A hierarchical Neural Network-based approach to VIRGO noise identification

F. Acernese[1], F. Barone[2,3], A.Eleuteri[3,8], F. Garufi[3], L. Milano[3,4], R. Tagliaferri[5,6,7]

[1] Facoltà di Scienze MM. FF. NN., Università di Salerno, via S. Allende, 84081 Baronissi (SA) Italia
[2] Dipartimento di Scienze Farmaceutiche, Università di Salerno, via Ponte Don Melillo, 84084 Fisciano (SA) Italia
[3] Istituto Nazionale di Fisica Nucleare, sez. Napoli, via Cintia, I-80126 Napoli Italia
[4] Dipartimento di Scienze Fisiche, Università di Napoli "Federico II"
[5] Dipartimento di Matematica ed Informatica, Università di Salerno
[6] INFM unità di Salerno, via S. Allende, 84081 Baronissi (SA) Italia
[7] IIASS "E. R. Caianiello", Vietri s/m Italia
[8] Facoltà di Scienze MM. FF. NN., Università di Napoli

Abstract: In this paper a hierarchical Neural Network-based approach is presented to identify the noise in the Virgo experiment to detect Gravitational Waves by means of a laser interferometer.

1 Introduction

Neural Networks (NN's) have become in the last years a very effective instrument for solving many difficult problems in the field of Signal Processing due to their properties like nonlinear dynamics, adaptability, self-organization and high speed computational capabilty (see for example [8] and the papers therein quoted).

Aim of this paper is to show the feasibilty of the use of NN's to solve difficult problems of signal processing regarding the so called VIRGO project. The main goal of the Virgo experiment is the direct detection of gravitational waves and, in joint operation with other similar detectors, to perform gravitational waves astronomical observations. In particular, the VIRGO project is designed for broadband detection from $10Hz$ to $10kHz$ [2].

A $3km$ arm-length Michelson interferometer with suspended mirrors (test masses) is used. The phase difference $\Delta\phi$ between the two arms is amplified using Fabry-Perot cavities of Finesse 50 in each arm. Aiming for detection sensitivity of $3.10^{-23}\,\frac{1}{Hz}\,at\,100Hz$. VIRGO is a very delicate experimental challenge because of the competition between various sources of noise and the very small expected signal. In figure 1 the overall sensitivity of the apparatus is shown. In this figure it is easy to see the contribution of the different noise sources to the global noise.

In the VIRGO data analysis, the most difficult problem is the gravitational signal extraction from the noise due to the intrinsic weakness of the gravitational waves, to the very poor signal-to-noise ratio and to their not well known expected templates. In this context we use a NN approach to identify the VIRGO system.

System identification consists in finding the input-output mapping of a dynamic system. A discrete-time Multi-Input Multi-Output (MIMO) system (or a continuous-time sampled-data system) can be described with a set of nonlinear difference equations of the form (*input-output* representation):

$$\begin{aligned}\mathbf{y}(n) &= \mathcal{F}[\mathbf{y}(n-1), \mathbf{y}(n-2), \ldots, \mathbf{y}(n-n_y), \\ &\quad \mathbf{u}(n-n_d), \ldots, \mathbf{u}(n-n_u-n_d+1)] \qquad (1)\end{aligned}$$

where $\mathbf{u} \in R^M$ is the input vector and $\mathbf{y} \in R^N$ is the output vector; n_u and n_y are the maximum lags of the input and output vectors, respectively, and n_d is the pure time-delay (or dead-time) in the system. This is called an NARMAX model (non linear autoregressive mobile average with exogenous inputs,) and it can be shown that a wide class of discrete-time systems can be put in this form [5].

Given a set of input-output pairs, one or more neural networks can be built which approximates the desired functional $\mathcal{F}[\cdot]$. If we build only one network , it has $n_y N + (n_u - n_d)M$ inputs and N outputs. An alternative is to design a system composed by a hierarchy of neural network: they can be trained following a top-down approach [6]), or a bottom-up approach. This latter one is simpler and cheaper but we need to train all the neural networks by different sources; moreover, if we have a linear composition of the sources, we can build a linear superposition of the outputs of the single neural nets, and optimize the weights by a simple least square technique.

A difficulty arises from the fact that generally we do not have information about the model order (i.e. the maximum lags to take into account) unless we have some insight into the system physics. Furthermore, the system is nonlinear. Recently [1] a method has been proposed for determining the so-called *embedding dimension* of nonlinear dynamical systems, when the input-output pairs are affected by very low noise. Furthermore, the lags can be determined by evaluating the *average mutual information* [1]. Such methodologies, although not always successful, can nevertheless be used as a starting point in model design.

2 A neural network-based model of the Virgo system

As we have seen in the introduction, the Virgo interferometer can be characterized by a sensitivity curve, which expresses the capability of the system to filter undesired influences from the environment, and which could spoil the detection of gravitational waves (such a noise is generally called seismic noise). The sensitivity curve has the following expression:

$$S(f) = \frac{S_1}{f^5} + \frac{S_2}{f} + S_3 \left[1 + \left(\frac{f}{f_k}\right)^2\right] + S_\nu \tag{2}$$

where $f_k = 500Hz$ is the shot noise cut-off frequency, $S_1 = 1.8 \cdot 10^{-36}$ is the pendulum mode, $S_2 = 0.33 \cdot 10^{-42}$ is the mirror mode , $S_3 = 3.24 \cdot 10^{-46}$ is the shot noise .

The contribution $S_\nu(f)$ of violin resonances is given by:

$$S_\nu(f) = \sum_i^n \frac{1}{i^4} \frac{f_i^{(c)}}{f} \frac{C_c \phi_i^2}{\left[\left(\frac{f}{i f_i^{(c)}}\right)^2 - 1\right]^2 + \phi_i^2} + \frac{1}{i^4} \frac{f_i^{(f)}}{f} \frac{C_f \phi_i^2}{\left[\left(\frac{f}{i f_i^{(f)}}\right)^2 - 1\right]^2 + \phi_i^2} \tag{3}$$

where the different masses of close and far mirrors are taken into account:
$f_i^{(c)} = i \cdot 327Hz$, $f_i^{(f)} = i \cdot 308.6Hz$,$C_c = 3.22 \cdot 10^{-40}$, $C_f = 2.82 \cdot 10^{-40}$, $\phi_i^2 = 10^{-7}$.
Samples of the sensitivity curve $\{S_i\}_i$ can be obtained by evaluating the expression (2) at a set of frequencies $\{f_i\}_i, f_i \in [10, 2048]Hz$. By an inverse Discrete Fourier Transform, samples of the system transfer function can be obtained in the time domain.

Assuming that the interferometer input noise is a zero mean gaussian process, by feeding it to the system (i.e. filtering it through the system transfer function) we obtain a coloured noise.

2.1 Experimental Results

The first step we followed is to design the NN archite cture: from a previous paper [2] we see that a single NN cannot predict alle the features of the VIRGO noise spectrum (e.g. the high frequency resonances cannot be modeled). The idea is to use one different NN for each noise source: we train each NN with a time series whose spectrum is a component of the global spectrum we want to model. Since the total noise is a linear combination of the single sources, we can use an ADALINE to combine the outputs of the lower-level NN's. In our experiments we used 4 MLP's corresponding to the following sources: shot noise, mirror noise, pendulum noise, violin noise.

The second step is model order determination of each of the NN's composing the hierarchy. To determine suitable lags which describe the system dynamics, we used the average mutual information (AMI) criterion

[1]. This can be seen as a generalization of the autocorrelation function, used to determine lags in linear systems. A strong property of the AMI statistic is that it takes into account the non-linearities in the system. Usually, the lag is chosen as the first minimum of the AMI function.

To find how many samples are necessary to unfold the (unknown) state-space of the model (the so called *embedding dimension* [1]) we used the False Nearest Neighbours algorithm. In order to test the NN's capability in solving the problem, we chose a width of 5, both for input and output (i.e. $n_y = n_u = 5$). In this way, we obtained a NN with a simple structure. Furthermore, some preliminary experiments showed that the system dead-time is $n_d = 0$; this gives the best description of the system dynamics.

Another fundamental issue is NN's complexity, i.e. the number of units in the hidden layers of the NN's. Usually the determination of NN complexity is critical because of the risks of overfitting. Since the NN's were trained following a Bayesian framework, overfitting was of no concern; so we directed our search for a model with the minimum possible complexity. In our case, we found an hidden layer with 8 tanh units is optimal for the first three MLP's, while 20 are sufficient for the MLP corresponding to the violin noise source (due to high frequency resonances).

The Bayesian learning framework (see [9] and [10]) allows the use of a *distribution* of NN's, that is, the model is a realization of a random vector whose components are the NN weights. The so obtained NN is the *most probable* given the data used to train it. This approach avoids the bias of the cross-validatory techniques commonly used in practice to reduce model overfitting [3]. To allow for a smooth mapping which does not produce overfitting, several regularization parameters (also called *hyperparameters*) embedded in the error criterion have been used. The approach followed in the application of the Bayesian framework is the "exact integration" scheme, that is we sample from the analytical form of the distribution of network weights.

The chosen models were trained using a sequence of less than eight hundred thousands patterns (we sampled the system at 4096Hz for 180s) normalized to zero mean and unity variance with a *whitening* process. Note that the input-output pairs were processed through discrete integrators to obtain pattern-target pairs. The NN's were then tested on several 1s long sequences composed by a signal (chirp, burst, sinusoids, Lorentz series, Henon maps) added to coloured noise.

The NN's were trained for 200 epochs, with the hyperparameters being updated every 15 epochs. The simulations were made using the MATLAB© language, the Netlab Toolbox [4] and other software designed by us.

In figure 3 the predictive capabilities of the model on a chirp signal is shown, while in figure 4 the correlations of the true signal and estimated one versus the signal-to-noise ratio are illustrated.

3 Conclusion

In this paper we have shown some preliminary tests on the use of a hierarchy of NN's for signal processing in the VIRGO Project. From our experience, the obtained results are much better than those obtained by a single MLP (also with Bayesian inference learning techniques), recurrent NN's (e.g. Elman NN) or linear AR models. Some observations can be elicited from the experimental results:

- By adopting a hierarchical model we have been able to simplify model training and also reach better results for each of the sources.
- The model can identify the VIRGO noise from its components up to $2048Hz$ with good performance as shown by the figures. Furthermore, with respect to our previous model, we have been able to model also the high frequency resonances (the violin noise).
- We can reach an high signal identification rate with signal-to-noise ratios up to $25db$ for many kind of signals.

The next steps in the research are:

- to extend the model with other noise sources (this can be simply achieved with our approach because the NN's are independet of each other);
- to test the models with a greater number of samples to obtain a better estimate of the system dynamics;

- to model the noise inside the system model to improve the system performance an to allow a multi-step ahead prediction (i.e. a output-error model)

References

[1] H. D. I. Abarbanel, Analysis of Observed Chaotic Data, Springer, 1996.

[2] F. Barone, R. De Rosa, A. Eleuteri, F. Garufi, L. Milano, R. Tagliaferri, A Neural Network-based ARX Model of Virgo Noise, in Neural Nets WIRN Vietri-99, M. Marinaro and R. Tagliaferri eds., pp. 171–183, Springer-Verlag, London, 1999.

[3] C. M. Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford, 1996.

[4] C. M. Bishop, I. T. Nabney, Netlab Toolbox, Neural Computing Research Group, Aston University, Birmingham, 1996.

[5] S. Chen, S. A. Billings, Modelling and Analysis of Nonlinear Time Series, Int. J. Control, Vol. 50, No. 6, pp. 2151–2171, 1989.

[6] M. I. Jordan, R. A. Jacobs, Hierarchical mixtures of experts and the EM algorithm, Neural Computation, vol. 6 no. 2, pp. 181–214, 1994.

[7] L. Ljung, System Identification Toolbox User's Guide, The MathWorks Inc, 1995.

[8] F. Luo, R. Unbehauen, Applied Neural Networks for Signal Processing, Cambridge University Press, Cambridge, 1997.

[9] D. J. C. MacKay, Hyperparameters: optimise or integrate out ?, Maximum Entropy and Bayesian Methods, Santa Barbara, Dordrecht, 1993

[10] R. M. Neal, Bayesian Learning for Neural Networks, Ph.D. thesis, University of Toronto, Canada, 1994.

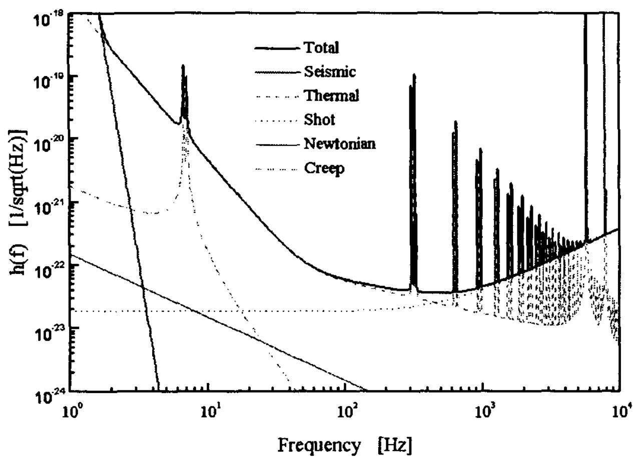

Figure 1: Sensitivity curve of the VIRGO interferometer

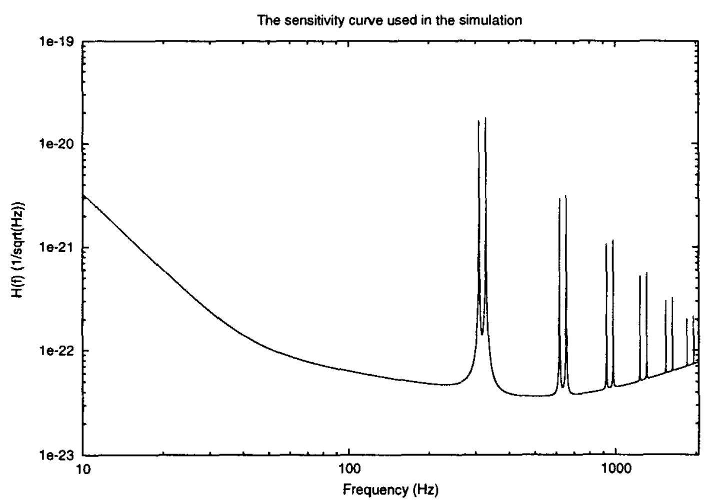

Figure 2: The simplified sensitivity curve used for the experiments

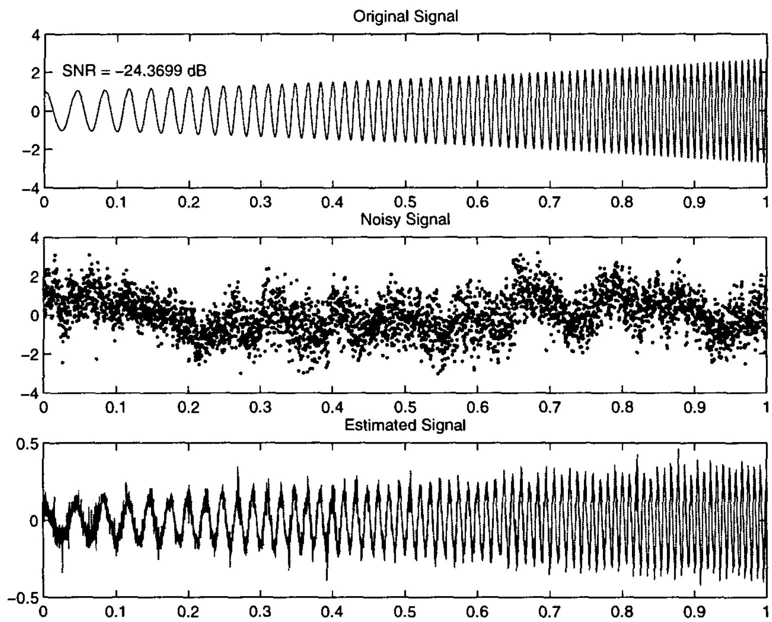

Figure 3: Test signal I: $\exp(-t/\tau) \times chirp$

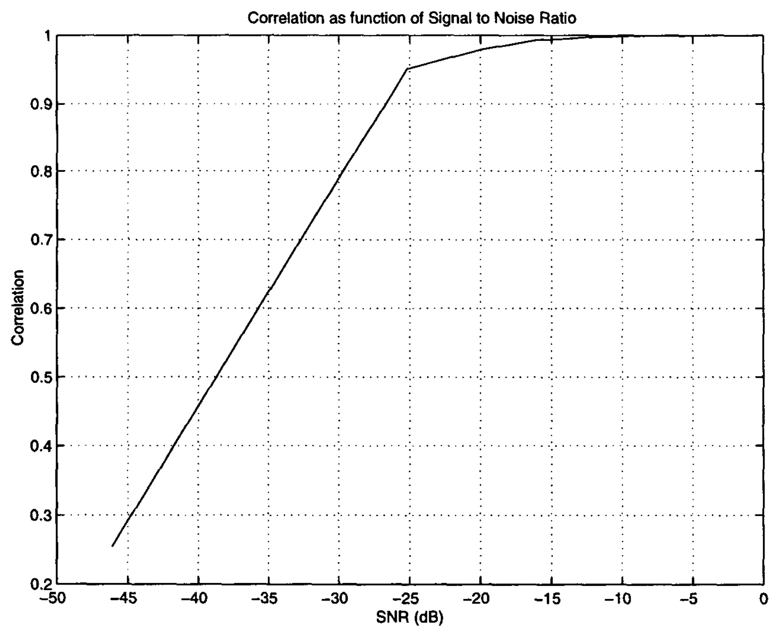

Figure 4: