Self-organizing Neural Network and Grey's Timbre Space

Giovanni De Poli, Paolo Tonella
CSC- DEI University of Padova, Via Gradenigo 6a, 53131 Padova, Italy
depoli@dei.unipd.it

Abstract

The target of this research is an exploration of timbre multidimensionality, developed with the aid of self-organizing neural networks. Such networks show the interesting capability of extracting the main dimensions in a multidimensional input, and implement a learning algorithm derived from models of clustering in the human brain. The starting point is Grey's experiment, where a three-dimensional timbre space was determined by muldimensional scaling of subjective similarity judgement. Using Kohonen feature maps a timbre space was produced inside a three-dimensional neural network, allowing timbre mapping and clusterization. The input data derive directly from the Grey's music signals, after a preprocessing phase. The clusterization obtained from subjective judgement and by neural networks are compared. The results encourage the use of neural tools for timbre analysis, and suggest future developments in the fields of signal pre-processing and neural net fine tuning.

1 Introduction

Timbre is a sound feature that can be analyzed with difficulty in physical and mathematical terms because it depends on a great number of parameters. The aim of our work is to simplify timbre multidimensionality following the lines of Grey's experiment, and to obtain similar results in terms of clusterization and timbre space. Grey[1975,1977] determined a three-dimensional (3D) space in which the sounds are mapped, by applying multidimensional scaling to subjective similarity judgments between the timbres of 16 traditional instruments. The interpretation of the coordinates explained the main factors affecting timbre discrimination. Experiments for timbre recognition through neural networks have been conducted recently[Feiten][Mourjopoulos]; our purpose is to use neural networks for the exploration of the topology of the timbre space. The tools we use are Kohonen self-organizing neural networks (KNN): they show a capacity of classification even outside the training set, because they are able to identify the dimensions of the input patterns characterized by the greatest variance. Another reason for their use comes from neurophysiology: the principles of self-organization used by Kohonen were derived from a model of the cerebral cortex behavior; it is therefore interesting to compare our results with the ones obtained by Grey from subjective judgments. To best approciate the problem, a 3D architecture of the KNN is proposed.

2 Grey's three-dimensional timbre space

The experiment carried out by Grey at Stanford University in 1975 makes use of 16 instrumental tones; these sounds were first equalized for subjective duration, pitch and intensity. The principal part of the experiment consists on the multidimensional scaling of the timbres; the subjective judgments of similarity expressed by 22 listeners, previously trained for the recognition of the instruments, were averaged in a similarity matrix. This similarity matrix was then processed using an MDS (MultiDimensional Scaling) algorithm; the result was the distribution of the various sounds in an n-dimensional space; moreover the same similarity matrix was processed using a hierarchical clustering algorithm based on the method of diameter, with the purpose to obtain a timbre clusterization. The very interesting result was that the clusters obtained applying the diameter method collected in the same group timbres located at a low distance in the three-dimensional timbre space produced by the MDS algorithm; this assert was not yet true for the bi-dimensional timbre space produced by the MDS algorithm. This fact leads to the conclusion that the three-dimensional space reveals to be the fittest for the purpose of timbre classification. A furthermore consideration comes to confirm the validity of this conclusion, as it is possible to give a meaning to the three spatial axes, while this is not possible for the two-dimensional space. The physical interpretation

that we can give to the three dimensions is the following: the first dimension can be interpreted as a spectral distribution of the energy; the second dimension can be interpreted as the presence of synchronicity in the attack stage through the harmonics; the third dimension is connected with the presence of high frequency inharmonic noise with low amplitude, during the attack segment.

3 3D Kohonen neural net

Kohonen neural networks (KNN) are inspired by the process that seems to be responsible for the map-organization of the cerebral cortex. From the biological process Kohonen [1984,1990] derived a self-organization algorithm for artificial neural networks, that gives a computational way to create a map of the input patterns that corresponds with the maps created by the human brain in the cerebral cortex. When input data are partially corrupted by noise KNN work well, because they show an interesting capability of feature extraction; KNN are able to render maximum the amount of information stored inside them, in presence of noise, because the network organizes itself satisfying two different contrasting requirements: to make maximum the variance of the outputs of all the neurons, with the purpose of recognizing the features mainly distinctive for the inputs, and to introduce a certain degree of redundancy, with the purpose of obtaining correct answers even in presence of noise. We can observe the capability of these networks of recognizing the dimensions of the inputs with the maximum of variance considering the distribution of the net weights after the self-organization process; the weights of the network tend to line up the input dimensions with greatest variance; if, e.g., the input patterns have a uniform distribution in a rectangle, the weights of the network line up the greatest dimension of the rectangle. KNN require an heuristical stage of fine tuning of some of their parameters, because no optimal values are known for them. Therefore it is necessary to make some experiments to determine the best values for these parameters, according to the features of the input signal.

Grey's experiment suggests the basic idea for the design of a neural network capable of classifying instrumental timbres. KNN have been used by Kohonen himself for the recognition of the phonemes from continuous speech; applications to timbre recognition have also been conducted, using data from the spectral analisys of the signals as input. The target of this work is not limited to recognizing the single timbres, but longs to revealing the metric relations existing between timbres and to comparing this metric with the subjective one. For a better comparison the original data of Grey's experiment were used; more precisely the line segment approximations of amplitudes and frequencies of the harmonics were considered for the 16 test sounds, after the equalization for subjective duration, pitch and intensity. Twelve tones (bassoon (BN), normal cello (S2), E flat clarinet (C1), flute (FL), french horn (FH), english horn (EH), muted cello (S3), oboe (O2), cello sul pont (S1), soprano sax (X3), trombone (TM), trumpet (TP)) are used in the training step; the others are used in the test step.

The ineffectiveness of Grey's two-dimensional solution and the incongruousnesses to which it leads induce to leave the idea of using a classical two-dimensional self-organizing neural network; it would generate a plane distribution of the timbres that hardly allows a good clusterization and classification. As the optimal dimensionality for timbre representation is 3D, that is a distribution in a three-dimensional space of the sounds, we were lead to develop a self-organizing neural network with a three-dimensional distribution of the neurons. This network is a generalization of Kohonen two-dimensional neural nets; in this network, in response to input stimulations, zones are activated with a distribution that corresponds to the distribution of input patterns, and that optimizes the representation of the metric relations existing between timbres, allowing a hierarchic clustering in agreement with the distances between points or groups of points in the three-dimensional space. In this way a corrispondence is estabilished between physical features of the input signals and neural map produced by the network; the trained net has the capacity of extracting the fundamental features of a timbre, producing a good classification even in presence of new input patterns, fed to the network in the test phase. Another reason for the use of a three-dimensional neural network comes from neurophysiology, because the cerebral surface, in human brain, ripples giving raise to a fractal surface; it was demonstrated that the fractal dimension of the cerebral surface is between 2 and 3, so that a three-dimensional artificial neural network leads to greater analogies with biological neural networks, and allows a more satisfactory model for the human mechanism of clustering that takes place in the cerebral cortex. A 3D network gives raise to a 3D metric which better corresponds to the subjective metric of timbres. A mathematical analisys of the dynamics of the KNN is extremely difficult; their properties were discovered through simulation experiments and practical applications. For this reason some preliminary experiments have been conducted to verify the extension of these properties from 2D-KNN to 3D-KNN; in particular the weight vectors tend to approximate the density function of the

input vectors in an orderly fashion. Our architecture consists of an arrangement of neurons and links along the three Cartesian axes where both Euclidean distances or length of the path between neurons were adopted.

Sometimes KNN show an undesired edge effect due to their finite dimensions, especially when they are small. To avoid this effect a toroidal structure can be adopted for the network. The edge effect pushes the bubbles of activity to the edges of the networks, because here the interference with other regions of activity is lower; in this way only the neurons on the frontier of the network are used to discriminate inputs, and their number can be considerably lower than the total number of neurons of the network. In a three-dimensional network the problem may become crucial because there is a great difference between the number of neurons on the surface of the cube containing the net, and the global number of neurons, including the neurons inside the cube. To solve this problem, in alternative to the classical metric, we adopted a metric that renders adjacent opposed faces of the net. The use of this metric introduces some differences between the neural timbre space and Grey's one, because Grey's space has not a toroidal structure, while in a network of this kind timbres located on opposed faces of the cube are adjacent. It should be pointed out that the purpose of the experiment is not to exactly reproduce Grey's timbre space, but to study the reduction of dimensionality produced by the KNN, and to obtain a metric between timbres corresponding to the subjective one.

4 Clusterization

The way we used Kohonen networks is quite particular because only few examples were available respect to the number of neurons in the network (a ratio of 1/50 is typical); this lack of examples causes a great sensitivity of the network final state to the initial random values of the weights. It happens that the neurons excited by the inputs, and the neurons in the close neighbourhood adapt themselves to the input patterns, while the other neurons remain fixed close to their initial random value. It is possible to reduce this sensitivity to the initial conditions considering the mean of the behaviours of the networks; we have studied the convergence properties of the average finding out the presence of a final mean configuration with low values of variance, and, subsequently, of the relative error (3% is a typical value). The average configuration is little sensitive to initial values of the weights because it is the configuration that results starting from many different points; the effect of the initial random weights is canceled by the average.

KNN are fed by numeric inputs; samples of the sound signal can be used, so that all of the processing is made by the neural network, or pre-processed data can be derived, so that the network is applied at the most critical stage, that is the classification stage. If pre-processing is chosen, it is important to accurately select it because it must not only allow the recognition, but also the construction of a topological structure for the timbres.

Since no input factor defines a particular orientation of the output 3D map, the latter is not invariant respect to rotation, and its orientation depends on the initial random values of the synapses; for this reason different maps with the same topological relations are obtained through successive trainings. To extract these relations from the final configuration of the map we applied the diameter algorithm, in analogy with Grey's experiment; in fact, this algorithm works well with data from perception. In our experiments the algorithm was little sensitive to the dimension of the network, and to the other parameters affecting the network, when the average matrix of the distances is considered.

From the training step we obtain a timbre space in which the timbres are located in the centers of the activity bubbles associated with them in the network. The distances between these centers of activity are processed with the diameter clusterization algorithm to obtain timbre clusters that can directly be compared with those of Grey.

The first group of experiments makes direct use of Grey's data set; for every timbre Grey's data give a line-segment representation of the variation in time of the amplitude of every harmonic, and of the variation in time of the frequency which the harmonic is located at; from this data set it is simple to generate the curves describing the amplitudes and the frequencies of the various harmonics, and to sample these functions, generating a file of samples which represents an heterodine analysis of the signal. This analysis contains all the information necessary for timbre reconstruction, and therefore it can be used to feed a self-organizing neural network. We report the clusterization obtained using for the input a file containing, for the first 20 harmonics, 10 samples of amplitude and 5 samples of frequency, and training a neural network with dimension 8*8*8 = 512:

{(BN FH) [TP (FL S2)] [S1 S3]}
{[(C1 EH TM) O2] X3}

where the parethesis indicate different levels in clusterization process. The analogies with Grey's results

{[(BN TP) FH] [(S2 S3) (FL S1)]}
[O2 TM]

{C1 (EH X3)}
are encouraging; the few differences deals more with different times at which grouping with other timbres takes place, then with substantial differences.

The second group of experiments applies some data reduction tecniques to Grey's data, that were inspired by Grey's observations about the physical factors affecting timbre perception. Our results show many analogies with Grey's clusterization; the main difference lies in the characterization of the clarinet, because it is grouped with timbres that are not subjectively close, producing a distorsion in the clusterization. As an example we present the results of using, for the first 9 harmonics, the following 5 parameters: time at which the amplitude reaches its maximum, maximum value of the amplitude, energy of the harmonic, maximum of the frequency variation of the harmonic and mean value of the frequency variation. The first two parameters describe synchronism in the attack stage between the harmonics; the third gives the energy distribution through the spectrum; the last two parameters concern with frequency micro-variation in the attack and sustain stage. The network dimension is 6*6*6 = 216. The resulting clusterization is:
{(BN FL) [(FH S2) S1]}
{[(O2 TM)] [(EH X3) TP] (C1 S3)}

Another data set were obtained applying the data reduction tecniques suggested by Charbonneau[1981]; he studied the effect of three kinds of data reduction: reduction of amplitude informations, reduction of frequency informations and reduction of temporal informations. Results from these kinds of data reduction show some differences with Grey's; it is probably that this preprocessing of the signals allows a good reconstruction of the timbres, but doesn't preserve metric information.

5 Conclusion

KNN seem to be an interesting tool for the classification of a data set belonging to a space with great dimensionality, where classical tools for the extraction of the dimensions with high variance fail; their capabilities of extracting the main dimensions from the input patterns in presence of noise lead to the reduction of the complexity, and, in consequence, to the construction of a meaningful map. The comparison between the map produced by the neural network and the map obtained through psycoacoustic sessions suggests that the model underlying the artificial networks principles of self-organization resembles, in a certain way, the features of biological organization of neurons to the purpose of reducing the multidimensionality of the input patterns and of creating a simplified cerebral map of them As regards the data-reduction tecniques, deeper studies are being conducted at Padova University; the best results have been obtained using pre-processing based on Grey's observations, while Charbonneau's preprocessing gave worse final configurations. An immediate and natural development of this work consists in the completion of the process of timbre recognition by substituting, at the initial stage, the etherodine analysis with a simulator of human ear; in this way the various operations made on input signals by biological organs and neurons is totally reproduced by an artificial system; experiments in this direction have already been conducted giving encouraging results, but the results are partial and need thorough examination. A second future development, which is also under our study, consists in the broadening of the input space from Grey's data set to a wider one; it is interesting to consider a timbre space where all instrumental sounds are represented, and to create a map of this space by the means of a hearing model and of a self-organizing neural network; the comparison between this map and the subjective perception of similarity between timbres could lead to new understandings in the field of biological signal processing and representation. It should be noticed that Grey's tone are of low sound quality, so that the usage of a higher quality sampling of the timbre space, with an adequate pre-processing will be adopted in our future developments.

References

[Charbonneau] Charbonneau, "Timbre and the perceptual effects of three types of data reduction", *Comp. Music J.*, 5(2): 10-19, 1981.

[Feiten] Feiten B., Frank R., Ungvary T., "Organizations of sounds with neural nets", *Proc. ICMC 91*, p. 441-444, 1991.

[Grey75] Grey J.M., *An exploration of musical timbre*, Rep. STAN-M-2, Stanford University, 1975.

[Grey77] Grey J.M., "Multidimensional perceptual scaling of musical timbres", *J. Acoust. Soc. Am.*, 61(5): 1270-1277, 1977.

[Kohonen84] Kohonen T., *Self-organization and associative memory.* Springer V., Berlin, 1984.

[Kohonen90] Kohonen T., "The Self-Organizing Map", *Proc. of the IEEE*, 78(9): 1464-1480, 1990.

[Mourjopoulos] Mourjopoulos J., Tsoukala D. ", "Neural network mapping to subjective spectra of music sounds". *J. Audio Eng. Soc.*, 40(4): 253-259, 1992