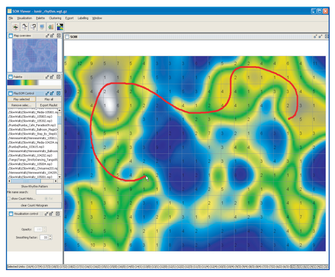
ATLANTIS
Or Towards a Multi-Modal Approach to Music Information Retrieval and its Visualisation

Informatik
Masterstudium
Intelligente Systeme
066 934

Robert Neumayer
neumayer@ifs.tuwien.ac.at
www.ifs.tuwien.ac.at/~neumayer

Institut für Softwaretechnik und
Interaktive Systeme
Information & Software Engineering Group (188-1)
Betreuer: Ao. Prof. Dipl.Ing. Dr.Andreas Rauber

Basics: SOM Clustering

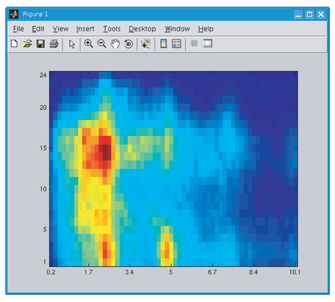


For clustering, the Self-Organising Map, an unsupervised neural network that provides a mapping from a high-dimensional input space to usually two-dimensional output space is used.

The Music Information Retrieval community has been gaining many insights into the area of abstract representations of music by means of audio signal processing. Recommendation engines provide novel ways of creating playlists based on users' preferences. Another im application of audio representation is automatic genre categorisation. However, for many applications audio features alone do not contain enough information. A song's lyrics often describe its genre better than what it sounds like, e.g. `Christmas carols'. Therefore, approaches for the combination of additional data like song lyrics, artist biographies, or album reviews for music recommendation are examined. Further, the application of the SOM for clustering to audio collections with respect to multi-modal feature sets is investigated. Additionally, a new visualisation for simultaneous display of multi-modal clusterings as well as cluster validation metrics are presented.

Basics: Audio Features

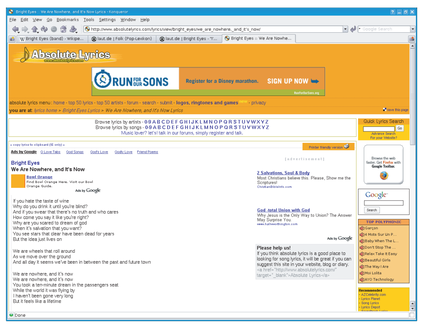
We use a modified version of the Rhythm Patterns features. Basedon that feature set Statistical Spectrum Descriptors yield good results at a lower dimensionality of 168 as compared to the original 1440 feature values.



Basics: Information Retrieval

In classic text categorisation low-level features are computed from a labelled training set of sufficient size. New documents can be assigned to the class represented by the most `similar' documents.

The basic idea is to treat text as a bag of words or tokens. This form of IR abstracts from any kind of linguistic information. Documents are represented as term vectors.

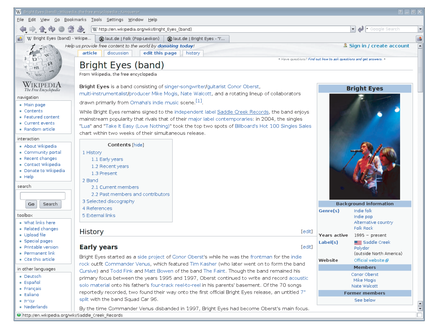


Additional data sources for the audio context:

- Album Reviews
- Genre descriptions
- Artist descriptions
- Song lyrics

Main research questions are:

- Multi-Modal similarity ranking
- Multi-Modal cluster visualisation

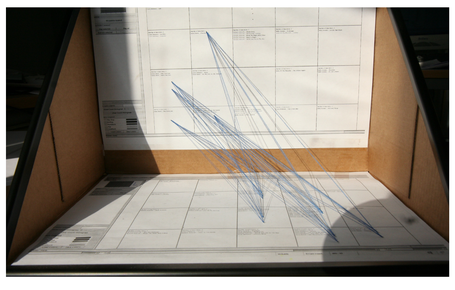


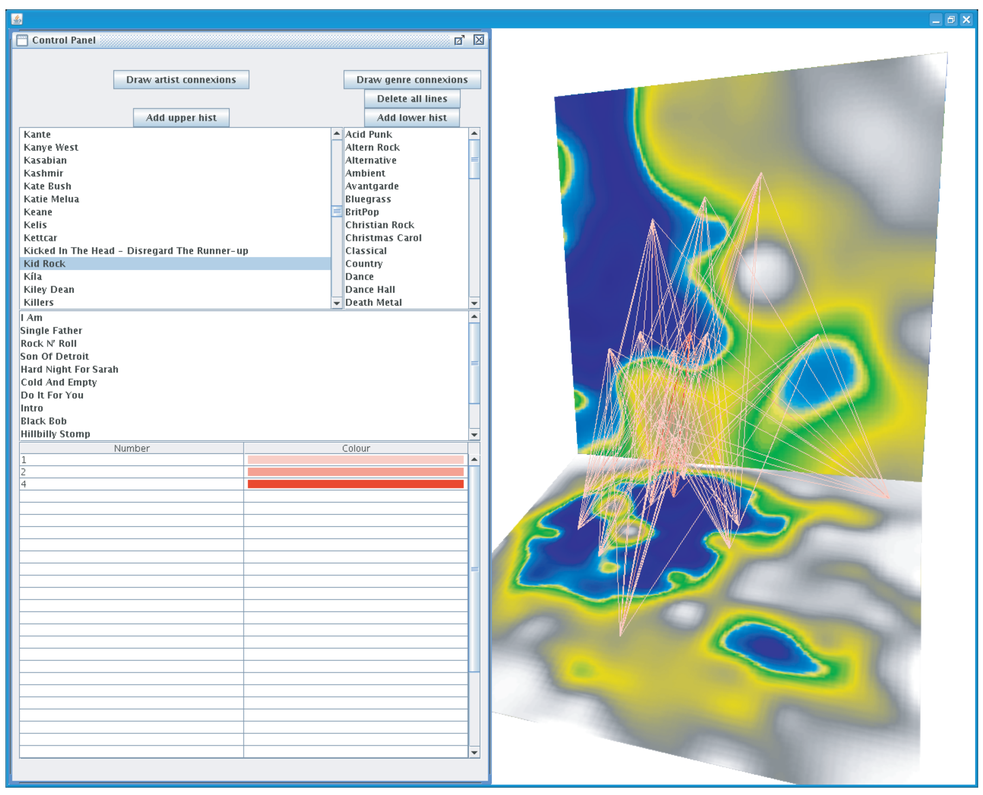
For genre classification and multi-modality experiments we used a parallel corpus of audio and song lyrics files for a music collection of 7554 titles organised into 52; genres were assigned manually. The collection contains songs from 644 different artists and 931 albums. Additional textual information was retrieved from the Internet for genre, artist, and albums.

Multi-Modal Visualisation

Identical instances, ie.e songs, mapped onto multiple clusterings can be linked.

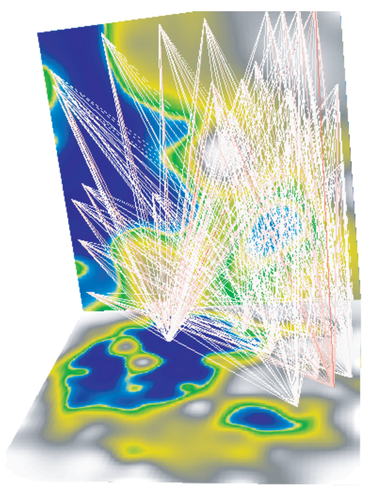
A prototype mock-up was `engineered' with paper, carton, and sewing cottons.



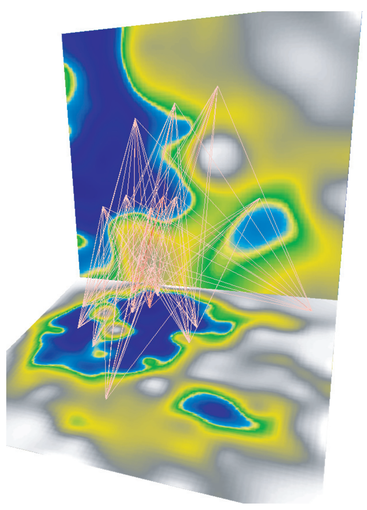


The main user interface of the Java implementation. The two SOMs are displayed on the right part. The 3D display offers the following functionalities:

- rotate the view
- pan and zoom in or out.
- select particular songs, artist or genres
- change the colour palette

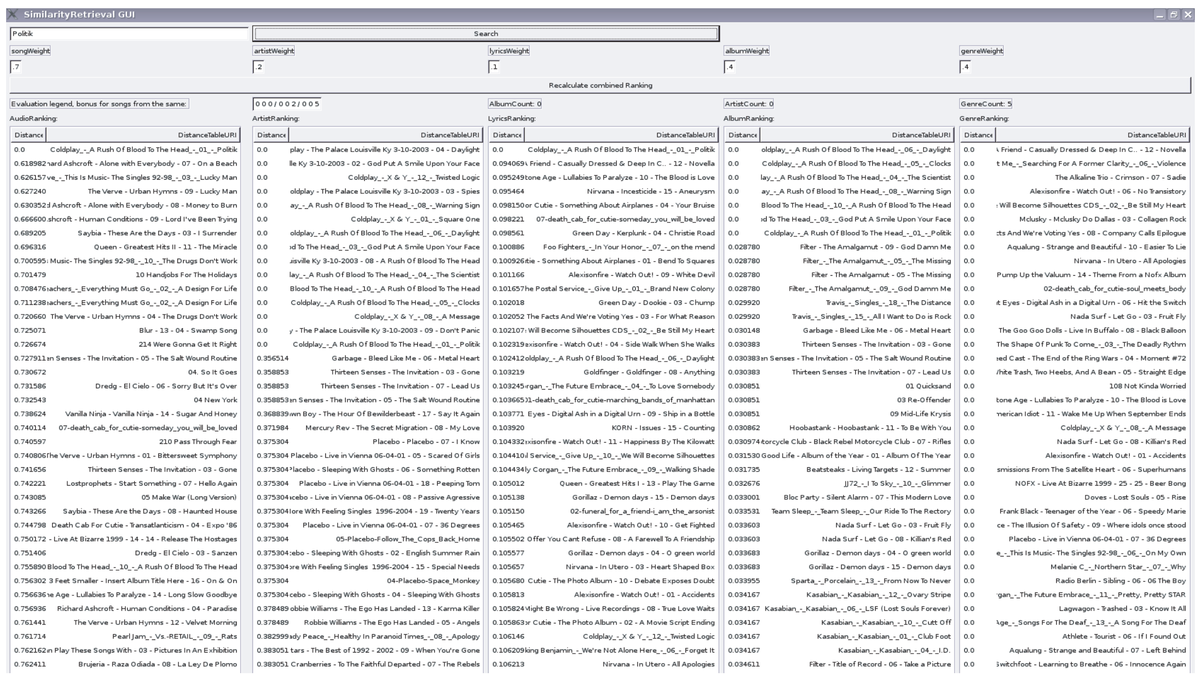


Connections for the 'Christmas Carol' genre with a much tighter distribution in the audio space.



Connections for the artist 'Kid Rock', the distributions are similar in both feature spaces..

Multi-Modal Similarity Ranking



The main user interface of an experimental system to evaluate the impact of different weighting strategies. The largest part of the GUI is occupied by the five different rankings, one for audio, artist, album, lyrics, and genre respectively.

Experimental and Qualitative Evaluation

Experiments were performetd to assess the quality of the multi-modal clustering, similarity ranking, as well as genre classification.

Genre	AC	LC	CR	ADR	ADR×CR
Speech	.0822	.0665	.8092	.3417	.2765
Christmas Carols	.0393	.0677	.5800	.7779	.4512
Reggae	.0392	.0413	.9495	.8475	.8047
Grunge	.0382	.0466	.8204	.9974	.8182
Rock	.0372	.0382	.9740	.9300	.9059

Same Album	**Top 5**	**Top 10**	**Top 20**
Category Subst.	1.84	2.79	3.52
Exclusion	1.91	2.78	3.55
Simple Avg.	2.36	3.41	4.07
Same Artist	**Top 5**	**Top 10**	**Top 20**
Category Subst.	2.43	4.09	5.83
Exclusion	2.41	3.95	5.53
Simple Avg.	2.97	5.11	7.15
Same Genre	**Top 5**	**Top 10**	**Top 20**
Category Subst.	1.55	2.85	5.15
Exclusion	1.64	2.87	4.92
Simple Avg.	1.90	2.91	4.43

We train one map representing the collection in terms of lyric similarity, one in terms of audio similarity. Those maps will henceforth be referred to as audio and lyrics map, respectively. Quality measures are given for a selection of genres, describing the differences in distributions across maps.

At first, the combined distances for each track in the collection to all other songs are computed. Then the first 5, 10 and 20 results are evaluated according to the number of songs belonging to the same artist, genre, and album.

For additional information and publications visit:
http://www.ifs.tuwien.ac.at/mir
http://www.ifs.tuwien.ac.at/~neumayer