

Cheng Zhi Anna Huang

ISE (CS) 575

Week 6

Samer Abdallah and Mark Plumbley (2004). "Polyphonic Music Transcription by non-negative sparse coding of power spectra." ISMIR

This paper presents a system that attempts to transcribe polyphonic piano music such as Bach's Fugue in G-minor. This is achieved through a method called adaptive spectral basis decomposition. It is adaptive because the system constructs a "learned dictionary" that holds "statistical summary of past experiences with power spectra, which enables the system to make better inferences about newly encountered spectra" during the transcribing process to determine pitch. It features spectral basis decomposition because when the system is trained on examples of polyphonic music, the spectrum is decomposed into a sum of note spectra, and the system then learns the spectral profile of each note. Spectral basis decomposition is achieved by using "the probabilistic model equivalent to a form of sparse coding with non-negativity constraints". This model forms a theoretical basis for spectral estimation and denoising. The sparse coding in this model assumes that only "a few" atomic features from the "dictionary" are needed to account for most observations from the spectra. The "dictionary" is a matrix, and the atomic features form the columns of the dictionary matrix.

Once pitches had been assigned to each of the pitched spectra in the dictionary, the transcription model can take action. The system first captures an input spectrogram from live acoustic recordings, and sparse codes it using the dictionary. Then, it reconstructs a 'schematized' version of the input where noise has been eliminated, and can be used as a basis for calculating the pitches.

Although the system is not able to provide a "complete" transcription, the "pitch transcription" given is very promising. The system can be adapted for cognition systems that are most interested in analyzing the interplay of pitch information in a polyphonic texture. Furthermore, the system gives a close estimate of the durational distance of onset pitches, though not yet a precise inter-onset interval (IOI).