AUTOMATIC TRANSCRIPTION AND PITCH ANALYSIS OF THE
BRITISH LIBRARY WORLD & TRADITIONAL MUSIC COLLECTIONS

Samer Abdallah¹, Aquiles Alencar-Brayner², Emmanouil Benetos⁵, Stephen Cottrell¹
Jason Dykes³, Nicolas Gold¹, Alexander Kachkaev³, Mahendra Mahey², Dan Tidhar⁴,
Adam Tovell², Tillman Weyde³, Daniel Wolff⁶
¹ Department of Computer Science, University College London, ² The British Library,
³ Department of Computer Science, City University London,
⁴ Department of Music, City University London
⁵ Centre for Digital Music, Queen Mary University of London
dml-owner@city.ac.uk - http://dml.city.ac.uk

1. INTRODUCTION

Music research, particularly in fields like systematic musicology, ethnomusicology, or music psychology has de-
veloped as “data oriented empirical research” (Parnccutt, 2007), which benefits from the development of computing
methods and infrastructure. In ethnomusicology there has been a recent growing interest in computational meth-
ods and their application to audio data collections (Gómez et al., 2013), (Canazza et al., 2010), and in the degree
to which such methods may reveal insights into musical practice which may not be evident from the participant-
observation paradigms that have otherwise characterised the discipline. For technological and legal reasons research
in this area was previously limited to small datasets, but this is changing in part due to the contribution of research
projects such as CompMusic¹ and cultural preservation projects such as Europeana Sounds².

The authors of this paper collaborate in the UK AHRC-
funded project Digital Music Lab - Analysing Big Music
Data (DML)³ to develop methods and technologies to
support the use of Big Data in musicology (Weyde et al.,
2014). As part of the project, we have developed soft-
ware/hardware infrastructure for exploring and analysing
large-scale audio collections, aiming to assist research in
systematic and empirical (ethno)musicology.

A major partner of the DML project is The British Li-
brary (BL), which holds several million audio recordings
in its Sound Archive, spanning oral history interviews, en-
vironmental & nature sounds, as well as over 3 million
recordings from classical, popular, world and traditional
music (of which approximately 10% are digitized). A por-
tion of these audio recordings (approx. 60k) are currently
available for online streaming⁴. Through the DML project,
a computing server was installed on-site at the BL, en-
abling storage and analysis for a collection of over 29k
recordings from its ‘World & Traditional Music’ corpus.

In this paper we present the collection of World & Tra-
ditional music that was curated and analysed as part of the
DML project; also presented will be methods for automatic
transcription and pitch analysis that were applied to record-
ings from that collection, and which were used as a basis
for creating an integrated tool/interface enabling musicoso-
logical enquiries and research in large music collections.

2. COLLECTION

The BL Sound Archive holds one of the world’s largest
collections of recordings variously described as traditional,
folk or ‘world’ music. The dataset drawn from this collect-
ion for the DML project consists of 29,198 audio record-
ings. It covers a large collection (8k) of English, Irish, and
Scottish folk songs; 1300 recordings from Oceania; 12k
recordings from Africa (covering West and South Africa,
as well as large collections from Uganda and Sudan); over
6k recordings from Asia (mostly from Nepal, India, and
Pakistan); 1100 recordings from the Middle East; and a
small collection of 47 recordings from the Americas (com-
prising music of indigenous Indians from Colombia). It
also contains collections of wax cylinders recorded by pi-
oneering fieldworkers, as well as more recent recordings
made on a range of formats as part of ethnographic re-
search. Recording dates span from 1898 (from the ethno-
graphic wax cylinders collection) up to 2014. It is worth
noting that several of these recordings contain segments of
speech as well as music, or even overlapped speech and
music. Information on the five largest collections can be
seen in Table 1.

The recordings are also accompanied with rich meta-
data in METS/XML format. Information present in the
metadata includes: title, collection ID, description, per-
former, recording engineer, recording date and temporal
information (e.g. Easter), language, geographic information,
as well as audio file information (duration, sampling
rate, resolution, URL for publicly available recordings).

Authors in alphabetical order. This work was supported by the
UK AHRC-funded project ‘Digital Music Lab - Analysing Big Music
Data’, grant no. AH/L01016X/1 and the UK AHRC funded project
‘An Integrated Audio-Symbolic Model of Music Similarity’, grant no.
AH/M002454/1. EB is supported by a Royal Academy of Engineering
Research Fellowship, grant no. RF/128.

¹ http://compmusic.upf.edu/
² http://www.europeanasounds.eu/
³ http://dml.city.ac.uk
⁴ http://sounds.bl.uk/
⁵ http://sounds.bl.uk/
Table 1: The five largest collections from the BL World & Traditional Music dataset used for the DML project.

<table>
<thead>
<tr>
<th>Title</th>
<th># recordings</th>
<th>Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob &amp; Jacqueline Patten English Folk Music Collection</td>
<td>6333</td>
<td>1953-2002</td>
</tr>
<tr>
<td>John Howson English, Irish &amp; Scottish Folk Music Collection</td>
<td>3498</td>
<td>1930-1999</td>
</tr>
<tr>
<td>Reg Hall English, Irish &amp; Scottish Folk Music &amp; Customs Collection</td>
<td>3195</td>
<td>1949-1996</td>
</tr>
<tr>
<td>Klaus Wachsmann Uganda Collection</td>
<td>1538</td>
<td>1949-1954</td>
</tr>
<tr>
<td>Peter Cooke Uganda Collection</td>
<td>1277</td>
<td>1964-1997</td>
</tr>
</tbody>
</table>

3. AUTOMATIC TRANSCRIPTION

As a first step towards musicological analysis for large audio collections, we use automatic music transcription (AMT) technology to convert a recording into machine-readable music notation (Klapuri & Davy, 2006). The vast majority of research in AMT technology is however limited to Western/Eurogenetic music, where an audio recording is converted into a MIDI-like representation.

For this work we used the model of Benetos & Dixon (2012), which can support the estimation of multiple pitches (along with onsets, offsets, and velocities) in a scale equal to the resolution of the input time/frequency representation (in our case, 20 cent resolution given as input a log-frequency spectrogram of 60 bins/octave), and ranked first in the MIREX 2013 evaluations for Multiple-F0 estimation and Note Tracking. This method is also publicly available as a VAMP plugin, which can be used in conjunction with software such as Sonic Visualiser.

Using the aforementioned method, the output is a probability distribution of pitches in 20 cent resolution over time: \( P(f, t) \) (\( f \) corresponds to pitch and \( t \) to the time index). This is post-processed and converted into a binary representation of note events, with a corresponding onset time, offset time, pitch, and velocity value. Apart from high-level musical applications the resulting transcription can serve as a way to store or visualise the content of a music recording (an example can be shown in Fig. 1). The resulting transcription files, along with other low-level features, can be downloaded for individual files through the semantic web server of the DML project.

4. PITCH HISTOGRAMS

Information extracted from the automatic transcriptions can be used to generate information on the pitch content of a recording, or a collection of recordings. Given a transcription \( P(f, t) \), a pitch histogram for that recording can be generated by: \( P(f) = \sum_t P(f, t) \). The above process can be extended for any number of recordings, for visualising the pitch content of collections. As part of the DML interface for browsing/interacting with large music collections (Kachkaev et al., 2015), pitch histograms for audio collections can be computed on-demand. The DML visualisation interface is available online. An example pitch histogram taken from that interface is shown in Fig. 2.

This allows the identification of the most frequently encountered pitches within a given collection, which in turn may suggest pitch hierarchies prevailing within the music culture and, potentially, commonly used scales. Cross-referencing with ethnographically-grounded research is likely to be necessary to confirm the cultural validity of the results. Nevertheless, the method allows large-scale examination of particular music cultures, comparison between collections to observe similar patterns of pitch distribution and, possibly, the automatic identification of otherwise unattributed recordings according to the pitch distribution found in them.

5. DISCUSSION

In this paper, we presented work carried out by the DML project on analysing over 29k recordings from the BL World and Traditional Music collection. This involved producing automatic transcriptions at a high pitch resolution and creating pitch histograms both for individual recordings and collections. These are supported by a semantic web server for downloading features for individual recordings and a visualisation interface for browsing through collections.

Questions remain about the use of additional high-level features for describing folk and traditional music recordings, given culturally-specific aspects such as modes, structure, and instrumentation. We believe that pitch analysis at a high-frequency resolution is the first step towards that...
Another issue relates to the integration of recordings from copyright-restricted collections; the recent UK Intellectual Property Bill has enabled researchers to use copyright-restricted data for research purposes, but computations still need to be done on-site. We believe this approach will enable musicians and musicologists unprecedented access to recordings, overcoming such restrictions. Finally, the problem of segmenting the recordings into speech and music parts is currently being investigated through the MuSpeak project. 

6. REFERENCES


10 http://mirg.city.ac.uk/muspeak