THE DIGITAL MUSIC LAB: A BIG DATA INFRASTRUCTURE FOR DIGITAL MUSICOLOGY

Tillman Weyde[1], Stephen Cottrell[1], Jason Dykes[1], Emmanouil Benetos[1,2], Daniel Wolff[1], Alexander Kachkaev[1], Simon Dixon[2], Stephen Hargreaves[2], Mathieu Barthet[2], Nicolas Gold[3], Samer Abdallah[3], Dan Tidhar[4], Mark Plumbley[5]

[1]City University London [2]Queen Mary University of London

[3]University College London [4] University of Cambridge [5] University of Surrey

{t.e.weyde,stephen.cottrell.1,j.dykes,emmanouil.benetos.1,daniel.wolff.1, alexander.kachkaev.1}@city.ac.uk; dut20@cam.ac.uk; m.plumbley@surrey.ac.uk; {s.e.dixon,s.hargreaves,m.barthet}@qmul.ac.uk; {n.gold,s.abdallah}@ucl.ac.uk

ABSTRACT

In music research, many tools and techniques for the analysis of individual pieces exist, while far fewer are available on the collection level. Such methods and tools are increasingly necessary as more data is being made available, and the Digital Music Library helps fill that gap. Based on parallel server-side processing with Music Information Retrieval techniques, Semantic Web knowledge representation, and specifically designed interactive visualisations, an infrastructure has been developed, which allows web-based processing, querying, visualisation, and musicological analysis of large music collections. By analysing large, user-defined collections, we enable the investigation of questions such as whether or not a geographical bias for orchestral tuning pitch exists or the examination of trends in musical key and tempo. We particularly support comparisons between multiple music collections.

1. INTRODUCTION

The Digital Music Lab infrastructure is the result of a collaborative AHRC funded project including the British Library and I Like Music Ltd. The aim of this project is to enable big music data analysis for musicology. By bringing together existing technologies and some new developments in a integrated system we enable the analysis of music in place, i.e. without compromising copyright, and in a distributed manner. Based on low-level feature extraction, the definition of collections by search criteria and the analysis and comparison of results are supported.

1.1 System

The DML system is consists of a back-end analytical computation server (CS), an Information and Computation Management System (ICMS), and a Web-based Visualisation front-end (VIS). The feature extraction happens on the system where the raw data is stored, so that copyright-protected data can be analysed. The ICMS supports the creation of connected systems for analyses of audio collections across organisations.

1.2 Features and Analysis

The analysis on the CS makes use of existing technologies, such as VAMP plug-ins, and allows for addition of new methods. The feature extraction includes automatic audio transcription, chord detection etc. and the data includes scores and their associated metadata. The user can define a collection based on a search query, selecting typically items based on metadata, as available in the supporting library. The analysis on the collection level then aggregates and computes statistics of the computed feature values.

1.3 Data and Interfaces

The system provides currently access to four different library collections with over 250,000 audio tracks, including ethnographic recordings, classical music in historic recordings, as well as jazz, rock, world, and western pop music. This data can be access in different views, including an RDF based browser and a specifically designed visual interface.

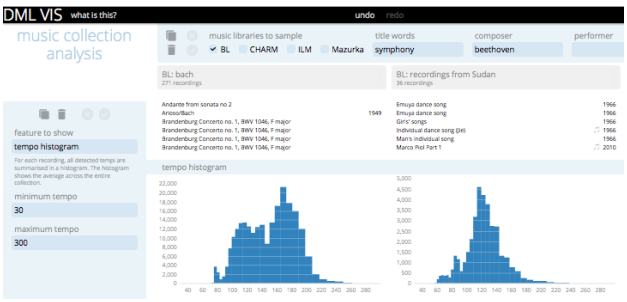

Figure 1. Visual interface (VIS) for collection definition and analysis.

Acknowledgement This works has been supported by AHRC grant AH/L01016X/1.