Multi-Modal Music Information Retrieval - Visualisation and Evaluation of Clusterings by Both Audio and Lyrics

Robert Neumayer and Andreas Rauber
{**neumayer,rauber**}**@ifs.tuwien.ac.at**

Vienna University of Technology
Institute for Software Technology and Interactive Systems
Favoritenstraße 9-11, 1040, Vienna, Austria

presented at RIAO'07
May 30, 2007

motivation for multi-modal analysis

generally increasing amount of digital audio
- private users
- commercial holdings

novel interfaces & music classification
- clustering/maps: PlaySOM, PocketSOMPlayer
- classification into categories: genres, emotions, situations,...

explore the influence of relevant information
- christmas songs
- love songs
- spoken documents

motivation for multi-modal analysis

generally increasing amount of digital audio
>   private users
>   commercial holdings

novel interfaces & music classification
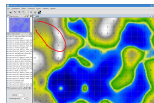>   clustering/maps: PlaySOM,
>   PocketSOMPlayer
>   classification into categories: genres,
>   emotions, situations,...

explore the influence of relevant information

>   christmas songs

>   love songs

>   spoken documents

motivation for multi-modal analysis

generally increasing amount of digital audio

     private users

     commercial holdings

novel interfaces & music classification

     clustering/maps: PlaySOM,

     PocketSOMPlayer

     classification into categories: genres,

     emotions, situations,...

explore the influence of relevant information

     christmas songs

     love songs

     spoken documents

introduction   music information retrieval and ml   multi-modality in mir   conclusions
○●
           ○○○○○○
                       ○○○○○○○○○○
                            ○

what to expect from this presentation

information retrieval
> text ir
> music ir

integration of both audio and text features

clustering
> self-organising map (som)
> multi-modal clustering

user interface

multi-modal cluster evaluation

audio features

computed from the audio waveform

abstract representation

can be computed for every piece of audio

a few feature sets

MPEG7 standard features

MARSYAS features

rhythm patterns (1440)

rhythm histograms (60)

statistical spectrum descriptors (168)

introduction
oo

music information retrieval and ml
o●oooo

multi-modality in mir
oooooooooo

conclusions
o

text features

plain text lyrics retrieval
    three lyrics portals are accessed
    missing values issues (e.g. lyrics cannot be retrieved)

'bag-of-words' approach
    stop word removal: yes
    stemming: no

tfidf weighting

still abstract, may yet be helpful
    interpretability
    content words / semantic categories

high dimensionality $\rightarrow$ reduction needed

self-organising map clustering

unsupervised neural network model

data mapping

> from high-dimensional input space
> to low-dimensional output space

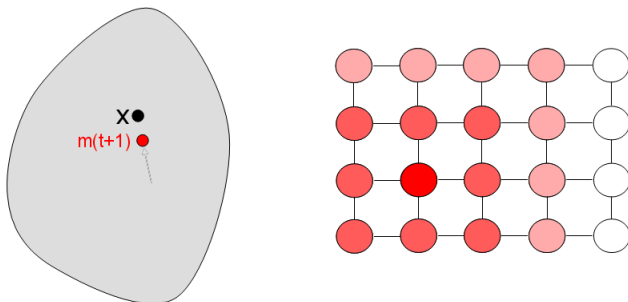topology preservation

simplification and visualisation

introduction
oo

music information retrieval and ml
ooo●oo

multi-modality in mir
oooooooooo

conclusions
o

som training process

Figure: self-organising map training algorithm
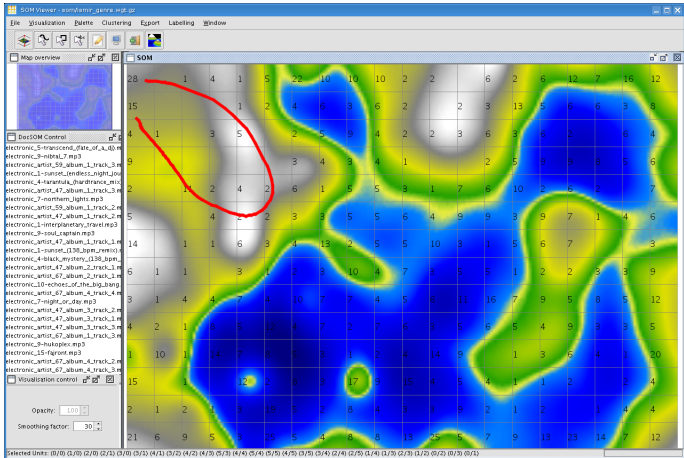
map based user interfaces

Figure: PlaySOM application

clustering music by lyrics

audio/lyrics collection (7500 songs / 54 genres)
som of 20×20 units
comprises a range of styles and genres:
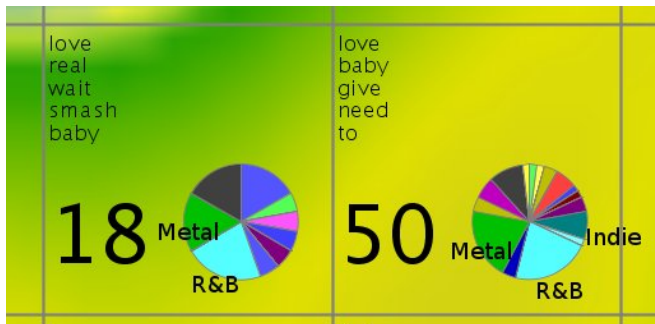   metal, r&b, indie , . . .

Figure: clustering of songs centred around the love topic

clustering and multiple data sources

why not cluster according to each modality?

connect instances/songs on both maps

identify differences in the data distributions on the map across clusterings

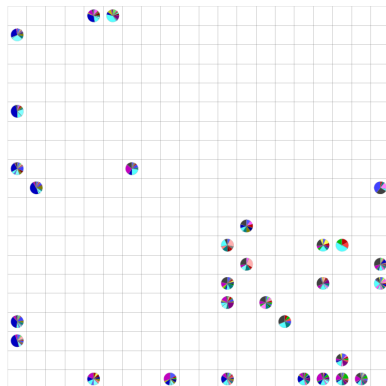# genre-wise distribution across mappings

audio clustering

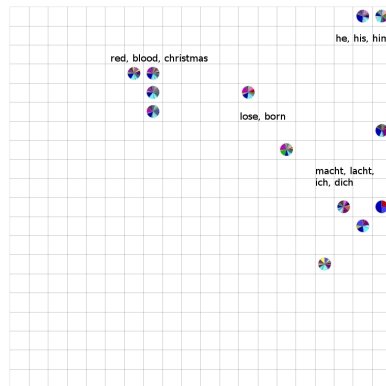Figure: clustering of Christmas
songs on the 2D audio map

lyrics clustering

Figure: clustering of Christmas
songs on the 2D lyrics map

introduction
oo

music information retrieval and ml
oooooo

multi-modality in mir
ooo●oooooooo

conclusions
o

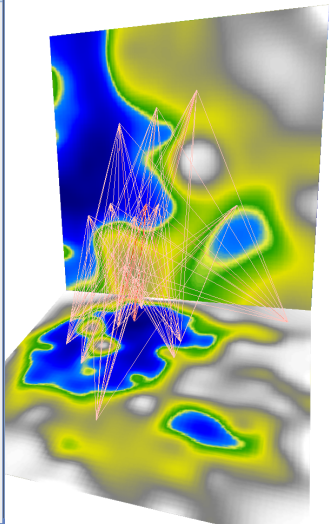Figure: full view of the visualisation prototype

Figure: full view of the visualisation prototype – the vertical map clusters songs by audio features, the horizontal map is trained on lyrics features

introduction
oo

music information retrieval and ml
oooooo

multi-modality in mir
oooo●ooooo

conclusions
o

detailed view of
connections

equally distributed
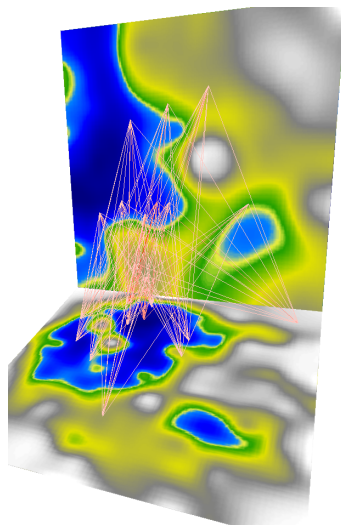artist 'Kid Rock'

colour-coded
connections

Figure: Kid Rock's songs

detailed view of connections

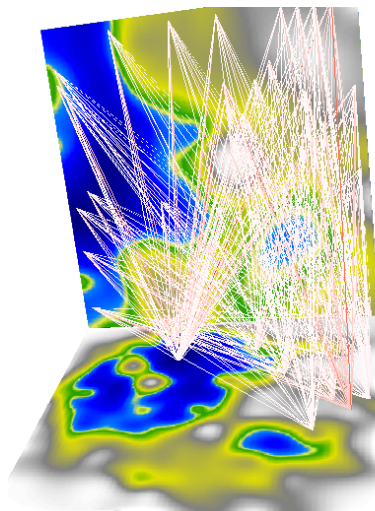genre 'Christmas Carols'

colour-coded connections

Figure: Christmas songs

introduction
oo

music information retrieval and ml
oooooo

multi-modality in mir
ooooooo●ooo

conclusions
o

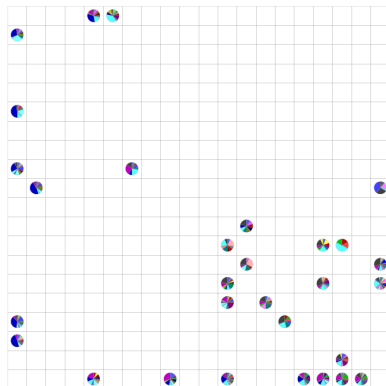# distribution across mappings

audio clustering

Figure: clustering of Christmas songs on the 2D audio map
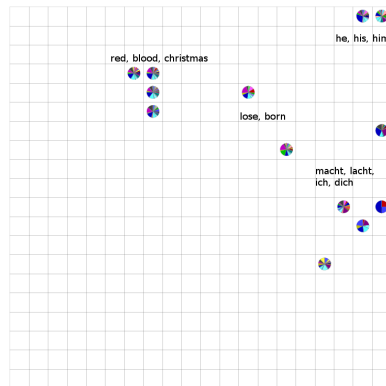
lyrics clustering

Figure: clustering of Christmas songs on the 2D lyrics map

quantitative evaluation

select instances belonging to one artist/genre

compute spreading factors on each map considering individual clusterings

integrate these values for both maps

quantitative evaluation by example

1	1		
1	1		

(a) Upper left corner (audio)

1			
	1		
		1	
			1

(b) Diagonally (lyrics)

average distance ratio

contiguity ratio

    bonus for continuous clusters

1	1		1	
		1		
			1	

(c) Non-Continuous (audio)

1	1		
			1
			1

(d) Sub-clusters (lyrics)

	$adr_{a,l}$	$cr_{a,l}$	$adr \cdot cr$
a/b	.48	.20	.06
c/d	.76	.73	.55

genre/artist-wise distribution measures

Artist	CR	ADR	ADR×CR
Sean Paul	.4152	.4917	.2042
Good Riddance	.8299	.7448	.6181
Shakespeare	.2626	.3029	.0795
Kid Rock	.9640	.9761	.9410

Genre	CR	ADR	ADR×CR
Speech	.8092	.3417	.2765
Christmas Carol	.5800	.7779	.4512
Reggae	.9495	.8475	.8047
Rock	.9740	.9300	.9059

recap

multi-modal clustering

plus evaluation

possible usage

>   artist genre identification
>   additional info for music information retrieval systems
>   quality metrics for cluster evaluation (focusing on music
>   context)

. . . and we'll be hosting the ISMIR conference in September!