Exploring Music Collections: An Interactive, Dimensionality Reduction Approach to Visualizing Songbanks

Oscar Gomez, Kaustuv Kanti Ganguli, Leonid Kuzmenko, Carlos Guedes
{oag229,kaustuvkanti,lk1640,carlos.guedes}@nyu.edu
New York University Abu Dhabi, Abu Dhabi, UAE 129188

ABSTRACT

This is an overview paper on an interactive music exploration interface for music collections. This interface is meant to help explore the cross-cultural similarities, interactions, and patterns of music excerpts from different regions and understand the similarities by employing computational audio analysis, machine learning, and visualization techniques. In our computational analysis, we used standard audio features that capture timbre information and projected them onto a lower-dimensional space for visualizing the (dis)similarity. There are two collections of non-Eurogenetic music under study. The 2-D and 3-D mappings are visualized through a dashboard application and also rendered in Virtual Reality space where users can interact and explore to get meaningful insights about the structural (dis)similarities of the music collections.

CCS CONCEPTS

• **Human-centered computing → User interface design**; • **Applied computing → Sound and music computing**.

KEYWORDS

Computational analysis, Ethnomusicology, Virtual reality, Interface design, Visualization techniques.

ACM Reference Format:
Oscar Gomez, Kaustuv Kanti Ganguli, Leonid Kuzmenko, Carlos Guedes. 2020. Exploring Music Collections: An Interactive, Dimensionality Reduction Approach to Visualizing Songbanks. In *25th International Conference on Intelligent User Interfaces Companion (IUI '20 Companion), March 17–20, 2020, Cagliari, Italy*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3379336.3381461

1 INTRODUCTION

Digital humanities is progressively coming to the forefront as a field of scholarly inquiry in the non-Western world. It also focuses on the qualitative study of the ways in which music and their attendant epistemologies (including those of the sonic digital humanities themselves), have been shaped by digital technologies. This is an extension of our previous work [3]. The goal of the current work is to explore the cross-cultural similarities, interactions, and patterns of the music excerpts from different music collections and understanding these similarities by employing visualization and dimensionality reduction techniques to the data. As a baseline model we extracted standard features, such as MFCCs to investigate how these features correlate with the music excerpts. MFCCs are spectral representations and can best used to describe the instrumentation and genre/style of the recordings and have been used in MIR extensively in the past. We tested our baseline model against an unsupervised analysis of the raw representation of the spectrogram that is fed to a deep autoencoder and investigate if its able to learn more complicated relationships and patterns of attributes of the music structure. In our deep learning model we are using a series of hidden layers that encode and decode the spectrogram to learn a compressed representation of the important features of the spectrogram.

We present two novel exploration interfaces for music based on the above approach. One is a dashboard application where users can listen to the clips in the collection as grouped by similarity and judge the similarity by themselves. They can also use selection tools to investigate meta-data of the selections such as the artists present in a cluster. The other is a similar Virtual Reality application that provides the same basic functionality but provides a much more engaging and immersive experience. We present a qualitative analysis two non-Eurogenetic music collections through these interfaces. The Music and Sound Cultures research group (MaSC) [1] has a particular focus on the Arab world and regions, including East Africa's Swahili coast, that have a long history of contact with the Arab world.

2 CORPORA

The corpora for analysis consist of two collections: the Eisenberg Collection and the Music Compendium from the Arab Mashriq. The former, an archival collection by MaSC, is the Eisenberg Collection of East African Commercial Sound Recordings. This collection contains 500 sound files and associated metadata of commercial recordings produced for East African Swahili coast audiences between the late 1920s and the first decade of the twenty-first century. Most of the sounds in the collection fall within the realm of Swahili-language urban popular music from the Swahili coast's major urban centers (Mombasa, Dar es, Salaam, and Zanzibar). There are also examples of rural music traditions, colonial-era, martial music, recited Swahili poetry, and Swahili comedy sketches. Using a collections [2] as data approach, the second corpus of this work consists of a digital compendium of 2827 recordings collected from the Library's collection of Arab audio on compact disc. The ethnic group and region of the digital compendium comes from Jordan, Kurdistan, Turkey, Lebanon, Morocco, Egypt, UAE, Bahrain, Yemen, Afghanistan, Beirut, and Azerbaijan.

[1] http://masc.hosting.nyu.edu/
[2] http://dlib.nyu.edu/findingaids/html/nyuad/ad_mc_035/index.html

IUI '20 Companion, March 17–20, 2020, Cagliari, Italy

ACM ISBN 978-1-4503-7513-9/20/03…$15.00
https://doi.org/10.1145/3379336.3381461

3 COMPUTATIONAL ANALYSIS OF AUDIO

We extracted standard feature extractors, such as Mel-Frequency Cepstral Coefficients (MFCC) to investigate how these features correlate with the music excerpts. We tested our baseline model against the extracted principal components from the mel-scale spectrogram and against an unsupervised analysis of the raw representation of the spectrogram that is fed to a deep autoencoder to investigate if these methods are able to learn more complicated relationships and patterns of attributes of the music structure. We used the bottleneck of this autoencoder layer as our final feature representation and fingerprint for the music excerpts.

3.1 User Interface Design

The next step was to convert the high-dimensional representations into 2-D and 3-D embeddings to visualize interesting clusters of the music excerpts using the feature representations learned from the model. Traditional dimensionality reduction techniques such as multidimensional scaling [1] are linear that focus on keeping the low-dimensional representations of dissimilar data points far apart. In our analysis, we used t-distributed stochastic neighbor embedding (t-SNE) for visualizing the resulting similarities of the feature representations [2]. Compared to methods discussed previously, t-SNE is capable of capturing much of the local structure of the high-dimensional data, while also revealing global structure such as the presence of clusters at several scales.

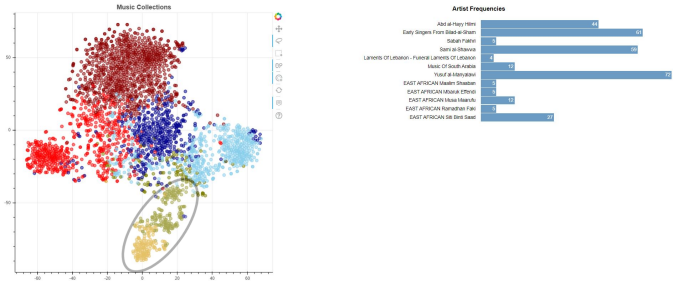

Figure 1: Dashboard application to explore 2-dimensional mappings of music collections. On the right, a histogram with the artists in the selection made on the left is presented

To get meaningful insights about the structural similarities of the different corpus, a dashboard browsing application was developed. Users could interact and explore the corpus by navigating through a 2-D similarity space. The user could select and hover over different points in the space and listen to the corresponding music excerpts. Points closer to each other reveal timbral and instrumentation similarities between different music excerpts. To offer a way to visualize clusters formed in the original, high-dimensional space, k-means was performed in this space and the clusters are colored in the final 2-D and 3-D visualizations. Figure 1 shows a t-SNE map with 6 clusters for the corpora.

3.2 VR Implementation

Virtual reality is a fast growing and powerful medium that allows the user to engage with the content on the next level of immersion by enabling a six depth-of-freedom motion inside the virtual environment. The scatter plot representation of the corpus was easily transformed from 2D space into virtual reality. Unreal Engine 4 was used as a development platform. The user could point at any nearby sphere and pull the trigger of the controller to activate the corresponding audio clip. To add more immersion to the experience, Steam Audio was used as a spatialization method. The virtual reality headset of choice was Oculus Quest. It is standalone and wireless, which means that it could be easily transported. In the future the experience could be shaped so that the user would be able to use their own hands instead. Figure 2 shows a screenshot of the VR rendering of the 3-D t-SNE space.

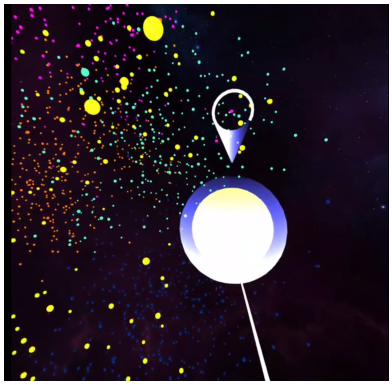

Figure 2: Screenshot of the VR rendering of the 3-D map.

4 CONCLUSION AND FUTURE WORK

This paper presents our first steps in developing an analysis of music collections through interactive listening and exploration interfaces. We have provided a short qualitative analysis on two non-Eurogenetic collections made possible by the use of these interfaces. However, the entire analysis pipeline can be applied to other collections. As the implementation is supported by open-source tools, our tool can be made available in the future to analyze other music collections. Furthermore, other interactivity can be added once metadata beyond the artists in the collection is made available, which will greatly enhance the user-experience.

ACKNOWLEDGEMENT

This research is part of project "Computationally engaged approaches to rhythm and musical heritage: Generation, analysis, and performance practice," funded through a grant from the Research Enhancement Fund at the New York University Abu Dhabi.

REFERENCES

[1] Joseph B Kruskal and Myron Wish. 1978. Multidimensional scaling. Number 07–011 in Sage University Paper series on quantitative applications in the social sciences.
[2] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, Nov (2008), 2579–2605.
[3] Konstantinos Trochidis, Beth Russell, Andrew Eisenberg, Kaustuv Kanti Ganguli, Oscar Gomez, Christos Plachouras, Carlos Guedes, and Virginia Danielson. 2019. Mapping the Sounds of the Swahili coast and the Arab Mashriq: Music research at the intersection of computational analysis and cultural heritage preservation. (2019).