# Mining oral history collections using music information retrieval methods

Article (Accepted Version)

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307

## Abstract

Recent work at the Sussex Humanities Lab, a digital humanities research program at the University of Sussex, has sought to address an identified gap in the provision and use of audio feature analysis for spoken word collections. Traditionally, oral history methodologies and practices have placed emphasis on working with transcribed textual surrogates, rather than the digital audio files created during the interview process. This provides a pragmatic access to the basic semantic content, but obviates access to other potentially meaningful aural information; our work addresses the potential for methods to explore this extra-semantic information, by working with the audio directly. Audio analysis tools, such as those developed within the established field of Music Information Retrieval (MIR), provide this opportunity. This paper describes the application of audio analysis techniques and methods to spoken word collections. We demonstrate an approach using freely available audio and data analysis tools, which have been explored and evaluated in two workshops. We hope to inspire new forms of content analysis which complement semantic analysis with investigation into the more nuanced properties carried in audio signals.

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



# Mining oral history collections using information retrieval methods

**Webb, S., Kiefer, C., Jackson, B., Baker, J. & Eldridge, A.**

Music Reference Services Quarterly (Taylor and Francis)

## 1 Introduction

The Sussex Humanities Lab is a multidisciplinary research program tasked with embedding digital humanities into research and teaching practices across the University of Sussex. As a multidisciplinary team we have unique access to varied expertise and skills that enable us to carry out experimental work in an agile and proficient manner. One experimental project problematized the predominant approach within digital humanities – a largely text based domain – to treat digital audio files as text.[1] We applied Music Information Retrieval (hereafter MIR) techniques to oral history interviews in order to develop new, complementary, approaches to text based methods of extracting semantic information from spoken word collections. As an established field, with established methods, the MIR community provides open source tools, code and libraries to work through our hypothesis, to treat audio as audio, and to help us work through and establish its practical application to spoken word collections. Having established the potential utility of MIR techniques to problems in both oral history and the digital humanities, we developed a workshop framework that aimed at exploring the utility of this approach for a variety of humanities scholars.

[1] Notable exceptions include, Tanya Clement and Stephen McLaughlin, "Measured Applause: Toward a Cultural Analysis of Audio Collections," *Cultural Analytics* 1, no .1 (2016), http://culturalanalytics.org/2016/05/measured-applause-toward-a-cultural-analysis-of-audio-collections/; Tanya Clement, Kari Kraus, Jentery Sayers and Whitney Trettien, "<audio>Digital Humanities</audio>: The Intersections of Sound and Method," *Proceedings of Digital Humanities 2014. Lausanne, Switzerland*, https://terpconnect.umd.edu/~oard/pdf/dh14.pdf

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



Taking oral history collections from the University of Sussex 'Archive of Resistance' as a test case, we led two distinct groups, at two separate workshops, through the process of using MIR approaches to categorize, sort and discover audio collections. This process enabled us to:

- Build a set of python workbooks that provide a conceptual and practical introduction to the application of MIR techniques (e.g. feature extraction and clustering) to spoken word collections.
- Work through, develop and amend use cases.
- Learn lessons, from two distinct communities and perspectives, about the potential benefits – or otherwise – of our approach.

Both workshops, the first at Digital Humanities 2016 (Krakow, July 2016) and the second at London College of Communication (March 2017),[2] provided points of clarification and discussion that enabled us to identify areas that require work. This article is therefore not a final report on our findings, instead it is an attempt to capture the hypothesis and problem statement, the experimentation and methodology used, and our preliminary findings. It also describes a method for workshop facilitation that utilizes a) virtual environments to reduce setup time for participants and facilitators and b) Jupyter Notebooks to enable participants to run sophisticated and complex code in a supported, learning environment.[3]

This article proceeds in five parts. First, in order to provide some context to this work, we provide some background information on the Sussex Humanities Lab. Parts two, three, and four

[2] 'Data-Mining the Audio of Oral History: A Workshop in Music Information Retrieval' at London College of Communication (March 2017) https://web.archive.org/web/20171003144121/http://www.techne.ac.uk/for-students/techne-events/apr-2015/data-mining-the-audio-of-oral-history-a-workshop-in-music-information-retrieval (accessed 3 Oct. 2017)

[3] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, Carol Willing, Jupyter Development Team, "Jupyter Notebooks – a publishing format for reproducible computational workflows," *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (2016). 87-90. doi: 10.3233/978-1-61499-649-1-87

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



consider our hypothesis and motivations, the workshops we developed and the technologies we used. The fifth and final part outlines our preliminary findings, both from mining oral history collections using audio feature analysis and from delivering workshops on MIR in a digital humanities context.

## 1.2 Background

The authors are current or former members of the Sussex Humanities Lab (hereafter SHL): a four-year university program, launched in 2015 at the University of Sussex, which seeks to intervene in the digital humanities. It is a team of 31 faculty, researchers, PhD students and technical and management staff, working in a state of the art space – the Digital Humanities Lab. SHL collaborates with a network of associates across and beyond the university nationally and internationally and is radically cross-disciplinary in its approach. The aim of SHL is to engage with the myriad of new and developing technologies to explore the benefits these offer to humanities research and to ask what will technology do to the arts and humanities? To achieve this, SHL is divided into four named strands of activity: digital history and digital archiving; digital media and computational culture; digital technologies and digital performance; digital lives and digital memory. However, the intention is to make sure that our research crosses and links these strands, to develop fruitful methodological and conceptual intersections. The work described here grows from this multidisciplinary ethos, since the project combines the diverse interests and expertise of the authors. It stems from the inherently collaborative environment facilitated by SHL and is influenced by two strands in particular: digital history and digital archiving, and digital technologies and digital performance.

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



## 2 Hypothesis: problem statement and motivation

Oral history best practice publications and resources often focus on the application and use of digital methods and tools to create, store and manage audio, audio-visual, and subsequent text files. They recommend, for example, standards for file formats, metadata and text encoding, software for audio to text conversion, and database and content management systems. And whilst the privileged position of text has been challenged,[4] the majority of oral history projects still rely on the creation of transcripts to carry out analysis using digital tools and methods. This focus on textual surrogates rather than audio sources denies – according to Alessandro Portellii – the 'orality of the oral source'.[5] It also denies – or at least underplays – the inherently *interpretative* nature of transcription.

Of course, textual encodings or transcripts of oral history interviews do have advantages: they are easier to anonymize, distribute, store and retrieve than digital audio files, and there are established techniques for analyzing them as text and/or data. But as a consequence of this privileging of the "text", a significant proportion of oral history collections and the tools provided to navigate and analyze them do not support navigation or analysis of the digital audio files captured during interviews. Instead, they focus on how to record oral history interviews, the management of digital files and the creation of transcripts using both semi-automated audio to text tools and manual transcription.[6]

---

[4] For example, Doug Boyd, "OHMS: Enhancing Access to Oral History for Free," *Oral History Review*, Winter-Spring, 40 no.1 (2013). doi:10.1093/ohr/oht031

[5] Ronald Grele, "Oral History as Evidence," in *History of Oral History: Foundations and Methodology*. Edited by Thomas L. Carlton, T.L., Lois E. Myers, L.E., & Sharpless, R. (UK, 2007), 69.

[6] For example, 'Oral History in a Digital Age' http://ohda.matrix.msu.edu/

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



While using text surrogates is an established tradition, the oral history community is beginning to question the privilege of this text based approach. This is evident, for example, in the UK's Oral History Society's 2016 conference call for papers, which stated the 'auditory dimension of oral history [has] for decades [been] notoriously underused'.[7] While this move is welcome, it is true that, as per Clement et al's 2014 survey of the field, currently 'there are few means for humanists interested in accessing and analyzing [spoken] word audio collections to use and to understand how to use advanced technologies for analyzing sound'.[8] Moreover, these technologies have the potential to help resolve some of the backlog in archives and libraries of 'un-described, though digitized, audio collections'.[9] It is from this context, therefore, that we decided to explore the potential for direct audio analysis of oral history interviews.

This work represents a move towards analyzing oral content in the context in which they were created. It also challenges the privilege of text, as it focuses on extracting information from the audio signal directly. We are particularly interested in how such techniques could complement the semantic content obtained through manual or automated transcription. On the basis that comparable methods have been developed for digital recordings of music for some time, we explored the field of MIR for possible solutions. We are explicitly carrying out a study of computational techniques for the analysis of oral history records with the aim of extracting quantitative results to assist research. The MIR techniques that we use create quantitative

---

[7] "Beyond Text in the Digital Age? Oral history, images and the Written Word" Oral History Society, 2016 Conference CFP: https://web.archive.org/web/20161214045140/http://www.ohs.org.uk/conferences/2016-conference-beyond-text-in-the-digital-age/ accessed 27th June 2017

[8] Tanya E. Clement, David Tcheng, Loretta Auvil and Tony Borries, "High Performance Sound Technologies for Access and Scholarship (HiPSTAS) in the Digital Humanities," *Proceedings of the Association for Information Science and Technology* 51 (2014) 1–10 doi:10.1002/meet.2014.14505101042

[9] Ibid

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



information (i.e. timestamps that locate specific features and/or events within the audio) that could enhance and stimulate new directions in the qualitative research of others.

MIR draws from digital audio signal processing, pattern recognition, psychology of perception, software system design, and machine learning to develop algorithms that enable computers to 'listen' to and abstract high-level, musically meaningful information from low-level audio signals. Just as human listeners can recognize pitch, tempo, chords, genre, song structure, etc., MIR algorithms – to a greater and lesser degree – are capable of recognizing and extracting similar information, enabling systems to perform extensive sorting, searching, music recommendation, metadata generation, transcription on vast data sets. Deployed initially in musicology research and more recently for automatic recommender systems, the research potential for MIR tools in non-musical audio data mining is being recognized but yet to be fully explored in the humanities.

We chose to develop a one-day workshop related to this topic because the approach allowed us to explore our hypothesis and methods on different users, both expert and novice, from different disciplines, digital humanities and oral history, and garner important, domain specific feedback.

## 3 Experimentation: workshops and method

The workshops were intentionally experimental in nature (especially from a content analysis perspective), but were developed and delivered with a number of use cases in mind. We framed these use cases around three distinct contexts: the digital object, the content of the interview and the environment in which the interviews were carried out. Upon completion of the

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



workshops we revisited these use cases. The following questions represent a series of potential applications for the use of MIR in the context of analyzing oral history collections. They are based on a synthesis of both our initial scoping work and our interactions with workshop attendees. A known problem within oral history and digital humanities is the time and resources intensive process of cataloguing and analysis oral history collections. Therefore, although for practical reasons small collections were used in the workshops, the use cases developed and methods adopted are fully scalable:

1. Context - the *object*:
    1.1. What technical metadata or technical information can we automatically extract from a digital audio file?
    1.2. Can this new information enhance what we know about an object and improve search and discoverability?
    1.3. Can we detect the use of different recording devices as a means of clustering and classifying two temporally distinct data sets?
2. Context - the *content* (i.e. what type of content analysis can we carry out):
    2.1. What descriptive metadata can we automatically extract from the digital audio file? For example, can we create a feature which distinguishes interviewer from interviewee? Could we use this to automatically detect a specific voice within a collection?
    2.2. Can we reveal anything about the relationship or dynamic between interviewer and interviewee? For example, can we detect overlaps or interruptions by the interviewer? Can this reveal anything about gender roles and/or behaviors?

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



2.3. Can we augment our ability to detect emotion by analyzing changes in rhythm, timbre, tone, tempo? Is it therefore possible to identify song, poetry, speech, crying, laughter, etc.?

2.4. Can we automatically cluster acoustically similar audio/material/objects? For which properties might this be most robust?

2.5. Can we use techniques from musical analysis to reveal structure in spoken audio, for example to pull apart different voices, and how might this be useful for oral history collections?

3. Context- the *environment:*

3.1. Can we detect any environmental features in the audio stream? What might this tell us about where the interview took place.

3.2. Can we use source separation, developed to separate parts (e.g. drums, vocals, keyboards in pop music), to pull apart intertwined 'voices' or 'noises'. Can we use this to remove background noise that provides context to recordings? How might this affect the analysis of interviews?

Enabling these kind of preprocessing and descriptive orientated steps affords new possibilities in oral history research and archival management. For example, these enable access to under described repositories such as the wealth of content created by the YouTube generation. This will enable new opportunities of empirical analysis and supporting qualitative research (e.g. gender studies).

The first workshop, 'Music Information Retrieval Algorithms for Oral History Collections', was facilitated by the authors in July 2016 at the Digital Humanities 2016

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



conference.[10] The workshops introduced participants to specialist software libraries and applications used by MIR researchers.[11] All tools used in the workshop are freely available and cross-platform, meaning our examples are extendable, reusable and shareable. We used open source Python libraries for audio feature analysis, maths and machine learning. Additionally we packaged all dependencies in a Virtual Machine (VM) for ease of accessibility (see section 3.3). All Python code was written and executed in the Jupyter Notebook environment. Jupyter enabled us to develop pre-written examples that supported participants from various backgrounds: those new to coding could immediately engage in working examples, whilst those with more technical experience could edit and explore the code as they wished. In these exploratory sessions we worked with digital audio files from the 'Archive of Resistance': a growing collection of oral history content related to forms of resistance in history (for example, British Special Operations Executive operations during World War II) that is held at the Keep, an archive located near the University of Sussex.

### 3.1 Introduction to MIR and technologies used

During the last decade content-based music information retrieval has moved from a small field of study, to a vibrant international research community[12] whose work has increasing application across music and sound industries.[13] Driven by the growth of digital and online audio

[10] See http://dh2016.adho.org/workshops/
[11] See 'Listening for Oral History' https://github.com/algolistening/MachineListeningforOralHistory and 'Music Information Retrieval Algorithms for Oral History Collections' in Zenodo (July 2016) available at https://zenodo.org/record/58336#.WdOghLzyt24
[12] http://www.ismir.net/
[13] http://the.echonest.com/

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



archives, tools developed in MIR enable musically-meaningful information to be extracted directly from a digital audio signal by computational means, offering an automated, audio-based alternative to text-based tagging (the latter of which is common to both spoken word and music collections).[14] For example, digital audio files can be automatically described using high level musical features such as melodic contour, tempo or even "danceability".[15] These features are designed to enable automatic genre recognition or instrument classification, which in turn support archive management and recommender services. Applications of these methods in musical research and in industry include:[16]

- Music identification (commonly associated with software applications such as Shazam and SoundHound), plagiarism detection and copyright monitoring to ensure correct attribution of musical rights, identification of live vs studio recordings, for database normalization and near-duplicate results elimination.
- Mood, style, genre, composer or instrumental matching for search, recommender and organization of musical archives.
- Music vs speech detection for radio broadcast segmentation and archive cataloguing.

Techniques are numerous and rapidly evolving, but most methods work by extracting low-level audio features and combining these with domain specific knowledge (for example, that hip-hop generally has less beats per minute than dubstep) to create models from which more musically-meaningful descriptors can be built and – in turn – tempo, or melody, and ultimately

[14] Downie, J. Stephen. "Music information retrieval." Annual review of information science and technology 37, no. 1 (2003): 295-340.
[15] http://the.echonest.com/app/danceability-index/
[16] Michael A. Casey, Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes and Malcolm Slaney, "Content-based music information retrieval: Current directions and future challenges," *Proceedings of the IEEE*96 4 (2008): 668-696.

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



genre, composer, etc. might be identified. Low level features are essentially statistical summaries of audio in terms of distribution of energy across time, or frequency range. Some features might equate to perceptual characteristics such as pitch, brightness or loudness of a sound; others, such as MFCC (Mel-frequency Cepstral Coefficients), provide computationally powerful timbral descriptions but have less obvious direct perceptual correlates. Such low level features can then be used to create methods to find sonically-salient events, such as an onset detector, to identify when an instrument or voice starts playing. This low level information can then be combined with domain specific knowledge – such as the assumption that note onsets occur at regular intervals in most music – to create a tempo detector. In turn, this might be used to inform musical genre recognition, in the knowledge – as above – that hip-hop generally has less beats per minute than dubstep.

Just as these low level features can be combined and constrained to create high-level, information with many applications in engaging with and managing music archives, we are interested in the possibility that information of interest to historians and digital humanists might be discoverable in a digital audio file, that would be missed by the analysis of semantic, textual surrogates alone. Whilst no off-the-shelf tools exist for such analysis yet, the open, experimental ethos of digital audio and machine learning research cultures means that there are many accessible software tool kits available which enable rapid experimentation.

### 3.2 Learning MIR in Jupyter Workbooks

Content based MIR combines methods of digital signal processing, machine learning and music theory which in turn draw upon significant perceptual, mathematical, programming

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



knowledge and experience. Together these are skills that can take years to acquire. We wished to provide sufficient insight into the core concepts and techniques so as to inspire the imaginations of humanities researchers – with very mixed technical experience and interests – in a single day workshop. Fortunately, many of the complex technical and conceptual underpinnings can be readily grasped with audio-visual illustration, especially if they can be interactively explored. We therefore chose a constructionist approach in the form of hands-on workshops where participants learned through exploring interactive workbooks containing a mix of text-based information, audio-visual illustration and executable, editable code. This meant participants could work through carefully designed examples and learn by editing and exploring the code, all without having to grasp the mathematical bases of the ideas.

**Figure 1: Screenshot of a workbook that introduces participants to some basic methods (reading and loading digital audio files).**

Jupyter Notebooks were used to present example code in interactive workbooks which combined formatted (mark-down) text, executable code and detailed, dynamic audio-visual illustration. Jupyter provides a rich and supportive architecture for interactive computing, including a powerful interactive shell, support for interactive data visualization and GUI toolkits, flexible, embeddable interpreters and high performance tools for parallel computing. For novice and expert users alike it offers an interactive coding environment ideal for teaching, learning and rapidly experimenting with and sharing ideas. Executable code was written in python. Python is a human-readable general purpose programming language which is fast becoming the primary choice in data science, as well as computer science education in general. A vibrant, active community of users contribute to well-maintained open-source libraries which we used in the workbooks. These include: librosa (for music and audio analysis), matplotlib and ipython display

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



(for visualisation), scikit-learn (for machine learning), and SciPy and NumPy mathematical libraries (see Resources for a full list).

### 3.3 Sharing workbooks in a Virtual Machine: Reducing barriers to participation

Workshop participants were humanities scholars from a range of backgrounds, with differing levels of programming experience and computing knowledge. The requirement to install and configure the necessary collection of developer tools on a disparate selection of participant laptops had the potential to consume significant amounts of workshop time, increase the difference in participant progression through the schedule of activities, as well as diminish the amount of time available to explore MIR techniques. To avoid this, we created server and virtual machine (VM) based Python development environments for the workshop sessions. This approach reduced technological barriers to participation.

VM images were created and distributed on USB memory sticks. Installation of the developer tools, sample digital audio files and a minimal host operating system (Lubuntu 32 bit) resulted in a VM image size of about 8GB. Oracle VM VirtualBox was selected as the technology to implement the VM as the software; it is free and cross platform.[17] The main drawbacks of the approach were the large amount of storage needed on user machines, and the requirement for authors to create content far enough ahead of the event so that it could be distributed with the VM image.

In response to the first drawback – the requirement of 8GB of available disk space is a barrier to adoption for some users – a server-based alternative was also developed. The local

[17] https://www.virtualbox.org/

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



computing requirement for the server-based approach is a modern web browser to run the Jupyter Notebooks and a terminal program capable of implementing the network communication method used for the service, such as a Secure Shell (SSH) tunnel. The reason for using SSH is that it simplifies security concerns relating to the provision of unrestricted access to a Python development environment across the internet. Tunneling through SSH is not necessary for secure access, in our case it provided a technique that did not impose restrictions on contributor code development methods. This is a trade-off between simplicity and restriction of user behavior.

## 4 Workbook content

The workbooks were designed to be taught across a full day. In the morning session, they were used to introduce participants to the key concepts of coding, digital audio and audio features. In the afternoon, they were used by participants to apply these ideas and methods to an illustrative example. Workbook One introduced basic Python and the Jupyter notebook, with interactive exercises to familiarize participants with navigating the environment, executing code, carrying out basic mathematical operations and getting help. Workbook Two introduced the fundamental practical tools and ideas necessary to work with digital audio. These included loading, playing and visualizing digital audio files and introducing both ways of understanding how audio is represented digitally and ways to visualize and analyze frequency content of audio files. Workbook Three used plotting and listening to develop an intuitive understanding of audio features, as well as introducing practical tools and existing libraries used to inspect digital audio files and extract audio features. The worked through example, in Workbook Three, demonstrates how simple, low-level audio features (spectral bandwidth and the average number of zero-

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



crossings) can be used to distinguish between recordings of female and male interviewees. The two interviews provided were 10 minute interviews from the 'Archive of Resistance': one of a French woman and one an English man. The participants used the workbook to split the digital audio files from these interviews into one second chunks and then extracted a range of illustrative audio features. Finally, unsupervised clustering (k-means) was applied and the results of different pairs of features plotted to see which most successfully separated the two files. We found that even without clustering, the files could be separated with just two audio features.

**Figure 2: Scatter plot showing spectral bandwidth versus zero-crossing for all 1200 one second chunks of two 10 minute interviews.**

Figure 2 is a scatter plot that shows spectral bandwidth versus zero-crossing for all 1200 one second chunks of two, ten-minute interviews. Segments from recordings of the male speaker are colored blue, the female speaker segments are red. The two clusters are quite distinct, making it simple to automatically separate the segments of the two files. This demonstrates how low-level features can be used to identify recordings according to distinct characteristics of speaker's voice. Note that both recordings also contain a male interviewer.

In this example, only two files were used, but the approach is scalable to large data sets, demonstrating how audio feature analysis might be used to sort and explore unlabeled archives. Large scale tests would be necessary to prove the generalizability of these results. Nevertheless, this example illustrates that simple feature analysis holds promise for meeting several of the use cases listed in Section 3. Because different recording devices create digital audio files with differing acoustic profiles, this approach has potential – for example – to reveal information about the content (use case 1.3). Identifying interviewee-specific characteristics suggests a route

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



to automated content analysis: the identification of gender provides useful metadata (use case 2.1) that could underpin further gender-specific analyses (use case 2.2) and be potentially extended other personal characteristics (use case 2.4).

The final workbook explored how changes in *textural* information within a sound could be analyzed to identify the points at which the speaker changed from interviewer to interviewee. In sound processing, the mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. The mel scale spaces frequency bands to approximate the human auditory system. The coefficients of the MFC (MFCCs) do not intuitively correlate with perceptual properties of sound, however they do consistently reflect timbral characteristics. If we calculate the MFCCs for short segments, or frames, of audio throughout the files, changes in values throughout the file reveal points of timbral, or textural change. By plotting a two-dimensional self-similarity matrix - in which the difference between each frame is compared to all other frames - we can visualize periods of similarity and change. In musical applications, this technique is used to identify structures such as changes between and repetitions of verse and chorus; in spoken word interviews, this allows us to observe changes in texture which reflect transitions between interviewer and interviewee.

This example demonstrates how changes *within* a single file might be used to reveal changes in speaker characteristics: in this case who the speaker is (use-case 2.4). In combination with successful gender identification, this could be applied to use case 2.2, enabling large scale analysis of interviewer-interviewee dynamics. A similar method could potentially be developed to identify changes in rhythm or timbre (use case 2.3).

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



## 5 Findings and lessons learned

The examples in the workbooks illustrated that relatively simple audio analysis could be used to provide useful extra-semantic insights into oral history interviews. However, the degree to which participants grasped the possibilities was directly related to their familiarity with firstly digital audio and secondly digital methods in general. Those participants who were familiar with basic digital audio concepts and programming techniques (as was the case with many participants at the Digital Humanities 2016 workshop), recognized the potential of this approach, particularly those who worked with large audio archives. Other participants, those who had not previously engaged with computational methods, or done any coding of any kind, found it more difficult to imagine wider usage. This was particularly true for those who worked with very small sets of recording, for which this type of analysis is largely irrelevant.

Whilst the process of developing and facilitating both workshops indicated that MIR methods and technologies can be usefully applied to digital audio files that contain spoken word, adoption of these techniques is likely to be amongst existing computationally literate communities. For some, understanding how to interpret the visual display of audio files (e.g. the spectrogram) was challenging: 'I found it hard to translate spectrograms and plots to observations about the interviews'.[18]

Our workbooks allowed participants to carry out sophisticated, complex analysis, yet many participants found it difficult to envisage or imagine questions beyond those that we posed or included in the workbooks. This difficulty is linked to a number of factors. First, the workbooks in effect hide the complexities of a number of different tasks, so that while

[18] MIR for Oral History Collections (feedback)

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



participants could execute a piece of code and get results, this reduced understanding of the methods and the capabilities of the software libraries used and the code developed. Second, while the workbooks scaffolded learning, some participants – especially those new to programming – experienced a steep learning curve. This was especially evident in the second cohort/workshop, which mostly consisted of PhD students and researchers using oral history as a method in their work. Indeed, a portion of this cohort had little to no experience of the term "digital humanities" , or indeed the methods used in this domain. From our perspective it was interesting to discover that many oral history practitioners in this session still use the audio to text method as standard procedure. Therefore, working with the audio or digital files in a computational manner was completely unfamiliar territory. However, even though some participants found the workbooks challenging, they indicated to us that by working through the concepts, ideas and indeed the use cases, they felt inspired to re-think their current forms of analysis and to investigate how they might incorporate new forms of computational analysis. One participant, working on an historical collection of oral history interviews, reported that they could see how using audio analysis might help to reverse engineer the methodology or order of the original interviews. Other participants noted that from an archival perspective, techniques used to cluster "related" content might help with cataloguing collections by creating new forms of metadata. Many participants remarked that although they did not envisage learning the skills necessary to carry out this kind of analysis, having seen the potential, they could see value in collaborating with data scientists to explore new approaches, something they had not previously considered.

These remarks made us reflect on how best such methods could be introduced into research communities with little or no prior engagement with computational methods. A

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



common solution is to create a package with a graphical user interface and presets, which users can employ without conceptual or technical knowledge, yet the real potential of such methods can only be realized through hands-on, bespoke experimentation with specific real-world research questions. Our decision to present participants with pre-written executable code was intended as a compromise between these two positions.

In terms of the technological set up, with further time spent on preparation it would be preferable to develop a service to support the workshop exercises without tunneling the connection, which would result in a more reliable delivery of the service. HTTP(S) communication is resilient to the fluctuating quality that is common in public Wi-Fi networks as it does not require an uninterrupted connection, instead connections are created and destroyed with every interaction.

Provision of server based development environments is a good fit for cloud based computing infrastructures. The cost of running the cloud servers used for the workshops was less than £1 (or $1.30 approx.) for each event. In hindsight this means that server allocation should have been increased to improve service reliability. During both workshops broken connections were experienced and servers crashed; however, user experience was preserved by monitoring the cloud servers, supporting the participants and reconnecting broken connections as quickly as possible. The lesson learned with regards to connection and server stability was the extent of the variation of computing resource requirements across the activities and participants. The lesson learned with regards to the provision of virtual machines is the choice to use container technology (Docker) instead. Docker overcomes variations in user software configuration without the need to distribute a full operating system to every user.

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



## Conclusion

Our overall aim for this experimental project was to help the digital humanities and oral history community explore alternatives to the use of textual surrogates in oral history. Using off-the-shelf tools, we created and disseminated online interactive workbooks which demonstrate how generic audio analysis methods can be used to extract extra-semantic information from digital audio files. This approach might be used to complement traditional semantic analyses, providing automation of existing methods (metadata) or potentially new levels of analysis, such as interviewee-interviewer dynamics. By running participatory workshops, we tested the response of a wide range of humanists interested in oral history collections. The workshops demonstrated that this approach might be of great interest to DH researchers working with large audio databases, but are unlikely to be rapidly taken up by those working with small data sets, or with preference for manual methods.

Our work suggest great potential for audio-analysis in oral history. Refinement of methods to meet the use cases outlined in Section 3 will require systematic research on a wide range of large oral history archives in order to establish how well this work can be generalized and extended. In terms of future adoption in digital humanities communities, as with all computational analyses, a balance must then be sought between providing ready-to-use tools with a low barrier to entry, or nurturing a wider understanding technically and conceptually, such that members of the community may build and develop their own methods. As computational literacy grows amongst research communities, we see potential for novel applications of these methods in the future.

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



## Bibliography

Bertin-Mahieux, T., Ellis, D.P., Whitman, B. and Lamere, P. "The Million Song Dataset." *ISMIR* 2, no. 9 (2011).

Boyd, Doug. "OHMS: Enhancing Access to Oral History for Free." *Oral History Review* 40, no. 1 (2013): 95–106 doi:10.1093/ohr/oht031

Casey, Michael A., Veltkamp, Remco, Goto, Masataka, Leman, Marc, Rhodes, Christophe and Slaney, Malcolm "Content-based music information retrieval: Current directions and future challenges." *Proceedings of the IEEE96* 4 (2008): 668-696.

Clement, Tanya, Kraus, Kari, Sayers, J. and Trettien, Whitney "<audio>Digital Humanities</audio>: The Intersections of Sound and Method." *Proceedings of Digital Humanities* 2014. Lausanne, Switzerland https://terpconnect.umd.edu/~oard/pdf/dh14.pdf (accessed August 14, 2017)

Clement, Tanya E., Tcheng, David, Auvil, Loretta and Borries, Tony "High Performance Sound Technologies for Access and Scholarship (HiPSTAS) in the Digital Humanities." *Proceedings of the Association for Information Science and Technology* 51 (2014):1–10 doi:10.1002/meet.2014.14505101042

Downie, J. Stephen. "Music information retrieval." Annual review of information science and technology 37, no. 1 (2003): 295-340.

Grele, Ronald J., "Oral History as Evidence." In *History of Oral History: Foundations and Methodology* Thomas L. Carlton, Lois E. Myers and Rebecca Sharpless (eds) (UK, 2007), 69.

Kluyver, Thomas, Ragan-Kelley, Benjamin, Pérez, Fernando, Granger, Brian, Bussonnier, Matthias, Frederic, Jonathan, Kelley, Kyle, Hamrick, Jessica, Grout, Jason, Corlay, Sylvain, Ivanov, Paul, Avila, Damián, Abdalla, Safia, and Willing, Carol (Jupyter Development Team) "Jupyter Notebooks – a publishing format for reproducible computational workflows." *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (2016):87-90. doi: 10.3233/978-1-61499-649-1-87

Tzanetakis, George and Cook, Perry "Musical genre classification of audio signals." *IEEE Transactions on speech and audio processing* 10, no. 5 (2002):293-302.

The Version of Record of this manuscript has been published and is available in the Journal, Music Reference Services Quarterly, November 2017, http://dx.doi.org/10.1080/10588167.2017.1404307



## Resources

All workbooks, data, slides from the workshops are deposited in both GitHub and Zenodo:

- 'Machine Listening for Oral History', GitHub
- Eldridge, A., Kiefer, C., Webb, S., Jackson, B., & Baker, J. (2016, July). Music Information Retrieval Algorithms for Oral History Collections. Zenodo. http://doi.org/10.5281/zenodo.58336

Python libraries used:

https://www.scipy.org/

http://scikit-learn.org/

https://matplotlib.org/

https://ipython.org/ipython-doc/3/api/generated/IPython.display.html

https://librosa.github.io/librosa/

Other:

https://www.docker.com/