Exploring music collections through automatic similarity visualization

Stefanie Nowak, Christoph Bastuck, Christian Dittmar

Fraunhofer Institute for Digital Mediatechnology, 98693 Ilmenau, Germany,

Email: {nwk, bsk, dmr}@idmt.fraunhofer.de

Introduction

Digital music collections are normally organized in folders, sorted corresponding to artists or genres, forcing the user to navigate through the folder hierarchy to find songs. A completely different paradigm for exploring music collections is the comprehensive search for similar music. Automatic methods from information retrieval as well as similarity ratings from experts exist. Nevertheless, the user still gets similar songs to one query listed and has to rerun the query to find other songs.

To allow for a comprehensive overview of the users music collection, we propose a fast and scalable visualization using the metaphor of a universe. Every song is presented as a star and the closeness of one star to another denotes their similarity. For an intuitive handling, the universe is composed of two dimensions and enables the user to zoom into dense star constellations.

Objective

The focus of the music visualization algorithms lies on the fast visualization of the whole music collection in one scenario. That means, that even for music archives containing thousands of songs the algorithm has to be able to position every song in the universe quickly. This requirement forbids utilizing commonly used methods for similarity visualization like multidimensional scaling (see [1]) or self-organizing maps (compare [2], [3]). Especially when a distance matrix depicting the similarity between each pair of songs is necessary as input, this results in a considerably large amount of computation time and storage needed for a huge music collection.

In this paper we present two methods; one completely based on low and mid-level features, the second one incorporating semantics by spanning the dimensions of the universe in terms of rhythm and instrument density. Both methods are automatically computable, scalable to a huge number of songs and determine the positions of the songs in the visualization space efficiently.

How to position the stars?

We utilize the metaphor of a 2-dimensional universe representing the musical space and stars acting as visual entities for the songs. The user can navigate through this universe finding similar songs arranged closely, sometimes even in star concentrations. There are two interesting research questions regarding this issue:

1. How can the similarity between songs of a whole music collection be determined?
2. Where should the stars, representing a special song, be positioned in the universe, so that the requirement of *closeness encodes similarity* is fulfilled?

Here we focus on the positioning of the stars. For further details concerning music similarity computation see [4].

Method 1: Holistic music visualization

Every song is represented through a song signature. This signature is a vector consisting of different low and mid-level features like MFCCs, Zero-Crossing-Rate or normloudness extracted from the audio signal. Details on the audio features can be found in [4].

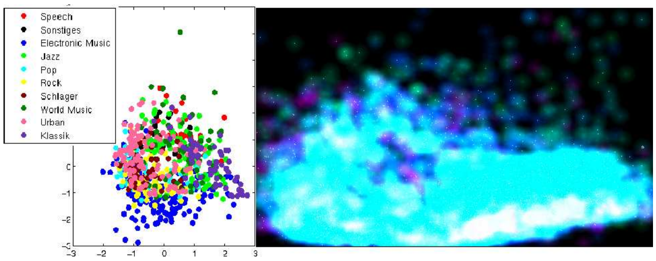

Abbildung 1: Left: Song positions after applying method 1. The plot shows that songs of the same genre are likely to be located near. Right: Final universe visualization of a music collection with more than 15.000 songs

A song is represented through 20 features spanning 806 dimensions. The idea behind this approach is to utilize a method from multivariate statistics to project the feature vectors of the songs into a lower dimensional space. Underlying is the assumption that the song's characteristics can be approximated through this feature vector and that vectors representing similar songs show related characteristics in their features and are therefore located near in the high-dimensional space. We use the Principal Component Analysis (PCA) to determine the three most important principal components and project the feature vectors onto the three resulting eigenvectors. The feature vectors are ordered as a kind of blob in the 3-dimensional space. To deskew the points in the blob, the feature vectors are decentralized from the center. The resulting feature vectors are reduced to two dimensions via PCA and their values denote the coordinates in the universe.

Figure 1 illustrates the results of this procedure. We used a testset of 774 songs sorted into 10 genres. Each point depicts its genre in the color. The genres are used as groundtruth for evaluation and show the ability of the algorithms to locate similar songs close. One has to keep in mind, that not all songs of one genre are necessarily similar to each other, but one can clearly see, that songs of the same genre are clustered by the PCA.

Method 2: Semantic dimensions

The second method focuses on visualizing songs in semantic dimensions. The visualization space is subdivided into several semantic regions. We chose rhythmic characteristics from slow to fast subdivided in five gradations as semantic entity for the x dimension and instrument density from sparse to full in three gradations for the y dimension. So we get a visualization space divided into 15 regions. For example region 1 covers all songs with sparse instrument density and slow tempo characteristics.

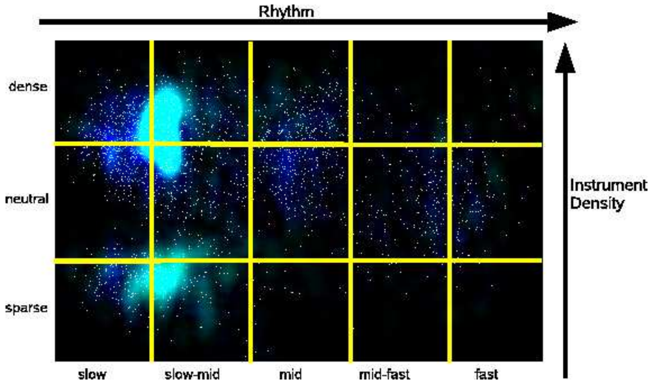

Abbildung 2: Grid partitioning and song positions in the universe with about 15.000 songs after applying method 2.

We created two reference music datasets, one for the rhythmic character and the other for the instrument density. Each reference set contains the feature vectors of three songs per gradation, so altogether fifteen feature vectors in the rhythmic reference set and nine feature vectors in the instrument density reference set. A similarity query from each song of the user's music collection to both reference sets is performed to position the songs. The music similarity retrieval system returns a resultlist with ascending distances between song signatures. So the song at position 1 denotes the most similar song to the query song. For both reference sets the winning song determines the position in the visualization space. In case of a mid-level tempo song as the winning song for the rhythmic dimension, the query song is positioned in the third gradation in x dimension. A neutral instrument density song results in the second gradation in y dimension. The exact position in the subregion is influenced by locally translating each song in the subspace in dependence from the mean and standard deviations of the song positions belonging to the same region.

Conclusion and Future Work

We propose two different techniques for visualizing music collections. Both approaches position similar songs in close proximity. For the user it is possible to explore his music archive and get an overall insight on the different aspects of the music. The introduced algorithms are already integrated in a commercial product for digital management of music archives including a content-based similarity search.

The advantage of the first approach is that all musical attributes inherent in the songs are used for the positioning in the visualization space as far as they are depicted by the applied features. The computationally most expensive step, the singular value decomposition that is necessary for computing the PCA, is independent of the size of the music collection as it is computed on the covariance matrix of the features. Unfortunately, the resulting two dimensions cannot be named by a semantic description. Thus it is not intuitive which musical characteristic changes when moving along the dimensions. As a solution one could integrate the genres as planets into the universe. The stars are then located near genre planets and serve as a guideline for the user which musical characteristics are typical for a specific direction.
In contrast, the second approach allows the exploration of music archives in two semantic dimensions. It is apparent which characteristics change by browsing along its axes. Actually the semantic dimensions could be exchanged by the user himself. It is only necessary to determine new reference sets and their granularity. Provided that the features are able to represent these characteristics, the user could invent his own semantic dimensions in dependence on his personal music taste. However music is not two dimensional, but multidimensional. So it is not possible to define the holistic impression of music along two semantic dimensions like rhythm and instrument density. One has to abstract that the songs are similar in the mentioned dimensions but regarding other musical aspects, neighboured songs can sound very different.

In the future we would like to enhance the considered dimensionality of music with the help of glyphs. So far the stars in the universe are color-coded depending on content-independent metadata. For example the size and the distortion of the star texture can also be taken into account for coding semantic attributes. Learning on a more detailed groundtruth depicting mood characteristics or more complex rhythmic attributes allows for a more sophisticated visualization. Furthermore a detailed evaluation on the quality of the introduced approaches is necessary.

Acknowledgement

This project was partially funded by MAGIX GmbH and is integrated into the MP3 Maker 14. Especially Jan Willkommen contributed to this work with the design of the universe.

Literatur

[1] Lamere,P.; EcK, D.: Using 3D visualizations to explore and discover music. In: Proc. of ISMIR 2007

[2] E. Pampalk et al.: Content-based organization and visualization of music archives. ACM Multimedia, 2002

[3] Leitich, S. and Topf, M.: Globe of Music - Music Library Visualization using Geosom, ISMIR 2007

[4] Dittmar, C.; Bastuck, C.; Gruhne, M.: Novel Mid-Level Audio Features for Music Similarity. Intern. Conf. on Music Communication Science, 2007