Multimed Tools Appl (2009) 44:449–468
DOI 10.1007/s11042-009-0275-y

Creating ambient music spaces in real and virtual worlds

**Jakob Frank · Thomas Lidy · Ewald Peiszer ·
Ronald Genswaider · Andreas Rauber**

Published online: 28 April 2009

Abstract Sound and, specifically, music is a medium that is used for a wide range of purposes in different situations in very different ways. Ways for music selection and consumption range from completely passive, almost unnoticed perception of background sound environments to the very specific selection of a particular recording of a piece of music with a specific orchestra and conductor at a certain event. Different systems and interfaces exist for the broad range of needs in music consumption. Locating a particular recording is well supported by traditional search interfaces via metadata. Other interfaces support the automatic creation of playlists via artist or album selection, up to more artistic installations of sound environments that users can navigate through. In this paper we present a set of systems that support the creation of as well as the navigation in musical spaces, both in the real world as well as in virtual environments. We show common principles and point out further directions for a more direct coupling of the various spaces and interaction methods, creating ambient sound environments and providing organic interaction with music for different purposes.

Keywords Music information retrieval · Virtual worlds ·
Mobile computing · User interfaces

J. Frank (✉) · T. Lidy · E. Peiszer · R. Genswaider · A. Rauber
Institute of Software Technology and Interactive Systems,
Vienna University of Technology, Favoritenstraße 9-11/188, 1040 Vienna, Austria
e-mail: frank@ifs.tuwien.ac.at

T. Lidy
e-mail: lidy@ifs.tuwien.ac.at

E. Peiszer
e-mail: mail@ewald-peiszer.net

R. Genswaider
e-mail: r.genswaider@gmx.at

A. Rauber
e-mail: rauber@ifs.tuwien.ac.at

1 Introduction

Music accompanies a large part of our daily life, in different degrees of prominence. This may start from almost unnoticed background sound environments as we find them for example in shops and restaurants. A somewhat more conscious choice of certain types of music is made when selecting certain radio stations at different times of the day, or when deciding to visit certain bars or clubs according to the style of music they are playing. Further along the line of selecting specific music we may consider putting together a playlist for a specific purpose, such as while learning or jogging or while travelling to work, or deciding to go to a specific concert—up to the very specific selection of a particular piece of music to listen to, possibly even in a very specific interpretation, played by a specific artist or with a selected conductor. The different styles of listening to music serve different purposes, from mere ambient sound via choosing a specific type of music for specific purposes or settings, forming a continuum with hardly any strict boundaries.

Music most frequently conveys emotions and feelings and has important social aspects [1]. For instance, many places such as restaurants and bars play a specific kind of music. As a consequence, people who meet there typically share a common taste for music. In this respect music has a context also with locations, but even more frequently with situations: People like to listen to different musical genres according to their mood and intentions, e.g. depending if they want to relax, do sports or go out and entertain themselves. For many people such activities are indispensable without music.

Different systems and interfaces support the interaction with, as well as the selection and consumption of music for these different purposes. These may range from databases with comprehensive metadata on the pieces of music for search purposes, via predefined structurings of music according to albums, artists, or musical styles, up to artistic installations, where users can interact with a musical space and influence it that way.

In this paper we present four such approaches that are all based on a fully automatic analysis of music, as briefly introduced in [4]. Following a feature extraction process that extracts music descriptors from an audiofile using psycho-acoustic models, the music is organized on a 2-dimensional plane according to perceived acoustic similarity forming a map of music. This organisation serves as the basis for four—potentionally interlinked—realisations. One of these is the *MediaSquare* with its *virtual MusicSOM Cafe*, where on each table of the cafe a specific type of music is being played and neighboring tables have similar sounding styles. User avatars can walk through the cafe perceiving gradual transitions of musical styles on the different tables and can choose the seat where they like the music most. The same system has also been realized in a prototypical *real-world MusicSOM Cafe* setup, allowing people to pick a table to sit at according to their musical preferences. Both approaches create localized ambient music environments that offer fascinating possibilities for integration and mutual interaction between avatars in a virtual world and persons in a physical cafe/installation. The third approach is utilizing a CAVE Automatic Virtual Environment to create an immersive *musicCAVE* landscape. Last, but not least, we show the implementation of the map-based music principles for mobile devices in various flavors of the *PocketSOMPlayer*, which—apart from being used for selecting and playing music—may also communicate playlists to the central server. There they are combined to facilitate the creation of community playlists in real-time, shaping the music that is played by a central system.

All these approaches offer multiple ways of integration and cross-connection to create musical spaces that are both individually controllable but also shaped by the community contemporarily present in that space, be it virtual or real life.

The remainder of this paper is structured as follows: Section 2 reviews various approaches and systems that influenced the design of the systems presented in this paper. Section 3 then focuses on the basic components of all systems, namely the extraction and computation of suitable descriptive features from music, as well as the basic concepts for creating the 2-dimensional maps of music using the Self-Organizing Map. Sections 4.1 and 4.2 then present realizations of these systems both in a virtual world based on a game engine, as well as in the real world. Section 4.3 presents a third realization of the resulting musical spaces in an immersive CAVE, while Section 4.4 again moves into a real-world setting, allowing users to utilize the music maps on mobile phones both for mobile music consumption via playback or streaming, but also and predominantly as a remote control creating and influencing centralized playlists. Section 5, finally, pulls these various approaches together and tries to identify directions for integrating these to form novel means of providing and interacting with musical spaces.

2 Related work

The ease of distribution over the Internet contributed to the pervasiveness of music. Yet, with massive collections of music new problems arise, with respect to selecting music from large repositories. A need for sophisticated techniques for retrieval emerged that goes beyond simple browsing or matching of metadata, such as artist, title, and genre. This need is addressed by the research domain of Music Information Retrieval (MIR). Recent research has resulted in “intelligent” methods to organize, recognize and categorize music, by means of music signal analysis, feature extraction from audio and machine learning. Downie [2] provides an overview of methodologies and topics in this research domain, with a review also of the early approaches. The article of Orio [26] contains a more recent review of the many different aspects of music processing and retrieval. Moreover, an overview of prototypical Music-IR systems is given.

One kind of system for retrieving music are Query-by-Humming systems that allow to search for songs by singing or humming melodies [6, 21]. While introduced already in the mid-1990s, today, this technique has reached a mature state and was implemented e.g. in the commercial online archive midomi.com.[1] Other applications allow users to explore areas of related music instead of querying titles they already know. Torrens proposed three different visual representations for music collections using meta-data, i.e. genres to create sub-sections and the date of the tracks for sorting them [32]. Tzanetakis and Cook introduced Marsyas3D, an audio browser and editor for collaborative work on large sound collections [33]. A large-scale

[1] http://www.midomi.com/

multiuser screen offers several 2D as well as 3D interfaces to browse for sound files which are grouped by different sound characteristics.

A specific approach of organizing music automatically, applied also in the scenarios described in this paper, is to cluster music according to perceived acoustic similarity, without the need of meta-data or labels. This is realized by (1) extracting appropriate "features" from the audio signal that describe the music so that it is processable and to a certain degree interpretable by computers and (2) applying a learning algorithm to cluster a collection of music. The set of features we use describes both timbre and rhythm of music and is called Rhythm Pattern (RP), covering critical frequency bands of the human auditory range and describing fluctuations with different modulation frequencies on them. The basic concepts of the algorithm were first introduced in [28] and enhanced later by the inclusion of psycho-acoustic models in [30]. The feature set has proven to be applicable to both classification of music into genres [15] and automatic clustering of music archives according to the perceived sound similarity [20]. Furthermore, the correspondence of the resulting organization with emotional interpretations of the sound in various regions of the map has been analyzed [1]. While the experiments in this paper are predominantly based on Rhythm Pattern features, other descriptors that capture sound characteristics may be used. These include the RP-related features Statistical Spectrum Descriptor (SSD) and Rhythm Histogram (RH) [15], but also features defined in the MPEG-7 standard [8]. These approaches can be extended by the inclusion of other features capturing further musical properties and also by approaches deriving musical notations from audio signals, as shown in [17]. Another prominent library to extract a range of features from audio is MARSYAS [34].

In order to cluster the music according to perceived sound similarity the Self-Organizing Map (SOM) algorithm is employed [10]. The SOM is a topology-preserving mapping approach, that projects high-dimensional input data—in our case the features extracted from audio—onto a 2-dimensional map space. It has frequently been used to organize information by similarity, predominantly in the textual domain [11, 25, 29], but also for other modalities such as images [12]. In the acoustic domain it has first been used for mapping short instrument sounds [3], before being first applied to the organization of entire music collections in the SOMeJB system [28]. The preservation of acoustical neighborhood in the music collection in the resulting map allows a number of applications, such as quick playlist creation, interactive retrieval and a range of further interesting scenarios, allowing for ambient music experience in real and virtual spaces, as we will describe in the course of this article.

Previously, we presented PlaySOM, a 2D desktop application offering interactive music maps and the PocketSOMPlayer, designed for small devices such as palmtops and mobile phones, that both allow users to generate playlists by marking areas or drawing trajectories on a music SOM [22]. Knees et al. transformed the landscape into a 3D view and enriched the units of the SOM by images related to the music found on the Internet [9]. The music is played back in an ambient manner according to the user's location in the 3D landscape and the vicinity to the specific clusters. Lübbers follows this principle of auralization of surrounding titles on a 2D music map application called SonicSOM [18]. Besides, he proposed SonicRadar, a graphical interface comparable to a radar screen. The center of this screen is the actual viewpoint of the listener. By turning around, users can hear multiple neighboring music titles, panning and loudness of the sounds describe their position relative to the user. Other projects built on the principle of SOM-based organization of music collections include the graphical tabletop interface by Hitchner et al. [7] and the Globe of music [13].

In contrast to these works, the applications presented in this article allow users to immerse into more familiar environments and enable to meet and interact with other people in a social environment. Ambient music experience is enabled by our scenarios in a virtual multi-user world, a real-world cafe, and an immersive CAVE environment. Besides enabling to carry one's musical space in everyday life to every possible place, our portable music map scenario also enables collaborative and social features, such as the conjoint creation of music playlists, an idea that is also exploited by the PublicDJ prototype [14].

3 Technical fundamentals

The application scenarios we are presenting in Section 4 make use of automatic spatial arrangements of collections of music. In this section, we describe the underlying fundamentals that are necessary to create this automatic arrangement, i.e. audio analysis with automatic feature extraction methods and clustering of pieces of music with Self-Organizing Maps. For the latter, the *PlaySOM* software is used.

3.1 Audio feature extraction

The research domain of *Music Information Retrieval* explores methods that enable computers to extract semantic information from music in digital form [2, 26]. Part of this research is the development of methods for the extraction of features from audio that on the low level capture the acoustic characteristics of the signal and on the higher level try to derive semantics such as rhythm, melody, timbre or genre from it. The extracted features, or descriptors, not only enable the computation of similarity between pieces of music, resembling the acoustic similarity perceived by a listener, but also allow an automated organization of music based on content or (semi-)automatic classification of music into genres.

In the context of the applications described in Section 4 we utilized the following feature extractors in order to derive measures of acoustic similarity from the audio content: Statistical Spectrum Descriptors, Rhythm Patterns and Rhythm Histograms. These extractors share a number of psycho-acoustic transformations based on a spectral representation that aims at reflecting perception by the human auditory system. First, a Short Time Fourier Transform (STFT) is applied to transform the audio waveform into a Spectrogram representation, whose frequency bands are then grouped into 24 critical bands, as defined by the psycho-acoustically motivated Bark scale [38]. The Spectrogram's values are then transformed into the Decibel scale, followed by the application of another psycho-acoustic model, the Phon scale, considering equal loudness curves, which equalize the different perception of loudness at different frequencies by humans [38]. Subsequently, the values are transformed into the unit Sone, reflecting the specific loudness sensation of the human auditory system in way that a doubling on the Sone scale sounds to the human ear like a doubling of the loudness. From this Sonogram representation a Statistical

Spectrum Descriptor (SSD) is derived by computing seven statistical measures on each of the 24 critical bands, capturing properties of fluctuations on these, including the evolution of the musical piece in time [15].

Applying a further Fourier Transform on the Sonogram representation, the magnitudes of modulation for different modulation frequencies are derived (a so-called "cepstrum"). After a weighting step for fluctuation strength and subsequent smoothing, a Rhythm Pattern [30] reflects the rhythmical structure of a piece of music and also includes information about the timbre.

A more simplified representation of rhythmic aspects is captured by a Rhythm Histogram, which aggregates the information contained in the ceptrum representation for all critical bands and thus gives a compact summarization of rhythmics [15].

Musical features are usually computed on a per-segment basis, in the case of the aforementioned descriptors for segments of 6 s, typically for every third segment in a piece of music. The final descriptor for a piece is then computed by taking the median of multiple segment descriptors.

Matlab modules as well as web services for computing the features described above are available at the project homepage.[2]

3.2 Self-organizing maps

There are numerous clustering algorithms that can be employed to organize music by sound similarity. One model that is particularly suitable, is the Self-Organizing Map (SOM), an unsupervised neural network that provides a topology preserving mapping from a high-dimensional input space to a usually two-dimensional output space [10].

A SOM is initialized with an appropriate number i of units (or nodes), proportional to the number of tracks in the music collection. Commonly, a rectangular map is chosen, although other forms are possible. The units are arranged on a two-dimensional grid. A weight vector $m_i \in \Re^n$ is attached to each unit. The input space is formed by the feature vectors $x \in \Re^n$ extracted from the music by an audio feature extractor. Elements from the high-dimensional input space (i.e., the input vectors) are randomly presented to the SOM and the activation of each unit for the presented input vector is calculated using a distance measure such as e.g. the Euclidean distance. In the next step, the weight vector of the unit showing the highest activation (i.e., having the smallest distance) is selected as the "winner" and is modified as to more closely resemble the presented input vector. The weight vector of the winner is moved towards the presented input vector. Furthermore, the weight vectors of units neighboring the winner are modified accordingly, yet to a smaller degree.

After numerous iterations the process results in a similarity map, in which music is placed according to perceived acoustic similarity: Similar sounding music is located close to each other, building clusters, while pieces with more distinct content are located farther away. If the pieces in the music collection are not from clearly distinguishable genres the map will reflect this by placing pieces along smooth transitions.

[2] http://www.ifs.tuwien.ac.at/mir/audiofeatureextraction.html

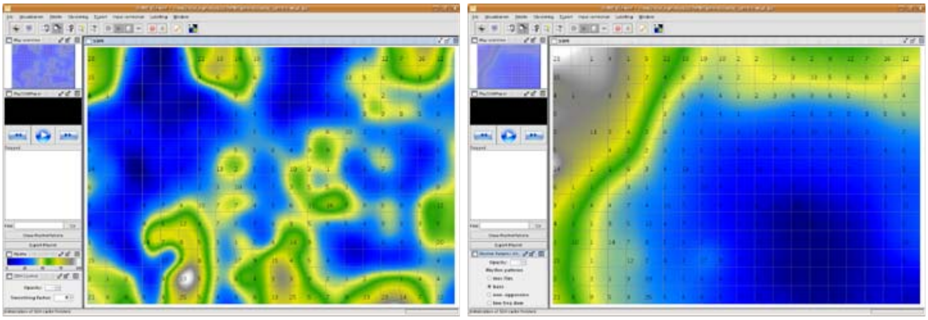
3.2.1 *Visualizations of a SOM*

Clusters and structures on a trained SOM are not inherently visible, therefore several visualization techniques have been developed to enable a deeper insight [16]. Various visualization algorithms aim at facilitating interpretation of the underlying structures of the SOM from different perspectives.

A standard visualization for SOMs is the U-Matrix [37] that visualizes distances between the weight vectors of adjacent map units. The local distances are mapped onto a color palette, gradually changing with distance. The P-Matrix is an modified version that incorporates local relative data densities based on the so-called Pareto-Radius around the prototype vectors [35]. The U*-Matrix [36] aims at showing cluster boundaries combines these two visualization algorithms, taking both the local distances between unit vectors and the data density into account, weighting the U-Matrix values according to the P-Matrix values.

Component planes visualize the distribution of particular components of the underlying feature set, allowing to investigate the influence of a particular feature (e.g. a statistical moment on a specific frequency band). Each unit on the map is color-coded, where the color reflects the magnitude of a particular component of the weight vector of each unit. Also, aggregations of feature attributes that exhibit certain semantics of the underlying data can be visualized with this method. For example, for the Rhythm Patterns feature set a range of semantically interpretable visualizations has been determined, exhibiting musically important parameters such as “fluctuation strength”, “non-aggressiveness” or “bass” (see Fig. 1b). With the appropriate color palette, this visualization is comparable to “Weather Charts” [27].

Smoothed Data Histograms (SDH) [31] are an approach to visualize the cluster structure of the data set in a more global manner than e.g. the U-Matrix that depicts unit distances. The concept of this visualization technique is basically a density estimation and resembles the probability density of the whole data set on the map. When a SOM is trained, each data item is assigned to the map unit which best represents it. However, by continuation of these distance calculations it is also possible to determine the second best, third best, and so on, matching units for a given feature vector. Based on a voting function, these additional matches are incorporated and accumulated in a histogram creating an SDH. The number of matches, or “levels”, can be adjusted by the user, so that the SDH offers a sort of



(a) Smoothed Data Histograms (SDH)

(b) “Bass”-Component Planes

Fig. 1 **a**, **b** PlaySOM with different visualizations of a music collection hierarchical representation of the cluster structures on the map. It is then visualized using spline interpolation and appropriate color palettes that create a visualization resembling *Islands of Music*, as depicted in Fig. 1a [27, 31].

Clusters of dense areas of similar sounding music are depicted by green and yellow islands, peaks by mountains, and less dense areas by water in the sea. These areas may still contain music items, but these are rather along smooth transitions from one musical genre to another.

An additional visualization, which is, however, not based on the SOM's structures but on external meta-data is the class visualization. Frequently, genre labels are available for music titles, alternatively, music collections may be manually sorted into different categories. The availability such class information allows the creation of a color-coded visualization in form of flood-filled areas or overlays, which assist in the analysis of the cluster structures on the map [19].

3.2.2 *PlaySOM*

PlaySOM is an interactive application that allows the creation of music maps using the SOM algorithm [22]. On top of that, it offers a range of visualizations and number of interaction features: It provides an easy-to-use 2D interface presenting a map as an interface to a music collection. The main window of PlaySOM consists of the map and allows the user to select songs for replay by drawing on it. Two modes of selection are available: Making a rectangular selection, entire clusters of music containing similar sounding pieces are selected. By drawing trajectories, one can quickly create playlists going from one musical genre smoothly to another, according to the path selected. The playlist window on the left shows the according selection of songs. Users can refine the playlist edit the list before sending it to a music player. Figure 1a shows the main screen of PlaySOM with an example music map, showing the SDH visualization and a trajectory selection.

A very important feature of PlaySOM is the implementation of various visualizations, some of which have been described in the previous section. These aim at helping the users to orient themselves on the map and aid in finding the desired music. To gain a more detailed view, semantic zooming provides different amounts of contextual information according to the zoom level.

PlaySOM also allows to export music maps—both the spatial arrangement (i.e., the clustering) and the graphical representation. It is used to generate the organization of the music maps for all of the applications presented in Section 4.

4 Applications in real and virtual worlds

All following applications aim to create music spaces, some in real world, some in virtual world. Users (or in some cases: visitors) can move or even walk through these spaces to experience music.

Audio files may come from a range of different sources, e.g., personal music libraries of the user. Many users store several thousands of tracks on their computer or MP3-player, which is an amount, where manual organization of music starts to become difficult and mere metadata-search is not sufficient any more. A second source are huge commercial music portals offering their catalogues as a music space in combination with a "music flat-rate" for on-demand streaming. In this scenario a small personal music collection could be additionally used to create a personal profile helping to orientate in and navigate through the vast music space offered by the provider. A third possibility could be temporarily shared music, e.g., at a party, where guests can bring their favorite music, which is then combined to a common music space.

4.1 The AudioSquare

Virtual three-dimensional environments yield the potential of reproducing interaction scenarios known from real-life situations. Avatars, for example, give users the opportunity for self-identification and encourage them to start social interaction with each other. Walking, running and jumping are navigation forms everyone is familiar with. In contrast to other human interface concepts, such as desktop applications or Web sites, navigation through a virtual space prevents users from experiencing visual cuts and, thus, loosing their context. A virtual continuum implicitly creates a mental map of the perceived environment.

The *MediaSquare* takes advantage of the virtual world paradigm for representing multimedia content. It consists of several buildings containing e.g. poster exhibitions (automatically arranged by topic similar to the organization of music as described above) [5] as well as slide-show presentation rooms and a MusicSOM Showroom. It is based on the *Torque Game Engine*.[3] With the free client application available from the *MediaSquare* Homepage,[4] the user can choose an avatar and enter the virtual world. The client-server approach of the application enables a social platform where users are encouraged to start conversations about the presented content through a simple text chat. All objects, avatars and the landscape designed for *The AudioSquare* are reminiscent of real-life scenes. This is based on the assumption, that users do not want to learn the principles of every virtual environment from the ground on. Rather, they are supported in quickly orienting themselves in a scenario that looks familiar to them and are able to focus on the main purpose of the virtual world.

The music within the Music Showroom is represented by 3D objects emitting spatial sound. Each of these objects is connected to a media server over the Internet, streaming several audio tracks consecutively. Users can explore the environment by walking around with their avatar and listening to the presented music streams. Standing in between multiple audio objects music is perceived as a mixture of multiple audio sources, depending on the distance of the avatar to the objects. When a user encounters an audio source, a head-up display (HUD) shows additional information about the currently playing audio track, which can be controlled by the user, e.g., by skipping to the next track. Since each audio source has a specific location, users can orientate themselves not only by visual feedback but also by perceiving the loudness and direction of the audio sources. The acoustic layer also helps users in creating a mental map of the environment.

The other buildings in the *MediaSquare* display posters and pictures mounted to walls and frames. Users can walk through and explore the presented materials

[3] http://www.garagegames.com/products/torque/tge/

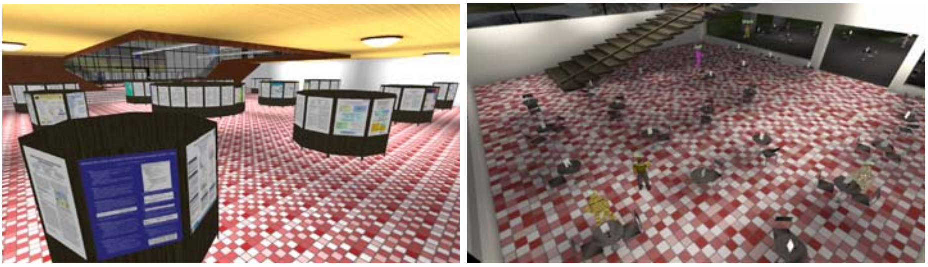
[4] http://www.ifs.tuwien.ac.at/mir/mediasquare/

like in an exhibition. When walking up to e.g., a presentation the user can browse forward/backward through the slides while the HUD shows additional information about the presentation such as the author or where the presentation was originally presented.

The workflow for creating *The AudioSquare*, respectively *The MediaSquare*, comprises three main steps. The first step is the creation of a basic environment with a terrain, buildings, interiors and other objects. Objects for representing the music are created in a separate step and stored as "assets" in a simple repository. Finally, *marker-areas* are placed in order to specify locations for music representation. An automatic process arranges the assets in the virtual world, whereby two different approaches for organizing and representing the underlying music archives are supported concurrently. In the first case, a SOM is used for automatic organization according to the sound characteristics of the music tracks, as described in Section 3. *Marker-areas* defined in the virtual world designate through their boundaries wherein the representation should take place. Each area refers to a section of a SOM as well as to a template from the repository displaying the content of the unit. This allows to distribute a SOM over several rooms and to assign different visual styles (see Fig. 3a). For other data-objects the SOM is based on representative textual characteristics, e.g., for posters, or image/video features for pictures or videos, respectively.

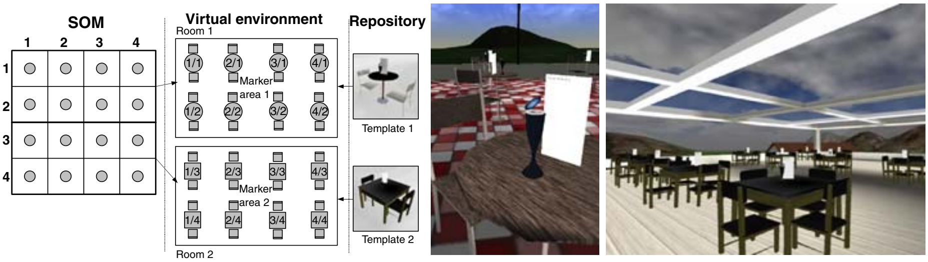
Alternatively to the automatic organization, a manual approach addresses the three-dimensional representation of a simple folder-hierarchy on the file system. Top-level folders stand for an umbrella term wherein further folders contain the audio files (or other media). Each folder contains a small text file that describes its contents by name, date and a short description. Within the virtual environment, the top-level folders are represented by buildings, while the sub-folders are represented by objects that are located inside these buildings. The descriptions provided in the text files are displayed on virtual signboards attached next to the respective objects.

Figures 2 and 3c depict the result of SOM-based organization in the virtual environment. It is a matrix of objects, each representing a unit of the SOM, which in turn can hold an arbitrary number of pieces of music. In this case, one unit is represented by a table with a small speaker on its top that plays the music stream and a playlist that represents the respective content.



(a) Poster presentation in the MediaSquare (b) The MediaSquare *MusicSOM Showroom*

Fig. 2 **a**, **b** Virtual showrooms for multimedia presentations



(a) setup scheme for the *AudioSquare* (b) detail view (c) overview

Fig. 3 **a**–**c** A virtual SOMCafe in the *AudioSquare*

4.2 The MusicSOM Cafe

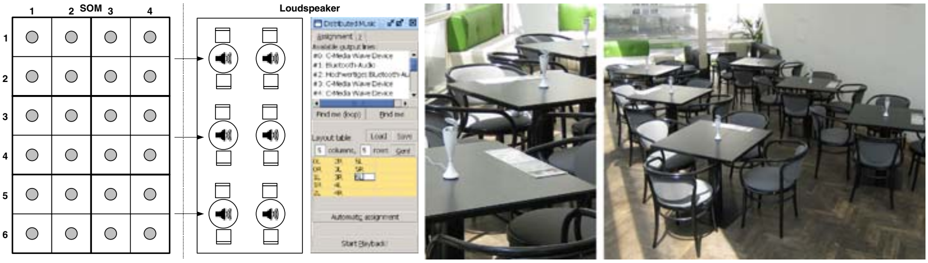
The principle of providing music of similar style on various locations, arranged by sound similarity, has been brought into a real-world scenario with the realization of the prototype of the *MusicSOM Cafe*. It is inspired by the fact that people with similar interests tend to associate with each other. Music is a common social catalyst: people like to meet in bars and clubs featuring a musical style they prefer. On the other hand, open-minded music listeners constantly explore the edges of their music universe to get to know new music. This is where the real-world MusicSOM installation, which basically is a real-world setup of the Music Showroom, comes in handy.

Basically, the concept consists of a public or private space with several loudspeakers distributed in it. Each speaker corresponds to one or more node(s) of a music map (created using *PlaySOM*) and plays the songs assigned to it/them. The speakers play simultaneously and are placed on (alternatively over or embedded into) tables alongside the “song menus”, which contain a list of the songs being played. Figure 4 illustrates the setup.

This concept meets both demands: First, it exhibits multiple areas with one specific music style (the surrounding of each table). There, one can meet people with similar music preference. Second, it allows for exploration of new music styles along adjacent tables by changing one’s position. The real-world MusicSOM is appropriate for both public installations, e.g., as a temporary or permanent art installation, as well as bars, cafes or lounges with a focus on individual and varied music experience.

To facilitate the setup of the installation a plugin for the *PlaySOM* application controls the SOM nodes to speaker assignment. It uses all audio devices available on the computer system regardless of the actual hardware: regular sound cards are as fine as USB audio devices or wireless Bluetooth speakers. Features of the plugin include:

- playing short audio clips to each speaker in a loop to help identification the sequence and spatial arrangement of the loudspeakers during setup
- both automatic and manual assignment of SOM nodes to speakers
- simultaneous playback of songs in random order to each speaker according to assignment



(a) setup scheme (b) detail view (c) complete installation

Fig. 4 **a**–**c** The *real-world MusicSOM Cafe*

The assignment is a three step process: First, the *layout table* is created, reflecting the arrangement of the speakers in the real world (e.g., if the *MusicSOM Cafe* has five tables in two rows, the layout table will be 5 × 2 cells). Second, the *speaker codes* are entered into the corresponding cells, referring to the channels provided by the Java API (right part of Fig. 4a). Note that not all layout table cells need to be used, some may remain empty. Third, a set of SOM nodes is assigned to each cell used, as illustrated in Fig. 4a. The songs that are represented by these nodes are played by the respective speaker in a random ordered loop. To avoid choppy sound due to too heavy CPU load not all songs are MP3-decoded and played on-the-fly. Rather, a parameter-controlled share of songs is decoded and saved as WAV file as a background process. Figure 4c shows an demonstrational installation with six tables.

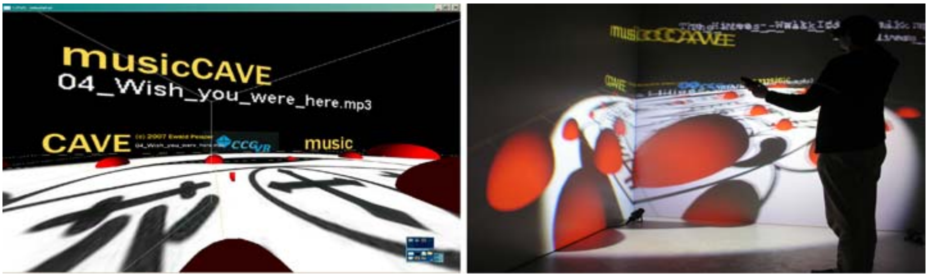
This allows now to establish a connection between the real world and the Music Showroom in the virtual world, which follows the same organization of music, and can also be designed to mirror even the visual appearance, c.f. Figs. 3b, c and 4b, c.

4.3 The musicCAVE

The *musicCAVE* is a hybrid of the *AudioSquare* and the *MusicSOM Cafe*. It demonstrates how music information retrieval and clustering techniques can be combined with Virtual Reality (VR) to create immersive music spaces. In our setup, a music map is displayed in a four-wall Cave Automatic Virtual Environment (CAVE) in stereoscopic projection, which consists of a front, left, right and bottom screen. The user can navigate with a wireless input device ("wand") and a head tracker. The ambient sound changes according to the (virtual) position of the user on the SOM. Thus, the user can explore a music collection in an immersive way. By moving to different positions the style of the music gradually changes. The prototype demonstration has been carried out at the Center for Computer Graphics and Virtual Reality at Seoul's EWHA Womans University.[5] A 6 × 6 SOM that organizes 120 songs was created using *PlaySOM* and Rhythm Pattern audio features.

Each SOM node is represented as a red hemisphere. The more songs are allocated to the node, the bigger the radius of the sphere is. All nodes are playing concurrently, one song each. A 3D sound engine is used to account for the attenuation of the songs

[5] http://ccgvr.ewha.ac.kr/



(a) The *musicCAVE* in the CaveLib simulator.

(b) One of the authors in the *musicCAVE*.

Fig. 5 **a**, **b** The *musicCAVE* creates an immersive music space

according to the distance to the user. Eventually, the user can listen to one specific song if standing right at a node's position. If he or she stands between two or more nodes, a mixture of these songs can be heard, each coming from the direction of the node it is located at. The title of the nearest sphere's song is visible on the wall of the virtual environment.

The position on the map can be changed by either walking around in the approximately 2.5 × 2.5 m sized CAVE or by using the wand joystick. The wand also offers a skip button and a shortcut button to return to the map's ground. The head tracker is used for perspective correction and 3D goggles are used to perceive the map in a stereoscopic view. Figure 5a is a screenshot of the CaveLib CAVE simulator, Fig. 5b is a photo of the application running in the actual CAVE.

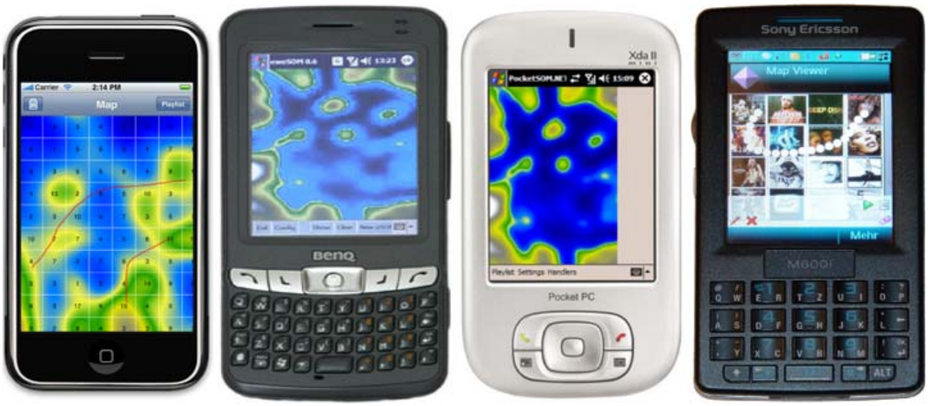
The navigation through the music collection in the virtual reality environment is an interesting experience. Experiments showed that the ideal (real world) distance between the hemispheres is the CAVE's edge length: in this case you can explore four SOM nodes by walking around. If the distance is lower the head tracker position resolution does not allow for a smooth transition between two nodes anymore.

4.4 Portable music maps

Music spaces are available also independently from certain locations or environments, thanks to mobile devices such as mobile phones, smart phones, PDAs or MP3-players. They allow users to listen to their preferred music wherever they go. The omnipresence of mobile Internet connections allows access not only to Internet radio streams but also online music stores as well as the personal audio collection on the computer at home. Yet, the music experience depends heavily on the interaction and access possibilities offered by the device.

Providing access to large audio collections using devices with limited interaction possibilities and small screen sizes is significant challenge currently under strong attention. One such solution are music maps, as proposed in [23], providing a graphical overview of a—potentially large—music collection. The *PocketSOM* software[6] is based on the automatic organization of music by a SOM and provides

[6] http://www.ifs.tuwien.ac.at/mir/pocketsom/



(a) iPhone (b) BenQ P50 (c) HTC Magican (d) Sony Ericsson M600

Fig. 6 **a–d** *PocketSOM* on different devices

intuitive interaction possibilities. Several different implementations are available, each based on and optimized for a specific platform:

- **iSOM** is the PocketSOM application for the iPhone (see Fig. 6a). It integrates perfectly into the iPhone OS using common gestures for map navigation such as panning and zooming, or clearing a playlist selected via drawing a path by shaking the device. A video demo is available via the project website.[7]
- **ePocketSOM** is based on the Eve VM, a java-like VM for portable devices which is available for several mobile platforms such as Windows Mobile, or Nokia N770/N880 (see Fig. 6b).
- **PocketSOM.NET** is based on the .NET Compact Framework. Since PocketSOM.NET is developed in a native environment, it offers better performance and integration into the operating system (see Fig. 6c).
- **mPocketSOM** is a pure JavaME (Java Micro Edition) implementation, thus potentially supporting a huge range of mobile devices (see Fig. 6d).
- **iPocketSOM** was the first prototype of PocketSOM for an iPAQ PocketPC and based on Java and SWT (Standard Widget Toolkit).

While all implementations are optimized for their specific platform, they share a set of common features: Users can listen to music by drawing a trajectory on the music map (“walking over the map”, to stay with the music space metaphor) where similar pieces of music are located together. Drawing is performed either by using a stylus, or using one’s fingers, if enabled by the touch-screen. The resulting playlist will start at one end of the trajectory, following the path to the endpoint containing tracks that are placed along the trajectory on the map, covering potentially multiple musical styles and smooth transitions between them. After creation of a playlist, it can be played back on the mobile device using locally stored music, or the music can be streamed over the Internet, if using a streaming service. This offers access to

[7] http://www.ifs.tuwien.ac.at/mir/pocketsom/isom/

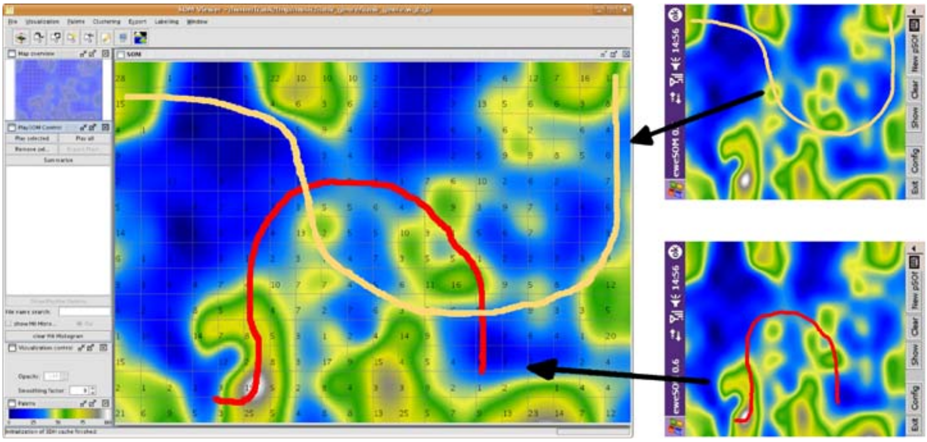


Fig. 7 PocketSOM sending a path to the PlaySOM application

large music portals via a simple portable interaction technique. Additionally, in some implementations the playlist can be sent to a remote server for playback, turning the mobile device into a remote control (e.g. for the personal computer in the living room at home).

A special enhancement of PocketSOM, available in iSOM and ePocketSOM, was realized in conjunction with the PlaySOM application. These implementations are able to directly connect to the PlaySOM application and download all necessary map data. With the "path-sharing" feature activated, every following playlist trajectory that is drawn on the mobile device is then directly transferred to the PlaySOM application, where further processing of the received path is possible (see Fig. 7). With multiple devices connected to one PlaySOM application it is possible to combine and merge multiple paths sent by different devices to one common playlist, which is played on a local HiFi-device. This enables collaborative playlist generation e.g., for a party, or the creation of a shared on-demand radio stream influenced by the audience. Multiple users send their paths and playlists to the central server connected to the HiFi-device playing the music. The server combines and merges the received paths thus adopting the playlist to best fit the demands of all users. Path merging is performed based on an algorithm originally developed for the MetroMap Visualisation of the SOM [24].

Another scenario for the PocketSOM is to connect to a central audio portal over the web. Depending on the current mood, the user can now connect to an existing "radio-stream" displayed on the map or draw a new trajectory on the map. This trajectory can instantiate a new live-streams or connect the user to a stream created by another user visiting similar regions on the map.

5 Conclusions and future work

In this paper we presented different ways to create, control and perceive ambient music spaces. Music consumption ranges from purposely selecting a specific track to almost unnoticed, but still present, background sound. Both extremes as well as any combination between them can be achieved and controlled with the systems presented. We described prototype systems that provide environments to experience music both in virtual and real spaces. We are currently working on a tighter coupling of these spaces. This will allow new forms of interaction to become possible. For example, users in the virtual and real world will be able to meet in discussion fora, focussing not only on communication aspects, but also experiencing the same sound environment, further strengthening the perception of being in a common place in the virtual world. Furthermore, users in the real and virtual worlds may contribute their music to joint playlists or feed it to the real and virtual music cafe, where it is being played at the appropriate locations. Music is controlled either collaboratively via a user's mobile devices or centralized via a DJ selecting streaming playlists in either the virtual or real world.

Future work will primarily focus on creating this coupling of real and virtual worlds, as well as on user studies to evaluate the usefulness of these systems in daily use. This is currently broken down into a set of individual user studies on playlist generation and evaluation, followed by group-based user studies on joint playlist generation. Additional investigations will focus on the perception of joint audio spaces in collaborative settings, e.g., in combination with chat environments.

Another initiative currently under investigation is the integration of game technology into the various musical spaces. While the present realizations offer themselves for—mostly aimless—consumption in different settings, and thus form a good basis for exploring and understanding different ways of music consumption, we aim at enriching these by providing goals and purposes. This will move music from a purely passive medium that is consumed to a more active role of attracting people (as it is currently used in shops and bars), as well as to more directly utilize it for social networking purposes. This may include the realization of non-user characters, i.e. avatars that are not controlled by users, but act on their own, who exhibit a certain behaviours, such as preferences for certain types of music at certain times of a day, preferences for certain group sizes to join socially, etc. This behaviours can be directly implemented into an avatar. Yet, it can also be modelled according to or learned from real users' behaviours. This will enrich interaction possibilities as well as the level of immersion by offering richer interaction possibilities and more realistic settings where music is consumed for different purposes. Furthermore, coupling these principles with music creation processes is an area to be investigated in more detail, with initial concepts aiming at a composition process where sound samples are combined by moving within a music space.

References

1. Baum D, Rauber A (2006) Emotional descriptors for map-based access to music libraries. In: Proceedings of the 9th international conference on Asian digital libraries. Kyoto, Japan
2. Downie JS (2003) Music information retrieval. In: Annual review of information science and technology, vol 37. Information Today, Medford, NJ, USA, pp 295–340
3. Feiten B, Günzel S (1994) Automatic indexing of a sound database using self-organizing neural nets. Comp Music J 18(3):53–65
4. Frank J, Lidy T, Peiszer E, Genswaider R, Rauber A (2008) Ambient music experience in real and virtual worlds using audio similarity. In: SAME '08: Proceeding of the 1st ACM international workshop on semantic ambient media experiences. ACM, New York, NY, USA, pp 9–16

5. Genswaider R, Berger H, Dittenbach M, Pesenhofer A, Merkl D, Rauber A, Lidy T (2008) Computational intelligence in multimedia processing: recent advances. A synthetic 3D multimedia environment. In: Studies in computational intelligence, vol 96. Springer, Berlin, pp 79–98
6. Ghias A, Logan J, Chamberlin D, Smith BC (1995) Query by humming: musical information retrieval in an audio database. In: Proceedings of the third ACM international conference on multimedia. ACM, New York, NY, USA, pp 231–236
7. Hitchner S, Murdoch J, Tzanetakis G (2007) Music browsing using a tabletop display. In: Proceedings of the 8th international conference on music information retrieval. Vienna, Austria, pp 175–176 (2007)
8. ISO (2002) Information technology—multimedia content description interface—part 4: audio. ISO/IEC 15938-4:2002. International Organisation for Standardisation
9. Knees P, Schedl M, Pohle T, Widmer G (2006) An innovative three-dimensional user interface for exploring music collections enriched. In: Proceedings of the 14th annual ACM international conference on multimedia. ACM, Santa Barbara, CA, USA, pp 17–24
10. Kohonen T (2001) Self-organizing maps. In: Springer series in information sciences, vol 30, 3rd edn. Springer, Berlin
11. Kohonen T, Kaski S, Lagus K, Salojärvi J, Honkela J, Paatero V, Saarela A (2000) Self-organization of a massive document collection. IEEE Trans. Neural Netw. 11(3):574–585
12. Laaksonen J, Moskela M, Oja E (1999) PicSOM: self-organizing maps for content-based image retrieval. In: Proceedings of the international joint conference on neural networks (IJCNN99). Washington, DC
13. Leitich S, Topf M (2007) Globe of music—music library visualisation unsing geosom. In: Proceedings of the 8th international conference on music information retrieval. Vienna, Austria, pp 167–170
14. Leitich S, Toth M (2007) PublicDJ—music selection in public spaces as multiplayer game. In: Proceedings of audio mostly 2007. Ilmenau, Germany
15. Lidy T, Rauber A (2005) Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In: Proceedings of the international conference on music information retrieval (ISMIR). London, UK, pp 34–41
16. Lidy T, Rauber A (2008) Machine learning techniques for multimedia. Classification and clustering of music for novel music access applications. In: Cognitive technologies. Springer, Berlin, pp 249–285
17. Lidy T, Rauber A, Pertusa A, Inesta JM (2007) Improving genre classification by combination of audio and symbolic descriptors using a transcription system. In: Proceedings of the international conference on music information retrieval (ISMIR). Vienna, Austria, pp 61–66
18. Lübbers D (2005) SoniXplorer: combining visualization and auralization for content-based exploration of music collections. In: Proceedings of the 6th international conference on music information retrieval (ISMIR 2005), pp 590–593
19. Mayer R, Aziz TA, Rauber A (2007) Visualising class distribution on self-organising maps. In: de Sá JM, Alexandre LA, Duch W, Mandic D (eds) Proceedings of the international conference on artificial neural networks (ICANN'07), LNCS, vol 4669. Springer, Porto, Portugal, pp 359–368
20. Mayer R, Lidy T, Rauber A (2006) The map of mozart. In: Proceedings of the international conference on music information retrieval (ISMIR). Victoria, Canada
21. McNab RJ, Smith LA, Witten IH, Henderson CL, Cunningham SJ (1996) Towards the digital music library: tune retrieval from acoustic input. In: Proceedings of the first ACM international conference on digital libraries (DL '96). ACM, New York, NY, USA, pp 11–18
22. Neumayer R, Dittenbach M, Rauber A (2005) PlaySOM and PocketSOMPlayer—alternative interfaces to large music collections. In: Proceedings of the international conference on music information retrieval (ISMIR). London, UK, pp 618–623
23. Neumayer R, Frank J, Hlavac P, Lidy T, Rauber A (2007) Bringing mobile based map access to digital audio to the end user. In: Proceedings of the 14th international conference on image analysis and processing (ICIAP 2007) - workshop on video and multimedia digital libraries (VMDL07). IEEE Computer Society, Modena, Italy, pp 9–14
24. Neumayer R, Mayer R, Pölzlbauer G, Rauber A (2007) The metro visualisation of component planes for self-organising maps. In: Proceedings of the international joint conference on neural networks (IJCNN'07). IEEE Computer Society, Orlando, FL, USA
25. Ong TH, Chen H, Sung W, Zhu B (2005) Newsmap: a knowledge map for online news. Decis Support Syst 39(4):583–597

26. Orio N (2006) Music retrieval: a tutorial and review. Foundations and Trends in Information Retrieval 1(1):1–90
27. Pampalk E, Rauber A, Merkl D (2002) Content-based organization and visualization of music archives. In: Proceedings of ACM multimedia 2002. Juan-les-Pins, France, pp 570–579
28. Rauber A, Frühwirth M (2001) Automatically analyzing and organizing music archives. In: Proceedings of the 5th European conference on research and advanced technology for digital libraries (ECDL 2001). Springer lecture notes in computer science. Springer, Darmstadt, Germany
29. Rauber A, Merkl D (1999) SOMLib: a digital library system based on neural networks. In: Fox E, Rowe N (eds) Proceedings of the ACM conference on digital libraries (ACMDL'99). ACM, Berkeley, CA, pp 240–241
30. Rauber A, Pampalk E, Merkl D (2002) 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 (ISMIR). Paris, France, pp 71–80
31. Rauber A, Pampalk E, Merkl D (2003) The SOM-enhanced jukebox: organization and visualization of music collections based on perceptual models. J New Music Res 32(2):193–210
32. Torrens M, Hertzog P, Arcos JL (2004) Visualizing and exploring personal music libraries. In: Proceedings of the 5th international conference on music information retrieval (ISMIR 2004)
33. Tzanetakis G, Cook P (2001) Marsyas3D: a prototype audio browser-editor using a large scale immersive visual and audio display. In: Proceedings of the international conference on auditory display
34. Tzanetakis G, Cook P (2004) Music analysis and retrieval systems for audio signals. J Am Soc Inf Sci Technol 55(12):1077–1083
35. Ultsch A (2003) Pareto density estimation: a density estimation for knowledge discovery. In: Baier D, Wernecke KD (eds) Innovations in classification, data science, and information systems - proceedings 27th annual conference of the German classification society (GfKL), pp. 91–100. Springer, Berlin (2003)
36. Ultsch A (2004) U*-matrix: a tool to visualize clusters in high dimensional data. Tech. rep., Department of Mathematics and Computer Science, Philipps-University, Marburg, Germany
37. Ultsch A, Siemon HP (1990) Kohonen's self-organizing feature maps for exploratory data analysis. In: Proceedings of the international neural network conference (INNC'90). Kluwer, Dordrecht, Netherlands, pp 305–308
38. Zwicker E, Fastl H (1999) Psychoacoustics—facts and models. In: Springer series of information sciences, vol 22. Springer, Berlin

Jakob Frank is a Research Assistant at the Department of Software Technology and Interactive Systems of the Vienna University of Technology (TU Vienna). He received his Bachelor in Computer Science from the Vienna University of Technology in 2006. His research focus is on music information retrieval, especially on mobile devices and multi-user audio interaction. He was co-organizer of the ISMIR 2007 conference and served as co-reviewer for several major international conferences.

Thomas Lidy is a Research Assistant at the Department of Software Technology and Interactive Systems of the Vienna University of Technology (TU Vienna). He received his MSc in Computer Science from the Vienna University of Technology in 2007. His research focus is on music information retrieval, in particular feature extraction methods for digital audio, music classification, and clustering and visualization of digital music libraries. He participates actively in the annual MIREX benchmarking campaign and was co-organizer of the ISMIR 2007 conference. He is author of numerous papers in refereed international conferences and workshops and served as co-reviewer for several major international conferences. In 2007, he was awarded the Distinguished Young Alumnus Award and also received a Microsoft Sponsorship Award.

Ewald Peiszer is a freelance web application and software developer with a strong scientific background. He received his MSc degree in Computer Science from Vienna University of Technology in 2007 with a master's thesis on automatic audio segmentation. Working towards combining Music Information Retrieval (MIR) techniques with Virtual Reality infrastructure he completed an internship at the Center for Computer Graphics and Virtual Reality, Ewha Womans University (Seoul). Occasionally, he (co-)authors articles on MIR topics which is also a focus of his freelance projects.

Ronald Genswaider graduated as Master of Economics in 2008 at the Department of Software Technology and Interactive Systems of the Vienna University of Technology (TU Vienna) as well as Master of Arts in the Department of Digital Arts at the University of Applied Arts in Vienna. He is working in Vienna as a free digital artist, Web developer and researcher. Currently he is working in various research projects in the R&D department at bwin and taking part in the exhibition "YOU_ser—Century of the consumer" at the ZKM in Karlsruhe, Germany.

Andreas Rauber is Associate Professor at the Department of Software Technology and Interactive Systems of the Vienna University of Technology (TU Vienna). He received his MSc and PhD in Computer Science from the Vienna University of Technology in 1997 and 2000, respectively. He is actively involved in several research projects in the field of Digital Libraries, focusing on text and music information retrieval, the organization and exploration of large information spaces, as well as Web archiving and digital preservation. He has published numerous papers in refereed journals and international conferences and served as PC member and reviewer for several major journals, conferences and workshops. He also co-organized the ECDL 2005 and ISMIR 2007 conferences.