Towards Augmenting Communication in Human-AI Music Improvisation

Firstname Lastname

Affiliation Information

email@email.com

Abstract

Creative collaboration involves a dynamic interaction between participating agents where communication (explicit and implicit) plays a critical role. In this paper we carry out a study that explores how to augment communication in human-AI creative partnerships in real-time music improvisation. Specifically, we are interested in understanding what are the ways in which humans and AI music systems can signal to each other their musical state or musical intention. In order to investigate this, we present a theoretical and experimental approach that enables us to explore the ways in which human musicians improvise with an AI sequencer. The experiments involve annotating the AI/human performances to indicate where communication of intention between AI and human would have improved the interaction and the resulting musical output. Through thematic analysis of the human experience, we identified key features of communication that users with different levels of expertise require in order to maintain a good interaction with an improvisation music system. In addition, we identified the use of *silence* as a novel medium of communication that can open up new possibilities for creative musicianship with AI in practice and performance.

Introduction

The idea of humans and machines improvising together in performance provides an array of challenges - musicians often have little or no sense of how machines work, machines can be limited in the musical information they can 'listen' and convey from/to their human partners, and there are limited opportunities for communication during an improvised performance (something human musicians learn to do naturally through a range of different linguistic and non-linguistic devices) Seddon (2005); Givan (2016); Walton *et al.* (2015).

The study of building AI systems to simulate new kinds of performance possibilities for musicians and audiences through AI-Human interactions is, and has been, a key challenge for AI designers. This is not only interesting in its own right but is also a critical field that enables the exploration of the nature of human creativity more generally Sawyer (2000, 2003). Not only does improvisation provide us with ways to explore the nature of creativity, but it is also regarded highly as an arts practice in itself. The ability to improvise on stage with others is considered to be one of the great achievements or examples of human creativity. It requires empathy, quick-wittedness, an awareness of the genre, an awareness of the audience and each other, and a huge source of patterns that can be used appropriately given the correct contextual settings. The success of improvising in a group further requires a huge range of non-linguistic communication to signal musical state and intention to others Seddon (2005).

Designing AI systems that can perform successfully is then a key challenge for demonstrating both the potential of AI, but also its philosophical and technical limitations. It is this notion of the musical experience of performance that has driven our recent work. We are interested in understanding to what extend we can use AI to challenge and provoke human creativity, and to understand how we might build systems that can enable new kinds of creative practices between humans and AI systems collaborating in creative contexts.

In this paper we present a study of human-AI co-creation that explores the role of communication in the artistic domain of music improvisation. We contextualise our work both within the context of what it is to musically improvise with other humans but also within a trajectory of music technology development which aims to create increasingly powerful tools for music performance. However, instead of looking to develop increasingly sophisticated models, our work is differentiated from previous work by focusing on the exploration of the kinds of interactions musicians want to have with these systems and how we can facilitate them.

We are particularly focusing on the role that communication can play during human-machine improvisation, as improvising agents are implicitly (and sometimes explicitly) required to explain any decisions they are making, both during improvisation and also in any dialogue after the improvisation, in order to build more understanding, more awareness and lead to more musically accomplished and satisfying performances. We believe that studying communication in music improvisation provides us with an ideal opportunity to explore notions of creativity Samek and Müller (2019), and how this can help to improve the human-machine interaction in music improvisation in particular.

The paper is structured as follows: first, we provide a brief description of research on the role of communication in human-human improvisation, and in human-AI settings. Then, we present the system infrastructure that enabled this study. We follow by discussing our methodological approach to understanding the role of communication in musical improvisational interactions and present a set of design considerations that highlight factors we should consider in the design of musical improvisational systems. We finish with some conclusions and remarks of future work.

Background on AI and Music Improvisation

When musicians perform together they communicate with each other primarily through the sonic output. However, other forms of communication take place between performers in an improvisation through the use of visual and verbal cues. To illustrate, a recent study investigated how timing information is included in musicians' cueing-in gestures Bishop and Goebl (2018), while in Bishop *et al.* (2019) it was concluded that periods of unpredictable behaviour, as well as familiarity between performers, motivated visual interaction. Other work has studied other more unusual communication elements; for instance, in Chew (2014) it was identified that human musicians often rely on breath cues to coordinate joint actions.

In human-machine partnerships, human performers and musical agents usually communicate directly through messages, or indirectly through the music Tatar and Pasquier (2019). Communication via messages relies on pre-defined protocols of how the system sends and receives messages, and makes corresponding responses. For example, if a human performer wishes the collaborative musical agent to play louder, the human performer can explicit request this through an interface with the system. Alternatively, this can also be captured by the system by identifying the raise of volume led by the human performer from the environment.

Alternative, and less conventional, mediums of communication in human-machine improvisations have been proposed. An example of this is provided in McCormack *et al.* (2019) where the authors equipped an AI musician with the ability to continuously communicate how confident it felt during an improvised performance. Human performers also implicitly communicated their confidence to the computer via biometric signals. The work showed that this type of simple, interpretable communication increased the flow within the human-AI collaboration and the quality of the music produced. More recently, Thelle and Pasquier (2021) proposed a communication strategy in the Spire Muse system through the use of a negotiating interface manipulated by the human musician through foot pedals, which, when operated, would send messages to the system in order to request a change of behavior or to signal positive feedback for the current interaction (this last type of message would keep the system in the same state for the next 30 seconds). However, active and explicit communication is mostly carried out by the human musician only. Other works like Biles (1994) and Brown (2018) also use real-time adjustments to the music and performance output as the method of communication between human performers and machines.

Other relevant works on human-AI musical improvisation systems have either not explored communication methods, or have done it in very subtle ways. The "SpeakSystem" Yee-King and d'Inverno (2016), an autonomous AI system that improvised with one of the UK's leading sax players, illustrates the former case. In this system there was no mechanism for signalling intention in either direction between machine and human, instead the human performer responded to the various musical challenges of being taken out of their comfort zone and not relying on any of the cues that they would get improvising with a fellow human. The reflexive looper Pachet *et al.* (2013), on the other hand, used a style of playing (bass lines, chords or a melody/improvised line) to 'guide' the looper into playing the missing parts. More subtly, the looper also keyed on energy (a function of the number of notes and their volumes). Whilst this was perhaps one of the most musically satisfying systems in a decade, the looper did not have autonomy as such and so did not signal musical state/intention back to the human musician.

Our research has partly been inspired by recognising that in order to build longer and more sustained performances we would need to build more sophisticated AI systems that include greater opportunities for communication; a research direction that has not only been identified for the context of music, but for creative domains in general Llano *et al.* (2020).

System Description

Technical framework

To undertake our experiments, we have developed an integrated application using the JUCE framework juc. We chose this framework as it is specialised for creating music software, and it allows the creation of cross-platform, fully native applications, allowing us to deploy MacOS, Windows and GNU/Linux versions in our study.

Our AI system is based around musical note processing, working with pitch, timing and expression, but not timbre. Hence, the application is essentially a MIDI processing system (MIDI, or Musical Instrument Digital Interface, is a ubiquitously adopted protocol for communication between digital musical instruments). The application receives MIDI from a connected MIDI keyboard, analyses it, then sends MIDI back out. We use MIDI information because it is straightforward to extract and generate polyphonic note data with this format, rather than analysing incoming audio.

The MIDI information generated by the system is used to generate music via a sample-playing software synthesiser. In the setup for our experiments, we used a piano sound for the human player and a vibraphone sound for the AI player. We chose these sounds as they are familiar, expressive and have distinct timbres, allowing the human player to easily differentiate their own playing from that of the AI. We next describe the AI system's technical implementation.

The algorithm

The AI system uses a variable-order Markov model (VMM) as its core algorithm Begleiter *et al.* (2004). Computer music researchers commonly use VMMs to model musical sequencesTatar and Pasquier (2019), and they are appropriate here as they are well understood and can be trained from scratch in real-time on live inputs. Our VMM

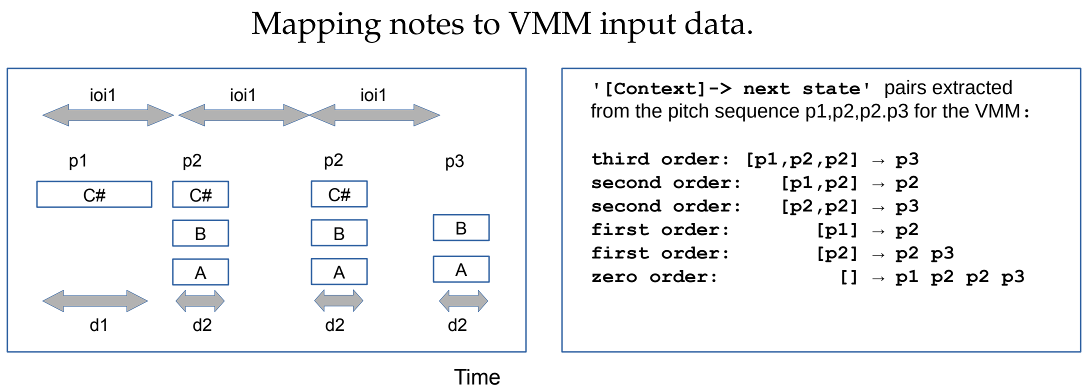

Figure 1. Left panel: illustration of how a note sequence is converted into inputs for the VMMs in the system - pitch, duration and inter-onset-interval. In this example all the IOIs are the same, there are 3 different pitch events and 2 different durations. The loudness of the notes, not shown, is fed into a fourth VMM model. Right panel: illustration of how the VMM extracts multi-order contexts from the simple four event pitch sequence. In this example, only the zero and first orders have more than one possible output.

builds a generative model mapping 'context' symbols to 'observation' symbols. Context is the last n musical events, and the observation is the musical event that follows that context. The model has variable order in that it maintains several contexts in parallel, with all values of n between 1 and N. The VMM addresses the zero-frequency problem, wherein some contexts have no observations, by increasing the number of available contexts.

In training, the model builds the context-observation pairs from incoming MIDI data generated by the human musician. In order to generate output, the model needs a context in which it can search for available observations. The generation context starts empty, and then as the model generates output events, these are used to build the context sequence. Each time the model wants to generate an output, it starts with the highest-order context. If it finds one or more observations for that context, it will select one using a random process wherein more frequent observations are more likely to be selected. If there are no observations, it reduces the order of the context iteratively until it finds an observation. If the model does not find any observations for any ordered contexts, it will pick randomly from all available observations (i.e. zero order).

The system uses four VMMs to discretely model four musical features: polyphonic pitch sequence, inter-onset interval (IOI), note duration and loudness. IOI is the elapsed time between the start times of the notes. Figure 1 illustrates how these features are extracted from a simple note sequence.

We have implemented the VMM in C++ and the source code is available in our github repository (link to be provided upon acceptance in order to maintain anonymity).

Methodology

Participants: We conducted a qualitative study with nine participants (n=9) with a different range of music experience. Four participants were female, and five were male, all aged between 20 and 50 years old. Five participants had more than 10 years of professional training on an instrument, while among the other 4 participants, 2 had less than 0.5 years of professional training and 2 varied between 3 and 9 years of training. Additionally, 2 participants had 0 years of music theory training, while the other 7 participants' training was distributed between 1 to 7+ years of music theory training. All of them experienced a regular, daily practice of a musical instrument for 4 to more than 10 years. While all participants could play at least 2 instruments, 3 considered that piano was the instrument they played best, 3 selected guitar and the other 3 were best with voice.

To obtain a more objective measure of the level of expertise in music, we used The Goldsmiths Musical Sophistication Index (Gold-MSI) Müllensiefen *et al.* (2014), which is a psychometric instrument that takes into account a wide range of musical activities, not limited to instrumental training, to give a more comprehensive measurement on musical skill. The Gold-MSI self-report questionnaire quantifies 5 specific factors of skilled musical behaviour, namely *active engagement*, *perceptual abilities*, *musical training*, *singing abilities* and *emotions*, and one general factor of *musical sophistication* that correlates the 5 specific factors. For our study we focused on the factors that referred to Perceptual Abilities and

	Scores of Participants for each Factor		
Percentile	**Musical Training**	**Perceptual Ability**	**Music Sophistication**
90>	94, 92, 90, 94	-	95, 90, 93
80-89	80, 83	84, 87	86, 87
70-79	72	77, 77, 77	75, 76, 78
60-69	66	65, 69	61
50-59	-	51, 56	-
40-49	41	-	-

Table 1. Results of the Gold-MSI for assessing the musical experience of the participants in the study. Three factors are assessed: Musical Training, Perceptual Ability and the general Music Sophistication factor. The table shows the scores of the participants that fall within a specific percentile range when compared to the general population for which the Gold-MSI assessement was created (n=147,63).

Musical Training, and the General Music Sophistication factor.

Table 1 shows the results of calculating the scores for the three selected factors using the *GMSI Scoring App* gms (b) and subsequently placing them within a percentile chart of the general population for which the index was generated gms (a); for instance, the first entry for the musical training score in the table means that 4 of our participants have more musical training than 90 percent of the population (from which the Gold-MSI assessment was created: n=147,636), the second entry in the perceptual ability column means that 2 participants have higher perceptual abilities than 80 percent of the population, and so on. These results show that we have a representative group of participants in terms of their musical experience.

Procedure: In each session the participants were asked to work together with the system to improvise a short piece of music. The improvised piece was recorded and uploaded into the online music annotation platform *Music Circle* mus. Following the improvisation, the participants were asked to comment on the improvised piece and their interaction with the system using *Music Circle*. Example annotations are shown in Figure 2. Participants could listen back over sections of the improvisation and add annotations at

Figure 2. Music Circle annotation platform showing participant annotations along a timeline of their collaborative improvisation with the system. Four audio wave forms are shown stacked vertically from top to bottom. In each wave form a small section of the audio recording has been highlighted and a comment is shown.

relevant points on the improvisation's timeline. We encouraged participants to include comments about their interaction with the system and their perceptions of the system's creative processes at any point during the improvisation.

A follow-up semi-structured interview was then undertaken to understand relevant observations further. The semi-structured interview followed a format in which participants were first asked questions to gauge their experience improvising with the system, and to comment on their experience *communicating* with the system during the improvisation. Follow-up questions were asked to probe the communication mechanisms the participant has used when improvising with other human musicians and to envision how some of these communication mechanisms could be adapted to improve communication with an AI system.

Thematic Analysis

Transcriptions from the semi-structured interviews and the participants' annotations in *Music Circle* were coded using Braun and Clarke thematic analysis methods Braun and Clarke (2006) to identify common themes. Two researchers independently coded each participant's annotations and interview transcripts to produce a list of preliminary codes, before comparing and combining their analyses and settling on the final set of themes (Table 2) described next. The corpus of data collected during this study has been made publicly available anonymised for review process (2021).

Theme	Codes	Code references (% of Theme Total)
Human Interaction	Musical interaction	58%
	Physical interaction	31%
	Past experiences	11%
Understanding the system	Expected musical knowledge	68%
	Experimenting	29%
	Understanding patterns	3%
Structured Improvisation	Leading and accompanying	67%
	Pre-performance baseline	18%
	Stages during performance	15%
Explicit Communication	Signal change to the system	39%
	Signal good interaction	22%
	Signal unwanted interaction	19%
	Signal end of performance	13%
	System signalling to user	7%
Implicit Communication	Musical cue	68%
	Physical cue	32%

Table 2. Themes and codes identified through thematic analysis of participant interviews and annotations.

Human Interaction

The theme **Human Interaction** explores the participant's previous experiences improvising with human musicians and highlights the types of communication mechanisms they have used in this context. Analysis of this theme identified two distinct styles of communication that participant's make use of when improvising with other human musicians, codes *Musical Interaction* and *Physical Interaction*, as well as identifying the importance of the musicians *Past experiences* improvising together. In this context, the code *Musical Interaction* refers to communication between musicians that occurs solely through their instruments and the music they create. *Physical Interaction*, on the other hand, refers to communication with another musician through a physical medium, including eye contact, body language and use of the voice. The code *Past Experiences* explores the idea that when musicians play and improvise together on a number of occasions, they start to learn aspects about the other musician's musical style that can improve their communication with one another.

Insights gained from understanding musicians' physical mechanisms when improvising together provide a valuable resource to design communication mechanisms that facilitate improvisational interaction between human and AI musicians. For example, a camera coupled with a computer vision system could give the AI the ability to identify visual cues from the human musician, such as eye contact or body language, and a visualisation of the AI presented on a screen next to the camera could provide visual feedback back to the human musician. Another possible interface might use a microphone to pick up on verbal cues from the human musician, with bi-direction audio communication implemented utilising a speaker to vocalise the AI musician's intentions back to the human musician. Additionally, understanding that communication between human musicians is a dynamic process that develops over time, and not a discrete set of rules that can be learnt in one sitting, provides a valuable context for the design of human-AI interactions. The development of a mental model that assists the AI in learning about the human's musical style could be implemented to emulate this principle. This mental model would be specific to each human musician interacting with the AI, and would be iteratively improved through sustained improvisational interactions.

Understanding the system

The theme **Understanding the system** explores the strategies participants used to attempt to understand the inner workings of the system and highlights the types of information that should be exposed to improve the participants' understanding of the system. Analysis of this theme uncovered that participants approached their interaction with the system with an expectation of its prior musical knowledge (code *Expected musical knowledge*). Participants expected that the system would have prior knowledge of musical theory. In fact, the system has been designed to have no prior knowledge of music theory, except what it learns from the participant during the improvisation session. This information was not explained to participants before the experiment as one of the study's objectives was to observe the mechanisms participants used to attempt to understand the system, which could have been biased if the participants were aware that the system had no prior music knowledge.

Within this theme, the code *Experimenting* explores participants' attempts to learn how the system works through experimentation. As the only way in which participants can interact with the system is by playing notes on the keyboard. The only feedback from the system to the musician is the system's musical output, this experimentation took the form of playing a series or pattern of notes and listening to how those notes influenced the musical output of the system. Through the process of experimentation, participants demonstrate their own form of the mental model via exploration of the system's responses to their given inputs. In this situation, the human musician is using experimentation to build and iteratively improve their own mental model of the AI system. The final code in this theme, *Understanding patterns*, explores participants' desire to have more information about the musical patterns the system is following. Visually representing aspects of the system's model and depicting on a screen the patterns of notes that the system has just performed, could provide additional context assisting the human musician to understand the system better.

Structured Improvisation

The theme **Structured Improvisation** explores the mechanisms participants identified that could provide more musical structure and help increase the quality of improvisational interactions with the system. Analysis of this theme identified three key structural concepts that are commonly used in an improvisational setting; *Pre-performance baseline*, *Leading and accompanying* and *Stages during performance*.

The code *Pre-performance baseline* refers to the practice of musicians discussing and agreeing upon a common set of rules before beginning an improvisation. Participants' experiences regarding the types of conversations and agreements that occur with another human musician before they begin musical improvisation were unpacked to understand how these concepts could be implemented to improve collaborative improvisation with an AI system. In jazz improvisation, for example, the *Pre-performance baseline* can take the form of a lead sheet which describes the basic structure of the melody and chord progressions to be used.

The code *Leading and accompanying* refers to commonly used roles that musicians adopt in improvisation, and which they move between during any improvisation episode with other musicians. For example, the musician in the lead role sets the mood, tempo and key of the music, freely improvising within the set of rules agreed upon before the improvisation. The musician in the accompanying role follows the lead role, providing a harmonic element to support the musical expression of the lead role. The code *Stages during performance* explores comments in which participants expressed a desire for the improvisation to follow a defined structure of musical stages, and to be able to communicate with the system as they progress through these pre-defined stages.

Explicit Communication

The themes **Explicit Communication** and **Implicit Communication** both refer to moments when the participants identified a need for more communication with the system. Understanding the communication mechanisms participants want to use when engaging with the system highlighted new communication pathways that could be implemented to improve musical collaboration with the system.

Explicit Communication refers to communicating specific intentions to the system through a direct channel, in order to produce a desired response from the system. Within this theme, four key intentions that participants wished to be able to communicate to the system were identified and explored in the codes below.

The code *Signal good interaction* refers to participants expressing a need to provide positive reinforcement to the system; while the code *signal unwanted interaction*, indicates stages in the improvisation when participants wanted to provide negative feedback to the system regarding an unwanted interaction. Giving the human musician the ability to positively or negatively reinforce particular musical decisions that the system produces would provide an additional feedback mechanism that the system could use to refine the model of musical patterns it has learned from the human musician.

The code *signal change to the system* refers to the participant's desire for a specific communication channel with the system to let it know an upcoming shift in the pace and/or direction of the music. The code *signal end of performance* was mentioned by every participant and indicates a strong need for participants to be able to communicate to the system that the improvisation is about to come to an end. The code *system signalling to user* explores the concept of bi-directional communication by giving the system autonomy to send signals to the human musician.

Implicit Communication

The theme *Implicit Communication* refers to communication with the system through an indirect mechanism, one in which desired outcome is not mentioned explicitly. Two categories of indirect methods of communication were highlighted during participant interviews: *Musical cue* and *Physical cue*. Although these codes resonate with the aforementioned codes in *Human Interaction* (*Musical interaction* and *Physical interaction*), the codes in this theme specifically relate to interaction with the system, and not another human musician.

The code *Musical cue* refers to participants communicating ideas with the system through the music that they play. As mentioned in the analysis of the theme *Human Interaction*, this style of communication through music is paramount when improvising with other human musicians and can convey a wide variety of musical ideas and suggestions. As the musical dialogue develops, it is only what the participants play interfaces with our AI system. In our experiments, participants attempted various methods of communicating with the system musically and expressed their desire for the system to respond to their musical suggestions.

The code *Physical cue* refers to communication with the system through physical interaction such as movements of the head, eyes, and body, or via voice commands. Understanding the ways in which physical cues are used to communicate ideas during improvisations can help to shape the possible methods of implementing indirect physical interaction with the system.

Discussion

> "We need a feeling for what is usual, if our attention is to be caught by the unusual" Ennals (2007)

The first aim of this paper was to understand the communication needs in human-machine improvisation for musicians with different levels of experience. Drawing on our group of participants, and on the scores from the Gold-MSI assessment, we found an inverse relation between the two individual factors of musical training and perceptual ability. Specifically, we found that for participants with higher musical training as well as for participants with lower perceptual abilities it is important to understand how the system works in order to **enhance predictability of the system's state in the performance**; while participants with lower musical training as well as participants with high perceptual abilities made more emphasis in the need for implicit and explicit communication mechanisms in order to provide alignment during the interaction and in this way **facilitate a dialog between the human and the machine**.

On the other hand, participants whose main instrument was voice, scored higher in the general factor of music sophistication. For these participants, understanding how the system works was also the most important feature when interacting with the system. While younger participants scored lower in music sophistication. These participants emphasized elements of human interaction that they considered necessary to sustained a good interaction; in particular, **building a connection through past experiences**, was at the top of their preferences.

Understanding how the system works was the main need among participants; therefore, **facilitating the discovery of the inner workings of the system** should be a design consideration for improvisation music systems. In general, understanding the system's capabilities from the outset would help users approach musical interactions with the system by setting up their expectations about how the interaction can develop. Attempts with this aim have been carried out by some systems but is not an standard practice in the field. Some examples are the Shimon marimba robot that moves its head to show excitement Hoffman and Weinberg (2010), or the use of emoticons to show the level of confidence McCormack *et al.* (2019). However, using expressive mediums to convey information of the inner working of music systems for improvisation is an area that still requires further work.

Additionally, a prominent observation, drawn from the weighting of codes 'Musical interaction' and 'Musical cue' in Table 2, is participants' expectation for the understanding of intentions to be achieved through the music. While this style of communication was often represented by the participants through the use of sound, e.g. "*if the AI detects that you're just playing the top register, maybe it can avoid playing notes in the area that you're playing*" or "*when someone's drumming, you know, they might start drumming a little harder, a little harder, you know, it's building up in it*", there was an element of the interaction that participants referenced throughout the different themes: the need for **space**. This is evidenced in the world cloud shown in Figure 3, which depicts the frequency of words used by participants in the semi-structured interviews – words such as the names of participants, the name of the transcription engine, etc. were added to the set of stopped words to be eliminated from the word cloud. In addition, we merged words that were used with a similar meaning by the participants; for instance, the words space, pause, rest and silence were merged into one word, 'space'; similarly improvisation and performance; person, musician, human and people, etc.; as well as words that were the same but that the transcription engine have added special characters to, such as apostrophes; e.g. idea and idea'. After this process we eliminated words whose frequency was less than 10.

The concept of space was very important among the participants, which they highlighted in several occasions when talking about how it worked in terms of structuring the interaction during a performance. We outline next how the concept of space is related to the concept of silence, and how it has been pointed in the literature to be an important medium of communication.

Figure 3. Word frequency cloud from participants' interviews.

The role of silence in human-machine communication

Pauses and rests, as well as moments of passive playing (which provide space for other voices/instruments to take the lead) are intrinsic constructs of music, not only as mediums to structure musical pieces, but also as mediums for *communication* between musicians: "a pause or a rest in the course of music has high information content because it becomes the "figure" against the "ground" of the constant musical expression" Beeman (2006). This use (and need) of space in musical interactions also came up repeatedly throughout the themes in our study, and with most of the participants, particularly highlighting moments when musical interactions with the system were either not working well: "*now there's a void where a dominant instrument was that's gone.*", or when participants were trying to work out how the system works: "*I guess I need just more space, room, more time to sort of listen to what the AI is offering me.*".

Scholars from different areas of the social sciences have categorised these constructs (pauses, rests and spaces in music) as types of *silence* Beeman (2006); Clair (2020), which carry out different meanings as they are framed based upon their placement within the context surrounding them, e.g. the musicians behaviours (movements, gestures, etc.), and the sounds that precede and follow them. According to this work, silence can take on different forms, from literal to figurative Clair (1998), have different valences Acheson (2008), and provide structural organisation, such as boundaries and turn taking Beeman (2006).

Despite the communicative nature of silence, this concept has been widely ignored in the study of machine learning systems. As put in Ennals (2007) ”we are denied an appreciation of silence”. However, researches in AI have highlighted the importance of the study into the meaning of silence for AI systems as it “can play a role in sending meaning between machine and human, specially in interactive communications” Kafaee *et al.* (2022). Similarly, here we propose the use of silence as *a mean of communication* in human-machine musical interactions. We describe this in more detail in the following sections.

Using silence to signal good/bad interactions

A valence value of silence in human-machine interactions could be determined as a function of the interplay between the human musician, and the sound surrounding the silence. As highlighted in Acheson (2008), silence is spatial and temporal, but it is also embodied in people’s behavior. When looking at the performances of some of the participants, we identified some embodied behaviours around silence. Participants would remove their hands from the keyboard and pause for a short period of time when they were not happy with the interaction, while short pauses with their hands remaining on the keyboard seem to signal the participants figuring out the way of joining what the system was doing. In a similar fashion, long pauses appear to signal participants giving space for the system to take the lead (we will expand on this in the next section).

Thus, analysing musician’s behavior, along with the temporal feature of silence, could be a path to assigning valence to moments of silence in a human-machine music interaction, and to signal in this way good and/or bad interactions.

Using silence to signal turn taking

Here the aim is to use moments of silence to facilitating a dialog between the human and machine as expressed by the *literal and figurative forms* of silence Clair (2020); Beeman (2006). To illustrate, Beeman (2006) highlights how Bach, in the partitas for violin and cello, does not indicate rests or pauses with literal silence, but with the concept of "relative volume" as an interpretative mechanism, by for instance de-emphasising some notes to "become the silence that separates each phrase". In our study, we could see two different behaviours during the performance of the participants with respect to taking turns when leading or following. In some occasions, this would be 'signaled' with long pauses (most times accompany with the removal of hands from the keyboard), while in other occasions, this was 'signaled' with the repetition of a note in the low registry (usually using only one hand while the other hand was removed from the keyboard).

We acknowledge that the concept of turn taking is not new in human-machine musical interactions. The SpireMuse system Thelle and Pasquier (2021), for instance, simulates different modes of turn taking, which are signaled by the human musician through the use of foot pedals during music interactions, or automatically selected by the system (as learned by the underlying model). However, we propose here the use of silence as a natural mechanism to communicate the need to change into a leading or following position.

Embodying silence

Alternative physical communication mechanisms were also identified as necessary fall backs when musical communication breaks down. As we consider the current limitations of state-of-the-art AI technologies, we argue that these alternative mechanisms are not only required in order to fully exploit the interaction with improvisational music systems, but are also part of how musicians communicate with each other in human-human settings.

Physical cues' are not only used to guide the interaction, but participants in the study also highlighted the process of improvisation to be *"a personal, intimate process, [in which] you are vulnerable and you are exposing your work"*, stating that it would be important for them if the system have some kind of 'presence' in order to establish a sense of closeness, and to enable both the human and the AI to better understand each other. For instance, one of the participants mentioned that "if there was some way that you could just see it there, banging away. You can sort of pick out little things that it's doing in there that make sense musically".

Ultimately, in any performative musical context there will be visual and (less frequently) vocal gestures which allow musicians understand the direction of the performance. Thus, the role of communication here is also to enhance predictability of the system's state in the performance. Additionally, communication through physical cues offers a lot of possibilities for the system and users to gather information that can enhance the mental models of each other.

This nuanced style of communication is an intrinsic element of co-creation in an improvisational setting, and it has been implemented in some music systems already, e.g. the Shimon marumba robot, which uses a human-like head that turns to show that it is 'listening' to the human performer. Here we propose that all music improvisations systems should employ some type of 'presence' that embodies the system in order to communicate aspects of the inner workings of the system, but also we propose that the interplay of moments of silence with some kind of embodiment, or for that matter, any kind of 'presence' that is implemented for an improvisational system (e.g. a digital visualisation), would play a significant role in conveying emotional meaning. We expand on this next.

The interplay between Embodiment and Silence to convey emotion

Embodying silence in a music system can give it a more powerful effect in communication, invoking reactions from musicians and audiences alike in the performance. In Acheson (2008), for instance, the author describes the case of a woman in the Warramunga tribe of Australia that didn't speak for 24 years after losing her husband. The cultural norms in the tribe require female relatives to remain silence for 2 years after a man's death, but the woman decided to remain silent for 22 years more. Acheson highlights that it is not only the duration of the silence that is significant, but as a bodily experience, imagining the woman walking through the village and people reacting in different ways (avoiding her, being sympathetic, etc.) made her statement even more powerful (either if this was a form of rejection to the oppression imposed to women in the tribe, or because she was honoring and mourning the lost of her loved one).

We propose here that, with a similar aim, providing a 'presence' to music systems (be this one an embody one or a digital one), can serve to enhance the meaning of silence, to improve the communication with human musicians, as well as to enhance the other roles of silence (to signal good/bad interactions as well as to signal turn taking).

Conclusions and Future Work

In this paper, we have detailed a study that explores the ways in which communication can support human-AI co-creation in a music improvisation setting. We recruited a group of musicians with varying degrees of experience in playing and improvising music and conducted an intensive experimental process. Participants were asked to improvise a piece of music with an AI system, which they later reviewed in an open interview and annotated on the resulting musical timeline. Within the annotations, participants were asked to describe their experience communicating musical intentions with their AI partner during the improvisation. Follow-up interviews with participants were coded through thematic analysis to understand the common themes among participants' experiences. Further analysis of the themes highlighted commonalities and differences, in terms of communication needs, among participants with different levels of music experience, this analysis culminated in the identification of *silence* as an important medium of communication within human-machine improvisation.

The concept of silence for communicative purposes has been overlooked in the study of machine learning systems, including in the domain of music. However, it has been extensively studied in the social sciences as a medium that carries meaningful and distinct information (depending on the context surrounding it). We further propose that the interplay between silence and the presence of the human and music system would provide rich ways for human-machine communication during improvisation.

Ultimately, this study has deepened our understanding of designing more musically nuanced human-machine communication during human/AI music improvisation performances. In future work, we will explore methods to extend AI music improvisation systems with the kinds of new communication opportunities we have explored in this paper.

References

Kris Acheson. Silence as gesture: Rethinking the nature of communicative silences. *Communication Theory*, 18(4):535–555, 2008.

Authors anonymised for review process. Human-ai music improvisation dataset, 2021.

William O Beeman. Silence in music. *Silence—the currency of power*, pages 23–34, 2006.

Ron Begleiter, Ran El-Yaniv, and Golan Yona. On prediction using variable order markov models. *Journal of Artificial Intelligence Research*, 22:385–421, 2004.

John A. Biles. Genjam: A genetic algorithm for generating jazz solos. In *International Conference on Mathematics and Computing*, 1994.

Laura Bishop and Werner Goebl. Beating time: How ensemble musicians' cueing gestures communicate beat position and tempo. *Psychology of Music*, 46(1):84–106, 2018.

Laura Bishop, Carlos Cancino-Chacón, and Werner Goebl. Moving to Communicate, Moving to Interact: Patterns of Body Motion in Musical Duo Performance. *Music Perception*, 37(1):1–25, 2019.

Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

Andrew R. Brown. Creative improvisation with a reflexive musical bot. *Digital Creativity*, 29(1):5–18, 2018.

Elaine Chew. Music interaction with others as conveyor of relational intent: A response to cross (2014). *Psychology of Music*, 42(6):826–838, 2014.

R.P. Clair. *Organizing Silence: A World of Possibilities*. G - Reference,Information and Interdisciplinary Subjects Series. State University of New York Press, 1998.

Robin Patric Clair. Silence. *The Handbook of Listening*, pages 427–438, 2020.

Richard Ennals. The meaning of silence. *AI & SOCIETY*, 21(4):625–632, 2007.

Benjamin Givan. Rethinking interaction in jazz improvisation. *Journal of the Society for Music Theory*, 22, 2016.

GMSI Documentation. https://media.gold-msi.org/test_materials/GMS/docs/gms_documentation_en.pdf. Accessed: 2022-12-29.

GMSI Scoring App. https://shiny.gold-msi.org/gmsiscorer/. Accessed: 2022-12-29.

Guy Hoffman and Gil Weinberg. Shimon: an interactive improvisational robotic marimba player. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 3097–3102. 2010.

Projucer. https://juce.com/. Accessed: 2022-12-28.

Mahdi Kafaee, Aliakbar Kouchakzadeh, and Shahriar Gharibzadeh. Silence: an ignored concept in artificial intelligence. *AI & SOCIETY*, pages 1–2, 2022.

Maria Teresa Llano, Mark d'Inverno, Matthew Yee-King, Jon McCormack, Alon Ilsar, Alison Pease, and Simon Colton. Explainable Computational Creativity. In *Proceedings of the Eleventh International Conference on Computational Creativity, 2020*, pages 334–341. Association for Computational Creativity (ACC), 2020.

Jon McCormack, Toby Gifford, Patrick Hutchings, Maria Teresa Llano Rodriguez, Matthew Yee-King, and Mark d'Inverno. In a silent way: Communication between AI and improvising musicians beyond sound. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2019.

Daniel Müllensiefen, Bruno Gingras, Jason Jiri Musil, and Lauren Stewart. The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, 9, 2014.

Music Circle. https://demo.museifi.com/. Accessed: 2022-12-29.

François Pachet, Pierre Roy, Julian Moreira, and Mark d'Inverno. Reflexive loopers for solo musical improvisation. In Wendy E. Mackay, Stephen A. Brewster, and Susanne Bødker, editors, *2013 ACM SIGCHI Conference on Human Factors in Computing Systems, CHI '13, Paris, France, April 27 - May 2, 2013*, pages 2205–2208. ACM, 2013.

Wojciech Samek and Klaus-Robert Müller. Towards explainable artificial intelligence. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, volume 11700 of

Lecture Notes in Computer Science, pages 5–22. Springer International Publishing, Cham, 2019.

R. Keith Sawyer. Improvisation and the creative process: Dewey, collingwood, and the aesthetics of spontaneity. *The Journal of Aesthetics and Art Criticism*, 58(2):149–161, 2000.

R Keith Sawyer. *Group creativity: Music, theatre, collaboration*. Erlbaum, Mahwah, NJ, 2003.

F. A. Seddon. Modes of communication during jazz improvisation. *British Journal of Music Education*, 22, 2005.

Kıvanç Tatar and Philippe Pasquier. Musical agents: A typology and state of the art towards musical metacreation. *Journal of New Music Research*, 48(1):56–105, 2019.

Notto J. W. Thelle and Philippe Pasquier. Spire Muse: A Virtual Musical Partner for Creative Brainstorming. In *NIME 2021*, apr 29 2021. https://nime.pubpub.org/pub/wcj8sjee.

Ashley E. Walton, Michael J. Richardson, Peter Langland-Hassan, and Anthony Chemero. Improvisation and the self-organization of multiple musical bodies. *Frontiers in Psychology*, 6, 2015.

Matthew Yee-King and Mark d'Inverno. Experience driven design of creative systems. In *Proceedings of the Seventh International Conference on Computational Creativity, ICCC'16*, pages 85–92. Sony CSL Paris, France, 2016.