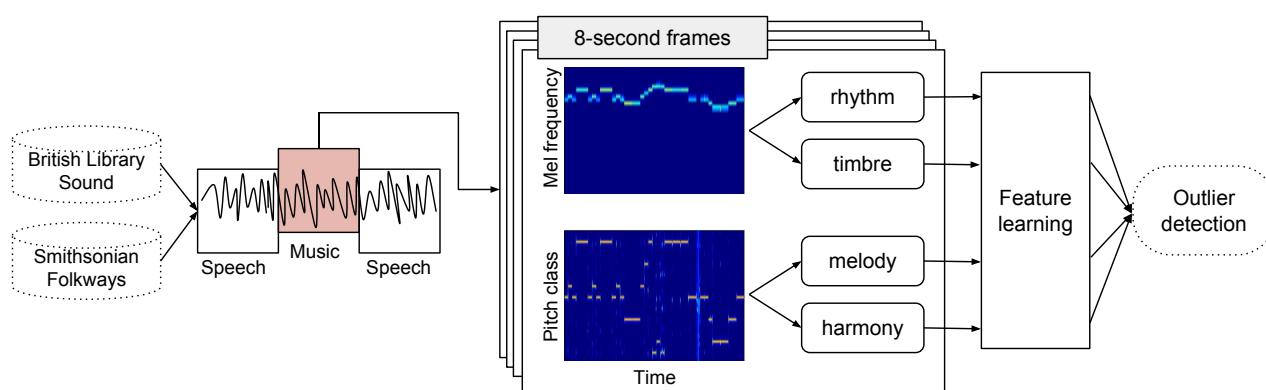


A large-scale comparison of world music corpora with computational tools

By Maria Panteli, Emmanouil Benetos, Simon Dixon from the Centre for Digital Music, Queen Mary University of London

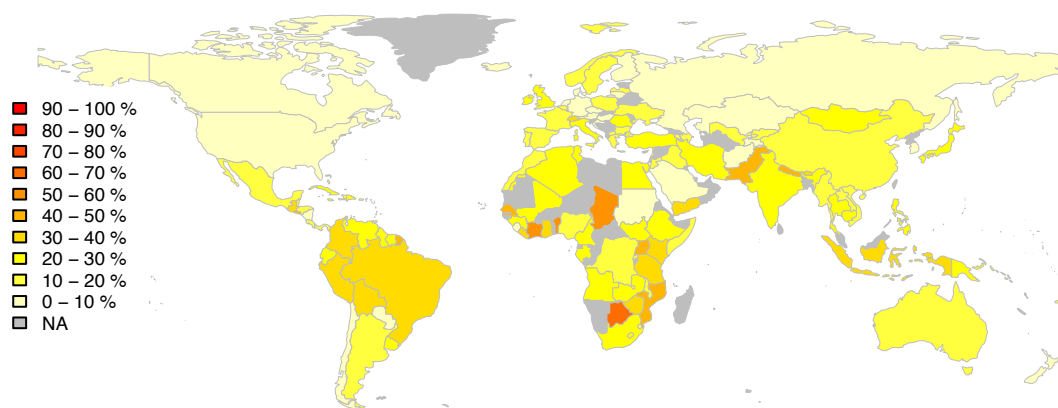
The comparative analysis of world music cultures has been the focus of several ethnomusicological studies in the last century. With the advances of Music Information Retrieval and the increased accessibility of sound archives, large-scale analysis of world music with computational tools is today feasible. We combine music recordings from two archives, the Smithsonian Folkways Recordings and British Library Sound Archive, to create one of the largest world music corpora studied so far (8200 geographically balanced recordings sampled from a total of 70000 recordings). This work was submitted for the 2017 British Library Labs Awards - Research category.

Our aim is to explore relationships of music similarity between different parts of the world. The history of cultural exchange goes back many years and music, an essential cultural identifier, has travelled beyond country borders. But is this true for all countries? What if a country is geographically isolated or its society resisted external musical influence? Can we find such music examples whose characteristics stand out from other musics in the world? By comparing folk and traditional music from 137 countries we aim to identify geographical areas that have developed a unique musical character.



Methodology: Signal processing and machine learning methods are combined to extract meaningful music representations from the sound recordings. Data mining methods are applied to explore music similarity and identify outlier recordings.

We use digital signal processing tools to extract music descriptors from the sound recordings capturing aspects of rhythm, timbre, melody, and harmony. Machine learning methods are applied to learn high-level representations of the music and the outcome is a projection of world music recordings to a space respecting music similarity relations. We use data mining methods to explore this space and identify music recordings that are most distinct compared to the rest of our corpus. We refer to these recordings as 'outliers' and study their geographical patterns. More details on the methodology are provided [here](#).



Distribution of outliers per country: The colour scale corresponds to the normalised number of outliers per country, where 0% indicates that none of the recordings of the country were identified as outliers and 100% indicates that all of the recordings of the country are outliers.

We observed that out of 137 countries, Botswana had the most outlier recordings compared to the rest of the corpus. Music from China, characterised by bright timbres, was also found to be relatively distinct compared to music from its neighbouring countries. Analysis with respect to different features revealed that African countries such as Benin and Botswana, indicated the largest amount of rhythmic outliers with recordings often featuring the use of polyrhythms. Harmonic outliers originated mostly from Southeast Asian countries such as Pakistan and Indonesia, and African countries such as Benin and Gambia, with recordings often featuring inharmonic instruments such as the gong and bell. You can explore and listen to music outliers in [this interactive visualisation](#). The datasets and code used in this project are included in this [link](#).

Outlier recordings per country (click on each point to listen to the audio). More info at: mpanteli.github.io/music-outliers/demo/



[Interactive visualisation](#) to explore and listen to music outliers.

This line of research makes a large-scale comparison of recorded music possible, a significant contribution for ethnomusicology, and one we believe will help us understand better the music cultures of the world.