Self-organizing Neural Nets and the Perceptual Origin of the Circle of Fifths

Nicola Cufaro Petroni[*] and **Matteo Tricarico**[†]

ABSTRACT: *One of the more surprising results of the use of artificial neural networks in the analysis of acoustic perceptions is the regularity of dispositions, on a Kohonen self-organizing map, of chord images: in fact these images clearly show the bent to spontaneously dispose themselves along closed, circular patterns reminiscent of the circle of fifths. In this paper we firstly reproduce these results and then test them against the hypothesis that the regularity of patterns could be the effect of the preprocessing and of the circular transpositions of the original signal images, rather than of their informative content referred to the tonal relations among the chords.*

1. INTRODUCTION

In recent years we have witnessed a consistent expansion of the researches devoted to the use of artificial neural networks in the study of acoustic perceptions (see for example Ref. [1], [2], [4], [5], [6], [10], [11], [12]) and a particular interest has been aroused by the works [10], [11], [12], [13] devoted to an understanding of the classification of auditory images by means of self-organizing Kohonen feature maps. In fact the ability of these neural nets in elaborating the signal information in characteristic patterns seems to give very stimulating cues about the tonal structure of (western) music. In Bharucha (1987) [1], for example, a network which *a priori* represents in its structure the relations among tones, chords and keys is activated by a musical context, and the pattern of activation at the equilibrium is analyzed: the results then show that the activation of units representing chords and keys decreases with increasing harmonic distance. On the other hand, in another series of papers ([11], [12], [13]) similar results are found with no initial hypothesis either on the structure of the net or on the relations among chords and keys. Here the computing units of the net do not *a priori* represent musical elements, but that notwithstanding an amazing regularity among the locations of the chord images fed to a Kohonen feature map has been found: in fact these images clearly show the bent to spontaneously dispose themselves on the map along characteristic circular paths which are strongly reminiscent of the well-known *circle of fifths*. This may be the result of the fact that the 12-numbers templates used as chord images (see Section 2)

---

[*] Dipartimento di Fisica dell'Universtitá di Bari, via Amendola 173, 70126 Bari (Italy); E-mail CUFARO@BARI.INFN.IT

[†] Via G. Pazienza 31/A, 71016 S. Severo (Italy); E-mail SGIROT10@FG.NETTUNO.IT

already contain inscribed in their structure all the relevant information about the more fundamental relations of tonal music, namely: the similarity and dissimilarity relations among the major (or minor) triads.

Moreover the results of these simulations seem to say even more since one of the interesting implications of this approach is the fact that, albeit the chord images can be calculated by means of several different pitch models, the outcomes of the experiments are always perfectly coherent. In other words "the conceptual difference between these [pitch] models ... are great, but their use in the tone semantic model have led to consistent and often surprisingly similar results (see Ref. [13] p. 44 and 71): no matter how you calculate your chord images to supply to the Kohonen feature map - provided that you use a reasonably musically meaningful pitch model - you always find something similar to a circle of fifths. Of course this is very interesting since one possible interpretation of this fact could be that the relevant information about the relations among triads is preserved in the preprocessing of the signals based on every one of these models: the peculiarities of every model seem in fact to be of no consequence for the outcomes of the simulation, although if it is it is possible to show that the use of one of the accepted pitch models is, after all, necessary.

However a deeper analysis of the way in which the numerical images of the chords to feed into the Kohonen feature map are produced will inevitably arouse the suspicion that the regularity of the map configurations could be a consequence of the cyclical rotations of the original template used to transpose it on the notes of the chromatic scale (see [13] p. 73), rather than of the informative content of the signal itself. In this paper we will scrutinize this hypothesis in order to show that in fact these suspicions are groundless: in particular in the following sections we will first of all reproduce in a simplified form the experiment of the original papers and then we will compare their outcomes with the results of a few simulations specially designed to put in evidence the meaning of the previous ones.

2. SIMULATION

We will describe here a computer simulated experiment about the learning of the relations among chords by means of a particular neural network known as Kohonen feature map. This is not the place to describe in detail how this algorithm in fact works: the interested reader is referred to an already wide literature about this type of computing devices (see [7], [8], [14], [16]). We will just remember here that these feature maps are self-organizing neural networks able to associate a spatial structure to statistical data organized in clusters. The idea is that the centers of the clusters are spatially correlated, so that their representatives on the grid become nearer when the clusters are more similar. A Kohonen feature map usually is a set of N processing elements (neurons), disposed in a (usualy) square grid, which receive as inputs a sequence of n-dimensional vectors $\mathbf{x}$ (considered as values drawn from a given random vector $\mathbf{X}$). Each i-th processing element then calculates the Euclidean distance $|\mathbf{x} - \mathbf{w}_i|$ between the example $\mathbf{x}$ and a given initial weight vector $\mathbf{w}_i$. At this point a competition starts to see which neuron has the $\mathbf{w}_i$ closest to $\mathbf{x}$: if this is the

j-th neuron a weight modification takes place in accordance with the so called Kohonen learning law:

$$\mathbf{w}'_k = \mathbf{w}_k + \alpha(\mathbf{w}_k - \mathbf{x})$$

for all the neurons k which are neighbours of j on the grid. Both the constant α and the definition of *neighbour* are allowed to change with time. This rule in some way draws the weight vector of the winning element closer to the input vector $\mathbf{x}$, and the topological aspect enters into the game since not just the best matching weight vector is updated, but also the weight vectors in the immediate vicinity.. After several rounds of iterations are performed on a given number of $\mathbf{x}$ samples, the $\mathbf{w}_i$ reach an equilibrium value disposing themselves in the space of the weight vectors in a way reproducing the original distribution of the $\mathbf{x}$'s. In other words the weight vectors become densest where the $\mathbf{x}$'s are more common, and become least dense where the $\mathbf{x}$'s hardly appear. The overall result is that every class of (in general n-dimensional) examples will be associated with a particular (in general 2-dimensional) region on the grid of computing elements following the pattern of the proximity between the weights $\mathbf{w}_i$ and the examples $\mathbf{x}$.

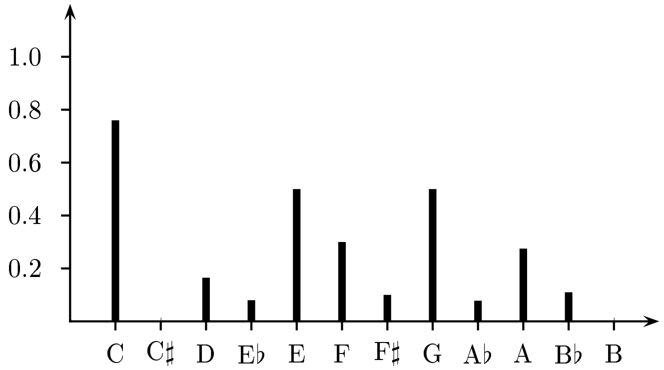

Fig. 1 Image (A) of the C major triad drawn from the Krumhansl measurements about similarities among keys (see for example [13] p. 14 and [9]).

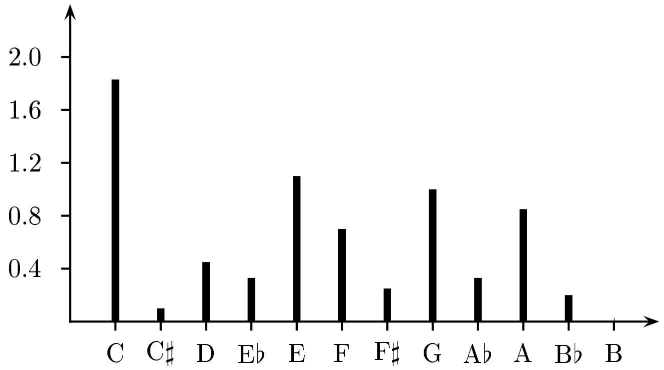

Fig. 2 Image (B) of the C major triad calculated by means of a simple pitch model as the subharmonic sum of Shepard tones (see for example [13] p. 44 and [15]).

In our simulation (performed using a program in C language written by us, on a Ms-DOS computer with CPU Intel 80386sx 25 MHz) the examples $\mathbf{x}$ are drawn froma a 12-dimensional space $\mathbf{R}^{12}$ of 12-number vectors which represent the images of tone centers obtained by means of different pitch models or, to ceck the results, by a suitable random generation. Following the discussion contained in [12] and [13] we have first of all prepared two different sets of examples just to reproduce the results of the previous simulations: the first (A) has been produced taking into account the results of psychoacoustic measurements performed by Krumhansl (see for example [9] and [13] p. 14) about similarities among keys; the second (B) comes from calculations based on a simple pitch model (see [13] p. 44 and [15]) where the resulting image is calculated as the subharmonic sum of the original representation of Shepard tones. The template of 12 numbers to be used as inputs are then obtained by transposing the template obtained for one (major, minor, diminished and so on) triad on the 12 steps of a chromatic octave. Fig. 1 and 2 show examples of these templates in the case of the C major triad: other templates have similar behaviour.

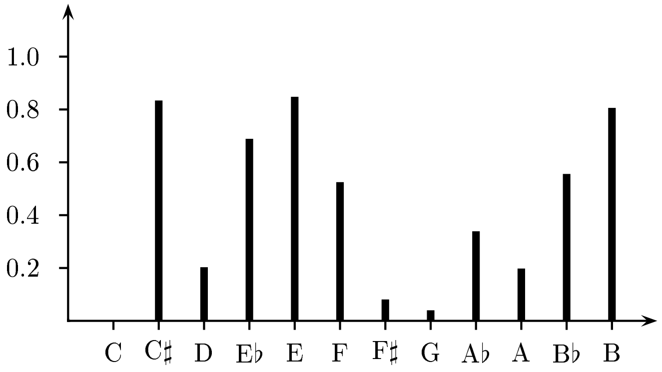

Fig. 3 Image (C) of the C major triad calculated by means of a random extraction of numbers.

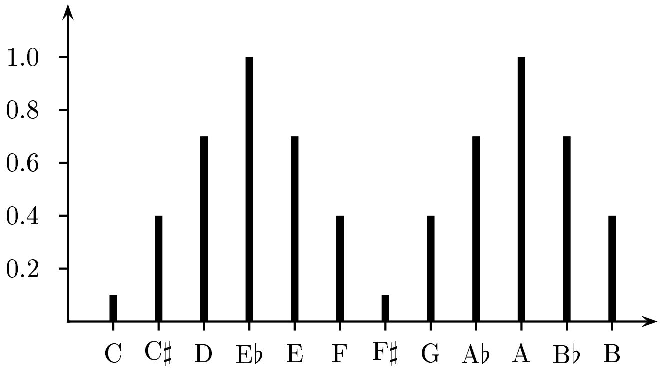

Fig. 4 Image (D) of the C major triad by means of an arbitrary, periodic choice of a template of numbers.

The set of examples is then built as a suitable mixture (taking into account the occurrence of chords in western, tonal music) of tone center images. In the original papers the set of different examples amounted to 115 different chord images (in whatever model you used to produce these images): in order to simplify the simulation we decided to reproduce the experiments with just

12 major and 12 minor triads plus 12 dominant seventh chords. This reduction in the number of proposed different examples is in fact justified by the remark that in any case the chosen chords can be credited for about the 84% of the total weight of the chords (see for example [3] and [12]). Indeed it must be pointed out that the set of examples is built by mixing up the chords in suitable proportions in order to take into account the weight of every particular chord in the usual musical production. Our simplification, which reduces the chords to the most relevant 36, allows one to correspondingly reduce the dimensionality of the grid of processing elements in the Kohonen feature map: instead of a 400 elements grid (a square of side 20) we could use a 100 elements grid (a square of side 10) because of the reduced number of objects to classify.

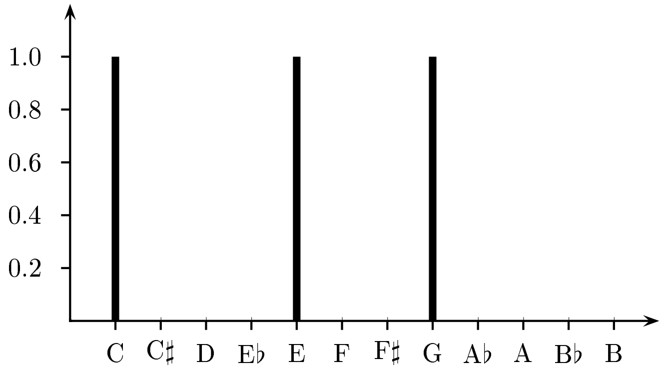

Fig. 5 Image (E) of the C major triad in the simple local representation which indicates just the pitches activated in the triad.

As will be seen in the next section the results are similar to that of previous simulations ([12], [13]); however the way of production of the sets of examples seems to be open to a criticism: we generated the tone center images by means of some model (either Krumhansl data or a pitch model) just once for every set of the 12 shifts of the template along the steps of a chromatic octave. This means that in reality the 12 templates of, for instance, the major triads are always the same template just transposed on another degree of the chromatic octave (with, of course, a folding of the exceeding elements on the opposite side of the template). Since the interesting result of these simulations is the fact that a particular *circular order* emerges among the activation regions of the triads (for example the the major triads are distributed along a closed path strongly reminiscent of the circle of fifths), a legitimate question is whether this underlying order of the chords could just be the product of the *circularity of the transpositrons* introduced by hand in our samples to represent every step of the chromatic scale. In other word it would be interesting to have a way to verify whether the order emerging out of the data is a property of the information encoded in the templates of numbers used as tone center images, or rather only a trivial consequence of our manipulation and preprocessing of the incoming signal. Moreover, since the results of the simulations seem to be largely independent from the particular model used to produce the input templates [13], a second, and related, question is whether the order of data can be achieved without any recourse to a model whatsoever, for example by feeding to the neural network a simple local representation of triads and chords.

In order to check these points, besides the examples of class (A) and (B) previously described, we have produced three other different sets of templates similar to the others in every respect except for the fact that now the first template in the series of 12 is not produced on the basis of a pitch model whatsoever. More precisely in the first series (C) we have produced an artificial template by means of 12 random numbers which are arbitrarily attributed as template for the C major triad; then the minor triad and the dominant reventh are generated from the major triad simply by reproducing the same differences among the corresponding templates in the Krumhansl data. Finally these three archetypal templates are transposed along the 12 steps of the chromatic octave. In the second series (D) we have calculated the templates in a way identical with (C), except for the fact that the first profile was not drawn at random, but was arbitrarily chosen to be higly regular and periodic, albeit dissimilar from the (A) and (B) cases. Finally the series (E) consists of images in simple local representation, with no pitch model or measurements involved, with the only indication of the pitches present in the triad or chord. Examples of these new type of templates are shown in the Fig. 3, 4 and 5.

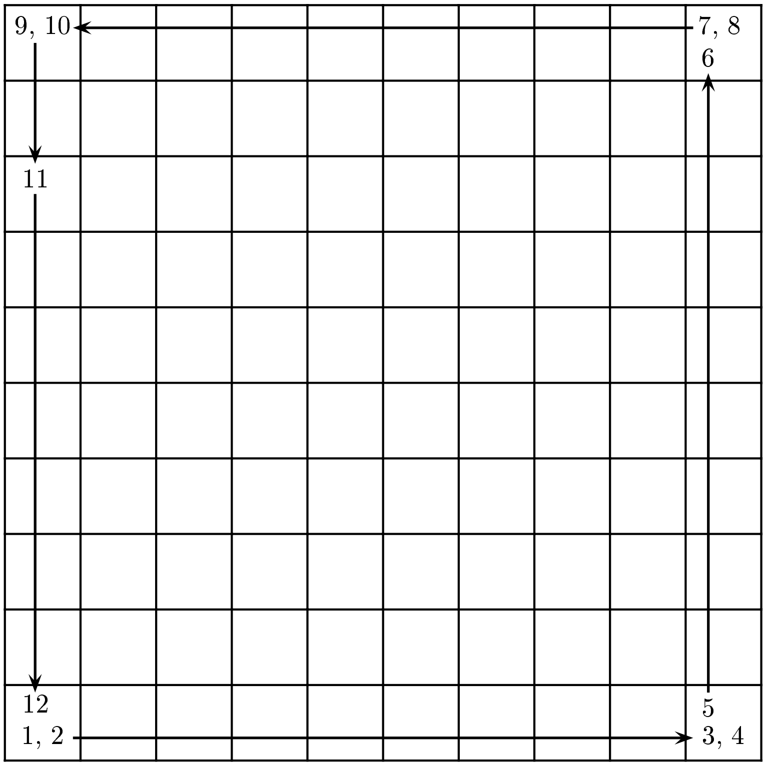

Fig. 6 Locations of the most activated neurons for major triads of the series (A) and indication of the sequence of triads in the circle of fifths.

We are now in the position to check the results of this simulation by comparing the activation patterns produced on the Kohonen feature maps by the sets of examples (A) and (B), generated either by an acceptable pitch model or by empyrical measurements, with the corresponding activation patterns produced by the sets (C), (D) and (E).

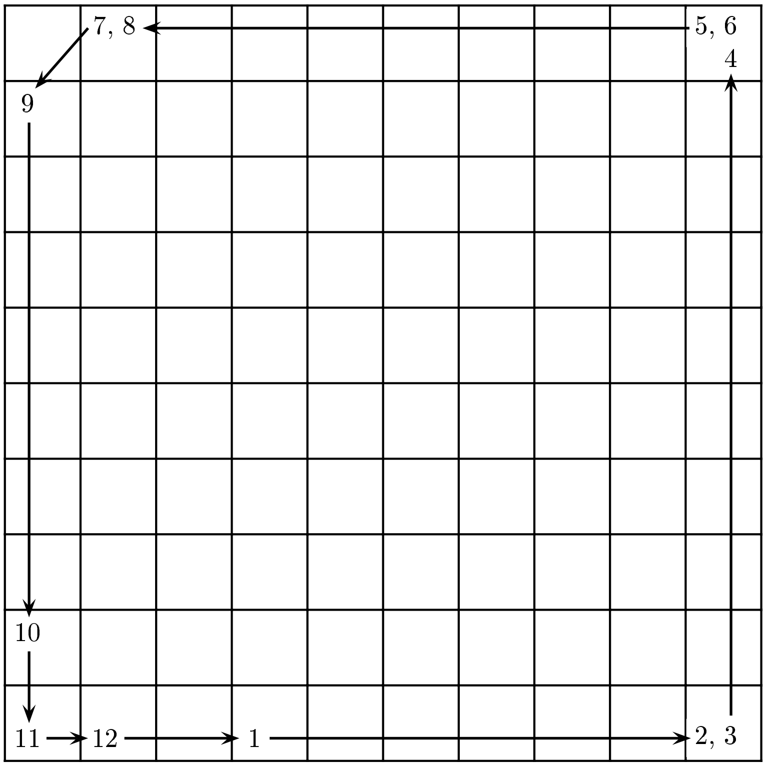

Fig. 7 Locations of the most activated neurons for major triads of the series (B) and indication of the sequence of triads in the circle of fifths.

3. RESULTS AND DISCUSSION

We performed the simulatin four times with different sets of examples (A), (B), (C), (D) and (E) on a square grid of processing elements (side 10 elements). Every set is composed of 120 replicas of the 36 chosen chords in a mixture which respects the composition used in the previous simulations (major triads about 50%, minor triads about 20% and dominat sevenths about 30%). We performed the training of the net varying a few parameters: number of iterations (from 100 to 200); learning rate α (from 0.2 to 0.3) and initial random values of the weight vectors, but we have always consistently found the same basic results, namely: *when the examples are generated either by means of a pitch model or by real measurements the activation regions tend to dispose themselves around closed circular paths which follow the well known circle of fifths.* In other words the Kohonen feature map recognizes a structure of *topological* similarities among the examples which is strongly reminiscent of *musical* similarities. On the other hand *when the examples are* *either drawn at random or arbitrarily fixed, or no model is used at all the similarity structure among the chords is completely scrambled up.* As a consequence a line joining, for example, the location of the major triads along the circle of fifths will appear as a closed but twisting curve on the self organizing map.

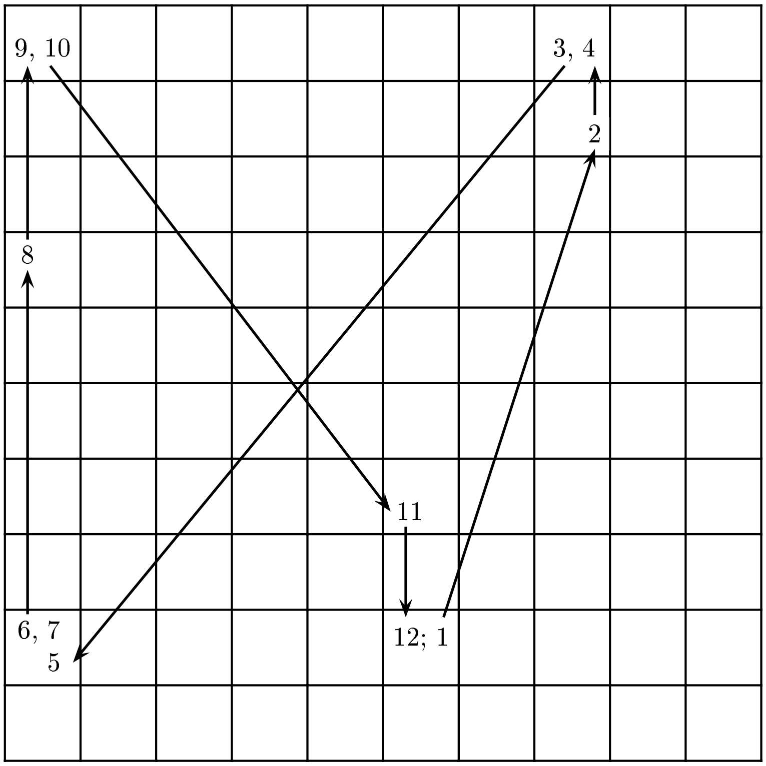

Fig. 8 Locations of the most activated neurons for major triads of the series (C) and indication of the sequence of triads in the circle of fifths.

We will not show here the complete results of the four rounds of simulations, and we will rather limit ourselves to the picture of the activation regions of the 12 major triads for the five classes (A), (B), (C), (D) and (E) of examples examinated, the disposition of minor triads and of dominant sevenths being quite similar. Remark that in the pictures displayed we indicate just the location of the most activated (nearest to the particular class of examples) processing element (neuron) corresponding to every triad or chord, since a more precise shadowing of the activation regions would inevitably hide the patterns. The association triad/most-activated-neuron is shown by numbers indicating the position of the chord in the circle of fifths according to the following scheme:

1	2	3	4	5	6	7	8	9	10	11	12
C	G	D	A	E	B	F♯	C♯	A♭	E♭	B♭	F

The lines drawn by us on the grid join in succession the location of most activated neurons following the circle of fifths and hence give an idea of the order implied by the disposition on the map. The most remarkable results are shown in Fig. 6 to 10. As can be seen from these diagrams the results are radically different in the (A) and (B) cases with respect to che (C), (D) and (E) cases. In fact when the chord images are calculated either from a pitch model or from real psychoacoustic data the major triads (but the same is true for the other types of chords) show the clear bent to dispose themselves around the closed, circular paths characteristic of the circle of fifths, indicating a clear relation between the topological proximity on the grid and the psychological similarity of the chords. On the other hand, when the examples are prepared either by randomly drawing the values of the first template, or by arbitrarily giving a regular profile, or just with no model at all, the line joining the 12 triads in the sequence of the circle of the fifths will show a twisted form which either criss-crosses its path or visits several times the same locations: this is a behaviour which is not coherent with a straight relation between topological proximity and psychological similarity.

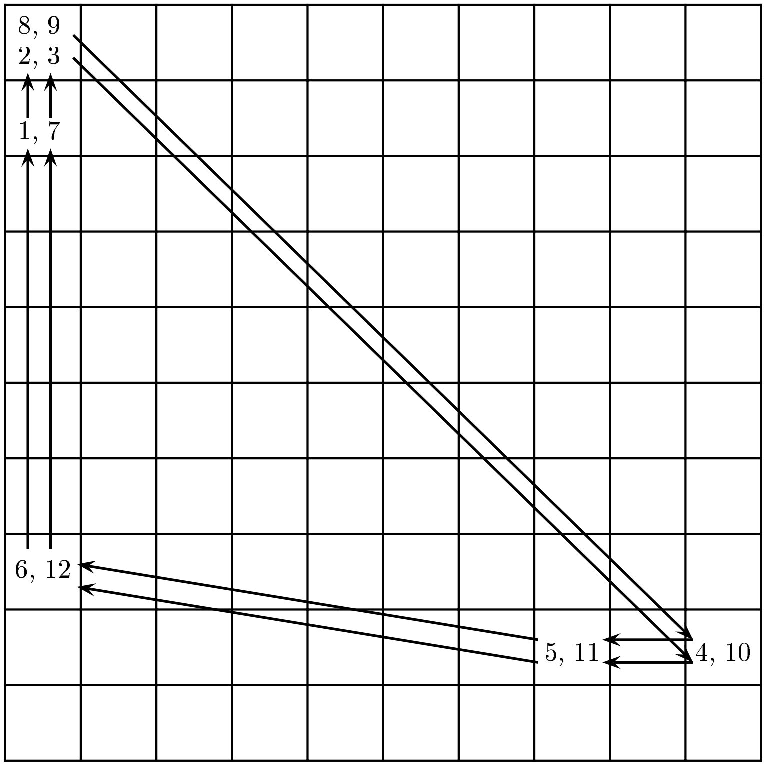

Fig. 9 Locations of the most activated neurons for major triads of the series (D) and indication of the sequence of triads in the circle of fifths.

Remark also that the structure of the circle of fifths in these simulations is different from that of previous papers [12], [13] which suggested a spiral structure, since here we have a plain circular structure. Moreover in our simulations (A) and (B) the chords show also the inclination to regroup themselves at the four corners of the square grid. Beside the oversimplification of our model, the reason for these differences can be found in the fact that in our feature map the four borders of the square grid are not folded together making the upper and lower borders, and the left and right borders coincide in a toroidal structure as in the pervious papers. It is evident that this toroidal structure would give rise to a different topology on the grid so that the circle of fifths would develop along a spiraling line rather than along a simple circular path.

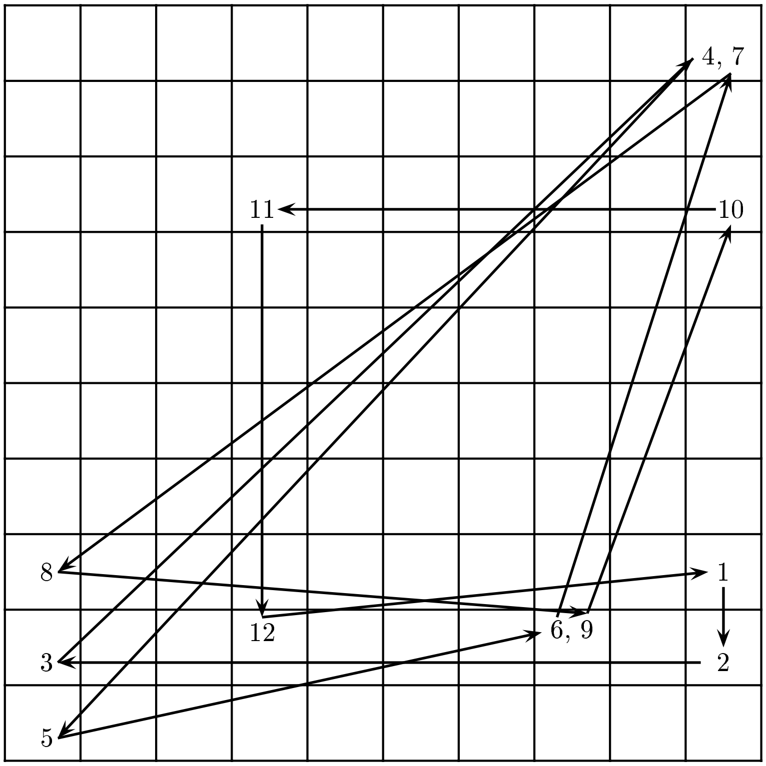

Fig. 10 Locations of the most activated neurons for major triads of the series (E) and indication of the sequence of triads in the circle of fifths.

4. CONCLUSIONS

The results summarized in this paper seem not equivocally indicate that the informational content carried either by the Krumhansl templates (A), or by the templates produced by a simple pitch model (B) is in fact musically relevant since their regular disposition in the analogous of the circle of fifths on a Kohonen feature map can not be achieved unless the archetypal template (which, by transpositions, generates all the templates of our exapmles) are the right ones. In fact we have tried to reproduce this regularity by generating anomalous examples which respect all the features of the usual examples, except for the fact that the first template is not generated by means of a pitch model or on the basis of empyrical data, and we have shown that this is in fact impossible. Since a complete mathematical theory of the Kohonen feature maps is still not at hand (see for example [7], p. 141) it could be risky to draw too sharp conclusions on a matter as complicated as this one. However in the opinion of the authors the results of the previous simulations [12], [13] and of this paper prove, beyond any reasonable doubt, the importance of the pitch models in this sort of investigations: indeed the relative insensitivity of the results to the details of these models (*robusteness*), as long as they are musically relevant, does not imply that these models are irrelevant (as the series (D) of simulations clearly indicates), but rather that the information about the tonal relations contained in every one of them is roughly the same.

REFERENCES

[1] Bharucha J.: *Music cognition and perceptual facilitation: a connectionist framework*, Music Perception 5 (1987) 1;

[2] Bharucha J.: *Neural nets and perceptual learning of tonal expectancies*, proceedings of the Conference on Music Perception and Cognition, Kyoto, October 1989.

[3] Bruhn H.: *Harmonielehre als Grammatik der Musik*; Psychologie Verlags Union (Munchen, 1988).

[4] Cufaro Petroni N., Degrassi F. and Pasquariello G.: *Detection of pitch in random acoustic signals by neural networks*; Journal of New Music Research 23 (1994) 369;

[5] Cufaro Petroni N., Guarino M. and Pasquariello G.: *Neural networks and pitch detection in random acoustic signals*; proceedings of the Conference on Engeneering Applications of Neural Networks, Otaniemi-Helsinki, August 1995, p. 389.

[6] D'Autilia R. and Guerra F.: *Qualitative aspects of signal processing through dynamical neural networks*; in *representation of musical signals*, G. De Poli, A. Piccialli and C. Roads eds., MIT Press (Cambridge, 1991) p. 447.

[7] Hecht-Nielsen R.: *Neurocomputing*; Addison-Wesley (Reading, 1990);

[8] Hertz J., Krogh A. and Palmer R.G.: *Introduction to the theory of neural computation*; Addison-Wesley (Reading, 1991);

[9] Krumhansl C.: *Cognitive foundations of musical pitch*; Oxford University Press (New York, 1990).

[10] Leman M.: *Symbolic and subsymbolic information processing in models of musical communication and cognition*, Interface 18 (1989) 141;

[11] Leman M.: *Emergent properties of tonality functions by self-organization*; Interface 19 (1990) 141;

[12] Leman M.: *The ontogenesis of tonal semantics: results of a computer study*; Report SM-IPEM 18 (Ghent University, 1990).

[13] Leman M.: *Music and schema theory*; Springer (Berlin, 1995).

[14] Murtagh F.: *Neural networks and related massively parallel methods for statistics: a short overview*, International Statistical Review, 62 (1994) 275.

[15] Parncutt R.: *Revision of Terhardt's psychoacoustical model of the roots of a musical chord*, Music Perception 6 (1988) 65.

[16] Ripley B.D.: *Pattern recognition and neural networks*; Cambridge University Press (Cambridge, 1996).