A machine learning approach for clustering western and non-western folk music using low-level and mid-level features

Neocleous Andreas[1], Panteli Maria[1], Rafaela Ioannou[1], Nicolai Petkov[2], Christos N. Schizas[1]

[1]University of Cyprus, Nicosia, Cyprus
neocleous.andreas@gmail.com, m.x.panteli@gmail.com,
rioann01@cs.ucy.ac.cy, schizas@ucy.ac.cy
[2]University of Groningen, Groningen, The Netherlands
n.petkov@tug.nl

Abstract. This work describes an unsupervised method for clustering folk music from four non-European eastern countries, western music and folk music of Cyprus. The folk music of Cyprus is consisted by Greek- Cypriot folk music and Turkish-Cypriot folk music. We used low-level and mid-level features for our investigations and we applied two unsupervised methods, the so-called k-means and the self organizing maps for creating cluster maps. From the results we observed that the Greek-Cypriot songs, the Turkish Cypriot songs and the Turkish makams are being clustered from the system in different clusters. Arabic and music from Iran are clustered in the same cluster and music from Syria is clustered together with music of Turkey. Western music is clustered in to two different clusters.

Keywords. Computational ethnomusicology, signal processing, timbre, tonality

1 INTRODUCTION

The aim of this work is to explore timbre and tonal similarities between folk music of Cyprus, western music and non-western music from, especially, East Mediterranean countries, using a computational approach. Models with K-means and self organizing maps (SOM) were created for clusters 1,2,..,10. in order to inspect the robustness of the results. We used the elbow method [5] for the identification of the optimal number of clusters for the particular problem. We built models using three different feature sets. The first feature set was consisted with low-level features, the second feature set was consisted with mid-level features and the third feature set was consisted with low-level and mid-level features. Studies exist that identify similarities between Cypriot folk music and Byzantine music [ex, 1]. These studies are based on manual analysis, a time consuming task that restricts the analysis on relatively small music collections. Related research in the area of computational ethnomusicology, has

In MML 2013: International Workshop on Machine Learning and Music, ECML/PKDD, Prague, 2013.

explored similarities/dissimilarities of western and non-western music traditions [2]. Non-western music, such as Cypriot, Turkish or Arabic music, uses different tuning systems than the equal-temperament system that is widely used in western music. Traditional music instruments of non-western music traditions exhibit timbre particularities that are interesting to study and contrasted to western music. Given the above particularities between Cypriot folk music, western and non-western music, a system is developed that investigates music similarity based on tonal and timbre features.

2 DATA

The music database consisted of a total of 242 monophonic songs from traditional music of Cyprus (including both Greek-Cypriot and Turkish-Cypriot performances), Turkey, Iran, Syria and other Arab countries, and western music. The instruments used in this music collection consist of wind and string instruments of the above regions and music traditions, as well as the singing voice. In particular, 37 songs were Greek Cypriot folk songs performed on the "pithkiavli" instrument [3] by Greek-Cypriot musicians Andreas Gristakkos and Giannis Zavros. The 75% of these songs were recorded specifically for the purposes of this research. Cypriot folk songs were also collected from Turkish-Cypriot performances. These included 40 pithkiavli melodies performed by the Turkish-Cypriot musician Enver Kavaz. Recordings from Turkey consisted of 44 songs, including improvisations, classical and folk melodies, performed on the "ney" instrument. Music from Iran consisted of a total of 40 religious and folk songs performed by instruments such as the ud, tar, kamanche, ney and the singing voice. Syrian music consisted of 12 songs and improvisations of the instruments ud, qanun, ney and the singing voice. A total of 24 recordings were folk, religious and improvisational music from other Arab countries including Yemen, Oman, Saudi Arabia, Egypt, and Morocco. These consisted of ud, qanun, and vocal recordings performed by various Arab music masters. For western music, a total of 46 recordings were collected consisting of songs and solo improvisations performed on the flute, bassoon, clarinet, oboe and saxophone. The songs performed with the flute were 4 movements from partita for solo flute by Bach, 11 fantasias for solo flute by Telemann, the songs "Syrinx" of Debussy, "Soliloquy for Solo Flute Op. 44" of Lowell Liebermann, "Image for solo flute" of Bozza, "Danse de la chevre" of Arthur Honegger, "Tango Etude" of Piazzolla, "Daphnis et Chloe" of Ravel, and nine solo improvisations played with flute. In our library we used also 3 bassoon solos, 3 pieces for solo clarinet from Stravinsky, two oboe solos, and eight saxophone solos.

3 METHODS

We used a frame-based approach to segment the audio signal into frames of 30ms, resulting in 30000 frames per second. For each frame, 25 low-level features have been extracted and from each feature vector we compute the mean and the standard deviation. We use these values as global features for describing the audio signals. The

low level features used are the following: 1) Zero crossing rate, 2) Spectral centroid, 3) Spectral brightness, 4) Spectral spread, 5) Spectral skewness, 6) Spectral kurtosis, 7) Spectral roll-off, 8) Spectral entropy, 9) Spectral flatness, 10) Spectral roughness, 11) Spectral regularity, 12) Spectral inharmonicity, 13-25) 13 MFCC coefficients. A set of 13 mid-level features has also been extracted gathering information from the pitch histograms. The pitch histograms are computed for a range of two octaves with bin resolution of 1200 per octave. Histogram peaks are detected and each histogram is aligned to the peak of highest amplitude. From the aligned histogram, we extract the 7 highest peaks and use their locations and amplitudes to describe the tonality of every song. Note, the location of the highest peak corresponds always to the first bin of the aligned histogram, and thus omitted. We trained the database with SOM and the k-means algorithm using Euclidean distance and for k=2,3,..,10. The two methods k-means and SOM had similar results, thus we present the results from the k-means algorithm. The elbow method has been applied in order to decide the optimal number of the clusters. For each cluster map we compute the quantization error:

$$J_k = \frac{1}{2}\sum_{\vec{\mu}}^{\vec{i}} \sum_{\vec{x} \in C(\vec{\mu})}^{\vec{j}} \left\| \vec{x}^{\,j} - \vec{\mu}^{\,i} \right\| \qquad (1)$$

where μ are the prototypes and x are the data points. Then we consider the function:

$$R(k) = J(1).k^{\frac{-2}{d}} \qquad (2)$$

where d is the number of dimensions. The function R(k) simulates the behaviour of the quantization error for a data set that is uniformly distributed in the domain occupied by the real data set. Next, we consider the function:

$$D(k) = \frac{R(k)}{J(k)} \qquad (3)$$

The optimal numbers of k are chosen where the function D(k) reaches the maximum value. Figure 1 shows the plot of this function, where we observe that the optimal clusters for this problem are 7.

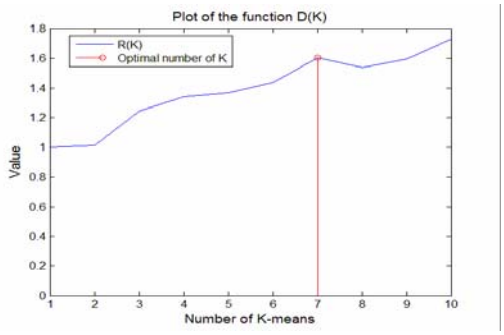

Figure 1: The function D(k). The optimal number of clusters for this problem are 7.

4 RESULTS

Table 1 shows the results from the clustering using the k-means algorithm for 7 clusters using the Euclidian distance. Low-level and mid-level feature sets were used in the results shown in table 1. We observe that Arabic music and music from Iran were clustered in the same cluster, while music from Turkey and Syria were cluster together in another cluster. Music from western was separated in to two clusters and Music from Cyprus was separated in to two clusters. The first cluster contains the Greek Cypriot music while the other contains the Turkish Cypriot music. The models built with low-levels features yield in to similar results and the models built with mid-level features were not able to create a meaningful cluster map.

k=7	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Arabic	0	18	1	0	4	1	0
Iran	0	30	9	0	0	0	1
Syria	0	2	8	0	2	0	0
G_Cypr	0	0	0	0	6	27	4
Turkey	0	7	34	0	0	0	3
West	1	2	5	0	21	0	17
T_Cypr	0	0	4	31	0	0	4

Table 1: Results of the k-means algorithm using the low-level and mid-level features.

5 CONCLUSIONS

This work focuses on the identification of similarities/dissimilarities between western and non-western music. Tonal and timbre features were used for our analysis. Results indicate that the features employed in this study are able to capture particularities of each music tradition. Limitations of this work are considered such as the different amount of recordings per music tradition, and the use of global features that bypass timbre and tonal nuances evolved throughout the time series. It is left to future work to model these explicitly and expand the features and music dataset.

REFERENCES

1. Kallinikos, T.: Kypriaki laiki mousa. Kypriaka Grammata, No. 9: 115- 126, (b) No. 21: 105-113. (1951)
2. Gómez, E., Herrera, P.: Comparative analysis of music recordings from western and non-western traditions by automatic tonal feature extraction. Empirical Musicology Review, Vol. 3, No. 3, pp. 140-156, (2008)
3. Michael, E.: The pithkiavli flute in Cyprus. Master thesis, Free University of Berlin (2008)
4. R. Tibshirani, G. Walther, T. Hastie.: Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(2), 411-423, (2000)