Creation and Exploration of Musical Information Spaces

Andreas Rauber
Department of Software technology and Interactive Systems
Vienna University of Technology
http://www.ifs.tuwien.ac.at/~andi

Key Words
Music Libraries, Genre Analysis, Audio Clustering, Information Discovery and Retrieval, Visualization

Abstract
With the creation of large audio collections, available on small portable devices or via Web portals, new ways for interacting with these collections need to be devised. Retrieving specific songs, or navigating through such audio repositories, relying on the way music sounds rather than on metadata tags, poses challenging problems. This paper presents the *SOMeJB* system, which builds on features extracted directly from sound signals such as CD recordings or MP3 files, to create a navigable information space of music according to perceived sound similarities of the individual titles.

1. Introduction
Recent advances both in storage as well as compression techniques make large collections of audio data available on small portable devices. People can fit their complete music collection onto single devices, having all their music available at their fingertips. Furthermore, given current broadband connections that reach increasingly into private homes, music offers itself for electronic distribution. While currently a large part of electronic music distribution is illegal sharing of files, it nevertheless points at the potential that music offers as an e-commerce good. We thus find the music industry starting to explore new business models offered by the electronic distribution of music (Premkumar, 2003).

All these scenarios require ways to interact with the resulting musical digital libraries, be it in the form of Web Portals, home entertainment centers, or as mobile devices. Currently, access to musical data is mostly limited to queries against metadata tags, such as artist information, song titles, and (often utterly wrong, unnecessarily complex, or non-descriptive) genre tags. Serendipitous search is, at its best, supported by the browsing of more or less complex genre hierarchies and alphabetical artist listings. Even with additional metadata becoming available due to its creation during the production cycle of music we may not expect this approach to satisfy the needs of customers and users who want to access music by its primary characteristic and content, i.e. the way it sounds.

First approaches to access music via its musical "content" are offered by query-by-humming systems, where users hum a melody in order to retrieve the appropriate song. While this allows users to locate a specific song in their collection even without knowing title or artist information, it does not support serendipity. What we would thus like to offer is a way for users to explore a music repository, to give them an idea of which kind of music is to be found where in the collection, to allow them to create a mental model of a music library. We want to provide them with spatial orientation similar to the possibilities offered by conventional music "repositories", such as their favorite record store or their private music collection. In both these scenarios spatial orientation, the knowledge where to expect and find what type of music, are key characteristics that add significantly to the "usability" of the store or organization of their private CD collection.

In this paper we present the *SOM-enhanced JukeBox (SOMeJB)* system, which aims at providing such an organization of music according to sound similarity, resembling musical style or genres. The *SOMeJB* Music Digital Library Project, as first outlined in (Rauber & Frühwirth, 2001), and described in more detail in (Rauber, Pampalk & Merkl, 2002), (Rauber, Pampalk & Merkl, 2003) aims at creating such a browsable music archive by combining a variety of technologies from the fields of audio processing, neural networks, and information visualization, to create maps of music archives. It consists of three layers, namely (1) feature extraction, (2) organization by sound similarity, and (3) a navigational interface. In order to be able to group music by sound similarity a set of features needs to be identified to allow the computation of similarity. We have devised a representation based on *Rhythm Patterns*, which allow the capture of characteristics of music incorporating psychoacoustic models (Rauber, Pampalk & Merkl, 2002). On top of this *Rhythm Pattern* representation we use unsupervised neural networks, particularly the *Self-Organizing Map (SOM)* (Kohonen, 1995) as well as its extended model, the *Growing Hierarchical SOM (GHSOM)* (Dittenbach, Rauber & Merkl, 2002) to create a map or atlas type organization and representation of a music collection. Several visualizations, such as the *Islands of Music (IoM)*, which are based on the *Smoothed Data Histogram* representation (*SDH*) (Pampalk, Rauber & Merkl, 2002a) of the *SOM*, as well as labeling provided in the form of *Weather Charts*, offer an appealing and intuitive visualization of the resulting maps, allowing users to explore a large musical space, to locate their favorite regions, and to find music according to their likings (Pampalk, Rauber & Merkl, 2002).

The remainder of this paper is structured as follows: Section 2 briefly reviews some related work in the field of automatic genre analysis of music. Section 3 presents the extraction of the *Rhythm Pattern* features from the sound signal, followed by a presentation of the *SOM* and *GHSOM* neural networks in Section 4. The *IoM* visualization is briefly introduced in Section 5. Experimental results based on two different collections of music are discussed in detail in Section 6, with some conclusions rounding off the paper in Section 7. Source Code for all modules of the *SOMeJB* system, as well as on-line demos are available via the *SOMeJB* project homepage at http://www.ifs.tuwien.ac.at/~andi/somejb.

2. Related Work

A significant amount of research has been conducted in the area of content-based music retrieval, cf. (Foote 1999);(Wold, Blum, Keislar & Wheaton, 1996). Methods have been developed to search for pieces of music with a particular melody. Users may formulate a query by humming a melody, which is then usually transformed into a symbolic melody representation. This is matched against a database of scores given, for example, in MIDI format. Research in this direction is reported in, e.g. (Bainbridge, Nevill-Manning, Witten, Smith & McNab, 1999); (Birmingham, et al., 2001). Other than melodic information it is also possible to extract and search for style information using the MIDI format. Yet, only a small fraction of all electronically available pieces of music are available as MIDI. A more readily available format is the raw audio signal, which all other audio formats can be decoded to. A system where hummed queries are posed against an MP3 archive for melody-based retrieval is presented in (Liu & Tsai, 2001).

Specifically genre based organization and detection has gained significant interest recently. One of the first works to incorporate psychoacoustic modelling into the feature extraction process and utilizing the SOM for organizing audio data is reported in (Feiten & Günzel, 1994). A first approach, classifying audio recordings into speech, music, and environmental sounds is presented in (Zhang & Zhong, 1995). A system performing trajectory matching using *SOMs* and MFCCs is presented in (Spevak & Favreau, 2002). Specifically addressing the classification of sounds into different categories, (Wold, Blum, Keislar & Wheaton, 1996) use loudness, pitch, brightness, bandwidth, and harmonicity features to train classifiers. A wide range of musical surface features is used by the Marsyas system (Tzanetakis & Cook, 2000),(Tzanetakis & Cook, 2002) to organize music into different genre categories using a selection of classification algorithms.

3. Features: Rhythm Patterns

The feature extraction process for the Rhythm Patterns is composed of two stages. Following some pre-processing steps the specific loudness sensation in different frequency bands is computed, which is then transformed into a time-invariant representation based on the modulation frequency. Due to space considerations we only present a brief summary of the feature extraction algorithm. Interested readers are referred to (Rauber, Pampalk & Merkl, 2003) for detailed discussion.

3.1 Preprocessing

Starting from a standard *Pulse-Code-Modulated (PCM)* signal, stereo channels are combined into a mono signal, which is further downsampeled to 11kHz. Furthermore, pieces of music are cut into 6-second segments, removing the first and last two segments to eliminate lead-in and fade-out effects, and retaining only every second segment for further analysis.

3.2 Feature Extraction: Specific Loudness Sensation

Using a Fast Fourier Transform (FFT), the raw audio data is further decomposed into frequency ranges using Hanning Windows with 256 samples (corresponding to 23ms) with 50% overlap to counter ringing or side-lobe effects, resulting in 129 frequency values (at 43Hz intervals) every 12 ms. These frequency bands are further grouped into so-called *critical bands*, also referred to by their unit *bark* (Zwicker & Fastl, 1999), by summing up the values of the power spectrum between the limits of the respective critical band, resulting in 20 critical-band values. A *spreading function* (Schröder, Atal & Hall, 1979) is applied to account for *masking effects*, i.e. the masking of simultaneous or subsequent sounds by a given sound. The spread critical-band values are transformed into the logarithmic *decibel* scale, describing the sound pressure level in relation to the hearing threshold. Since the relationship between the dB-based sound pressure levels and our hearing sensation depends on the frequency of a tone, we calculate *loudness levels*, referred to as *phon*, using the equal-loudness contour matrix. From the loudness levels we calculate the specific loudness sensation per critical band, referred to as *sone*.

3.3 Feature Extraction: Amplitude Modulation

To obtain a time-invariant representation, reoccurring patterns in the individual critical bands, resembling rhythm, are extracted in the second stage of the feature extraction process. This is achieved by applying another discrete Fourier transform, resulting in *amplitude modulations of the loudness* in individual critical bands. These amplitude modulations have different effects on our hearing sensation depending on their frequency, the most significant of which, referred to as *fluctuation strength* (Fast, 1982), is most intense at 4Hz and decreasing towards 15Hz (followed by the sensation of *roughness*, and then by the sensation of three separately audible tones at around 150Hz). We thus weight the modulation amplitudes according to the fluctuation strength sensation, resulting in a time-invariant, comparable representation of the rhythmic patterns in the individual critical bands. To emphasize the differences between strongly reoccurring beats at fixed intervals a final gradient filter is applied, paired with subsequent Gaussian smoothing to diminish unnoticable variations. The resulting 1.200 dimensional feature vectors (20 critical bands times 60 amplitude modulation values) may optionally be reduced down to about 80 dimensions using PCA. These *Rhythm Patterns* are further used for data signal comparison.

4. SOM and GHSOM-based Organization

The *SOMeJB* system uses the topology-preserving capabilities of the *Self-Organizing Map* (*SOM*), as well as its extended model, the *Growing Hierarchical SOM (GHSOM)*, to create a map of a music collection, where similar pieces of music are located next to each other. The *SOM* (Kohonen, 1995) consists of a set of units i, which are arranged according to some topology, where the most common choice is a two-dimensional grid. Each of the units i is assigned a model vector m_i of the same dimension as the input data, $m_i \in \Re^n$, initialized e.g. to random values. During the training process input signals x are presented to the map in random order. An activation function based on some metric (e.g. the Euclidean Distance) is used to determine the winning unit (the `winner'). In the next step the weight vector of the winner is modified following some time-decreasing learning rate α in order to represent the input signal more closely. Apart from the winner, units in a time-varying and gradually shrinking neighborhood region h_{ci} around the winner are adapted as well, cf. Equation 1. This enables a spatial arrangement of the input patterns such that alike inputs are mapped onto regions close to each other in the grid of output units.

$$m_i(t+1) = m_i(t) + \alpha(t) \cdot h_{ci}(t) \cdot [x(t) - m_i(t)] \quad (1)$$

As two deficiencies of the standard *SOM* model we have to note its static architecture, which has to be defined prior to the training process, as well as the impossibility to faithfully reflect the hierarchical structure inherent in data. With the *GHSOM* (Dittenbach, Rauber & Merkl, 2002) we proposed a novel neural network model that addresses both deficiencies. It uses a hierarchical structure of multiple layers, where each layer consists of a number of independent *SOMs*. One *SOM* is used at the first layer of the hierarchy, representing the respective data in more detail. For every unit in this map a *SOM* might be added to the next layer of the hierarchy. This principle is repeated with the third and any further layers of the *GHSOM*.

To overcome the *SOMs* limit of a predefined network size we use an incrementally growing version of the *SOM*. This relieves us from the burden of predefining the network's size, which is rather determined during the unsupervised learning process. We start with a layer 0 consisting of only one single unit. The weight vector of this unit is initialized as the average of all input data. The training process then starts with a small map of, say, *2x2* units in layer 1, which is self-organized according to the standard *SOM* training algorithm. This training process is repeated for a fixed number λ of training iterations. Ever after λ training iterations the unit with the largest deviation between its weight vector and the input vectors represented by this very unit is selected as the error unit. Either a new row or a new column of units is inserted between the error unit and the neighboring unit most dissimilar in input space. The weight vectors of these new units are initialized as the average of their neighbors. This results in an unbalanced hierarchy of maps, where each map represents a subspace of the complete data set at increasing levels of granularity.

5. Visualization

The resulting maps offer themselves as interfaces to explore a music archive. Advanced cluster visualization techniques based on the SOM, such as the *U-Matrix* (Ultsch & Siemon, 1990), may be used to assist in cluster identification. A specifically appealing visualization based on *smoothed data histograms (SDH)* (Pampalk, Rauber & Merkl, 2002a) are the *Islands of Music*, which use the metaphor of geographical maps, where islands resemble styles of music, to provide an intuitive interface to music archives. Furthermore, attribute aggregates are used to create *Weather charts* that help the user in understanding the sound characteristics of the various areas on the map.

With *SDH* each data item, when presented to the map, ``votes'' for the map units which represent it best. All votes are accumulated for each map unit and the resulting distribution is visualized on the map. As voting function we use a robust ranking where the map unit closest to a data item gets *s* points, the second *s*-1, the third *s*-2 and so forth, for the *s* closest map units. All other map units are assigned 0 points. The parameter *s* can interactively be adjusted by the user. The concept of this visualization technique is basically a density estimation, thus the results resemble the probability density of the whole data set on the 2-dimensional map. The main advantage of this technique is its low computational cost. For a detailed discussion and evaluation of these visualizations, see (Pampalk, Rauber & Merkl, 2002).

6. Experiments

In this section we present results from two sets of experiments. The first setting is a rather small collection of 77 pieces of music (total playing time: approx. 5 hours) from a variety of genres. It will serve as a demonstrator of the basic capabilities of the *SOMeJB* system and the resulting IoM visualization, focusing on the explorative possibilities offered by our approach.

The second experiment is based on a much larger collection of 1.129 pieces of music with a total playing time of about 56 hours, structured into the 10 dance categories of the International Dance Sport federation (IDSF), namely the five *Latin-American* Dances Samba*, Cha-Cha-Cha, Rumba, Paso-Doble*, and *Jive*, as well as the five *Ballroom* Dances *Slow Waltz, Tango, Viennese Waltz, Slow Foxtrot*, and

Quickstep. Apart from the much larger size of this collection, and the resulting scalability issues being a focus of analysis of the system, this setting can also be used for a quantitative analysis of the system's organizational capabilities, as we can compare the organization created by the *SOMeJB* system with the pre-defined classification of the titles into the respective dance categories.

6.1 Experiment 1: Collection-77

This collection contains 77 pieces of music from some very diverse genres, ranging from classical music to heavy metal. Due to its size it offers itself for detailed presentation and discussion of the resulting map. The experiments presented in this section, including audio samples, are also available on the Web for interactive exploration.[1]

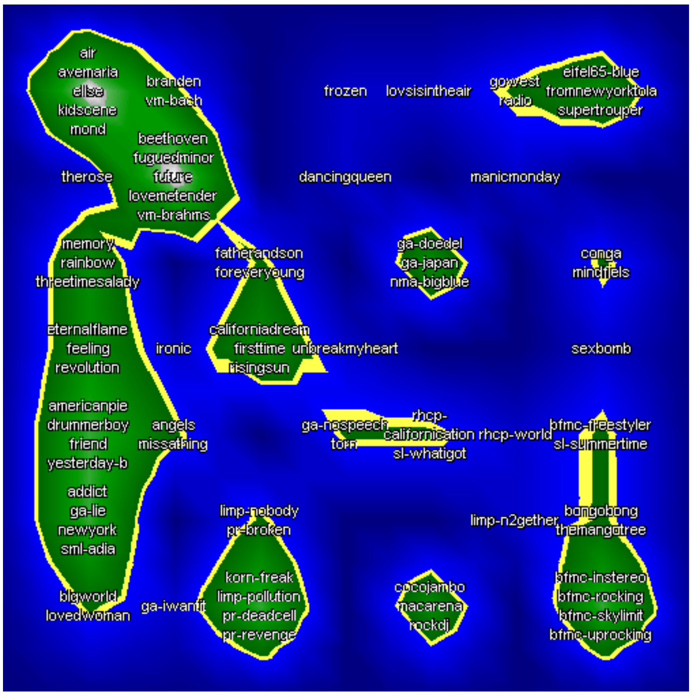

Figure 1: IoM Visualization of 7x7 SOM of Collection-77

Organizing these 77 pieces of music on a 7x7 *SOM* results in the map depicted in Figure 1 in the *Islands of Music (IoM)* representation. When analyzing the basic layout of the map, we find a rather large island on the left part of the map, consisting of basically two connected islands. The one in the upper left corner has two high peaks, depicted as snow-capped mountain tops. When looking at (or listening to) the titles mapped onto the respective units of this cluster we find them to be soft classical music, specifically *Air* from *Bach's Orchestersuite #3 , Ave Maria* by *Schubert, Für Elise* and the *Mondscheinsonate* by *Beethoven*, as well as *Schumann's Fremde Länder und Menschen*, on the corner unit. All of these titles are very soft, peaceful classical titles. Moving to the neighboring unit to the right takes us to a cluster of two titles, namely the *Andante* from the *Brandenburgisches Konzert #2* by *Bach*, as well as the *Bach Partita #3 in E for Solo Violin*. These are again clearly classical pieces of music, yet somewhat more dynamic.

Moving one unit down from this takes us to the second peak, which still is made up of predominantly classical music, yet again more dynamic, intensive, and basically orchestral pieces, as opposed to the

1 http://www.ifs.tuwien.ac.at/~andi/somejb/experiments/somejb2_col77/

predominantly solos constituting the clusters discussed so far. Titles here are the *1st Movement of the 5th Symphony* by *Beethoven*, the *Toccata and Fugue in D Minor* by *Bach*, which is a voluminous organ piece, *Brahms' Scherzo in C Minor*, or the end credits of the film *Back to the Future II*. Also on this unit, but not a strictly classical piece of music, and kind of announcing the transition to the next island, is the song *Love Me Tender* by *Elvis Presley*. As we move further down this second part of the island moving again one row to the left, we find more titles with vocals, such as *Memory* by *Barbara Streisand*, *Over the Rainbow* by *Judy Garland*, and *Three Times a Lady* by *Lionel Richie*.

Rather than continuing further south along this island, let us take a look at the island on the opposite lower right corner of the map. This is a smaller island, representing drastically heavier music, namely four titles by the band *Bomfunk MCs*, *In Stereo, Rocking just to make ya move, Sky's the Limit*, and *Uprocking Beats*. This music is characterized by the very strong and rhythmic bass beats as the dominant characteristic.

If we move to the left along the lower part of the map, we find two islands that exhibit less strong bass beats, but gradually faster beats that spread across a larger range of the spectrum. The first island we come across contains the titles *Cocojambo* by *Mr President, Macarena* by *Los del Rio*, and *Rock DJ* by *Robbie Williams*, while the next island is constituted by more aggressive titles with very fast rhythmic activities, especially in the higher frequency domains, by *Limp Bizkit, Papa Roaches*, and *Korn's Freak on a Leash*.

As a last example, we find Samba-Style music on the island in the upper right corner of the map, with titles such as *From New York to LA* by *Stephany McKay*, or *Super Trouper* by *A-Teens*.

We thus find the resulting *SOM* to produce a sensible organization and grouping of music according to sound similarities. The map can be explored by moving to various islands, each of which contains music of largely similar style. Due to the topological representation of the music archive a user can get accustomed to the representation, knowing which type of music to expect in which part of the map, and where to go for her or his favourite styles.

6.2 Experiment 2: Dance Music

As opposed to the predominantly qualitative evaluation of the previous small-scale example, we present here a partially quantitative evaluation of a large collection of dance music, organized into the 10 different IDSF dances. This setting is intended to demonstrate two characteristics of the *SOMeJB* system, namely its basic ability to organize music according to different styles of music, as well as the fact that – for creating this organization – the system uses more than the mere beat or dominant rhythm information. The former can be evaluated by the degree to which the system organizes the titles according to the pre-defined classification scheme, while the latter can be qualitatively verified by analyzing in how far sound similarity dominates over pure beat similarity of several pieces of music.

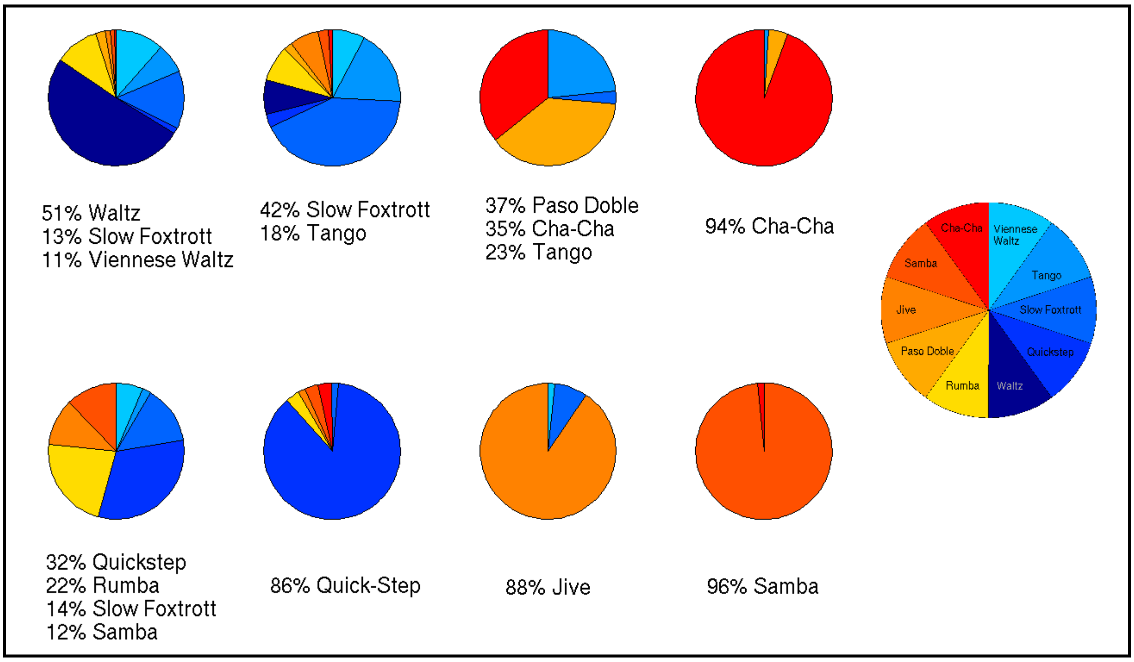

Figure 2: Cluster purity for each of the 2x4 units of the top-layer map

Due to the size of the data collection we decided to train a *GHSOM* rather than a flat map representation. The top-layer map of the *GHSOM* evolved to 4x2 units, with each of the units being expanded onto a more detailed second-layer map. Figure 2 depicts the cluster purity of each of the units. We find especially the Latin-American dances to be well separated in the right half of the map, with the upper and lower corner units having cluster purities of 94% for *Cha-Cha-Cha* and 96% for *Samba* music. Next to the bottom right corner unit of *Samba* music we find a unit representing 66 pieces of music, of which 88% are *Jive*, next to the first cluster of *Ballroom* music, of which 86% percent are *Quickstep*. This clear separation can be attributed to the clear dominance of rhythmic characteristics of each of these dances. For the major part of *Ballroom* music, the separation on the top-level map is not as clean, as these dances are not as strongly characterized – from a sound perspective – by their rhythmic characteristics, but rather by different styles, instrumentations, etc. Still, we find the upper left corner unit to be made up of 51% of *Waltz* music, followed by 13% of *Slow Foxtrot* and 11% of *Viennese Waltz*. The neighboring unit to the right contains 42% of *Slow Foxtrot*, followed by 18% of *Tango*. The neighboring third unit in the top row constitutes the border between the *Ballroom* and *Latin-American* sections of the map, and thus acts as an interpolating unit, containing *Tango* and *Cha-Cha-Cha* music with respect to its position in between these two dominant clusters to the left and to the right, as well as almost all *Paso Doble* titles.

If we drill down this unit, which in total represents 99 titles, and take a look at the corresponding second-layer map depicted in Figure 3, we find the individual dances to be well separated, as well as the global topology preserved: The upper-left part of the map contains almost exclusively *Tango* music, oriented to the neighboring map to the left, while *Cha-Cha-Cha* is locate don the lower right part of the map, neighboring the pure *Cha-Cha-Cha* cluster in the top-right corner of the first layer map. The *Paso-Doble* is spread across the upper right quarter of the map.

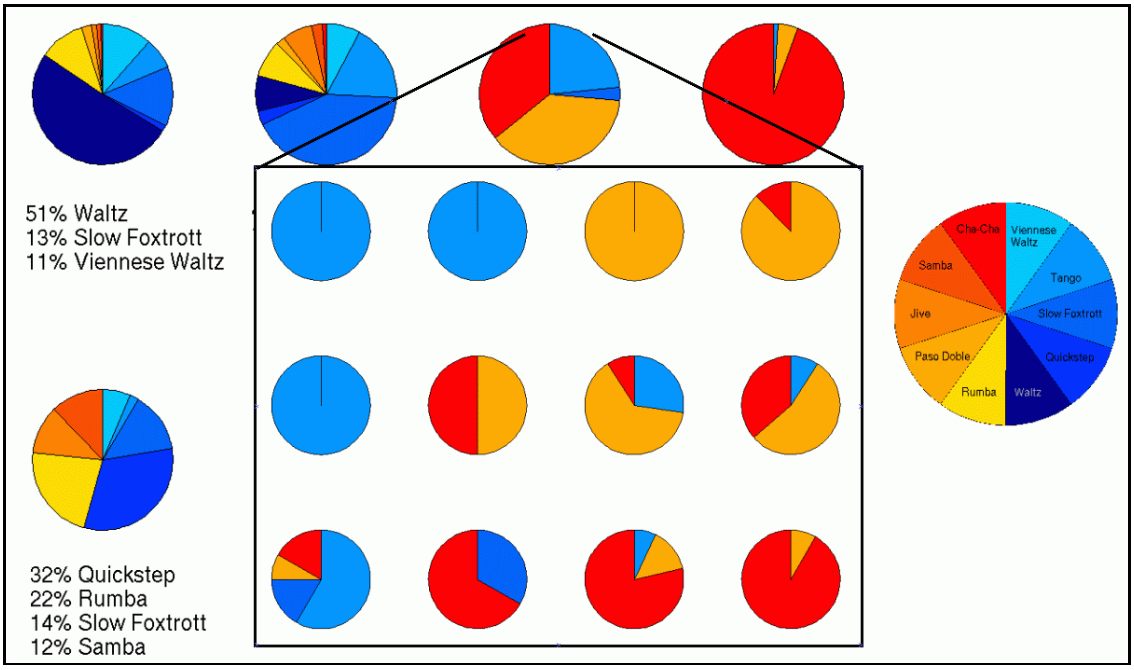

Figure 3: Cluster purity for each of the 4x3 units of the second-layer map

Apart from evaluating the cluster purity of the individual units, we can also take a look at how the individual dances are distributed across the various map units, as shown in Figure 4. Again, we can find a strong grouping of the individual dances on the map, following largely the trends observed during the evaluation of cluster purity, with most dances being located predominantly on one unit, with a few titles spread onto the neighboring units. An exception to this is *Quickstep*, which is distributed to almost equal shares across two neighboring units in the lower left corner.

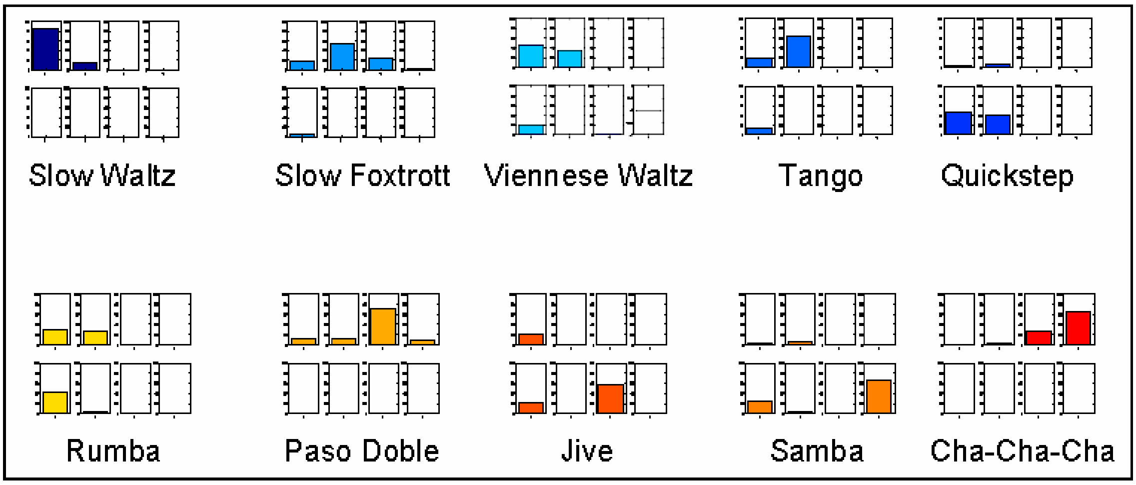

Figure 4: Distribution of the 10 Dances across 2x4 Top-Layer GHSOM map

However, two dances, namely *Samba* and *Cha-Cha-Cha*, are spread to notable degrees across two non-neighboring units. This raises the question, why titles from the same dance genre should be split far apart in spite of their identity in the dominant beat structure, hinting at the existence of subgroups within each of these dances. Analyzing a few titles from each of the sub-maps reveals that this is in fact the case. While typical *Samba* music is located on the bottom right corner unit, we find *Samba* titles interpreted and played with an instrumentation that makes them almost indistinguishable from *Rumba* music, with which these titles are co-located on a second layer map originating from the bottom left corner unit. In fact, the *Samba* rhythm is only faintly audible, while the dominant melodic part prevails in a *Rumba* style.

This situation becomes even clearer with a second example from the *Jive* genre, where a *Jive*, played in Big-Band instrumentation, is collocated with some *Quickstep* music, and which could, or possibly even should, be considered as a *Quickstep* both from the melodic as well as rhythmic point of view.

7. Conclusions

Experiments have demonstrated the capability of the *SOM-enhanced JukeBox (SOMeJB*) system to organize music according to sound similarities, offering a convenient interface to music repositories. It facilitates interactive exploration and navigation through the holdings of the collection, and can thus serve both as an interface to collections of audio files on portable devices, as well as for on-line music portals. It can be used to locate music according to ones likings by projecting preferred pieces of music onto the map space, where similar style music will be found in the immediate neighborhood. It can also serve as a convenient way of playlist generation by selecting a region containing a certain style of music on the map, rather than manually constructing playlists based on artist names or coarse and disputed genre metadata tags.

While the current system provides remarkable quality, the incorporation of additional features, specifically aiming at the capture of sound texture information, should offer room for further improvement and are currently under investigation. Additionally, while the focus of the proposed system definitely is on content-based features, i.e. features extracted directly from the sound signal of a piece of music, the integration of metadata information as well as social indexing is possible and may offer additional fascinating possibilities.

8. Bibliography

Bainbridge, D. , Nevill-Manning, C., Witten, H., Smith, L., & McNab, R. 1999.
Towards a digital library of popular music.
In Fox, E. and Rowe, N. editors, *Proceedings of the ACM Conference on Digital Libraries (ACMDL'99)*, pages 161-169, Berkeley, CA, August 11-14 1999. ACM.

Birmingham, W., Dannenberg, R., Wakefield, G., Bartsch, M., Bykowski, D., Mazzoni, D., Meek, C., Mellody, M., & Rand, W. 2001.
MUSART: Music retrieval via aural queries.
In *Proceedings of the 2nd Annual Symposium on Music Information Retrieval (ISMIR 2001)*, Bloomington, ID, October 15-17 2001.

Dittenbach, M., Rauber, A., & Merkl, D. 2002.
Uncovering hierarchical structure in data using the growing hierarchical self-organizing map.
Neurocomputing, 48:199-216, September 2002.

Fast, H. 1982.
Fluctuation strength and temporal masking patterns of amplitude-modulated broad-band noise.
In Hearing Research, 8:59--69, 1982.

Feiten, B. & Günzel, S. 1994.
Automatic indexing of a sound database using self-organizing neural nets. *Computer Music Journal*, 18:53-65, 1994.

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

Kohonen, T. 1995.
Self-organizing maps.
Springer-Verlag, Berlin, 1995.

Liu, C. & Tsai, P. 2001.
Content-based retrieval of mp3 music objects.
In Proceedings of the 10th International Conference on Information and Knowledge Management (CIKM 2001), pages 506-511, Atlanta, Georgia, 2001. ACM.

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

Pampalk, E., Rauber, A. & Merkl, D. 2002a.
Using smoothed data histograms for cluster visualization in self-organizing maps.
In *Proceedings of the International Conference on Neural Networks (ICANN 2002)*, pages 871-876, Madrid, Spain, August 27-30 2002. Springer.

Premkumar, G.P. 2003
Alternate Distribution Strategies for Digital Music.
In Communications of the ACM, 46(9):89-95, September 2003.

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, Darmstadt, Germany, Sept. 4-8 2001. Springer.

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 3rd International Conference on Music Information Retrieval*, pages 71-80, Paris, France, October 13-17 2002.

Rauber, A., Pampalk, E. & Merkl, D. 2003.
The SOM-enhanced JukeBox: Organization and Visualization of Music Collections based on Perceptual Models.
In: *Journal of New Music Research (JNMR)*, 32(2):193-210, June 2003.

Rauber, A., Dittenbach, M. & Merkl, D. 2002.
The Growing Hierarchical Self-Organizing Map:Exploratory Analysis of High-Dimensional Data.

In: *IEEE Transactions on Neural Networks*, 13(6):1331-1341, November 2002.

Schröder, M., Atal, B. & Hall, J. 1979.
Optimizing digital speech coders by exploiting masking properties of the human ear.
Journal of the Acoustical Society of America, 66:1647-1652, 1979.

Spevak, C. & Favreau, E. 2002.
Soundspotter - a prototype system for content-based audio retrieval.
In *Proceedings of the 5. International Conference on Digital Audio Effects (DAFx-02)*, Hamburg, Germany, September 26-28 2002.

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

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

Ultsch, A. & Siemon, H. 1990.
Kohonen's self-organizing feature maps for exploratory data analysis.
In *Proceedings of the International Neural Network Conference (INNC'90)*, pages 305-308, Dordrecht, Netherlands, 1990. Kluwer.

Wold, E., Blum, T., Keislar, D. & Wheaton, J. 1996.
Content-based classification search and retrieval of audio.
IEEE Multimedia, 3(3):27-36, Fall 1996.

Zhang, H. & Zhong, D. 1995.
A scheme for visual feature based image indexing.
In *Proceedings of the IS&T/SPIE Conference on Storage and Retrieval for Image and Video Databases*, pages 36-46, San Jose, CA, February 4-10 1995.

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