Self-Organizing Topological Timbre Design Methodology Using a Kohonen Neural Network

Marcelo Caetano[1,2], César Costa[2], Jônatas Manzolli[2], and Fernando Von Zuben[1]

[1]Laboratory of Bioinformatics and Bio-inspired Computing (LBiC)
[2]Interdisciplinary Nucleus for Sound Studies (NICS)
University of Campinas (Unicamp), PO Box 6101 - 13083-970, Brazil

{caetano,vonzuben}@dca.fee.unicamp.br; {jonatas,cesar}@nics.unicamp.br

Abstract. *Generating sounds for music composition with the desired timbral characteristics has been a challenge ever since the dawn of electroacoustic music. Timbre is a remarkably complex phenomenon that has puzzled researchers for a long time. Actually, the nature of musical signals is not fully understood yet. In this paper, we present a sound synthesis technique that uses Kohonen's one-dimensional self-organizing map to generate neuronal-sounds to respond to a fixed and predefined set of stimulus-sounds, producing timbral variants with the desired characteristics. The self-organizing algorithm provides maintenance of topology so that the intended aesthetical result is properly achieved by avoiding the formal definition of the timbral attributes. To evaluate the obtained results we propose crossing a mathematical/subjective spectral distance from the neuronal-sounds to the stimulus-sounds with the method of timbral classification using Kohonen's two-dimensional self-organizing map.*

1. Introduction

Computer music is an ever-growing field partly because it allows the composer such great flexibility in sound manipulation when searching for the desired result. In the particular case of music composition, once the search space and the goals are defined, a technique for achieving the final product is required. Within the frame of this work, when we consider music improvisation, there is no goal and no such thing as a final result. It is the actual path through the search space that is of interest. Many different approaches have been proposed to meet the requirements of the process, i.e. creating interesting music, with results that vary from the unexpected to the undesired, depending upon a vast number of factors and on the methodology itself. Traditional sound synthesis techniques present limitations especially due to the fact that they do not take into consideration the subjective and/or the dynamic nature of music, by using processes that are either too simple or not specifically designed to handle musical sounds [Caetano et al. 2005 a].

Musical timbre is the characteristic tone quality of a particular class of sounds. Timbre is much more difficult to characterize than either loudness or pitch. No one-dimensional scale – such as the loud/soft of intensity or the high/low of pitch – has been postulated for timbre, because there exists no simple pair of opposites between which a scale can be made. Because timbre has so many facets, computer techniques for multidimensional scaling [Grey 1975; Grey and Moorer 1977] have constituted the first major progress in quantitative description of timbre, since the pioneering work of Hermann von Helmholtz (1885) in the nineteenth century. Since then, researchers have determined a more accurate model of natural sound. Digital recording has enabled the contemporary researcher to show that the waveform (and hence the spectrum) can change drastically during the course of a tone. Risset and Wessel (1982) observed that complex sounds have dynamic spectra and the evolution in time of the sound's spectrum plays an important part in the perception of timbre [Grey and Moorer 1977]. Timbre variations are perceived, for example, as clusters of sounds played by a particular musical instrument, or said by a particular person, even though these sounds might be very distinct among themselves, depending upon its pitch, intensity or duration. In fact, the concept of timbre has always been related to sounds of musical instruments or speech, and it is in this scope that the majority of researches on timbre have been developed. These works identified innumerable factors that form what is called timbre perception.

Many researchers have recently suggested the creation of Bio-Inspired and Artificial Intelligence (AI) based systems for music composition and improvisation. Applications of Bio-Inspiration and AI in music composition involve artificial neural networks [Chen and Miikulainen 2001], cellular automata [Burraston et al. 2004], artificial immune systems (AIS) [Caetano et al. 2005 a], particle swarms [Blackwell and Young 2004] and evolutionary computation (EC) [Biles 1994; Horowits 1994; Caetano et al. 2005 b]. Refer to the work of Santos et al. (2000) for a detailed review of the application of EC in music systems. As a preliminary step toward the current proposal, Caetano et al. (2005 a,b) suggested the use of AI to pursue stationary/fixed target sounds that are considered the user's desired timbral outcome. The reported results can be interpreted as a sort of spectral blend between the initial and target sounds.

In this work, we are focusing primarily on the production of sounds that present complex spectral dynamic features for musical applications taking self-organization as paradigm. We propose the use of a one-dimensional Self-Organizing Map (SOM) in our approach to timbre design, founded on unsupervised learning and on the ability of SOM to orderly arrange the original soundspace in a cyclical fashion, proposing a tentative timbral improvisational scale.

SOMs are the most commonly used strategy in Artificial Neural Networks (ANNs) for unsupervised learning [Lippman 1987]. During the training process the neurons tend to represent statistical properties of the input data, preserving the topology of the input space (soundspace), even though it is unknown. It is a handy tool for feature analysis of high-dimensional data, allowing visualization in a low-dimensional neuron layer space.

The main application of SOM is data clustering using two-dimensional mapping [Kohonen 1984 b]. One-dimensional SOM has been applied to the Travelling Salesman Problem (TSP) [Gomes and Von Zuben 2003] and as a topological ordering method of multidimensional data. We found several different proposals of timbre taxonomical classification making use of SOM as a feature extraction tool [Damiani et al. 1995; Cosi et al. 1994 a,b; De Poli and Tonella 1993; De Poli and Prandoni 1997; Loureiro et al. 2004; Feiten and Gunzel 1994]. The authors did not find any applications of one-dimensional SOMs in timbre design in the literature.

The subsequent sections describe the fundamentals of SOMs and the way they are related to the development of our timbre design technique. Experimental results are then described and analyzed. Finally, concluding remarks and perspectives for further research are considered.

2. Kohonen's Self-Organizing Feature Maps

One important organizing principle of sensory pathways in the brain is that the placement of neurons is orderly and often reflects some physical characteristic of the external stimulus being sensed [Kandel and Schwartz 1985]. For example, at each level of the auditory pathway, nerve cells and fibers are arranged anatomically in relation to the frequency which elicits the greatest response in each neuron. This tonotopic organization in the auditory pathway extends up to the auditory cortex [Moller 1983]. Although much of the low-level organization is genetically pre-determined, it is likely that some of the organization at higher levels is created during learning by algorithms which promote self-organization. Kohonen (2000) presents one such algorithm which produces what he calls self-organizing feature maps (SOMs) similar to those that occur in the brain.

Kohonen's algorithm creates a mapping of high-dimensional input data into output nodes arranged in a low-dimensional grid, characterizing a vector quantizer [Lippmann 1987]. Output nodes are extensively interconnected with many local connections. During training, continuous-valued input vectors are presented either sequentially in time or in batch without specifying the desired output. This is called unsupervised learning. In addition, the weights will be organized such that topologically close nodes are sensitive to inputs that are physically similar. Output nodes will thus be ordered in a natural manner. This may be important in complex systems with many layers of processing because it can reduce lengths of inter-layer connections. After enough input vectors have been presented, weights will specify clusters or vector centers that sample the input space such that the point density function of the vector centers tend to approximate the probability density function of the input vectors [Kohonen 1984 b]. Kohonen demonstrates how SOMs can be used in a speech recognizer as a vector quantizer [Kohonen 1984 a].

2.1 SOM's algorithm

After the synaptic weights initialization, the learning procedure enters upon an episodic loop that just stops when a defined final condition is achieved. Each epoch corresponds to a learning procedure whereby every input data is presented to the network. For each data input the procedure is divided into three processes. In the competitive process, the output node with the shortest distance to the input data, called Best Matching Unit (BMU), is selected to learn the input. Euclidean distance is normally used as a distance metric [Kohonen 2000]. In the cooperative process, the nodes that support the BMU's victory are also selected to learn, but in a lower level related to the nodes' help effort. The degree of cooperativeness of a node is defined by a neighborhood function that is monotonically decreasing with the nodes' distance to the BMU [Kohonen 1984 b]. At last, it's in the adaptive process that the learning takes place. The weights of each node are updated by the learning procedure shown in equation (1). At each epoch (n), the weight vector (ω) of each node (j) is changed in the direction of the input data vector (x). The learning degree is obtained by the product of the current global learning rate (η) and the neighborhood function (h) of the BMU ($i(x)$), considering the node being updated.

$$\omega_j(n+1) = \omega_j(n) + \eta(n) \times h_{j,i(x)}(n) \times (x - \omega_j(n)) \tag{1}$$

The global learning rate and the dispersion of the neighborhood function decreases exponentially in time [Gomes and Von Zuben 2003]. This policy grants two different stages on the map’s development: a rough and fast convergence with initially high learning values; a fine tune with the decrease of the learning values. The initial values for rate and dispersion and their time constants act as control parameters to the dynamics of SOM’s generation and to the quality of the final mapping.

2.2 Clustering

A class is defined as a data group with similar properties, as illustrated in Figure 1a. Different classes have non related properties. The topology preserving feature implies that correlated data are mapped into close regions in the array of neurons. Therefore, an output node will be closer to nodes related to data from the same class than to output nodes that represent data from other classes. A cluster can be identified as a group of output nodes nearly located in terms of their weight vectors produced by the learning phase.

2.3 U-Matrix

The U-Matrix is a useful tool for clustering visualization in SOMs. It represents an average picture of the distance profile between the weight vector of each neuron and the weight vector of its immediate neighbors. High values in the U-matrix indicate neighbor neurons with distant weight vectors, and low values indicate neighbor neurons with high-correlated weight vectors, so that they will be stimulated by similar input patterns [Ultsch 2003]. As seen in Figures 1b and 1c, valleys denote neurons with similar behavior, being an indication of a cluster. High areas indicate that neighbor neurons have dissimilar weight vectors, revealing transition between two distinct clusters.

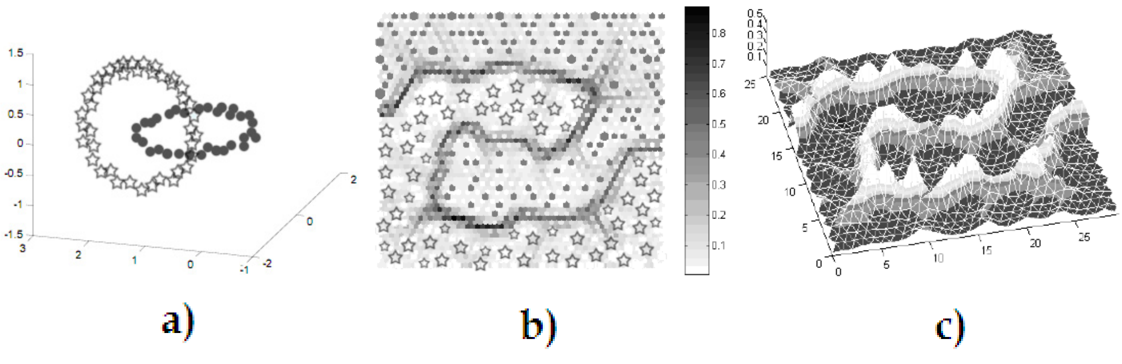

Figure 1 U-Matrix for a two-dimensional map. (a) 3D two-class data set; (b) U-Matrix visualization after clustering (gray levels indicate height and data hit marks identifies winning neurons); (c) 3D visualization of U-Matrix, with two contiguous valleys (dark areas) and peaks (light areas) characterizing the frontier between the two valleys. After Zuchini (2003).

3. Neural Network Timbral Improvisation

Here, we present a timbre design method that allows the composer to express a certain degree of subjectivity by simply choosing the number of neurons and setting the parameters adequately, according to aesthetical preferences. The user is enabled to find candidate solutions that meet certain musical requirements by using a set of waveforms (stimuli) as examples of the desired timbres (Figure 2 a). Instead of describing the sounds using numerical parameters or any other linguistic tool, we used a set of waveforms to characterize timbre. Smalley (1990) declared that the information contained in the frequency spectrum cannot be separated from the time domain, because "*spectrum is perceived through time and time is perceived as spectral motion*". Thus, by specifying the target waveforms (stimuli), the user is also specifying the spectral contents and the timbral characteristics of the tones. Grey (1975) discusses the advantages of time domain representation. We aim at sound design by means of the specification of the spectral contents using time-domain representation and manipulation.

Timbre soundspace is unknown and there is no consensus ordering or classification (Figure 2 a). Due to the self-organizing feature of SOM, it is possible to propose timbral arrangements that respect the original topology (Figure 2 b). The key feature of SOM that allows this process is that stimuli with similar characteristics trigger neurons in close regions of the one-dimensional mapping that represents the topological neighborhood in the original soundspace. In our application, self-organization gives rise to two different musically profitable phenomena. Firstly, one stimulus might trigger more than one neuron, causing the result to represent timbral variations of the original (stimulus) sounds (Figure 2 c). Secondly, more than one stimulus-sound might trigger the same neuron (zoomed-in areas in Figure 2 b). The expected result is a timbral merger of the stimuli corresponding to the neuronal-sounds that responded to these inputs. The resultant one-dimensional arrangement corresponds to a cyclic ordering of the stimulus-sounds that can be regarded as a proposal of a sort of timbral scale. The concept of timbral improvisation emerges from the possibility of following this orderly self-organizing path provided by the method in much the same way scales are used in traditional music improvisation. Moreover, the very dynamic convergence process of the neuronal sounds from the initialization to the final result can be sequentially played. This process would reveal the timbral neurological-induced transformation resulting from the path followed by each neuron during the self-organizing process.

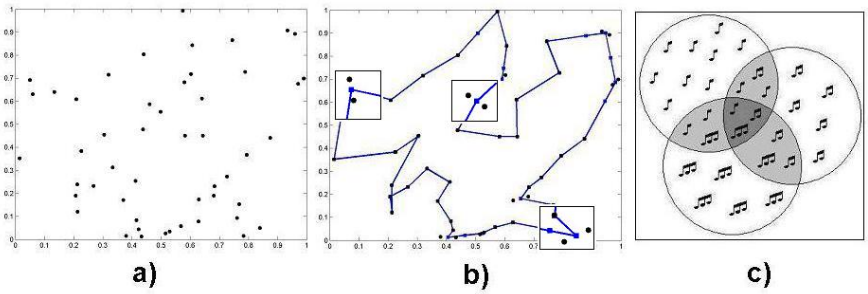

Figure 2 Depiction of the vector quantization ability of SOMs. Part (a) shows the original data topology; part (b) shows the resultant one-dimensional SOM representing the original data topologically arranged; and part (c) illustrates the common timbral features of three variants of sounds.

3.1. Representation

The input parameters of the present implementation are shown in Table 1. Each individual is codified as a vector composed of L samples of a given waveform at a sampling frequency of *SF* samples per second. The individuals are, thus, represented in time domain, as vectors in $\mathbb{R}^L$. The individuals are arranged as a one-dimensional circular SOM with the desired number of output nodes (neuronal-sounds). After training, the weight vectors represent the input data in the same vector space (with the same number of dimensions). The procedure used is similar to a TSP solution obtained via SOMs [Gomes and Von Zuben 2003].

Table 1. Input parameters that can be controlled by the user

Parameter	Description
L	Number of samples per individual
SF	Sampling rate
G	Number of Stimuli
N	Number of Neurons (weights)
Lr	Initial Learning Rate
Radius	Initial Radius
Tlr	Exponential decrease learning rate time constant
Tr	Exponential radius decrease rate time constant
Epoch	Number of epochs

4. Experimental Results

The objective of the experiments described is to analyze input/output correlation to verify the vector quantization capability of our proposal. We simulated the method using a set of seven different natural sounds as stimuli and an output layer of fourteen neurons (output data). The neurons outnumber the stimuli so as to force some neurons to never be BMU and simply be pushed around by their neighbors during the self-organizing process. In other words, this procedure tends to maximize the timbral merger and timbre variation cases. Table 2 shows the chosen stimulus-sounds. They were selected to try and maximize dissimilarity in order to explicit individual features of the stimulus-sounds blended in different resultant neuronal-sounds. The weights were initialized with a random Gaussian variable (white noise) and were expected to converge to sounds correlated to the stimulus-sound set. The parameters used are shown in Table 3. The stopping criterion was achieved either by reaching the maximum number of epochs or by a learning rate lower than 0.01.

Table 2 Instruments adopted as stimulus-sounds

Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7
Alto sax	Electric bass	Guitar	Piano chord	Harmonica	Voice	Whistle

We aim at showing that each neuronal-sound is correlated with at least one stimulus-sound, representing a variant. In cases when neuronal-sounds are correlated to more than one stimulus-sound, we wish to show that it represents a timbral merger of the related stimuli by means of blending their features. The classification procedure consists of evaluating the correlation of the neuronal-sounds with the stimuli by means of the estimation of the distance between stimulus-neuron pairs in the timbral soundspace. Short distances are to be interpreted as showing that the particular stimulus-neuron pairs are highly correlated. Caetano et al. (2005 a) define a spectral metric that was applied to each input/output (stimulus/neuron) pair. A similarity table was constructed using this metric (Table 4), associating dissimilarity to spectral distance.

Then, the same procedure was done using a subjective similarity criterion. Five musically untrained subjects were presented to all the stimulus/neuronal-sound pairs and were asked to define a distance value between 1 and 5, 5 being maximum similarity. For each pair, the stimulus-sound was played first, followed by one of the neuronal-sounds. Table 5 shows the average and standard deviation values for this evaluation.

Finally, a two-dimensional SOM was used to generate a topological representation of the input/output resultant soundspace [Damiani et al. 1995; Cosi et al. 1994 a,b; De Poli and Tonella 1993; De Poli and Prandoni 1997; Loureiro et al. 2004; Feiten and Gunzel 1994]. A U-Matrix topological map similar to Figure 1b was generated. All data was presented to the map during learning. Sounds with similar properties were expected to be mapped into the same region, while others should be mapped in a different class valley. This way, neuronal-sounds mapped near stimulus-sounds may represent a similarity relation (Figure 4).

Table 3 Parameters for the experiments

L	*SF*	*G*	*N*	*Lr*	*epoch*	*Radius*	*Tlr*	*Tr*
4096	44100	7	14	0.2	451	3	150	150

The convergence dynamics of one specific neuron is shown in Figure 3. Snapshots of one stimulus-sound and the corresponding BMU (neuronal-sound) are plotted at different stages of the self-organizing process. The goal is to highlight the rapid convergence early in the process (exploration of soundspace), followed by a fine tuning stage due to the decreasing of the neighborhood along the learning dynamics (exploitation of promising areas). Notice how the noisy, highly uncorrelated neuron learns to respond to the stimulus, representing its timbral features.

The results of the mathematical and subjective distance evaluations are shown, respectively, in Tables 4 and 5. Values in bold represent a high correlation between the respective input-output pair, i.e. stimulus-neuron. In Table 4, the distances must be interpreted relatively to all the values in the same column, once it is not normalized. Low values mean small distance and thus high correlation. Table 5 shows the average value and standard deviation estimated by the subjects. Here, the subjects estimated the similarity between the input-output pairs. Therefore, high values imply high correlation.

Finally, Figure 4 shows the resultant mapping using a two-dimensional SOM of all the stimulus-sounds and neuronal-sounds together. The gray scale represents topological distance, white being the shortest. The input data (stimulus-sounds) are plotted as black spots and output data (neurons) are plotted as white spots. All data is labeled. Here we expected the same patterns that emerged from Tables 4 and 5 to reveal in this two-dimensional clustering. Highly correlated input/output sounds should be mapped in the same region. Thus, the relations made explicit in Table 6 were also expected to arise in this analysis. The neuronal-sounds that were considered very close to a given stimulus-sound should have been mapped in the same light region.

The results of the experiment discussed here, as well as other significant results, can be found in http://www.dca.fee.unicamp.br/~caetano/SBCM.html.

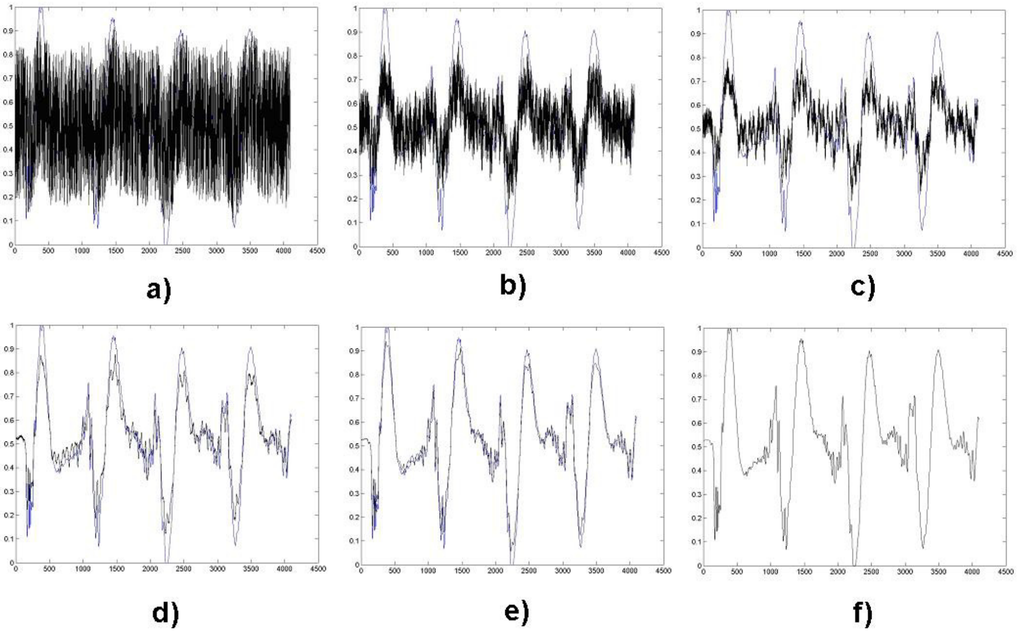

Figure 3 Depiction of the convergence dynamics after different epochs: Stimulus-sound × Neuronal-sound. Part a) 1 epoch; b) 3 epochs; c) 5 epochs; d) 10 epochs; e) 50 epochs; f) 100 epochs;

5. Discussion

Crossing the result of the subjective and mathematical similarity evaluations (Tables 4 and 5) it is possible to infer a relation between the input and output data. Table 6 shows the outputs (neurons) with maximum similarity to each of the inputs extracted from the values in bold in Tables 4 and 5. Interestingly enough, both the spectral distance and subjective similarity values vary, revealing different degrees of correlation between the pairs. Notice that although the subjective estimation presents maximum similarity evaluations, no spectral distance resulted in zero (refer to [Caetano et al. 2005 a] for the definition of the spectral distance metric).

Table 4 Spectral distance: Input x Output evaluation

	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7
Output 1	34.7606	140.7773	**37.6053**	117.6477	34.4166	33.3402	44.7520
Output 2	66.3417	144.5577	**7.1025**	123.9408	53.3262	53.3984	60.8768
Output 3	65.7337	141.9019	**8.3761**	123.6183	52.4717	52.5984	60.1636
Output 4	127.6824	**18.9327**	129.2990	165.1742	119.4274	120.5729	123.7715
Output 5	129.9995	**16.4692**	131.9691	166.9731	121.8598	123.0056	126.1364
Output 6	127.6733	**19.0678**	129.7479	163.7705	119.3749	120.5557	123.6851
Output 7	107.3686	167.9962	113.6441	**15.6206**	97.3370	99.4349	99.5542
Output 8	**6.7819**	143.9433	68.7541	119.4962	38.6160	36.9926	47.6565
Output 9	70.1802	149.4947	80.6534	**61.0166**	55.0873	58.3408	53.5189
Output 10	50.4559	141.3996	65.7026	113.7396	28.8807	33.9405	**3.4725**
Output 11	45.2253	139.1221	61.4320	112.2853	15.1187	25.3820	**17.0501**
Output 12	43.9312	138.1801	60.1144	112.5711	**1.4608**	22.8388	31.5043
Output 13	41.3641	138.9457	59.6131	114.1093	21.3422	**2.5079**	35.2750
Output 14	**5.1640**	144.0565	69.1358	119.7276	39.2329	37.2416	48.0991

Table 5 Result of subjective similarity Input x Output evaluation

	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7
Output 1	3.0±1.58	1.6±0.89	**3.8±0.45**	2.2±0.84	1.8±0.84	2.4±1.34	1.6±0.89
Output 2	2.0±1.41	1.4±0.89	**4.8±0.45**	2.0±0.71	1.6±0.89	1.6±0.89	1.4±0.89
Output 3	2.0±1.41	1.6±0.89	**4.6±0.55**	2.0±0.71	1.2±0.45	1.8±0.84	1.4±0.89
Output 4	1.2±0.45	**4.2±1.30**	1.8±0.84	1.4±0.55	1.0±0.00	1.0±0.00	1.2±0.44
Output 5	1.2±0.45	**4.8±0.45**	1.8±0.84	1.2±0.45	11.0±0.00	1.6±0.89	1.2±0.44
Output 6	1.2±0.45	**4.8±0.45**	1.6±0.89	1.2±0.45	1.0±0.00	1.2±0.45	1.2±0.44
Output 7	1.2±0.45	1.0±0.00	1.4±0.55	**5.0±0.00**	1.6±0.89	1.0±0.00	1.2±0.44
Output 8	**4.8±0.44**	1.2±0.45	1.2±0.45	1.0±0.00	1.8±0.84	1.4±0.89	2.0±0.71
Output 9	1.2±0.45	1.0±0.00	1.6±0.89	**4.6±0.55**	1.4±0.55	1.0±0.00	1.4±0.89
Output 10	2.0±1.00	1.2±0.45	1.0±0.00	1.0±0.00	1.8±0.84	1.2±0.45	**5.0±0.00**
Output 11	1.4±0.55	1.2±0.45	1.2±0.45	1.0±0.00	3.0±2.00	1.2±0.45	4.0±0.71
Output 12	1.6±0.89	1.0±0.00	1.8±0.84	1.0±0.00	**5.0±0.00**	1.0±0.00	1.6±0.55
Output 13	1.4±0.89	1.4±0.55	1.4±0.55	1.0±0.00	1.2±0.45	**5.0±0.00**	1.2±0.45
Output 14	**4.4±0.89**	1.4±0.89	1.6±0.89	1.2±0.45	1.8±1.30	1.8±1.30	1.6±0.89

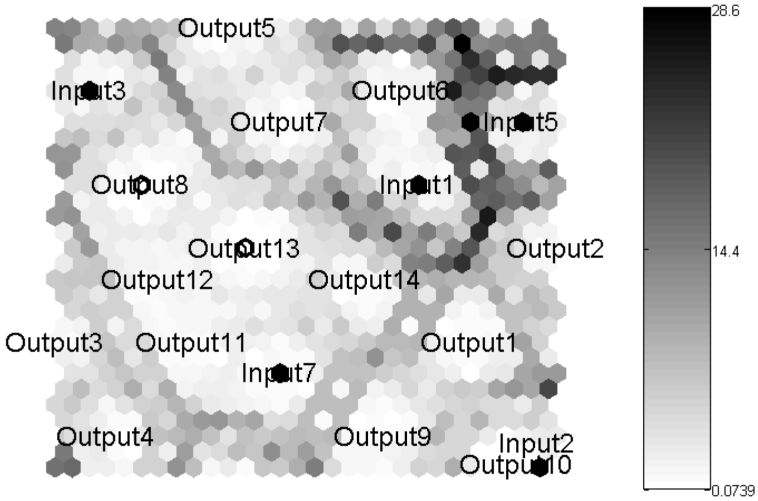

Figure 4 U-Matrix visualization. Light areas represent clusters associated with similar sounds. Dark areas represent cluster borders. The distance scale is shown beside the map.

One can conclude that, despite the subjective evaluation estimations imply in some cases that some outputs are exactly the same as the inputs, the distance metric reveals that it actually represents a variant. Intermediate distance (similarity) values can also be interpreted as resulting from the blending of timbral features present in one or more stimuli. The spectral distance metric revealed very adequate, matching the subjective estimation at every instance.

Table 6 Input-Output relation inferred from subject data

Input	1	2	3	4	5	6	7
Related Output	8,14	4,5,6	1,2,3	1,7,9	12	13	10,11

On the other hand, a direct comparison of the U-Matrix visualization (see Figure 4) does not have the same effect. The similarity found by the above mentioned analysis, explicited by Table 6, ceases to exist here. Outputs that were mapped together by the U-Matrix are shown in the same valley (white regions). Dark regions represent cluster borders. The map confirms some relations but it fails to relate a representative number of others. In fact it does suggest new ones. This may be due to two distinct factors. Either different classes were found in Figure 4, implying that the inputs and outputs can be clustered in different ways than the distance evaluations results show, according to different criteria; or it is simply impregnated with topological violations, probably due to the great effort of maintaining the topology of such high-dimensional vector space ($\mathbb{R}^{4096}$!) represented by the stimuli, projecting it into a two-dimensional space. Differently from the results reported in the literature using two-dimensional SOM to classify sounds according to timbre [De Poli and Tonella 1993], we found the method inappropriate for the purpose of timbral classification in high-dimensional timbral soundspaces.

Here we should stress the important fact that this is hardly the first proposal for a measure of timbral topological relations or classification. Many other techniques are available, including multidimensional scaling [Grey 1975; Grey and Moorer 1977] and subjective analyses [Caetano et al. 2005 b], among others.

The experiments show that SOM is capable of producing sounds that have the desired spectral content with flexibility and robustness. The method makes possible to avoid the burden of trying to describe the desired result in terms of timbral attributes or to exhaustively search the entire soundspace for the desired result interactively, as is the case for Interactive Genetic Algorithms [Biles 1994].

6. Conclusion

A novel method of timbre design was presented, which utilizes SOM, a connectionist clustering technique, in the task of obtaining sounds that are topologically arranged using self-organization. These sounds possess a set of desired timbral characteristics that are inherent to musical sounds and that cannot be precisely described due to the intrinsic multidimensional nature of timbre and the subjective characteristics involved. There is no consensus on how many or what these dimensions are, let alone their subjective relation to the spectral contents of the tone.

The input-output similarity was tentatively measured to base the resultant arranged timbral improvisation cycle respecting topology of the original timbral soundspace. A spectral measure of distance was crossed with a subjective similarity analysis to classify the outputs as being most closely related to one input, representing a variant. Posteriorly, this result was compared with a two-dimensional SOM clustering technique well documented in the literature for timbre classification [Cosi et al. 1994 a,b]. Crossing the results of both evaluations, we found that SOM fails to properly represent the topological relations of the sounds, incurring in topological violations probably due to the high dimensionality of the vector space the sounds were represented in.

The method presented is adjustable according to the input parameters and leads to interesting variations and mixtures of the stimulus-sounds (inputs). The characteristics of maintenance of topology and unsupervised learning provided by SOMs are essential in the results.

Many extensions can be envisaged and tested. It can be used to compose soundscapes, as a timbre design/improvisation tool or in live electroacoustic music where a neurological timbre is generated, which evolves in real time along with other music materials. Future trends might include using the technique in AI-based musical systems and adapting the method for dynamic environments, i.e. using time-varying stimulus-sounds.

7. Acknowledgments

The authors wish to thank FAPESP (process no. 03/11122-8 and 04/11742-9) and CNPq (process no. 306672/2004-9 and 308765/2003-6) for their financial support.

8. References

Biles, J. A. (1994) GenJam: A Genetic Algorithm for Generating Jazz Solos, Proceedings of the 1994 International Computer Music Conference, (ICMC'94), pp. 131-137.

Blackwell, T. and Young, M. (2004) Swarm Granulator. In G. R. Raidl et al. (Eds): EvoWorkshops, Lecture Notes in Computer Science 3005, pp 399-408.

Burraston, D., Edmonds, E. A., Livingstone, D. and Miranda, E. (2004) Cellular Automata in MIDI based Computer Music. Proceedings of the International Computer Music Conference, pp. 71-78.

Caetano, M., Manzolli, J. and Von Zuben, F. J. (2005 a) Application of an Artificial Immune System in a Compositional Timbre Design Technique. In C. Jacob et al. (Eds): ICARIS 2005, Lecture Notes in Computer Science 3627, pp 389-403.

Caetano, M., Manzolli, J. and Von Zuben, F. J. (2005 b) Interactive Control of Evolution Applied to Sound Synthesis. in Markov, Z., Russel, I. (eds.) Proceedings of the 18th International Florida Artificial Intelligence Research Society (FLAIRS), Clearwater Beach, Florida, EUA, pp. 51-56.

Chen, C. J. and Miikkulainen, R. (2001) Creating Melodies with Evolving Recurrent Neural Networks, Proceedings of the International Joint Conference on Neural Networks (IJCNN-01), 2241-2246.

Cosi, P., De Poli, G. and Lauzzana, G. (1994a) Auditory modelling and self-organizing neural networks for timbre classification. Journal of New Music Research 23, 71-98.

Cosi, P., De Poli, G. and Lauzzana, G. (1994b) Timbre classification by NN and auditory modeling. In Marinaro, M. and Morasso, P. G., editors, Proc. ICANN'94, International.

Damiani, F., Pérez, M. J., Fornari, J. E. (1995) Reconhecimento de timbres musicais através da rede neural auto-organizável de Kohonen . XV Congresso da Sociedade Brasileira de Computação, II Simpósio Brasileiro de Computação e Música. Pag. 107-113.

De Poli, G. and Prandoni, P. (1997) Sonological models for timbre characterization. Journal of New Music Research 26, 170-197.

De Poli, G. and Tonella, P. (1993) Self-organizing neural networks and Grey's timbre space. In Proceedings of the 1993 International Computer Music Conference, pp. 441-444.

Feiten, B. and Gunzel, S. (1994) Automatic indexing of a sound database using self-organizing neural nets. Computer Music Journal 18(3), 53-65.

Gomes, L. C. T. and Von Zuben, F. J. (2003) Multiple criteria optimization based on unsupervised learning and fuzzy inference applied to the vehicle routing problem. Journal of Intelligent & Fuzzy Systems, IOS Press, vol. 13, no. 2-4, pp. 143-154.

Grey, J. M. (1975) An Exploration of Musical Timbre. Doctoral dissertation, Stanford Univ.

Grey, J. M. and Moorer, J. A. (1977) Perceptual Evaluations of Synthesized Musical Instrument Tones. Journ. Ac. Soc. Am., 62, 2, pp 454-462.

Helmholtz, H. (1885) On the Sensations of Tone. London, Longman.

Horowits, D. (1994) Generating Rhythms with Genetic Algorithms. Proceedings of the 1994 International Computer Music Conference (ICMC'94), pp. 142-143.

Kandel, E. R. and Schwartz, J. H. (1985) Principles of Neural Science, Elsevier.

Kohonen, T. (1984 a) Phonotopic Maps – Insightful Representation of Phonological Features for Speech Representation. Proceedings IEEE 7th Inter. Conf. On Pattern Recognition, Montreal, Canada.

Kohonen, T. (1984 b) Self-Organization and Associative Memory. Springer-Verlag.

Kohonen, T. (2000) Self-Organizing Maps. Springer.

Lippmann, R. P. (1987) An Intoduction to Computing with Neural Nets. IEEE ASSP Magazine.

Loureiro, M. A., de Paula, H. B. and Yehia, H. C. (2004). Timbre classification of a single musical instrument. Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004), Barcelona, Spain.

Moller, A. R. (1983) Auditory Physiology. Academic Press.

Risset, J. C. and Wessel, D. L. (1982) Exploration of timbre by analysis and synthesis. In D. Deutsch (ed.) *The Psychology of Music* (pp. 26-58). New York: Academic.

Santos, A., Arcay, B., Dorado, J., Romero, J. and Rodríguez, J. (2000) Evolutionary Computation Systems for Musical Composition. International Conference on Acoustic and Music: Theory and Applications (AMTA 2000). vol 1. pp 97-102.

Smalley, D. (1990) Spectro-morphology and Structuring Processes. In the Language of Electroacoustic Music, ed. Emmerson, pg. 61-93.

Ultsch, A. (2003) U*-Matrix: a tool to visualize clusters in high dimensional data. Technical report, Departement of Mathematics and Computer Science, Philipps-University Marburg.

Zuchini, M.H., (2003) Aplicações de mapas auto-organizáveis em mineração de dados e recuperação de informação. Dissertação Mestrado, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas.