Reference: E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri, “Automatic music transcription: Challenges and future directions,” J. Intelligent Inform. Syst., vol. 41, no. 3, pp. 407–434, 2013.

Accuracies reported in recent years have reached a limit as can be seen in MiReX. Current methods use general purpose models which are unable to capture the rich diversity found in music signals.

**Ways forward**

* Tailor algorithms to specific use cases
* Semi-automatic approaches like score informed audio aligned

**AMT tasks**

1. MPE – finding multiple f0 in a time frame
2. Note onset/offset
3. Loudness estimation/quantisation
4. Instrument recognition
5. Extraction of rhythmic information
6. Time quantisation

**Methods**

**MPE**

They can either be joint/iterative approaches. Iterative approaches work by extracting the most prominent f0 in each iteration until no additional pitches can be found. Accumulates errors at each iteration step. Joint approaches evaluate f0 combinations at the expense of computational cost. Most approaches are joint estimations.

**Best approaches**

**Reference for evaluation of MiReX**

Bay, M., Ehmann, A.F., Downie, J.S. (2009). Evaluation of multiple-F0 estimation and tracking systems. In 10th int. society for music information retrieval conf. (pp. 315–320).

**FEATURE BASED**

A pitch candidate set score function or pitch salience function is used to estimate f0 from time-frequency representations.

**Best approach (2012)**

Dressler, K. (2012). Multiple fundamental frequency extraction for MIREX 2012. In Music information retrieval evaluation eXchange. http:www.music-ir.org/mirex/abstracts/2012/KD1.pdf.

Based on FFT multi resolution analysis. Each spectral bin is multiplied by the bins instantaneous frequency. Pitch estimation is made by identifying spectral peaks and performing pair-wise analysis. Rank peaks by harmonicity, smoothness, appearance of intermediate peaks and harmonic number. System also tracks tones over time using an adaptive magnitude and a harmonic magnitude threshold.

**STATISTICS BASED**

Viewed as a MAP (maximum a posteriori) estimation problem. Maximum likelihood models based on gaussians that represent the partials.

Synchronous evolution of Gaussian partials are modelled by Gaussian mixtures.