Making Sense of Sound and Music

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

CREST Symposium on Human-Harmonized Information Technology
Kyoto, Japan 1 April 2012

Queen Mary
University of London

centre for digital music

Overview

- Separating sounds
- Extracting musical notes
- Following beats
- Visualisation and manipulation
- Non-speech non-music sounds

Queen Mary
University of London

centre for digital music

Many thanks to …

- Samer Abdallah
- Emmanouil Benetos
- Thomas Blumensath
- Chris Cannam
- Matthew Davies
- Mike Davies
- Simon Dixon
- George Fazekas
- Dimitrios Giannoulis
- Chris Harte
- Fabio Hedayioglu
- Jason Hockmann
- Anssi Klapuri
- Chris Landone
- Andrew Nesbit
- Marcus Pearce
- Josh Reiss
- Andrew Robertson
- Mark Sandler
- Adam Stark
- Dan Stowell
- Geraint Wiggins

and many others…

And thanks to funders:
EPSRC, Leverhulme Trust, EU Framework 7, TSB, ...

Queen Mary
University of London

centre for digital music

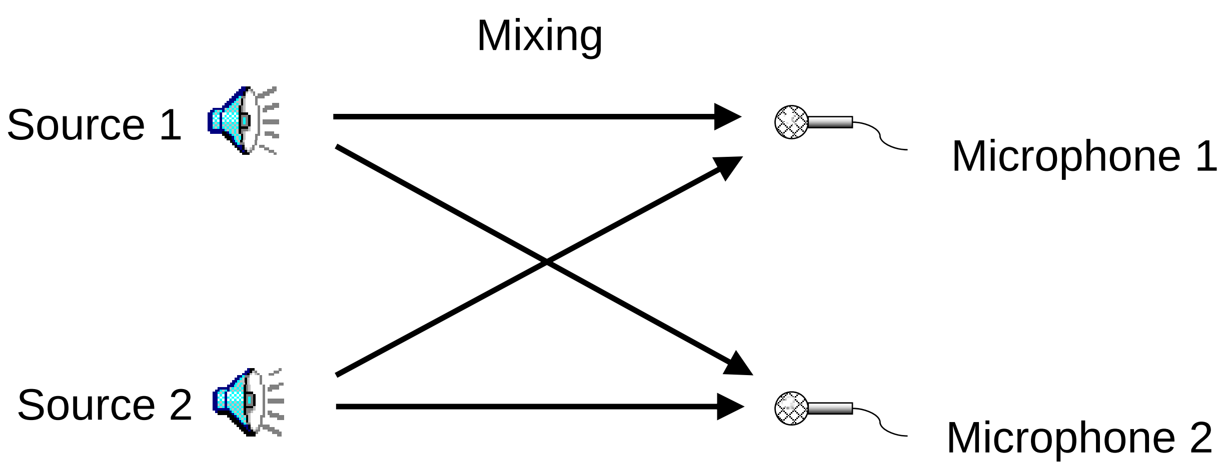
Separating Mixed Sounds

Queen Mary
University of London

centre for digital music

The “Cocktail Party” problem

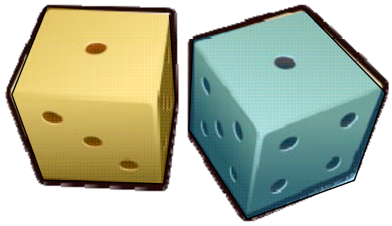
Simpler “Cocktail Party” problem



(In Maths: Microphones = Mixing x Sources **x** = **As**)

Problem: How can we “unmix” these,
if we only have the microphone signals?

Try something simpler



Let's try with dice instead of sounds

2 coloured dice, one Amber (A) and one Blue (B)

1. Secretly add some of A to some of B. Call this X
   Example: **X = ½ x A + 3 x B**
2. Do again with different amounts. Call this Y
   Example: **Y = 2 x A + 1 x B**
3. Roll the dice and write down the numbers X and Y
4. Give me the numbers.

Can I work out A and B, and how you mixed them?

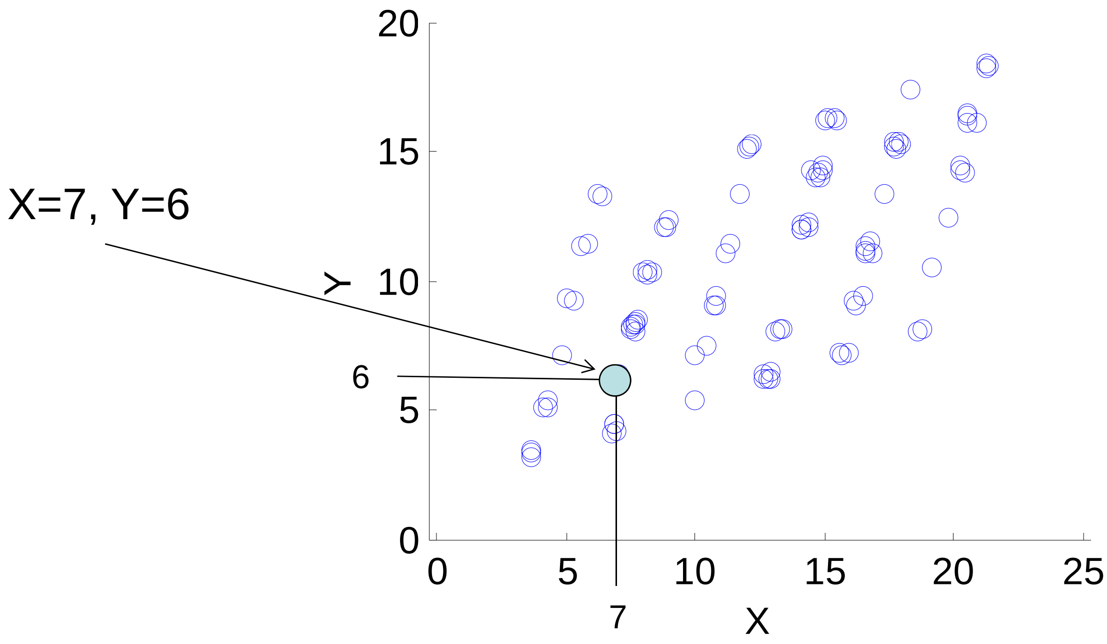
Mixed die rolls

You give me these:

X	Y
7	6
10.5	7
6	9
18.5	12
(etc...)	(etc...)

Let's plot them...

Scatter plot of the data points

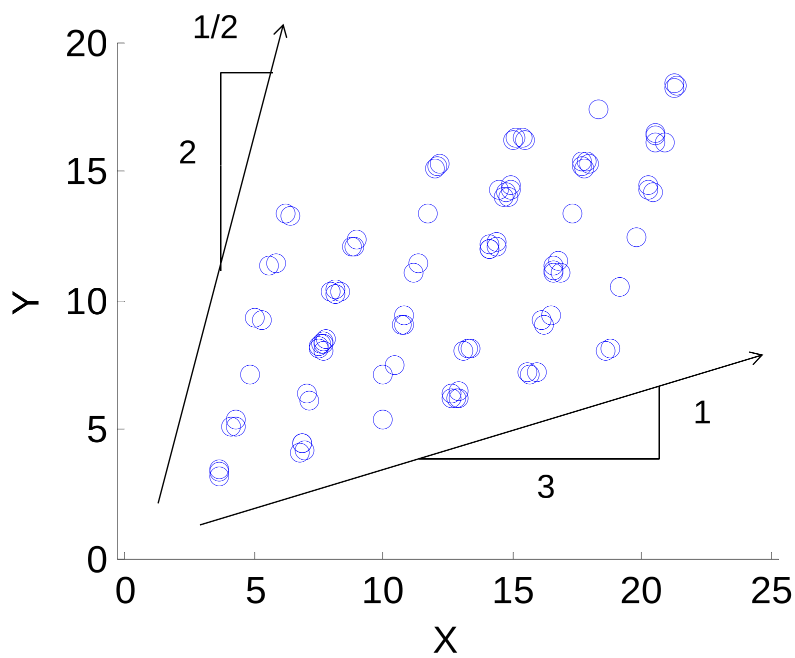


Plotted data points create a lozenge shape. Hmm...

Queen Mary
University of London

centre for digital music

Draw lines parallel with the shape



Queen Mary
University of London

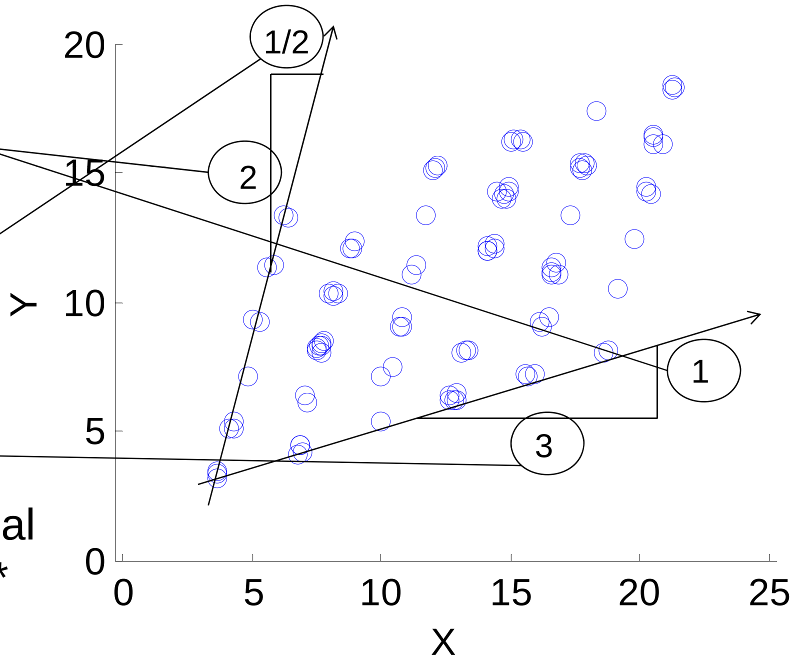
centre for digital music

Read off mixtures for A and B

Y = 2 x A + 1 x B

X = ½ x A + 3 x B

These are the original mixing equations! *

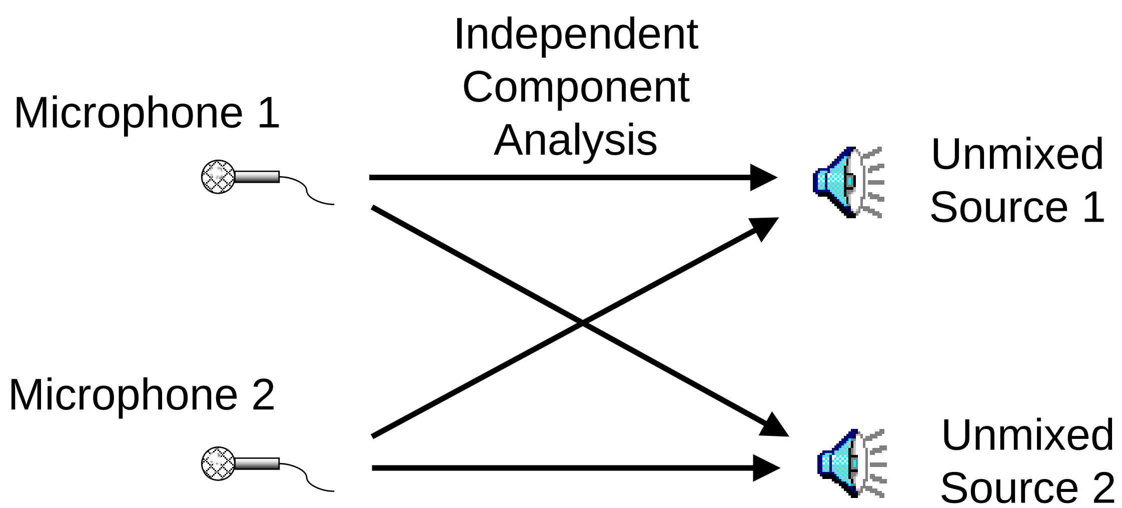


Then use School Math to solve for A and B. **Done!**

* We might have swapped A and B, or scaled them

Do the same for audio signals

Method called: “Independent Component Analysis” (ICA)



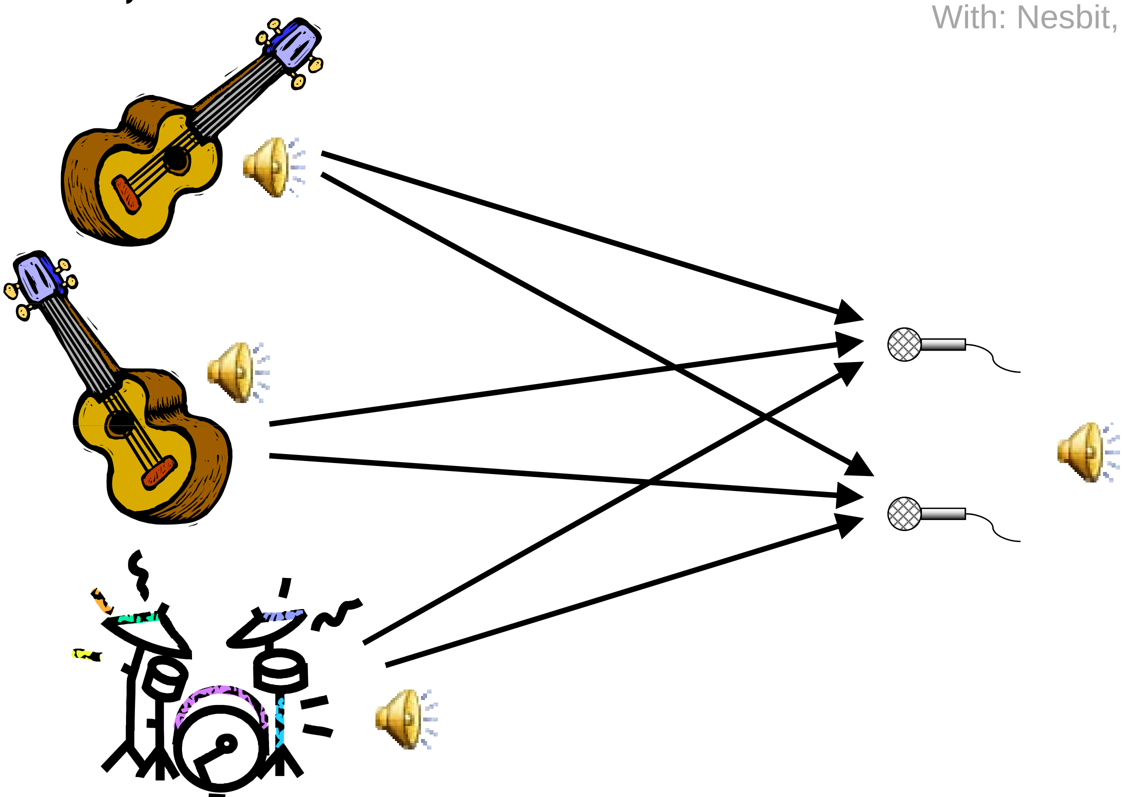
Separating More Mixed Sounds

Queen Mary
University of London

centre for digital music

Stereo, but more than two sources

With: Nesbit, Jafari



Queen Mary
University of London

centre for digital music

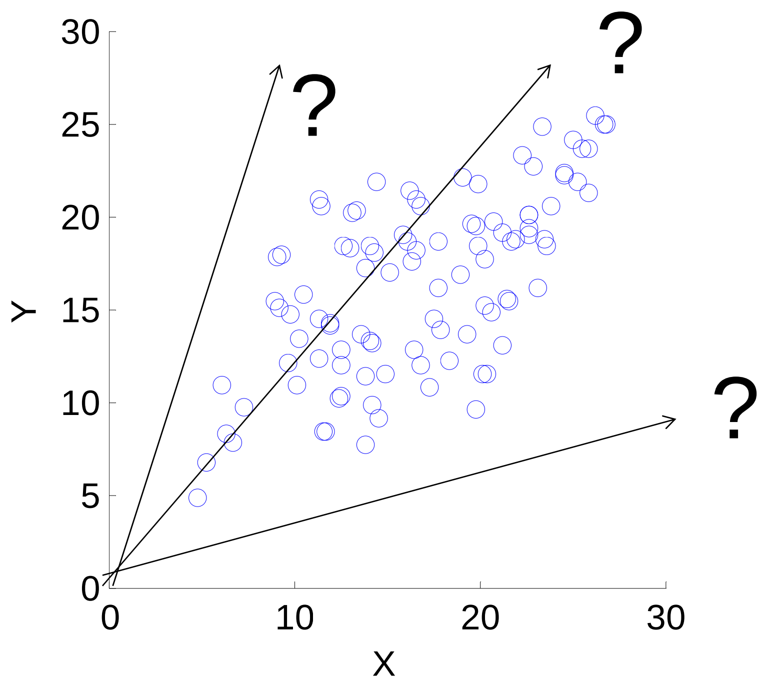
Simple scatter plot doesn't work

Try our dice again:
3 dice, 2 mixtures

We can't see three obvious lines!

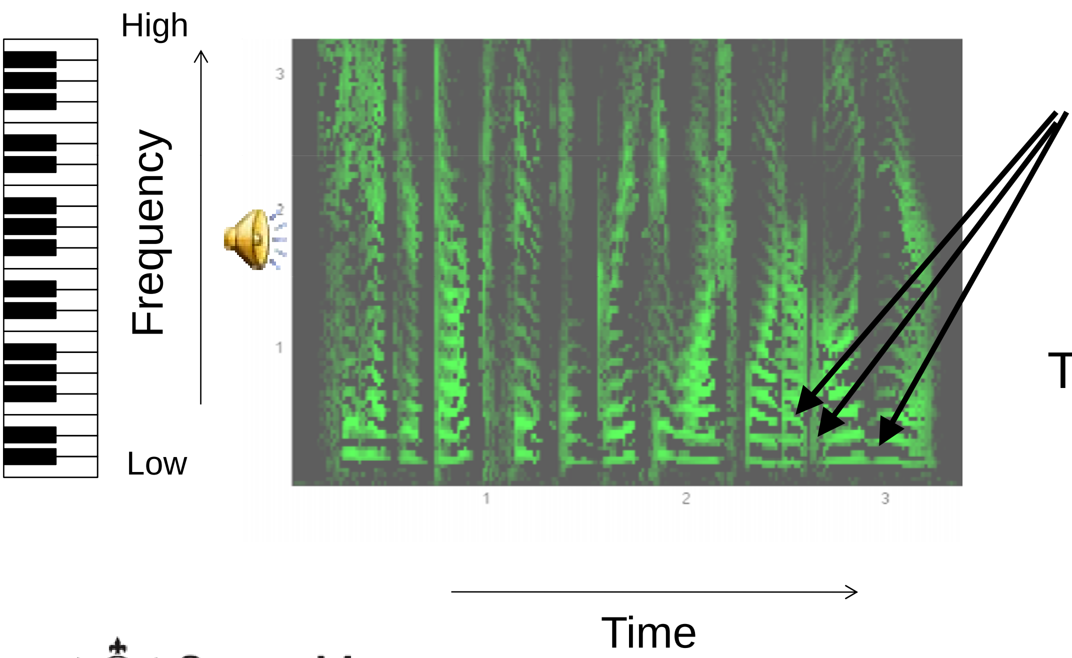
What can we do?

Change the game...



Sounds have changing frequencies

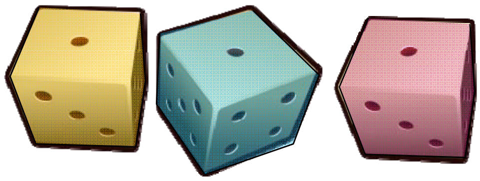
Let's look at how the frequencies change over time
(a "spectrogram")



Sounds only use **a few** frequencies

This is called **<u>"sparse"</u>**

Sparse dice example

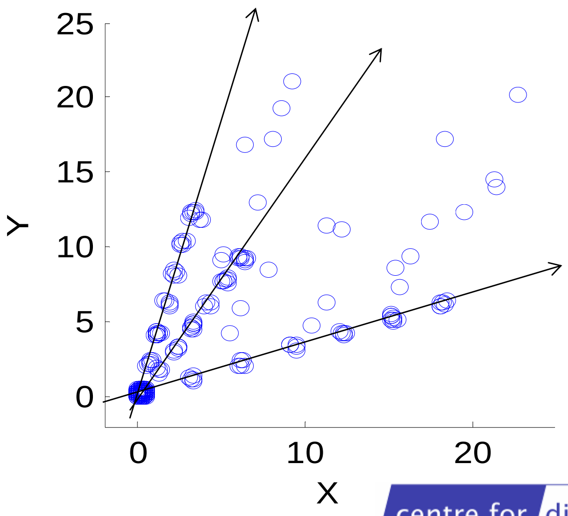


Imagine you have to throw a 6 before you can score **anything.**

4 -> 0
2 -> 0
4 -> 0
6, 4 -> 4
1 -> 0
(etc.)

Most scores are zero -> “**Sparse**”

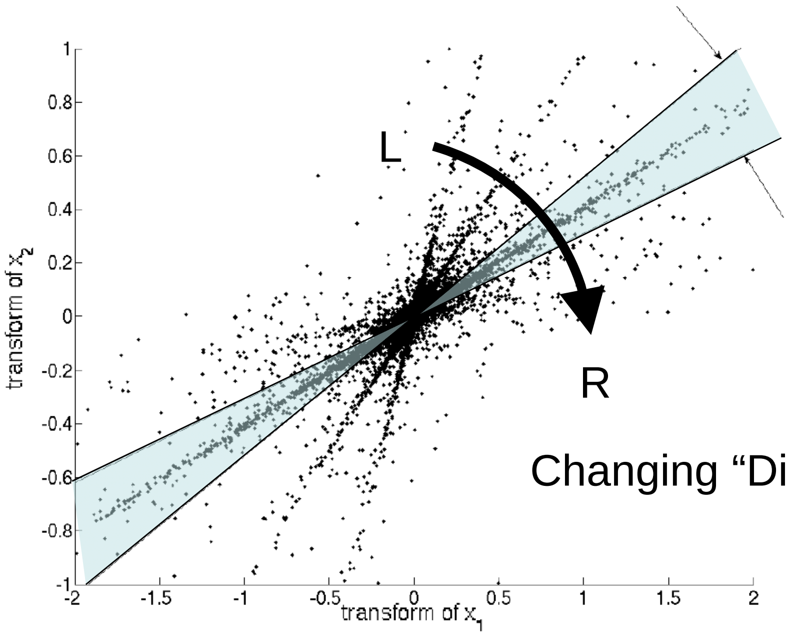
But if we mix them, we can now easily see the mixture lines!



Scatter plot of spectrogram

Make the scatter plot trick again,
but this time with the numbers from the spectrogram

Left microphone

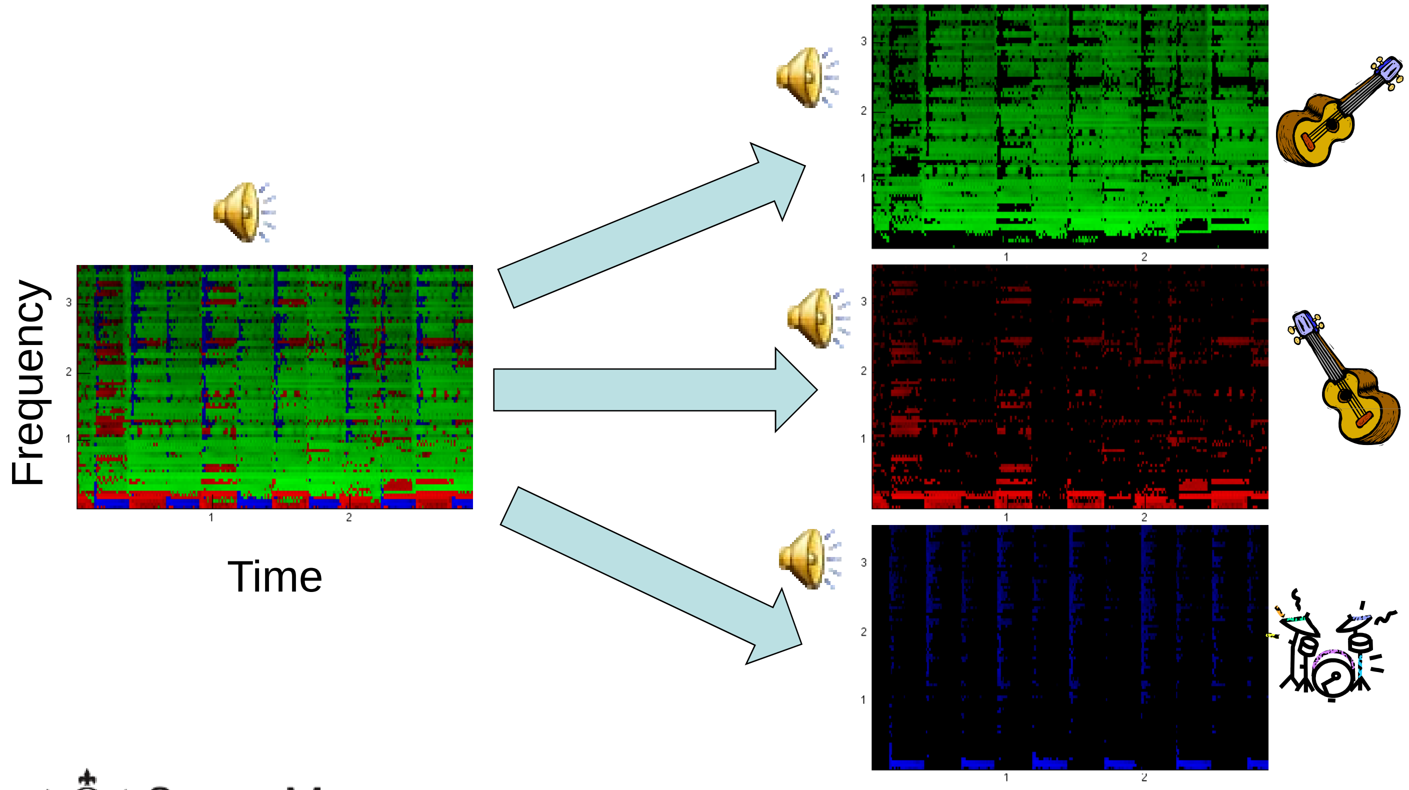


Colour all these patches the same colour

Changing “Direction of Arrival”

Right microphone

Colour-coded spectrogram

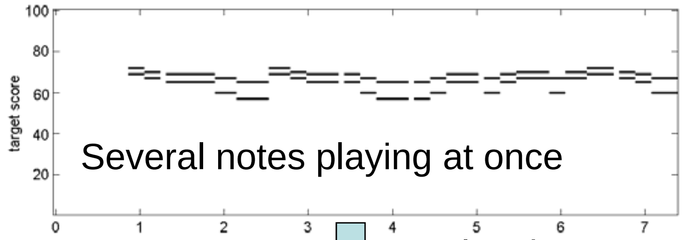


Extracting Musical Notes

Automatic Music Transcription

With: Abdallah

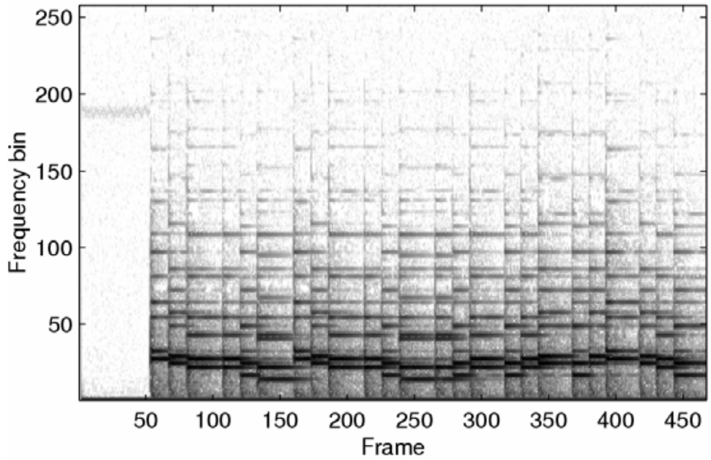
MIDI
("Piano Roll")



Play the notes

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

Problem:
Extract notes from this



Spectrogram

Queen Mary University of London

centre for digital music

How can we do this?

Musical notes are very sparse

- Out of 88 notes on a piano, only a few are played at once

So idea:

1. Search for ways to turn our spectrogram into something **even sparser**
2. Then we have found the notes (we hope!)

- We are looking for a "Sparse Representation"

Queen Mary University of London

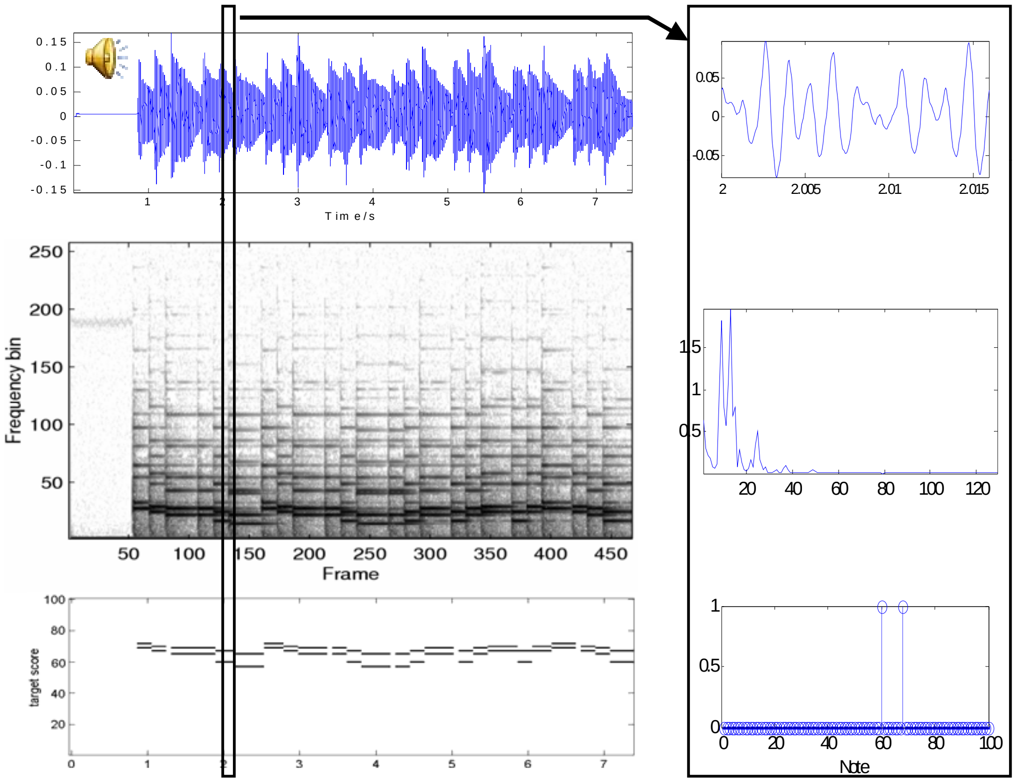
centre for digital music

Getting to "even sparser"

Audio signal:
Mostly non-zero.
Not sparse

Spectrogram:
Many small values
Fairly sparse

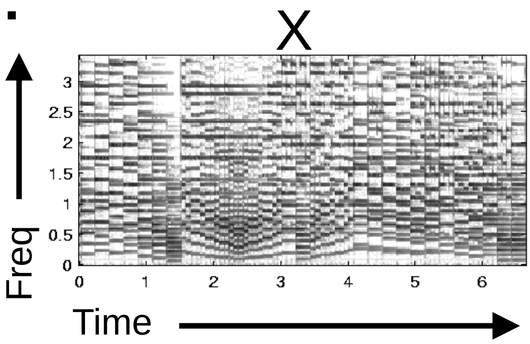
Notes:
Mostly zero
Very sparse



Example:

Observations (Spectrogram of Music)

Harpsichord music:
Bach Partita
in A Minor BWV827



Notes are **sparser** than spectrogram

Sparse Decomposition

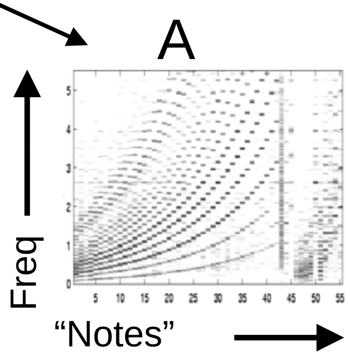
$X = A \times S$

Note frequencies

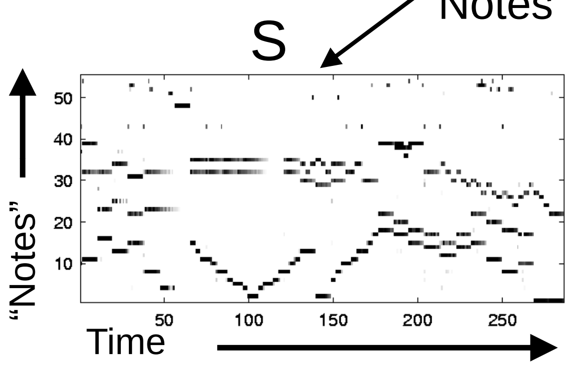
Notes

$=$

"Notes" are **discovered**



$\times$



Following Beats

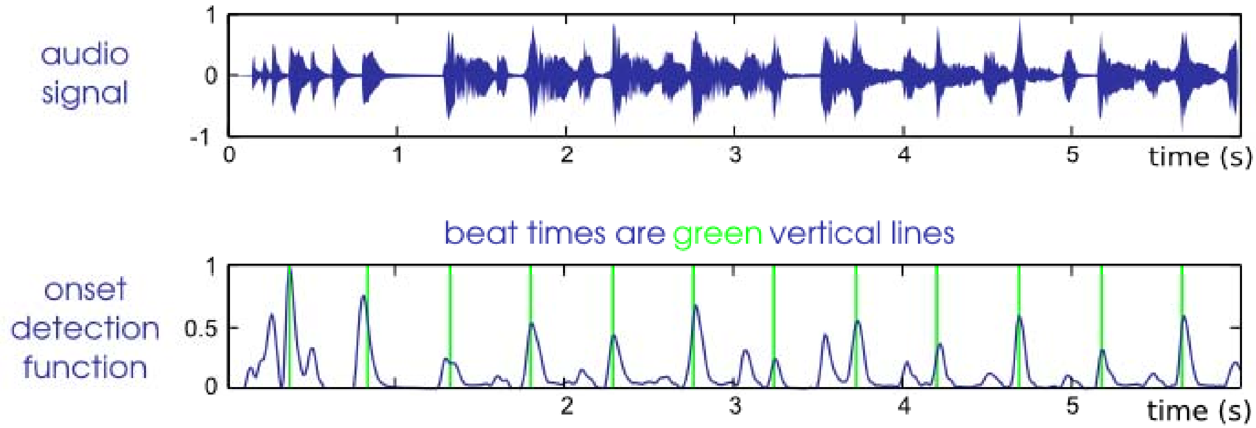
Queen Mary University of London

centre for digital music

Beat Tracking

With: MEP Davies

1. Measure how much the audio signal changes
   -> Biggest at note onsets ("Onset Detection Function")
2. Find regular pattern of peaks -> Beat Locations



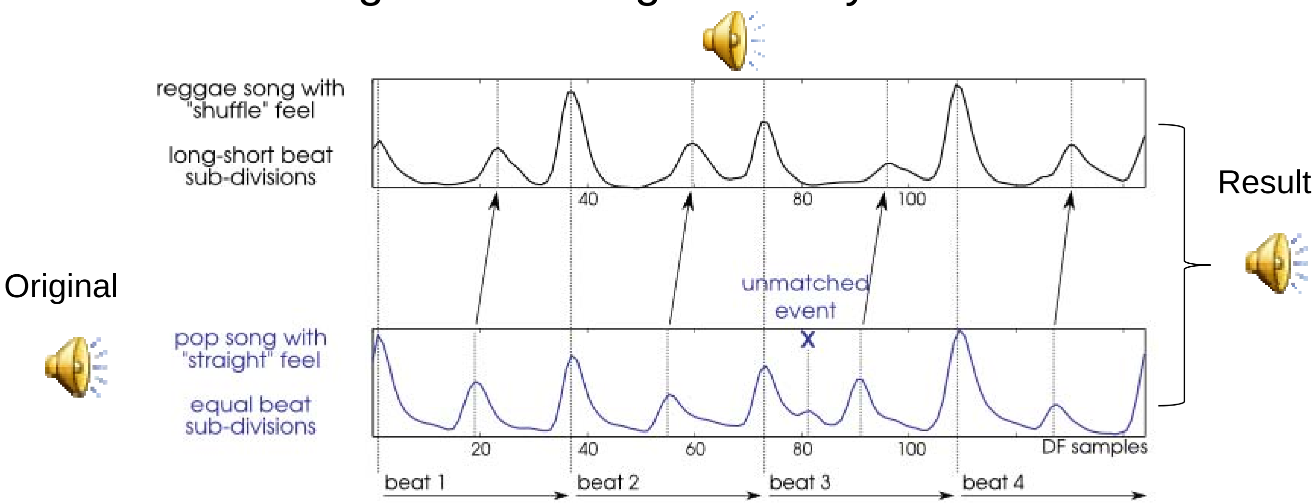
Now, the computer can "tap along to the music" ...

Queen Mary University of London

centre for digital music

Rhythm Transformation

- Extend Beat Tracking to Bar level: Rhythm Tracking
- Rhythm Tracking on model (top) and original (bottom)
- Time-scale segments of original to rhythm of model

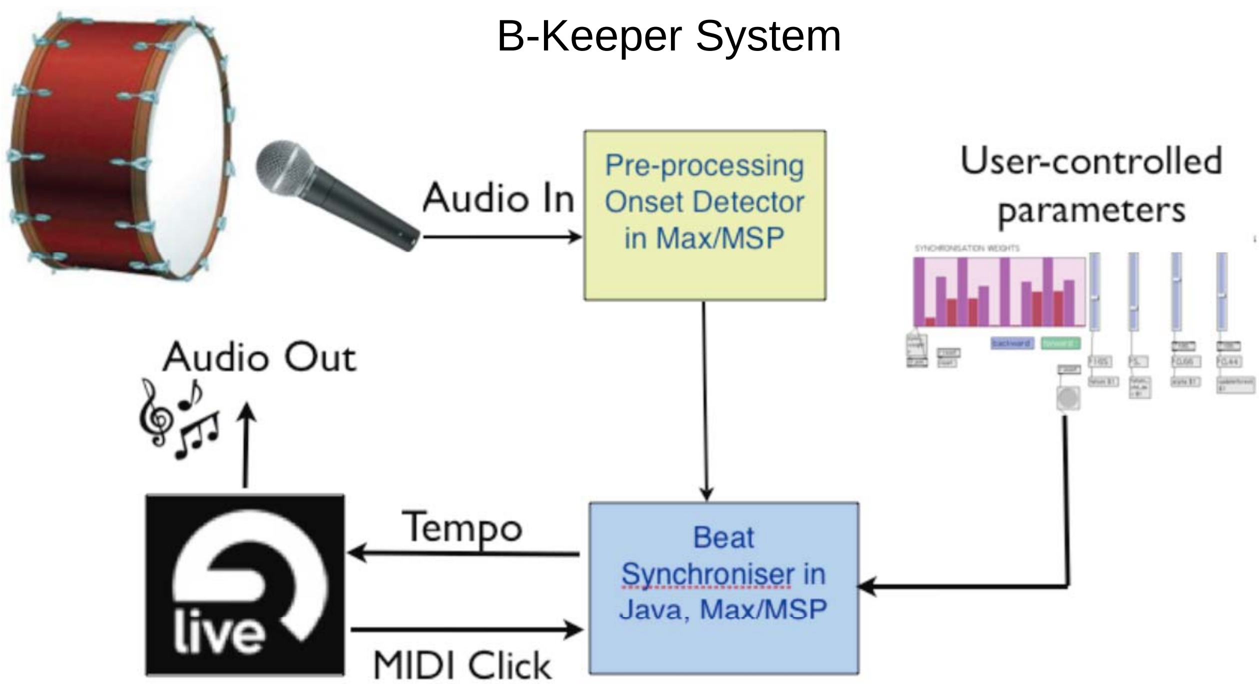


Queen Mary University of London

centre for digital music

Live beat tracking: accompaniment

With: Robertson



Queen Mary University of London

centre for digital music

B-Keeper video

- B-Keeper System: Andrew Robertson
- Drums: David Nock
- Glockenspiel: Dave Meckin

Queen Mary University of London

centre for digital music

[Video]

Music and Information

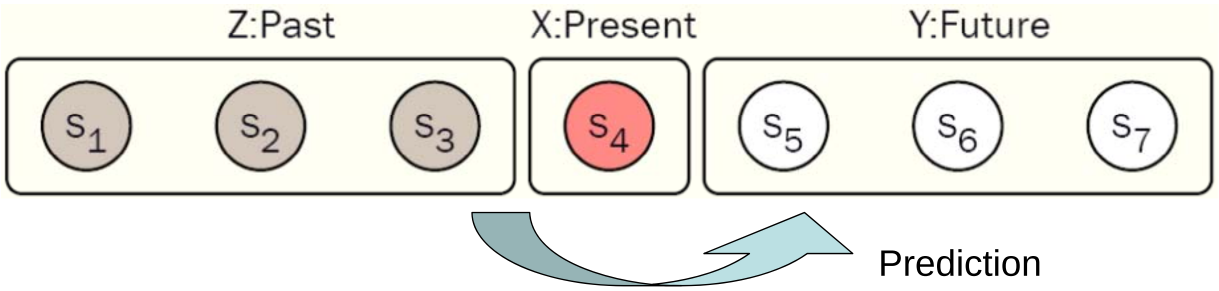
Queen Mary University of London

centre for digital music

Music and Information

With: Abdallah, Pearce, Wiggins, ...

- Idea: Listening to notes gives information* about music
- Notes are: "surprising" (high information)
  or "not surprising" (low information)
- Each note can tell us something new about the future
- Not "following" the music, but "predicting" the music!

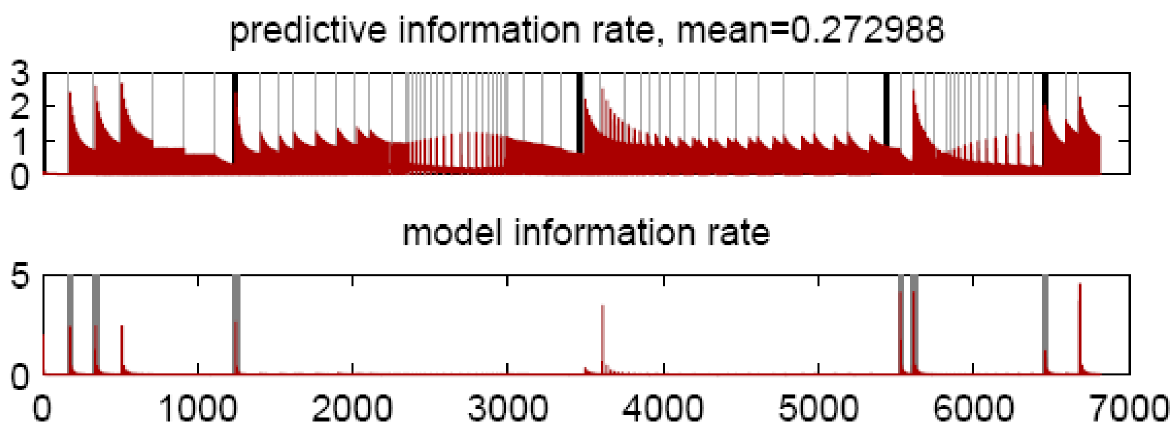


* - Information Theory: same as communications engineers use

Queen Mary
University of London
centre for digital music

Find the boundaries in music

- Measure how much the prediction has to change



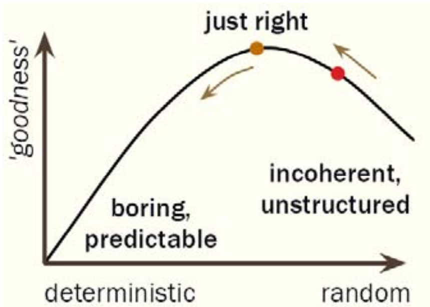
Philip Glass: Two Pages

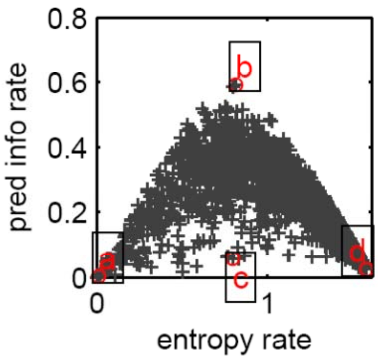
Queen Mary
University of London
centre for digital music

Help explain music people like?

- Too predictable from the start? -> boring ☹
- Can't make any sense of it? -> sounds random ☹
- Music unfolds bit by bit? -> just right ☺

"Wundt curve"





Also use this idea to build models of the music

Visualisation & Manipulation

Sonic Visualiser

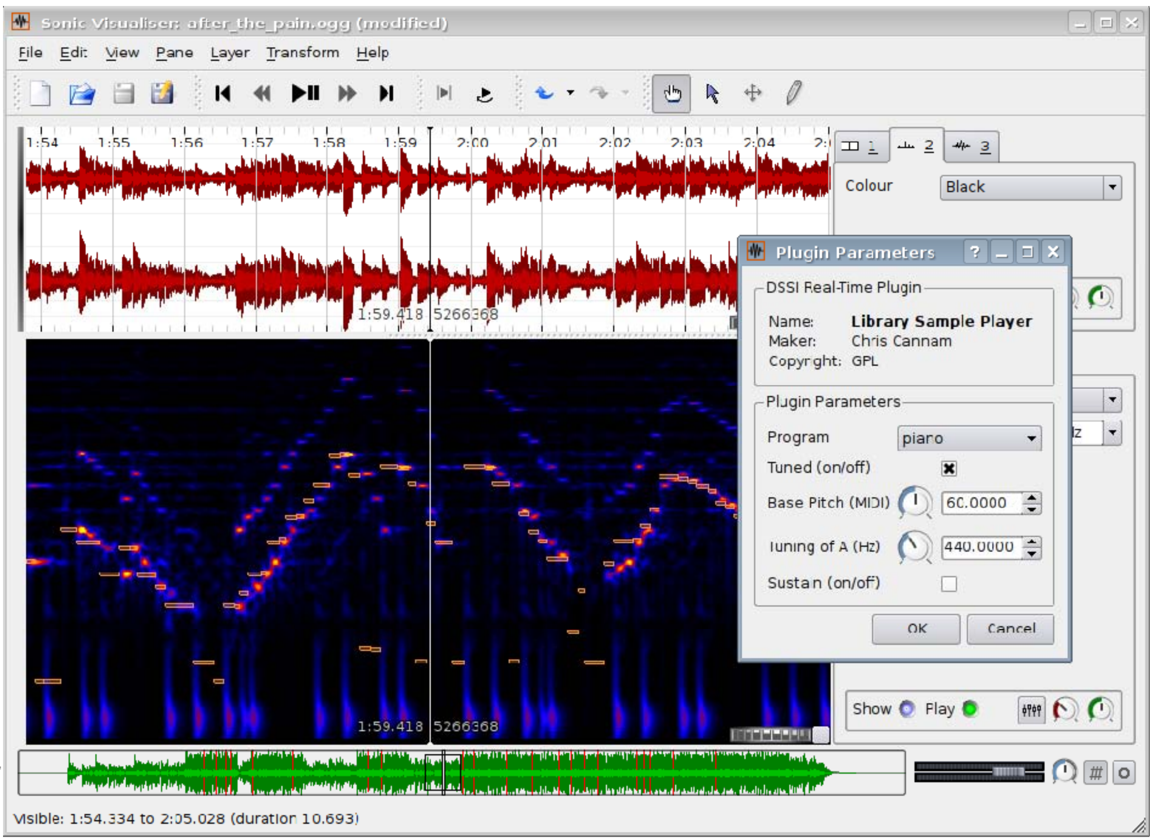
Cannam, Sandler, ...

- Visualise and edit content-derived metadata (low-level audio features and semantic descriptors)
- Open source
- VAMP plug-in API for creating new feature extractors
- Plug-ins for onset, beat, structural segmentation, key, transcription, etc
- Contribute/consume “Web 2.0” Linked Open Data
- Used by MIR researchers, musicologists, etc. (> 200,000 downloads)

Queen Mary University of London

centre for digital music

Sonic Visualiser

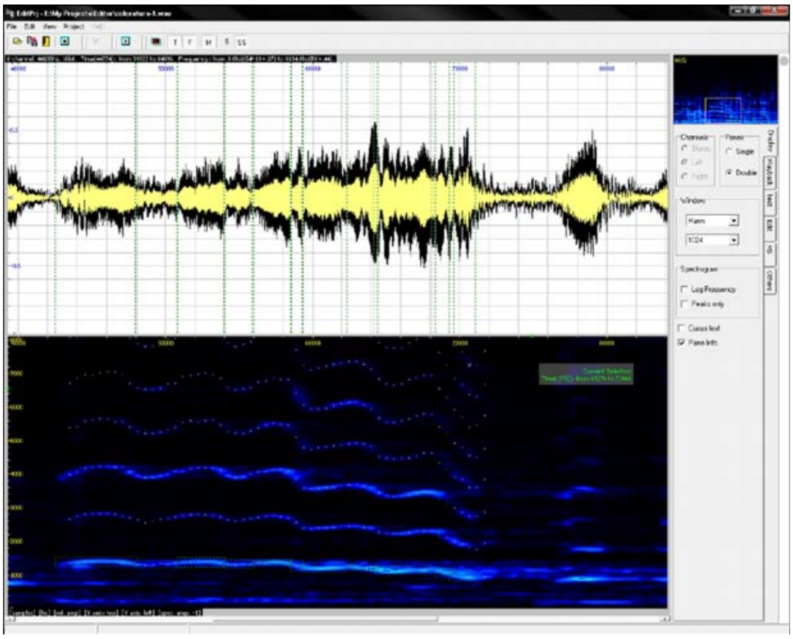


[Video]

Harmonic Visualiser

Wen, Sandler

- Signal modelled as quasiharmonic sinusoids plus residual
- Handles inharmonicity, captures complete note
- Models vibrato, enabling modification
- Musicologists: “what-if” analysis
- Studio: edit pitch and vibrato, remove notes, etc.



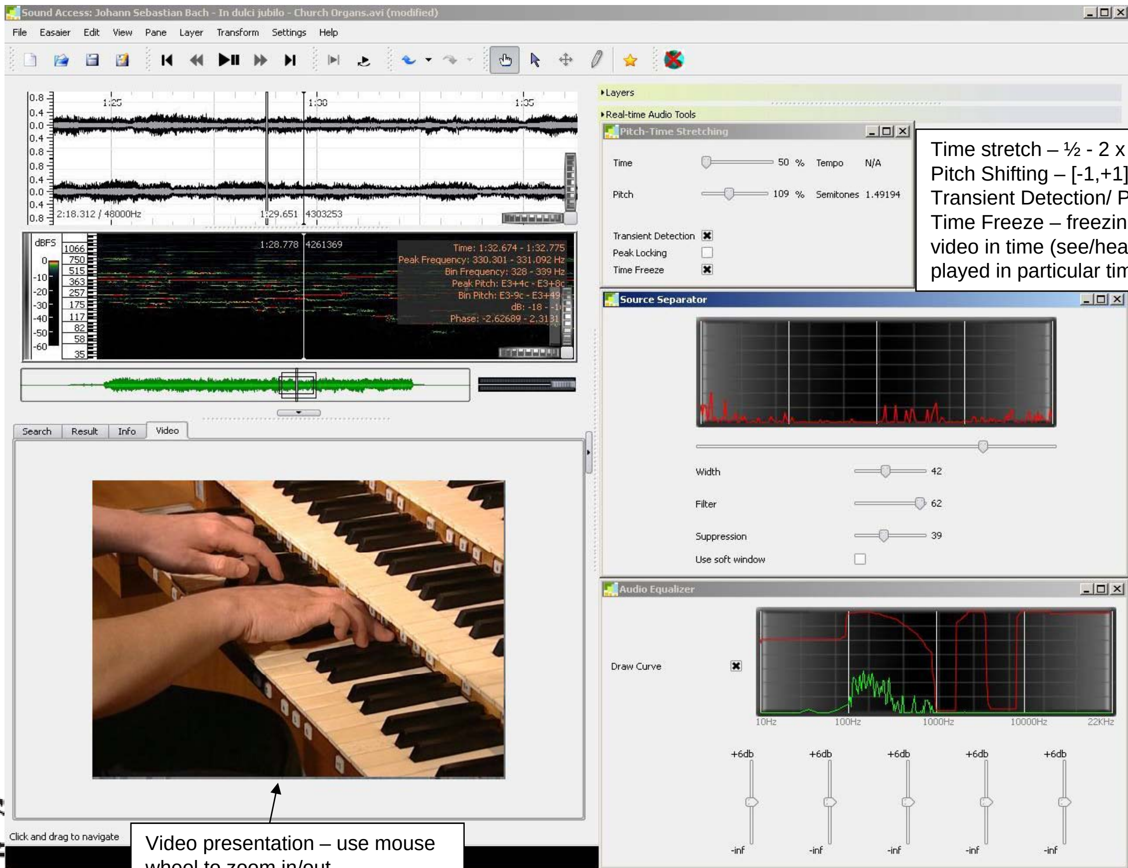
Queen Mary University of London

centre for digital music

[Video]

EASAIER: Audio-Visual Tool

Zhou, Reiss



Time stretch – ½ - 2 x
Pitch Shifting – [-1,+1] octave
Transient Detection/ Peak Locking
Time Freeze – freezing audio and video in time (see/hear chord played in particular time instance)

Video presentation – use mouse wheel to zoom in/out

[Video]

Non-speech Non-music sounds

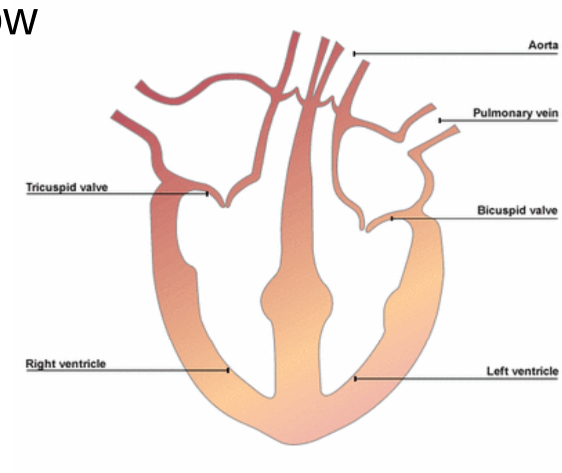
Separating Heart Sounds

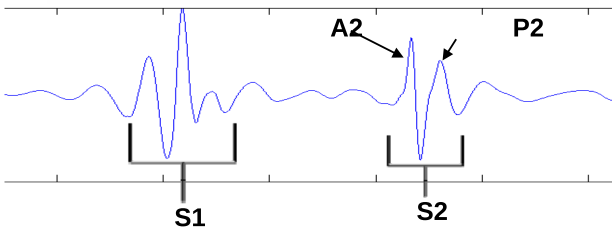
Queen Mary
University of London
centre for digital music

Medical Sounds: Heart Sounds

With: Hedayioglu, Jafari, Coimbra, Mattos

- Produced by valves and blood flow
- Important screening tool
- Difficult to hear
- May just hear “boomp - boomp”
- But the second “boomp” has two important parts:
  - A2 (aorta) – to body
  - P2 (pulmonary artery) – to lungs
- Can we separate them?



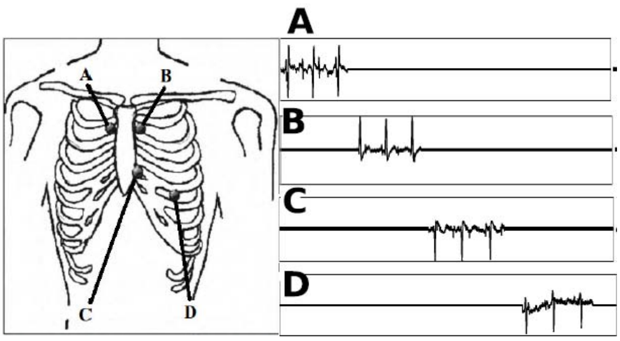


Queen Mary University of London

centre for digital music

Only one microphone (stethoscope)

- Clinician listens to 4 different places (“auscultation”)
- Each sound is a slightly different mixture

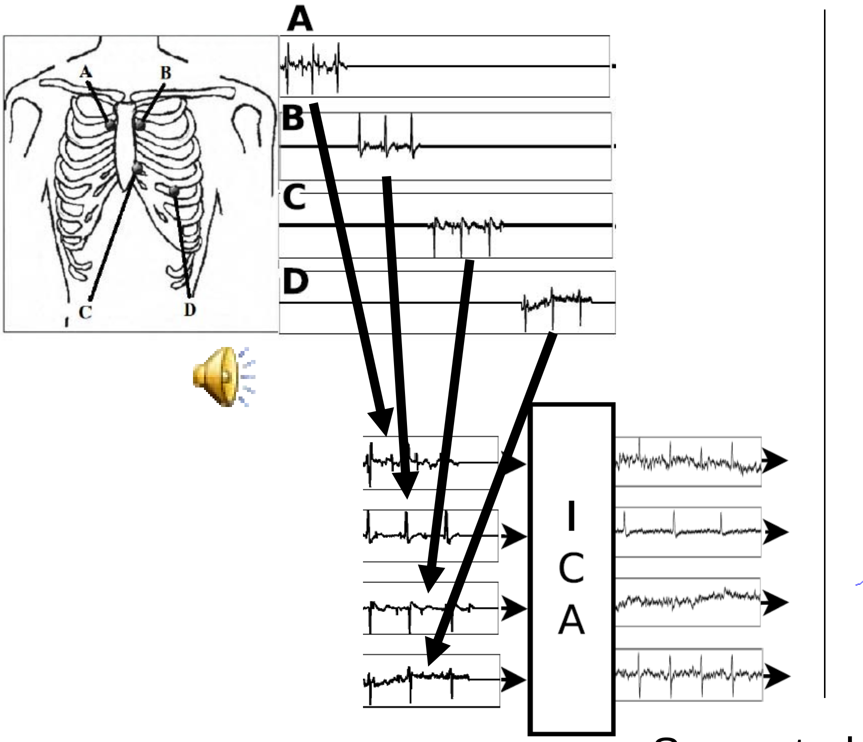


- The heart sound repeats, so:
  1. Line them up
  2. Use Independent Component Analysis (ICA)

Queen Mary University of London

centre for digital music

Line up and separate...

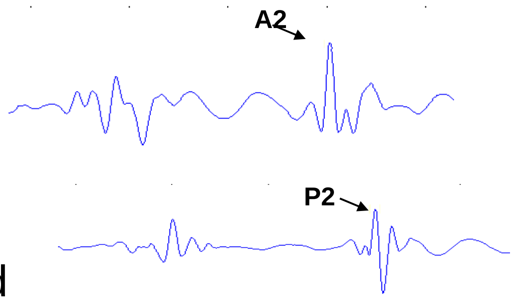


Separated sounds

Play both
with P2 delayed

Hear both A2 & P2

“boomp - ba-doomp”



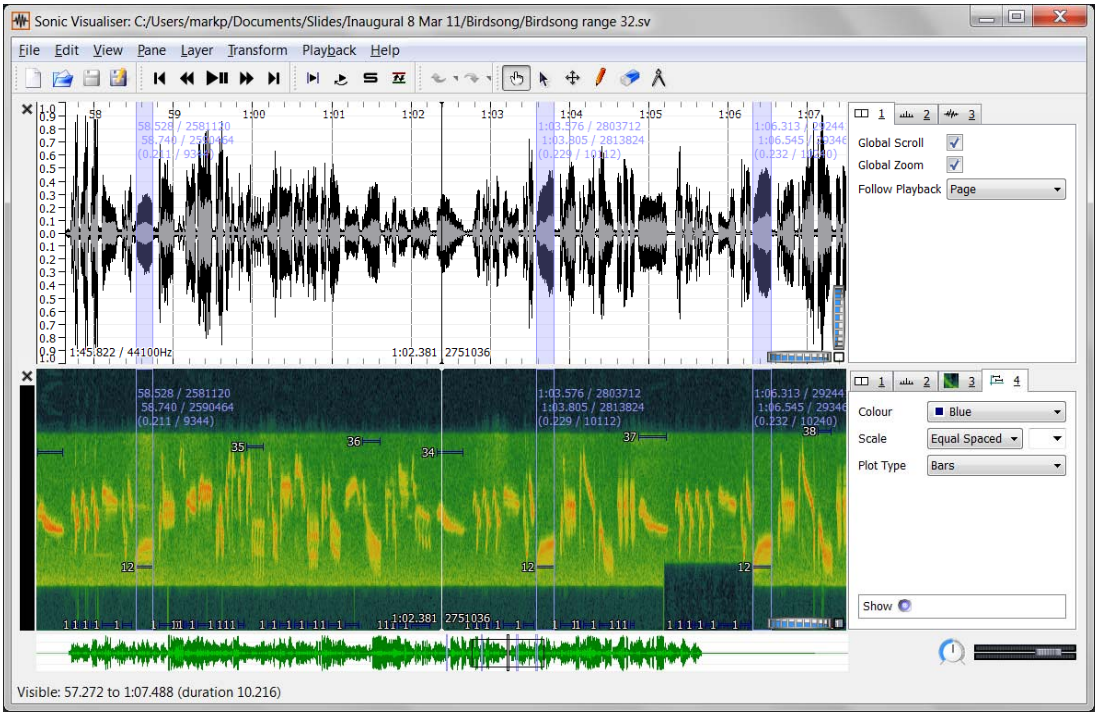
Queen Mary University of London
centre for digital music

Natural Sounds: Birdsong

Queen Mary University of London
centre for digital music

Birdsong Segmentation & Clustering

With: Stowell, Briefer, McElligott



Queen Mary University of London

centre for digital music

[Skylark demo]

Conclusions

- Separating sounds, extracting notes, beats, ...
- Visualization and manipulation

Where next?

- Digital Media everywhere
  - Personal devices, social networking, audio and video
- More interaction with music – not just passive consumers
- “Non-speech non-music audio”
  - Medical sounds, Environmental sounds, Urban sounds

Queen Mary University of London

centre for digital music