

# Experiments on musical instrument separation using multiple-cause models

J Klingseisen and M D Plumbley\*  
Department of Electronic Engineering  
King's College London

\* - Corresponding Author - mark.plumbley@kcl.ac.uk

## Abstract

Over the last few years, interest has been growing in neural network circles in the separation of independent sources, using techniques such as blind source separation and independent component analysis (ICA). A related technique is the 'Multiple-Cause Model' of Saund [Neural Computation, 7, 51-71, 1995]. In this technique, a neural network is trained to model the observed pattern as a composition of several underlying 'causes', in contrast to the more traditional 'winner-takes-all' neural networks which can handle only a single 'cause'. In this paper, we report on experiments which use a simple multiple-cause model to separate different instruments and notes from audio spectral representations such as perceptually scaled power spectra and wavelets. We will consider the implications of this approach for audio music analysis and compression.

## 1. Introduction

Human perception of sounds is much more advanced than any technical system so far created. A human listener is able to distinguish different tones in a complex sound structure such as a number of different human voices or musical instruments.

In this paper, we report on an approach which tries to learn patterns of sounds. Different tones, such as a violin playing the note 'A', are presented in the form of an audio signal: the goal of the system would be to recognize these tones, without any prior knowledge.

A technique that has proved to be successful at recognition of patterns like these is that of neural networks. In particular, we shall use a neural network which uses feedback connections, the *multiple cause model* (Saund, 1995). This model searches for representations of the underlying causes of the input signal by attempting to take account of all of these underlying causes. The audio signal will be pre-processed using a suitable transform (e.g. FFT or wavelets) before being passed to the multiple cause model (Figure 1).

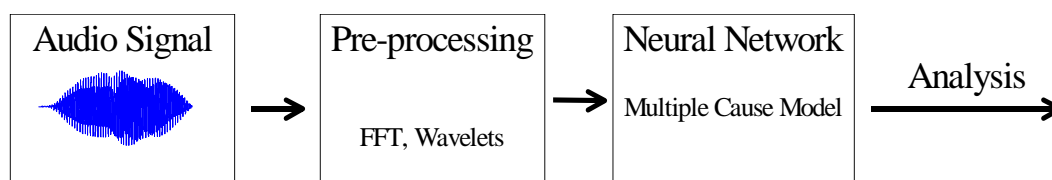


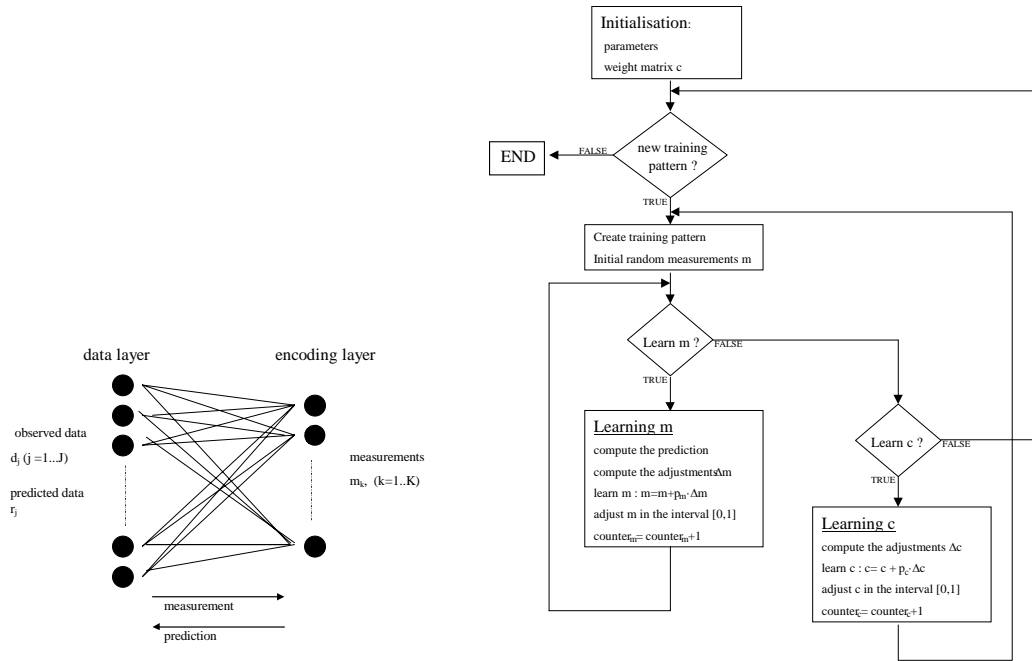
Figure 1: Analysis of an audio signal

Ultimately, this analysis of an audio signal into its underlying causes could be used for compression, since good compression would be achieved when separate parts of the signal are separated and compressed, or for audio transcription, where the separate components could be recognized (perhaps by another neural network).

## 2. The Multiple Cause Model

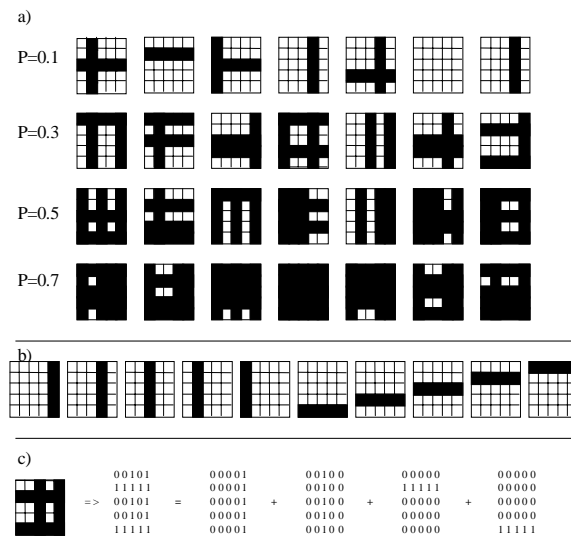
Learning in neural networks can be either *supervised* or *unsupervised* (Haykin, 1994). For supervised learning, a teacher is provided which provides a target output for the neural network, and the network is trained to minimize the error between the actual output and the target output. For unsupervised learning, no teacher is provided, and the neural network learns from input data alone. A well-known example of a neural network that can be trained using unsupervised learning is the Kohonen self-organizing map (SOM) (Kohonen, 1990). Here, any input to the neural network causes one output unit (the 'winner') in the SOM to become active: this is called a *winner-take-all* network. The SOM is very useful for extracting a low-dimensional representation of a single cause within a high-dimensional data space.

Saund (1995) introduced an alternative type of network, the *multiple-cause model* (Figure 2). This type of network is designed to cope with input data which is composed of several causes active at the same time. This network does not operate in a simple feed-forward manner: rather the encoding layer and connections are adjusted until the encoding forms a good reconstruction of the observed data. For details of the algorithms used, see Klingseisen (1999).



**Figure 2: Multiple cause model architecture and algorithm**

A simple demonstration of the capability of the multiple-cause model is on the Bars problem. Here an image is composed of a white (0) background with horizontal and vertical black (1) bars, each of which may appear with some probability  $P$ . Where two bars overlap, black (1) is the result: this is a non-linear OR-type 'write-black' imaging model (Figure 3).



(a) Data set consisting of horizontal and vertical bars in respect of the probability  $p$  of each basic pattern to appear. Because of this probability it might happen that no black pixel appears or that the whole pattern is black. (b) The basic patterns for the data in a. (c) The translation of the black and white pixels into numbers (black represented by 1, white represented by 0) and how this pattern is mixed up of its basic patterns. The mixing function is the logical or.

**Figure 3: Bars problem**

### 3. Dealing with non-binary data

In the “Bars” problem, the data used was binary basic patterns, with binary amounts (1 or 0) of each, combined with an OR mixing function to produce a binary image. However, we wish to use the multiple-cause model to deal with basic which have continuously variable (non-binary) levels (e.g. power spectra), and also have continuous amounts (e.g. volumes).

Initially we introduced only one of these non-binary elements, constructing a dataset with grey-level basic patterns (in the interval  $[0,1]$ ), but with the linear “images” created with binary amounts (present with probability  $P$ ), added with a linear mixing function. (This might be equivalent to the spectrum of a musical instrument which could either be played at a constant volume, or not at all.) With a simple modification to the multiple-cause model architecture, separation of 8 patterns with overlap of up to about 50% is possible. Above this, some patterns come to be recognized as a partial activation of other patterns, leaving a small error that is insufficient to drive the learning algorithm (Klingseisen, 1999).

In the multiple-cause model, since the reconstruction is the product of the measurements  $m$  and the weights  $c$ , there is some redundancy between these parameters. If the probability of occurrence,  $P$ , is above 0.5, we found that it was helpful to learning to constrain the interval of the mean value of the weights  $c$  for each basic pattern. For example, if we know that the smallest basic pattern has mean 0.22 and the largest has mean 0.26, we could constrain the mean to be in the interval  $[0.2, 0.3]$ , and re-scale whenever the weights of a learned basic pattern fall outside of this range. We found that this made learning more successful. Typical learning time for these grey patterns, consisting of 8 patterns, each of 16 components, is 20 minutes to an error level of 0.1 using Matlab on a Pentium II (350 MHz).

Varying the probability of appearance of each pattern had a significant effect on the learning time for each dataset. For a given error threshold, we observed that both high and low probabilities of occurrence gave rise to longer learning times than probabilities around 0.4-0.6.

So far the measurements  $m$  (volumes) have all been binary: for real music signals we would need to release this constrain to allow constantly varying volumes. To this end, we released some of the measurements  $m$  (33%, 50% and 100%) so that they could vary between 0 and 1, and the  $c$  values were also constrained to lie in the interval  $[0, 1]$ . We found that the more of the measurements that were allowed to vary, the longer was the time taken to learn the patterns, so that fastest learning is obtained when as many measurements as possible are fixed as 0 or 1 only, and the probability of occurrence is in the approximate interval  $[0.3, 0.6]$ . For more details, see (Klingseisen, 1999).

### 4. Music-based signals

To work towards audio signals, we next applied the multiple-cause model to artificial spectra produced from synthesized sounds. In the first instance, we trained the model on spectra, downsampled to 30 bins, of a clarinet playing one of 8 notes ( $G_3, C_4, A_3, D_4, F_4, G_4, A_4, E_4$ ). The training set was composed of linear additions of these basic spectra (NB: not mixed in the time domain), and needed about 800 presentations of training patterns for successful learning. Separation of patterns composed from spectra of different instruments playing the same note also worked. For six instruments, about 600 presentations was needed, with about 3000 presentations for 10 instruments (Figure 4).

Combining these two approaches, patterns composed from the spectra of three instruments (Clarinet, Oboe, Trumpet) playing each of three notes was also possible: this was successful after about 800 presentations of the training patterns.

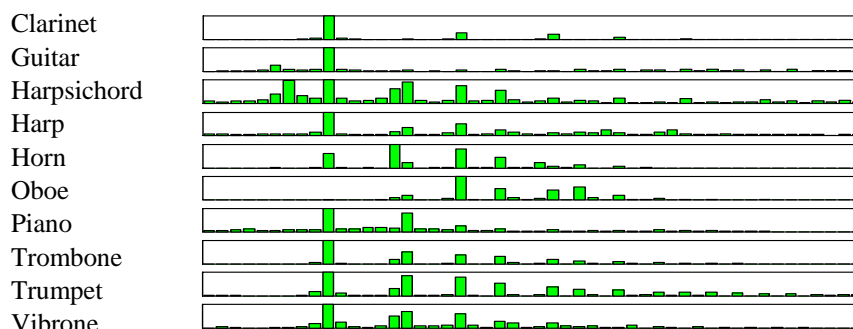
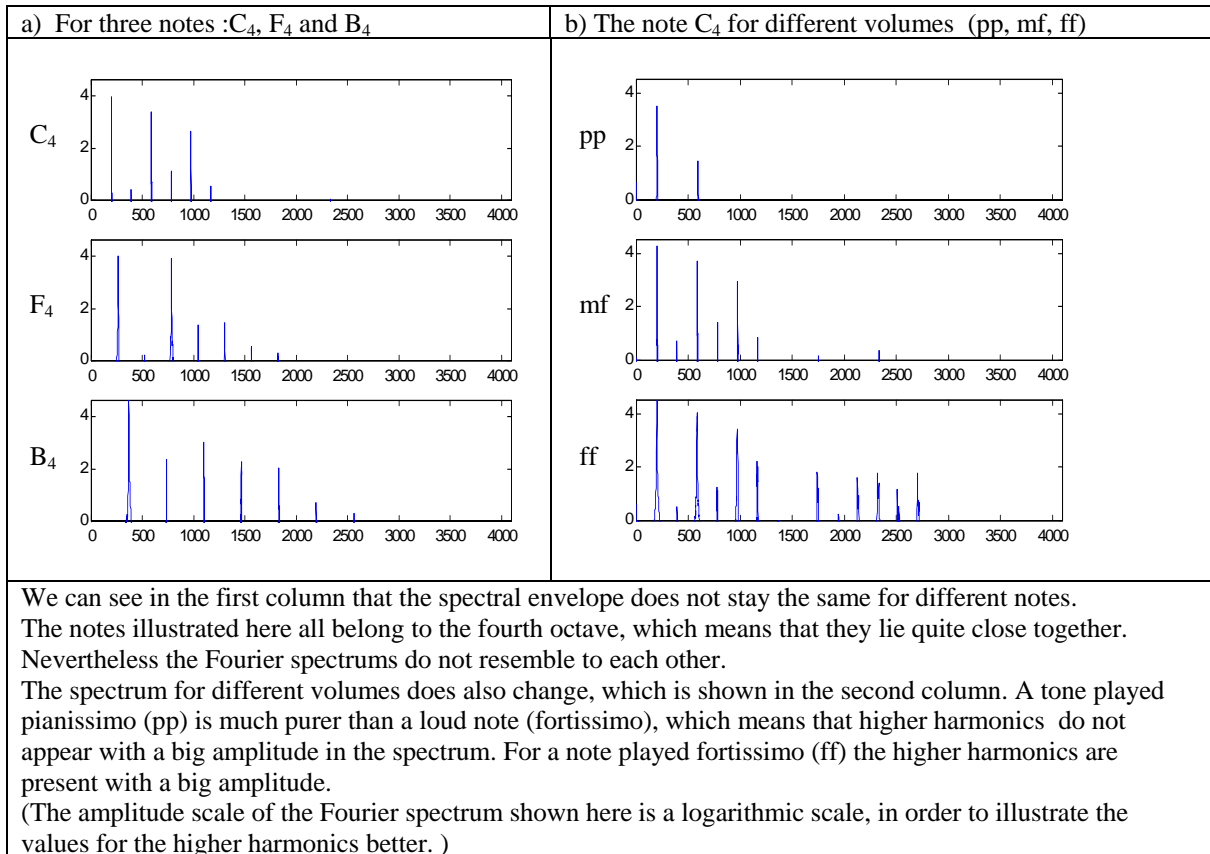


Figure 4: The basic patterns for 10 different instruments

## 5. Real sounds

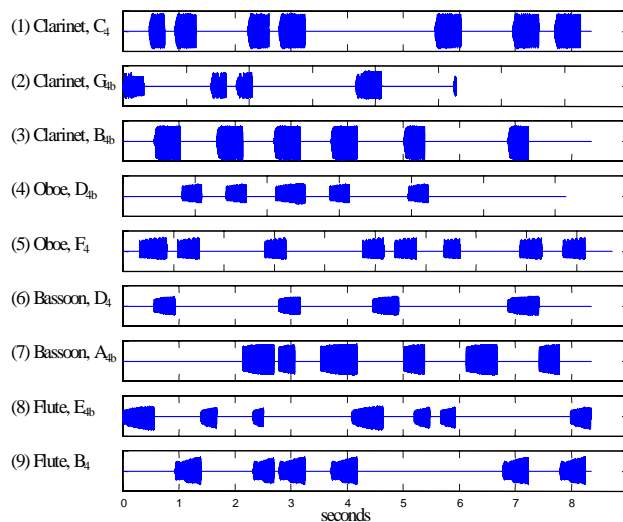
In the experiments reported on so far, we analysed synthesized sounds, with artificial “spectra” composed by linear addition of underlying spectra. We also assumed that the spectra are essentially unchanged by volume or tone changes.

For the sounds of real instruments (Iowa, 1999), the situation is more complicated (Figure 5). Volume and tone can both change the spectra, so that simple shifting or scaling is not sufficient.



**Figure 5: Fourier spectrum of notes played on a clarinet**

For this experiment, we constructed an audio signal composed of the addition of pulses of notes played on different instruments (Figure 6).



**Figure 6: Waveforms of nine notes played on different instruments**

The algorithm was adapted to use a large number of input units (spectrum of 2048 inputs), with 0.1 seconds used to generate each spectrum. No attempt was made at windowing, so patterns at the onset and offset of notes will have disrupted spectra. The algorithm found the nine underlying patterns after 300 presentations of the training patterns, equivalent to 20 hour's learning.

## **6. Conclusions**

In this paper, we have reported on initial work investigating the use of Saund's "multiple-cause model" neural network applied to audio signal separation.

While there is still a long way to go, our initial results are promising, and we feel that future work in this direction will be fruitful.

## **References**

- Haykin, S. (1994) *Neural Networks: a comprehensive foundation*. Macmillan College Publishing Company.
- Iowa (1999) *Instruments Sample of Real Sounds*; University of Iowa Musical Instrument Samples internet page; URL <http://theremin.music.uiowa.edu/~web/sound/>
- Klingseisen, J (1999) *Audio Analysis using Multiple Cause Neural Networks*. Project Report. Audio & Music Technology Lab, Department of Electronic Engineering, King's College London.
- Kohonen, T (1990) The Self-Organizing Map. *Proc. IEEE*, **78** (9), 1464-1480.
- Saund, E (1995) A Multiple Cause Mixture Model for Unsupervised learning. *Neural Computation*, **7** 51-71.