# Demo of Esteso: an AI Music Duet Based on Extended Double Bass Techniques

**Domenico Stefani[1], Matteo Tomasetti[1], Filippo Angeloni[2], Luca Turchet[1]**

[1] *Department of Information Engineering and Computer Science, University Of Trento, Italy*
[2] *Independent - Composer, Musician, Italy*

## Abstract

*Esteso* is an interactive improvisational system for double-bass based on player-idiosyncratic extended techniques. This system was created in collaboration with the contemporary double-bass player and composer Filippo Angeloni and tailored for his personal vocabulary of extended techniques. In *Esteso*, AI agent and the player engage in a duet, taking turns in the performance. The system replies with a manipulation of the real double-bass, achieved live through a timbre-transfer neural network, granular synthesis, and reverb. The timbre-transfer network was trained on a public double-bass dataset, resulting in a peculiar *hybrid* sound. Machine listening is integrated through a classifier of extended techniques played on the double-bass, whose output controls sound processing to affect various techniques differently. We present a demonstration of a performance where the double-bass player interacts with Esteso, creating a back-and-forth interplay between the acoustic and virtual elements.

## 1 Project Description

Esteso was presented at the Interational conference on New Interfaces for Musical Expression, Utrecht, The Netherlands ([Stefani et al., 2024]). This demo will however allow attendants to see the inner workings of the system in detail.

The system demoed here is related to the "player" paradigm outlined by Rowe [Rowe, 1992], which defines an "artificial" player able to interact with human players.

Videos of the system during a test session and a live performance can be found online[1,2]. The software can be found on Github[3]. This demo fits the theme of this year's DIMMI as *contamination* can be seen in the multifaceted nature of the sonic output of the system, which steams from the musician's signal, his way of playing, and the sound that the timbre-transfer model learned from generic data. In addition, this represents a multidisciplinary collaboration between a musician and researchers where different interests merge into one output.

**Figure 1:** *Esteso during a live performance.*
*Picture by Alberto Boem*

The system was implemented as a Cycling'74 Max patch. It is organized into three parts: (1) a technique recognition section, (2) a mechanism to manage the duet-like interplay, and (2) a sound processing stage. The purpose of the recognition section is to detect the use of different extended techniques from the musician and subsequently affect the system's response. Secondly, the duet-mechanism is responsible for the

[1] www.youtube.com/watch?v=HEhJXAgFiXM
[2] www.youtube.com/live/Vrywo3fpALw at 1:31:00s
[3] www.github.com/domenicostefani/Esteso

**Figure 2:** *Player interacting with Esteso*

action-reaction nature of the duet, enforcing simple rules that start and stop the system's response. Finally, the sound processing stage is responsible for the sonic nature of the system's response. Figure 3 depicts the architecture.

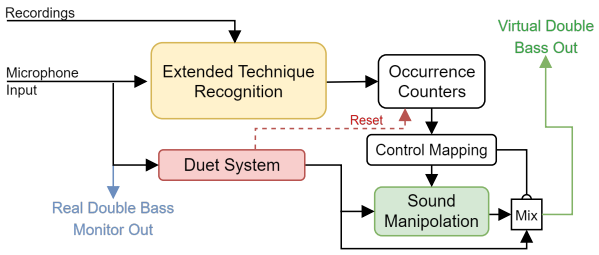


**Figure 3:** *Architecture of* Esteso

The technique recognition system is composed of a feature extractor, an onset detector, and a machine-learning classifier. The encoder of a RAVE model was used as a feature extractor, while a simple peakamp object served as an onset detector. Repeated samples from the latent space of RAVE are collected upon onset detection and fed to the classifier. The classifier used was a K-nearest neighbor method. We focused on detecting the three extended techniques (i.e., "*Brushed*" *Jeté*, *Sfregato con legno*, and *Percussive*) and used the classification result to select different parameters for the sound processing stage.

During the performance, the duet-mechanism enforces a basic rule: *Esteso* starts listening as soon as sound is detected by the microphone; audio is recorded to a short buffer (ten seconds); when the signal from the double-bass is quieter than a set silence threshold for one second, the system feeds the recorded buffer to the sound processing stage. The processed buffer constitutes the response of the agent to the human player. Buffer length and silence thresholds were found through experimentation with the double-bass player.

Finally, the response of the AI agent is obtained through the alteration of the short recordings produced by the duet mechanism. The processing pipeline is composed by a granular synthesizer[4], a timbre-transfer model (RAVE [Caillon & Esling, 2021]), and a reverb effect[5]. The granular synthesizer was chosen as it can morph the temporal structure of input recordings, providing the musician with novel responses. Secondly, the timbre transfer model, trained on generic double-bass recordings, generated unusual sounds that only somewhat resembled an actual double-bass. We chose the model as we felt its sound was interesting and conceptually made it similar to self-sabotaged instruments [Dannemann et al., 2023]. We then used reverb to grant *Esteso* the feel of different acoustic spaces, in contrast with the dry double-bass sound. Selected parameters of the sound processing pipeline are affected by the results of the extended technique recognition classifier. This was mapped through a sound design process, where dry double-bass recordings were fed through the effect pipeline to parameterize it in different ways that would highlight each technique. Therefore, sets of parameter values were mapped to each possible output of the classifier.

Apart from the performance proposed here, the system was formally evaluated through three performance sessions aimed at understanding the musician's reaction and testing different sound processing mappings. New Interfaces for Musical Expression (nime) practices and techniques were adopted in the design of the experiments, the formal evaluation, and the extraction of themes from the musician's comments.

## References

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

[Dannemann et al., 2023] Dannemann, T., Bryan-Kinns, N., & McPherson, A. (2023). 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.

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

[Stefani et al., 2024] Stefani, D., Tomasetti, M., Angeloni, F., & Turchet, L. (2024). Esteso: Interactive ai music duet based on player-idiosyncratic extended double bass techniques. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME'24)*.

[4]The Max/MSP package "Petra": https://github.com/CircuitMusicLabs/petra

[5]"Reverb-2" from the BEAP module package, by Matthew Davidson: https://github.com/stretta/BEAP