SOMBA : MULTIUSER MUSIC CREATION USING SELF-ORGANIZING MAPS AND MOTION TRACKING

Steven R. Ness, George Tzanetakis

University of Victoria
Department of Computer Science, Victoria, Canada
sness@sness.net,gtzan@cs.uvic.ca

ABSTRACT

SOMba is a system where multiple users create new rhythms and music by moving around a physical space while being tracked in real time using the infrared sensor on the Wii remote control. The physical space they move around in is mapped to a 2D Self-Organizing map. This SOM is created using the Marsyas audio processing framework from a collection of aligned 1-bar Samba rhythms of a variety of Brazilian musical instruments. A user is mapped to a unique point in the SOM, and this point contains a single rhythm. As the user moves around the 2D space, different rhythms are played. Multiple users can move around a space, and each would generate a different rhythm.

1. INTRODUCTION

Since the dawn of music in pre-historic tribal societies, music has often been accompanied by dance. In some traditions, for example in native american tribes, the dancers themselves would participate in the music making experience by wearing noise making artifacts. However, in most situations, music making and dancing have been relatively separate activities. When Samba bands perform in Brazil, musicians are joined by large numbers of dancers, who only indirectly participate in the music making experience.

Self-organizing maps [1] are a dimensionality reduction technique where a high dimension dataset is mapped to a lower, typically 2-dimensional surface. SOMs and other such dimensionality reduction tools have been used to allow users to visualize data relationships within complex datasets. There have been many applications of SOMs in Music Information Retrieval including work by automatically analyzing and organizing music archives [2] [3], visualizing music genres [4] and music recommendation [5].

A less explored frontier of SOMs and music is the production of new music. Until recent times, the production of music has been solely by highly trained musicians. Recently, commericial systems like Rock Band and Guitar Hero along with many academic research projects have attempted to put the control of the musical experience in the hands of less experienced users. In addition ystems like Dance Dance Revolution and other academic projects have begun to bridge the gap between dancing and music creation.

Our current research project proposes to let dancers create their own music, by generating music collaboratively with other dancers. We achieve this by tracking the positions of each of the dancers in two dimensions on a theatre stage. We can then translate these stage positions into positions in a 2D SOM. We have used the SOM methodology because it provides a straightforward way to map the high dimensional audio features to a two dimensional surface.

In the field of embodied music cognition it has been noted [6] that people seek involvement with music in order to experience a behavioural resonance with physical energy. Our system attempts to project both the spatial motion of the dancer, and her interactions with other dancers, into music. This allows dancers to become instrumentalists and intermedial composers, simultaneously interpreting the generative music composition. In traditional full-body tracked music generation systems, the gestures and movements of dancers are transformed into music, but in our system, we are more interested in relationships between the dancers within not only a physical space, but also within their socially constructed environment.

The SOM that we build contains aligned rhythmic phrases of different drum patterns from Samba music. Samba music is built from many overlayed drum patterns, with each sounding on different beats of a typically 2/4 metrical pattern. It is a highly syncopated music style with each instrument, from the large surdo drum, to the small hand held clave, playing on different beats of the bar. Although each pattern in a typical Samba song is simple, the patterns that result from all the instruments playing together can be quite complex and interesting.

1.1. Related Work

Self organizing maps have been used extensively in the visualization of data for audio based music information retrieval [7]. They have been used to analyze and organize music archives [2] [8] [3] [9], and to visualize the resulting music collections [10] [11] [12]. A particularily relevant study was that of Palmalk in his paper Islands of Music

[4]. SOMs have also been used for audio retrieval, browsing and constructivist learning in several papers [13] [14] [15]. While the previous mentioned studies concentrated on organizing whole classes of music, SOMs have also been applied to smaller audio segments, including timbre [16], energy-spectrum [17], and musical time series analysis [18].

There has been considerable research on the automatic generation of music, including the generation of background music [19] [20], the creation of rhythmic patterns [21] [22] and more general automated music generation systems [23] [24]. An excellent study was conducted back in 1970 by Howe [25] about compositional considerations when creating electronic music.

A particularily relevant study was recently conducted [22], in this paper the authors describe a self-organizing map system that allows users to create rhythms co-creatively and interactively. Also closely related was the [26] SENEgal project, which used genetic algorithms to create rhythms from western Africa.

Often the previous described systems have their user interaction paradigm centered on the computer system. In this study we are interested in bringing the creation of music into the physical environment. The creation of new methods of interacting with the computer has seen much activity, including projects using the Radiodrum [27] [28] , SmartSkin [29] and Cyber composer [30]. A particularily relevant paper involved the use of large numbers of giveaway sensors in a large rave dance setting [31].

Our project extends and simplifies this by using the infrared sensor capabilities of the Wii remote control, or wiimote. The wiimote connects to a computer via Bluetooth, and using the cwiid [32] library on Linux, we are able to easily access the data it provides. The wiimote has buttons and accelerometers, and also has the capability to track up to 4 infrared sources. The positions of these sources are tracked by the wiimote.

2. THE SOMBA SYSTEM

The SOMba system consists of three distinct parts, the generated music tracks for each of the different instruments, a process to map this audio onto a two dimensional representation, and a way to interact with the grid that is created. For each of these parts, we used the Marsyas [1] programming framework, a toolkit written in C++ that allows for the creation, analysis and output of audio.

2.1. Creation of Musical Tracks

We obtained transcriptions of authentic Samba music from Brazil and converted them into a simple string based machine readable representation. We also obtained from the RWC [33] library audio samples of some of the musical instruments used by Brazilian samba bands, including surdo, agogo bells, tambourim, shaker and quijada. A program was written to generate audio files of equal lengths using the transcription along with an audio sample for each instrument. The resulting audio was found to be too mechanical in feel, so we added the ability to add a small amount of random time jitter to each sample, and created several versions of each transcription with slightly different timings of each beat. We also created versions of each audio track with different audio effects, including high pass and low pass filters, as well as phaser and flanger effects.

2.2. Music Feature Extraction

Each audio track is represented as a single feature vector. Even though much more elaborate audio track representations have been proposed in the literature we have found that a single feature vector per audio clip is well suited to use in machine learning in general and the SOM algorithm in particular. It has been shown that such song-level features perform quite well [34].

The features used in our approach are Spectral Centroid, Roll-Off, Flux and Mel-Frequency Cepstral Coefficients (MFCC). To capture the feature we compute a running mean and standard deviation over the past M frame. This results in a feature vector of 32 dimensions at the same rate as the original 16-dimensional one. A more detailed description of the features can be found in Tzanetakis and Cook [35].

This process extracts audio features from the tracks, and to more accurately capture the rhythmic differences between tracks, we use the original string representation of the track and calculate the Hamming distance between it and a string that is composed solely of rests. We then add this measure to the feature vector. We also take the average jitter value for all the notes in the track and add this as another component of the feature vector.

2.3. Self-Organizing Map Generation

The resulting audio tracks were then mapped to a two dimensional space using Self-Organizing Maps (SOMs). SOMs are a technique for transforming data of multiple dimensions into a lower number of dimensions, typically two. They were first introduced in 1982 by Teuvo Kohonen [1] and were based on Artificial Neural Networks. SOMs are a form of dimensionality reduction, which is a broad term for any technique that transforms a higher dimensional space into a lower dimensional space.

We used the existing SOM implementation within Marsyas to transform the high dimensional feature vectors into a two-dimensional represention. For the neighbourhood function, we started with a value of 0.17, and decreased this by a factor of 0.98 for each round of training. We found that a 12x12 grid worked well for this collection of feature vectors and

[1] `http://marsyas.sourceforge.net`

training parameters and resulted in a final grid with all cells containing at least one track of audio.

2.4. Interaction

The Wii remote, or wiimote, is a multimodal interface device developed by Nintendo for use with the Wii game system. The wiimote has traditional buttons and rumble functionality, but also contains a speaker, an 3-dimensional accelerometer, and an Infrared (IR) sensor. This IR sensor has the ability to track up to 4 independent sources of infrared light, and reports back the positions and intensities of the detected points. All data from the wiimote is sent back to the computer via Bluetooth. We then use the OSC protocol to transmit messages recieved from the Wiimote controller to the SOMba program. In our current system we use the Wiimote for position tracking of the dancers, but the use of OSC allows us to use a variety of other position trackers quite easily.

Using OSC, we have added functionality that allows for the use of multiple Wiimotes at once, each one returning the positions of four dancers at once. In the canonical SOMba system, we use two wii-motes, allowing up to 8 dancers to be tracked, but this number could easily be expanded. It is important to align the sensors of the wiimotes accurately with the performance space and with each other, so that an individual dancer is only tracked by one wiimote at once. In real performance settings static and dynamic reflections and poor camera resolution deteriorate the tracking results. State of the art tracking technologies provide built-in multi-camera calibration and elaborate 3D point tracking algorithms to solve these problems. To help alleviate this problem in our simple system, we have found that it was important to use very bright infrared emitters, and to limit the intensity of stage lights on the dancers.

3. CONCLUSIONS

Dancing and music have long been close companions, and recent advances in technology can allow us to further blend these together into a new form of musical and physical articulation. By allowing dancers to interact with and create their own music, we hope to create new exciting opportunities for creative expression. We are also interested in the spatial representation of music, and by transferring Self-Organizing maps from the computer screen into physical space, we anticipate to discover new ways of interacting with and visualizing data. Another part of this research is to use self-organizing maps not just to organize different songs, but different rhythmic patterns. Such a project might prove interesting for creation of music using different interfaces.

4. ACKNOWLEDGEMENTS

We would like to thank Gabrielle Odowichuck and Manjinder Benning for assistance in this project.

5. REFERENCES

[1] Teuvo Kohonen, *Self-Organizing Maps*, Springer, Berlin, Heidelberg, 1995.

[2] A. Rauber and M. Fruhwirth, "Automatically analyzing and organizing music archives," in *Proc. Research and Advanced Technology for Digital Libraries. ECDL*, 2001, pp. 402–14.

[3] A. Rauber and D. Merkl, "Creating an order in distributed digital libraries by integrating independent self-organizing maps," in *Proc. International Conference on Artificial Neural Networks. (ICANN)*, 1998, vol. 2, pp. 773–8.

[4] E. Pampalk, S. Dixon, and G. Widmer, "Exploring music collections by browsing different views," in *Proc. Int. Conf. on Music Information Retrieval (ISMIR)*, 2003.

[5] S. Vembu and S. Baumann, "A self-organizing map based knowledge discovery for music recommendation systems," in *Computer Music Modeling and Retrieval*, 2005, pp. 275–284.

[6] Marc Leman, *Embodied Music Cognition and Mediation Technology*, The MIT Press, September 2007.

[7] Matthew Cooper, Jonathan Foote, Elias Pampalk, and George Tzanetakis, "Visualization in audio-based music information retrieval," *Computer Music Journal*, vol. 30, no. 2, pp. 42–62, 2006.

[8] A. Rauber and D. Merkl, "Organization of distributed digital libraries: a neural network-based approach," in *Intelligent Data Engineering and Learning (IDEAL)*, 1998, pp. 283–8.

[9] A. Rauber, E. Pampalk, and D. Merkl, "Content-based music indexing and organization," in *ACM Special Interest Group on Information Retrieval (SIGIR)*, 2002, pp. 409–410.

[10] A. Rauber, E. Pampalk, and D. Merkl, "The SOM-enhanced JukeBox: Organization and visualization of music collections based on perceptual models," *Journal of new Music Research*, vol. 32, no. 2, pp. 193–210, June 2003.

[11] Pampalk E., Dixon S., and Widmer G., "Exploring music collections by browsing different views," *Computer Music Journal. Summer 2004; 28(2): 49–62*, 2004.

[12] A. Rauber, E. Pampalk, and D. Merkl, "Using psycho-acoustic models and self-organizing maps to create a hierarchical structuring of music by sound similarities," in *Proc. Int. Symposium on Music Information Retrieval (ISMIR)*, 2002.

[13] Pedro Cano, Martin Kaltenbrunner, Fabien Gouyon, and Eloi Battle, "On the use of fastmap for audio retrieval and browsing," in *Proc. Int. Conf. on Music Information Retrieval (ISMIR)*, 2002.

[14] Markus Frühwirth and Andreas Rauber, "Self-Organizing Maps for Content-Based Music Clustering," in *Proc. Italien Workshop on Neural Nets (WIRN01)*. 2001, Springer.

[15] Timo Honkela, Teemu Leinonen, Kirsti Lonka, and Antti Raike, "Self-organizing maps and constructive learning," in *Proc. of ICEUT*, 2000, pp. 339–343.

[16] P. Toirvainen, "Optimizing self-organizing timbre maps: two approaches," *Music, Gestalt, and Computing. Studies in Cognitive and Systematic Musicology*, pp. 337 – 50, 1997.

[17] M. Masugi, "Energy spectrum-based analysis of musical sounds using self-organizing map," *IEICE Transactions on Information and Systems*, vol. E86-D, no. 9, pp. 1934 – 8.

[18] O.A.S. Carpinteiro, "A self-organizing map model for analysis of musical time series," in *Proc. Brazilian Symposium on Neural Networks*, 1998, pp. 140 – 5.

[19] C. Rui, L. Zhang, F. Jing, W. Lai, and W.Y.Ma, "Automated music video generation using web image resource," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2007, vol. 2, pp. 737–740.

[20] M.J. Yoo, I.K. Lee, and J.J. Choi, "Background music generation using music texture synthesis," in *Proc. Conf. in Entertainment Computing*, 2004, pp. 565 – 70.

[21] N. Labordus, "Design techniques for rhythm generators. II," *Electronics Australia*, vol. 45, no. 12, pp. 78 – 81, 1983.

[22] S. Kasahara, R. Saegusa, and S. Hashimoto, "Rhythm generation with associative self organizing maps," *Trans. of the Information Processing Society of Japan*, vol. 48, no. 12, pp. 3649 – 57, 2007.

[23] J. Erb, "Musicalc, the program for music fans," *HC Mein Home Computer*, vol. 1, no. 11, pp. 98 – 100, 1984.

[24] T. Unemi and E. Nakada, "A tool for composing short music pieces by means of breeding," in *IEEE Int. Conf. on Systems, Man and Cybernetics*, 2001, vol. vol.5, pp. 3458 – 63.

[25] H.S. Howe, "Compositional considerations in electronic music," in *Journal of the Audio Engineering Society*, 1970, vol. 18, p. 690.

[26] D. Tzimeas and E. Mangina, "SENEgaL: a GA system for generating rhythms of western africa," in *Int. MultiConference of Engineers and Computer Scientists*, 2007, vol. 1, pp. 25 – 30.

[27] M.S. Benning, M.M., and P. Driessen, "Improved position tracking of a 3-d gesture-based musical controller using a kalman filter," in *Proc. Int. Conference on New Interfaces for Musical Expression (NIME)*, 2007, pp. 334–337.

[28] J. Murdoch and G. Tzanetakis, "Interactive content-aware music browsing using the radio drum," in *Proc. Int. Conf. on Multimedia and Expo, (ICME)*, 2006, pp. 937 – 940.

[29] Jun Rekimoto, "Smartskin: an infrastructure for free-hand manipulation on interactive surfaces," in *Proc. of the SIGCHI conference on Human factors in computing systems (CHI)*, 2002, pp. 113–120.

[30] H.H.S. Ip, K.C.K. Law, and B. Kwong, "Cyber composer: hand gesture-driven intelligent music composition and generation," in *Proc. Int. Multimedia Modeling Conf.*, 2004, pp. 46 – 52.

[31] M. Feldmeier and J.A. Paradiso, "An interactive music environment for large groups with giveaway wireless motion sensors," *Computer Music Journal*, vol. 31, no. 1, pp. 50 – 67.

[32] L. Donnie Smith, *cwiid : Tools for interfacing to the Nintendo Wiimote*, 2007, Software available at `http://abstrakraft.org/cwiid/`.

[33] Masataka Goto and Takuichi Nishimura, "Rwc music database: Music genre database and musical instrument sound database," in *in ISMIR*, 2003, pp. 229–230.

[34] Michael Mandel and Daniel Ellis, "Song-level features and support vector machines for music classification," in *Proc. Int. Conf. on Music Information Retrieval (ISMIR)*, 2005.

[35] G. Tzanetakis and P. Cook, "Musical Genre Classification of Audio Signals," *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 5, July 2002.