Automatic Transcription: An Enabling Technology for Music Analysis

Simon Dixon

simon.dixon@eecs.qmul.ac.uk

Centre for Digital Music
School of Electronic Engineering and Computer Science
Queen Mary University of London

Queen Mary University of London

centre for digital music

Research Councils UK Digital Economy Transforming Business and Society

JISC

Semantic Audio

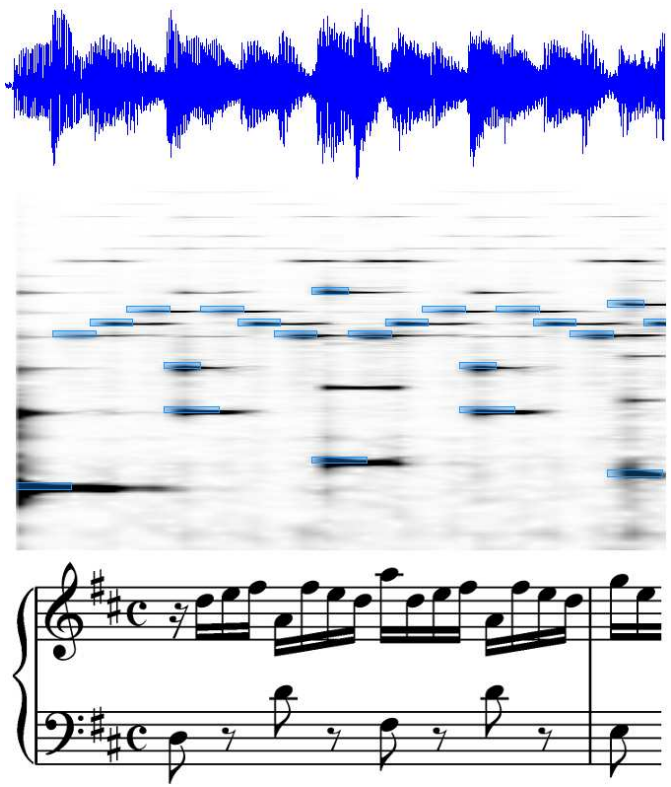
- Extraction of symbolic “meaning” from audio signals: How would a person describe the audio?
  - Solo violin, D minor, slow tempo, Baroque period
  - Bar 12 starts at 0:21.961
  - 2nd verse starts at 1:43.388
  - Quarter comma meantone temperament
- Annotation of audio
  - Onset detection
  - Beat tracking
  - Segmentation (structural, ...)
  - Instrument identification
  - Temperament and inharmonicity analysis
  - Automatic transcription
- Automatically generated annotations are stored as metadata
  - Matched against content-based queries
  - Also useful for navigation within a document

centre for digital music

Queen Mary University of London

Automatic music transcription (AMT)

- Audio recording → Score
- Applications:
  - Music information retrieval
  - Interactive music systems
  - Musicological analysis
- Subtasks:
  - Pitch estimation
  - Onset/offset detection
  - Instrument identification
  - Rhythmic parsing
  - Pitch spelling
  - Typesetting
- Still remains an open problem



centre for digital music

Queen Mary University of London

Background (1)

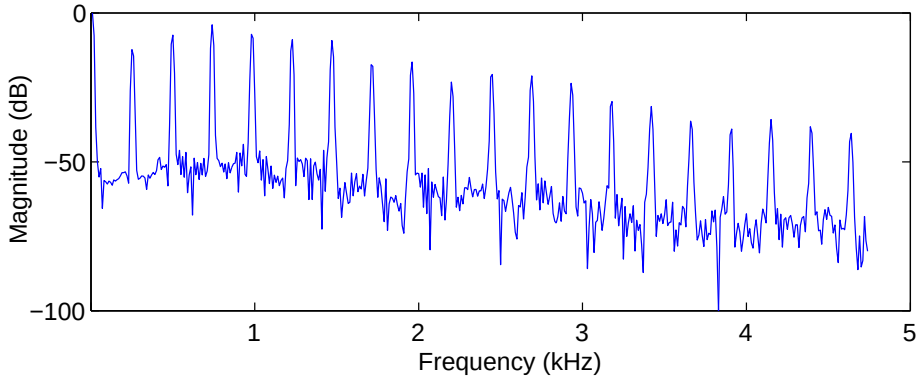
- Core problem: Multiple-F0 estimation
- Common time-frequency representations:
  - Short-Time Fourier Transform
  - Constant-Q Transform
  - ERB Gammatone Filterbank
- Common estimation techniques:
  - Signal processing methods
  - Statistical modelling methods
  - NMF/PLCA
  - Sparse coding
  - Machine learning algorithms
- HMMs commonly used for note tracking

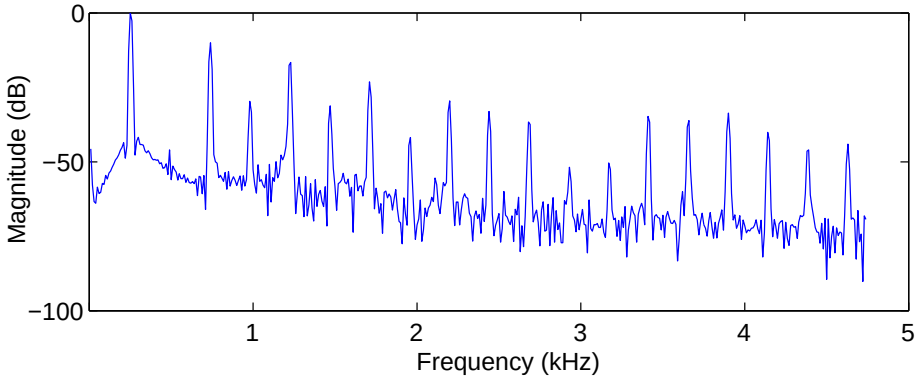
centre for digital music

Queen Mary University of London

Background (2)

- **mirex** evaluation:
  - Best results achieved by signal processing methods
- Multiple-F0 estimation approaches rely on assumptions:
  - Harmonicity
  - Power summation
  - Spectral smoothness
  - Constant spectral template



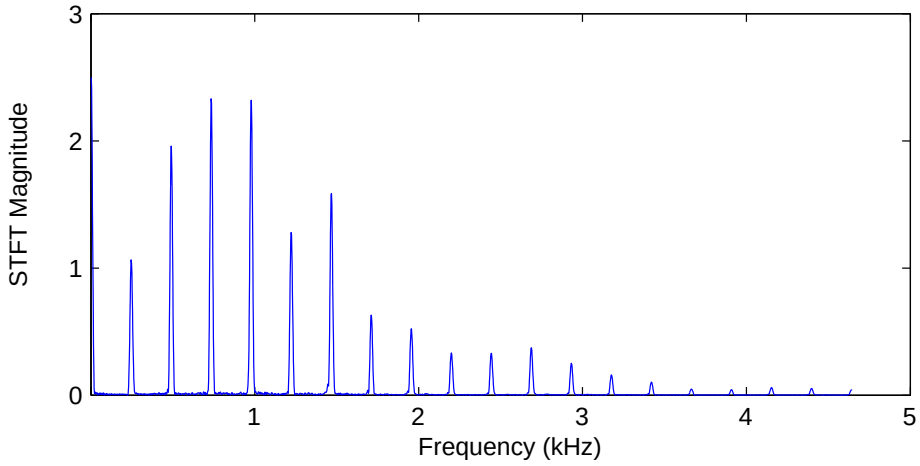


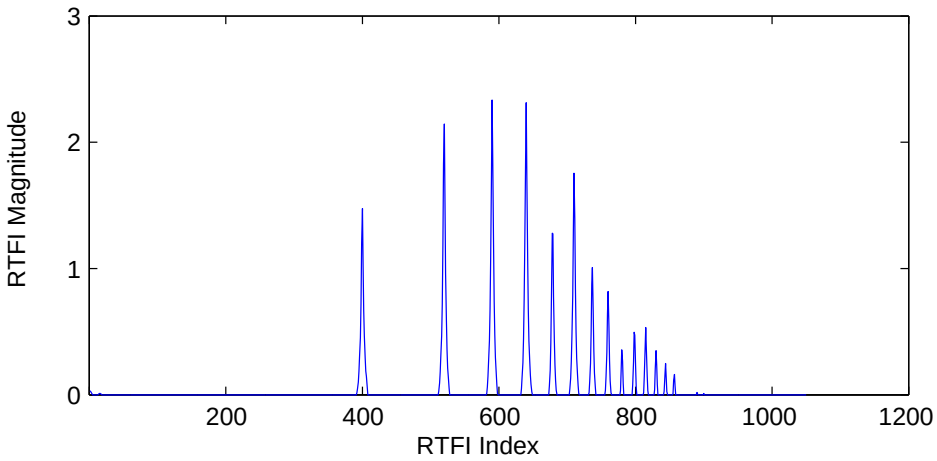
centre for digital music

Queen Mary University of London

Background (3)

- Design considerations:
  - Time-frequency representation
  - Iterative/joint estimation method
  - Sequential/joint pitch tracking





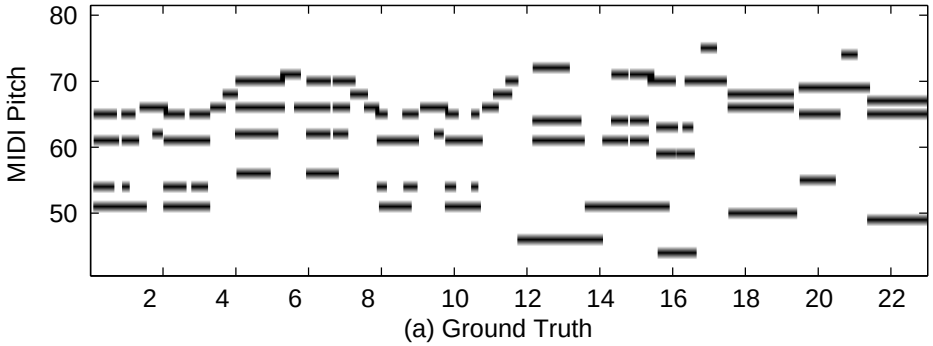
centre for digital music

Queen Mary University of London

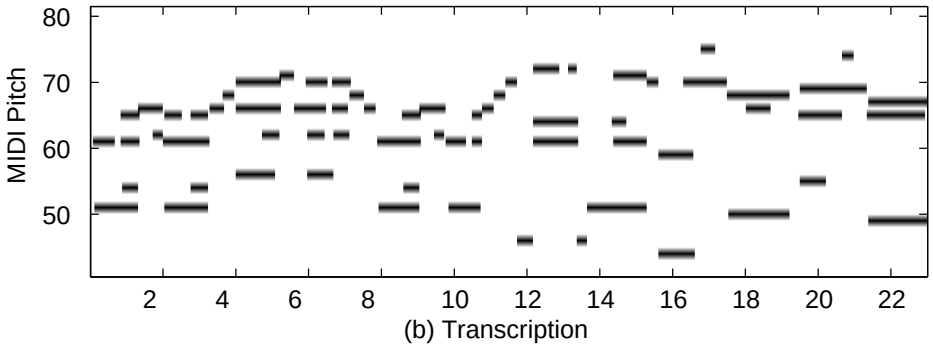
Signal Processing Approach (ICASSP 2011)

- RTFI time-frequency representation
- Preprocessing: noise suppression, spectral whitening
- Onset detection: late fusion of spectral flux & pitch salience
- Pitch candidate combinations are evaluated, modelling tuning, inharmonicity, overlapping partials and temporal evolution
- HMM-based offset detection

(a) Ground-truth guitar excerpt 'RWC MDB-J-2001 No. 9' ♩♩



(b) Transcription output of the same recording ♩♩



centre for digital music

Background: Non-negative Matrix Factorisation (NMF)

- $X = WH$
  - X is the power (or magnitude) spectrogram
  - W is the spectral basis matrix (dictionary of atoms)
  - H is the atom activity matrix
- Solved by successive approximation
  - Minimise $|| X - WH ||$
  - Subject to constraints (e.g. non-negativity, sparsity, harmonicity)
- Elegant model, but ...
  - Not guaranteed to converge to a meaningful result
  - Difficult to extend / generalise

centre for digital music

Queen Mary University of London

PLCA-based Transcription

(Ref: Benetos and Dixon, SMC 2011)

- Probabilistic Latent Component Analysis (PLCA): probabilistic framework that is easy to generalize and interpret
- PLCA algorithms have been used in the past for music transcription
- Shift-invariant PLCA-based AMT system with multiple templates:

$$P(\omega, t) = P(t) \sum_{p,s} P(\omega|s,p) *_{\omega} P(f|p,t) P(s|p,t) P(p|t)$$

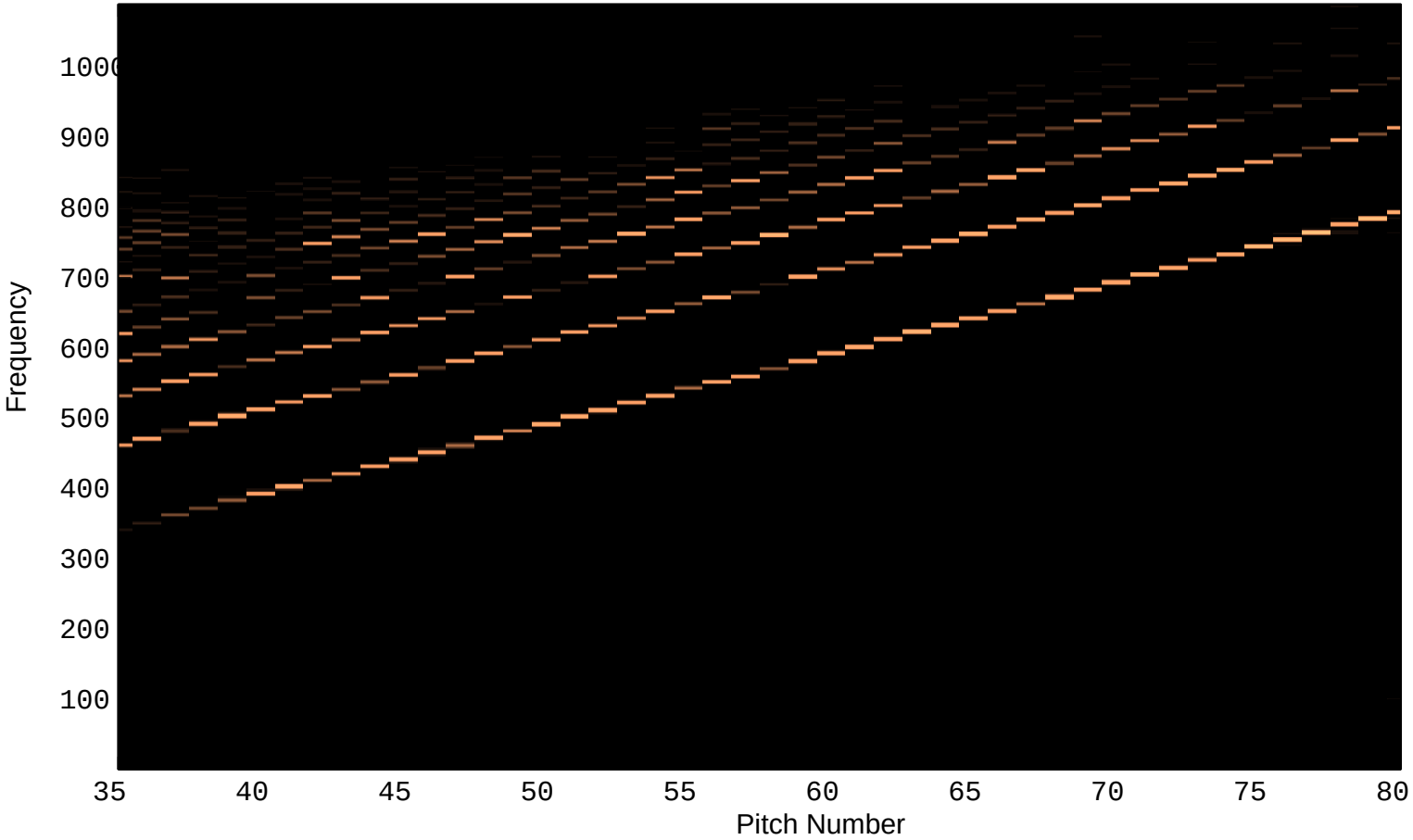
$P(\omega, t)$ is the input log-frequency spectrogram,
$P(t)$ the signal energy,
$P(\omega|s,p)$ spectral templates for instrument s and pitch p,
$P(f|p,t)$ the pitch impulse distribution,
$P(s|p,t)$ the instrument contribution for each pitch, and
$P(p|t)$ the piano-roll transcription.

centre for digital music

Queen Mary University of London

PLCA-based Transcription (2)

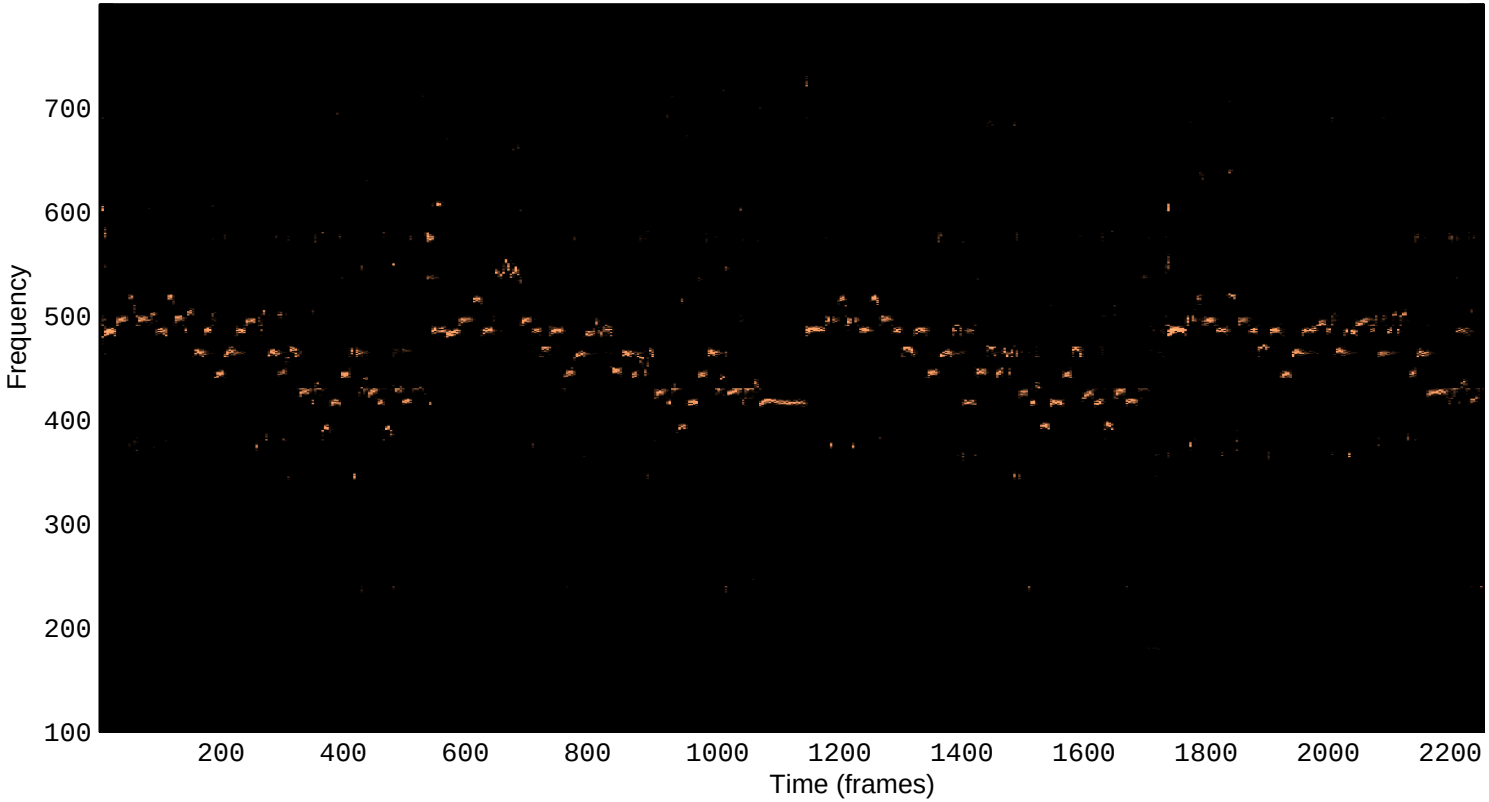
System is able to utilize sets of extracted pitch templates from various instruments (Figure: violin templates)



PLCA-based Transcription (3)

Figure: transcription of a Cretan lyra excerpt

Original:  Transcription: 



centre for digital music

Queen Mary University of London

Application 1: Characterisation of Musical Style

(Ref: Mearns, Benetos and Dixon, SMC 2011)

- Bach Chorales: audio and MIDI versions
- Audio transcribed and segmented by beats
- Each vertical slice is matched to chord templates based on Schoenberg's definition of diatonic triads and 7ths for the major and minor modes
- HMM detects key and modulations
  - Observations are chord symbols, hidden states are keys
  - Observation matrix models relationships between chords and keys
  - State transition matrix models modulation behaviour
  - Various models based on Schoenberg and Krumhansl — comparison of models from music theory and music psychology

	Transcription	Chords	Key
Audio	33%	56%	64%
MIDI	(100%)	85%	65%

centre for digital music

Queen Mary University of London

Application 2: Analysis of Harpsichord Temperament

(Ref: Tidhar et al. JNMR 2010; Dixon et al. JASA 2011, to appear)

- For fixed-pitch instruments, tuning systems are chosen which aim to maximise the consonance of common intervals
- Temperament is a compromise arising from the impossibility of satisfying all tuning constraints simultaneously
- We measure temperament to aid study of historical performance
  - Perform "conservative" transcription
  - Calculate precise frequencies of partials (CQIFFT)
  - Estimate fundamental and inharmonicity of each note
  - Classify temperament

	Synthesised	Acoustic (real)
Spectral Peaks	96%	79%
Conservative Transcription	100%	92%

centre for digital music

Queen Mary University of London

Application 3: Analysis of Cello Timbre

(Ref: Chudy, Benetos and Dixon, work in progress)

- Expert human listeners can recognise the “sound” of a performer
  - Not just the recording
  - Not just the instrument
- Can this be captured in standard timbral features?
  - Temporal: attack time, decay time, temporal centroid
  - Spectral: spectral centroid, spectral deviation, spectral spread, irregularity, tristimuli, odd/even ratio, envelope, MFCCs
  - Spectrotemporal: spectral flux, roughness
- In order to estimate features, it is necesary to identify the notes (onset, offset, pitch, etc.)
  - As a manual task, this is very time consuming
  - Conservative transcription solves the problem: high precision, low recall
  - System is trained on cello samples from RWC database

centre for digital music

Queen Mary University of London

Take-home Messages

- Automatic transcription enables new musicological questions to be investigated
- Many other applications: music education, practice, MIR
- The technology is not perfect but music is highly redundant
- Shift-invariant PLCA is ideal for instrument-specific polyphonic transcription

- Proposed procedure for folk music analysis:
  - Extract templates for desired instruments (e.g. using NMF)
  - Perform shift-invariant PLCA transcription
  - Estimate features of interest
  - Interpret results according to estimates obtained

centre for digital music

Queen Mary University of London

Any Questions?

- Acknowledgements
  - Emmanouil Benetos, Lesley Mearns, Magdalena Chudy: the PhD students who do all the work
  - Dan Tidhar, Matthias Mauch: collaboration on the harpsichord project
  - Aggelos, Christina and EURASIP for the invitation

centre for digital music

Queen Mary University of London