AMT RESEARCH

# IEEE Volume 36 January

# Automatic Music Transcription

## Overview

The nature of music signals, which often contain several sound sources that are highly correlated over both time and frequency, AMT is still considered an open problem in the literature. Usually and AMT system takes an audio waveform as input, computes a time-frequency representation and outputs pitches over time or ideally a typeset music score. Percussion and unpitched sounds will be outside the scope of this research. [1], [2]

## Applications

A successful AMT system would aid in musical education, music creation, music production, music searching and musicology. It is considered a fundamental problem in the field and is closely related to audio source separation [3] and music information retrieval [4] because knowing the contents of the piece can greatly assist with these tasks. It provides the main link between music signal processing and symbolic music processing.

AMT has close relations with speech processing as both tasks involve converting acoustic signals to symbolic sequences. [5] Both disciplines benefit from language modelling components that are combined with acoustic components. The methodologies used are also very similar in both fields. [6] One of the key differences of the two disciplines are that musical sources are highly correlated in time and in frequency.

Furthermore, AMT is related to image processing and computer vision as musical objects can be recognized in two-dimensional time-frequency representations.

## Key Challenges

1. Polyphonic music contains a mixture of simultaneous sources. Inferring musical attributes from the mixture signal is an underdetermined problem
2. Overlapping sound events exhibit harmonic relations with each other. For any consonant musical interval, the fundamental frequencies form small integer ratios, so that their harmonics overlap in frequency.

For example the fundamental frequency ratio of its three notes C:E:G is 4:5:6.

1. Statistical independence of sources does not apply in music source separation, due to the synchronization of onsets and offsets between different voices.
2. The annotation of ground truth transcriptions for polyphonic music is time consuming and requires high expertise. The lack of such annotations has limited powerful supervised learning techniques to specific contexts. There are some approaches to circumvent this problem [7] but they require professional music performers and thorough pre- and postprocessing. It is also noted sheet music is often considered a weak label for a number of reasons: they are not time-aligned to the audio signal, there are different versions and interpretations of musical pieces.

## Issues with current AMT systems

* Octave errors
* Semitone errors
* Missed notes
* Extra notes
* Merged/fragmented notes
* Incorrect offsets/onsets
* Misassigned streams

## Overview of AMT methods

Most approaches are designed to achieve an intermediate goal in AMT which does not actually resemble musical notation.

### MPE ( Multiple Pitch Estimation/ Frame Estimation)

The estimation of the number and pitch of notes that are present in each time frame ( ~10ms). This is usually performed independently in each frame although contextual information is considered in filtering estimations in a post processing stage. A number of approaches operate at this level including:

* Traditional Signal processing methods [11], [12]
* Probabilistic modeling [8]
* Bayesian approaches [13]
* NMF [14] –[17]
* Neural Networks [18],[19]

These types of methods do not form concepts of musical notes and rarely model any high-level musical structures.

## Commercial AMT software

* Melodyne - <http://www.celemony.com/en/melodyne>
* AudioScore- <http://www.sibelius.com/>
* ScoreCloud- <http://scorecloud.com/>
* AnthemScore - <https://www.lunaverus.com>
* Transcribe! - <https://www.seventhstring.com/xscribe/>