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

Samer Abdallah[1], Aquiles Alencar-Brayner[2], Emmanouil Benetos[5], Stephen Cottrell[4] Jason Dykes[3], Nicolas Gold[1], Alexander Kachkaev[3], Mahendra Mahey[2], Dan Tidhar[4], Adam Tovell[2], Tillman Weyde[3], Daniel Wolff[3]
[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 developed as "data oriented empirical research" (Parncutt, 2007), which benefits from the development of computing methods and infrastructure. In ethnomusicology there has been a recent growing interest in computational methods 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 CompMusic [1] and cultural preservation projects such as Europeana Sounds [2] .

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 software/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 Library (BL), which holds several million audio recordings in its Sound Archive, spanning oral history interviews, environmental & nature sounds, as well as over 3 million recordings from classical, popular, world and traditional music (of which approximately 10% are digitized). A portion of these audio recordings (approx. 60k) are currently available for online streaming [4] . Through the DML project, a computing server was installed on-site at the BL, enabling 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 & Traditional 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 recordings from that collection, and which were used as a basis for creating an integrated tool/interface enabling musicological 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 collection for the DML project consists of 29,198 audio recordings. 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 (comprising music of indigenous Indians from Colombia). It also contains collections of wax cylinders recorded by pioneering fieldworkers, as well as more recent recordings made on a range of formats as part of ethnographic research. Recording dates span from 1898 (from the ethnographic 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 metadata in METS/XML format. Information present in the metadata includes: title, collection ID, description, performer, 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/

Title	# recordings	Dates
Bob & Jacqueline Patten English Folk Music Collection	6333	1953-2002
John Howson English, Irish & Scottish Folk Music Collection	3498	1930-1999
Reg Hall English, Irish & Scottish Folk Music & Customs Collection	3195	1949-1996
Klaus Wachsmann Uganda Collection	1538	1949-1954
Peter Cooke Uganda Collection	1277	1964-1997

Table 1: The five largest collections from the BL World & Traditional Music dataset used for the DML project.

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 [5]. This method is also publicly available as a VAMP plugin [6], which can be used in conjunction with software such as Sonic Visualiser [7].

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 musicological 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 [8].

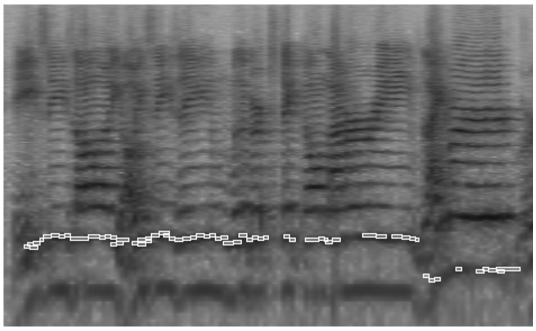

Figure 1: The spectrogram for a segment of 'The Blackbird' from the Bob & Jacqueline Patten Collection. Overlayed as white boxes are detected pitches.

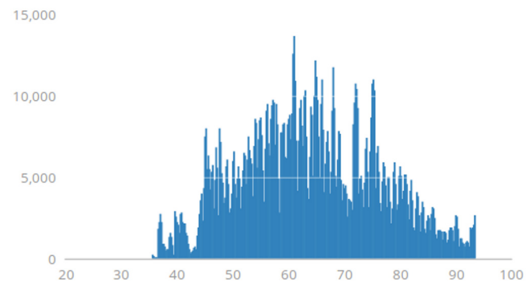

Figure 2: The aggregated pitch histogram for 3,022 Ugandan recordings, as shown in the DML VIS interface.

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 [9]. 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

[5] http://www.music-ir.org/mirex/wiki/MIREX_HOME
[6] https://code.soundsoftware.ac.uk/projects/silvet/files
[7] http://sonicvisualiser.org/
[8] http://mirg.city.ac.uk/cp/
[9] http://dml.city.ac.uk/vis/

goal. 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 [10] .

6. REFERENCES

Benetos, E. & Dixon, S. (2012). A shift-invariant latent variable model for automatic music transcription. *Computer Music Journal*, *36*(4), 81–94.

Canazza, S., Camurri, A., & Fujinaga, I. (2010). Special section: Ethnic music audio documents: From the preservation to the fruition. *Signal Processing*, *90*(4), 977–1334.

Gómez, E., Herrera, P., & Gómez-Martin, F. e. (2013). Special issue: Computational ethnomusicology. *Journal of New Music Research*, *42*(2).

Kachkaev, A., Dykes, J., Abdallah, S., Barthet, M., Benetos, E., Cottrell, S., Dixon, S., Gold, N., Hargreaves, S., Tidhar, D., Wolff, D., & Weyde, T. (2015). Small multiples for big data a framework for comparison using open web technologies with music collections. In *IEEE Information Visualization Conference 2015, submitted*.

Klapuri, A. & Davy, M. (Eds.). (2006). *Signal Processing Methods for Music Transcription*. New York: Springer-Verlag.

Parncutt, R. (2007). Systematic musicology and the history and future of western musical scholarship. *Journal of Interdisciplinary Music Studies*, *1*, 1–32.

Weyde, T., Cottrell, S., Dykes, J., Benetos, E., Wolff, D., Tidhar, D., Gold, N., Abdallah, S., Plumbley, M., Dixon, S., Barthet, M., Mahey, M., Tovell, A., & Alencar-Brayner, A. (2014). Big data for musicology. In *Digital Libraries for Musicology Workshop*, (pp. 85–87)., London, UK.

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