Ambient Music Experience in Real and Virtual Worlds Using Audio Similarity

Jakob Frank, Thomas Lidy, Ewald Peiszer, Ronald Genswaider, Andreas Rauber
Department of Software Technology and Interactive Systems
Vienna University of Technology
Favoritenstraße 9-11, 1040 Vienna, Austria
{frank,lidy,rauber}@ifs.tuwien.ac.at, mail@ewald-peiszer.net, r.genswaider@gmx.at

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 may 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 on 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 creation of playlists via artist or album selection, up to more artistic installation of sound environments that users can navigate through. In this paper we will 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 some common principles and point out further directions for a more direct coupling of the various spaces and interaction methods.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation (e.g., HCI)**]: Multimedia Information Systems—*Artificial, augmented, and virtual realities*; H.5.5 [**Information Interfaces and Presentation (e.g., HCI)**]: Sound and Music Computing—*Systems*

General Terms

Human Factors, Design

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 it 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 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 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, 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, namely the creation of a *virtual MusicSOM Cafe* in the *AudioSquare*, 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.

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 setting 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, concerning appropriate selection of music from large repositories. A need for sophisticated techniques for retrieval emerged that go 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 [14] contains a more recent review of the many different aspects of music processing and retrieval and considers also the role of the users. Moreover, an overview of prototypical Music-IR systems is given.

One kind of systems for retrieving music are Query-by-Humming systems that allow to query songs by singing or humming melodies [4, 11]. While introduced already in the mid-1990s, today, this technique has reached a mature state and was implemented 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 [17]. Tzanetakis and Cook introduced Marsyas3D, an audio browser and editor for collaborative work on large

sound collections [18]. A large-scale 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, covering critical frequency bands of the human auditory range and describing fluctuations with different modulation frequencies on them. The algorithm was first introduced in [15] and enhanced later by the inclusion of psycho-acoustic models in [16]. The feature set has proven to be applicable to both classification of music into genres [7] and automatic clustering of music archives according to the perceived sound similarity [10]. Furthermore, the correspondence of the resulting organization with emotional interpretations of the sound in various regions of the map has been analyzed [1].

In order to cluster the music according to perceived sound similarity the Self-Organizing Map (SOM) algorithm is employed [6]. The SOM is a topology-preserving mapping approach, that maps high-dimensional input data – in our case the features extracted from audio – to a 2-dimensional map space. 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 paper.

Previously, we presented PlaySOM, a 2D desktop application offering interactive music maps and the PocketSOM-Player, 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 [12]. 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 [5]. 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 [9]. 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. In contrast to these works, the applications presented in this paper allow users to immerse into more familiar environments and enable to meet and interact with other people in a social environment (virtual multi-user world, real-world cafe, scenarios with mobile phones).

3. TECHNICAL FUNDAMENTALS

The application scenarios we are presenting in Section 4 make use of automatic spatial arrangement 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 content extraction meth-

---

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

ods 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, 14]. Part of this research is the development of feature extraction methods 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 organization of music based on content or automatic classification of music into genres.

One such method suitable for describing acoustic characteristics of music is the Rhythm Pattern feature extractor [16]. A Rhythm Pattern describes fluctuations on critical frequency bands of the human auditory range. It reflects the rhythmical structure of a piece of music and also contains information about the timbre. The algorithm for extracting a Rhythm Pattern is a two stage process: First, from the spectral data the specific loudness sensation in Sone is computed for 24 critical frequency bands. Second, this Sonogram is transformed into a time-invariant domain resulting in a representation of modulation amplitudes per modulation frequency.

A Rhythm Pattern is typically computed for every third segment of 6 seconds length in a song, and the feature set for a song is computed by taking the median of multiple Rhythm Patterns. A Rhythm Pattern constitutes a comparable representation of a song, which can be used in clustering and classification tasks or for similarity retrieval.

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 [6].

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 an activation function. The Euclidean distance between the weight vector of the unit and the input vector is frequently used for the activation function. 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 as compared to the winner. This process is repeated for a large number of
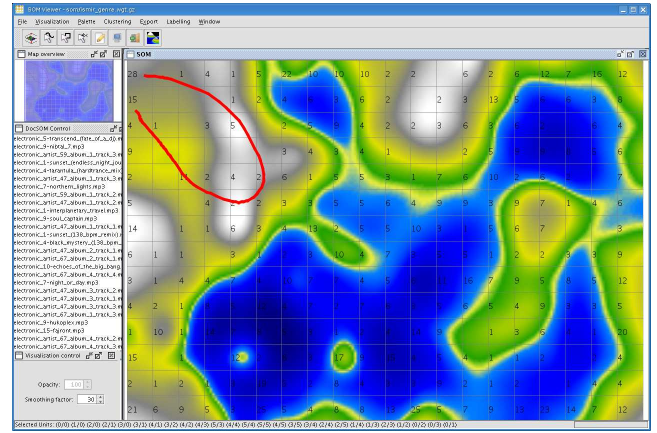

Figure 1: PlaySOM interactive 2D music map application.

iterations, presenting each input vector in the input space multiple times to the SOM.

The result is 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.

PlaySOM is an application that allows the creation of music maps using the SOM algorithm [12]. On top of that, it offers a number of interaction features: It provides an easy-to-use 2D desktop interface presenting a music map to access a music collection that was previously processed by an audio feature extractor. The main window of PlaySOM consists of the music map and allows the user to select songs for replay by drawing on it. Two modes of music 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 one or various others, 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 1 shows the main screen of PlaySOM with an example music map and a trajectory selection. A very important feature of PlaySOM is the implementation of various visualizations [8], which aim at helping the users to orient themselves on the map and aid in finding the desired music. To gain a more detailed view, the users can use the semantic zooming feature providing 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. Regarding the former, PlaySOM is used to generate the organization of the music maps for all of the applications presented in Section 4. For music maps on mobile devices also the graphical representation is used.
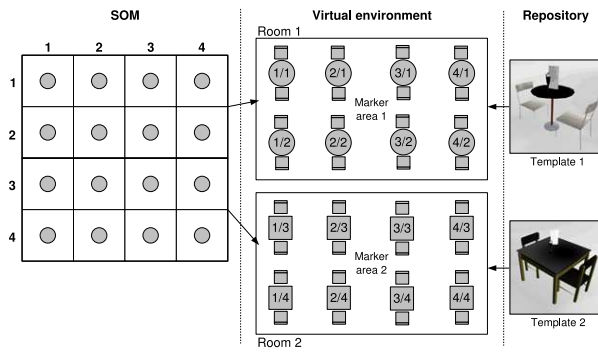
(a) The setup scheme of the *MusicSOM Showroom*

(b) Overview of the virtual *MusicSOM Showroom*

Figure 2: The *AudioSquare*'s *MusicSOM Showroom*: Automatic organization of music based on a SOM.

4. APPLICATIONS IN REAL AND VIRTUAL WORLDS

All following applications aim to create music spaces, some in real world, some in virtual world. However, 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 the organization of music starts to become difficult and mere metadata-search is not sufficient any more. A second source could be a huge commercial music portal offering its catalogue 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 created 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 (ignoring for the scope of this paper any legal aspects).
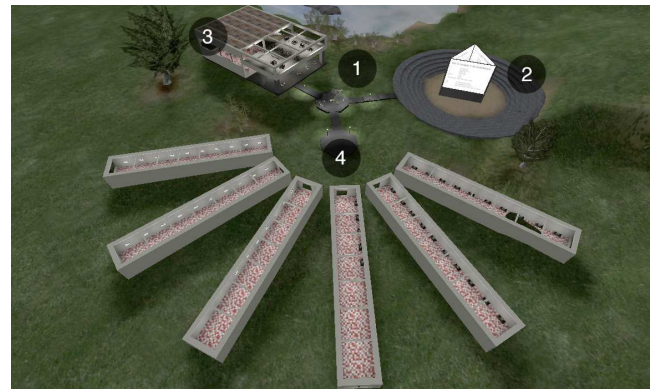
4.1 The AudioSquare

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 interactions 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 AudioSquare takes advantage of the virtual world paradigm for representing music archives. With the client application, 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. For the development of *The AudioSquare* the *Torque Game Engine*[2] has been used, providing many features, such as indoor as well as outdoor rendering, multi-user support, avatars and spatial audio playback.

The music within *The AudioSquare* 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 a their avatar and listening to the presented music streams. 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, a user can orientate not only by the visual feedback but also by perceiving the loudness and direction of the audio sources. The acoustic layer also helps the users in creating their mental map of the environment.

This environment can also be extended by including objects containing other media, such as images, presentations or video, as shown in *The MediaSquare* [3].

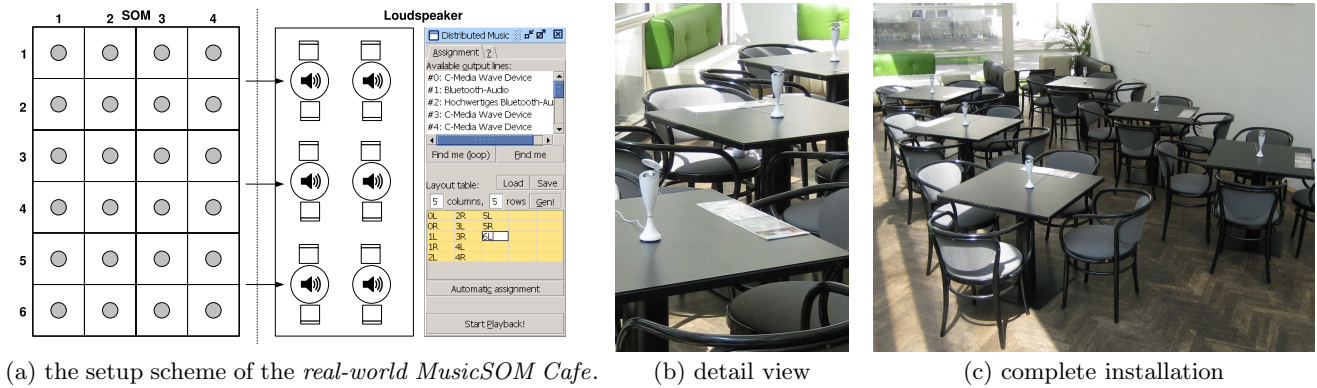**Figure 3: Bird's-eye view on the *AudioSquare*.**

---

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

(a) the setup scheme of the *real-world MusicSOM Cafe*.   (b) detail view   (c) complete installation

Figure 4: The *real-world MusicSOM Cafe*.

The workflow for creating *The AudioSquare* 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 Figure 2(a)).

Figure 5: Inside a building for the directory-based representation.

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. 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 rep-

resented 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.

An overview of the current implementation of *The AudioSquare* is depicted in Figure 3. Users enter the virtual world on a small welcome area (1) from where they can go to three different places. A big information screen (2) gives the users a quick overview of what they can do within the environment. The *MusicSOM Showroom* (3) presents music organized by a SOM. Finally, the *Manual Showrooms* (4) represent music organized manually in a folder-structure.

Figure 2(b) depicts the result of SOM-based organization in the virtual environment. It is a matrix of objects, each representing a unit of the SOM. 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. The directory-based approach is depicted in Figure 5. The user is located inside one of the arranged buildings. The objects on the left side represent the different audio streams. The HUD on the left shows information about the current audio track, the right one displays the full playlist for the closest audio source, while on the top the text chat is displayed.

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 are into. 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 our real-world MusicSOM, which basically is a real-world setup of the *AudioSquare*, installation 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.

(a) Bird's-eye view of *musicCAVE* in CaveLib simulator.
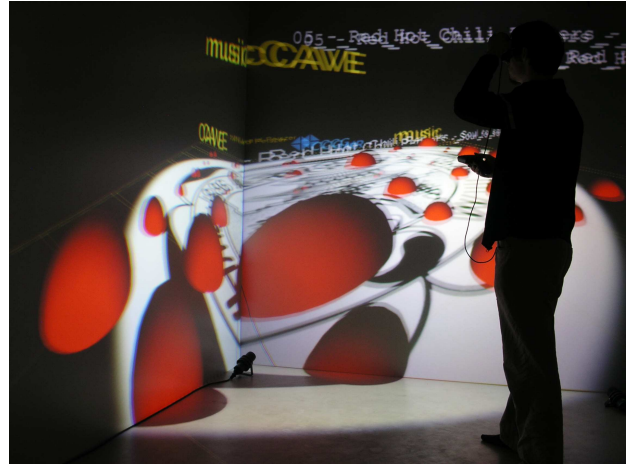
(b) One of the authors immersed in the *musicCAVE*.

Figure 6: The *musicCAVE* provides an immersive environment for music experience.

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, and private owned bars, cafes or lounges with a focus on individual and varied music experience.

To facilitate the setup of the installation we developed a plugin for the *PlaySOM* application that 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. We use the open source Tritonus implementation[3] of the Java Sound API. Features of the plugin include:

- playing short audio clips to each speaker in a loop to help identification during setup

- both automatic and manual assignment of SOM nodes to speakers

- saving and loading of assignments

- simultaneous playback of songs in random order to each speaker according to assignment

- 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.

The assignment is a 3 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 5 tables in 2 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 Figure 4(a)). Note that not all layout table cells need to be

used, some can remain empty. Third, a set of SOM nodes is assigned to each cell used, as illustrated in Figure 4(a). The songs that are represented by these nodes are played by the respective speaker in a random ordered loop. Figure 4(c) shows an demonstrational installation with six tables.

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 4-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[4]. 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 according to the distance of 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 two ore more songs can be heard, each coming from the direction of the node it is located in. The song 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 meter 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.

---

[3] http://tritonus.org/

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

(a) *PocketSOM* on BenQ P50

(b) *PocketSOM* on the iPhone

(c) *PocketSOM* on Nokia 7710

Figure 7: *PocketSOM* on different devices

The head tracker is used for perspective correction and 3D goggles are used to perceive the map in a stereoscopic view. Figure 6(a) is a screenshot of the CaveLib CAVE simulator, Figure 6(b) 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, which allow users to listen to their preferred music with them 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 personal computer at home. Yet, the music experience depends heavily on the interaction and access possibilites offered by the device.

Providing access to large audio collections using devices with limited interaction possibilities is a task that is currently under strong attention. One way is the usage of music
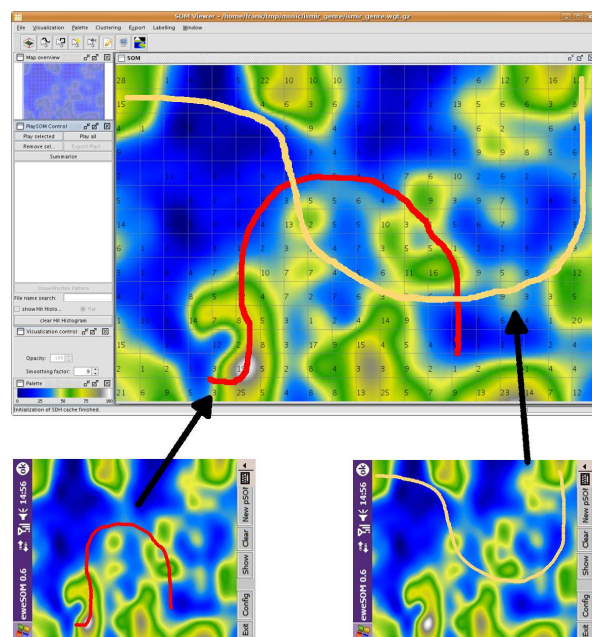

Figure 8: PocketSOM sending a path to the PlaySOM application

maps on these devices, as proposed in [13]. The *PocketSOM family* (see Figure 7), a series of implementations for different platforms, provides simple and intuitive SOM-based access to large audio collections.

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. 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. After creation, the playlist can be played on the mobile device using locally stored music or receiving the music via a web stream. Additionally, the playlist can be sent to a remote server for playback, turning the mobile device into a remote control.

A special enhancement of the PocketSOM was realized in conjunction with the PlaySOM application. The PocketSOM is 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 transfered to the PlaySOM application, where further processing of the received path is possible (see Figure 8). Moreover, 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 replayed 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.

Another scenario for the PocketSOM is to connect to a central audio portal over the web to receive the map from there. The user can then again draw a trajectory on the map to generate a live audio stream, or connect to another stream already created on the map.

5. CONCLUSIONS AND FUTURE WORK

In this paper we presented different ways to create, control and perceive ambient music experience. Music consumption ranges from purposely selecting a track to 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. By combining these spaces, new forms of interaction become possible. For example, users in the virtual and real world might meet in discussion fora, focussing not only on communication aspects, but also experiencing the same sound environment, further strengthening the perception of a physical 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 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 will be 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.

6. REFERENCES

[1] D. Baum and A. Rauber. Emotional descriptors for map-based access to music libraries. In *Proceedings of the 9th International Conference on Asian Digital Libraries*, Kyoto, Japan, November 27-30 2006.

[2] J. S. Downie. *Annual Review of Information Science and Technology*, volume 37, chapter Music information retrieval, pages 295–340. Information Today, Medford, NJ, USA, 2003.

[3] R. Genswaider, H. Berger, M. Dittenbach, A. Pesenhofer, D. Merkl, A. Rauber, and T. Lidy. *Computational Intelligence in Multimedia Processing: Recent Advances*, volume 96 of *Studies in Computational Intelligence*, chapter A Synthetic 3D Multimedia Environment, pages 79–98. Springer Berlin / Heidelberg, April 23 2008.

[4] A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith. Query by humming: musical information retrieval in an audio database. In *Proceedings of the third ACM international conference on Multimedia*, pages 231–236, New York, NY, USA, 1995.

[5] P. Knees, M. Schedl, T. Pohle, and G. Widmer. An innovative three-dimensional user interface for exploring music collections enriched. In *Proceedings of the 14th annual ACM International Conference on Multimedia*, pages 17–24, Santa Barbara, CA, USA, 2006.

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

[7] T. Lidy and A. Rauber. Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 34–41, London, UK, September 11–15 2005.

[8] T. Lidy and A. Rauber. *Machine Learning Techniques for Multimedia*, chapter Classification and Clustering of Music for Novel Music Access Applications, pages 249–285. Cognitive Technologies. Springer, Berlin Heidelberg, February 2008.

[9] D. Lübbers. 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)*, pages 590–593, 2005.

[10] R. Mayer, T. Lidy, and A. Rauber. The map of mozart. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Victoria, Canada, October 8-12 2006.

[11] R. J. McNab, L. A. Smith, I. H. Witten, C. L. Henderson, and S. J. Cunningham. Towards the digital music library: tune retrieval from acoustic input. In *Proceedings of the first ACM International Conference on Digital Libraries (DL '96)*, pages 11–18, New York, NY, USA, 1996.

[12] R. Neumayer, M. Dittenbach, and A. Rauber. PlaySOM and PocketSOMPlayer – alternative interfaces to large music collections. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 618–623, London, UK, September 11-15 2005.

[13] R. Neumayer, J. Frank, P. Hlavac, T. Lidy, and A. Rauber. 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)*, pages 9–14, Modena, Italy, September 10-14 2007.

[14] N. Orio. Music retrieval: A tutorial and review. *Foundations and Trends in Information Retrieval*, 1(1):1–90, September 2006.

[15] A. Rauber and M. Frühwirth. 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, Darmstadt, Germany, September 4–8 2001. Springer.

[16] 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 (ISMIR)*, pages 71–80, Paris, France, October 13–17 2002.

[17] M. Torrens, P. Hertzog, and J.-L. Arcos. Visualizing and exploring personal music libraries. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*, 2004.

[18] G. Tzanetakis and P. Cook. Marsyas3D: A prototype audio browser-editor using a large scale immersive visual and audio display. In *Proceedings of the International Conference on Auditory Display*, 2001.