The Neural Impulse Response Filter

Volker Tresp[a], Ira Leuthäusser[a], Martin Schlang[a], Ralph Neuneier[b], Klaus Abraham-Fuchs[c] and Wolfgang Härer[c]

[a]Siemens AG, Central Research and Development, Otto-Hahn-Ring 6, 8000 München 83, Germany

[b]Fachbereich Informatik der Universität Kaiserslautern, Germany

[c]Siemens Medical Engineering Group, 8520 Erlangen, Germany

Abstract

The neural network filter architecture presented in this paper is well suited for a restricted class of nonlinear adaptive filter applications. Our filter can model systems with long time responses and it is able to learn to take into account additional parameters that influence the filter response nonlinearly. Our filter was successfully used in a biomedical application for the removal of the cardiac interference from magnetoencephalographic (MEG) data.

1   Introduction

Neural networks have been used in a number of adaptive nonlinear filter applications such as interference cancellation [1], prediction of time-series [2, 3, 4] and speech recognition [5]. The neural architectures which were used in these applications are very general. Our paper demonstrates that it is sometimes advantageous to relinquish some of this generality for an architecture that is more efficient in a particular application. Recent work on on complexity and generalization has demonstrated clearly that optimal performance can only be achieved if the architecture and number of free parameters are chosen according to the problem at hand: a system cannot be modeled by a network with insufficient resources but, on the other hand, an architecture that is too complex requires a large amount of training data to sufficiently specify all parameters in the network and, in addition, it might generalize badly on new data.

We propose a neural network architecture that we have coined the neural impulse response (NIR) filter, which is an efficient model for certain systems with complex responses. Such systems can be found in biomedical applications (the heart beat is a complex response to a heart trigger signal), industrial applications (adaptive noise cancellation) and physiology (modeling the reaction to the administration of drugs). Our filter was successfully used in a biomedical application for the removal of the cardiac interference from magnetoencephalographic (MEG) data.

2 The NIR Filter

2.1 The Architecture

We consider a linear filter which can be described by

$$g(n) = \sum_{i=0}^{M} f(n-i) \ k(i) + \sum_{i=1}^{K} g(n-i) \ r(i). \tag{1}$$

The output of the filter is a linear superposition of present and past inputs $f(n-i)$ and past outputs $g(n-i)$. If the output depends only on the present and past inputs, the equation describes a finite impulse response (FIR) filter. A generalization to a nonlinear filter is described by

$$g(n) = \mathcal{N}[f(n)...f(n-M), g(n-1)...g(n-K)]. \tag{2}$$

Here, the next sample depends nonlinearly on the inputs and outputs. If $\mathcal{N}$ is approximated by a neural network, one obtains as special cases the filter architecture of Lapedes and Farber [2] or the time-delay neural network of Waibel [5].

We consider systems in which an input affects the output of the system over a long time-period. A recursive network can model systems with long responses. The alternative is to choose M to be equal to the length of the desired system response. This requires a large number of inputs to the network and therefore results in inefficient training.

We suggest that for some systems with long responses, the neural impulse response (NIR) filter architecture can be employed to generate models that can be trained efficiently. The output of the NIR filter can be described by

$$g(n) = \sum_{i=0}^{M} \mathcal{N}[f(n-i), i, \mathbf{p}(n-i)] \tag{3}$$

Similar to the FIR filter, the output of the filter is the sum of responses to previous inputs to the system. The difference from an FIR filter is that the responses depend nonlinearly on the input $f(n-i)$, the time i that has passed since that input occurred and the parameter vector $\mathbf{p}$. The relation to an FIR filter becomes even more apparent if we consider the following variation. If we assume that the $f(n-i)$ linearly influences the output, we obtain

$$g(n) = \sum_{i=0}^{M} f(n-i) \ \mathcal{N}[i, \mathbf{p}(n-i)] \tag{4}$$

and if the filter is independent of $\mathbf{p}$

$$g(n) = \sum_{i=0}^{M} f(n-i) \mathcal{N}[i]. \tag{5}$$

In the last equation, there is only one input to the network and that is the time that has passed since the input $f(n-i)$ occurred. It describes an FIR filter with a filter response that is modeled by a neural network. This, by itself, has certain advantages. If M is large (in the following application around 400) it requires a lot of training data to specify all the coefficients in an FIR filter. If the neural network has fewer than M parameters, it can give a more efficient description of the filter function and therefore require less training data to learn to approximate the filter response. A related fact is

that the neural network imposes an implicit complexity constraint on the filter response. The complexity of an FIR filter is typically reduced by decreasing M. This can be done by either decreasing the time window (which is what we do not want since we want the filter to have a long response) or by decreasing the sample period, thus reducing the bandwidth of the filter. The complexity of the neural network, on the other hand, is defined by the number of neurons and weights in the network. A network has the ability to assign resources wherever the data indicate that the function to be approximated is complex, and assign few resources where the data indicate that a good approximation can be achieved without many resources. In the application that is described later, for example, the response has a narrow peak (QRS-complex[1]) but little complexity anywhere else.

In Equation 4, the parameter vector $\mathbf{p}$ is an additional input to the network. Through $\mathbf{p}$, the response can be influenced by external, typically slowly changing parameters. In medical applications one of these parameters can be respiration, and in other applications it might be the time of day, the temperature or the humidity. Here, the efficiency of the NIR filter is even more important. For a neural network, a new parameter simply requires an additional input. If, on the other hand, for every possible $\mathbf{p}$ a new FIR filter needs to be trained, the number of coefficients and, therefore, also the number of required training data quickly becomes enormous.[2]

Equation 3, finally, describes the most general form of the NIR filter in which the input $f(n-i)$ influences the response nonlinearly.

2.2 Training

In supervised learning, the task is to minimize

$$E = 1/2 \sum_n (g_m(n) - g(n))^2. \tag{6}$$

Here, $g_m(n)$ denotes the desired output of the system at time n. E can be minimized using gradient descent, if a weight w in the network is updated such that

$$\Delta w \propto \sum_n (g_m(n) - g(n)) \frac{dg(n)}{dw} \tag{7}$$

with (considering Equation 3)

$$\frac{dg(n)}{dw} = \sum_{i=0}^{M} \frac{\partial \mathcal{N}[f(n-i), i, \mathbf{p}(n-i)]}{\partial w}. \tag{8}$$

2.3 Implementation

An implementation of an NIR filter requires an architecture similar to an FIR filter (Figure 1A). In a buffer, we store the last M samples and the current sample of the input f and the parameter vector $\mathbf{p}$. For the calculation of a new output $g(n)$, we successively present $f(n-i)$, $\mathbf{p}(n-i)$ and i ($i : 0...M$) to the input of the neural network and sum the corresponding outputs in the accumulator. For the next sample, the contents of the

---

[1] The QRS-complex is the predominant peak in a typical electro-cardiogram (ECG). See Figure 2.

[2] In a way, this corresponds to the representation of a regression function by either a look-up table (FIR) or a neural network (NIR).
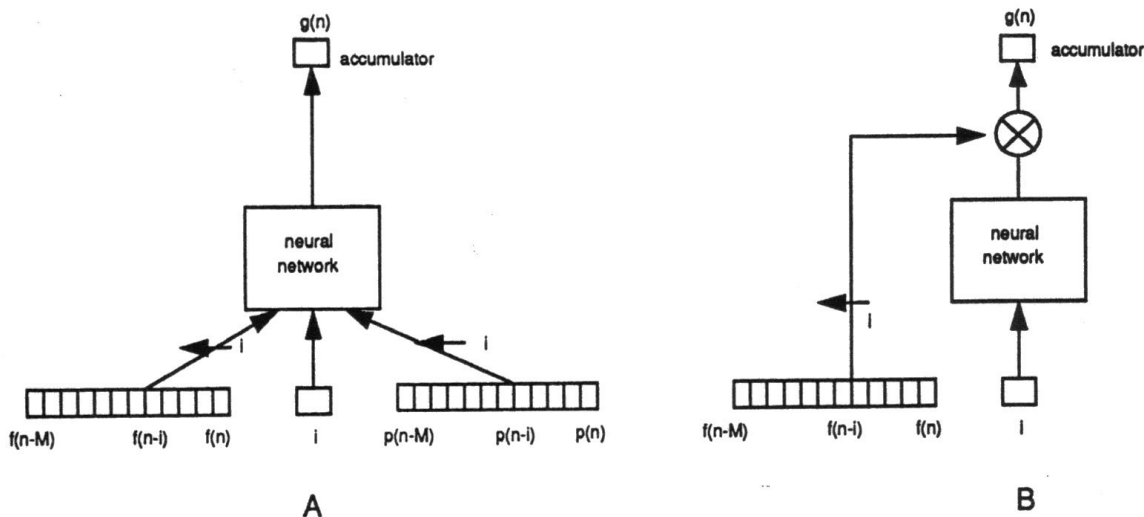
Figure 1: NIR filter architecture. The architecture on the left corresponds to the filter described in Equation 3 and the architecture on the right to the filter described in Equation 5.

buffers are shifted by one sample and a new sample of the input and the parameter vector is entered into the buffer. As can be seen, the calculations are more involved than with an FIR filter. Although a neural network should be used in the training phase, the implementation might sometimes be more straightforward if the mapping which the neural network has learned is stored in a look-up table, in particular, if the simplified filter described in Equation 5 is used. For that filter, the architecture is shown in Figure 1B. Here, $f(n - i)$ is not applied to the network but multiplies the output of the neural network. In the next section, we will show how the NIR architecture can be utilized in a practical application.

3   Magnetoencephalography

In Magnetoencephalography (MEG), an array of highly sensitive superconducting SQUID detectors is employed to scan noninvasively the minute magnetic fields produced by the brain, e.g. with the Siemens KRENIKON$^{\circledR}$. These measurements are used to localize sources of strong neural activity such as centers of epilepsy [7]. Since the magnetic field of the earth is eight orders of magnitude larger than the magnetic field produced by the brain, the measurements have to be performed in a magnetically shielded room. The magnetic field produced in the cardiac muscle is a source of interference which cannot be shielded as easily and also interferes with the measurement.

Removing the cardiac-interference is a difficult task[3] since neither MEG nor cardiac-interference can be measured independently and a mathematical model of the physical and physiological processes which generate the cardiac-interference are impossible to derive from first principles. But since we can acquire an independent measurement of the source of the interference signal in form of the electro-cardiogram (ECG), it is possible to train a neural network to form a model of that process. In the training phase, the input to

---

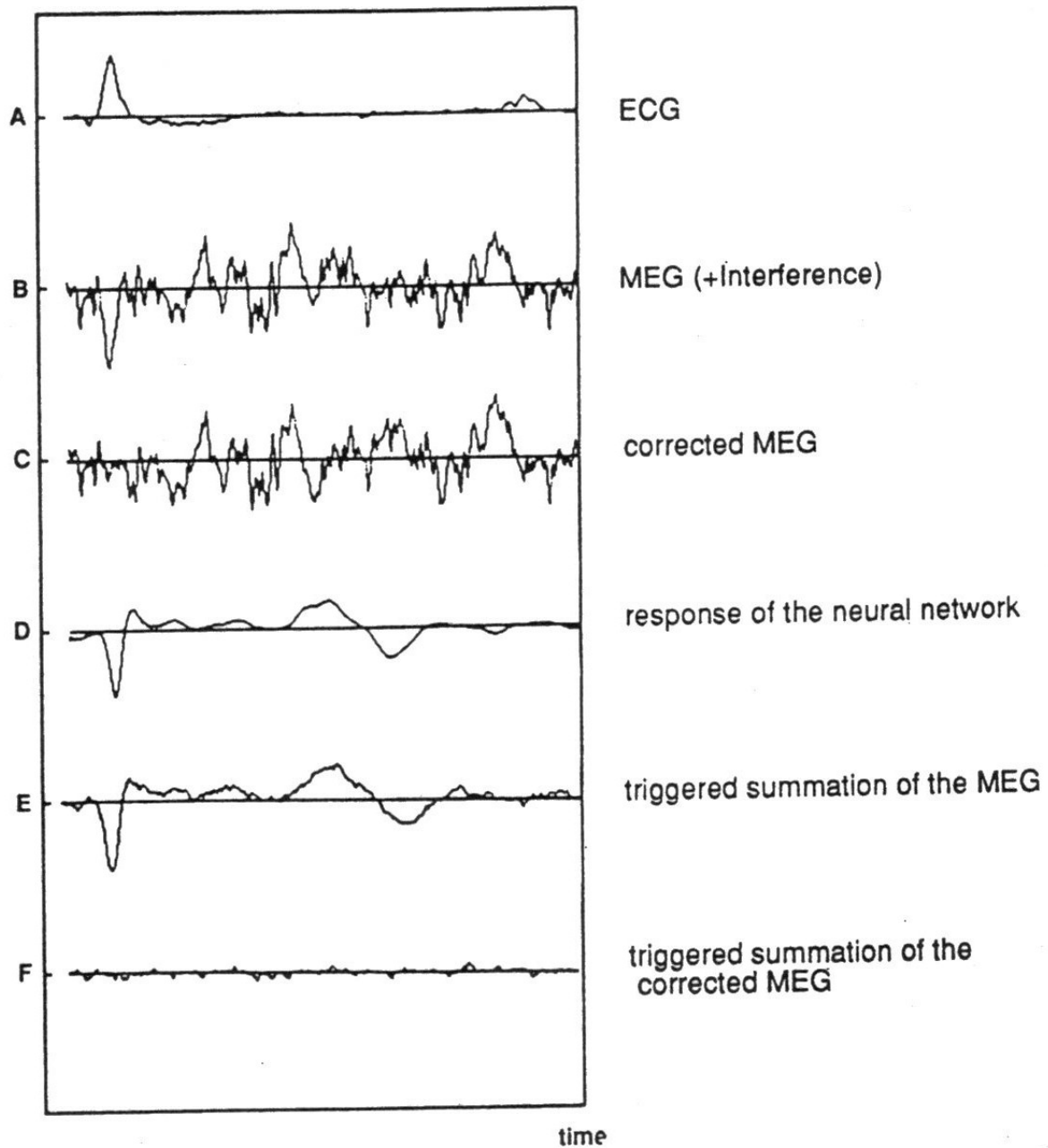[3]An alternative method for the removal of the cardiac-interference is described in [6].

Figure 2: Removal of the cardiac-interference. C is the corrected MEG (C=B-D). A comparison between E and F demonstrates how well the interference was removed. The time window is approximately one heart beat (1 second).

the neural network is the ECG and the desired output is the measured MEG signal. In the recall phase, the output of the neural network describes an estimate of the cardiac-interference which then can be subtracted from the measured MEG. As has been already shown by Widrow [1], necessary conditions for the neural network to be able to learn to predict the cardiac-interference are

- desired signal (MEG) and interference are uncorrelated.

- reference signal and interference are correlated.

- the measured signal is a linear superposition of the interference and the desired signal.

In the first experiments, we trained a linear filter and a time-delay neural network to learn to predict the interference from samples of the EKG. Both architectures failed, mainly because of the high order of the required filter and the strong nonlinearities.

In our approach, we simplified the problem by first nonlinearly extracting a delayed estimate of the heart trigger from the ECG by using a QRS-detector. The input to the NIR filter has a unit sample at the peak of the QRS-complex and is equal to zero elsewhere. A further input was the amplitude of the QRS-complex of the most recent heart beat. Another relevant input would be the respiration because the heart moves during breathing but this was not yet considered.

In the experiments we used a radial-basis-function neural network with typically 20 hidden units. Figure 2C shows the MEG signal after removal of the interference. In a test, we averaged the MEG triggered on T_H. If the cardiac-interference is not completely removed, it will add up in phase and should become visible after a number of averages. Figure 2F shows that after interference removal the averaged signal consists mostly of random noise.

4   Conclusion

The viability of the NIR filter was demonstrated on a difficult interference canceling problem. The results demonstrate that a restricted problem-specific architecture should generally be preferred over more general approaches. The NIR filter can generate complex responses and should therefore also find applications in robotics to train and reproduce complex time sequences, such as intricate movements of a joint.

Appendix

Here, we briefly describe a model of the generation of EKG and cardiac-interference C. Both ECG and cardiac-interference can be thought of as complex responses to the last heart trigger: $ECG(n) = \sum_i f_{ECG}(i) \, T_H(n-i)$ and $C(n) = \sum_i f_C(i) T_H(n-i)$. Here, $T_H(n-i) = 1$ if at time $n-i$ a heart trigger occurred and zero otherwise. The filter has to learn the complex mapping $f_F = f_C \, f_{ECG}^{-1}$. The filter function f_F can be simplified under certain circumstances (for example, if $f_C = f_{ECG}$) but not in this application. The peak of the QRS-complex occurs at a time interval after T_H, therefore, a QRS-detector can be used to generate a delayed estimate of T_H.

References

[1] B. Widrow, S. D. Stearns. *Adaptive signal processing.* Prentice Hall, Englewood Cliffs, N. J., 1985.

[2] A. Lapedes, R. Farber. How neural nets work. In: *Neural information processing systems.* D. Z. Anderson (ed), New York, American Institute of Physics, pp. 442 - 456, 1988.

[3] J. Moody, C. Darken. Fast learning in networks of locally-tuned processing units. *Neural Computation*, Vol. 1, pp. 281-294, 1989.

[4] A. S. Weigend, D. E. Rumelhart, B. A. Huberman. Generalization by weight-elimination with application to forecasting. In: *Advances in Neural Information Processing Systems*, R. R. Lippmann *et al.* (eds.), Morgan Kaufmann, 1991.

[5] A. Waibel. Modular construction of time-delay neural networks for speech recognition. *Neural Computation*, Vol. 1, pp. 39 - 46, 1989.

[6] P. Strobach, K. Abraham-Fuchs, W. Härer. Event-synchronous cancellation of the heart wave in biomedical signals. *Biomedical Engineering.* Accepted for publication.

[7] S. Schneider, E. Hoenig, H. Reichenberger, K. Abraham-Fuchs, G. Daalmans, W. Moshage, A. Oppelt, G. Röhrlein, H. Stefan, J. Vieth, A. Weikl, A. Wirth. A multichannel biomagnetic system for high-resolution functional studies of brain and heart. *Radiology 167*, pp. 825-830, 1990.