Analysing Musical Audio

Mark Plumbley
Centre for Digital Music
Queen Mary, University of London

Outline

A snapshot / lightning tour of some of our work in:

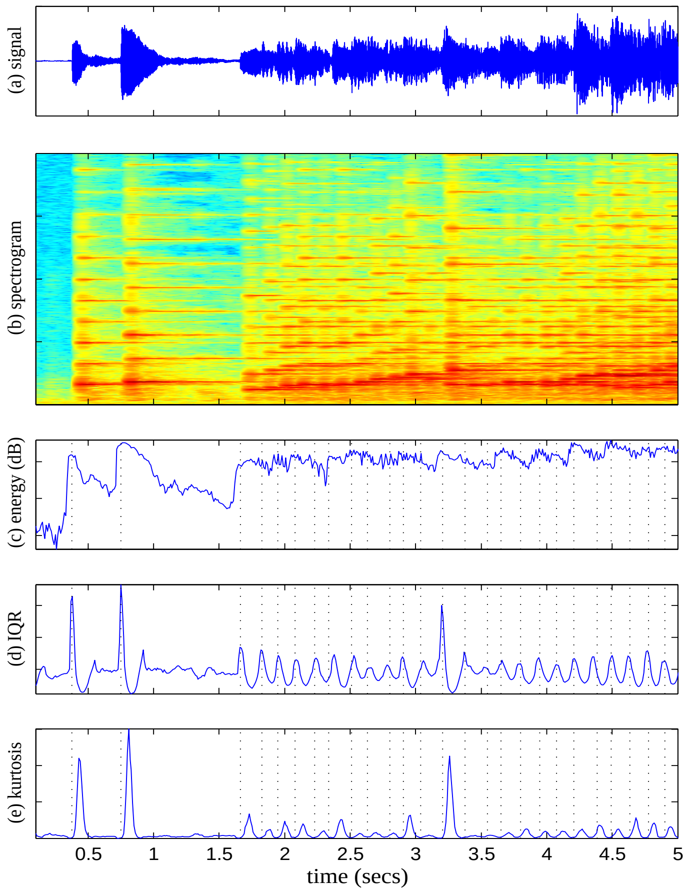
- Musical Audio Analysis
  - Beat Analysis
  - Music Transcription
  - Visualisation
  - Interaction
- Audio Source Separation

Part 1. Musical Audio Analysis
Beat analysis

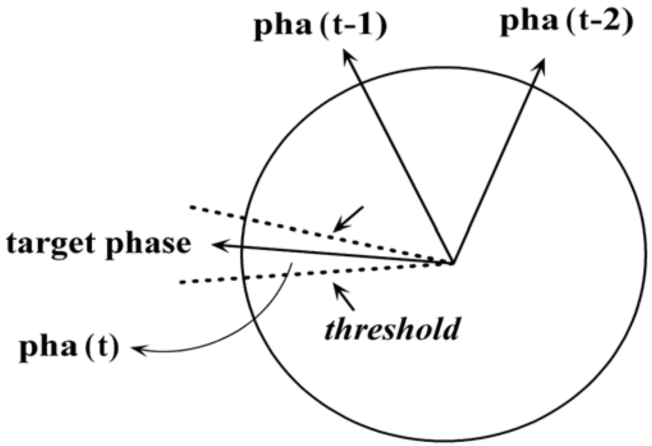
Queen Mary
University of London

centre for digital music

Step 1: Onset Detection



1. Take the Fourier Transform of frames
2. Measure the phase
3. Predict next phase from last 2 phases
4. Diff = onset detection function (DF)



Queen Mary University of London

centre for digital music

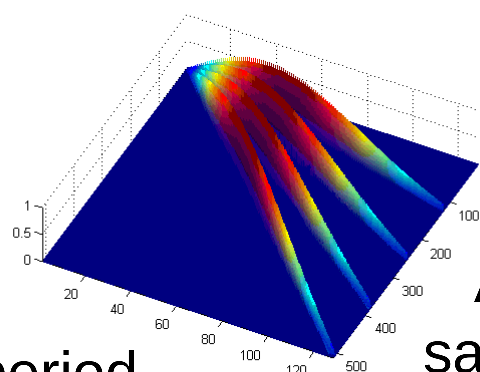
Step 2: Beat Period

How estimate beat period?

-> Peaks in autocorrelation (ACF) of Detection Function (DF)

But…how choose the right level?

-> weighted comb filter

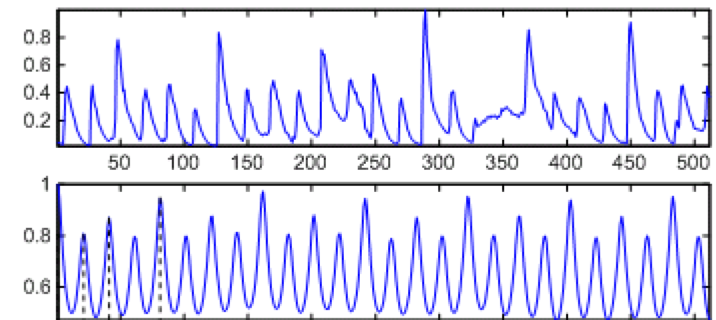


Beat period

ACF samples

DF

ACF



$\tau_1 = 20$ tempo = 260 bpm

$\tau_2 = 40$ tempo = 130 bpm

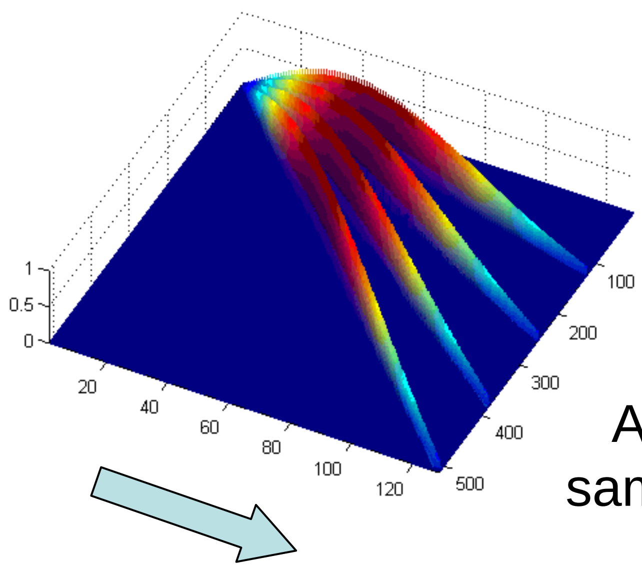
$\tau_3 = 80$ tempo = 65 bpm

Emphasis on *causal* implementation
(Don't use future information)

Queen Mary
University of London

centre for digital music

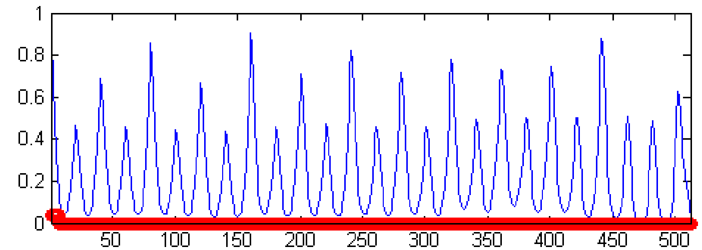
Beat Period



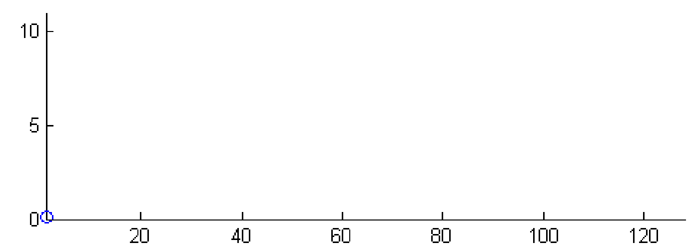
ACF samples

Beat period

ACF



Output of comb filterbank



Beat period (bpm)

Queen Mary University of London

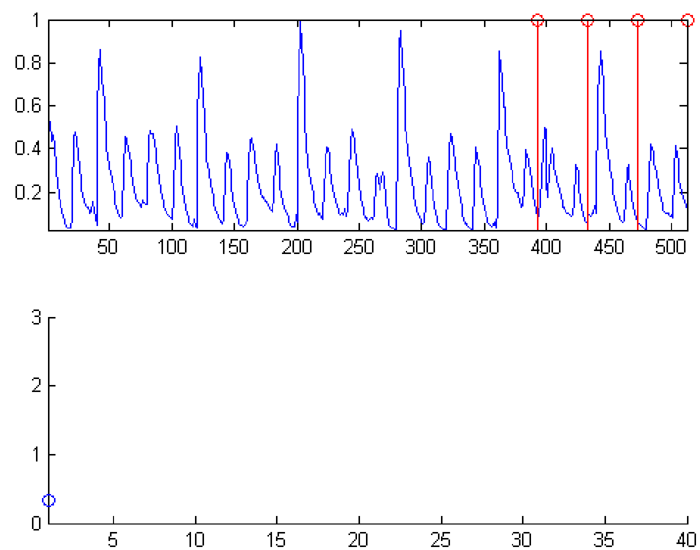
centre for digital music

Step 3: Alignment & Prediction

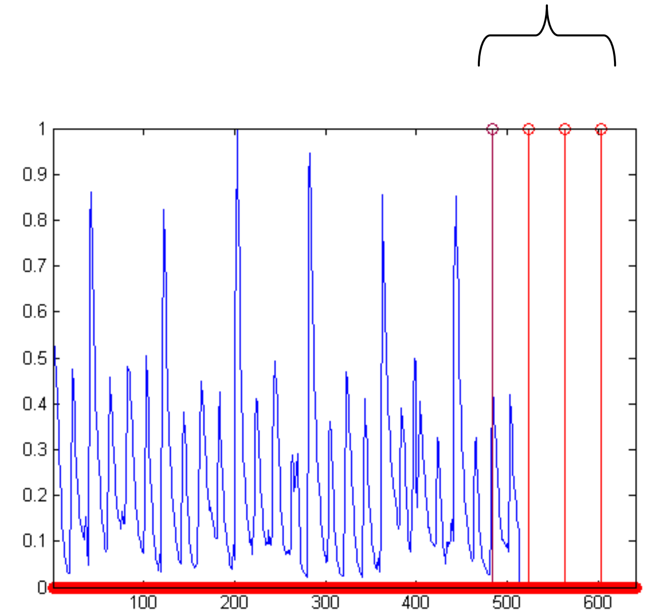
(1) Align comb at beat period with strongest DF peaks

DF

Comb filter output



(2) Predict at beat period intervals



Queen Mary
University of London

centre for digital music

Not the end of the story…

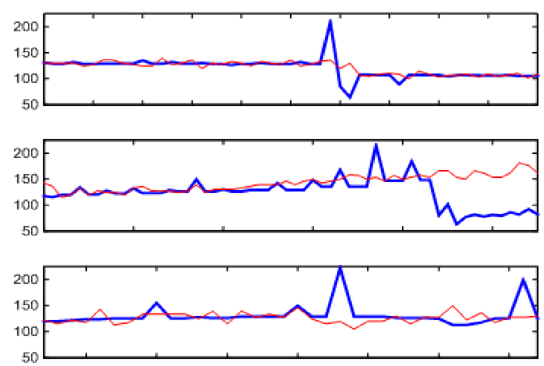
Simple model:

Changes are hardest:

Step, Ramp, Expressive

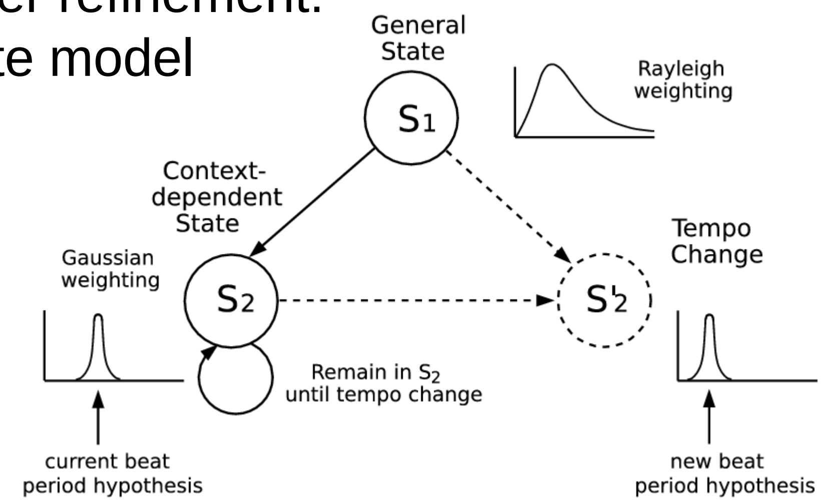
Human tapping
vs algorithm

bpm

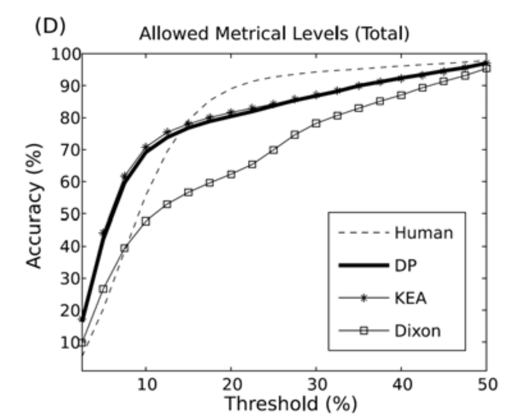


time

Further refinement:
2-state model



Performance close to state-of-art (Klapuri et al, 2006) but much less complex.



Queen Mary
University of London

centre for digital music

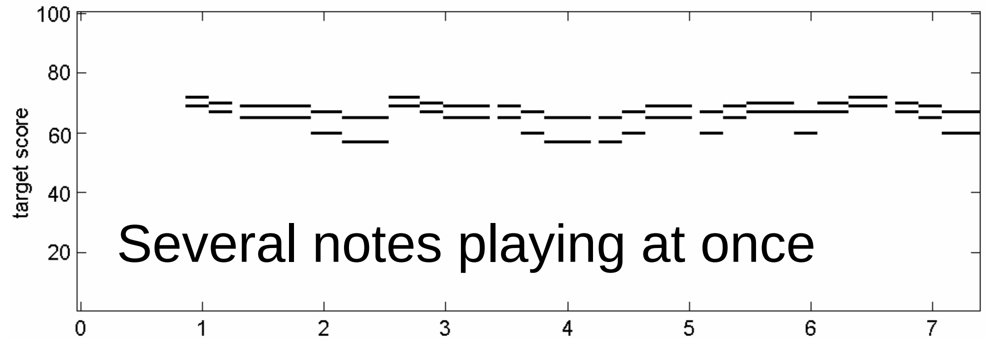
Music Transcription

Queen Mary
University of London

centre for digital music

Polyphonic Transcription Problem

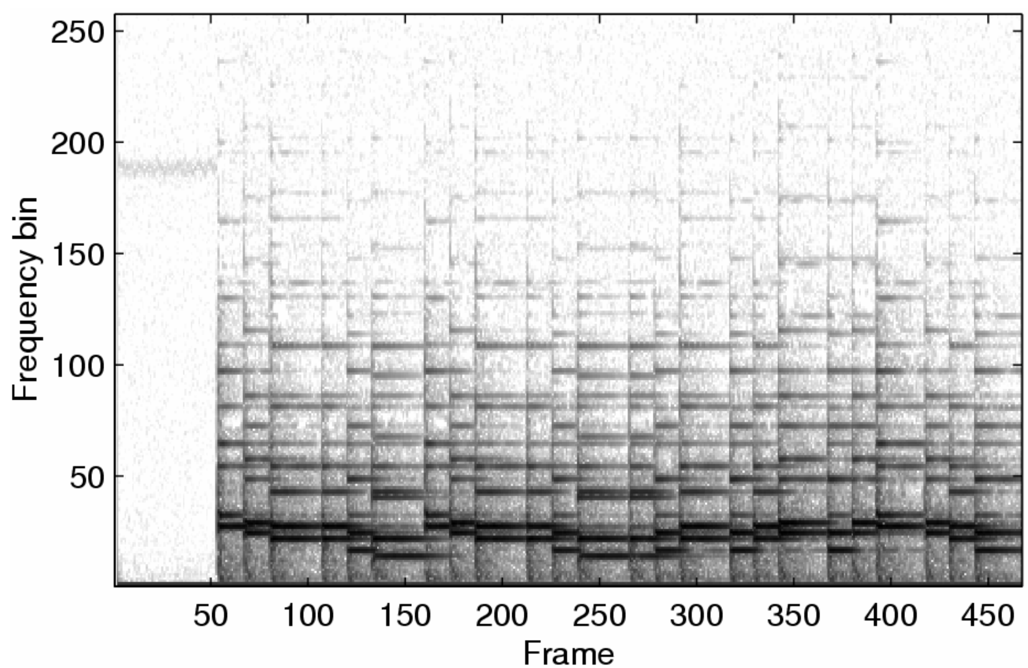
Notes → MIDI (“Piano Roll”)



Synth/sample/FT

(Liszt: Etude No. 5 aus Grandes Etudes de Paganini. MIDI from Classical Piano Midi Page http://www.piano-midi.de, copyright Bernd Krueger)

Task: Extract notes from e.g. this →



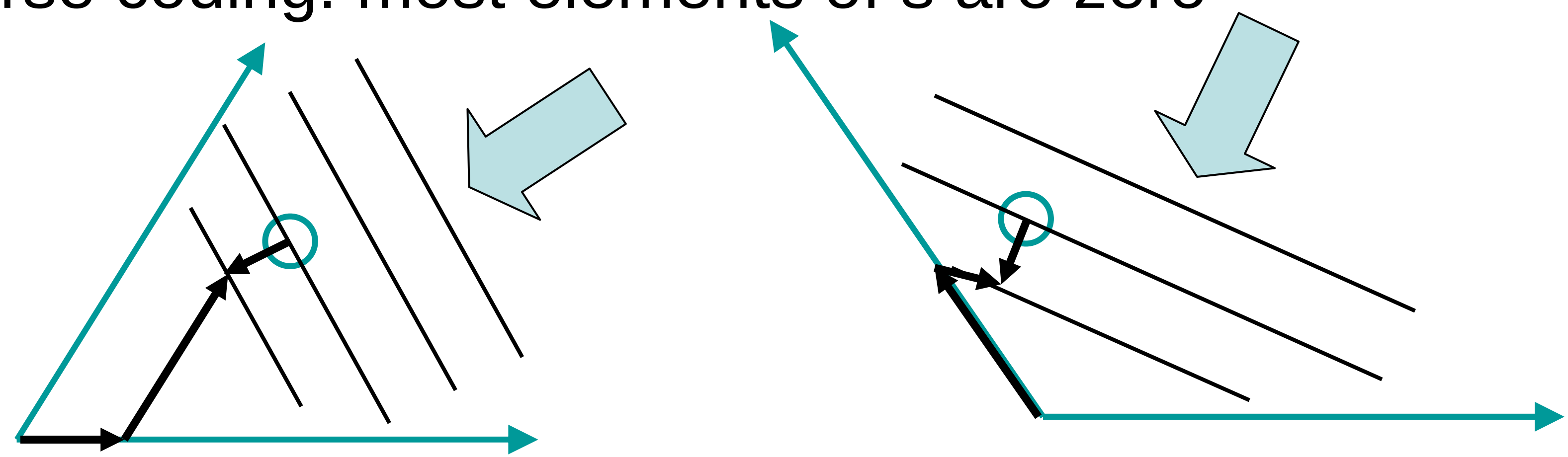
Queen Mary
University of London

Generative Model

Spectra x are weighted sum A of source s plus noise e

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{e}$$

Simplest case, s = activity of individual notes
Assume gaussian iid e with sparse priors $p(s_i)$
Inferred representations s shrink towards zero
-> Sparse coding: most elements of s are zero



(Also have to handle learning the matrix A)

Queen Mary
University of London

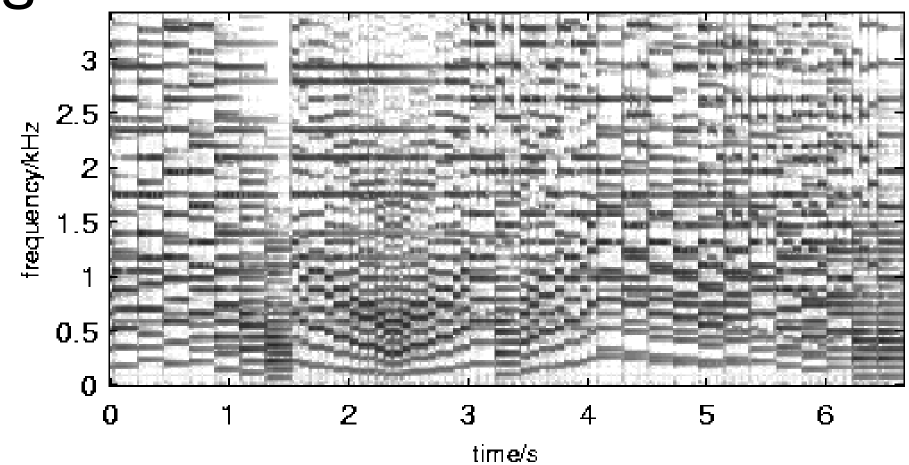
centre for digital music

Example: Synthetic Harpsichord

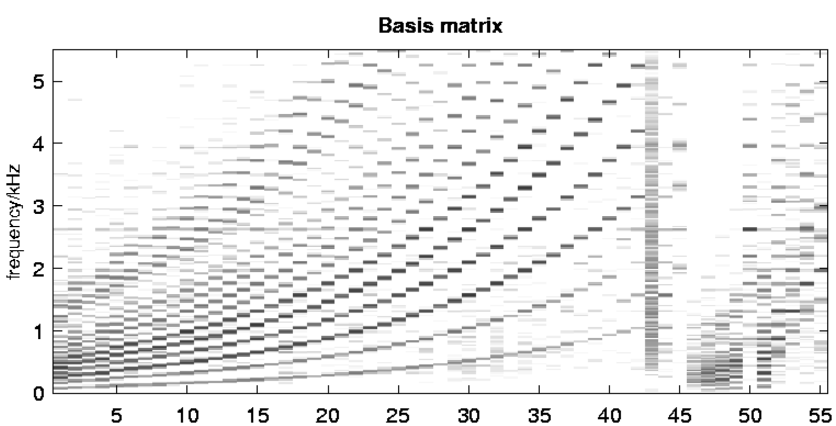
Result: 54 non-zero vectors

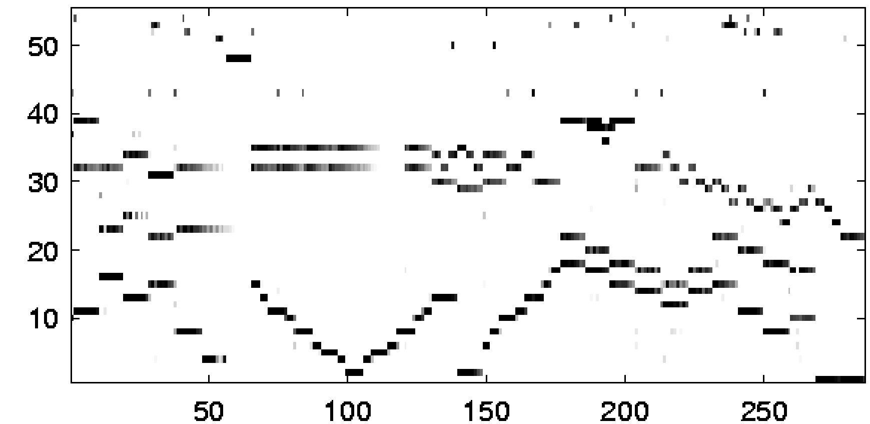
Original audio

MIDI Bach Partita in A Minor BWV827



Dictionary matrix





Resynth MIDI Piano

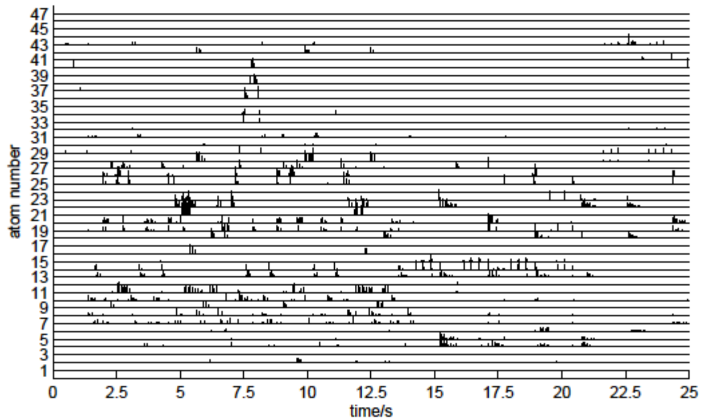
Extracted sources

Queen Mary
University of London

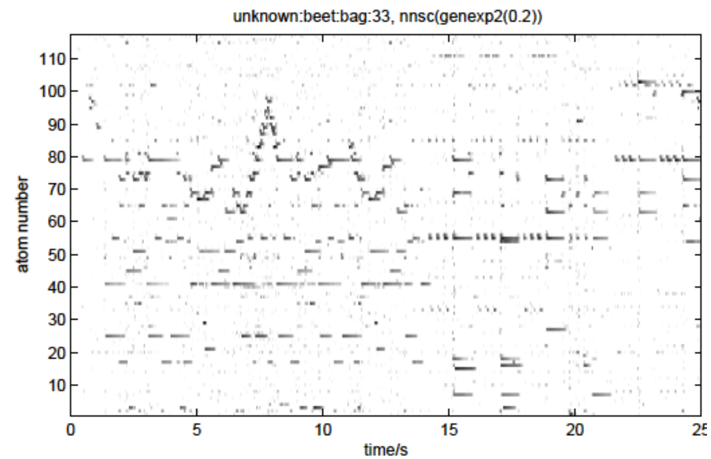
centre for digital music

Example: Real Piano

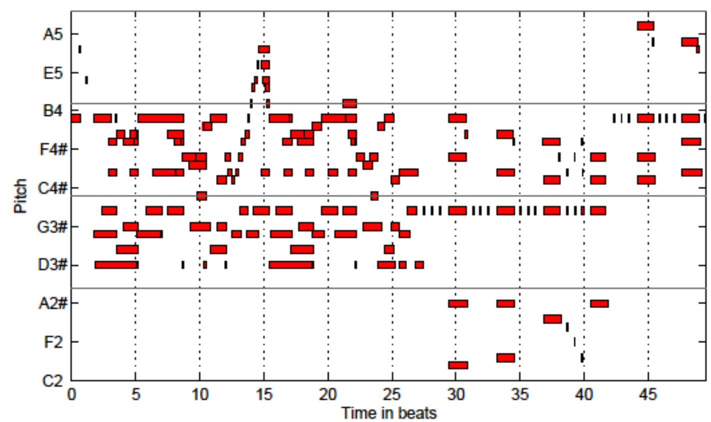
Time domain sparse coding



Freq domain sparse coding



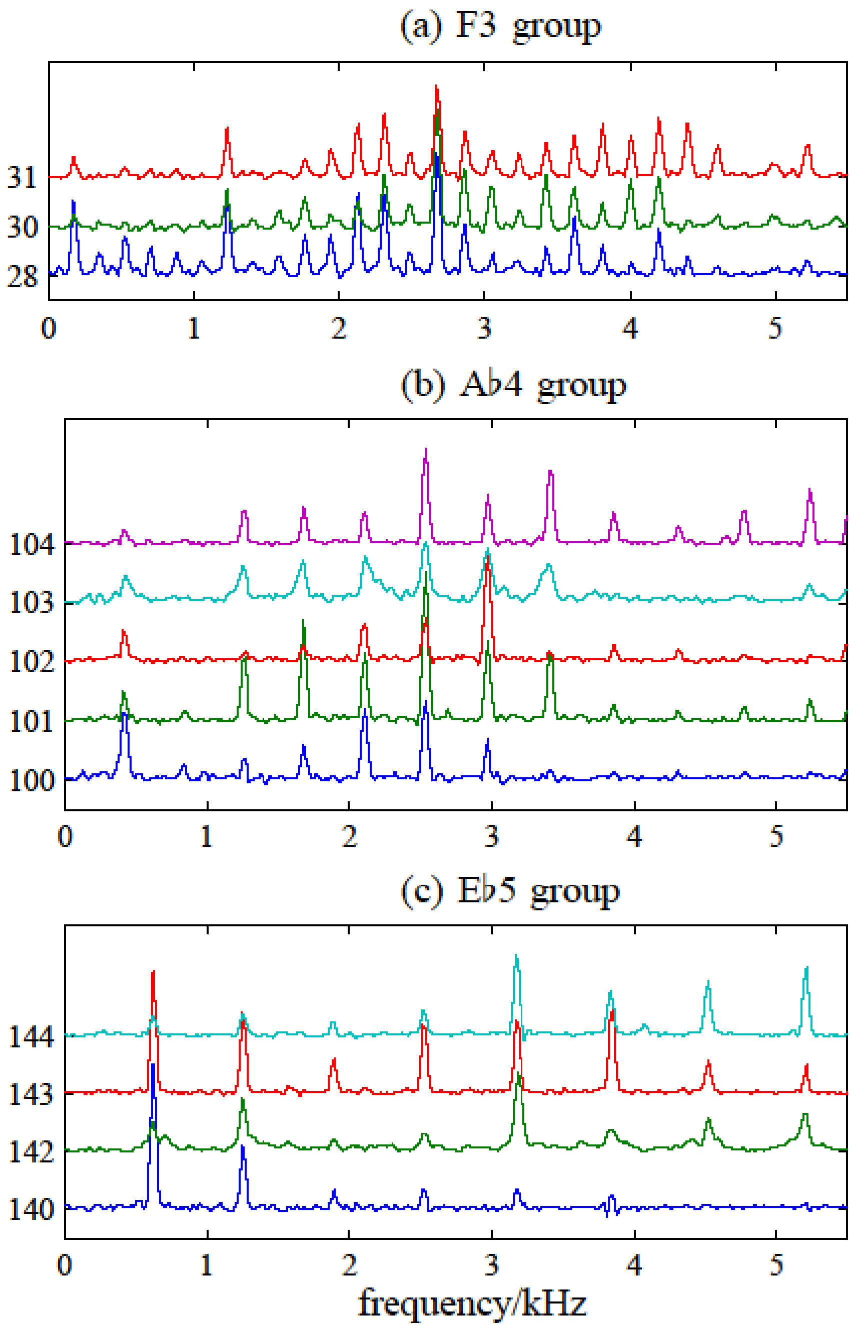
Beethoven:
Bagatelle, Opus 33
No. 1 in E*b* Major



Original
MIDI

Queen Mary
University of London

centre for digital music

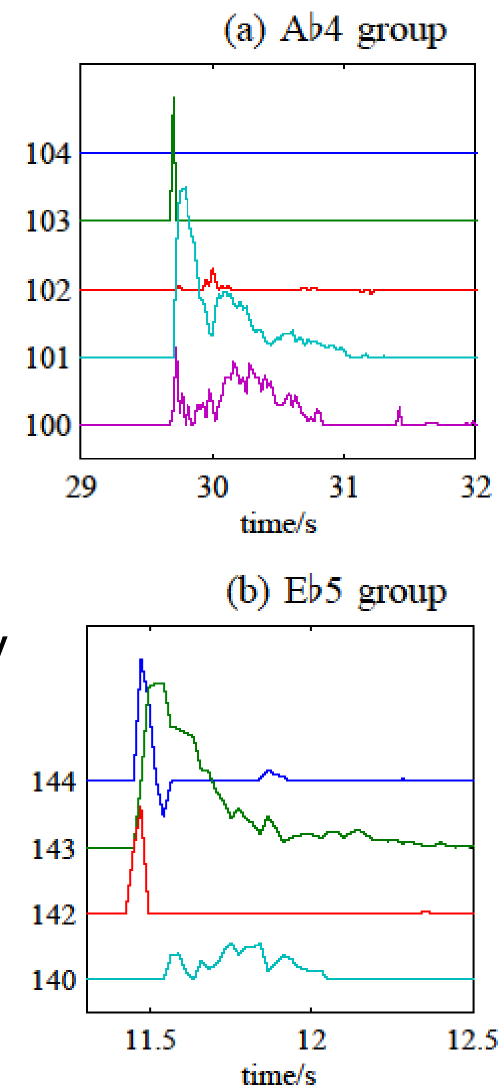


Pitch groups

A single musical "note" is made up from a subspace of underlying dictionary vectors.

Dictionary vectors

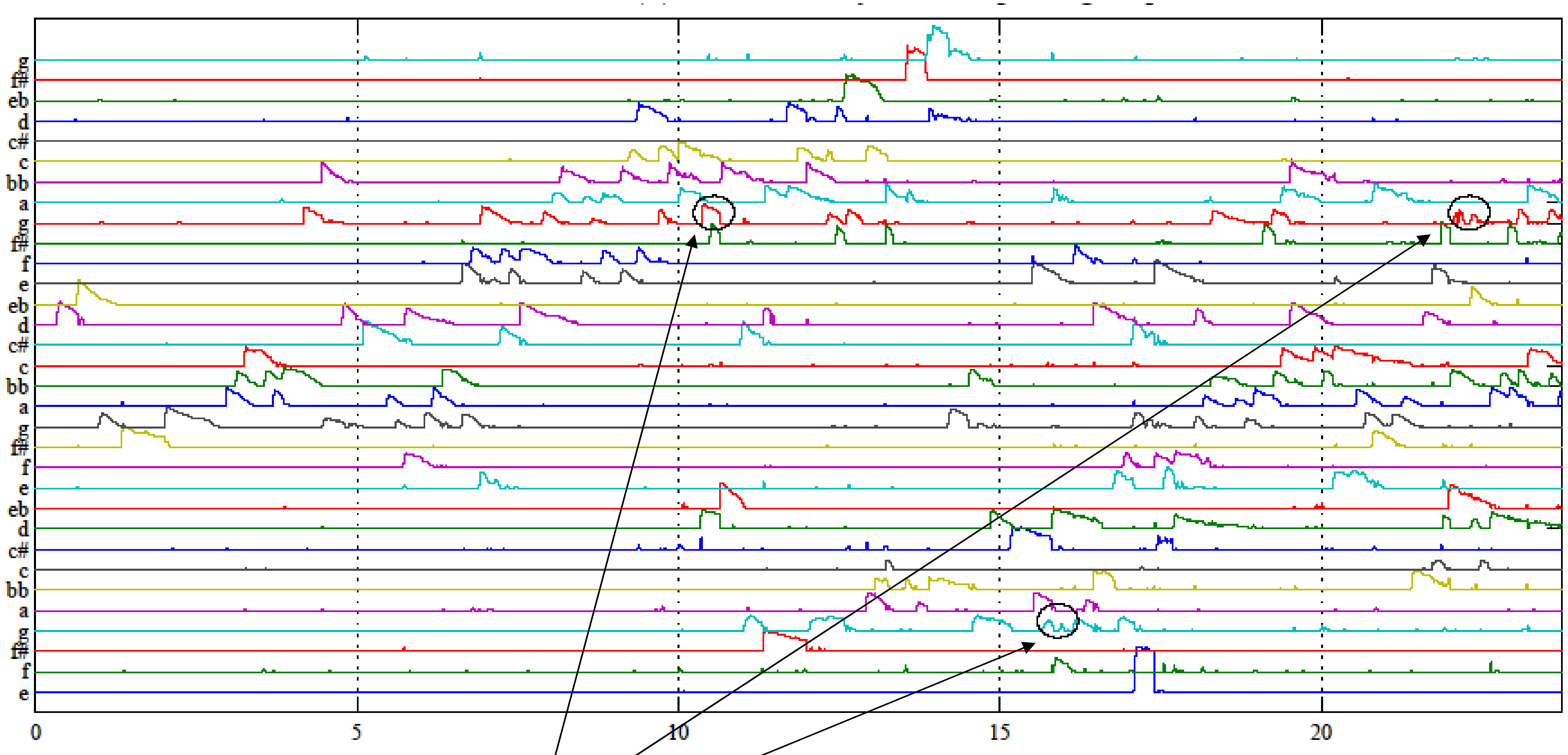
Time Activity



Queen Mary University of London

centre for digital music

Pitch group activities



Transcription errors

Queen Mary University of London

centre for digital music

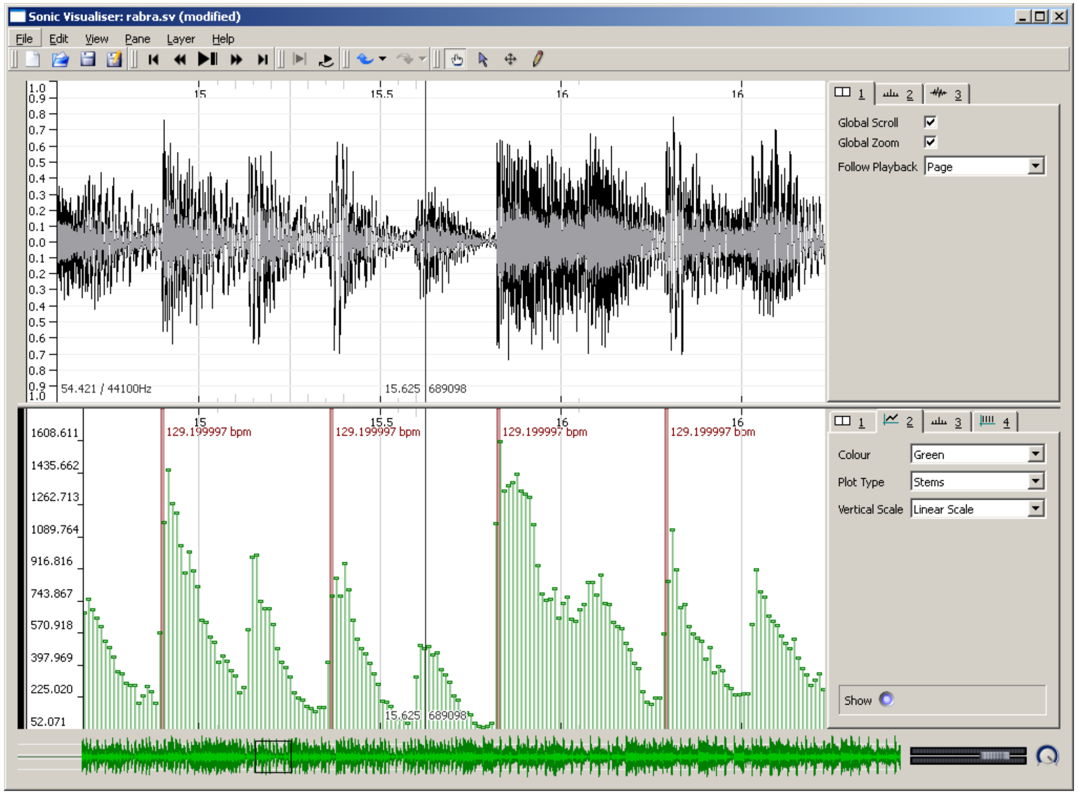
Visualisation

Queen Mary
University of London

centre for digital music

Sonic Visualiser

- Viewing & editing audio semantic descriptors
- Overlaying descriptors
- Independent zooming with linked scrolling
- Open source



Queen Mary
University of London

centre for digital music

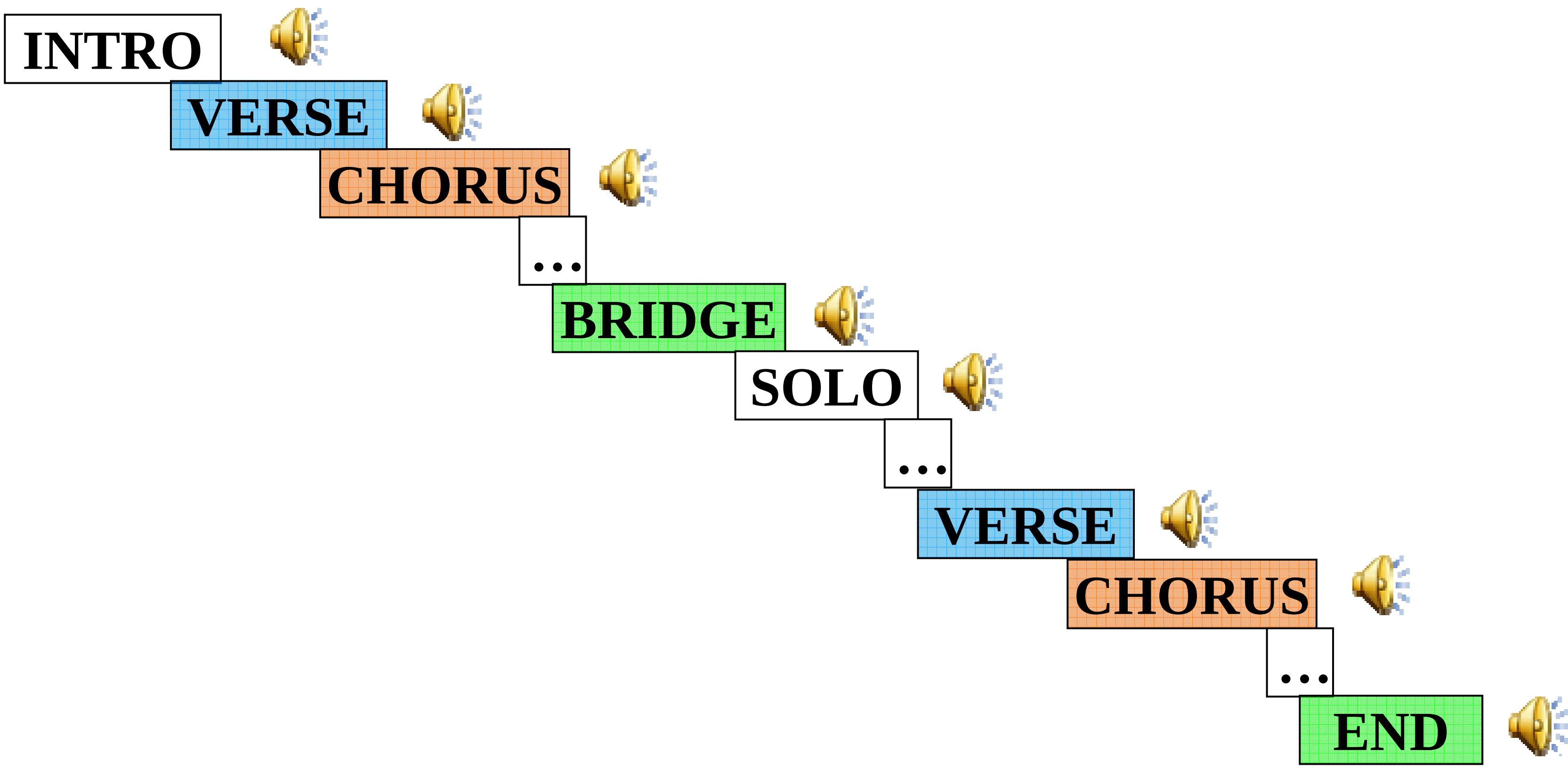
Analysis: Segmentation

Queen Mary
University of London

centre for digital music

Structure in Music

The Beatles – Let it Be

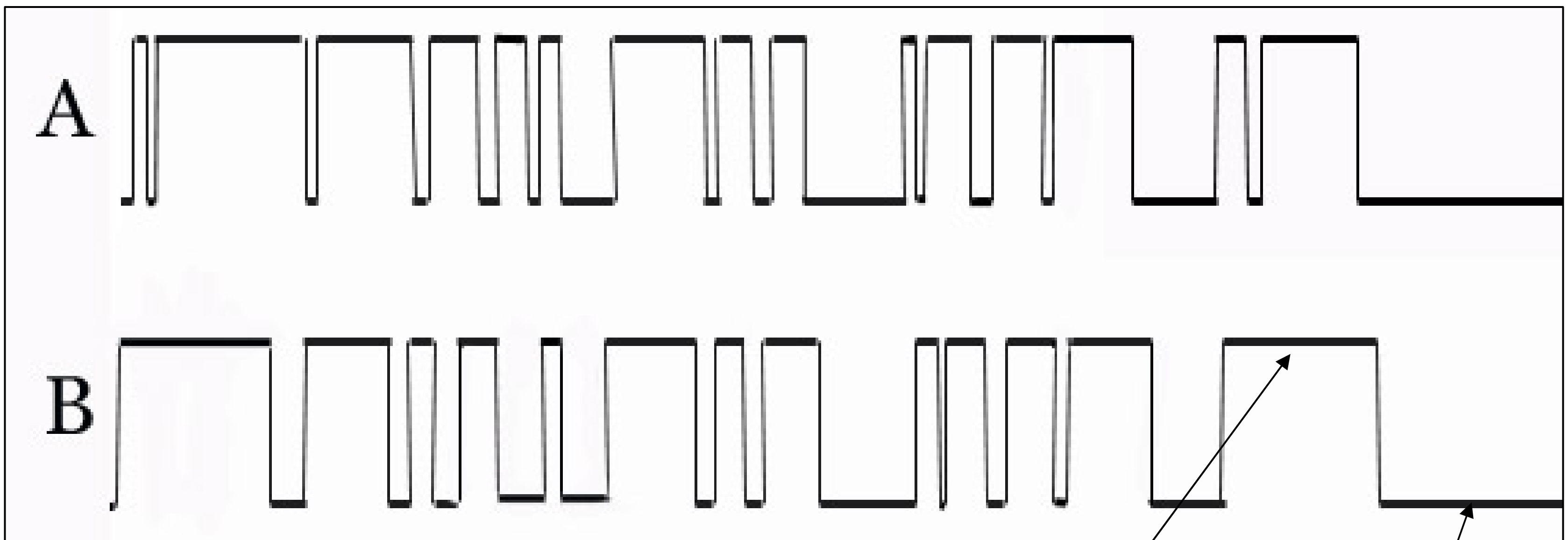


Queen Mary
University of London

centre for digital music

Finding repeating patterns

Look for repeating changes of texture (timbre)
Large approximate pattern: the verse



Thumbnailing

Composite summary of a piece

Piano + voice

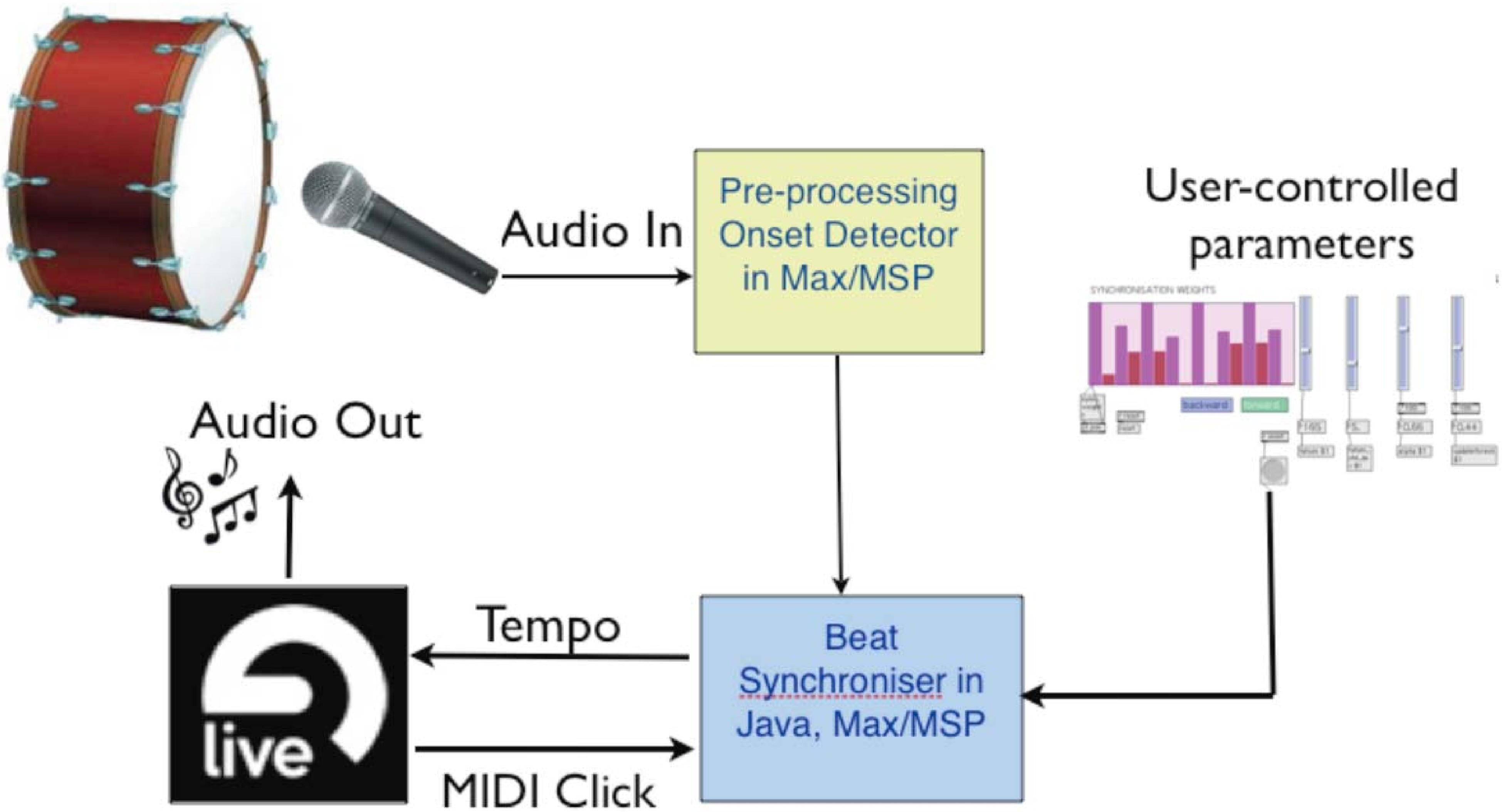
Piano alone

Interaction: B-Keeper

Queen Mary
University of London

centre for digital music

B-Keeper



Audio In
Pre-processing Onset Detector in Max/MSP
User-controlled parameters
Audio Out
Tempo
Beat Synchroniser in Java, Max/MSP
MIDI Click
live

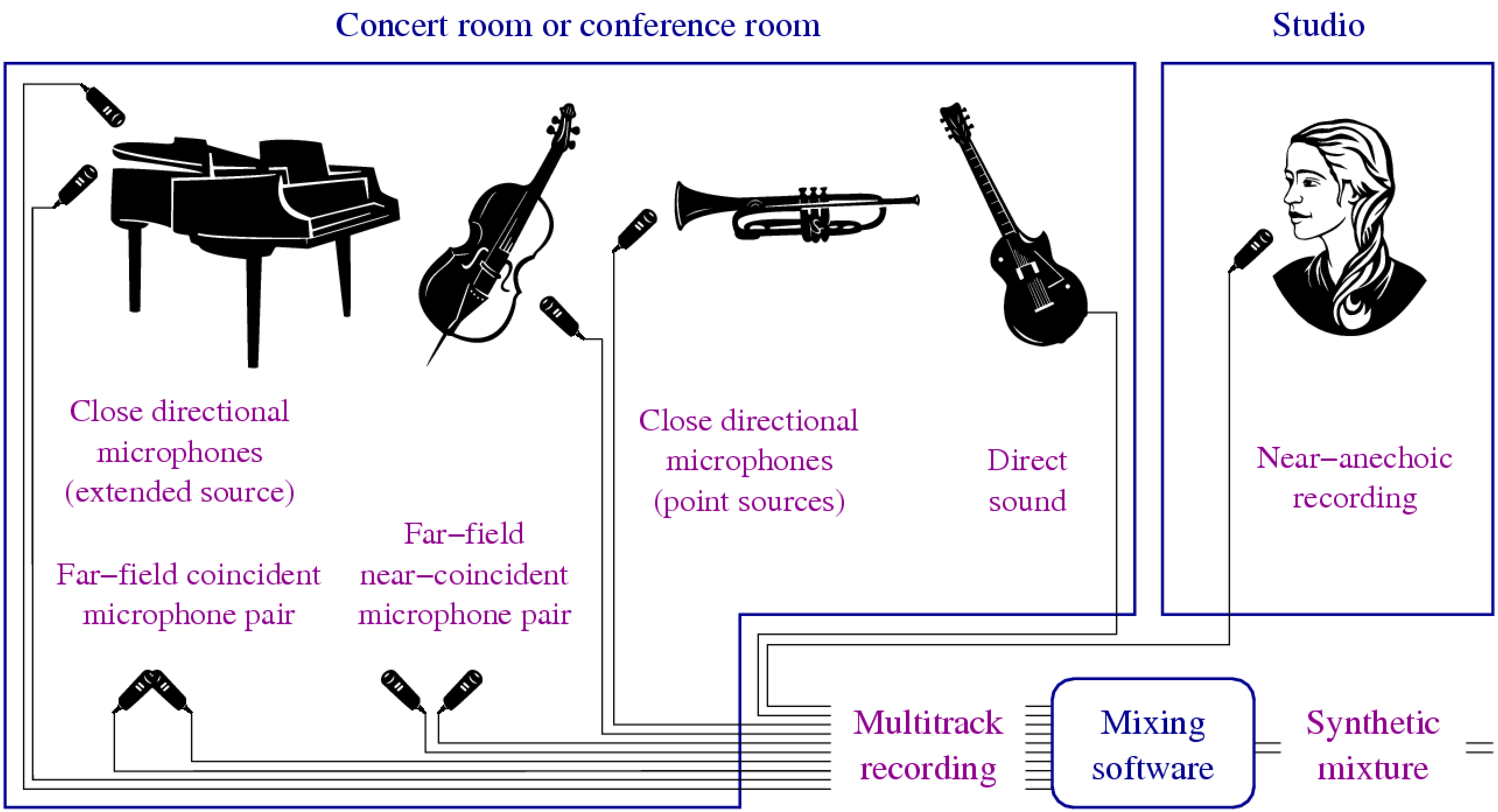
[Video]

Part 2. Source Separation

Queen Mary University of London

centre for digital music

Mixing musical audio

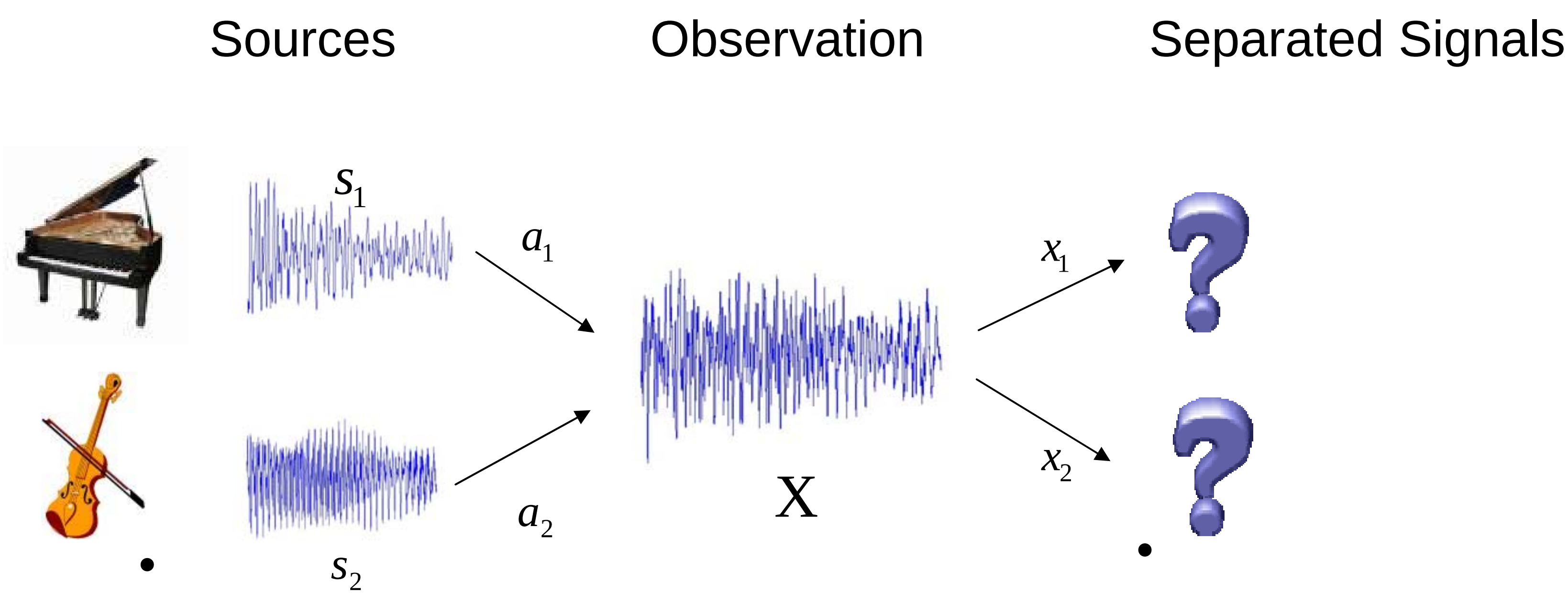


Queen Mary University of London

centre for digital music

Source Separation from Mono Mix

If no of sources = no of microphones:
-> use independent component analysis (ICA) and variants.
But what if fewer microphones – or only one?



$$s_1 a_1 + s_2 a_2 + \ldots = X$$

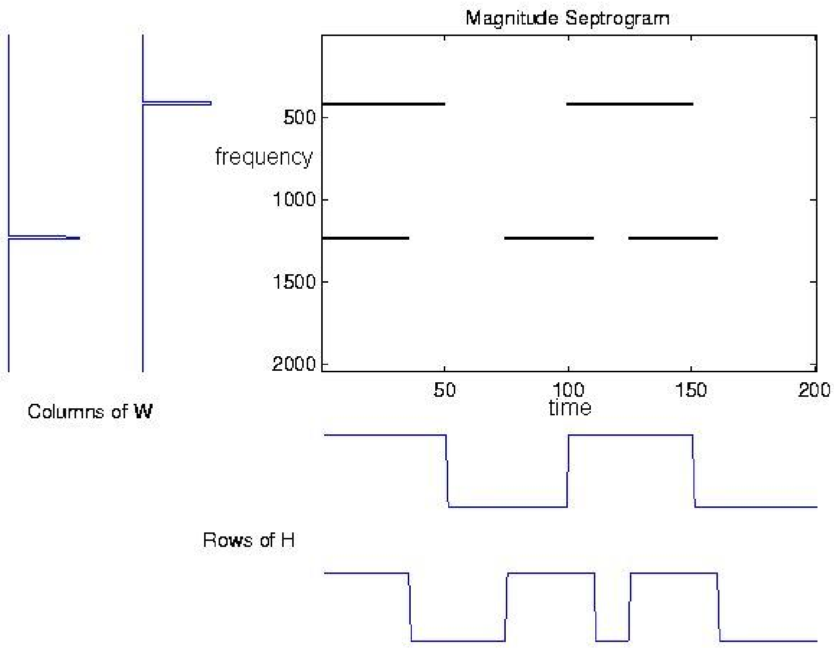
One approach: Nonnegative Matrix Factorization (NMF) of spectrograms.

Nonnegative Matrix Factorization

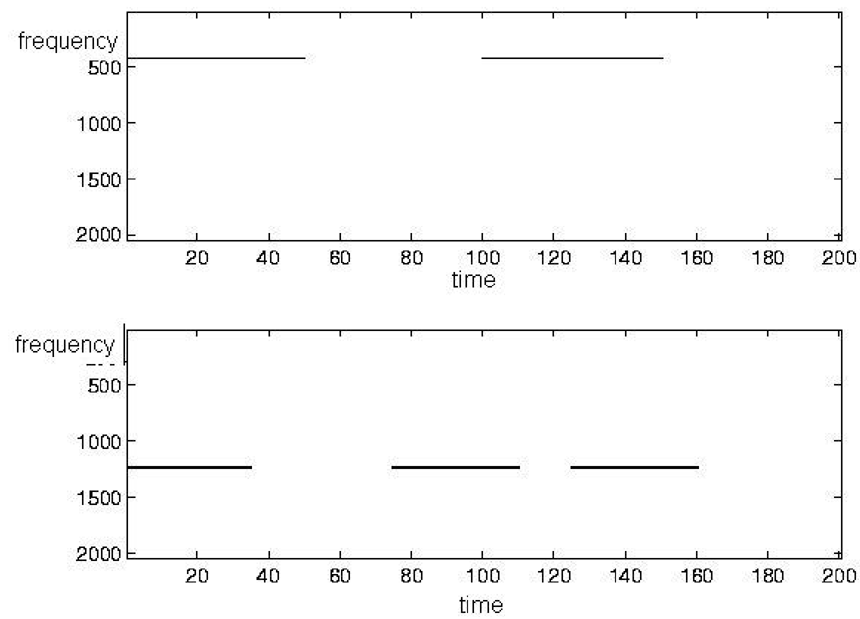
Lee & Seung (2000)

Decompose spectrogram matrix V into nonnegative product: $\mathbf{V} \approx \mathbf{WH}$

Decomposition:

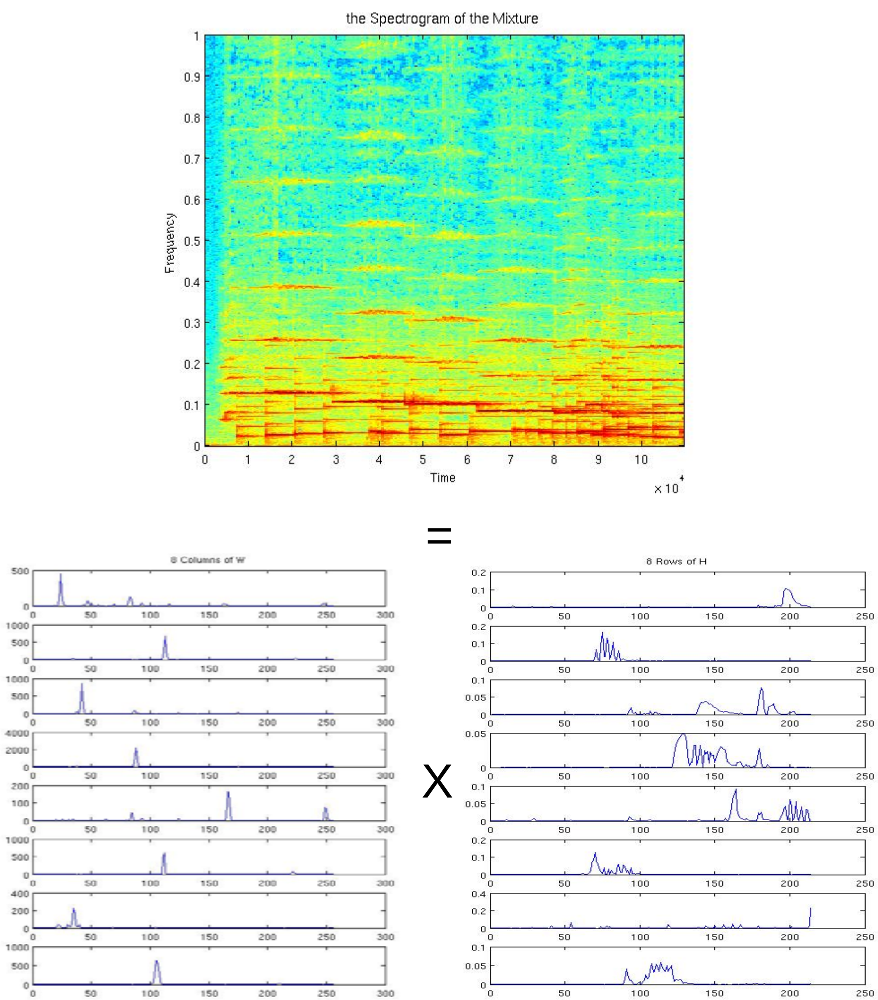


Separation:

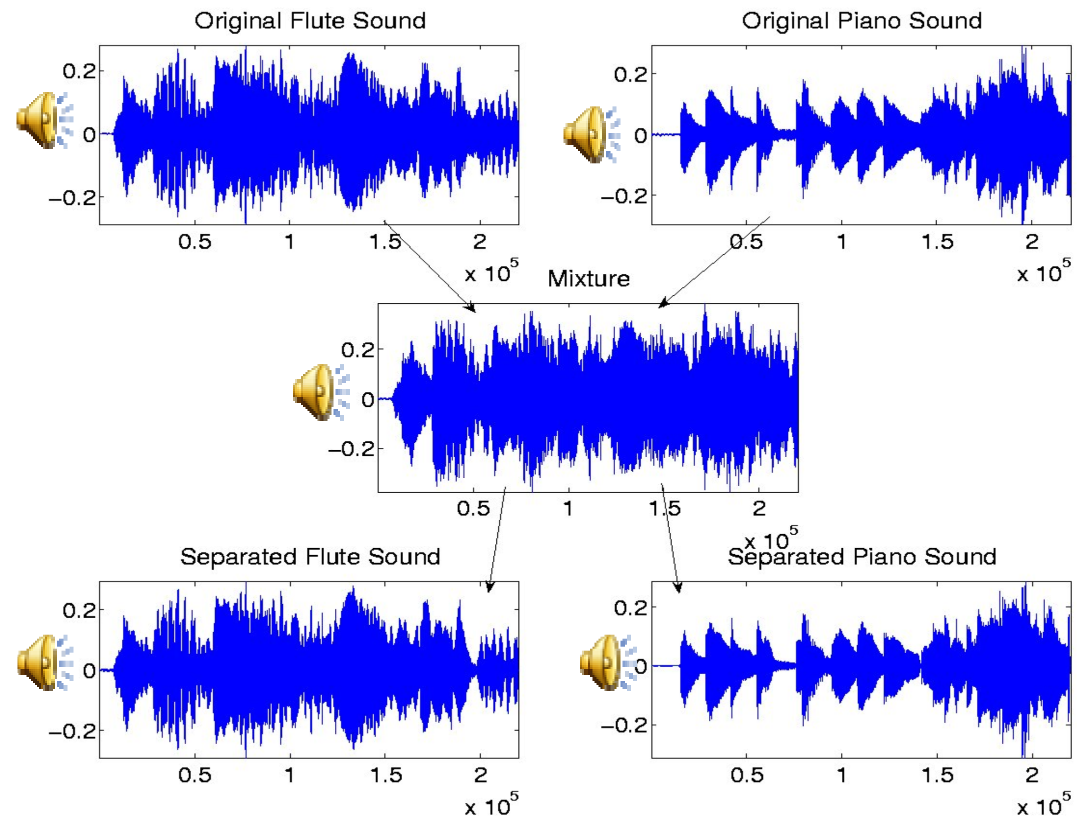


NMF Example

Separate the audio using *time-freq masking.*
Basis vectors can be grouped by hand.
Also investigating source-directed grouping.



Artificial mixture

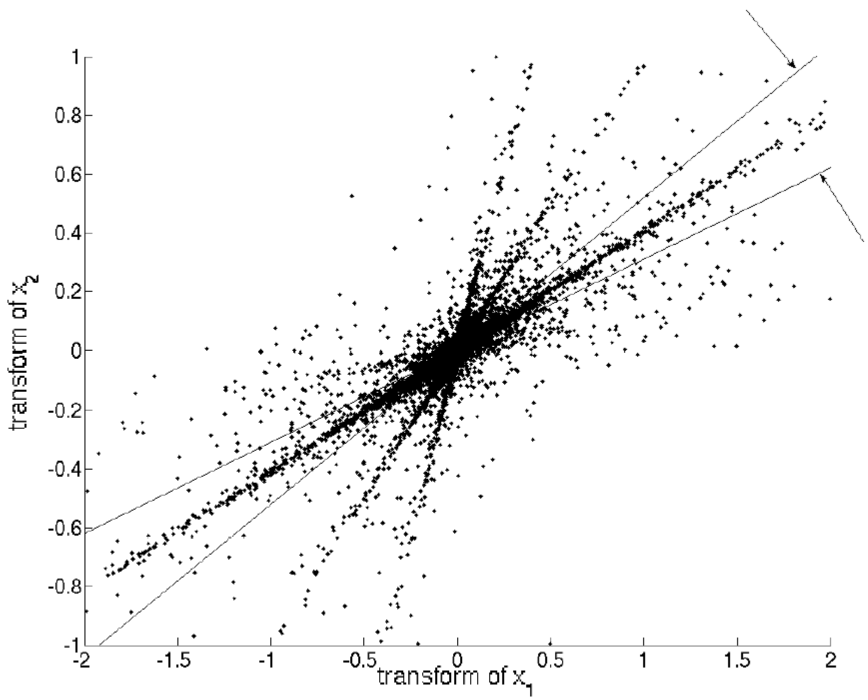


Source Separation from Stereo

For pan-potted stereo, sources have "angle": $\theta_i = \arctan(x_2 / x_1)$

Transform into a *sparse* representation of the sources.

Typically short-time Fourier transform (STFT) used.



4 sources mixed to stereo (2 channels)

Separate *i*th source by keeping only components arriving near its "angle"

Development of DUET algorithm (Yilmaz & Rickard, 2004)

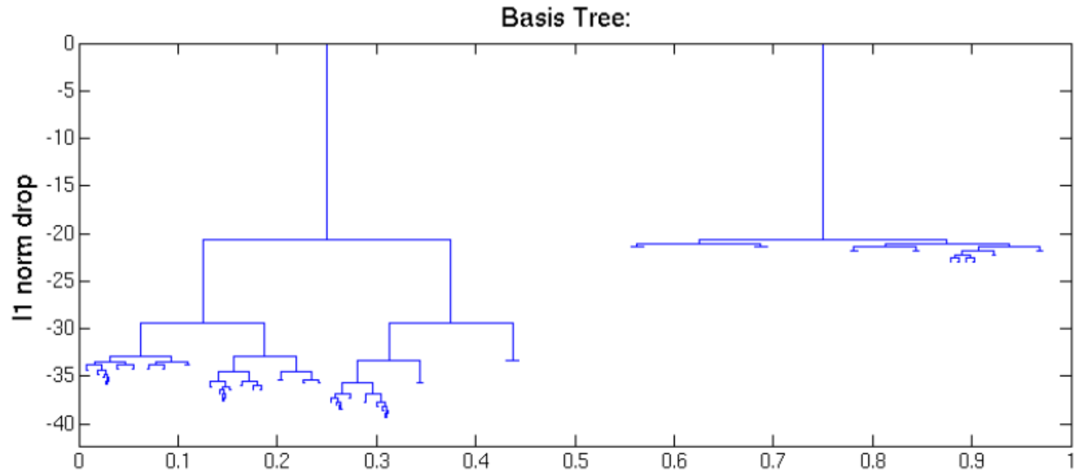
Queen Mary University of London

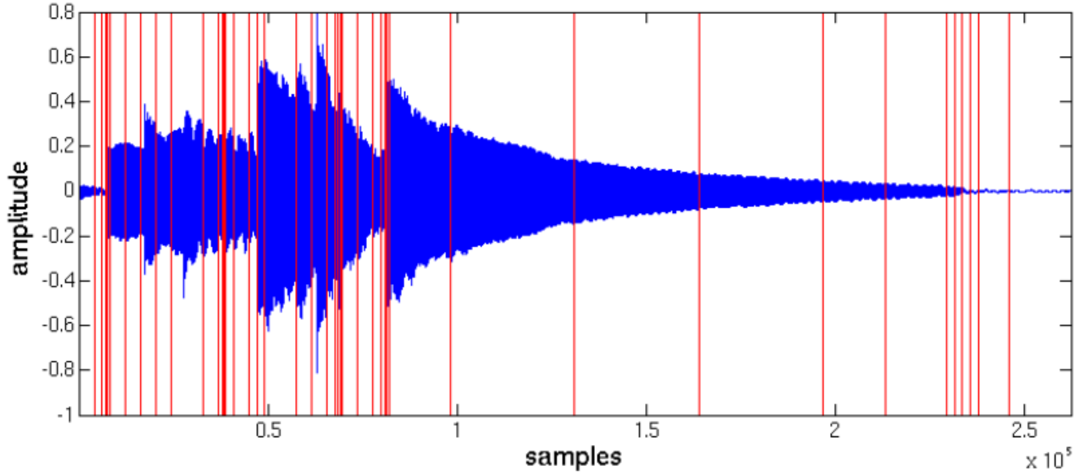
centre for digital music

Cosine Packet Trees

Alternative to STFT…
Adapt frame size to signal.

Can do this efficiently with Cosine Packet Tree / Best Basis (Coifman & Wickerhauser, 1992)





Queen Mary
University of London

centre for digital music

Stereo Separation Example

A synthetic mixture of "real" tracks (*Personalized Perfection* by Another Dreamer)

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.90 & 0.71 & 0.50 & 0.28 \\ 0.09 & 0.29 & 0.50 & 0.72 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{pmatrix}$$

Original Sources

Stereo Mix

Estimated Sources

Percussion

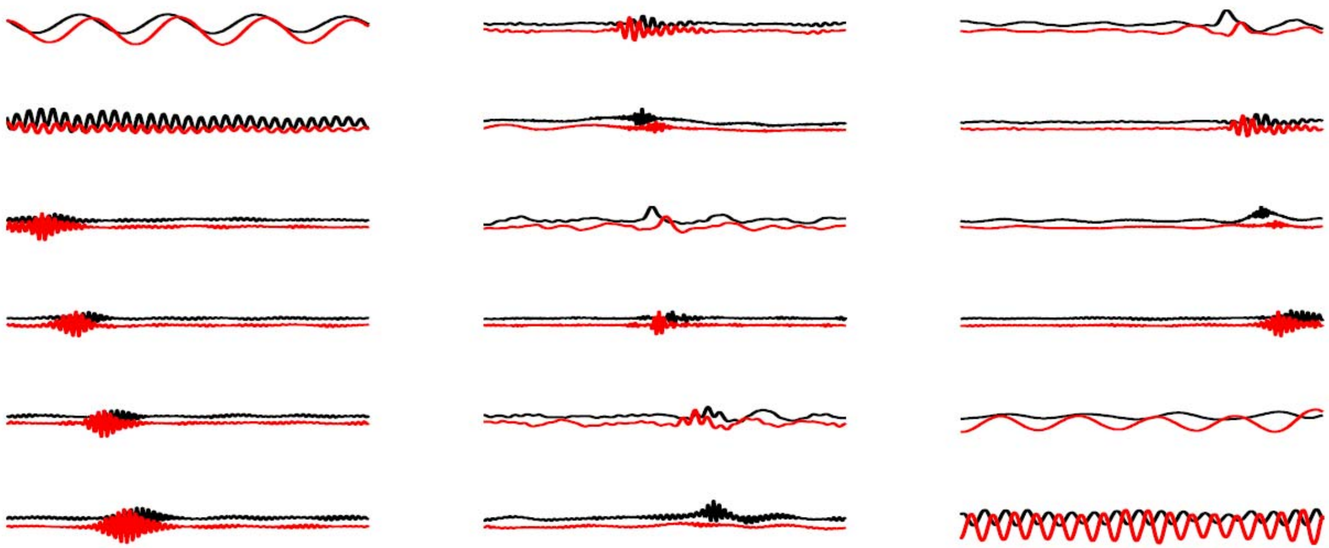
Guitar 1

Vocal

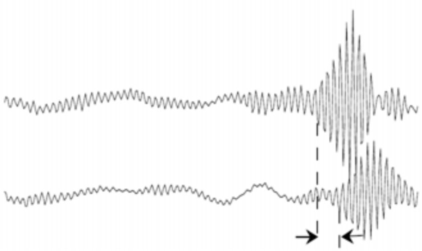
Guitar 2

Convolutive Stereo Mixtures

- Sparsifying transform in place of STFT
- Finds basis functions with delays



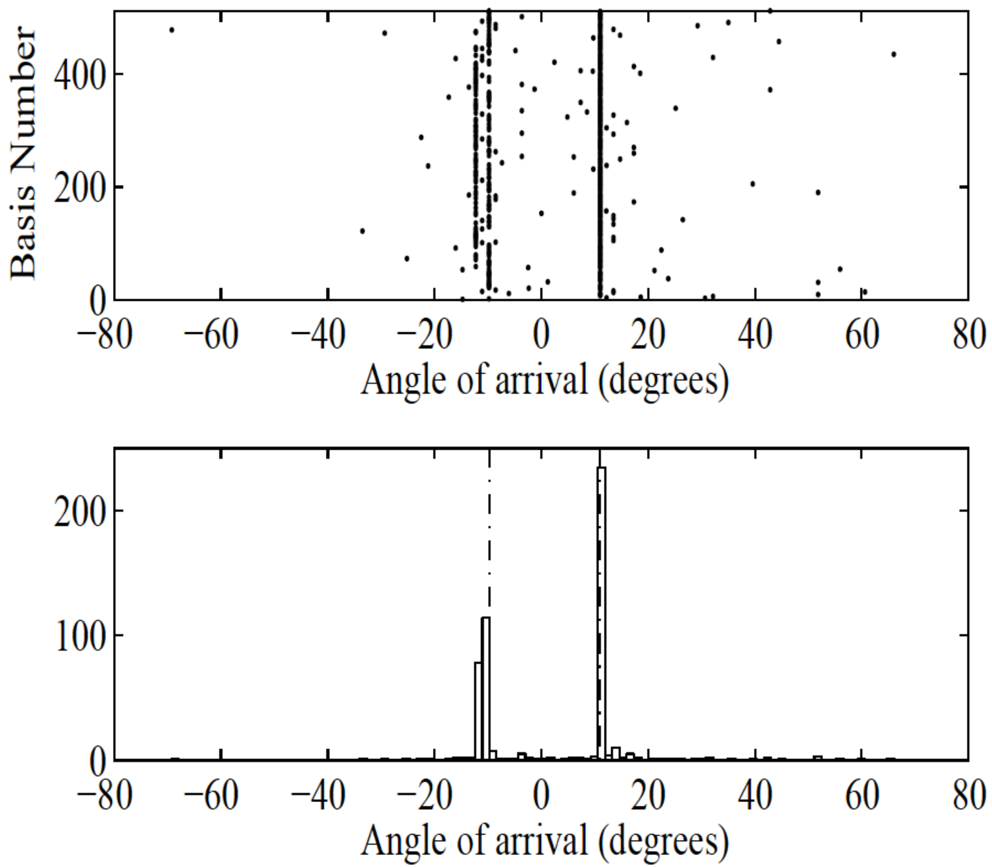
Measure left-right delay to find direction of arrival (DOA)



Queen Mary
University of London

centre for digital music

Cluster delays to find sources



Results

Mixtures

Freq Domain ICA
(Frame size=256)

Sparsified Basis
(Frame size=256)

Summary

- Interesting (and hard!) problem domain
- Seen some approaches to:
  - Music audio analysis:
    - Onset detection & beat tracking
    - Source separation
    - Music transcription
    - Visualisation
  - Source Separation
    - Single Channel
    - Multi channel

Many thanks to…

- Samer Abdallah
- Juan Bello
- Paul Brossier
- Chris Cannam
- Matthew Davies
- Mike Davies
- Maria Jafari
- Chris Harte
- Chris Landone
- Andrew Nesbit
- Mark Sandler
- Emmanuel Vincent
- Beiming Wang

and others…

Queen Mary
University of London

centre for digital music