

**Department of Electrical and Computer Engineering**

**University of Rochester, Rochester, NY**

**Ph.D. Public Defense**

**Friday, December 15, 2017**

**1:00 PM**

**Computer Studies Building, Room 426**

## **Toward a Human-Centric Automatic Piano Music Transcription System**

Andrea Cogliati

Supervised by  
Professor Zhiyao Duan

Music Transcription is the process of notating a music piece solely by hearing. It is the highest skill acquired by musicians during the ear training process, as it involves the identification of pitches, intervals, melodies, chords, and rhythms, along with a mature understanding of music theory. Some of the skills involved in transcription, such as recognizing melodies and rhythms, are natural for most people, but extremely difficult for machines. The goal of my research is to design a computer system that automatically transcribes a piece of music into standard music notation, which is the most natural notation for musicians.

I call this system *human-centric* as its purpose is mainly to assist musicians in being more precise and effective and to speak to them in the language they are most familiar with, *i.e.*, common music notation. Amateur musicians might use the system to improve their skills, while professional musicians, such as composers, might appreciate the ability to quickly notate improvisation sessions on a piano.

The core of the proposed method is a high-accuracy music transcription system for piano music in a context-specific setting. That is, the system has to be trained on the same instrument and in the same acoustic context, *i.e.*, room and microphone position, that will be used for the transcription. This constraint allows to increase the accuracy of the transcription above the accuracy of state-of-the-art transcription systems, which cannot be reliably used for practical applications.

The system displays the transcription in music notation in an interactive editor, so that the transcription can be edited and corrected. Music notation output is not common for automatic music transcription systems, which are mostly focused on the *parametric transcription*, *i.e.*, determining the played pitches, their onsets and offsets. As a consequence, there is no way to objectively evaluate the accuracy of a transcription in music notation. For my research, I designed an algorithm to convert a parametric transcription into music notation. I also propose an objective evaluation metric that can predict human evaluation of a music notation transcription.

Finally, I outline a possible generalization of the transcription system, in order to relax the context-specific constraint and transcribe music played on different instruments.