

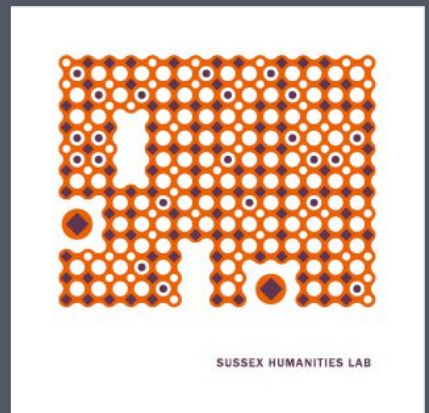
# Music Information Retrieval Algorithms for Oral History Collections

Sharon Webb, Chris Kiefer, Ben Jackson,  
Alice Eldridge, James Baker,

US

University of Sussex

Sussex Humanities Lab

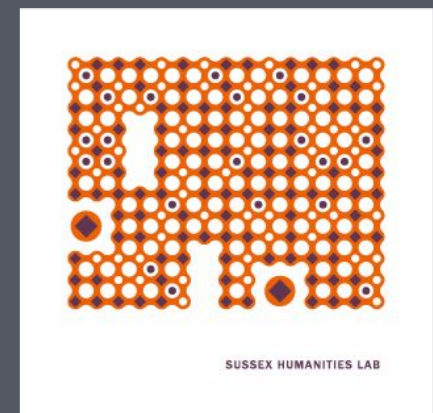


# Introductions

We are part of a 32 strong team of researchers and lecturers at the **Sussex Humanities Lab**, which is a £3 million investment by the **University of Sussex** to establish a Digital Humanities research programme.



To find out more about us and our programme (which includes a visiting fellow scheme) pop along to our poster session on Wednesday or follow us on Twitter @SussexHumsLab and/or Facebook



# Why are we here?

1. Digital humanities typically treats audio files as texts, retrieving semantic information in order to categorise, sort, and discover audio.
2. Text encodings or transcripts of oral history interviews have their obvious advantages, however approaches that privilege text should be questioned given the ever increasing availability of computation.
3. Oral History and DH could both benefit from digital tools for audio analysis - to treat audio as audio.

# What is MIR?

- Music Information Retrieval algorithms enable the abstraction of high-level musical information from low-level audio data - ‘machine listening’.
- Algorithms are based on digital audio signal processing (DSP), pattern matching and machine learning methods, informed by auditory psychology and music theory.
- We are interested in exploring how these can be adapted into non-musical, auditory contexts such as oral history archives.

# What do we get from today?

- **Outcomes for you**
  - Basic understanding of digital audio
  - Insights into what information MIR tools can provide
  - Inspiration as to how these could be used in humanities research
  - Motivation to explore further
- **Outcomes for us**
  - Gauge appetite for MIR
  - Rich insight into oral history modes of enquiry
  - Build use cases for MIR
  - Feed into ongoing project development

# Technical Specification

- The workshop is designed for participants with basic coding skills
- We use Python and the Jupyter notebook environment which embeds code, graphics and sound in a single interactive document (but you don't need to know how to use any of this right now to enjoy the workshop!)
- The course runs in a Linux virtual machine, inside VirtualBox, so only minimal setup is required

# Schedule

<b>0930 - 1000</b>	Intro
<b>1000 - 1200</b>	0. Digital Audio & MIR 101 1. Jupyter notebooks - intro 2. Digital audio - representation and visualising 3. Audio Features
<b>1200 - 1300</b>	Lunch
<b>1300 - 1500</b>	4. Feature extraction 5. Classifiers
<b>1500 - 1600</b>	Q&A

*Note that depending on the pace at which we work, we are happy to adapt and change the schedule to suit!*