

# BACHDUET: A HUMAN-MACHINE DUET IMPROVISATION SYSTEM

**Christodoulos Benetatos**

**Zhiyao Duan**

c.benetatos@rochester.edu zhiyao.duan@rochester.edu

Department of Electrical and Computer Engineering, University of Rochester

## EXTENDED ABSTRACT

**Summary.** Back in the days of what today we refer to as *Baroque period*, improvisation was a key element of music performance and education. Great musicians such as J. S. Bach, were better known as improvisers than composers. Today, however, there is a lack of improvisation culture in classical music performance and education; classical musicians either do not know how to improvise, or cannot find other people to jam with. Motivated by this observation, we developed *BachDuet*, a system that enables real-time counterpoint improvisation between a human and a machine. This system uses a recurrent neural network to process the human musician’s monophonic performance on a MIDI keyboard and generates the machine’s monophonic performance in real time. We hope that it will serve as both an entertainment and practice tool for classical musicians to develop their improvisation skills.

**Proposed System.** We design this system to support two kinds of human-machine music improvisation: a) free counterpoint improvisation in a duet setting and b) conditional counterpoint improvisation in a duet setting over a given bass part. In both settings, the system predicts the next note of the machine’s voice based on the past notes of both the machine’s and the human’s voices, as well as the pre-defined bass part in the second setting (Figure 1). This causal way of music generation allows us to implement this system with a real-time GUI, where the user interacts with the system by playing on a MIDI keyboard. The system operates with a preset steady tempo to ensure human-machine synchronization. A demo video can be found here <sup>1</sup>.

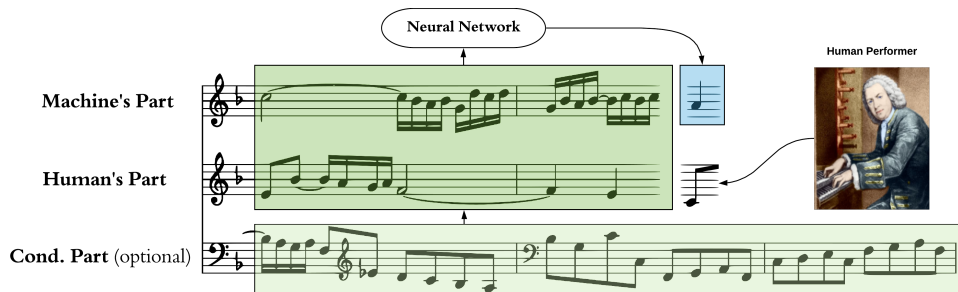


Figure 1. System overview. The system predicts the next note of the machine’s voice based on the past notes of both the machine’s and the human’s voices, and a pre-defined bass part, when available.

The core algorithm of *BachDuet* consists of an LSTM network augmented with an stack structure serving as extra memory. We divide the music time into sixteenth note grid and assume that a note always starts and ends on grid positions (time steps). At every grid position, the network predicts the note of the machine’s voice in the next time step as well as the musical key of the current time step. To encode each note, we use a combination of the MIDI number and the onset indicator. For example, a C4 quarter note spans four time steps and is encoded with four tokens: [60\_1, 60\_0, 60\_0, 60\_0], where the first token denotes the onset

<sup>1</sup><https://www.youtube.com/watch?v=eZb0wA7k0NU>.

