# Artificial Neural Network Based Processing in Cochlear Spaced Sub-Bands for Adaptive Speech Enhancement

*Amir Hussain and Douglas R. Campbell*
Department of Electronic Engineering and Physics
University of Paisley, High St., Paisley PA1 2BE, Scotland U.K.
Corresponding Author's email: amir@diana22.paisley.ac.uk

## ABSTRACT

A general class of single-hidden layered, linear-in-the-parameters feedforward Artificial Neural Networks is proposed for processing band-limited signals in a multi-microphone sub-band adaptive speech enhancement scheme. The sub-band spacing within the adaptive speech enhancement system is set according to a published cochlear function. Comparative results achieved in simulation experiments demonstrate that the proposed sub-band scheme is capable of significantly outperforming conventional full-band and sub-band noise cancellation methods employing linear processing, in the presence of non-linear interference.

## 1. INTRODUCTION

Humans are capable of detecting and understanding speech at low Signal-to-Noise Ratios (SNR) without prior knowledge of the speech, the noise or the environment [1]. In speech enhancement research, new ideas are often stimulated by a study of the human auditory system. An important and much studied property is that of the filterbank present in the cochlea, which splits incoming signals into a large number of band-limited signals prior to further processing. In practice, the use of sub-band based speech enhancement systems has been shown to give the important benefit of supporting adaptive diverse parallel processing in the sub-bands [3] [16]. It allows signal features within the sub-bands, such as the noise power, the coherence between the in-band signals from multiple sensors and the convergence behaviour of an adaptive algorithm, to influence the subsequent processing within the respective frequency band.

Classical speech enhancement methods based on full-band multi-microphone noise cancellation implementations [4] which attempt to model acoustic path transfer functions can produce excellent results in anechoic environments with localized sound radiators [5], however performance deteriorates in reverberant environments [2]. Multi-band processing has been found to be important in combating reverberation effects [3] [6]. Adaption is necessary to compensate for changing noise fields [7] due for example to, non-Gaussian sources, source/sensor motion, or time-varying acoustic paths. Multi-sensor methods are necessary to compensate for reverberation and speech/noise spectral overlap [3].

In previous multi-band noise cancellation systems, filter-bank or transform methods are first employed to provide a set of contiguous sub-bands within which continuous time signals can be processed. The subsequent processing within each sub-band is performed using linear adaptive filters often using the Least Mean Squares (LMS) algorithm. Significant performance improvements have been demonstrated (in terms of error convergence speed and output SNR) over classical full-band Finite Impulse Response (FIR) filter based noise cancellers spacing for speech enhancement in both simulated and real reverberant environments [2] [3] [8]. However, in cases where the acoustic path transfer functions to be modelled are not linear, the conventional linear adaptive filters will not be able to optimally cancel the non-linear interference. Additionally, acoustic signals of interest cannot generally be modeled as Gaussian processes (e.g., speech amplitude density functions are usually approximated by Laplacian or Gamma densities), and Knecht et al [14] have recently demonstrated performance improvements through the use of non-linear filtering in the full-band case.

Over the past decade, there has been an increasing interest in the use of "biologically inpired" Artificial Neural Networks (ANNs) for solving complex real-world problems [9][10]. This is mainly due to their ability to effectively deal with non-linearity, non-stationarity and non-Gaussianity [10]. The conventional feedforward neural networks include the category of multi-hidden layered, Multi-Layered Perceptron (MLP) type structures [9], and the single-hidden layered, Radial Basis Function [9], Volterra Neural Network (VNN) [11] and newly developed Functionally Expanded Neural Network (FENN) [12] type structures. All have been shown to be capable of forming an arbitrarily close approximation to any continuous non-linear mapping. However, the multi-layered MLP type networks have highly non-linear-in-the-parameters structures, and require computationally expensive non-linear updating algorithms (such as back-propagation) which are very slow and consequently less suitable for on-line adaptive applications [12]. On the other hand, the RBF, VNN and FENN have linear-in-the-parameters structures giving the relative advantages of ease of analysis and rapid adaption.

In this paper, we investigate the use of a class of general non-linear adaptive FIR type filters based on single hidden-layered linear-in-the-parameters ANNs, for processing the band-limited signals in a multi-band speech enhancement system. The spacing of the sub-bands within the adaptive speech enhancement scheme is set according to a published cochlear function. We show for a real speech signal corrupted with simulated non-linear interference that non-linear ANN based processing in the sub-bands can significantly enhance the performance of conventional multi-band speech enhancement systems.

## 2. STRUCTURE OF ANN BASED ADAPTIVE NON-LINEAR FILTER

The general structure of the proposed non-linear FIR type filter illustrated in Figure 1, is based on single-hidden layered, linear-in-the-parameters feedforward ANNs. It employs an input expander which transforms the $n$ inputs $[x_1, \ldots, x_n]$ (representing lagged values of the input signal $x$ passed through a ($n$-1)-th order tapped delay line) into a non-linear intermediate (hidden) space of increased dimension $N$. The expanded input terms (termed the basis functions) are then weighted and linearly combined to form the adaptive filter output $y$. The overall mapping of the adaptive FIR filter is thus $\boldsymbol{R}^n \to \boldsymbol{R}^N \to \boldsymbol{R}$. The non-linear expansion model is completely general and can employ for example:

**(i)** Any of the non-linear basis functions commonly employed in RBF neural networks [9], such as the:

**(a)** Thin-plate spline basis functions of the $n$ inputs:

$$f_i(\mathbf{u}_i) = \mathbf{u}_i^2 log(\mathbf{u}_i) \qquad (2.1)$$

where $\mathbf{u}_i = ||\mathbf{x} - \mathbf{c}_i||$ for $i$=1,...,$N$ ; $\mathbf{x}=[x_1 \ . \ . \ . \ x_n]$ is the input vector, $f_i(.)$ are the $N$ non-linear basis functions of the inputs, $||\ .\ ||$ denotes the Euclidean norm, $\mathbf{c}_i$ are the centres of the basis functions, and $N$ is the number of RBF centres. The centres are some fixed points in the $n$-dimensional input space, which they must sample.

**(b)** The multi-quadratic activation functions:

$$f_i(\mathbf{u}_i) = (\mathbf{u}_i^2+\sigma^2)^{1/2} \qquad (2.2)$$

where $\sigma$ is a real constant usually termed the width of the basis function.

**(c)** The inverse multi-quadratic functions:

$$f_i(\mathbf{u}_i) = 1/(\mathbf{u}_i^2+\sigma^2)^{1/2} \qquad (2.3)$$

**(d)** And the most widely used Gaussian basis functions:

$$f_i(\mathbf{u}_i) = exp(-\mathbf{u}_i^2/\sigma^2) \qquad (2.4)$$

**(ii)** The sigmoidal basis functions employed in MLP networks [10]:

$$f_i(\mathbf{x}) = tanh(\mathbf{x}) \qquad (2.5)$$

**(iii)** The Volterra (polynomial) expansion employed in the hidden layer of the conventional VNN [11]:

$$\mathbf{f}(\mathbf{x}) = [1, x_{i1}, x_{i1}\, x_{i2}, \ldots, x_{i1}\, x_{i2} \ldots x_{ik}] \qquad (2.6)$$

for $i1, i2, \ldots, ik = 1, \ldots, n$; and $\mathbf{f}(.)=[f_1 \ldots f_N]$. The above represents a $k$-th order polynomial expansion of the $n$ inputs.

**(iv)** A *hybrid* functional expansion employed in a newly developed Functionally Expanded Neural Network (FENN) [12]:

$$\mathbf{f}(\mathbf{x}) = [1, \mathbf{x}, sin(i\mathbf{x}), cos(i\mathbf{x}), x_j sin(x_k), x_j cos(x_k), x_{i1}x_{i2}, \ldots, x_{i1}x_{i2} \ldots x_{ik}] \qquad (2.7)$$

for $i$=1,2,3; $j \neq k$ and $j,k$=1,...,$n$ ; $i1 \neq i2, \ldots, \neq ik$, and $i1, i2, \ldots, ik = 1, \ldots, n$. The above expansion comprises a combination of sigmoidal shaped, Gaussian shaped and polynomial subset activation functions.

An additional benefit of employing the FENN's functional expansion model (like the VNN's polynomial expansion) is that the use of the original network inputs within the expansion model, also enables efficient modeling of linear dynamical transfer functions [12].

### 2.1 Adaptation Algorithms

(1) Compute the filter output at time k, as

$$y(\mathrm{k}) = \mathrm{F}^T(\mathrm{k})\, \mathrm{W}(\mathrm{k}-1) \qquad (2.8)$$

where F(k) defines the $[N,1]$ hidden layer vector comprising the enhanced input functions:

$$\mathrm{F}(\mathrm{k}) = [\, \mathrm{f}_1(\mathrm{k})\ \mathrm{f}_2(\mathrm{k}) \ldots \mathrm{f}_N(\mathrm{k})\, ]^T$$

where $\mathrm{f}_i(\mathrm{k})$, i=1, ..., $N$ represent the basis functions described above, superscript T denotes vector transpose, and W(k-1) is the $[N,1]$ filter weight vector given by:

$$\mathrm{W}(\mathrm{k}-1) = [\, \mathrm{w}_1(\mathrm{k}-1)\ \mathrm{w}_2(\mathrm{k}-1) \ldots \mathrm{w}_N(\mathrm{k}-1)]^T$$

(2) The output error for the filter output is:

$$e(\mathrm{k}) = d(\mathrm{k}) - y(\mathrm{k}) \qquad (2.9)$$

where $d$(k) is the desired or reference signal. The Mean Squared Error (MSE) is therefore (where E(.) denotes the expectation operator and T denotes matrix transpose):

$$\mathrm{E}(e(\mathrm{k})^2) = \mathrm{E}(d(\mathrm{k})^2) - 2\mathrm{W}(\mathrm{k}-1)^T\, \mathrm{E}(d(\mathrm{k})\mathrm{F}(\mathrm{k})) + \mathrm{W}(\mathrm{k}-1)^T \mathrm{E}(\mathrm{F}(\mathrm{k})\mathrm{F}(\mathrm{k})^T)\, \mathrm{W}(\mathrm{k}-1) \qquad (2.10)$$

The corresponding minimum MSE (MMSE) for the ANN based filter can thus be readily written as [12] (with superscript -1 denoting matrix inverse and assuming that the autocorrelation matrix $\mathrm{E}(\mathrm{F}(\mathrm{k})\mathrm{F}(\mathrm{k})^T)$ is non-singular):

$$\mathrm{MMSE} = \mathrm{E}(d(\mathrm{k})^2) - \mathrm{E}(d(\mathrm{k})\mathrm{F}(\mathrm{k}))^T\, \mathrm{E}(\mathrm{F}(\mathrm{k})\mathrm{F}(\mathrm{k})^T)^{-1}\, \mathrm{E}(d(\mathrm{k})\mathrm{F}(\mathrm{k})) \qquad (2.11)$$

which includes as a special case the best linear (Wiener) MMSE for $\mathrm{F}(\mathrm{k})=[\mathrm{x}_1(\mathrm{k}) \ldots \mathrm{x}_n(\mathrm{k})]^T$. The advantage of this particular non-linear filter structure is that linear adaptive filter theory can be readily applied for on-line adaptation.

The above MSE expression (2.10) guarantees that there will be no local minima, since the MSE is a quadratic function of the filter weights W. Fast and certain convergence may be obtained in practice by use of conventional stochastic gradient LMS type or faster but computationally more expensive least squares based RLS type adaptation algorithms, all with well established convergence properties. Other computationally efficient Fast RLS (FRLS) type algorithms and recently reported class of algorithms linking the LMS and RLS can also be readily employed to adapt the ANN based filter [12].

Thus, once the full expansion model at the single hidden layer of the ANN based filter has been specified, any of the above algorithms can then be used to provide an efficient means for real time adaptation of the filter weights. This will give these non-linear FIR filters a significant advantage over multi-layered (MLP type) neural network based filters [14] recursive applications.

## 3. SUB-BAND SCHEME BASED ON COCHLEAR MODELING

In previous work [3][17], the sub-band filters within a multi-microphone sub-band adaptive (MMSBA) speech enhancement system illustrated in Figure 1, were spaced linearly in the frequency domain. The human cochlea, which evolved to deal with all sounds available to the human ear, has been modeled by Ghitza [1] who proposed use of the logarithm function for approximating the cochlear distribution of filters. However, Greenwood [15] has presented the following more accurate function for the spacing of the filters in the *mammalian* cochlea:

$$F(x) = A(10^{ax} - k)\ \text{Hz} \qquad (3.1)$$

where $x$ is the proportional distance from 0 to 1 along the cochlear membrane, $A$, $a$ and $k$ are constants based on empirical knowledge of the cochlea, and $F(x)$ are the upper and lower cut-off frequencies for each filter obtained by the limiting value of $x$. For the human cochlea, values of $A$=165.4, $a$=2.1 and $k$=0.88 are used, and this is confirmed by Allen [16]. The number of filters within the cochlear filterbank is not accurately known, and different researchers have suggested various numbers of filters within their models. In this work, the sub-bands are achieved by modifying the spectra of the FFT (or DCT) of the input signals, and the number of filters is therefore limited by the size of the FFT.

## 4. SIMULATION RESULTS

A multi-band version of the classical noise cancellation system is illustrated in Figure 2, where the processed signal e(k) represents the sum of the errors between the band-limited primary signal and the output of the adaptive filter within each Sub-Band Processing (SBP) block. The complete system was implemented in MATLAB. The filter-bank was realized using the real-valued Discrete Cosine Transform (DCT) method. The sub-bands were distributed according to the cochlear function illustrated in equation 3.1.

The desired signal at the primary channel was a real anechoic speech signal s(k) sampled at 10kHz, and the reference noise signal was $n(k) = 0.285 \sin(2\pi 1000k)$. This noise was passed through a non-linear transfer function to produce the correlated noise signal $n'(k)$:

$$n'(k) = 0.3n(k) + 0.6n(k-1)^2 + 0.9n(k-2)^3 - 0.6n(k-3)^2 - 0.3n(k-4)$$

for addition to the speech in the primary channel. The above transfer function was arbitrarily chosen in order to provide a test case with a strong non-linearity. The SNR at the primary input was approximately -1.4dB. Ten thousand samples (representing one second) of the reference signal n(k) and the primary signal s(k)+n'(k) were used.

The enhancement performance of the conventional full-band linear FIR (FBLFIR) based noise canceller was then compared with that of the multi-band linear FIR (MBLFIR) based noise canceller and the proposed multi-band non-linear FIR (MBNLFIR) based noise canceller. The input expansion model within the non-linear FIR filter employed for the SBP, was chosen to comprise the Volterra Series expansion model (2.6), employing a truncated 2nd order polynomial expansion of the filter inputs. The exponentially weighted RLS algorithm was used for adapting the weight coefficients of all the noise cancellers.

In order to make the comparisons as fair as possible, an attempt was made to balance the computational complexity of the three algorithms. The order of the FBLFIR filter was set to 84. For this demonstration, the number of sub-bands in the MBLFIR system was set to four and the order of the linear FIR filter within each band was thus chosen as 21. In the case of the MBNLFIR system with 4 sub-bands, the order of the non-linear VNN based FIR filter within each band was set to 7. A truncated 2nd order polynomial expansion of the sub-band NLFIR filter inputs was employed comprising the actual sub-band filter inputs, their square terms and 2nd order cross-product terms, which resulted in a total of 21 terms (basis functions).

The Mean Squared Error (MSE) achieved by the various noise cancellers over the last nine and a half thousand samples (allowing the first five hundred samples for convergence) is compared in Table 1. The SNR improvements are also shown in parenthesis.

| | FBLFIR | MBLFIR | MBNLFIR |
|---|---|---|---|
| MSE (SNR improvement) | $8.6\times10^{-4}$ (7.9dB) | $8.5\times10^{-4}$ (8dB) | $5.1\times10^{-5}$ (20.2dB) |

Table 1: Performance comparison of various Adaptive Noise Cancellers.

As can be seen from Table 1, for this test case the use of a MBLFIR system gives similar performance in cancelling the non-linear interference compared to the conventional FBLFIR filter. However, the use of the proposed MBNLFIR system can be seen to produce a much greater performance increment over the MBLFIR system. Informal listening tests also showed the MBNLFIR processed speech to be both enhanced in SNR and of significantly better perceived quality than that obtained by the other methods.

## 5. CONCLUSIONS

A class of general ANN based adaptive non-linear FIR type filters has been presented together with the adaptation algorithms employed in a sub-band adaptive speech enhancement scheme which is based conceptually on the human auditory system. Initial comparative results achieved in simulation experiments using highly non-linear distortion, demonstrate that the use of ANN based non-linear filters within a multi-band noise cancellation system can significantly enhance performance compared to the conventional linear filtering based multi-band and full-band noise cancellers. The superior performance of the MBNLFIR system is due to the use of non-linear ANN based processing within the sub-bands, which enables a more efficient modeling of the non-linear noise transfer function. Furthermore, for the case study considered in this paper the complexity of the MBNLFIR system was forced to be comparable to that of the MBLFIR system, and it can be further reduced by employing for example, a self-structuring LMS type algorithm. Note that although a particular simulated non-linear noise transfer function has been chosen for the case study, in practice, any non-linear dynamical transfer function can be efficiently modelled using the MBNLFIR approach, since all the ANN based non-linear expansion models employed in the proposed adaptive non-linear filters are universal approximators [12]. Current experiments are using anechoic speech signals corrupted with realistic reverberated noises using both linear and non-linear sub-band distributions in order to further investigate the performance of the proposed MBNLFIR noise canceller. Initial results have been very encouraging and will be reported elsewhere.

## 5. ACKNOWLEDGEMENTS

During this work, Dr. A.Hussain was supported by the U.K. EPSRC Project Ref. GR/K48907, and Prof. D. R. Campbell was supported by a Leverhume Fellowship.

## 6. REFERENCES

[1] O.Ghitza, *Auditory models and human performance intasks related to speech coding and speech recognition,* IEEE Trans. on Speech and Audio Proc., Vol.2, pp.115-132, 1994.
[2] D.R.Campbell, *Speech Enhancement for Hearing Aids,* Proceedings EUSIPCO, pp.467-470, Trieste, Italy, 1996.
[3] E.Toner and D.R.Campbell, *Speech Enhancement using sub-band intermittent adaption,* Speech Communication, Vol.12, pp.253-259, 1993.
[4] B.Widrow, S.D.Stearns, *Adaptive Signal Processing* , Prentice-Hall, 1985.
[5] H.S.Dabis, T.J.Moir and D.R.Campbell, *Speech enhancement by recursive estimation of differential transfer functions,* Proceedings ICSP, Beijing, pp.345-348, 1990.
[6] M.M.Goulding and J.S.Bird, *Speech enhancement for mobile telephony,* IEEE Trans. On Vehicular Technology, Vol.39 no.4, pp.316-326, 1990.
[7] R.B.Wallace and R.A.Goubran, *Improved tracking adaptive noise canceller for non-stationary environments,* IEEE Trans. on Signal Processing, Vol.40, no.3, pp.700-703,
[9] D.R.Hush and B.G.Horne, *Progress in Supervised Neural Networks: What's new since Lippmann* , IEEE Signal Processing Magazine, pp.9-39, Jan'1993.
[10] S.Haykin, *Neural Networks expand Signal Processing's Horizons,* IEEE Signal Processing Magazine, pp.25-49,1996.
[11] P.J.W.Rayner, M.R.Lynch, *A new connectionist model based on a non-linear adaptive filter* , IEEE ICASSP, pp.1191-1194, Glasgow, 1989.
[12] A. Hussain, *A new Neural Network Structure for Temporal Signal Processing,* Proceedings IEEE ICASSP, Munich, Germany, 20-24 April 1997.
[14] W.G. Knecht, M.E.Schenkel and G.S. Moschytz, *Neural Network Filters for Speech Enhancement,* IEEE Transactions on Speech and Audio Processing, Vol.3, No.6, November'95
[15] D.D.Greenwood, *A cochlear frequency-position function for several species-29 years later,* J. Acoustic Soc. Amer, Vol.86, No.6, pp.2592-2605, 1990.
[16] J.B.Allen, *How do humans process and recognise speech?,* IEEE Trans. Speech and Audio Proc., Vol.2, No.4, pp.567-577, 1994.
[17] A.Hussain and D.R.Campbell, *A Multi-microphone Sub-band Adaptive Speech Enhancement System employing Artificial Neural Network Based Non-linear Filters* , Proceedings ESCA-NATO Workshop, Pont-a-Mousson, France, 17-18 April 1997.

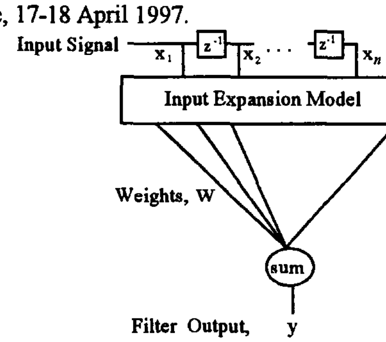


Figure 1: General structure of the proposed ANN Based Adaptive non-linear Filter

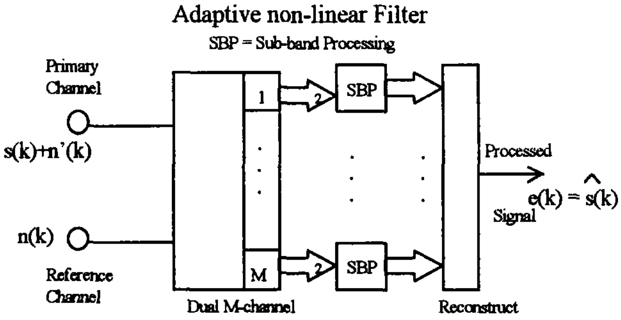


Figure 2: The multi-band version of the classical noise canceller