TOWARDS A SOCIO-CULTURAL COMPATIBILITY OF MIR SYSTEMS

Stephan Baumann
German Research Center for Artificial Intelligence
Erwin Schrödinger Str.
67663 Kaiserslautern
Germany

Tim Pohle
German Research Center for Artificial Intelligence
Erwin Schrödinger Str.
67663 Kaiserslautern
Germany

Vembu Shankar
Technical University of Hamburg

21071 Hamburg
Germany

ABSTRACT

Future MIR systems will be of great use and pleasure for potential users. If researchers have a clear picture about their "customers" in mind they can aim at building and evaluating their systems exactly inside the different socio-cultural environments of such music listeners. Since music is in most cases embedded into a socio-cultural process we propose especially to evaluate MIR applications outside the lab during daily activities. For this purpose we designed a mobile music recommendation system relying on a trimodal music similarity metric, which allows for subjective on-the-fly adjustments of recommendations. It offers online access to large-scale metadata repositories as well as an audio database containing 1000 songs. We did first small-scale evaluations of this approach and came to interesting results regarding the perception of song similarity concerning the relations between sound, cultural issues and lyrics. Our paper will also give insights to the three different underlying approaches for song similarity computation (sound, cultural issues, lyrics), focusing in detail on a novel clustering of album reviews as found at online music retailers.

Keywords: Socio-cultural issues in MIR, multimodal song similarity, ecological validation.

1. INTRODUCTION

We propose a socio-cultural compatibility of MIR systems and achieved promising results by evaluating such an application in the field of mobile music recommendations. We included the following aspects:

1. The musical work of artists is examined from the perspective of a music-consuming society.
2. Optionally, users may add personal information about age, gender, musical education, personal taste which reflects belonging to social peer groups.
3. Subjective music-listening behavior in socio-cultural environments is collected and evaluated with an ecological approach.
4. Long-term observations are undertaken using a plugin for Winamp MP3 software.
5. Aspects of the artist's creative intention being partially represented in sound, orchestration, production environment, selection of singer and lyrics are covered by audio analysis and information retrieval methods.

We are well aware of the fact that such a holistic approach needs for a significant amount of research. Nevertheless other authors [1] have proposed similar approaches emphasizing the socio-cultural dimension. Our activities and the presented paper focus on the aspects (1) and (3) (in contrast to our previous publication [2] which included no details about the clustering techniques). Point (5) is described very shallow and (2), (4) are considered in future work.

Figure 1. Ecological evaluation.

2. RELATED WORK

Our research asks how we might add to our understanding of perception of music similarity through an 'ecological' approach. This means studying how people perceive music similarity in their normal lives beyond the artificial world of lab-based experiments. To this end we want to find new ways of observing users' interaction with our systems as they go about their everyday activities. Cognition in the wild means studying cognitive phenomena in the natural contexts in which they occur. This approach relates to the insight that what people do in labs may not be 'ecologically valid': experimental results may be artefacts of the lab situation, failing to represent people's behaviour in the 'ecologies' of their normal lives. While the lab-based approach can tell us about perception of music similarity [3], we feel it is also important to look beyond the lab and its artificial experimental setups, to music users' spontaneous perception of music similarity in real situations as part of their everyday lives. This ecological approach might reveal, for example, how perception changes with time, location, or activity, in ways, which could have implications for how systems generate recommendations. For this purpose we designed a mobile music recommendation system relying on a trimodal music similarity metric, which allows for subjective adjustments on the fly. We did first small-scale evaluations of this approach and came to interesting results regarding the perception of song similarity concerning the relations between sound, cultural issues and lyrics. We will introduce this multimodal similarity metric in the following two sections and present the results in section 5.

3. MULTIMODAL SONG SIMILARITY

Our multimodal song similarity measure is realized as a weighted linear combination of three different local similarity metrics, namely timbral similarity, similarity of lyrics and cultural similarity:

$$S = wso * Sso + wly * Sly + wst * Sst \qquad (1)$$

This section will give insights to the three different underlying approaches for similarity computation.

3.1. Timbral similarity

The computation of timbral similarity has meanwhile a long tradition in the MIR community. Objective evaluations based on genre, artist and album metadata have been performed and also being compared against each other by different authors [4,5]. A basic finding is that MFCCs and a GMM or k-means clustering and Earth Moving Distance behave pretty well for predicting similar sounding songs for given anchor songs.

# of Neighbours	#of songs in the same album	#of songs of the same artist	# of songs in the same genre
1	0,30	0,41	0,45
3	0,75	0,99	1,17
5	1,05	1,40	1,78

Table 1. Experimental results of our best timbral operator: 13 MFCCs, 16 clusters, EMD-KL (see [4]).

We added to our own operator bank recently an approach relying on ICA of spectral features to see if we could increase previous performance. We could not achieve better results according to the objective evaluation. There seems to be an upper limit which was reported also in a recent evaluation by Aucouturier [6].

3.2. Cultural similarity

The processing of cultural descriptions in the context of MIR systems has been introduced by [7]. In our previous work we implemented a minor improvement of this original work to generate artist recommendations [2]. From our findings we decided to expand this idea by accessing album reviews from the Amazon web site and apply a clustering instead of using a simple vector space model for similarity computation between artists. The basic idea of our approach is to spatially organize these reviews that are in the form of textual documents using an unsupervised learning algorithm called Self-Organizing Maps (SOM) [8] and thus be able to give recommendations for similar artists by making use of the model built by the algorithm. The use of SOMs in the field of text mining and for audio-based MIR applications is well understood [9]. The SOM is an unsupervised learning algorithm used to visualize and interpret large high-dimensional data sets. The map consists of a regular grid of processing units called “neurons”. Each unit is associated with a model of some high dimensional observation represented by a feature vector. The map attempts to represent all the available observations with optimal accuracy using a restricted set of models. Map units that lie nearby on the grid are called *neighbours*. After the formation of a map for a particular data set, the model vectors are arranged in such a manner that nearby map units represents similar kind of data and distant map units represent different kinds of data. The literature on Information Retrieval provides techniques to pre-process and represent textual documents for mining operations. The documents are then represented in the form of a bag-of-words where each document is considered as a point (or vector) in an n-dimensional Euclidean space where each dimension corresponds to a word (term) of the vocabulary. The i^{th} component d_i of the document vector expresses the number of times the word with index i occurs in the document, or a function of it. Furthermore, each word can be assigned a weight signifying its importance. Commonly used weighting strategy is the tf * idf (term frequency – inverted document frequency) scheme, e.g. by a variant such as

$$w_{ij} = tf_{ij} * idf_i = tf_{ij} * \log_2 (N/ df_i) \qquad (2)$$

where

w_{ij} is the tf-idf weight for the i^{th} word in j^{th} document in a collection of N documents,

tf_{ij} is the term frequency of the i^{th} word in the j^{th} document and

$idf_i = \log_2 (N/df_i)$ is the inverse document frequency of the i^{th} word over the entire collection.

The tf * idf weighting scheme described above does not take into consideration any domain knowledge to determine the importance of a word. But when trying to find similarities between two documents in a musical context, it is desirable to exploit any domain knowledge that is inherently present in the documents. We propose one such mechanism to accomplish this by introducing the concept of a modified weighting scheme in the musical domain or context. Therefore, in addition to the weighting importance given to a word by the tf *idf scheme, it would be worthwhile to increase the weight of a word by a certain factor if it is pertaining to the musical domain. We came up with such a word list of 324 from the genre taxonomies of the *All Music Guide*. The modified weighting scheme gives rise to a new weight for musical words that is given by

$$w^{m}_{ij} = tf^{m}_{ij} * idf^{m}_{i} = tf^{m}_{ij} * \log_2 (N/df^{m}_{i}) * \alpha \qquad (3)$$

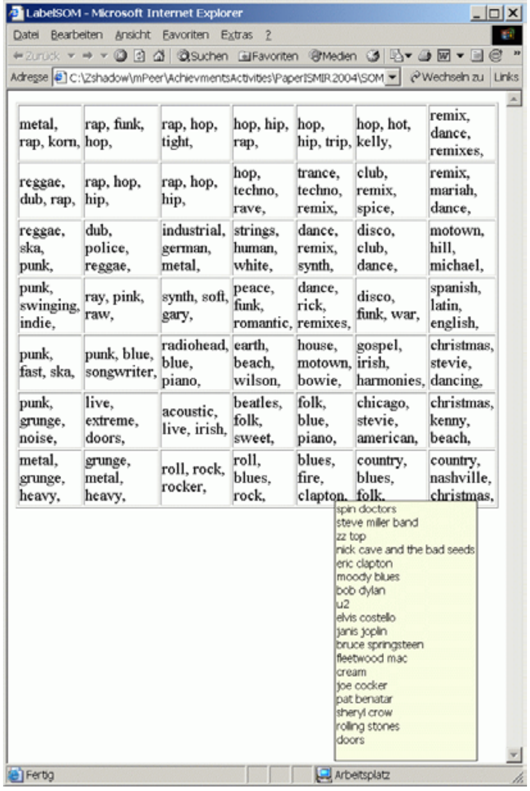
where the superscript m indicates words belonging to the musical context and α is the weighting increment. The pre-processed textual documents represented in the form of n-dimensional vectors can be used to train a SOM in an unsupervised way. The learning starts with a set of reference vectors also called the model vectors that are the actual map units of the network. As the learning proceeds, the model vectors gradually change or arrange themselves so as to approximate the input data space. The final arrangement is such that the model vectors that are nearby are similar to each other. The model vectors are usually constrained to a two-dimensional regular grid, and by virtue of the learning algorithm, follow the distribution of the data in a non-linear fashion. The model vectors are fitted using a sequential regression process. Given a sample vector x(t) at iteration step t the model vector $m_i(t)$ with index *i* is adapted as follows:

$$m_i(t + 1) = m_i(t) + h_{c(\mathbf{x}),i}(t)[x(t) - m_i(t)] \qquad (4)$$

where the index of the "winner" model, c for the current sample is identified by the condition,

$$\forall i,\ \|x(t) - m_c(t)\| \leq \|x(t) - m_i(t)\| \qquad (5)$$

$h_{c(\mathbf{x}),i}(t)$ is called the *neighborhood function*, which acts as a smoothing kernel over the grid, centered at the "winner" model $m_c(t)$ of the current data sample. The neighborhood function is a decreasing function of the distance between the *i*th and *c*th nodes on the map grid. The regression is usually reiterated over all the available samples. Thus, with this unsupervised learning algorithm we can spatially arrange all the documents i.e. the album reviews of all the artists, resulting in a topological ordering of the artists. In addition to this, the SOM algorithm also obtains a clustering of the data onto the model vectors wherein the artists present in a particular cluster are similar to each other. Labeling plays an important role in the visualization of the SOM. We employed a simple labeling technique where a map unit is represented by a label or a keyword that has a higher weight, as calculated by the tf * idf weighting scheme, when compared to other words that appear in the map unit. The modified weighting scheme described in the previous sections also aided in labeling the map units. Since we increase the weight of a word that pertains to the musical context, many, if not all, of the labels that we obtained were from the musical list of words. This is indeed desirable when we are labeling an SOM of artists as we would like to see labels that are musical words like *rap, rock, metal, blues* and not plain English words. Figure 2 shows a few distinct sections of the map with their respective labels. As can been seen from the results, we were able to obtain a clear categorization of artists based on different musical genres.


LabelSOM - Microsoft Internet Explorer
Datei Bearbeiten Ansicht Favoriten Extras ?
Zurück
Suchen
Favoriten
Medien
Adresse
C:\Zshadow\mPeer\AchievmentsActivities\PaperISMIR2004\SOM
Wechseln zu
Links
Fertig
Arbeitsplatz

metal, rap, korn,	rap, funk, hop,	rap, hop, tight,	hop, hip, rap,	hop, hip, trip,	hop, hot, kelly,	remix, dance, remixes,
reggae, dub, rap,	rap, hop, hip,	rap, hop, hip,	hop, techno, rave,	trance, techno, remix,	club, remix, spice,	remix, mariah, dance,
reggae, ska, punk,	dub, police, reggae,	industrial, german, metal,	strings, human, white,	dance, remix, synth,	disco, club, dance,	motown, hill, michael,
punk, swinging, indie,	ray, pink, raw,	synth, soft, gary,	peace, funk, romantic,	dance, rick, remixes,	disco, funk, war,	spanish, latin, english,
punk, fast, ska,	punk, blue, songwriter,	radiohead, blue, piano,	earth, beach, wilson,	house, motown, bowie,	gospel, irish, harmonies,	christmas, stevie, dancing,
punk, grunge, noise,	live, extreme, doors,	acoustic, live, irish,	beatles, folk, sweet,	folk, blue, piano,	chicago, stevie, american,	christmas, kenny, beach,
metal, grunge, heavy,	grunge, metal, heavy,	roll, rock, rocker,	roll, blues, rock,	blues, fire, clapton,	country, blues, folk,	country, nashville, christmas,

spin doctors
steve miller band
zz top
nick cave and the bad seeds
eric clapton
moody blues
bob dylan
u2
elvis costello
janis joplin
bruce springsteen
fleetwood mac
cream
joe cocker
pat benatar
sheryl crow
rolling stones
doors

Figure 2. SOM in HTML format: 7x7 grid, highlighted is the rectangle with similar artists of the unit labeled with *blues,fire,clapton [7th row, 5th column]*.

We show in the following two examples of the data feed from Amazon. We used uppercase to indicate words, which appear as labels, and words in bold to indicate the occurrence of the word in the crisp feature set:

[Artist: Eric Clapton Album: Unplugged]
Clapton caught the "unplugged" trend just at the right time, when the public was hungry to hear how well ROCK stars and their material can hold up when stripped of elaborate production values. Clapton himself seemed baffled by the phenomenon, especially when picking up the armload of Grammys Unplugged earned him, including Record and Song of the Year for "Tears in Heaven," the heart-rending elegy to his young son, Conor. That song and a reworked version of "Layla" got most of the attention, but the rest of the album has fine versions of ***acoustic*** *BLUES numbers such as "Malted Milk," "Rollin' & Tumblin', and "Before You Accuse Me" that make it worth investigating further.*

[Artist: Bob Dylan Album: Highway 61 Revisited]
*****3/4 "Highway 61 Revisited" is an amazingly original record which finds Bob Dylan moving effortlesly between* ***folk-*** *ROCK, BLUES and flat-out garage ROCK. The songs are among the best and most energetic in Dylan's catalogue, and the band, which features Michael Bloomfield, Al Kooper and BLUES drummer Sam Lay, careen through the* ***classic*** *"Like A*

Rolling Stone", the acerbic "Ballad Of A Thin Man", the stylish BLUES "It Takes A Lot To Laugh, It Takes A Train To Cry", and the blistering "Highway 61".One of a handful of truly essential mid-60s ROCK records, and one of Bob Dylan's very best and most cohesive albums.

The results of our experiments were validated using *www.echocloud.net*, a web-based artist recommendation engine. It works by crawling peer-to-peer networks (*soulseek, gnutella*) to capture users' file lists in order to discover correlations between musical artists. The similarity model uses a database of around 120k artists, which represents kind of *open world* approach in contrast to our reduced test set of 398 different artists. We compared the Top 10 recommendations from Echocloud with our Top 10 recommendations for all the artists. We also compared Echocloud recommendations with the artists that are present in the 3 and 5 Best Matching Units (BMUs) of the artist in question. A Best Matching Unit for an artist is the SOM map unit that best models the set of according textual album reviews. The Euclidean distance measure is used in finding the BMUs for an artist.

	Total number of recommendation matches	Average recommendation match in percentage
Top 10	482 / 3980	12.1 %
3 BMUs	685 / 3980	17.2 %
5 BMUs	982 / 3980	24.7 %

Table 2. Validations of our results for 398 artists with Echocloud's Top 10 recommendations without modified weighting scheme

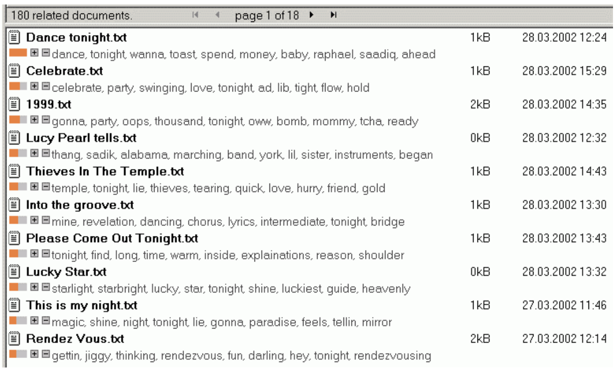
	Total number of recommendation matches	Average recommendation match in percentage
Top 10	493 / 3980	12.4 %
3 BMUs	785 / 3980	19.7 %
5 BMUs	1038 / 3980	26.1 %

Table 3. Validations of our results for 398 artists with Echocloud's Top 10 recommendations with modified weighting scheme

As can be seen from the results in Tables 2 and 3, the quality of the recommendations increases by using the modified weighting scheme incorporating domain knowledge. Nevertheless the performance gain is rather moderate which could be an indication that we already reached the upper bound of such kind of approaches. In order to use this artist recommendation approach as a cultural facet for the song-based similarity engine, we decided to implement a simple inference step: the similarity values between two artists are propagated to each of their songs. In this way they are accessible for the multimodal similarity computation at song level. In the future we plan to cluster album reviews separately by the SOM to deliver a more fine-grained similarity model to propagate song recommendations. We are well aware of the fact that in the long-term we should incorporate an ontology-based reasoning engine at this stage. The explicit semantic relationships between the artist, album and song as basic ontological entities and their diverse sub-categories (e.g. "compilations", "concept albums", "best of <artist, decade, genre>albums") are rather difficult to map onto this approach.

3.3. Similarity of lyrics

The last aspect of our multimodal similarity engine covers the aspect of song lyrics. We included the same approach as presented in our previous work [2]. It is based on the standard tf*idf weighting to represent lyrics as document vectors (see formula 2) and the cosine metric to perform similarity computation.



180 related documents. page 1 of 18

Dance tonight.txt dance, tonight, wanna, toast, spend, money, baby, raphael, saadiq, ahead	1kB	28.03.2002 12:24
Celebrate.txt celebrate, party, swinging, love, tonight, ad, lib, tight, flow, hold	1kB	28.03.2002 15:29
1999.txt gonna, party, oops, thousand, tonight, oww, bomb, mommy, tcha, ready	2kB	28.03.2002 14:35
Lucy Pearl tells.txt thang, sadik, alabama, marching, band, york, lil, sister, instruments, began	0kB	28.03.2002 12:32
Thieves In The Temple.txt temple, tonight, lie, thieves, tearing, quick, love, hurry, friend, gold	1kB	28.03.2002 14:43
Into the groove.txt mine, revelation, dancing, chorus, lyrics, intermediate, tonight, bridge	1kB	28.03.2002 13:30
Please Come Out Tonight.txt tonight, find, long, time, warm, inside, explainations, reason, shoulder	1kB	28.03.2002 13:43
Lucky Star.txt starlight, starbright, lucky, star, tonight, shine, luckiest, guide, heavenly	0kB	28.03.2002 13:32
This is my night.txt magic, shine, night, tonight, lie, gonna, paradise, feels, tellin, mirror	1kB	27.03.2002 11:46
Rendez Vous.txt gettin, jiggy, thinking, rendezvous, fun, darling, hey, tonight, rendezvousing	2kB	27.03.2002 12:14

Figure 3. *Profiler* application.

From the technical point of view we used this time a commercial tool providing such standard information retrieval functionalities via a JAVA programming interface. The out-of-the-shelf application was used for fast subjective evaluation during implementation phase (see Figure 3 for an example of some similar "dancy" lyrics to the query song *dance tonight*).

4. SERVICE-ORIENTED ARCHITECTURE

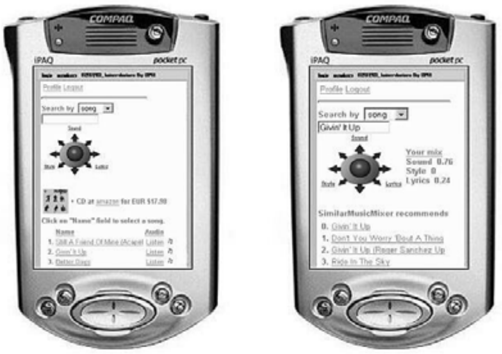
We selected a service-oriented architecture in order to combine our internal services and external services from Amazon and Echocloud. The album reviews for artists can be accessed from the Amazon site using the Amazon Web Service interface that is available as a standard development kit. It supports either web service SOAP messages or using XML over HTTP. Echocloud offers an XML-based web service for artist recommendation, which we also included into our framework. In addition to the abovementioned services we plan to integrate further symbolic and semantic web-related services in the near future. By following this strict architecture we are able to offer our own multimodal similarity engine to be invoked from external services via standard web service description.

4.1. Wireless LAN for ecological evaluation

For an embedding of the evaluation into most natural music listening situations of everyday people we decided to equip people with a mobile device being connected to the described server. Since several hotspots offering free wireless LAN access are available at our campus site and in the city we only had to build a standard browser-based application to set up the prerequisites for the ecological evaluation.

4.2. MYMO: Mobile application

The central web site of our services has been optimized for small screen sizes of PDA devices.

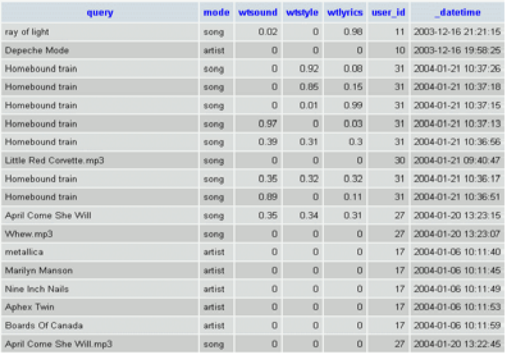


Profile Logout
Search by song
Sound
Style
Lyrics
CD at amazon for EUR $17.98
Click on "Name" field to select a song.
Name Audio
1. Still A Friend Of Mine (Acapel Listen
2. Givin' It Up Listen
3. Better Days Listen

Profile Logout
Search by song
Givin' It Up
Sound
Your mix
Sound 0.76
Style 0
Lyrics 0.24
Style
Lyrics
SimilarMusicMixer recommends
0. Givin' It Up
1. Don't You Worry 'Bout A Thing
2. Givin' It Up (Roger Sanchez Up
3. Ride In The Sky

Figure 4. MYMO application

It is possible to search for artists, albums and song titles as well as accessing individual items in a top-down selection mode. To allow the user interactive feedback in the song recommendation mode, a virtual joystick was included that can be easily accessed using the pen of the PDA. The recommendation engine uses the song similarity measure described above. The position of the joystick has a direct influence on the individual weights in the linear combination. In this way the user can select different settings and find his favorite combination on the fly. The logging at the server allows for storage of the individual interactions with the device.



query	mode	wtsound	wtstyle	wtlyrics	user_id	_datetime
ray of light	song	0.02	0	0.98	11	2003-12-16 21:21:15
Depeche Mode	artist	0	0	0	10	2003-12-16 19:58:25
Homebound train	song	0	0.92	0.08	31	2004-01-21 10:37:26
Homebound train	song	0	0.85	0.15	31	2004-01-21 10:37:18
Homebound train	song	0	0.01	0.99	31	2004-01-21 10:37:15
Homebound train	song	0.97	0	0.03	31	2004-01-21 10:37:13
Homebound train	song	0.39	0.31	0.3	31	2004-01-21 10:36:56
Little Red Corvette.mp3	song	0	0	0	30	2004-01-21 09:40:47
Homebound train	song	0.35	0.32	0.32	31	2004-01-21 10:36:17
Homebound train	song	0.89	0	0.11	31	2004-01-21 10:36:51
April Come She Will	song	0.35	0.34	0.31	27	2004-01-20 13:23:15
Whew.mp3	song	0	0	0	27	2004-01-20 13:23:07
metallica	artist	0	0	0	17	2004-01-06 10:11:40
Marilyn Manson	artist	0	0	0	17	2004-01-06 10:11:45
Nine Inch Nails	artist	0	0	0	17	2004-01-06 10:11:49
Aphex Twin	artist	0	0	0	17	2004-01-06 10:11:53
Boards Of Canada	artist	0	0	0	17	2004-01-06 10:11:59
April Come She Will.mp3	song	0	0	0	27	2004-01-20 13:22:45

Figure 5. User logs.

5. ECOLOGICAL, SUBJECTIVE EVALUATION

A group of 10 subjects reflecting the current distribution of Internet users in Germany was selected by means of varying gender, age, education, and musical background. We used a within-subjects design with two conditions: the lab, and 'the wild'. In each condition, each subject was asked to find the optimal joystick setting that would return an acceptable block of 5 recommendations for a given anchor song. This position produces a particular trimodal weighting. Subjects were instructed that if they did not like the results of a given weighting, they could change that weighting immediately to produce an alternative result. They were also asked, if they liked the results of the weighting to select their favourite recommendation. If subjects ended up finding nothing, they selected nothing. In order to avoid learning effects different sets of songs were used for each condition and the order of presentation of conditions was randomised. We gathered quantitative and qualitative data. Quantitative data was generated by the logging mechanism, which collected the joystick settings that led to the user-intended results. Qualitative data consisted of observations and interviews.

5.1. Anchor songs and session statistics

People had to rate 10 anchor songs and blocks of Top5 recommendations in the lab and in the "wild". By adjusting the joystick to different positions each subject rated at average 1000 recommendations per session.

Rock	Homebound Train
German Rock	Jetzt geht s los
Folk	April Come She Will
HipHop	Real Love
German HipHop	Michi Beck In Hell
Soul	Caligula
German Soul	Aus der Dunkelheit
NuSoul	Guidance
German Pop	Mensch aus Glas
Funk	Eye To Eye
Electronica	Frozen
Acid Jazz	Stay This Way

Table 4. Anchor songs.

5.2. Quantitative findings

By averaging the weightings of the different facets over all sessions we received the results shown in table 5. Our candidates seem to rely most of the time on an equally rated mixture of sound and cultural aspects (0,41 and 0,36) while lyrics play a minor role (0,22). The lab vs. „wild" results shows a decrease of the sound facet in the ecological environment, maybe because of the noisy environment in the wild settings (WLAN powered public restaurants). The findings were statistically significant (with error rate 0.01 using a paired sign test), but indeed we have to work on large-scale experiments.

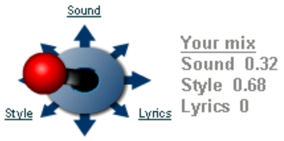
ALL	Sound	Style	Lyrics
Avg	0.41	0.36	0.22
LAB	Sound	Style	Lyrics
Avg	0.44	0.34	0.21
WILD	Sound	Style	Lyrics
Avg	0.38	0.38	0.24

Table 5. Experimental results of lab-based vs. ecological evaluation.

5.3. Qualitative findings: typology of users

Within our small group of subjects we could identify 3 different types of user by averaging their joystick settings and performing post-experiment interviews asking for their introspective view on the experiment.

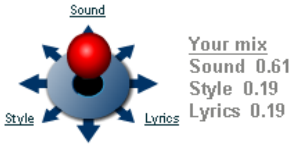
Type 1 users ignored the lyrics and showed a clear bias to the cultural (style) facet, if they knew the recommended songs, they even did not listen to the proposed recommendations.



Your mix
Sound 0.32
Style 0.68
Lyrics 0

Figure 6. Type1 users prefer "cultural agreements".

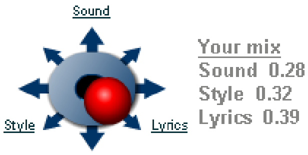
Type 2 users appreciated very much the capabilities of the timbral similarity and made heavy use of the fact to find songs sounding similar. They used the cultural and lyrics facet only to make small corrections.



Your mix
Sound 0.61
Style 0.19
Lyrics 0.19

Figure 7. Type2 users prefer "what they hear".

We found one type 3 user who was interested in finding new and unexpected things. He loved to be surprised by new sound and cultural unusual material. Therefore he used the lyrics facet to explore the song searchspace most of the time by this dimension.



Your mix
Sound 0.28
Style 0.32
Lyrics 0.39

Figure 8. Type3 users prefer "to experiment with unusual facets, e.g. the lyrics".

At this point it becomes obvious that a recommendation engine does not necessarily have to be build on the notion of "similarity"!

6. CONCLUSION AND FUTURE WORK

Our initial experiments have resulted in previously unknown findings about individual and common ratings of song similarity based on subjective evaluation. We have shown that there are statistically significant differences between lab-based and ecological validations. We have described and validated how the perception of music as a socio-cultural product can be supported by MIR technology relying on web mining. We find surprising aspects by exploring interviews with the subjects being engaged in the ecological experiments. As a consequent next step we want to see if we can find relations between users background (education, gender, age, etc. which we already collected) and personal preferences using our data collection from the ecological experiment. Additionally we want to work on stable wireless access settings (e.g. replacing Wireless LAN by UMTS) to improve the idea of ecological evaluation. We will conduct further large-scale experiments within these set-ups. Using a plugin to Winamp will support the long-term observation of user preferences. These efforts have recently been started. We will collect data from users listening behaviour for subsequent data mining in order to extract sequence structure for recommending similar songs. Finally we will open the implemented web-service framework for interested researchers in order to integrate truly semantic services being related to music. We believe that the "socio-cultural compatibility" is a fruitful perspective for future research in MIR.

7. REFERENCES

[1] Leman, M., "Semantic Descriptions for Musical Audio-Mining and Information Retrieval", invited talk *at CMMR 2004,* Esbjerg, Denmark, May, 2004.

[2] Baumann, S., Halloran, J., "An ecological approach to multimodal subjective music similarity, *Proc. of the First CIM 2004,* Graz, Austria, 2004.

[3] Allamanche, E. et al., "A multiple feature model for musical similarity retrieval", *Proc. of the ISMIR 2003*, Baltimore, USA, October, 2003.

[4] Baumann, S., Pohle, T., "A Comparison of Music Similarity Measures for a P2P Application", *Proc. of the 6th DAFX*, London, UK, September 8-11, 2003.

[5] Pampalk, E., Dixon, S., Widmer, G. "On the Evaluation of Perceptual Similarity Measures for Music", *Proc. of the 6th DAFX-03*, London, UK, September 8-11, 2003.

[6] Aucouturier, J.-J., Pachet, F., "Improving Timbre Similarity: How high is the sky?". *Journal of Negative Results in Speech and Audio Sciences, 1(1),* 2004.

[7] Whitman, B., Lawrence S., "Inferring Descriptions and Similarity for Music from Community Metadata", *Proc. of the 2002 ICMC*, Göteborg, Sweden, 16-21 Sep.2002, pp 591-598.

[8] Kohonen, T., *Self-Organizing Maps*, Springer, Berlin, Heidelberg, 1995.

[9] Pampalk, E., Dixon, S., Widmer, G., "Exploring Music Collections by Browsing Different Views", *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR'03)*, Baltimore, MD, October 26-30, 2003, pp 201-208.