Identification of acoustic signals by artificial neural networks

Nicola Cufaro Petroni
INFM – Unità di ricerca di Bari e
Dipartimento Interateneo di Fisica
Università di Bari
Via Amendola 173, 70126 Bari, Italy
cufaro@ba.infn.it

Keywords: Backpropagation neural networks; Kohonen selforganizing maps; Acoustic signals.

Abstract

In this paper we present the results of several investigations devoted to explore the possibilities offered by the neural networks in the analysis of acoustic signals. Particular attention is given to the validation of the models used in the compression of the information in the preprocessing of the signals. Both supervised and unsupervised learning rules are adopted.

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 [1], [2], [3], and a particular interest has been aroused by the works devoted to the identification and the classification of auditory images by means of either self-organizing Kohonen feature maps [4], [5], [6], [7], [8], or backpropagation neural networks [9], [10]. It is known for example that the backpropagation neural networks are especially useful to approximate non linear regression functions [11], [12] and that they can learn to identify classes of objects in a statistical ensemble of data. This remark has led to the idea that these artificial networks could simulate the detction of a pitch in an acoustic signal. The idea of pitch is a psychological one meaning the perception of the height of an acoustic signal: whereas in a musical tone we can detect a precise pitch, in a noise we can not. However, even if signals with or without a well recognizable pitch exist in nature, there are others such that this characteristic is much more ambiguous: they constitute the more interesting challenge to these models. The performed simulations show that in fact the network can learn to distinguish between tones (with pitch) and noises (without a pitch) and is also able to sort out the particular pitch of the signal when these are classified in acoustically meaningful classes even when the signal is preprocessed (on the basis of a model) to compress the information to part relevant to the pitch identification.

2 Detection of pitch by supervised learning

The experiment [9], [10] is based on acoustic data completely simulated: the neural network has a 12 neurons input layer, a 12 neurons output layer, and one hidden layer. In the training phase we have randomly generated training sets of sample signals by mixing tones and noises in a suitable way to implement a sufficiently realistic acoustic environment. Every sample is composed of 24 numbers: 12 inputs and 12 target outputs. If we generate a tone in a particular pitch class, the target is a vector where only one value is 1 and the others are 0; if we generate a noise, the target is a vector with 12 zeros. To test the performance of the trained network we used new sample sets of tones and noises, including harmonic tones lacking the first partials in order to check the generalization power of the model: they are typical signals connected with the empirical phenomena of the residue pitch [13].

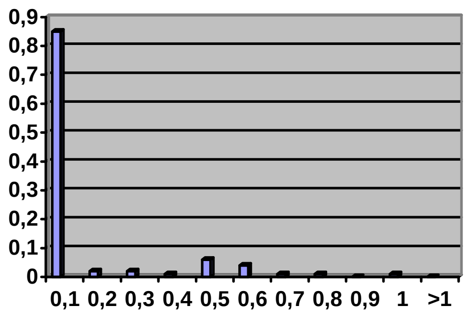

Fig. 1 Typical histogram of the errors for harmonic test tones

To evaluate the results in the test phase we compare the required and the actual outputs and we calculate a mean square error for every sample. The error will be 1 when a pitch is misclassified as a different pitch; is 0.5 when a tone is misclassified as a noise (or vice versa); is 0 when the classification is correct. Sometimes we get errors greater than 1 as spurious results. The performances are very satisfactory since the essential information about the pitch of a signal is preserved also in its reduced form. Of course the quality of the detection is better for signals of a type included in the training sets, but it is acceptable even for unheard before signals indicating that the residue pitches can still be classified on the basis of previous different experiences on complete harmonic tones. In fact the results of these simulations show the same qualitative behaviour presented in the case of the complete signals, but for the fact that the detection of the residue pitch is less clear than the perception of the complete pitch, as also happens in the real world.

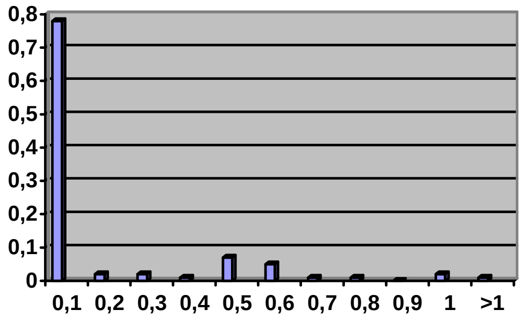

Fig. 2 Typical histogram of errors for tones with residue pitch.

Hence our conclusion is that this simple model of the real acoustic world makes sense since at least it allows one to give on account of some phenomena (the residue pitch detection) by means of simpler facts (the ability to detect the pitch of complete signals).

3 The *Circle of Fifths* by unsupervised learning

The ability of the self-organizing Kohonen feature maps in elaborating the signal information in characteristic patterns seems to give very stimulating cues about the tonal structure of (western) music. For example in [5], [6], [7] amazing regularities among the locations of the chord images fed to a Kohonen feature map have 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 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. 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, rather than of the informative content of the signal itself. We have scrutinized [8] this hypothesis in order to show that in fact these suspicions are groundless: in particular we have reproduced in a simplified form the experiment of the original papers and then we have compared their outcomes with the results of a few simulations specially designed to put in evidence the meaning of the previous ones. We have used the fact that the Kohonen feature maps are self-organizing neural networks able to associate a spatial structure to statistical data organized in clusters. In our simulation the examples are drawn from a 12-dimensional space of 12-number vectors which represent the images of musical chords obtained by means of different pitch models. 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. The sets of examples, albeit simplified with respect to the original simulations, are then built as a suitable mixture taking into account the occurrence of chords in western, tonal music. We performed the training of the net varying a few parameters: number of iterations (from 100 to 200); learning rate alpha (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: see for example Fig. 3 where for simplicity just the locations of the activations for the 12 major triads are shown (the sequence of numbers reproduces the sequence of the major triads along the circle of fifths).

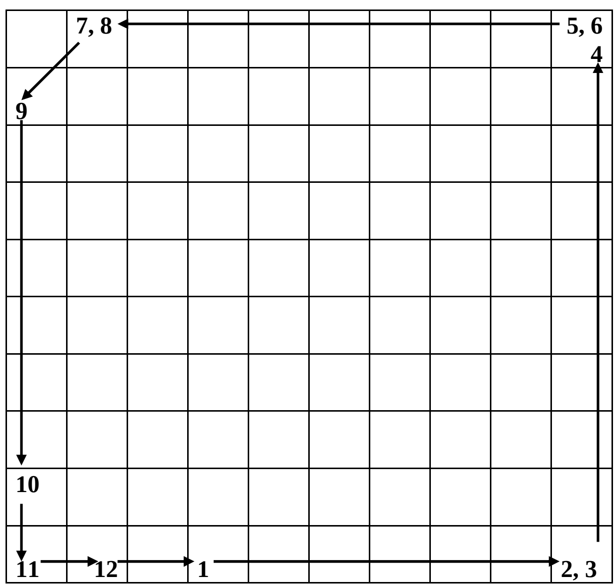

Fig. 3 Locations of the most activated neurons for major triads and indication of the sequence of triads in the circle of fifths.

However, as already remarked, we generated the tone center images by means of some 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 which work, for instance, as images for the major triads are always the same template just transposed on another degree of the chromatic octave. Since the interesting result of these simulations is the fact that a particular circular order emerges among the activation regions of the triads, a legitimate question is whether this underlying order of the chords could just be the product of the circularity of the transpositions introduced by hand in our samples to represent every step of the chromatic scale. In order to check this point we have produced new 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 produced not on the basis of a pitch model whatsoever, but for instance by means of 12 random numbers which are arbitrarily attributed as template for the first triad. The result is that when the examples are either drawn at random or arbitrarily fixed, and no auditory model is used 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 as in Fig. 4.

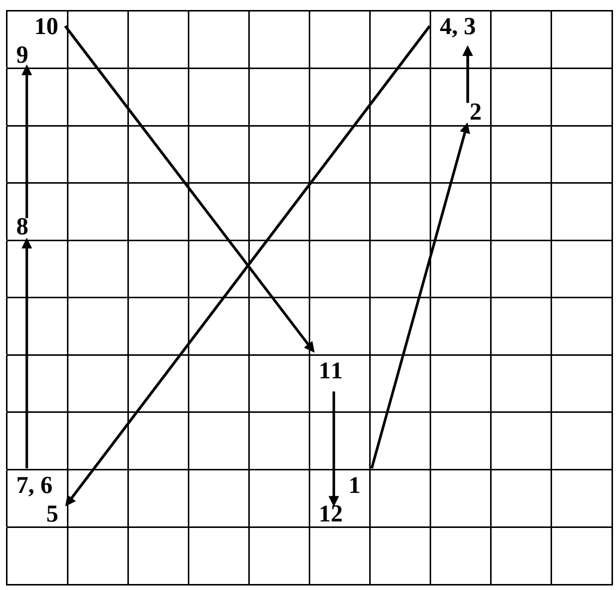

Fig. 4 Locations of the most activated neurons for randomly generated major triads.

These results seem not equivocally indicate that the informational content carried by the usual pitch models is in fact musically relevant since the regular disposition of images in the analogous of the circle of fifths on a Kohonen feature map can not be achieved unless the archetypal templates (which, by transpositions, generates all the templates of our exapmles) are the right ones. In fact we found impossible to reproduce this regularity by generating anomalous examples similar to 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. Since a complete mathematical theory of the Kohonen feature maps is still not at hand (see for example [11], 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 these and of the previous simulations prove, beyond any reasonable doubt, the importance of the pitch models in this sort of investigations.

References

[1] J. Bharucha. Music Cognition and Perceptual Facilitation: a connectionist framework. *Music Perception*, 5: 1, 1987.

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

[3] R. D'Autilia and F. Guerra. Qualitative aspects of signal processing through dynamical neural networks. In *Representation of musical signals*, G. De Poli et al Eds. MIT Press: 447, 1991.

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

[5] M. Leman. Emergent properties of tonality functions by self-organization. *Interface*, 19: 141, 1990.

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

[7] M. Leman. *Music and Schema Theory*. Springer, Berlin, 1997.

[8] N. Cufaro Petroni and M. Tricarico. Self-organizing neural nets and the perceptual origin of the Circle of Fifths. In *Music, Gestalt and Computing*, M. Leman Ed. Springer: 169, 1998.

[9] N. Cufaro Petroni, F. Degrassi and G. Pasquariello. Detection of pitch in random acoustic signals by neural networks. *J. New Music Res.* 23: 369, 1994.

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

[11] R. Hecht-Nielsen. *Neurocomputing*. Addison-Wesley, Reading, 1990.

[12] N. Cufaro Petroni and M. Nitti. Regression functions by neurocomputing. Preprint of the Bari University, 1996.

[13] E. Zwicker and H. Fastl. *Psychoacoustics, Facts and Models.* Springer, Berlin, 1990.