Esteso: Interactive AI Music Duet Based on Player-Idiosyncratic Extended Double Bass Techniques

Domenico Stefani
Dept. of Information Engineering and Computer Science
University of Trento
Trento, Italy
domenico.stefani@unitn.it

Matteo Tomasetti
Dept. of Information Engineering and Computer Science
University of Trento
Trento, Italy
matteo.tomasetti@unitn.it

Filippo Angeloni
Composer, Musician
Castelvecchio, Italy
filippo.angeloni89@gmail.com

Luca Turchet
Dept. of Information Engineering and Computer Science
University of Trento
Trento, Italy
luca.turchet@unitn.it

ABSTRACT

Extended playing techniques are a crucial characteristic of contemporary double bass practice. Players find their voice by developing a personal vocabulary of techniques through practice and experimentation. These player-idiosyncratic techniques are used in composition, performance, and improvisation. Today's AI methods offer the opportunity to recognize such techniques and repurpose them in real-time, leading to new forms of interactions between musicians and machines. This paper is the result of a collaboration between a composer/double-bass player and researchers, born from the musician's desire for an interactive improvisational experience with AI centered around the practice of his extended techniques. With this aim, we developed *Esteso*: an interactive improvisational system based on extended technique recognition, live electronics, and a timbre-transfer double-bass model. We evaluated our system with the musician with three duet improvisational sessions, each using different mapping strategies between the techniques and the sound of the virtual double bass counterpart. We collected qualitative data from the musician to gather insights about the three configurations and the corresponding improvisational duets, as well as investigate the resulting interactions. We provide a discussion about the outcomes of our analysis and draw more general design considerations.

Author Keywords

Extended Playing Techniques, Double Bass, Music AI, Interactive Music Systems, NIME

NIME'24, 4–6 September, Utrecht, The Netherlands.

Figure 1: Evaluation session of *Esteso*.

1. INTRODUCTION

Contemporary double bass practice is rooted in the use of extended playing techniques [2, 31, 6]. Extended techniques refer to the experimentation of new instrumental techniques to foster diversification in the timbral palette of instruments and compositions [33]. Double-bass compositions and practice have been a thriving place for extended techniques, from the early uses of *col legno* and *pizzicato* to more advanced techniques, which can be attributed to the genesis of the free improvisation movement in the 1960s. This led to many extended techniques being developed as recently as the 80s and 90s [2].

This is the musical background of the third author: a professional double-bass player and composer who has been playing the instrument for over 17 years and exploring the use of extended techniques in his practice. With this sentiment, we (the remaining authors) were approached by the third author, who asked about the possibility of integrating Artificial Intelligence (AI) into his practice, with particular attention to his use of extended techniques. This fostered a conversation about technical possibilities and the improvisational practice of the musician. As a result, we acknowledged how some of the techniques in the player's vocabulary (or how they are played) are born from his experimentation and are hard to find in the literature. Moreover, when reading the literature on interactive musical systems, we found only few that took into consideration playing techniques, let alone extended techniques. Lastly, we found a lack of interactive systems targeting extended techniques peculiar to the double-bass.

In this paper, we describe the development, evaluation, and outcomes of *Esteso*[1]: an interactive improvisational duet system for double-bass based on the player-idiosyncratic extended techniques that define the musician's playing style. *Esteso* is a system based on machine listening where the musician and an AI counterpart take turns in an improvisational duet. The system's response is produced through manipulation of the real double-bass improvisation. This is achieved through live electronics and a timbre-transfer neural network trained on double-bass recordings. Machine listening is integrated in the form of a real-time classifier of extended techniques, whose output controls the sound manipulation process to affect differently the various techniques. The extended techniques to train the classifier were chosen from the player's musical vocabulary. The system was evaluated through several free duet/improvisational sessions, and the musician's comments were collected and analyzed.

The remainder of this paper is organized as follows. In Section 2, the double-bass player presents the underlying artistic motivation. Section 3 describes works related to interactive improvisational systems and extended double-bass techniques, while Section 4 illustrates the technical implementation of *Esteso*. Section 5 presents the methodology of the user experiments and a formal analysis of the results, along with our findings. Subsequently, we discuss the results in Section 6. Finally, we draw our conclusions in Section 7.

2. ARTISTIC MOTIVATION

As a composer and double bass player, I experience research as an unavoidable constant in my actions. Absolutely convinced that there is always something new to be discovered (on all fronts, including acoustic), I have approached new expressive techniques on my instrument. I explored both the established scenarios of acoustic and electronic music since I think that the individual components of acoustic sound, as much as those of electronic sound, are increasingly converging in one direction. Beyond my research, my studies with composers such as Marco Momi, Matteo Franceschini, Aurelio Samori, and Gianluca Verlingeri have further reinforced this idea of convergence. I therefore consider the advancement of new extended techniques indispensable. On this front, the double-bass is equaled by few instruments and has given proof of high technical-timbral possibilities, made possible in part by its physical size and extended register.

Starting from the existing literature, I found interest in investigating open-string techniques, and the types of "thin" noises and timbres that can be extracted from the double bass when treated with subtlety. The physical size of the double bass in contrast to a performer's hand is one of the focuses of this research. The vast surface area of the sound box, as well as of the scroll, bridge, and tailpiece offers vast possibilities for timbral research. The open-string technique chosen for this collaboration is "Sfregato col legno', which has a very different nature from many open-string techniques and could be considered as a daughter of the historically established "Col legno battuto" (Clb) technique.

Secondly, I opted for a technique of a different nature: Jeté is one of the most widely used bow strokes in the centuries-old tradition of string instruments. Here I proposed a brushed version of Jeté, executed by taking advantage of the tension of the horsehair and bow angle. "Brushed Jeté" takes full advantage of the shape and structure of the double bass by producing definite timbres depending on the point touched.

Finally, I chose percussive techniques as they are an excellent means of physical and timbral exploration of the instrument's body. Depending on the hit location and materials, as well as the part of the hands used to strike, they offer their unique timbre.

Table 1 describes the extended double bass techniques identified for this study. Audio samples for these can be found in the project's repository[2].

Table 1: Extended double-bass techniques selected from the musician's personal vocabulary.

Name	Description
"Brushed" Jeté	Jeté (or Ricochet technique [19]) with brushed characteristics, introduced with careful use of bow actions and angle. See Figure 2.
Sfregato con legno	Open strings are struck and the wood part of the bow is slid along the string from the bridge, barely making contact with it. It introduces a peculiar buzzing sound. To the best of the authors' knowledge, it originated from the musician's exploration. See Figure 3.
Percussive	Collection of different percussive hits [31, 32] on the instrument's body. The hand parts used are fingertips, palms, knuckles, and closed fists. The parts hit are the wooden body, plastic tailpiece, and fingerboard. See Figure 4.

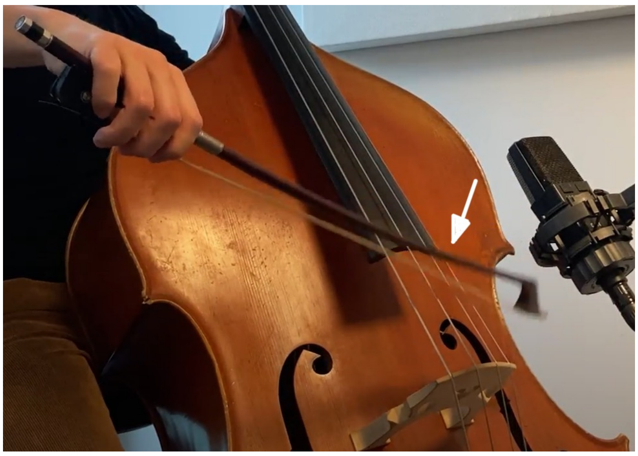

Figure 2: The musician executing brushed Jeté. The white arrow indicates the direction of the bow movement.

[1]Demo video: `https://youtu.be/HEhJXAgFiXM`

[2]`https://github.com/CIMIL/Esteso`

Figure 3: The musician executing *Sfregato con legno*. The white arrow indicates the direction of the bow movement. Note the angle of rotation with which the bow is held in the player's hand, which makes the "legno" (wooden part of the bow) come into contact with the string.

Figure 4: The musician executing the percussive technique.

3. BACKGROUND

Improvising with computers

The proposed system is affine to the "player" paradigm defined by Rowe [45], which defines an "artificial" player able to interact with human players. The study of computer and musician interactions involving machine listening was pioneered by Lewis with his early works [42] and ongoing composition/system *Voyager* [30], where musicians collaborate with non-human improvising agents. McCormack *et al.* [35] explored the possibilities and interaction methods in musical improvisation, drawing design principles for designing collaborative systems.

Furthermore, Jordà presented a survey of his works on music improvisations with computers [25, 23, 24]. Other examples of interactive music improvisational systems are approaches that use multiple input modalities such as Lepri's *InMuSIC* [29], Ciufo's environment for sonic improvisation [7], Erdem *et al.* *CAVI* [12], or McCormack *et al.* *In a Silent Way* [36].

Recent studies on musician-computer interaction through machine listening can be found in the extensive work of Gioti [13, 14, 15, 18, 17]. Symbiosis [13] is relevant to our study as it focused on the interaction of a double-bass player and interactive system that reacted to the player in an action/reaction cycle [13]. Gioti's system replied to the musician with their own recorded sound, which was heavily modified through non-linear processes, achieving a non-deterministic response. Our work differs in that we focused on the artist's request to have extended techniques as the basis of the interaction. Moreover, a core part of the processing in our approach is the use of a timbre-transfer neural network that was trained on double-bass sounds, resulting in a "hybrid instrument"[3] with double-bass features along with novel sounds that are alien to acoustic instruments.

Gioti [15] also incorporated a playing technique classifier into an interactive composition. However, the author designed this system for the interaction with a saxophone, and the classification of rather broad sound categories (i.e., single notes, multiphonics, air tones, and slap tones), rather than extended techniques from the personal vocabulary of the musician.

Additional Musician-Computer interactive systems outside the scope of this study include robotic players [16, 21, 49].

Extended Double Bass Techniques

In-depth historical context about the double-bass, especially as a solo instrument and free improvisations, can be found in Botting's thesis [2] The interest in extended double-bass techniques can be attributed to studies and performances by influential players such as Fernando Grillo, Stefano Scodanibbio, and Barry Guy [2, 31, 6]. These studies show how crucial the nature and development of a personal technique vocabulary is for players [2].

Additional resources on double-bass contemporary and extended techniques include Meyers' dissertation [37] and Hartley's book "Double Bass Solo Techniques" [20].

4. SYSTEM DESIGN

Esteso was implemented as a Max/MSP patch [41], and its structure can be broken down into three parts: an extended technique recognition section, the *duet-mechanism*, and a sound manipulation stage. A diagram of the system architecture is found in Figure 5.

4.1 Extended Technique Recognition

At the core of *Esteso* is a playing technique recognition system composed of a feature extraction stage, an onset detection stage, and a classifier. The purpose of this section is to capture the use of extended techniques, which are then used to influence the system's response.

In place of common signal representation or timbral features (e.g., Spectrograms, Mel-frequency cepstrum), we employed the encoder part of a *RAVE* model [4] that we trained on double-bass recordings from the OrchideaSOL database [5]. RAVE is a variational autoencoder for neural audio synthesis tasks such as timbre-transfer. Autoencoders are neural networks that compress complex inputs down to a small latent space (through an *encoder*), which is then expanded back to the original size (through a *decoder*). Autoencoders are trained to reconstruct input samples, but in trained models, the encoder and decoder can be used independently. The size of the latent space of the model trained

[3] For lack of a better term, *hybrid* in the sense of part acoustic and part digital/electronic.

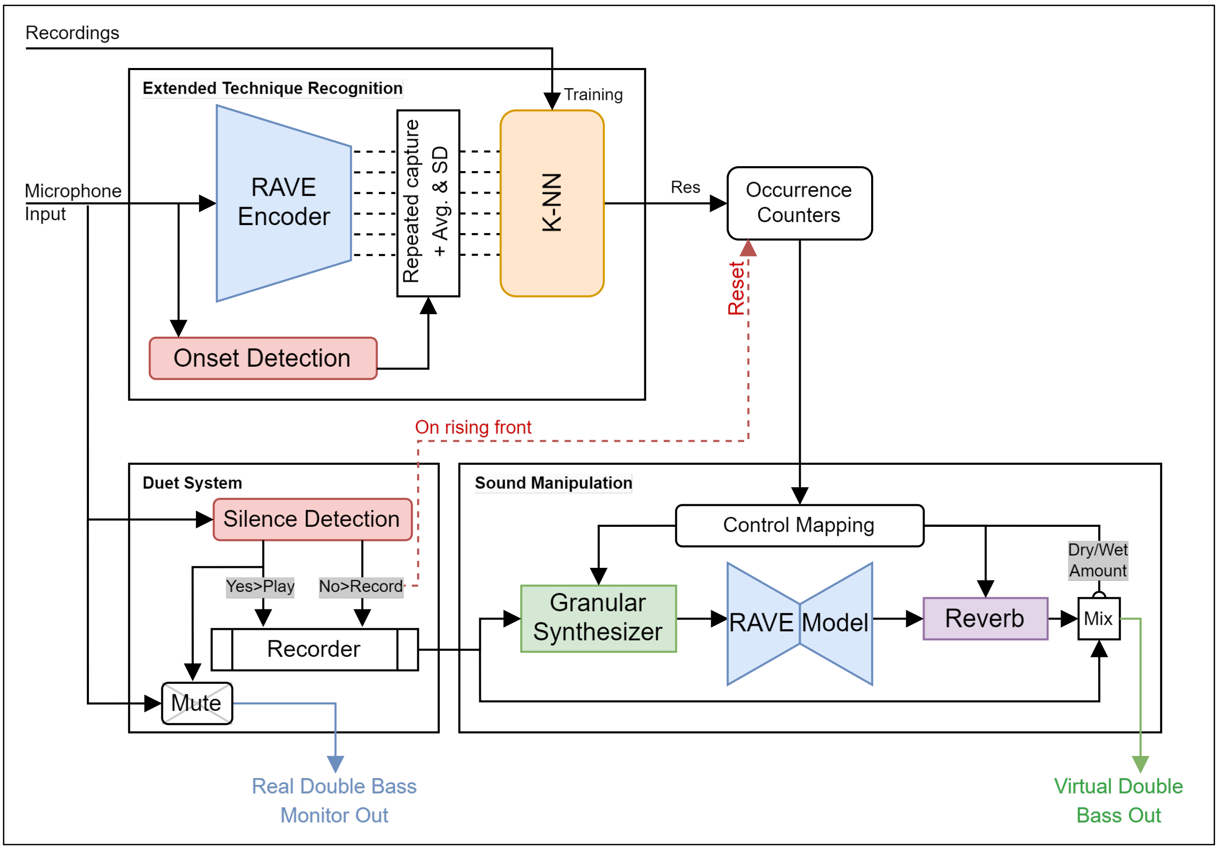

Figure 5: System architecture of *Esteso*. The system is divided into three parts: (1) extended technique recognition, (2) the duet system, and (3) a sound manipulation stage. The first part (top left) is devoted to recognizing and counting occurrences of extended techniques from the audio coming from the musician. The duet system (bottom left) manages the action-reaction mechanism by recording the player's audio and playing back the recorded buffer when detecting silence for a set period (1.5 seconds). Finally, the sound manipulation stage processes the audio recording coming from the duet system, producing the response to be played through the speakers.

Figure 6: Presentation mode view of the Max/MSP patch for *Esteso*.

was 16. The reasoning behind the use of RAVE is that it is trained to compress musical sound down to few parameters, which will eventually encode salient features regarding pitch and timbre. Consequently, the lower number of features paired with our hypothesis of high-quality description can make for a low-noise input for simple machine learning classifiers such as K-nearest neighbors (KNN) or Support Vector Machiness (SVMs).

Here, the latent representation is sampled 10 times at 10-millisecond intervals upon onset detection, capturing a section of the beginning of sounds in the signals. Subsequently, the mean and standard deviation of the captures for each variable are fed to a KNN classifier. For onset detection, we employed the `peakamp` Max/MSP object.

We employed the KNN classifier from the FluCoMa toolkit [46], which was trained with recordings of the three extended techniques performed by the musician, plus a fourth class for "other" techniques. We used the default K value of `fluid.knnclassifier` (3 neighbors). We tested recognition with new recordings for each technique, resulting in a rather low occurrence of incorrect classifications. A more formal cross-validated analysis and parameter search with the system will be the scope of future work.

Finally, the predictions from the classifier are fed to a system to count the occurrences of each technique. The majority technique is read, and the counters are reset according to the duet mechanism.

4.2 Duet Mechanism

The action-response nature of the interactive duet was managed by a simple rule: the system starts listening (therefore recording) whenever sound coming from the microphone is detected; audio is recorded to a 10-second buffer; when audio from the microphone stops for 1.5 seconds, the system plays the recorded buffer, which is fed to the sound manipulation stage to produce the response. Buffer length and silence length threshold were found through experimentation.

Whenever the recording is started, the occurrence counters for the technique recognition system are reset. Additionally, whenever the response is triggered, the system computes the technique that occurred the most from the counters and feeds it to the sound manipulation system.

4.3 Sound Manipulation

In *Esteso*, the response to the musician's phrases is obtained by manipulating the brief recordings provided by the duet mechanism. For this, we employ a granular synthesizer (GS), a full RAVE timbre-transfer model (i.e., both encoder and decoder), and a reverb effect.

The majority technique obtained from the classifier and counters is used to affect the parameters of both the GS and reverb. This was mapped through a sound design process (by the first and second authors), where recordings of the musician were used to find effect parameters that would highlight each technique. The result was four parameter presets for each effect, to be controlled by the majority technique. The mappings can be found in the Esteso patch in the project's repository[4].

For the GS, we used the Max/MSP package "Petra" [5]. Petra's primary application is polyphonic granulation, drawing inspiration from the concepts of asynchronous granular synthesis [44]. Petra incorporates an audio object for real-time input granulation, which was used for *Esteso*. This feature incorporates a circular buffer with adaptable and optionally randomized control for delaying the entire buffer duration. This enables the control of several parameters of the granular synthesis, which are the initial position of the circular buffer, the length and pitch of the grains, the pan position for each grain, and the gain for individual grains. For the mapping of different techniques to parameters of the GS, its parameter space was explored through trial-and-error.

We used granular synthesis because it is a form of synthesis that can dismantle temporal relationships within the double-bass recordings. This is to be attributed to the core concepts of granular synthesis [43]. With this, we introduced heavy-handed manipulation of the sound to change it at the temporal level.

Furthermore, we employed a RAVE timbre-transfer model trained on double-bass recordings from the Orchidea-SOL database [5]. While we intended to only use the encoder for classification, we were struck by the peculiar sound obtained by the model, which only partially resembled a real double-bass. We felt the model's sound to be interesting and akin to self-sabotaged instruments [10], sharing some similarities with De Souza's class of redesigned-instruments [11]. We may refer to our use of this "virtual" double-bass as an intentional use of an unintentionally and arbitrarily sabotaged instrument. The process with which the virtual double-bass comes to be can also be an example of *ergomimesis*[6] [34] driven by data and probabilistic processes in the neural training. Audio samples can be found in the project repository[4].

We also used three low-frequency oscillators (LFOs) to influence different values of the latent representation of RAVE, further shifting the hybrid nature of the virtual double-bass away from acoustic sounds.

For the reverb, we used 'Reverb-2' from the BEAP modules by Matthew Davidson [7]. We used the reverb to give a sense of physical space to *Esteso* to differ from the dry double-bass sound.

We worked at parametrizing the reverb similarly to the GS. We acted on the size, decay, diffusion, and mix parameters of Reverb-2. For instance, we chose less reverberate sound (with size, decay, and diffusion set to low values) for percussive hits and wider spaces for other expressive techniques.

Given the complex manipulations, we (the first, second, and fourth authors) decided to have the musician (third author) test three different configurations of *Esteso* to collect comparative comments and better understand the perceived qualities of the sound. These configurations were the following:

1. *Granular+RAVE+Reverb*: we used the GS, followed by the RAVE timbre-transfer model, and reverb. The LFOs were not used, so the RAVE model was in its best configuration to emulate the double-bass sound, although the GS was highly morphing the input sounds.

2. *RAVE+LFOs+Reverb*: we removed the GS and summed three LFOs to arbitrarily chosen parameters of RAVE's latent space. The reverb was used with its mappings. This was meant to be a less extreme manipulation in terms of temporal content but more marked in terms of other sound characteristics due to the modifications to the latent space.

[4] `https://github.com/CIMIL/Esteso`

[5] `https://github.com/CircuitMusicLabs/petra`

[6] The action of miming patterns in the process of transduction where characteristics from established instruments are translated in new digital instruments

[7] `https://github.com/stretta/BEAP`

3. *Granular+RAVE+Reverb with 50% mix*: the third configuration closely matched the first, with the use of the GS, timbre-transfer model, and reverb, and no arbitrary manipulation of the latent space. Additionally, the manipulated sound was mixed with the double-bass sound from the record buffer for a less extreme sound. Each of the two components contributed to 50% of the mix.

5. EVALUATION

5.1 Methodology

Esteso was evaluated by the third author within three sessions of 15 minutes each. Each session involved one of the sound configurations above (Section 4.3). Each session consisted of having the musician improvise with the system (Figure 1). The improvisation took place as a duet, where the musician played a short phrase followed by silence and then listened to a *response/reaction* from the system. The mutual exclusivity of the duet was suggested but not enforced, and, at times, the musician intentionally overlapped their sound with the system's.

Similar to autobiographical approaches [39] within and outside the NIME community [9, 48, 47], *Esteso* was evaluated with a single musician that was closely involved with the project. However, here the musician and subject of the experiments (third author) had a part in expressing his interests and artistic needs; and was decoupled from the designers of the system and experiment (remaining authors).

Collaborations between researchers and a single musician are found often in literature [1]. Musician-researcher collaborations often occur to cater to the specific needs of a musician, and adopt technology that is available to the researcher for that purpose [1]. Besides autobiographical studies, the practice extends to studies where the musician is responsible for the artistic input or evaluation and, in some occasions, participatory design [22]. Single musician-researcher collaborations are also common for accessible instruments [27, 28, 8].

For our study, the musician was kept unaware of the system's inner workings. The musician was only aware of his motivations and techniques, previously presented to us in a conversation (Section 2). This was done to evaluate unbiased interactions between the musician and *Esteso*. Besides thoughts of general relevance, we intended to collect comparative comments between the sessions and, therefore, the different configurations.

The experiment started with a brief pilot phase to get the musician partially acquainted with *Esteso* and finalize the system with slight corrections [26]. Subsequently, we had the musician interact with the system for three sessions with adequate pauses in between.

To gather comments about each session, we considered methodologies with little interaction from the experimenters, such as think-aloud. Similar to other studies from the NIME community [9] and given the incompatibility of think-aloud alone with music tasks [26], we employed **think aloud** concept with **post-task walkthrough** [40]. To achieve this, we video-recorded each session, allowing the musician to review each session immediately afterward. When reviewing, the musician was asked to verbally express any thought that came to mind. Audio recordings from the think-aloud sessions were captured and transcribed.

After the procedure was repeated for all three sessions, we used thematic analysis [3], which encompassed the generation and categorization of codes into three subgroups - Concept of time, dynamics and timbre, and comparative comments - each representing themes reflecting discernible patterns.

For the experimental setup, we used a laptop where the *Esteso* Max/MSP patch was running, to which we connected an RME Fireface UFX soundcard to capture the sound of the musician's double-bass through an AKG C414 microphone. The soundcard was also connected to two studio monitors to play the system's response. The musician's double-bass sound was acoustically diffused. Each session was video-recorded with a smartphone for playback during the think-aloud phases.

5.2 Findings

5.2.1 Concept of time

Different concept of time: The musician felt that the system did not have a concept of the duration of the performance. To further explain, the musician specified that he knew the performance had a set duration and, therefore, that at a certain point, he would have had to execute a climax, pauses, or fade-outs. On the contrary, he felt that the system did not have this concept of the musical form (*"I know that the performance lasts 15 minutes and therefore I have my [concept of] time when executing climaxes and closing acts [...], while it interprets time differently because probably [...] it has no desire for interpretation"* and *"It was absurd because I was preparing to close the improvisation, and instead, towards the end, it seemed to reactivate. It was surprising because, at that moment, I kept playing without feeling the reaction of the AI, and then, at a certain point, it answered me"*).

Active silence as if the system was listening: The musician reported his perception of silences in the system's response as *active silences*. He defined these as intentional pauses as opposed to a lack of reaction because of the system's negligence in listening to him. He expressed the feeling that the system opted not to react because it understood that the musician was finishing his phrases (*"I had the feeling that it was listening more, so I felt the AI silences as more active. I felt it was listening more, and therefore it didn't always have to react necessarily. I considered that pause an active silence because it was as if it was listening to me and giving me my space where I could perform. If I must say, I felt it was more intelligent, that's it"*).

Independent behavior: The musician commented that, at times, he would have liked the system to respond in a more independent/personal way. For instance, the musician would have liked the system to react to a small gesture with a longer, "denser" musical idea or vice versa (*"I would have liked the system to be even more personal in its processing, in the sense that I would have liked it to respond to a small gesture of mine with a longer-lasting or denser processing and vice versa, to create different musical situations"*).

5.2.2 Dynamics and timbre

No reaction to *Pianissimo*: The musician felt that the system did not react when he was playing with *Pianissimo* dynamics. At times this appealed to the musician as musically useful (*"When I articulated musical situations around a dynamic close to Piano, and the system does not react to it, allows me to build an additional musical situation, to have greater space and greater freedom in interacting with it"*). Furthermore, on some occasions, the musician appreciated a *Fortissimo* reaction because he could keep playing dynamically opposite ideas (e.g., with a *Pianissimo* dynamic) to create musical superpositions. Contrarily, at times this reaction felt limiting (*"It would undoubtedly be very nice if it *could, at times, also play with dynamics in the Pianissimo range [...] it plays in a range that goes from mezzo-forte to fortissimo and, at times, I would like it to be able to move and respond even to dynamics close to piano"*).

Evident musical intention and timbre: The musician reported noticing a clear musical intention and timbre from the system, which influenced him throughout the performance (*"It suggests the musical atmosphere to me. The AI was very clear both in terms of musical intention and timbre and as a result, it always clearly expressed something to me"*).

Changing musicality: At times, the system's response felt more musical than in previous instances. The musician reported not knowing whether this feeling was affected by him getting used to the system, but he felt that the sound processing changed over time (*"It is as if its way of processing changes from time to time. For example, I suddenly received this sound material that was more grainy than the previous in terms of timbre. It is as if it turned from a rock to many small stones, metaphorically speaking"*).

5.2.3 Comparative comments

Session 1: Aggressive behavior, but stable: The musician told us that the first session had the most *aggressive* and "*anarchic*" response. Initially, he felt a great musicality in the pitches of the responses. Yet, later, this characteristic was lost to greater timbral variations (*"Even though I suggested [to the system] a certain variety of techniques [...] it always responded in an aggressive and tense manner. It was anarchic in response and attitude. In the beginning, however, I felt a certain musicality in it, which was then lost because its timbre had too much variety in materials"*). However, the musician reported feeling a stable and not very unpredictable temporal response (*"Despite its anarchic sound nature, it was always stable in responding and communicating with me...it didn't have an unpredictable nature; I could communicate with it in a very stable way"*).

Session 2: Softer and cleaner behavior, but unpredictable: The musician reported feeling a cleaner and softer response in terms of timbre in the second session. However, the response was very unpredictable from a temporal perspective. He reported being more involved and intrigued by this, prompting him to experiment more with the techniques (*"Sometimes it didn't answer me, sometimes it listened to me, sometimes it answered me in its way. It had this degree of unpredictability, which was an asset for me as a player because it suggested and involved me a lot. It also made me explore the techniques more because I wanted to somehow try to connect more with it. Despite this, the response was always clean and soft regarding timbre"*).

Session 3: More control, less participation from the system: In the third session, the musician had more musical control precisely because he felt much less participation from the system, as he noticed that the system responded by replicating the last part he had played as it was. This made him explore less the techniques and become less involved (*"On the one hand, I had the ability to build something; on the other hand, I felt less its identity, partly because I felt that its personality was 50%, while the other 50% was a replication of what I did. There was, therefore, a component that partly reflected me; there was a mirror. As a result, my exploration has been much lower and limited. In my opinion, this is conditioned by the fact that the AI material was, let's say, the same substance as what I was doing in part, and therefore, this [...] made me stay where I was rather than explore or go looking for other clues or techniques to understand how to interact"*).

6. DISCUSSION

From the musician's comments, we noted his focus on the long-term concept of time and the structure of improvisations. This was in contrast with his perception of the system's abilities, where he felt a lack of consideration of musical form. Additionally, this matches the occasional lack, perceived by the musician, of independent behavior from the system. We believe this is to be attributed to the duet-mimicking behavior. This can serve as a guideline for designers of interactive music systems, prompting them to impart a concept of musical form to the behavior of these systems or foster its natural emergence in AI processes.

On the upside, some of the peculiarities of the temporal behavior of *Esteso* were praised as interesting and human-like, such as the use of silences, which was repeatedly perceived as active listening. Moreover, the short-term temporal behavior of the first two sessions was appreciated.

In terms of dynamics, the musician reported contrasting feelings about both a lack of response to his *Pianissimo* sounds and responses with widely different dynamics than his. From time to time, pauses were felt as a way to give the musician performing space, while the dynamic range used by the system felt limiting from time to time. At the same time, the responses from *Esteso* that had a completely different dynamic than the musician's short-term improvisations were praised as forms of independent behavior. Similarly to the temporal evolution of a musical form, the variety and perceived intentionality of dynamics emerge as a crucial aspect of the perception of the quality of an improvisation, which can inform further developments.

In terms of timbre, the musician appreciated a certain timbral identity in *Esteso*, often referring to it with language usually used to refer to human players and their actions or intentionality. On the other hand, this sometimes led the musician to find a contrast between the figure of a coherent virtual player and his desire for an extremely varied and inspiring "AI". This could highlight two potentially opposite ways an interactive system could be developed and presented: one as an actual virtual player made to closely mimic human behavior, and the other as a completely virtual and non-human-like improvisation machine.

In terms of comparative comments, the first configuration felt aggressive and somewhat anarchic (in terms of timbre). However, certain stability in musical communication was felt, partly due to the temporal nature of short-term responses. In contrast, the musician found the second configuration softer and cleaner in terms of timbre but more unpredictable temporally and in communication. Unpredictability felt inspiring, allowing for improvisational exploration. This second configuration was the most appreciated. The third configuration was found sterile in terms of improvisation and discovery. Interestingly, the musician expressed the perception of a 50%/50% contribution of his sound and the system's sound in the responses without being aware that it was precisely the mix setting used. The musician also stated that he felt much more in control than the others, at times finding the interaction similar to that with a "loop machine", which, while positive in some contexts, was negative in an improvisational context. This suggests the detrimental effect of a high level of control in a system for interactive improvisations.

Contrary to the sound-design intentions, the second configuration with RAVE and a reverb was the most appreciated. Interestingly, it was less influenced by extended techniques than other configurations, and it included arbitrary manipulations that gave it a less predictable nature. At the same time, although the system was based on extended techniques, and the musician used them heavily during the experiences, the musician's thoughts focused on the improvisational and musical nature of the system's responses. This can suggest how, contrary to what we thought, microscopic properties of the phrases, such as the use of extended techniques, may be less important than the macroscopic context of an improvisational music form in interactive music systems. This could, however, be attributed to the strong effect of the GS and requires further investigation.

Interestingly, however, in both the first two sessions, the musician did not notice how each system's response was the product of manipulation of his preceding phrase. Part of this can be attributed to the timbre-transfer model, which, especially in session 2, revealed to produce far enough sounds from the original to be perceived as new. The use of extended techniques by the musician is a primary cause of this behavior, as the model can only learn from data and the available double-bass datasets hardly contain the wide variety of possible extended techniques, let alone very personal ones. As a result, in the presence of unknown sounds and techniques, the model produced interesting incoherent tones that defined *Esteso*. Therefore, ultimately the use of personal extended techniques shaped the response of the system.

Finally, the sometimes contradictory nature of the musician's comments and desires reminded us of the interaction between two improvising human musicians who are not in complete synchrony or do not share the same state of mind.

7. CONCLUSIONS

In this paper, we presented *Esteso*, an interactive improvisational system for double-bass based on the personal extended techniques of the third author. We evaluated different configurations of the system, gathering insights that can inspire other designs of interactive music systems and further improve *Esteso* with feedback from the player. Ultimately, the experience with *Esteso* was appreciated by the musician, who nevertheless, expressed extremely useful thoughts regarding the limitations of this first stage of its design. His perspective highlighted the importance of more investigation in extended technique integration, incorporating concepts of musical form, and encouraging independent behavior within improvisational music systems. In addition, the discussion surrounding temporal dynamics, timbre, and the balance of control underscored the multifaceted nature of the needs of an improvisation musician. All these concepts will benefit from further investigation with more players. In addition, the first two system configurations complemented each other in many regards and are currently being combined and tested with the double-bass player.

8. ETHICAL STANDARDS

This study followed all ethical and data protection guidelines from the University of Trento. This paper complies with the NIME Conference standard [38]. The code is publicly available and no external users were involved, as only the authors participated in the experiments. We declare no conflict of interest.

9. REFERENCES

[1] K. Andersen and D. Gibson. The Instrument as the Source of New in New Music. *Design Issues*, 33(3):37–55, 07 2017.

[2] T. Botting. *Developing a Personal Vocabulary for Solo Double Bass Through Assimilation of Extended Techniques and Preparation.* PhD thesis, Sydney Conservatorium of Music, The University of Sydney, 2019.

[3] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.

[4] A. Caillon and P. Esling. RAVE: A variational autoencoder for fast and high-quality neural audio synthesis. *CoRR*, abs/2111.05011, 2021.

[5] C.-E. Cella, D. Ghisi, V. Lostanlen, F. Lévy, J. Fineberg, and Y. Maresz. Orchideasol: a dataset of extended instrumental techniques for computer-aided orchestration. In *Proceedings of the International Computer Music Conference*, pages 420–429. International Computer Music Association, 2020.

[6] C. P. Chesanek. *Invention through the Harmonics of Stefano Scodanibbio: A Method of Creative Improvisation for the Contemporary Double Bassist.* PhD thesis, The University of Nebraska-Lincoln, 2020.

[7] T. Ciufo. Beginner's Mind: an Environment for Sonic Improvisation. In *Proceedings of the International Computer Music Conference*, 2005.

[8] A. G. D. Correa, I. K. Ficheman, M. do Nascimento, and R. de Deus Lopes. Computer assisted music therapy: A case study of an augmented reality musical system for children with cerebral palsy rehabilitation. In *2009 Ninth IEEE International Conference on Advanced Learning Technologies*, pages 218–220, 2009.

[9] F. A. Dal Rì and R. Masu. Exploring Musical Form: Digital Scores to Support Live Coding Practice. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, jun 9 2022. https://nime.pubpub.org/pub/ex3udgld.

[10] T. Dannemann, N. Bryan-Kinns, and A. McPherson. Self-sabotage workshop: a starting point to unravel sabotaging of instruments as a design practice. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Mexico City, Mexico, May 2023.

[11] J. De Souza. *Music at hand: Instruments, bodies, and cognition.* Oxford University Press, 2017.

[12] C. Erdem, B. Wallace, and A. Refsum Jensenius. CAVI: A coadaptive audiovisual instrument–composition. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, The University of Auckland, New Zealand, jun 2022.

[13] A.-M. Gioti. From live to interactive electronics. symbiosis: a study on sonic human-computer synergy. In *Proceedings of the International Computer Music Conference*, pages 572–578, 2016.

[14] A.-M. Gioti. Machine listening in interactive music systems: Current state and future directions. In *Proceedings of the International Computer Music Conference*. Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 2017.

[15] A.-M. Gioti. Neurons: An interactive composition using a neural network for recognition of playing techniques. In *Proceedings of the 2018 Musical Metacreation Workshop*, 2018.

[16] A.-M. Gioti. Imitation game: Real-time decision-making in an inter-active composition for human and robotic percussionist. In *Proceedings of the International Computer Music Conference*, 2019.

[17] A.-M. Gioti. A compositional exploration of computational aesthetic evaluation and ai bias. In *Proceedings of the International Conference on New*

Interfaces for Musical Expression, Shanghai, China, June 2021.

[18] A.-M. Gioti. Converge/diverge: Collaborative emergence in a composition for piano, double bass and interactive music systemt. In *Proceedings of the International Computer Music Conference*, 2021.

[19] K. Guettler. A guide to advanced modern double bass technique. *(No Title)*, 1992.

[20] K. Hartley. *Double bass solo techniques: a book of orchestral excerpts*. Oxford University Press, 2008.

[21] G. Hoffman and G. Weinberg. Shimon: an interactive improvisational robotic marimba player. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, page 3097–3102, New York, NY, USA, 2010. Association for Computing Machinery.

[22] T. Hunter, P. Worthy, B. Matthews, and S. Viller. Using participatory design in the development of a new musical interface: Understanding musician's needs beyond usability. In *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*, pages 268–271, 2019.

[23] S. Jordá. A real-time midi composer and interactive improviser by means of feedback systems. In *Proceedings of the International Computer Music Conference*, pages 463–463. International Computer Music Association, 1991.

[24] S. Jordà. Afasia: the ultimate homeric one-man-multimedia-band. In *Proceedings of the 2002 conference on New interfaces for musical expression*, pages 1–6, 2002.

[25] S. Jordà. Improvising with computers: A personal survey (1989–2001). *Journal of New Music Research*, 31(1):1–10, 2002.

[26] C. Kiefer, N. Collins, and G. Fitzpatrick. HCI Methodology For Evaluating Musical Controllers : A Case Study. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 87–90, 2008.

[27] J. V. Larsen, H. Knoche, and D. Overholt. A longitudinal field trial with a hemiplegic guitarist using the actuated guitar. In *New Interfaces for Musical Expression 2018*. NIME, 2018.

[28] J. V. Larsen, D. Overholt, and T. B. Moeslund. The prospects of musical instruments for people with physical disabilities. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 327–331, Brisbane, Australia, 2016. Queensland Conservatorium Griffith University.

[29] G. Lepri. InMuSIC: an Interactive Multimodal System for Electroacoustic Improvisation. In *Proceedings of the International Computer Music Conference*, 2016.

[30] G. E. Lewis. Too Many Notes: Computers, Complexity and Culture in Voyager. *Leonardo Music Journal*, 10:33–39, 12 2000.

[31] A. J. Long. *The creative application of extended techniques for double bass in improvisation and composition-Volume 1*. PhD thesis, Cardiff University, 2020.

[32] A. J. Long. (website) Percussive Techniques, Hands (non pitch specific). `https://www.themoderndoublebass.org.uk/percussive-hands-non-specific.html`, 2020. Accessed: 2024-02-02.

[33] V. Lostanlen, J. Andén, and M. Lagrange. Extended playing techniques: the next milestone in musical instrument recognition. In *Proceedings of the 5th International Conference on Digital Libraries for Musicology*, DLfM '18, page 1–10, New York, NY, USA, 2018. Association for Computing Machinery.

[34] T. Magnusson. *Sonic writing: technologies of material, symbolic, and signal inscriptions*. Bloomsbury Publishing USA, 2019.

[35] J. McCormack and M. d'Inverno. Designing improvisational interfaces. In F. Pachet, A. Cardoso, V. Corruble, and F. Ghedini, editors, *Proceedings of the Seventh International Conference on Computational Creativity*, pages 98–105. Sony CSL, July 2016.

[36] J. McCormack, T. Gifford, P. Hutchings, M. T. Llano Rodriguez, M. Yee-King, and M. 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.

[37] M. Meyer. *Contemporary Double Bass Techniques: An Advanced Technical Approach*. PhD thesis, University of North Texas, 2018.

[38] F. Morreale, N. Gold, C. Chevalier, and R. Masu. NIME Principles & Code of Practice on Ethical Research, Jan. 2023.

[39] C. Neustaedter and P. Sengers. Autobiographical design in hci research: designing and learning through use-it-yourself. In *Proceedings of the Designing Interactive Systems Conference*, pages 514–523, 2012.

[40] P. G. Polson, C. Lewis, J. Rieman, and C. Wharton. Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. *International Journal of Man-Machine Studies*, 36(5):741–773, 1992.

[41] M. Puckette. The Patcher. In *Proceedings of the International Computer Music Conference*, pages 420–429. International Computer Music Association, 1988.

[42] C. Roads. *Composers and the Computer*. William Kaufmann, 1985.

[43] C. Roads. Introduction to granular synthesis. *Computer Music Journal*, 12(2):11–13, 1988.

[44] C. Roads. *Microsound*. The MIT Press, 2004.

[45] R. Rowe. *Interactive music systems: machine listening and composing*. MIT press, 1992.

[46] P. A. Tremblay, G. Roma, and O. Green. Enabling Programmatic Data Mining as Musicking: The Fluid Corpus Manipulation Toolkit. *Computer Music Journal*, 45(2):9–23, 06 2021.

[47] L. Turchet. Smart mandolin: autobiographical design, implementation, use cases, and lessons learned. In *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion*, AM '18, New York, NY, USA, 2018. Association for Computing Machinery.

[48] Y. Wang and C. Martin. Cubing Sound: Designing a NIME for Head-mounted Augmented Reality. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, jun 16 2022. https://nime.pubpub.org/pub/w82of2do.

[49] G. Weinberg and S. Driscoll. The design of a robotic marimba player – introducing pitch into robotic musicianship. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 228–233, New York City, NY, United States, 2007.