PLAYSOM AND POCKETSOMPLAYER, ALTERNATIVE INTERFACES TO LARGE MUSIC COLLECTIONS

Robert Neumayer
Vienna University of Technology
Department of Software Technology
and Interactive Systems
Favoritenstr. 9-11 / 188
A - 1040 Wien, Austria
robert.neumayer@univie.ac.at

Michael Dittenbach
eCommerce Competence Center
iSpaces Group
Donau-City Strasse 1
A-1220 Wien, Austria
michael.dittenbach@ec3.at

Andreas Rauber
Vienna University of Technology
Department of Software Technology
and Interactive Systems
Favoritenstr. 9-11 / 188
A - 1040 Wien, Austria
andi@ifs.tuwien.ac.at

ABSTRACT

With the rising popularity of digital music archives the need for new access methods such as interactive exploration or similarity-based search become significant. In this paper we present the *PlaySOM*, as well as the *PocketSOMPlayer*, two novel interfaces that enable one to browse a music collection by navigating a map of clustered music tracks and to select regions of interest containing similar tracks for playing. The *PlaySOM* system is primarily designed to allow interaction via a large-screen device, whereas the *PocketSOMPlayer* is implemented for mobile devices, supporting both local as well as streamed audio replay. This approach offers content-based organization of music as an alternative to conventional navigation of audio archives, i.e. flat or hierarchical listings of music tracks that are sorted and filtered by meta information.

Keywords: User Interaction, Music Collections, Information Discovery and Retrieval, Audio Clustering, Audio Interfaces, Mobile Devices.

1 INTRODUCTION

The increasing popularity and size of digital music repositories drives the need for advanced methods to organize those archives for both private as well as commercial use.

Similarity-based organization of music archives allows users to explore pieces of music that are similar to ones they know and like. Moreover, it provides a clear and easy navigation for music collections that users are familiar with and allows users to abstract from manually assigned genre information which is, at least in private collections, often inappropriate.

Overcoming traditional genre boundaries can improve search results, e.g. concerning tracks from samplers or movie soundtracks which do not have any (reliable) genre assigned at all. Further, single songs that are very different from the rest of an album could distort the result for a query if relying on genre information, as it is common for all songs of an album to be assigned the same genre. This could lead to problems for albums containing remixes or rather inhomogenous songs. Concerning the access to rapidly growing and changing collections, the similarity-based organization is much more satisfying than conventional search methods because users do not have to know new songs by name as they are offered within their usual queries. This problem becomes more important with the size of a collection. The browsing of a few hundred songs with which a user is familiar might not be much of a problem using metadata, but navigating through thousands of songs one is not familiar with may lead to restrictions, preventing the user from gaining access to the majority of songs.

This paper describes two novel interfaces for accessing music collections, organizing tracks spatially on a two-dimensional map display based on the similarity of extracted sound features. Our work focuses types of interaction itself, as we present user interfaces for both desktop applications and mobile devices.

Section 2 briefly reviews the related work followed by an introduction to the fundamentals of the *Self-Organizing Map*, a neural network-based clustering algorithm and the *Rhythm Patterns* feature extraction model that were used for our experiments in Section 3. We then describe the experimental results of clustering the collection of the ISMIR04 genre contest and describe the presented user interfaces in detail in Section 4 and Section 5 provides some conclusions.

2 RELATED WORK

Scientific research has particularly been conducted in the area of content-based music retrieval (Downie, 2003; Foote, 1999). Recently, content analysis for similarity-based organization and detection has gained significant interest. The *MARSYAS* system uses a wide range of musical surface features to organize music into different genre categories using a selection of classification algorithms (Tzanetakis and Cook, 2000, 2002). This paper will use the *Rhythm Patterns* features to cluster a music collection, previously used in the *SOMeJB* system (Rauber et al., 2002).

Regarding intelligent playlist generation, an exploratory study using an audio similarity measure to create a trajectory through a graph of music tracks is reported in Logan (2002). Furthermore, many applications can be found on the Internet that are not described in scientific literature. An implementation of a map-like playlist interface is the *Synapse Media Player*[1]. This player tracks the user's listening behavior and generates appropriate playlists based on previous listening sessions and additionally offers a map-interface for manually arranging and linking pieces of music for an even more sophisticated playlist generation. Another example of players offering automatic playlist generation is the *Intelligent Multimedia Management System*[2] which is based on tracking the user's listening habits and recommends personalized playlists based on listening behavior as well as acoustic properties like BPM or a song's frequency spectrum. A novel interface particularly developed for small-screen devices was presented in Vignoli et al. (2004). This artist map-interface clusters pieces of audio based on content features as well as metadata attributes using a spring model algorithm. The need for advanced visualization to support selection of audio tracks in ever larger audio collection was also addressed in Torrens et al. (2004), whereby different representation techniques of grouping audio by metadata attributes using Tree-Maps and a disc visualization is presented.

3 SELF ORGANIZING MAPS FOR CLUSTERING AUDIO COLLECTIONS

3.1 Self-Organizing Map

For clustering we use the *Self-Organizing Map (SOM)*, an unsupervised neural network that provides a mapping from a high-dimensional input space to usually two-dimensional output space (Kohonen, 1982, 2001). A *SOM* consists of a set of i units arranged in a two-dimensional grid, each attached to a weight vector $m_i \in \Re^n$. Elements from the high-dimensional input space, referred to as input vectors $x \in \Re^n$, are presented to the *SOM* and the activation of each unit for the presented input vector is calculated using an activation function (usually the Euclidean Distance). In the next step, the weight vector of the winner is moved towards the presented input signal by a certain fraction of the Euclidean distance as indicated by a time-decreasing learning rate α. Consequently, the next time the same input signal is presented, this unit's activation will be even higher. The result of this learning procedure is a topologically ordered mapping of the presented input signals in two-dimensional space. A *SOM* can be trained using all kinds of feature sets. For our experiments we will use the *Rhythm Patterns* features as input data.

3.2 Audio Feature Extraction Using Rhythm Patterns

The feature extraction process consists of two main stages, incorporating several psycho-acoustic transformation (Zwicker and Fastl, 1999). First the specific loudness sensation in different frequency bands is computed. This is then transformed into a time-invariant representation based on the modulation frequency.

The audio data is decomposed into frequency bands, which are then grouped according to the Bark critical-band scale. Then, loudness levels are calculated, referred to as phon using the equal-loudness contour matrix, which is subsequently transformed into the specific loudness sensation per critical band, referred to as sone.

To obtain a time-invariant representation, recurring patterns in the individual critical bands are extracted in the second stage of the feature extraction process. These are weighted according to the fluctuation strength model, followed by the application of a final gradient filter and gaussian smoothing. The resulting 1.440 dimensional feature vectors capture rhythmic information up to 10Hz (600bpm), more detailed descriptions of this approach can be found in Rauber et al. (2003)

3.3 Visualization Techniques of the SOM

Several visualization techniques have been developed to visualize a trained *SOM*, the most appealing in this context being the visualization of component planes.

Here, only a single component of the weight vectors is used to color-code the map representation.

In other words, the values of a specific component of the weight vectors are mapped onto a color palette to paint units accordingly allowing to identify regions that are dominated by a specific feature.

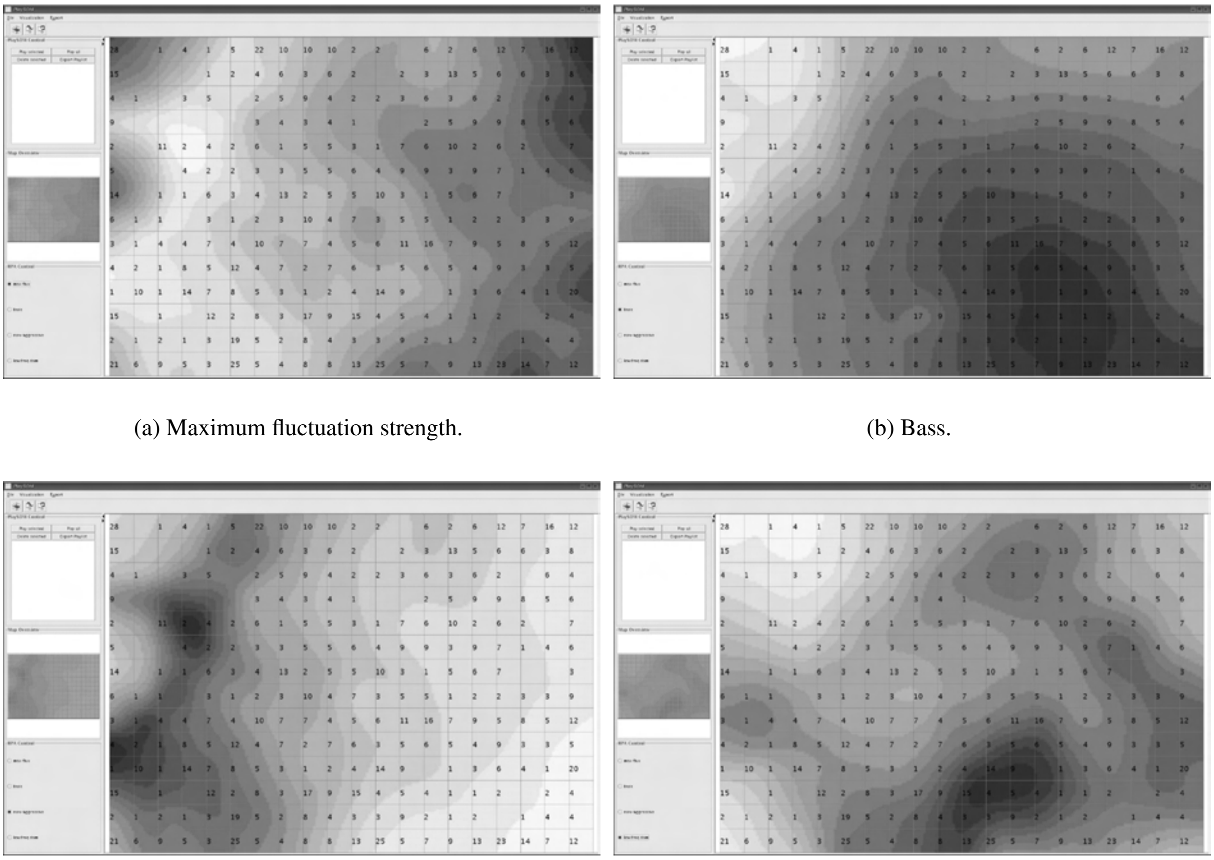
Since single component planes do not directly translate into psychoacoustic sensation noticed by the human ear, the *Rhythm Patterns* uses four combinations of component planes according to psychoacoustic characteristics (Pampalk et al., 2002). More precisely, *maximum fluctuation strength* evaluates to the maximum value of all vector components representing music dominated by strong beats. *Bass* denotes the aggregation of the values in the lowest two critical bands indicating music with bass beats faster than 60 beats per minute. *Non-aggressiveness* takes into account values with a modulation frequency lower than 0.5Hz of all critical bands. Hence, this feature indicates rather calm songs with slow rhythms. Finally, the ratio of the five lowest and highest critical bands measures in how far *low frequencies dominate*. These characteristics can be used to color the resulting map, providing weather-chart kind of visualizations of the music located in different parts of the map. Figure 1 shows examples for all four kinds of visualizations.

4 PLAYSOM AND POCKETSOMPLAYER

We present two interfaces to digital music collections that are based on the *Self-Organizing Map* clustering algorithm and allow interactive exploration of music collections according to feature similarity of audio tracks. The *PlaySOM* and *PocketSOMPlayer* applications both enable users to browse collections, select tracks, export playlists as well as listen to the selected songs. The *PlaySOM*

[1] www.synapseai.com

[2] www.luminal.org



(a) Maximum fluctuation strength.

(b) Bass.

(c) Non-aggressiveness.

(d) Low frequencies dominant.

Figure 1: *PlaySOM* interface with different visualizations of Rhythm Patterns.

presents a full interface, offering different selection models, a range of visualizations, advanced playlist refinement, export to external player devices or simply playback of selected songs. The *PocketSOMPlayer*, on the other hand, offers a slim version of the desktop application, optimized for the *PocketPC* platform, implemented for an iPaq using Java and SWT to be used in a streaming environment.

4.1 Data Collection and Trained SOM

The audio collection used in the ISMIR 2004 genre contest comprises 1458 titles, organized into 6 genres, the major part of which is *Classical* music (640), followed by *World* (244), *Rock_Pop* (203), *Electronic* (229), *Metal_Punk* (90) and *Jazz_Blues* (52). Yet, these genre labels are only used as an indicator during evaluation, as the kind or ofganization provided by the SOM is intended also to overcome the restrictions of manually assigned genre information. The *Rhythm Patterns* of that collection of songs were extracted and its songs were mapped onto a *Self-Organizing Map* consisting of 20×14 units.

The assessment of clustering quality is generally difficult due to the highly subjective nature of the data and the broad spectrum of individual similarity perception. We still try to provide an overview of the map-based organization of this collection and pick some sample areas of the map to demonstrate the results based on the interfaces.

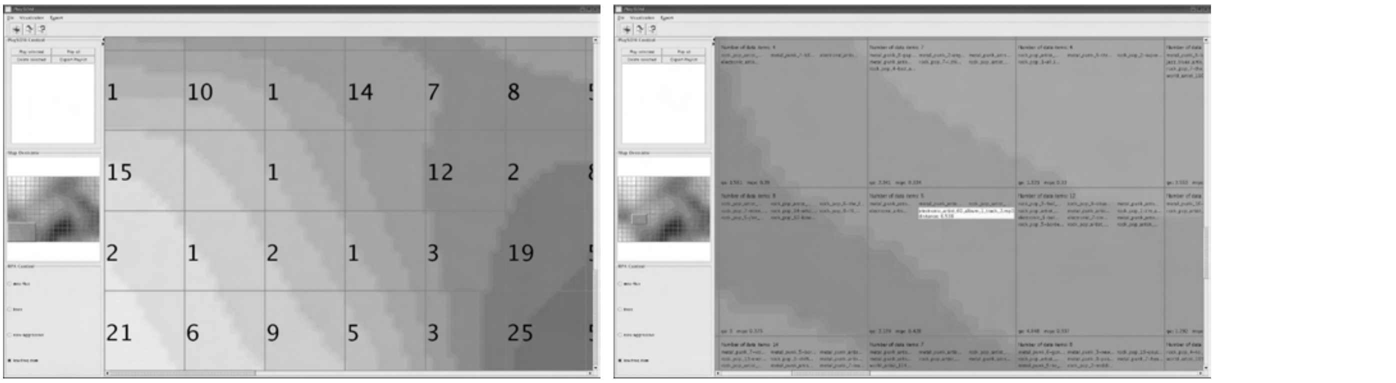
4.2 PlaySOM

Figures 1(a)-(d) show the complete map visualizing the four different *Rhythm Patterns* sub-groups described in the previous section.

A linear gray scale comprising 16 colors from dark gray to white representing feature values that range from low to high is used for printing purposes. (For on-screen use, we emphasize the map metaphor by using a fine-grained color palette ranging from blue via yellow to green reflecting geographical properties similar to the *Islands of Music* (Pampalk, 2001)).

The organization of the songs according to the *maximum fluctuation strength* feature is clearly visible in Figure 1(a) where pieces of music having high values are located primarily on the left-hand side of the map. Especially *Metal_Punk* and *Rock_Pop* as well as some of the *Electronic* songs that are less bass-dominated can be found there. Contrarily, songs with low values are located on the map's right-hand side. Some examples of rather tranquil music are tracks belonging to the genres *Classic* or *World* as well as single *Pop_Rock* songs.

Figure 1(b) shows that the feature *bass* is concentrated



1	10	1	14	7	8
15		1		12	2
2	1	2	1	3	19
21	6	9	5	3	25

(a) Low level of detail - the number of songs mapped are written on the respective units.

(b) High zooming level - song names are displayed on the respective units.

Figure 2: Semantic zooming and its impact on the displayed data.

on the upper left corner and basically consists of bass-dominated tracks belonging to *Electronic* genre. This cluster is the most homogenous on the map (along with a cluster of classical music) according to genre tags, almost no other genres are found in this area.

Finally, a small cluster where *low frequencies dominate* is located in the upper left of the map as shown in Figure 1(d) and corresponds to the results of *bass* setting, leading to low values in this region.

The different types of classical music are a good example of similarity-based clustering that overcomes genre boundaries. Whereas many songs from operas are located on the lower left-hand side of the map, many other tracks that also belong to the *Classical* genre, but sound very different from operas are located on the upper right. This mapping is based on the fact that many songs from the *World* genre share many characteristics with slow pieces of classical music, but differ from operas, a possibility which is not captured by static genre assignments whatsoever.

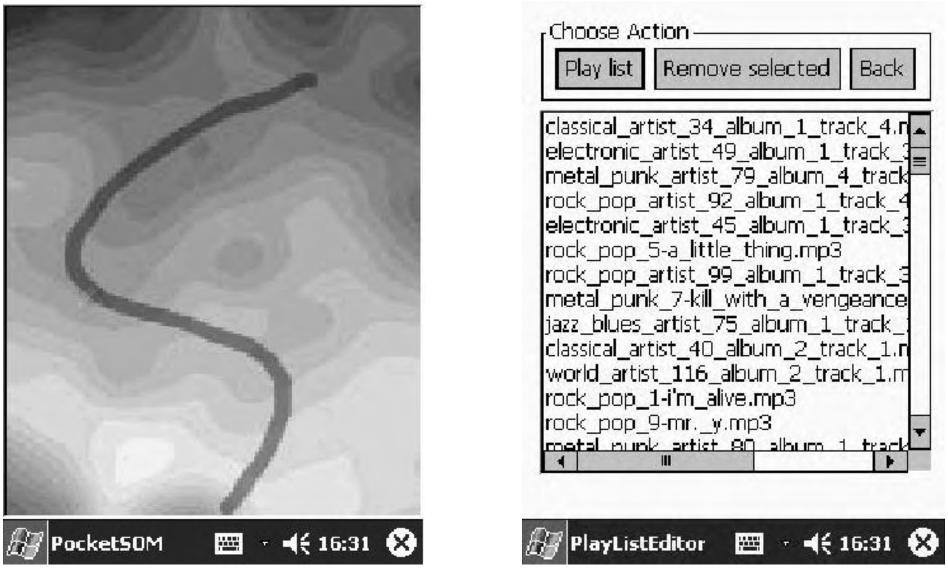
The *PlaySOM* allows users to interact with the map mainly by panning, semantic zooming and by selecting tracks. Users can move across the map, zoom into areas of interest and select songs they want to listen to. They can thereby browse their private collections of a few thousand songs, generating playlists based on track similarity instead of clicking through metadata hierarchies, and either listening to those selected playlists or exporting them for later use. Users can abstract from albums or genres which can often lead to rather monotonous playlists often consisting of complete albums or many songs from one genre. This approach enables users to export playlists based on the track itself not on metadata similarity or manual organization.

The main *PlaySOM* user interface's largest part is covered by the interactive map on the right, where squares represent single units of the SOM. Controls for selecting different visualizations and exporting the map data and the current visualization for the *PocketSOMPlayer* are part of the menubar on the top. The left hand side of the user interface contains (1) a playlist of currently selected titles, (2) a birds-eye-view showing which part of the potentially very large map is currently depicted in the main view on the right and (3) controls for the currently selected visualization (as demonstrated by the different settings of the *Rhythm Patterns* in Figure 1).

The icons on the upper left allow the user to switch between the two different selection models and to automatically fit the map to the current screen size. The *PlaySOM* currently supports two interaction models. The rectangular selection model allows the user to drag a rectangle and select the songs belonging to units inside that rectangle without preserving any order of the selected tracks. This model is used to select music from one particular cluster or region on the map.

On the other hand, the line selection model allows selection of songs below a trajectory in its specific order. In this case the sequence of selected units is of particular importance, because this line chooses a variety of songs according to their position on the map, i.e. their similarity. Hence the line selection model makes it possible to generate playlists that provide smooth transitions between clusters of tracks. This might be of specific interest when browsing very large music collections or when rather long playlists shall be generated (for example if a playlist for several hours should be generated and several changes in musical style shall occur over time, similar to an *auto-dj* functionality).

Another vital aspect of the interface is that it supports semantic zooming, i.e. the zooming level influences the amount and type of data displayed. As outlined in Figure 2, the higher the zooming level, the more information is displayed ranging from information about the number of songs mapped to a particular unit (Figure 2(a)) to detailed information about the tracks (Figure 2(b)), i.e. artist- and trackname. Furthermore, the main *PlaySOM* application can easily and efficiently be used on a Tablet PC and used as a touch screen application because of its portable Java implementation (a live demo is shown in 4(b)).



classical_artist_34_album_1_track_4.n
electronic_artist_49_album_1_track_3
metal_punk_artist_79_album_4_track
rock_pop_artist_92_album_1_track_4
electronic_artist_45_album_1_track_3
rock_pop_5-a_little_thing.mp3
rock_pop_artist_99_album_1_track_3
metal_punk_7-kill_with_a_vengeance
jazz_blues_artist_75_album_1_track_1
classical_artist_40_album_2_track_1.n
world_artist_116_album_2_track_1.m
rock_pop_1-i'm_alive.mp3
rock_pop_9-mr._y.mp3
metal_punk_artist_80_album_1_track

(a) The *PocketSOMPlayer*'s main panel showing a trajectory selection.

(b) *PocketSOMPlayer* user refinement panel.

Figure 3: The *PocketSOMPlayer* interface showing different interaction views.

4.3 PocketSOMPlayer

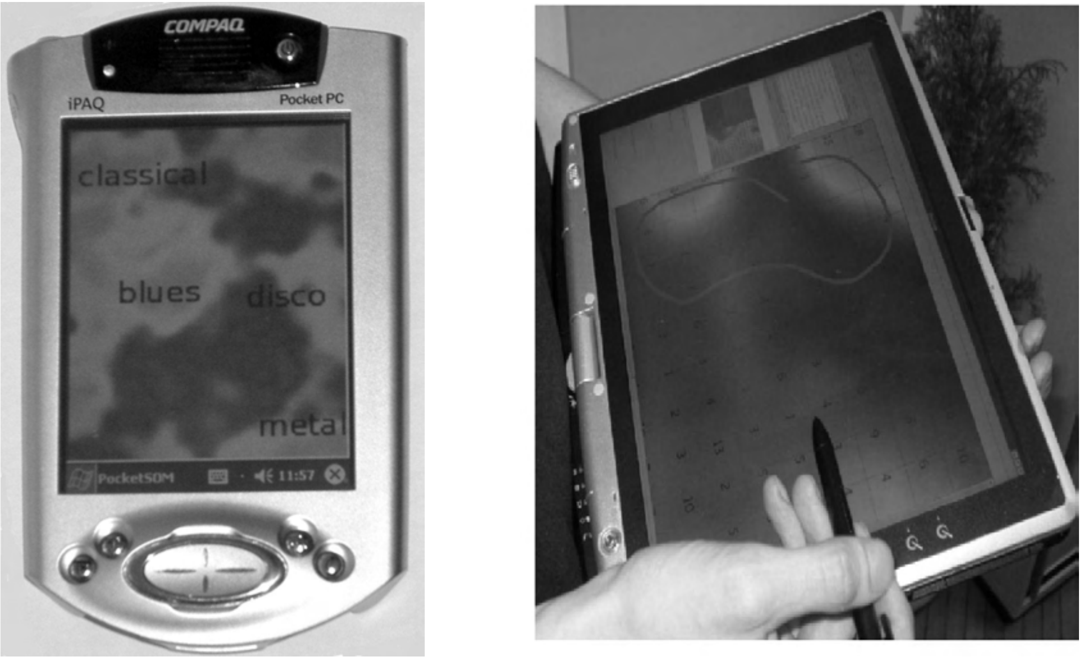
The *PocketSOMPlayer* application offers similar but simplified functionality as the *PlaySOM* is designed for mobile devices such as PDAs or Smartphones. Therefore it provides only the basic functinality of selecting by drawing trajectories and a simplified refinement section, omitting means to zoom or pan the map. Its operational area is likely to be a client in a (wireless) audio streaming environment for entertainment purposes. Regarding the current memory restrictions of PDAs, the use of a streaming server as a music repository seems even more appealing than for the desktop application. Nevertheless, the mobile interface could be synchronized with its desktop pendant to take the role of a mobile audio player within the PDA's memory limits.

Figure 3(a) shows the *PocketSOMPlayer*'s main interface, a trajectory selection with an underlying map. Its user refinement view which allows the user to modify the previously selected playlist before listening to the result is depicted in Figure 3(b). (Due to the anonymized format of the ISMIR collection we emphasized on genres instead on individual track names. In real application scenarios, filenames or ID3-tag information would be used for displaying information on the map.) The main panel allows the user to draw trajectories and to select the units underneath those trajectories. All songs mapped to the selected units are added to the playlist. The user refinement panel pops up as soon as a selection is finished and provides similar functionality as the *PlaySOM*'s playlist controls, namely the user can delete single songs from the playlist to refine her/his selection. The resulting playlist can then be played, retrieving the MP3s either from the local storage or a streaming server.

Figure 4(a) shows the *PocketSOMPlayer* running on an iPaq PDA without a trajectory selection. The map describes a music repository located on a streaming server running on another machine, accessible via WLAN, in contrast to keeping the music files locally (note that labels are manually assigned to clusters according to the most prominent genres in this example). Selecting tracks via drawing of trajectories on a touch schreen is straightforward, easy to learn and intuitive as opposed to clicking through genre hierarchies and therefore particularly interesting for mobile devices and their handling restrictions.

5 CONCLUSIONS

We presented the *PlaySOM*, a novel user interface to map representations of music collections created by training a *Self-Organizing Map*, i.e. a neural network with unsupervised learning function using automatically extracted feature values to cluster audio files. The interface allows user interaction and interactive exploration based on those maps, which was described in detail in our experiments. The *PlaySOM* offers a two-dimensional map with spatial organization of similar tracks and is especially appealing for large or unknown collections. The application allows users to browse their collections by similarity and therefore find songs similar to ones they know by name in contrast to metadata-based approaches. Moreover, we introduced a PDA application offering similar functionality. Both user interfaces are well suited for interactive exploration of collections of digital music due to their different levels of interactive features, including semantic zooming or on-the-fly playlist generation.



(a) The *PocketSOMPlayer* application running on an iPaq PDA.

(b) *PocketSOMPlayer* running on a Tablet PC.

Figure 4: Both presented interfaces running on an iPaq and Tablet PC respectively.

ACKNOWLEDGEMENTS

Part of this work was supported by the European Union in the 6. Framework Program, IST, through the DELOS NoE on Digital Libraries, contract 507618, and the MUSCLE NoE on Multimedia Understanding through Semantics, Computation and Learning, contract 507752.

REFERENCES

J. S. Downie. *Annual Review of Information Science and Technology*, chapter Music information retrieval, pages 295–340. Information Today, 2003.

J. Foote. An overview of audio information retrieval. *Multimedia Systems*, 7(1):2–10, 1999.

T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.

T. Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, Berlin, 3rd edition, 2001.

B. Logan. Content-based playlist generation: Exploratory experiments. In *Proc. 3rd Ann. Symp. on Music Information Retrieval (ISMIR 2002)*, France, 2002.

E. Pampalk. Islands of music: Analysis, organization, and visualization of music archives. Master's thesis, Vienna University of Technology, December 2001.

E. Pampalk, A. Rauber, and D. Merkl. Content-based organization and visualization of music archives. In *Proceedings of ACM Multimedia 2002*, pages 570–579, Juan-les-Pins, France, December 1-6 2002. ACM.

A. Rauber, E. Pampalk, and D. Merkl. Using psycho-acoustic models and self-organizing maps to create a hierarchical structuring of music by musical styles. In *Proceedings of the International Conference on Music Information Retrieval*, pages 71–80, Paris, France, October 13-17 2002.

A. Rauber, E. Pampalk, and D. Merkl. The SOM-enhanced JukeBox: Organization and visualization of music collections based on perceptual models. *Journal of New Music Research*, 32(2):193–210, June 2003.

M. Torrens, P. Hertzog, and J. L. Arcos. Visualizing and exploring personal music libraries. In *ISMIR 2004, User Interfaces*, pages 421–424, Barcelona, Spain, October 2004.

G. Tzanetakis and P. Cook. Marsyas: A framework for audio analysis. *Organized Sound*, 4(30), 2000.

G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, July 2002.

F. Vignoli, R. van Gulik, and H. van de Wetering. Mapping music in the palm of your hand, explore and discover your collection. In E. Fox and N. Rowe, editors, *ISMIR 2004, User Interfaces*, pages 409–414, Barcelona, Spain, October 2004.

E. Zwicker and H. Fastl. *Psychoacoustics, Facts and Models*, volume 22 of *Series of Information Sciences*. Springer, Berlin, 2 edition, 1999.