# Human-AI musical interaction under an ethnomusicological perspective: the case of Djazz

XXX
XXX

**Abstract:** Although research on the use of computational tools in ethnomusicology has been expanding, there remains a lack of ethnomusicological studies specifically exploring the human-AI musical interaction. In this article, using the collaborative Djazz project as a case study, we discuss the performative, discursive, and sensory aspects of the interactions between several musicians and the eponymous improvisation software, shedding light on the meanings, aesthetic reactions, and epistemic disputes surrounding the AI-generated musical improvisations.

**Key words:** Artificial intelligence; Audiovisual ethnomusicology; Human-AI Interaction; Metaphor; Sensoriality.

**A interação musical humano-IA em uma perspectiva etnomusicológica: o caso do Djazz**

**Resumo:** Embora a realização de pesquisas sobre o uso de ferramentas computacionais na etnomusicologia esteja em expansão, ainda se fazem necessários estudos etnomusicológicos que explorem especificamente a interação musical humano-IA. Neste artigo, utilizando o projeto colaborativo Djazz como estudo de caso, discutimos os aspectos performativos, discursivos e sensoriais das interações entre diversos músicos e o software de improvisação homônimo, buscando lançar luzes sobre os significados, reações estéticas e disputas epistémicas em torno das improvisações musicais geradas pela IA.

**Palavras-chave:** Inteligência artificial; Etnomusicologia audiovisual; Interação Humano-IA; Metáfora; Sensorialidade.

## Introduction

Musical improvisation with artificial intelligence is a field that has been growing in recent years. Several articles provide an overview of work in this field (Tatar; Pasquier, 2019; Siphocly; El-Horbaty; Salem, 2019; Gifford et al., 2018). Dimitris Smailis (2020) develops an analysis specifically devoted to the OMax, ImproteK and Djazz software discussed here. Two other projects deserve special mention. One of these is the *Voyager* system by XXX (2021), who pioneered the field back in the mid-1980s. Based on rules describing XXX's own behavior as an improviser, it did not make use of machine learning technologies, unlike all subsequent improvisation systems. The other project is the Shimon robot, which implements a marimba player (Hoffman and Weinberg, 2011). Equipped with a camera, he watches his musical partners to coordinate with them. This visual dimension is of

considerable importance in the anthropological approach to computer-based improvisation, as we shall see in this article. There is already some research documenting experiment with musicians in this field. This can take the form of workshops with musicians and robotics engineers aimed to study how human and machine improvisation can be explored through interdisciplinary design research (Carter; Papalexandri-Alexandri; Hoffman, 2021). The visual aspect of improvisation is taken into account in work investigating the importance of intuitive visual communication channels (Greer, Fleig and Dubnov, 2024), and in work involving semi-structured interviews conducted with musicians to study visual representations during AI-driven musical performances (Hopkins et al., 2023).

Although research on the use of computational tools in ethnomusicology has been conducted in recent years (Egan, 2021; Gómez; Herrera; Gómez-Martin, 2013; Morales; Perez; Tabuena, 2024), there remains a lack of ethnomusicological studies specifically exploring the interaction between humans and musical AI. To address this gap, we focus on the interactions that musicians with diverse backgrounds, experiences, and aesthetic preferences who have been engaging in over the last decade with the improvisation software Djazz, developed by XXX at XXX. In the first section of this article, we present issues related to the functioning of Djazz and the musical, anthropological and interpersonal factors that facilitated collaborations with each musician. In the second section, we examine the discursive aspects of these interactions, paying close attention to the expressed and implicit metaphors employed by musicians when discussing AI. Finally, in the third section, we rely on audiovisual materials to explore the sensory dimensions of these interactions.

**Djazz: genesis and developpement**

Djazz is an artificial intelligence dedicated to musical improvisation. It is the result of a research project that began in the early 2000s at XXX. At the time, XXX was working on stylistic simulation in musical composition in collaboration with XXX (XXX., 1999). He used the computer to compose a variant of Bach's Musical Offering. XXX was working on jazz harmonic substitutions, a technique used by jazzmen to enrich chord progressions. Together with XXX, he created an improvisation software prototype (XXX).

At the end of 2001, XXX and XXX created OMax, a software program that improvises by capturing what a musician is playing and recombining the recorded phrases. The name OMax came from the association of two programming environments, Max/MSP on the one hand used for real-time interaction and OpenMusic (OM) on the other hand, a Lisp

library connected to Max/MSP for implementing the artificial intelligence model. The first version uses MIDI, meaning it works with notes rather than instrument sound, but XXX then adds an audio version that can capture the musician's sound itself (XXX). The optimized version of OMax was developed with the help of XXX and distributed by XXX. The software has since spread all over the world, and has been used in concerts featuring the great jazz musician Bernard Lubat.

In 2009, at XXX in XXX, XXX created a variant of OMax called ImproteK, able to play in rhythm by synchronizing to a pulse. Actually, ImproteK derived from a previous version of OMax called “OMax beat” that was not included in the distributed version of OMax. ImproteK was developed by him and XXX (XXX). It gave birth to a variant called DYCI2 implemented by XXX based on the refinement of scenarios in order to improve the control of the improvisation process (XXX). At the same time, XXX created SoMax with XXX to harmonize on the fly a melody played by a musician. XXX and XXX use the ImproteK and SoMax software in concert with Bernard Lubat in the context of free improvisation. More recently, SoMax was developed to become one of the main improvisation systems of the XXX project with contribution by XXX (XXX) and currently by XXX for his PhD. It is now freely available on github.

In 2016, XXX created Djazz, a variant of ImproteK with specific controls that enable it to be integrated into a jazz or traditional orchestra playing pulse-based music. He began working mainly with musicians from Madagascar, and he experimented a kind of virtual duo with Malagasy zitherist Velonjoro. He also started performing on stage with Malagasy guitarist Charles Kely Zana-Rotsy's jazz-world band, and more recently with great zither player Justin Vali. Djazz is the other main improvisation system designed for the REACH project. An update has been realized by XXX in 2022 and a new version is currently being implemented by XXX for an upcoming free distribution. It is entirely rewritten in Max/MSP and as in the previous versions, it relies on a special object for beat following called Antescofo.

Djazz is based on machine learning. It can capture what a musician is playing live and recombine the recorded sequences on the fly, taking into account rhythm and harmonies. The software can also include elements from a pre-recorded database. Its learning model is purely combinatorial (finite automata, pattern-matching algorithms). Thus its calculation is traceable and reproducible. It is important to notice that there is no deep learning in Djazz as it is not utilizing neural networks with multiple layers that need to be trained with a great amount of data. This is partly due to the fact that we are interested in improvisation which needs very

efficient, light and reactive systems. As soon as a few notes are given by a musician to train Djazz, then it can improvise on its own. This is why we define our use of artificial intelligence as "lean AI". The adjective "lean" refers to the fact that the software only learns from selected data, more precisely data provided by the musicians we are working with, or at least accepted by them. We do not use big data and in this sense, there is no decolonization required for this type of AI (see Adams, 2021 for a discussion on this subject).

The idea of the “acceptability” of the software's output by the musicians involved in the Djazz project is at the heart of our approach. It refers to cultural anthropology questions related to the boundaries of cultures. Can our software play music from a given cultural context, for instance Malagasy context? The answer is yes because one of the main properties of Djazz is its “agnosticism”. Obviously we do not use this term in its philosophical and religious meaning, but in a more specialized musical meaning. By “agnostic” we mean that Djazz has very little a priori musical knowledge. It just takes an input sequence marked with beats and labels that may correspond to chords or whatever, and it can improvise with it. Initially, labels were supposed to be chord labels since Djazz was designed in the context of a research dealing with jazz. But there is no assumption about the nature of these labels. Any user can define his own labels with their specific musical meaning. This feature is quite important since it allows Djazz to be adapted to very different contexts. In doing so, we are faced with this question: are the outputs of this software acceptable from the point of view of the members of a given culture?

This question is central in the anthropological investigation carried out in parallel with the development of Djazz, which requires audiovisual recording of work sessions with the musicians as we shall see. But there is another characteristic of Djazz that must be presented in relation to the artistic and cultural objectives that have been set. Djazz's generative capacities mean that its outputs are partly unpredictable. What to do when the output is wrong in some sense (i.e. not good from an artistic point of view, or not acceptable from a cultural one)? Since we are dealing with improvisation, we need to react instantly. Thus we have introduced two layers in the design of Djazz. Besides the automatic layer related to the recombination process, there also is a manual layer based on interfaces connected to the computer that allow us to control its automatic computation and to transform its result. We use standard interfaces such as slider and button sets (Korg NanoKontrol), lighted key matrices (Launchpad), or more gestural devices (Genki ring). This manual layer gives the software a high degree of reactivity. It enables the user to take instant action during improvisation.

Actions available through the interface are of different types. The first type concerns positioning in time. As we are interested in pulse-based music, time is measured precisely by the pulse. It can happen that the software is out of sync with the other musicians, and a way is needed to resynchronize it. Part of the Launchpad represents the time structure, and by pressing the appropriate key, you can resynchronize the computer with what is played around. The second type of action concerns the control of recombination. Djazz's data memory is divided into different sections, accessed via special keys from the Launchpad, and the user can choose which part of the memory he wants to be used. The third type of actions is real-time transformations applied to the output of the recombination process. One can change the playback speed of the recorded data, loop a slice with variable length measured in number of pulses, and make register jumps to octaves higher or lower.

The manual control of Djazz relies on our idea of co-creativity (XXX; for another approach see Mogensen, 2019). Indeed, we speak of co-creativity between human and artificial agents to emphasize that (1) creativity is an emergent dynamic resulting from complex interactions between heterogeneous players, (2) it cannot be understood by considering the production of each agent separately. It neutralizes the aporias associated with the question: Can artificial intelligence be creative? The idea of artificial creativity, or artificial intelligence in the arts, inevitably leads to the problem of intentionality. This philosophically difficult question is defused if we consider that creativity is shared between the agents involved. As a result, creativity is not a state, but rather a dynamic effect of interaction in a complex system. In the REACH project, various degrees of co-creativity are explored with different balances between the roles of humans and machines, depending on the level of autonomy entrusted to the latter. From this perspective, the importance of manual control in the Djazz software places it more on the side of the “intelligent instrument” than the “autonomous agent”. Thus it reintroduces the body into man-machine interaction, encouraging the development of appropriate gestures through the interfaces. This compensates for the visual impoverishment of artificial intelligence, which in a way ignores the body. In fact, during training, the software captures a musician's playing, but paradoxically it doesn't take into account the musician's bodily engagement while playing, which is very important, especially in jazz (XXX; Saint-Germier; Canonne; Fiorini, 2024).

The anthropological dimension of this research mentioned above relies on collaboration with musicians that was initiated right from the start of this project. Since the early 2000s, we've been working with Bernard Lubat. Pianist, drummer, accordionist and vocalist, he is a major figure in French jazz history. He has played with the great saxophonist

Stan Getz and organist Eddy Louiss. Since the 1970s, he has been one of the leading figures in the development of the free improvisation scene in Europe. In his hometown of Uzeste, near Bordeaux, he has created a highly multidisciplinary festival combining music, theater, visual arts and politics, a testament to his open-mindedness. This attitude explains why he was very interested in taking part in the experiments carried out with XXX and XXX on the first versions of the improvisation software in the early 2000s.

The experiment with Malagasy musicians came from a fieldwork conducted by XXX in 2000. He had the opportunity to take part in trance rituals, where he was fascinated by the music of Velonjoro, a traditional zither player. Several years later, he began a collaboration with Velonjoro, using sensors on the zither to capture his rhythmic virtuosity, and to convert it into MIDI data. In 2000, he also met guitarist Charles Kely Zana-Rotsy during a party in Antananarivo. Charles was playing in renowned zitherist Rajery's band, and they were already touring the world. In the mid-2000s, Charles moved to France to pursue his solo career and, in 2015, reunited with XXX at the XXX in XXX. They began collaborating and playing as a duet in bars to test the Djazz software in concert situations. They then played in larger concert halls with Charles's band including Julio Rakotonanahary on bass and Fabrice Thompson on percussion. More recently, XXX began a fruitful collaboration with Justin Vali, a world famous Malagasy zitherist who came to France in the mid-1980s where he met Paddy Bush, brother of singer Kate Bush, and played with her on two songs "The Red Shoes" and "Eat The Music" from her album *The Red Shoes* (1993). He recorded his own album *Ny Marina* (1994) at Real World Studios under Peter Gabriel's label. The collaboration with Justin took the form of a duet with the computer playing both Djazz and another piece of software called Ableton Live playing grooves based on Indian Ocean rhythmic patterns conceived in collaboration with electronic musician Mike Insula from Reunion Island.

**Discourses about AI**

A key factor of the development of Djazz was the conducting of a multivocal survey, through which the musicians, each with distinct backgrounds, experiences, and aesthetic preferences, talked about their experience with AI. In this sense, if alterity is a fundamental component of (ethno)musicological research (Marti, 2014, p. 646), feedback from musicians is capable of creating an "alterity of alterity", since the musicians are not only the "research subjects", but are themselves evaluators of the musical performance of an other—even if this other has a different ontological status.

One of the most striking aspects of the dialogue with musicians about artificial intelligence is the use of metaphors, understood as "understanding and experiencing one kind of thing in terms of another" (Lakoff; Johnson, 2003, p. 5) Although it is commonly understood as a figure of speech, it goes far beyond the lexical aspect. As George Lakoff and Mark Johnson point out,

> human thought processes are largely metaphorical. This is what we mean when we say that the human conceptual system is metaphorically structured and defined. Metaphors as linguistic expressions are possible precisely because there are metaphors in a person's conceptual system. [...] The concepts that govern our thought are not just matters of the intellect. They also govern our everyday functioning, down to the most mundane details. Our concepts structure what we perceive, how we get around in the world, and how we relate to other people. Our conceptual system thus plays a central role in defining our everyday realities. If we are right in suggesting that our conceptual system is largely metaphorical, then the way we think, what we experience, and what we do every day is very much a matter of metaphor (Lakoff; Johnson, 2003, p. 5).

The field of music is especially fertile for illuminating the potential of metaphor in shaping the way we relate to and interpret the world. In his proposal for a tripartite model of ethnography of musical experience composed of the dimensions of temporality, locality and metaphor, Rice (2003, p. 63) explains that rather than being abstract mental constructs, metaphors about music find resonance in bodily, sensory and interactive aspects, as well as in broader social structures. Thus,

> [the metaphorical dimension] consists of beliefs about the fundamental nature of music expressed in metaphors in the form "A is B," that is, "music is x." These beliefs then become the basis for discourses about music, musical behaviors (including all aspects of creativity, reception, performance, and institution-alization), and strategies for deploying these beliefs and behaviors in self-interested ways.

This dimension is important because it highlights cognitive mechanisms that guide, often tacitly, the conceptions of music taken for granted in a given context or by a given individual. Thus, music can be seen as art, as social behavior, as a commodity, as a cure, and so on. Being aware of the metaphorical dimension is therefore fundamental to better understanding conflicting views of the same phenomenon.

An ethnomusicological approach that takes seriously the metaphors mobilized explicitly or tacitly by the musicians with whom we relate is capable of refining our understanding of the relationship between humans and artificial intelligence—by the way, a

metaphorical term par excellence, since, although it highlights artificiality, it projects human cognitive attributes onto machines whose functioning is entirely different.

Unlike Rice, who pays attention to metaphors about music in a more global way, for the purposes of this work we will focus on metaphors specifically associated with improvisation. The first of these is that of improvisation as recombination, which guided the very way in which Djazz was developed. In this sense, more than just a technical device, recombination can be thought of as a concept regarding the nature of improvisation.

Although common sense associates improvisation with spontaneous creation in the heat of a performance, it is well known that jazz musicians often draw on materials that have been previously heard, studied, and memorized. Paul Berliner gives a clear account of this when discussing the process of learning improvisation:

> Some musicians, when learning new solos from recordings, can imitate consecutive patterns with the ease of "catching a ball and throwing it back" (CP) [Charli Persip]. More typically, they commit endless hours to the task. With each replay, they reinforce the memory of previous musical fragments, and they work on matching the next several pitches by singing or playing them. Tommy Flanagan recalls sudden displays of virtuosity by musicians like Art Tatum. "I'd say, 'Wow, what was that?' and I'd put the needle back to that spot." To master technically advanced passages, many divide them into "smaller parts," just as they might learn "long words by studying their syllables separately and recombining them" (HO) [Harold Ousley] (Berliner, 1994, p. 96).

Similarly, the treatment of small musical materials can serve as a way to avoid the literal recurrence of certain musical elements, which could be perceived as a break in spontaneity, a notion particularly valued by the jazz community:

> Artists can avoid such problems [of repeating previous presented materials] by treating planned phrases flexibly—creatively displacing, varying,and recombining them as underpinnings for new ideas. Additionally, repeating phrases in relation to particular segments of a lengthy progression separated by stretches of contrasting material presents risks different from those an artist faces if repeating phrases within short forms, where similar phrases are more likely to be noticed (Berliner, 1994, 271).

In this view of improvisation as recombination, the possibility of creation is acknowledged as not occurring "ex nihilo" but as deeply dependent on what is already known. In this sense, the existence of a stock or vocabulary of musical materials becomes fundamental, functioning as a database from which the musician—or the machine—organizes

their solo. It is no coincidence, therefore, that such a metaphor of improvisation as recombination was used for the development of Djazz; after all, these processes closely resemble computing codes, which are recombined to generate new functions within a given system.

Just like the jazz community, the feedback from musicians who performed with Djazz reveals an apprehension regarding the use of pre-existing materials, sometimes presented literally or recombined into small fragments – but always appearing unpredictably. During a recording session at EHESS, titled *XXX*[1], when asked about what interested him most when playing with the computer, Charles said: "Especially when XXX takes the musicians' phrases in there. Ray Charles, Jimmy Smith […]. I kind of feel like I'm playing with these people, in quotation marks. They make beautiful melodies, beautiful phrases, so it inspires me." On the other hand, Fabrice Thompson, a drummer who accompanied Charles in that same session, was more reticent about using pre-established phrases, highlighting his unease with transposing phrases created in one context into another:[2]

> It's strange. I'm not used to it yet, it's not easy. It's an experiment. Let's say these are choruses that come from Ray Charles, but they aren't there. That is to say, they are transposed into a piece with tempo, because he [XXX] taps with his finger there […], it's transposed into the music, and I play along with it. […] These are solos already made in one circumstance, and here they are put into another dimension. At the same time, you have to listen to what Ray Charles is doing, follow along, and make music simultaneously with where we are, so… […] We're in the groove, in Charles's musical world, and bam! A Ray Charles riff comes in. We look left and right and open our ears wide. […] He manages to bring Ray Charles into Madagascar!

Indeed, the use of recognizable phrases from the American repertoire in music from other traditions raises questions about the cultural relevance of these improvisations. One initial approach to addressing these issues was, naturally, to focus the discussion on the repertoire of American standards. Nevertheless, questions regarding the cultural appropriateness of Djazz's improvisations remained.

In our conversations with Bernard Lubat, it became clear that a simple and pure recombination of the musical materials from the database is not enough for the sounds produced by the computer to be considered musical. For him, it requires the action of the

1 The video can be found at the following link (the mentioned interaction begins at 47:38): https://www.ehess.fr/fr/media/live-session-djazz-improvisations-num%C3%A9riques-et-traditions-malgaches

2 Watch the video from the previous note starting at 51'08". Beyond the textual elements transcribed here, there are several bodily gestures by Fabrice Thompson that clearly demonstrate his discomfort in interacting with Djazz. This underscores the importance of audiovisual media in capturing aspects as significant as, or even more so than, textual ones—a point we will discuss further ahead.

person controlling the machine, so that it is possible not only to recombine but also to "retrombine:"

> Instead of recombining, one should "retrombine", that is, *tiiiioupdadadi*... like we do when the finger slips or plays an unexpected appoggiatura that we make something out of. In every chorus [i.e., a solo performed by a musician over a harmonic structure], you have that, except for the bad ones, which are perfect. This is the challenge for making the machine truly play. The places where it's not interesting are those with naïve mistakes, whereas when the improviser makes mistakes, they subjectivize it—they don't apologize.

An important aspect emphasized by Lubat in this quotation is that the process of recombination involves a recontextualization of data. In jazz music in particular, time is treated elastically, with musicians playing more or less ahead or behind the beat. If a passage is extracted from one context and placed in another, rhythmic placement can be radically disrupted.

As Berliner (1994) and Monson (1996) show, the metaphor of conversation is the one most used by jazz musicians to deal with aspects of improvisation—no wonder Monson's book is entitled "Saying Something." However, another key metaphor is that of improvisation as struggle, covering everything from aspects linked to learning improvisation to the demand to create something new in the heat of the moment. It is worth highlighting a few excerpts from Berliner's work (1994):

> Under the pressures of thinking in motion and amid the rapid-fire interplay and continually shifting frames of their conceptions, improvisers routinely contend with a variety of challenging, potentially daunting, experiences (p. 205)

> Abdullah Ibrahim likens the improvising musician to 'a samurai warrior,' implying that it is the 'fearless' confrontation of almost limitless challenges under tremendous pressure that leads to artful mastery and the ultimate achievement of 'bliss' (p. 218).

> The sense of exhilaration that characterizes the artist's experiences […] is heightened for jazz musicians as storytellers by the activity's physical, intellectual, and emotional exertion and by the intensity of struggling with creative processes under the pressure of a steady beat (p. 220).

> Young improvisers struggle to develop competency in the fundamental language of jazz and to assert their individuality within a world of criticism... the gradual mastering of the art often comes with moments of discouragement followed by breakthroughs (p. 284).

Monson (1996) also highlights this aspect:

> These moments of rhythmic interaction could also be seen as negotiations or struggles for control of musical space. One player's interjection, for example, might be experienced by another as an interruption or a challenge. Peterson told me that there were times when Allen felt uncomfortable with some of his interjections and that they had discussed the difference between enhancing a solo and obliterating it (p. 80).

> Musicians sometimes feel that other musicians are not sufficiently worthy to play with them. An "unworthy" musician can transform this image by performing well, which may include responding well to the potentially confusing musical interjections of other musicians. Such interjections challenge the soloist in much the same way as an opening insult in a verbal duel. An improviser can choose to respond to or ignore these musical ideas, but some of the most magical moments in jazz have occurred when the soloist is, as clarinetist Don Byron has put it, "poised enough to do something about them" […]. This poise is essential for defending and defining one's reputation within the fiercely competitive world of professional jazz (p. 94).

This metaphor is also present in the speeches of musicians who have interacted with Djazz. Bernard Lubat, for example, considers that not only improvised musical performance, but music as a whole is like a rugby match, and vice versa:

> Music is played like rugby: with opposition. Music that doesn't involve opposition is mere commerce. When I play, I first oppose myself when I miss a pass, and when I fail to pass the ball to my teammate, I am responsible for that failure. We cannot understand or think about freedom without being aware of our responsibility at the same time as our right to make mistakes. [...] For me, a rugby match is an avant-garde concert because it is experimentation, exploration. Each team prepares everything, and once it faces the other, it must also be ready to adapt its game based on the opposition from the opponent. For me, art and life can be like that. If in life we don't encounter opposition, we are dead; we're alone in a pickle jar soaked in vinegar.

Charles, on the other hand, imagines the interaction between musicians and the audience as a karate fight:

> People are going to be difficult to impress if you give them time to understand and decipher everything you do. It's a bit like in a fight, you say to the opponent, "Watch out, I'm going to clench my fist and I'm going to put it on your face. Watch out, it's coming!" He has time to understand everything you're doing, and he's not impressed, instead of being really impressed if you do "tac, tac, tac" [rapid mouvements].
> [XXX]: It's very funny that you use this image, because Bernard Lubat, when we get him talking like that, he says, "the music I prefer is rugby."

> Because what he likes in music is precisely confrontation, interaction, and he says his favorite music is rugby.
> [Charles]: In that case, I prefer karate.

In the specific case of the interaction between a musician and a musical artificial intelligence, there are intersting aspects that this metaphor of improvisation as a struggle can illuminate. The first relates to the fact that AI constantly introduces unexpected elements—with varying degrees of aesthetic success, it is worth saying. In this struggle, the musician often feels powerless against an AI that seems to follow its own designs and operates with an inaccessible internal logic, leaving them with no choice but to respond to it. Fabrice Thompson admitted that one thing that bothers him is that when he sees a musician, he can anticipate what they are going to do, whereas here, he cannot anticipate what XXX is doing on the pad. As he himself says:

> [When I play with a musician] I can hear him and think, 'Ah, he's going there, this is happening, he's doing something […],'because there is interaction. Whereas the musician [in the database] has already done their thing. And we play based on what they've done, on their chorus, on their interventions […]. So, we adapt to it.

From Thompson's statement, it becomes clear that the "fight" here should not be understood as a competition to defeat the AI, as in an evaluative contest to determine who performs better—like the famous 1997 duel between chess world champion Garry Kasparov and the supercomputer DeepBlue. Instead, the goal is to be in harmony with the other. In this sense, the act of "retrombiner" is another manifestation of this metaphor, as it demands constant attentiveness from the human musician in directing and correcting the AI so that it aligns with their aesthetic-musical goals.

At the same time, the struggle with AI can push the musician out of a habitual response logic, broadening their aesthetic and creative horizons. Bernard Lubat clearly speaks to this in one of his performance feedbacks, attributing an unexpected musicality to the AI:

> That's very good, and even if there are two or three strange spots, they aren't foreign notes—they are foreign to the harmony, but good. That's excellent. [...] There, it's very good, the chorus is remarkable. [...] I don't understand, it creates a chorus that is musically magnificent. It isn't outside the harmony; it is strangely musical.

It’s worth mentioning one last metaphor brought by Justin Vali. Born and grown up in the rural rice fields near Antananarivo, in a family of instrument makers, he is very attached to his peasant origins, and has a deep knowledge of nature, which not only inform the materials needed to build a bamboo zither, but also his view of music:

> Everything that comes out of the computer, just as everything that is acoustic, has the same source: it's nature that gives the sound. If you don't have the raw materials of the earth, it doesn't come out. Before sound technologies, we used to play in caves or other places where it sounded good. The Globe is the great sound system, and it's the same departure for technology: everything that comes from the Earth, gives satellites to capture the universe, gives communication, gives music. I don't think technology harms nature at all, because we use the same magnetism.

This metaphor is especially important because the use of artificial intelligence in the context of traditional music can give rise to problems of negative representations by fans of this music about the unexpected appearance of technology in this context. But contrary to intuition, traditional musicians are not hostile to technology—quite the contrary. It is the unveiling of this metaphor that help us understanding the full commitment of Justin Vali with such experimentations with AI, which otherwise would be seen just as an appropriation or misrepresentation of Malagasy traditional music.

**Sensoriality**

Since the early days of developpement of Djazz, it was common for XXX to make audiovisual recordings of the work sessions with each of the musicians mentioned here. The main purpose of these recordings was to later revisit the musical performances and participants’ feedback with the aim of making possible refinements to the AI. In 2021-2022, when XXX carried out a research internship at EHESS, we decided to undertake a new use for the audiovisual media: instead of just being a research film, “the gathering of audiovisual data collected with a camera or camcorder during fieldwork, with the purpose of being used as ethnographic-musical data for research, analysis and restudy” (D’Amico, 2020, p. 41), we would make, simultaneously with conducting an ethnography centered on Charles Kely Zana-Rotsy (XXX), an ethnomusicological film in which aspects related to his music and his life story would be dealt with. From this process came the film *Open Gasy*, which, through presenting Charles's achievements and challenges as a “world musician,” offers a glimpse into the workings of the French world music scene (XXX). The experience with the making

o f *Open Gasy* was crucial for later making an ethnographic film about Djazz, recently released[3].

The painstaking process of revisiting the filmed material, selecting it and organizing it into a narrative plot made us consider that the audiovisual media could make an important contribution to a sui-generis look at the relationship between musicians and musical artificial intelligences, both from an interactive, material and sensory point of view, aspects that would be very difficult to access if we only paid attention to the verbal-discursive level.

It is certainly possible to imagine a textual description of the musical, kinetic and material aspects of a performance with such detailed AI that a large part of its qualities could be accessed by the reader—and, it should be remembered, it is precisely in this imaginative process that much of the process of reception of a literary work resides (Jauss, 1979). Even so, something of Velojoro's smile, to cite an iconic moment in the interaction with Djazz, to be discussed later, would be lost. In this sense, by capturing the gazes, gestures and expressions of the participants, the audiovisual is able to broaden the sensory perception of what it means to use AI, revealing the complexity of its use not only in terms of its ethical and aesthetic implications, as discussed in the first part of this article, but also in the ways we experience it bodily (Pink, 2009). Thus, rather than a commitment to a supposed ethnographic truth, this attention to sensoriality is capable of making us identify with the characters in a visceral way. After all, films "stretch the boundaries of four consciousnesses and create affinities with bodies other than our own" (MacDougall, 2006, p. 17).

The work sessions with Bernard Lubat were full of these moments when gestures or facial expressions shed decisive light on what was happening musically in the interaction with the AI[4]. A good example of this situation is an experiment session in February 2013 where Bernard Lubat was improvising on a tune entitled *Cécile ma fille*, which is a song by famous French singer Claude Nougaro based on a 3/4 signature. After his solo, Bernard leaves the piano and comes close to the computer to listen to how what he has played is taken up by the machine. The video shows him dancing the waltz (at 2:11). Compared with other situations where Bernard has grimaced at the computer's less-than-successful productions, this choreography provides a clear indication that what the computer was playing at that moment was just right for him.

Later in the video, the launchpad interface is used to transform the results of the recombination process with accelerations and superimpositions that break up the sequences

3 The film can be viewed at the following link: XXX
4 LINK

produced. These transformations of his own material could have been perceived negatively by Bernard Lubat. The opposite was true. Bernard Lubat began to make gestures that mimicked the processing carried out by the machine (at 4:00). His subsequent comments proved that he was interested in these unusual transformations and he elaborated on the idea of "transgressing" the rules which is allowed by the machine through the transformations offered by the interface. Thus, these transformations were encouraged by him, and they made it possible to concretize the notion of *trombination* introduced above in the way of interacting with the machine.

Another example of facial expression captured by the video which is of great importance in the experimentation with musicians is a smile by Velonjoro that was captured during a duet with him and the computer in 2016. Such a detail may be accused of being over-interpreted and we must be careful not to engage in quite elaborate speculation about what that smile meant. We have made the assumption that this smile indicated Volonjoro's satisfaction at hearing the computer join him on the right pattern after a break formula (XXX). This speculation would of course require more concrete feedback from Malagasy experts. Moreover, from a postcolonial point of view, one might consider that a northern researcher studying the music of a less economically privileged society should be cautious in their interpretation of a practitioner's response, without that practitioner having a say and without a systematic attempt to solicit the opinions of a larger number of local practitioners. But the fact is, traditional musicians don't necessarily talk about music. At least, they don't describe music in terms of academic theoretical language. Facial expression, gestures and mimicry are often the only way to access what the musician has in mind. Obviously, such signals obtained from video recordings need to be collected in sufficient numbers and compared with each other if we are to be able to derive a sufficiently credible interpretation. We defend the idea that these signals are a fundamental aspect of ethnomusicological investigation, since musicians don't talk much about music, but it makes them react a lot. For instance, in another video of a duet with the computer in 2014[5], Velonjoro gives a thumbs-up at 1:22 as a sign of acquiescence about the computer output.

The fact that this duet was validated by the musician's gesture indicates that the computer was able to reproduce his playing convincingly according to him, which was confirmed by subsequent playback listenings. As it happens, Velonjoro passed away in 2017, but his playing data has been preserved. This raises new questions about the role such data can play in heritage conservation.

5 LINK

A third example of reaction to the computer's improvisation concerns the guitarist Charles Kely Zana-Rotsy. In the video captured during a concert in Uzeste in August 2023[6], he turns to the computer at 04:51, as the artificial intelligence generates a phrase that goes too far into the high register. Here again, the musician's reaction can be clarified by adding contextual elements and comparing it with other reactions in similar situations. Charles had previously criticized the AI's improvisation in the piece *Masoala* which is played at that moment in the video. The model is trained on data from the great jazz pianist Joe Zawinul that we have transcribed taken from his concert at Munich Piano Summer in 1985 (at 13:03, Partie I)[7]

In this passage, Zawinul is playing alternatively "in" and "out", according to the terms used in jazz context to describe this phenomena, which means that he improvises on the notes of the tonality and sometimes outside the tonality. When recombined by the computer, the musical logic of these sequences alternating "in" and "out" passages may be disturbed. This is what Charles criticized about another video of Justin playing *Masoala* in duet with the computer (Djazz solo at 2:50 based on the same Zawinul's data). On Charles's viewing of the *Masoala* video[8], he noticed Justin's grimaces as an indication of his discomfort with the machine's output. He wrote in an email: "He's starting to change his face, [...] then he comes back normal, there are a few seconds where you can't see the face. Then maybe it's me too, but as I said in my previous message, he's out right away and almost all the way through, that's too much dissonance, at least for me". Actually, Justin seems a bit surprised when hearing an "out" passage (at 2:55), but his face becomes smiling again when the solo goes back to an "in" passage (at 3:02), and he is moving his body in synchronization with the computer solo. Charles himself admits that his judgment may be subjective, and that other people may perceive the solo differently. As we have said, these indications need to be confirmed by multiple occurrences in order to provide a correct interpretation of the phenomenon.

With the exception of fictional films such as *I, Robot* (dir. Alex Proyas, 2004) and *Ex Machina* (dir. Alex Garland, 2015) that depict artificial intelligence in humanoid forms in a dystopian future scenario, artificial intelligence is generally associated with the realm of the immaterial or virtual (algorithms, data flows, etc.). However, it should be remembered that it is often linked to a set of material realities which, rather than providing the physical

6 LINK
7 LINK
8 LINK

infrastructure for its operation, condition the ways in which it is conceived, implemented and, above all, experienced. This is especially relevant in the case of Djazz, since, unlike an AI whose material identity is fragmented as it is incorporated into different contexts and devices —which is being foreseen, by the way—its use, to date, is intrinsically associated with control by the user and its creator, behaving in a certain way like a conventional instrument.

The first tangible aspect that stands out in the case of Djazz is the presence of a computer and a Korg NanoKontrol controller. While in other contexts, such as in a studio, the controller functions as a practical instrument, here it takes on another dimension. First, the importance of the fingers must be highlighted. It is through them that the reference tempo is established. From the perspective of a spectator in a concert, this might seem like a trivial detail. However, when hyper-focused through the lens of a camera, the viewer can understand how the metaphor of improvisation is enacted not merely through discourse but corporeally. Beyond a purely functional aspect, what is revealed is the localized, repetitive, and obsessive effort, as well as the tension and focus required to ensure the pulse is precise—a fundamental factor for the synchronization not only of musical timing but also of "their energies, their intentions, and their moments of inspiration" (Borgo, 2022, p. 10), which can lead to an effective, pleasant and meaningful interaction between the AI and the musicians.

**Conclusion**

Unlike ChatGPT and other generative AI based on language models, Djazz evidently is not capable of articulating a discourse about itself in words. Therefore, when considering the human-IA relationship, the importance of the human discourse becomes evident not only for producing narratives and meanings about the interactions or the "works" created but also for attributing agency to the AI itself. It is precisely this attribution of agency through discourse that lies at the heart of the relationship between humans and AI. After all, it is this that allows the former to move beyond the complexities of computational technicality and explore aesthetic, social, and cultural issues—far more engaging aspects of interaction than what a purely structural approach could offer.

At the same time, as we have seen, much of what constitutes a performance remains in the realm of the unspoken, conveyed only through the spontaneous and often unconscious reactions of the participants. In this regard, the field of audiovisual ethnomusicology (D'Amico, 2020), by enabling the exploration of the gestural, sensory, and material aspects

of a performance, can provide access to aesthetic judgments in a much more refined and subtle manner than verbal discourse could achieve.

Even though rapid technological progress may, in the not-too-distant future, enable AI to create and improve its own systems, human sensitivity will remain an essential factor in the cultural evaluation of AI-generated products. In this sense, we believe that a multivocal ethnography, based on specific uses and anchored in long-term experience (Suchman, 2007) as well as attentive to the performative, discursive, and sensory dimensions of musicking, can reveal ways of using and reflecting on technology as diverse as its users. By focusing on the performative, discursive, and sensory features of human-AI musical interaction, ethnomusicology can uncover a fertile field of investigation, offering significant contributions to a more nuanced understanding of a phenomenon that is only just starting to become pervasive.

**Funding**

This research is supported by the European Research Council (ERC) REACH project, under Horizon 2020 programme (grant n. XXX); by Agence Nationale de la Recherche (ANR) project MERCI (grant n. XXX); and by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (grant n. XXX).

**References**

ADAMS, Rachel. Can artificial intelligence be decolonized?. *Interdisciplinary Science Reviews*, v. 46, n. 1-2, p. 176-197, 2021.

XXX

XXX

XXX

XXX

XXX

BERLINER, Paul F. *Thinking in jazz: The infinite art of improvisation*. Chicago: University of Chicago Press, 1994.

BORGO, David. *Sync or Swarm : Improvising Music in a Complex Age*. New York, NY: Bloomsbury Academic, 2022.

CARTER, Anthonia, PAPALEXANDRI-ALEXANDRI Marianthi; HOFFMAN, Guy. Lessons From Joint Improvisation Workshops for Musicians and Robotics Engineers. *Frontiers in Robotics and AI*, v. 7, p. 576702, 2021.

XXX

XXX

XXX

XXX

D'Amico, Leonardo. *Audiovisual Ethnomusicology:* Filming Musical Cultures. Bern: Peter Lang, 2020.

EGAN, Patrick. Insider or outsider? Exploring some digital challenges in ethnomusicology. *Interdisciplinary Science Reviews*, v. 46, n. 4, p. 477-500, 2021.

GIFFORD, Toby et al. Computational systems for music improvisation. *Digital Creativity*, v. 29, n. 1, p. 19-36, 2018.

GÓMEZ, Emilia; HERRERA, Perfecto; GÓMEZ-MARTIN, Francisco. Computational Ethnomusicology: perspectives and challenges. *Journal of New Music Research*, v. 42, n. 2, p. 111-112, 2013.

GREER, Ross; FLEIG, Laura; DUBNOV, Shlomo. Creativity and Visual Communication from Machine to Musician: Sharing a Score through a Robotic Camera. 13th EAI ArtsIT Conference 2024. *Proceedings* […]. New York University, Abu Dhabi, Nov. 2024.

HOFFMAN, Guy; WEINBERG, Gil. Interactive improvisation with a robotic marimba player. *Autonomous Robots*, v. 31, p. 133–153, 2011.

HOPKINS, Torin et al. Virtual AI Jam: AI-Driven Virtual Musicians for Human-in-the-Loop Musical Improvisation. The International Conference on AI and Musical Creativity (AIMC) 2023. *Proceedings* […]. University of Sussex, 2023.

JAUSS, Hans Robert. A Estética da Recepção: colocações gerais. *In*: LIMA, Luiz Costa (org.). *A literatura e o leitor*: textos de estética da recepção. Rio de Janeiro: Paz e Terra, 1979. p. 43-61.

LAKOFF, George; JOHNSON, Mark. *Metaphors we live by*. Chicago: University of Chicago Press, 2003.

XXX

MACDOUGALL, David. The corporeal image: film, ethnography, and the senses. Princeton: Princeton University Press, 2006.

MARTÍ, Josep. Music and alterity processes. *Humanities*, v. 3, n. 4, p. 645-659, 2014.

MOGENSEN, René. Human-machine improvisation: an analytical approach to human-computer co-creativity in improvised music. In: Computational Creativity Symposium at AISB Convention 2019. *Proceedings* […]. Falmouth University, 2019.

MORALES, Glinore Santiago; PEREZ, Mary Leigh Ann Corpus.; TABUENA, Almighty Cortezo. Artificial Intelligence and the Integration of the Industrial Revolution 6.0 in Ethnomusicology: Demands, Interventions and Implications. *Musicologist*, v. 8, n. 1, p. 75-107, 2024.

SIPHOCLY, Nermin Naguib J.; EL-HORBATY, El-Sayed M.; SALEM, Abdel-Badeeh M. Analysis of Computational Intelligent Techniques of Music Generation. *Egyptian Computer Science Journal*, v. 43, n. 3, p. 49-64, 2019.

XXX

XXX

XXX

PINK, Sarah. *Doing Sensory Ethnography*. London: SAGE, 2009.

XXX

RICE, Timothy. Time, place, and metaphor in musical experience and ethnography. *Ethnomusicology*, v. 47, n. 2, p. 151-179, 2003.

RICE, Timothy. *Ethnomusicology*: A very short introduction. Oxford: Oxford University Press, 2014.

SAINT-GERMIER, Pierre; CANONNE, Clément; FIORINI, Marco (2024). The corpus' body. Embodied Interaction from Machine-Learning in Human-Machine Improvisation. The International Conference on AI and Musical Creativity (AIMC) 2024. *Proceedings* […]. University of Oxford, 2024.

SILVERS, Michael. Attending to the nightingale: on a multispecies ethnomusicology. *Ethnomusicology*, v. 64, n. 2, p. 199-224, 2020.

SUCHMAN, Lucille Alice. *Human-machine reconfigurations: Plans and situated actions*. Cambridge: Cambridge University Press, 2007.

SMAILIS, Dimitris. *Artificial Intelligence Music Generators in Real Time Jazz improvisation: a performer's view*. Master thesis. National and Kapodistrian University of Athens, 2020.

TATAR, Kıvanç; PASQUIER, Philippe. Musical agents: A typology and state of the art towards Musical Metacreation. *Journal of New Music Research*, v. 48, n. 1, p. 56–105, 2019.