

Automatic Detection of Outliers in World Music Collections

Maria Panteli

Emmanouil Benetos

Simon Dixon

Queen Mary University of London, United Kingdom

In big data collections it is often important to identify outlier behaviour that should be filtered out or treated differently. In music this could help us identify sound recordings that stand out in a recorded music collection. We call these recordings ‘outliers’ and perform a computational analysis to detect them. We focus on world and traditional music collected from available sound archives. Using signal processing tools we extract audio features that denote rhythmic, melodic, harmonic, and timbral aspects of each recording in our collection. Outliers in the dataset are detected with data mining techniques. A sound recording is an outlier depending on how distinct its musical characteristics are compared to other recordings in the collection. However, outlier recordings can also be detected in cases where the descriptors have failed to capture the correct attributes. To evaluate our findings we perform a listening test with music experts. From preliminary results we are able to capture, amongst other, songs with distinct patterns of timbre and rhythm but also speech samples as outliers. The proposed methodology can help identify sound recordings that have a unique musical character or filter irrelevant audio from music collections.