musical experience is complex, and can be compared, in the game design process, to the narrative vs. interaction dilemma; it is not simple to devise efficient mechanisms that tell stories without compromising the interactive model (and vice-versa). Commonly, one ceases for the other to appear, as we can observe in cinematic cutscenes. In music video games, as discussed before, the compositional element is commonly reduced to musical mini-games in order for the experience to be "fun" while audio variations emerge from the player's interaction with game mechanics, thus not allowing for new music to be created from scratch.

As an extension of our evaluation, we would like to discuss further the creative freedom and support for collaborative creation in partially autonomous systems, as well as other elements regarding our system and approach:

General considerations about meta-interactivity. Our approach dialogues with the concept of Metacreation of art using the paradigms of artificial life, or a-life, as discussed by Whitelaw, M. [69]. Metacreation is an interdisciplinary science focused on artificial systems that mimic the properties of living systems, explored in the 90's by contemporary artists whom appropriated and adapted these techniques to create novel forms of art through the conception of "Cybernatures", which are interactive computational systems that simulate or "mimic" ecosystems in virtual worlds. This concept relates to our approach in the attempt to bring music performance and its physical manifestations into abstract virtual environments, bringing considerations about how algorithms can efficiently work translating ideas and endeavors into something novel and engaging. Games already do that; they establish metaphors for actions that should be performed by an avatar that represents the player within the experience. And there are protocols, that are well conceived models that already became patterns, such as having directional buttons in the joystick (or any other device) for the player to move its character in the world. For Virtual Reality Musical Instruments (VRMIs), however, there are no pre-established interactive models or protocol to do so, and new patterns must be created in order to foster a bridge between musical instruments and its anatomic designs into novel ways to map it through common interactive devices, such as those utilized in games.

One element that separates interactive musical systems, such as Electroplankton [54] and Bubble Sounds to some form of ludic musical instruments is that it is harder to foster live improvisation with a band. This is not a creative freedom issue by itself, but a recurrent limitation in terms of collaborative creation, for example, when we think about these systems working in harmony in a live performance environment. Many works, like in Roberts et al. [55] attempted to experiment with autonomous approaches that learns from other instrumentalist's input to provide musical accompaniments for live performance, but it is not the kind of live improvisation that a musician would have when soloing over a harmonic rhythmic base in a free jazz concert, for example. It is not spontaneous and on the fly, it demands time for training. Interactive musical systems, to the best of our knowledge, still work upon the creation of an editor for a musical corpus to emerge. These before-mentioned interactive systems are not as bureaucratic and complex to manipulate as audio tools interfaces, such as FL Studio [10] or Pro Tools [19], but also not as responsive as musical instruments such as guitars or any other string instruments, that work as an extension of the human body, providing a quick conversion of ideas and feelings into a musical mass. They are somewhere in between.

Taking a meta-interactive approach to live performance is a very powerful way to experiment and test new interactive interfaces for musical creation, and definitely a path we will undertake in the future as an extension of this work. We believe that a more "performatic" version of Bubble Sounds, using both the VR device and the microphone input goes towards a more natural way to create music, allowing musicians on stage to work and interact with other musicians in harmony.

However, we also understand that evaluating the system at its maximum interactive capacity involves a considerable higher effort and investment to be implemented, which also points to a satisfactory and successful experiment carried out remotely in this version, which allowed us to glimpse many positive characteristics and some limitations of our approach.

We believe we advanced in regards of creating a ludic musical instrument by allowing users to manipulate all the 12 possible notes in a scale (considering its accidents) through our color translation approach, thus providing more resources for all kinds of users, regardless of their expertise level, to compose music. Past works, such as Martins et al. [43] and Escarce et al. [34] also proposed complex music to emerge from a playful cooperation between the user and a system, but the user has less control of the notes that can be chosen at certain times, since they use a dynamic grid on the environment's floor for notes to be triggered (thus, sometimes a determined note might not be at reach in the neighboring arrangement of notes, considering the user's current position in the grid).

According to Serafin et al. [59], the NIME (New Interfaces for Musical Expression) community has not shown the proper attention to the imagetic side of VRMIs. According to the authors, one of the reasons why VRMIs have not drawn its deserved attention in the HCI community might be due to the fact that musicians, sound designers and enthusiasts rely mostly in some specific points when designing this kind of interactive models, such as auditory and tactile feedback, as well as in ways in which performers interact with the audience during live performances. In addition, another reason pointed out is that only recently portable visualization devices have become accessible in a matter of cost. In this regard, we highlight the importance of the convergence between VRMIs and videogames, such as Toshio Iwai's work in Electroplankton [54] and also Bubble Sounds, especially when the goal is in proposing more ludic interactive models that aim on approximating the general public to artistic activities, such music composition. We believe this work contributed showing that visual interfaces combined with sound interfaces can create powerful experiences in terms of improving the quality of outcomes generated by partially-autonomous systems.

It is also important to encourage the thinking about new ways to conceive minimalistic interfaces for interactive musical systems, different from audio interfaces such as Pro Tools [19] and FL Studio [10], for example, that presents very complex panels with many options for the composer to create music (i.e. not friendly for lay users), as the one we proposed on Bubble Sounds, intending on providing to novice users the possibility to create complex outcomes without the necessity to manipulate many instructions at once or excessive time investment in training. Our evaluation shows that minimalistic interfaces for music creation generated very positive results in the quality of non-expert compositions if compared to experts, even with experts acknowledging non-experts production as better music in some cases. In this way, we believe we achieved interesting results in providing novel engaging ways to interact with musical systems, but there are still many obstacles to overcome. Thus, further experiments on new approaches are necessary.

Future evaluation. The version of Bubble Sounds used in our assessment was an adaptation from the original version, that can be used through a regular mouse. However, we believe this adapted version, that preserves all the original and intended functionalities of the original version, has also advantages. For instance, the usual game convention of using the mouse as control for the camera (replacing the VR device accelerometer) provides a simpler way to learn how to manipulates notes in the vortex. We assume this based on the fact that the VR devices are still not so popular, many people might have never experienced virtual reality using wearable devices.

Regardless of the interactive mode chosen, it is important to emphasize that the model of interaction (i.e. microphone or mouse) did not play a major factor in our evaluation, since our main focus at this stage was to assess the music created (and how well our system guided novice users).

For future evaluation sessions, we also intend on recording the experiments in video to perform analysis and explore in-depth aspects of participants' interactions with the system. We believe a lot can be learned from their spontaneous reactions along their experience.

The role of sentiment in the emergence of art. Our evaluation contributed to a discussion of whether or not it is possible to dissociate music production in fully and partially autonomous systems from human sentiment. After all, is it possible for a music piece to exist and be relevant without the human factor?

As presented in Figure 11 (b), we observed an expressive preference among users who participated in our evaluation process for the music created by human agents over the music generated through our random procedure. Even on a system such as Bubble Sounds, which organizes the musical corpus in a matter of temporal structure (thus fostering rhythmically coherent music to emerge even from chaotically arbitrary interaction), the perception of external listeners still acknowledged the human production as better music in both experts and non-experts scenarios.

This result empowers our system in its attempt to be a ludic musical instrument, since, differently from autonomous approaches, such as [1, 7, 29, 55], users expressivity plays a major importance in the experience. In our case, system and user depend on each other for the emergence of music, so feeling and sentiment will always be a preponderant factor in the compositions to emerge from our approach. Thus, on Bubble Sounds, the algorithm has less to do with the music creation itself, and more as a support for users to achieve more complex results in their creative process.

Additional feedback from the 1st stage of the evaluation. During the assessment sessions with users who experimented with the system, we received feedbacks regarding the interface for composition, and these feedbacks will also be taken into account when planning the next steps on the development. For instance, it was mentioned the possibility of dynamic generation of circular tracks in the vortex (for drumming patterns) in order to better guide non-expert users on the construction of the rhythmic structure of the music. Also, the dynamically generated images in the background are randomly generated in the current version, and this was noticed by some users. Indeed, at this moment, this resource is not yet being fully explored. Now, it does not yet perform the function of a compositional element (it is not yet triggered by the user's action, but as a collaborative complement from the system). This is a very promising module that can be improved in this work to help achieve better results, allowing the users to create even more engaging musical outcomes. We intend to create a mechanic guideline to propose the emergence of micro-narratives associated with sound qualities, making it become a stronger imagetic compositional element for the user to use in his/her expressivity (thus, empowering our translation approach).

Regarding outcomes from the assessment sessions. Although we could verify that the background knowledge did not play an influence in the results (as discussed in Q8 of the 1st assessment stage in the Results section), a variance may be observed if we consider other parameters such as age and nationality, for example. For instance, in Q5 of the first stage of the assessment in the Results section, we presented the different perceptions of the users in their attempt to designate a feeling to their own composition, and we intend, in the future, to track down what triggers these feelings in the human agents. Also, we intend on developing an effective way to assess how well the users understood the system features (such as color/tone relation) in order to measure how well they learned and applied the system's translation approach (i.e. color to tone) in their production.

Past Bubble Sounds experiences. We have had the opportunity to present Bubble Sounds on digital ateliers on the Fine Arts School of the University of Minas Gerais (UFMG), and its sounding outcome has been acknowledged by many artists that had an experience with it, either by listening or interacting with the system, as an "ambience music", a concept-genre defined by Eno [17] as a sound mass capable to properly mix with ambient sounds, creating soundscapes.

Since this is not a very widespread genre among music enthusiasts, given its experimental nature, it can, in some way, generate a feeling of "strangeness" in people more accustomed to the traditional musical model developed through the western's notation system. We intend to evaluate Bubble Sounds on musical performance environments, where it could actually be used as a musical instrument that work in harmony with other instruments, like in a band. At this stage, also due to limited human interactions, we only evaluated the system on its solo capacities, but it will be very engaging for us to evaluate it on musical workshops along with other voices. This kind of experimentation will certainly help us plan the next system's iterations.

Similar approaches. Similar goals to our meta-interactive proposal have been attempted using different AI approaches, such as machine learning. For instance, Wekinator [20] can generate a wide range of new instruments and ways to interact with them through gestural input (users can fully customize the way the compositional mechanic should work). Our approach, on the other hand, is more focused on allowing users to create coherent harmonic structures in a more "gamelike" way, through a minimalistic interface.

Additionally, Crosscale [6] employed VR devices to propose a virtual musical instrument interface that captures gesture input from users and presents customized instrumental mappings to allow performers from different expertise levels to play complex songs. In this system, musical notes are arranged in a grid that adjusts the distance between notes across the scale, thus fostering interesting progressions. This system relates with Bubble Sounds in many aspects, but we highlight the interactive simplicity; both systems aim on improving the learning curve for users from different backgrounds to efficiently produces high quality outcomes. The way on which both systems propose the experience is rather different, however, both in their mechanics to produce music and also in the way they are presented as a VRMI. Crosscale proposes a musical experience that resembles a piano, where left hand is used for chords and the right hand is expected to play isolated notes, like those contained in a scale. In addition, this system uses a grid for the arrangement of notes, in the same fashion as the one proposed by Escarce et al. [34]. On the other hand, Bubble Sounds uses the notion of projecting a musical score in a cylindrical structure, and presents a metaphor of a vortex that matches the idea of musical loops with different tempos in each layer. It is also detached from the analogous mechanics of a common known musical instrument, and go towards a NIME (New Interface for Musical Expression), resembling, in this way, works such as Iwai's Tenori-On and Electroplankton [31, 54].

As previously mentioned, we believe gamified interactive models should be used to empower users in more playful ways, thus allowing for a more engaging experience that can guide users to accomplish complex tasks. In the sense of proposing more playful interfaces, the concept of playification [42] was developed and presented to address this purpose and achieve more engaging results. However, it is not necessarily used as a tool to mitigate, engage and extend user's productivity in regards to the creation of works of art, like meta-interactivity. Thus, we believe interesting outcomes may emerge out of hybrid approaches of these techniques combined.

Different layers of interactivity. We can observe that the meta-interactive approach can occur and be observed in many semiotic levels. For instance, we are now focusing on a translation approach capable to transform a user proficiency into another, by giving a ludic interface that resembles something from our daily lives, such as games, and appropriate from its mechanic and aesthetic to generate a side-effect behavior, that is emergent music.

However, on a system such as Bubble Sounds, we can actually observe that the dialogue between the user and the system goes both ways; the user "plays" with bubbles and its colors to produce music, and by doing so, he/she also produces a musical corpus that emerges in the concrete reality by clapping its hands in a cadenced fashion. Thus, an input from the user in the concrete world generated changes in the state of the virtual world, but also, the virtual world ends up dictating gestural performance that also transforms the concrete reality.

Usually, users use headphones for experimenting with systems such as Bubble Sounds, and the sound outcome might be restricted only to the individual engaged in the experience. However, since clapping hands takes place in the concrete world, an observer unaware of what is happening in the virtual world might also find interesting, or at least curious, that a new sort of rhythmically coherent clapping of hands emerges out of the in-game experience.

8 CONCLUSION

The meta-interactive approach has presented an immense potential to provide novice users in a given task to achieve good performance through a more ludic dialogue with partially autonomous system. As discussed, this kind of subverting behavior has been observed in classical art in the past, as Kandinsky proposed when transforming abstract forms into musical notation systems. Variants of this model, however, are not observed in digital media, which allows a fertile experimental field for the development of new approaches.

Our experiment shows that non-expert users were capable to produce quality music through our system, with an external audience composed of experts and non-experts assessing and perceiving the sounding outcome as a coherent musical piece. As we improve our system with more functionalities, adapting it to produce multitrack music parts (such as rhythmic and melodic structures besides the harmonic), it will be necessary to run experiments with a larger pool of human subjects, especially in the first stage, with participants that actually experiment with the system in order to better confirm the interface efficiency to work as an imagetic notation system, capable to support the user expressivity without limiting it in a way of changing the composer original intent.

We believe meta-interactivity can take the human-computer relation to another level. This approach has shown to be very promising in proposing novel forms of interactive experiences, as well as proposing new perspectives to discuss art, music and games in their many cultural manifestations, especially considering its technological interventions. Our experiments showed that art can not be dissociated from the human aspect (i.e. it relates to human emotions), therefore, we believe this work can contribute on novel ways to think and design musical interactive systems, capable of supporting individuals in any form of creative effort, such as musical composition, without imposing excessive restrictions that may affect users' artistic expression.

Acknowledgments: *The authors would like to thank Andrezza Gonzalez, Lucca Bottino, Rafael Sodré, Guilherme Augusto, Washington Ramos, Jalver Bethonico for their support in the development of this work, more specifically on the recruiting process and in the development of the assessment tools used in the evaluation with human subjects of this work. We are also grateful for the studentship granted by the Faculty of Science and Technology from Lancaster University, that allowed us to conduct this research. Thank you!*

REFERENCES

[1] Nipun Agarwala, Yuki Inoue, and Axel Sly. 2017. Music composition using recurrent neural networks. *CS 224n: Natural Language Processing with Deep Learning, Spring* 1 (2017), 1–10.

[2] Willi Apel. 2003. *The Harvard dictionary of music*. Vol. 16. Harvard University Press, Harvard University.

[3] Mark Applebaum. 2012. The mad scientist of music. TEDxStanford Conferences.

[4] Florent Berthaut, Victor Zappi, and Dario Mazzanti. 2014. Scenography of immersive virtual musical instruments. In *2014 IEEE VR Workshop: Sonic interaction in virtual environments (SIVE)*. IEEE, Minneapolis, MN, USA, 19–24.

[5] BitFinity. 2016. Tadpole Treble.

[6] Marcio Cabral, Andre Montes, Gabriel Roque, Olavo Belloc, Mario Nagamura, Regis RA Faria, Fernando Teubl, Celso Kurashima, Roseli Lopes, and Marcelo Zuffo. 2015. Crosscale: A 3D virtual musical instrument interface. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, IEEE Xplore, Arles, France, 199–200.

[7] Pablo Samuel Castro. 2019. Performing Structured Improvisations with pre-trained Deep Learning Models. *arXiv preprint arXiv:1904.13285* (2019), 7.

[8] Mihaly Csikszentmihalyi. 1990. *Flow: The Psychology of Optimal Experience*. Harper & Row, USA.

[9] Magdalena Dabrowski. 1995. Kandinsky Compositions: The Music of the Spheres. *MoMA* 5, 4 (1995), 10–13. http://www.jstor.org/stable/4381285

[10] Didier Dambrin. 1997. FLStudio.

[11] N. Davis, C.-Pi. Hsiao, K. Yashraj Singh, L. Li, and B. Magerko. 2016. Empirically Studying Participatory Sense-Making in Abstract Drawing with a Co-Creative Cognitive Agent. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. Association for Computing Machinery, Sonoma, California - USA.

[12] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: defining "gamification". In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*. Association for Computing Machinery – ACM, Tampere, Finland, 9–15.

[13] Darina Dicheva, Christo Dichev, Gennady Agre, and Galia Angelova. 2015. Gamification in education: A systematic mapping study. *Journal of Educational Technology & Society* 18, 3 (2015).

[14] Mark d'Inverno and Jon McCormack. 2015. Heroic versus Collaborative AI for the Arts. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*. IJCAI 15, Buenos Aires, Argentina.

[15] Jonathan Duckworth, Nick Mumford, Karen Caeyenberghs, Ross Eldridge, Scott Mayson, Patrick R Thomas, David Shum, Gavin Williams, and Peter H Wilson. 2015. Resonance: an interactive tabletop artwork for co-located group rehabilitation and play. In *International Conference on Universal Access in Human-Computer Interaction*. Springer, Germany, 420–431.

[16] Eran B Egozy, Daniel Teasdale, Brian Thomas Stephens, Daniel Schmidt, John R Eskew, Rafael Baptista, and Kasson Crooker. 2014. Systems and methods for simulating a rock band experience. US Patent 8,663,013.

[17] Brian Eno. 2004. Ambient music. *Audio Culture. Readings in Modern Music* 9497 (2004).

[18] Andromeda Entertainment. 2020. SoundSelf: A Technodelic.

[19] Peter Gotcher Evan Brooks. 1991. Pro Tools.

[20] Rebecca Anne Fiebrink. 2011. *Real-time human interaction with supervised learning algorithms for music composition and performance*. Citeseer, Princeton University, New Jersey, United States.

[21] William W Gaver, John Bowers, Andrew Boucher, Hans Gellerson, Sarah Pennington, Albrecht Schmidt, Anthony Steed, Nicholas Villars, and Brendan Walker. 2004. The drift table: designing for ludic engagement. In *CHI 04 – Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, United States, 885–900.

[22] Sarah Victoria Gentry, Andrea Gauthier, Beatrice L'Estrade Ehrstrom, David Wortley, Anneliese Lilienthal, Lorainne Tudor Car, Shoko Dauwels-Okutsu, Charoula K Nikolaou, Nabil Zary, James Campbell, et al. 2019. Serious gaming and gamification education in health professions: systematic review. *Journal of medical Internet research* 21, 3 (2019), e12994.

[23] Darrell Gibson and Richard Polfreman. 2019. A framework for the development and evaluation of graphical interpolation for synthesizer parameter mappings. In *Proceedings of the 16th Sound and Music Computing Conference*. SMC2019, Nottingham, United Kingdom, 8.

[24] Google. 2016. Chrome Music Lab.

[25] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does gamification work?–a literature review of empirical studies on gamification. In *2014 47th Hawaii international conference on system sciences*. IEEE, Waikoloa, HI, USA, 3025–3034.

[26] Harmonix. 2007. Rockband.

[27] Simon Holland, Katie Wilkie, Paul Mulholland, and Allan Seago. 2013. Music interaction: understanding music and human-computer interaction. In *Music and human-computer interaction*. Springer, London, UK, 1–28.

[28] Amy K Hoover, Paul A Szerlip, and Kenneth O Stanley. 2014. Functional scaffolding for composing additional musical voices. *Computer Music Journal* 38, 4 (2014), 80–99.

[29] Allen Huang and Raymond Wu. 2016. Deep Learning for Music. *ArXiv* abs/1606.04930 (2016), 8.

[30] Andy Hunt and Ross Kirk. 2000. Mapping strategies for musical performance. *Trends in gestural control of music* 21, 2000 (2000), 231–258.

[31] T. Iwai. 2005. Electroplankton.

[32] Mikhail Jacob and Brian Magerko. 2015. Interaction-based Authoring for Scalable Co-creative Agent. In *Proceedings of the Sixth International Conference on Computational Creativity*. ICCC, Utah, USA, 236–246.

[33] Sergi Jordà, Martin Kaltenbrunner, Günter Geiger, and Ross Bencina. 2005. The reacTable. In *In Proceedings of the International Conference on Computer Music – ICMC*. Association for Computing Machinery – ACM, New York, NY, United States.

[34] Mário Escarce Junior, Georgia Rossmann Martins, Leandro Soriano Marcolino, and Yuri Tavares dos Passos. 2017. Emerging Sounds Through Implicit Cooperation: A Novel Model for Dynamic Music Generation. In *Proceedings of the Thirteenth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment)*. AAAI, Snowbird, Little Cottonwood Canyon, Utah, USA, 186–192. https://aaai.org/ocs/index.php/AIIDE/AIIDE17/paper/view/15887

[35] Samantha Kalman. 2014. Sentris.

[36] David Kanaga and Ed Key. 2013. Proteus.

[37] Anssi Klapuri and Manuel Davy. 2007. *Signal processing methods for music transcription*. Springer-Verlag US, United States. 440 pages.

[38] Hiroyuki Kotani. 2007. Patapon.

[39] Andrés Lucero, Evangelos Karapanos, Juha Arrasvuori, and Hannu Korhonen. 2014. Playful or Gameful? Creating Delightful User Experiences. *Interactions* 21, 3 (May 2014), 34–39. https://doi.org/10.1145/2590973

[40] Teemu Mäki-Patola, Juha Laitinen, Aki Kanerva, and Tapio Takala. 2005. Experiments with virtual reality instruments. In *Proceedings of the 2005 conference on New interfaces for musical expression*. New Interfaces for Musical Expression (NIME05), Vancouver, BC, Canada, 11–16.

[41] William P Malm. 1996. *Music cultures of the Pacific, the Near East, and Asia*. Vol. 2. Pearson College Division, New Jersey, USA.

[42] Elena Márquez Segura, Annika Waern, Luis Márquez Segura, and David López Recio. 2016. Playification: the physeear case. In *Proceedings of the 2016 annual symposium on computer-human interaction in play*. Association for Computing Machinery – ACM, New York, NY, United States, 376–388.

[43] Georgia Rossmann Martins, Mário Escarce Junior, and Leandro Soriano Marcolino. 2016. Jikan to Kukan: A Hands-On Musical Experience in AI, Games and Art. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. AAAI, Phoenix, Arizona, USA, 4377–4378. http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12123

[44] Masaya Matsuura. 1999. Vib Ribbon.

[45] J. Kent McAnally. 1995. The Bulletin of Historical Research in Music Education. www.jstor.org/stable/40214882. *SAGE Publishing* 20, 13 (1995), 19–58.

[46] Gary E McPherson. 1997. Giftedness and talent in music. *Journal of Aesthetic Education* 31, 4 (1997), 65–77.

[47] Chase Mitchusson. 2020. *Indeterminate Sample Sequencing in Virtual Reality*. Ph.D. Dissertation. Louisiana State University – LSU, USA. https://doi.org/10.5281/zenodo.4813332

[48] Shigeru Miyamoto. 1983. Mario Bros.

[49] Tetsuya Mizuguchi. 2001. Rez.

[50] Tetsuya Mizuguchi. 2011. Child of Eden.

[51] Julian Moreira, Pierre Roy, and François Pachet. 2013. Virtualband: Interacting with Styslistically Consistent Agents. In *Proceedings of the 14th International Society for Music Information Retrieval Conference*. ISMIR, Curitiba, Brazil.

[52] François Pachet, Pierre Roy, Julian Moreira, and Mark d'Inverno. 2013. Reflexive loopers for solo musical improvisation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery – ACM, New York, NY, United States, 4.

[53] Simon Penny. 2017. *Making sense: cognition, computing, art, and embodiment*. MIT Press, USA. 498 pages.

[54] Martin Pichlmair. 2007. Electroplankton revisited: A Meta-Review. *Eludamos – Journal for Computer Game Culture* 1, 1 (2007), 5.

[55] Adam Roberts, Jesse Engel, Curtis Hawthorne, Ian Simon, Elliot Waite, Sageev Oore, Natasha Jaques, Cinjon Resnick, and Douglas Eck. 2017. Interactive musical improvisation with Magenta. In *Proceedings of the Thirtieth Annual Conference on Neural Information Processing Systems*. NIPS, Barcelona, Spain. (Demonstration).

[56] Juan G Roederer. 2008. *The physics and psychophysics of music: An introduction*. Springer Science & Business Media, University of Alaska, Fairbanks, AK, USA.

[57] Katja Rogers, Matthias Jörg, and Michael Weber. 2019. Effects of Background Music on Risk-Taking and General Player Experience. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. Association for Computing Machinery – ACM, New York, NY, United States, 213–224.

[58] Katja Rogers, Maximilian Milo, Michael Weber, and Lennart E Nacke. 2020. The Potential Disconnect between Time Perception and Immersion: Effects of Music on VR Player Experience. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. Association for Computing Machinery – ACM, Virtual Event Canada, 414–426.

[59] Stefania Serafin, Cumhur Erkut, Juraj Kojs, Niels C. Nilsson, and Rolf Nordahl. 2016. Virtual reality musical instruments: State of the art, design principles, and future directions. *Computer Music Journal* 40, 3 (2016), 22–40.

[60] Miguel Sicart. 2014. *Play matters*. MIT Press, London, England.

[61] Inc. Survious. 2018. Survios.

[62] Harmonix Music Systems. 2005. Guitar Hero.

[63] Phosfiend Systems. 2014. Fract OSC.

[64] Koray Tahiroğlu, Thor Magnusson, Adam Parkinson, Iris Garrelfs, and Atau Tanaka. 2020. Digital Musical Instruments as Probes: How computation changes the mode-of-being of musical instruments. *Organised Sound* 25, 1 (2020), 64–74.

[65] David Temperley. 2007. *Music and probability.* MIT Press, USA.

[66] Ubisoft. 2011. Rocksmith.

[67] Jan Ilavsky Vladimir Hrincar. 2018. Beat Saber.

[68] Lucas Waltenberg. 2016. Novas configurações do álbum de música na cultura digital: O caso do aplicativo "Biophilia". *Revista Crítica de Ciências Sociais* 109 (2016), 185–202. https://doi.org/10.4000/rccs.6276

[69] Mitchell Whitelaw. 2004. *Metacreation: Art and Artificial Life.* MIT Press, USA. 296 pages.

[70] Zhuoming Zhou, Elena Márquez Segura, Jared Duval, Michael John, and Katherine Isbister. 2019. Astaire: A Collaborative Mixed Reality Dance Game for Collocated Players. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play.* Association for Computing Machinery – ACM, New York, NY, United States, 5–18.

Received February 2021 ; revised June 2021 ; accepted July 2021