Introduction

* Commercial applications and interests in DSP for music signal processing

What is AMT ?

The capability of transcribing music audio into music notation is a fascinating example of human intelligence. It involves perception (analyzing complex auditory scenes), cognition (recognizing musical objects), knowledge representation (forming musical structures), and inference (testing alternative hypotheses). Automatic music transcription (AMT), i.e., the design of computational algorithms to convert acoustic music signals into some form of music notation, is a challenging task in signal processing and artificial intelligence. It comprises several subtasks, including multipitch estimation (MPE), onset and offset detection, instrument recognition, beat and rhythm tracking, interpretation of expressive timing and dynamics and score typesetting.

*[E. Benetos, S. Dixon, Z. Duan, S. Ewert, “Automatic Music Transcription”, in IEEE SPS Journal Vol. 36 January]*

Applications

A successful AMT system would enable a broad range of interactions between people and music, including music education (e.g., through systems for automatic instrument tutoring), music creation (e.g., dictating improvised musical ideas and automatic music accompaniment), music production (e.g., music content visualization and intelligent content-based editing), music search (e.g., indexing and recommendation of music by melody, bass, rhythm, or chord progression), and musicology (e.g., analyzing jazz improvisations and other nonnotated music). As such, AMT is an enabling technology with clear potential for both economic and societal impact.

*[E. Benetos, S. Dixon, Z. Duan, S. Ewert, “Automatic Music Transcription”, in IEEE SPS Journal Vol. 36 January]*

* Introduce DSP concepts
  + Sampling theorem

The sampling theorem is a consequence of digitizing analogue signals. Sampling an analogue signal stores quantized values of the amplitude of a continuous signal at regular intervals determined by the sampling rate of the ADC system. The result of this conversion will be an array stored in memory with N total samples that represent the analogue signal.

The sampling theorem says that to avoid higher frequency components aliasing as lower frequencies components after sampling the following must be satisfied.

Fs > 2B

Where B is the highest frequency expected in the signal. Frequently a sampling rate of 44.1 kHz is used in audio recording because the range of human hearing is from 20-20kHz.

* + DFT
    - Frequency resolution
    - Zero padding
  + STFT

As in other audio-related applications, the most popular tool for describing the time-varying energy across different frequency bands is the short-time Fourier Transform (STFT), which, when visualized as its magnitude, is known as the spectrogram.

Formally, let x be a discrete-time signal obtained by uniform sampling a waveform at a sampling rate Fs Hz. Using a N-point tapered window w (eg. Hamming w[n] = 0.5-0.46cos(2\*pi\*n/N) for n element of [0: N-1]) and an overlap of half a window length we obtain the STFT

With t element of [0 : T-1] and k element of [0:k]. Here, T determines the number of frames , K = N/2 is the index of the last unique frequency value as dictated by the Sampling Theorem. Thus X(t,k) corresponds to the window beginning at time t\*N/(2\*Fs) in seconds and frequency

in Hz.

*[A. Klapuri and M. Davy, Eds., Signal Processing Methods for Music Transcription. New York: Springer-Verlag, 2006.]*

* + - Windowing

To improve the performance of the STFT several different windowing functions can be used to smooth discontinuities between each frame. In audio recordings this is extremely important to ensure the integrity of the signal is conserved.

* + - Frequency and time resolution trade off

The frequency resolution achieved by an STFT is directly influenced by the window length L and the sampling frequency Fs. The time resolution possible is the inverse of the frequency resolution.

Given that humans can discern changes in a signal as small as 10ms, the frequency resolution has imposed limits.

* + - Log frequency spectrograms
* Introduce AMT and pitch detection

Here, an important question is why a trained musician has no problem in analyzing a chord containing two notes one octave apart. Better understanding in which what two overlapping partials interact and how their amplitude can be precisely estimated is certainly of central importance.

*[C. Yeh and A. Roebel, ‘The expected amplitude of overlapping partials of harmonic sounds’ in Proc. Int. Conf. Acoust., Speech, Signal Process (ICASSP’09), Taipei, Taiwan, 2009, pp. 316-319]*

* State of the art approaches and challenges faced
  + NMF
  + NN
  + Traditional Signal Processing Methods

System Design

* Libraries used with analysis
* Main approaches
* System block diagram illustrating filters in even library functions such as librosa

Results

* Accuracy, precision and recall results of pitch detection
* STFT graphs showing predictions
* NMF
* NN
* Traditional Signal Processing Methods (CEPSTRUM, ACF) etc.

Future Work

* Areas to improve upon
* Problems unsolved
* Most promising avenues for further research