Speech Enhancement through Nonlinear Adaptive Source Separation Methods

Nikos Doukas, Tania Stathaki and Patrick Naylor

Signal Processing Section, Dept of Electrical Engineering,Imperial College,
Exhibition Road, London SW7 2BT, United Kingdom.
e-mail: n.doukas@ic.ac.uk

Abstract

In this paper, a new method that exploits the ideas of independent source separation in the context of Speech Enhancement in single sensor signals, is developed and tested in various situations. The channel distortions of the two sensor case are artificially reproduced by suitable linear and nonlinear filters. Separation is implemented via a Lagrange neural network. Results on speech signals are shown.

1. Introduction

Recently there has been considerable work on the problem of source separation (see e.g [7], [8], [10]). In its simplest form the problem is given a linear mixture of signals (sources), to separate the contribution of each of the sources present assuming they are independent. Other interesting work in the area has been presented in [3], [6] and [9]. Previous research has focused mainly on multisensor approaches to the problem where different mixtures of the source signals arrive at each one of the sensors. Such approaches are difficult to use in practice, because of the increased complexity imposed by the presence of an array. The approach of our work is to produce estimates of the signals present using just one sensor. The different distortions normally suffered by the signals in the channel are modelled locally by suitably filtering the received signal. A Lagrange minimisation problem is formed to be solved by a Lagrange programming neural network ([11]). The results of the application of the method on contaminated speech signal are included.

2. The Source Separation Problem

Consider two independent signals x_1 and x_2 propagating in the same medium and two sensors, each receiving a different mixture of the two signals, i.e. $y_1 = a_{11}x_1 + a_{12}x_2$ and $y_2 = a_{21}x_1 + a_{22}x_2$.

It can then be shown that the initial signals can be recovered as ([1]):

$$s_1 = b_1 x_1 = w_{11}y_1 + w_{12}y_2 \quad (1)$$

$$s_2 = b_2 x_2 = w_{21}y_1 + w_{22}y_2 \quad (2)$$

where b_1 and b_2 are constant gains and the w_{ij} depend only on the a_{ij}s. This recovery may be performed provided that $a_{11}a_{22} - a_{12}a_{21} \neq 0$.

Since the a_{ij}s are of course not known, the w_{ij}s must be estimated through some kind of optimisation procedure. The two signals are by assumption idependent, zero mean implying that their odd powered cross moments are zero. This fact can be exploited for this optimisation. Examples of ways to estimate these moments are given in [1] and [7]. The method of estimation used in our work will be presented later on in this paper.

A typical block diagram of a source separation apparatus, is given in figure 1. The first part of the circuit (marked as 'CHANNEL') reproduces the distortions that would normally be suffered by the signals in the channel. The second part (marked as 'NEURAL NET') is the one that recovers the mixed signals. The weights w_{ij} are controlled by some adaptive mechanism, specific to each method.

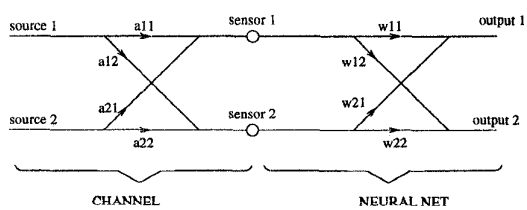


Figure 1: Standard source separation setup

0-8186-7576-4/96 $5.00 © 1996 IEEE

3. The single sensor case

The modified arrangement for the new method is depicted in figure 2. In our case there is only one signal available, namely the noise contaminated signal (marked as 'sensor'). A two sensor simulation can be made in such a manner that the distortions that the signal would undergo when travelling through a channel are modelled by passing it through two different filters (shown in figure 2 as $H1$ and $H2$). Some guidelines for chosing these filters are given later in this paper. This produces two pseudo-sensor signals, shown as "sensor 1" and "sensor 2". These two signals are then used as substitutes for the signals from the two sensors.

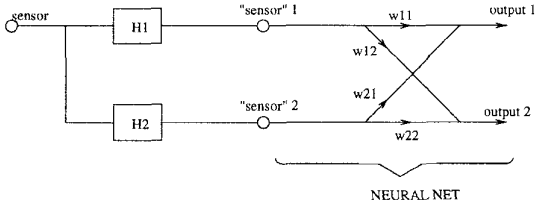


Figure 2: Block diagram of the setup used for the new method

The adaptation mechanism is further assisted by the introduction of constraints. A constrained optimisation problem is set up and its solution implemented through the use of Lagrange Programming Neural Networks. This type of neural networks are based on the Lagrange minimisation theory. They were chosen because they permit the introduction of constraints, but exhibit further advantages in terms of speed of convergence, ability to readapt and good stability. Details about them are given in [11] and [4].

It has already been mentioned that odd power cross moments of the outputs must be zero, and the function to be minimised is therefore taken to be

$$J = \sum_{i,j} \left(E[s_1^{2i+1} s_2^{2j+1}] \right)^2 \qquad (3)$$

subject to the constraint that $s_1 + s_2 = y$ where y is the received signal. This gives the following Lagrange function to be minimised:

$$J = \sum_{i,j} \left(E[s_1^{2i+1} s_2^{2j+1}] \right)^2 + \lambda(s_1 + s_2 - y) \qquad (4)$$

The update equations for w_{ij} and λ can be obtained by using (1) and (refequ2) and differentiating the above expression. A steepest descent adaptation is then performed.

In this study i and j are restricted so that:

$$(i, j) \in \{0, 1\}^2$$

For reasons of simplicity only the two source case is considered.

4. Implementation Issues

The received signal which is assumed to be a linear mixture of the two source signals is passed through two separate filters. The two outputs are used in our setup in the manner of a standard source separation problem ([5], [7]). These filters should not have high stopband attenuation so that both the outputs convey information about all frequency components of the signals. Further investigations as to the choice of these filters are currently under way.

It can be easily seen that the following modification to the objective function, reduces the computational load considerably:

$$J = \left(\sum_{i,j} E[s_1^{2i+1} s_2^{2j+1}] \right)^2 + \lambda(s_1 + s_2 - y) \qquad (5)$$

Possible further implications of this modification are currently under investigation.

Several alternative methods for estimating the cross moments of the signals have been investigated. Clearly, since we are dealing with higher order moments, a large number of samples must be used for reducing the variance of the estimation. The fact however that the signals can not be assumed stationary poses a limit on the number of past samples that can be meaningfully used in the estimation. For these reasons the following recursive formula was used:

$$\left(E[s_1^i s_2^j] \right)_n = \phi \times \left(E[s_1^i s_2^j] \right)_{n-1} + (1-\phi) \times s_{1,n}^i s_{2,n}^j \qquad (6)$$

where

$$\left(E[s_1^i s_2^j] \right)_n$$

is the estimate for the moment at time n, and $s_{k,n}$ is the value of signal s_k at time n and ϕ is a forgetting factor. Equation (6) provides an unbiased estimate for the moments. Clearly it produces good estimates of the value of the moments, since a large number of samples is involved. Additionally, with a suitable choice of ϕ, it can quickly respond to changes in the statistics of the signal.

A variable gain adaptation was used to give better stability and eliminate oscillations of the weights in a dynamic Lagrange neural network realisation. For stationary environments the adaptation gain modification is taken to follow the rule:

$$\mu = \mu_0 \frac{1}{(iteration\ number)^\beta} \tag{7}$$

where β is a positive constant. Typically $0 \leq \beta \leq 2$. This update method is used in current literature ([1]). It gives an initial, near optimal solution quickly and then convergences with small missadjustment.

Solutions for non-stationary cases are currently being explored.

5. Results

Convergence is fast and due to the variable gain there are no weight oscillations after the final values are reached. Sample convergence curves for the weights of the neural network can be seen in figure 3.

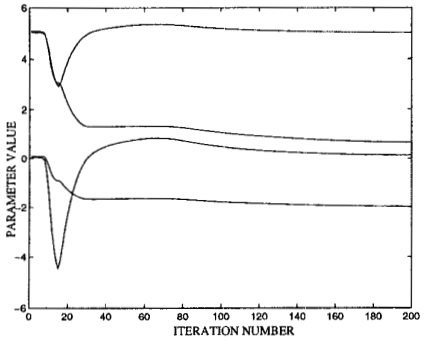


Figure 3: Sample convergence curve for the weights

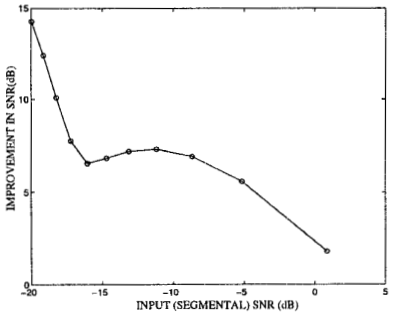
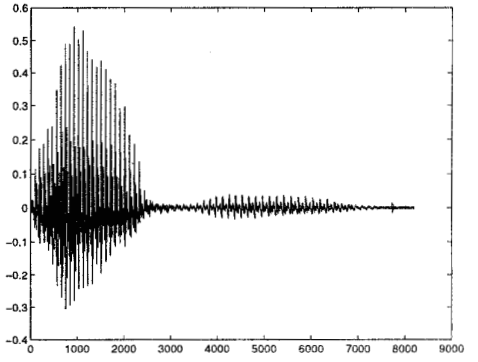
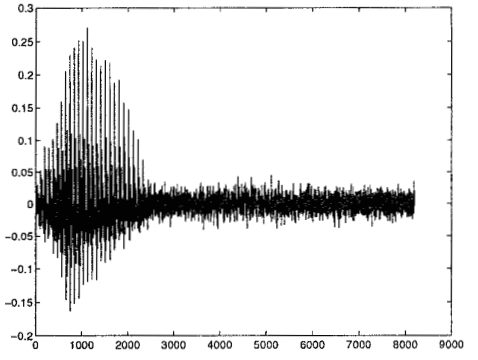


Figure 4: Improvement in SNR after processing versus input SNR (both measured as segmental SNR)

The tests were performed on single sinusoid plus white, zero–mean, gaussian noise, speech plus sinusoid and speech plus white, zero–mean, gaussian noise. Sample results for speech plus white noise, can be seen in figures 5 (the original and the contaminated signals) and 6 (the reconstructed signals).



(a)

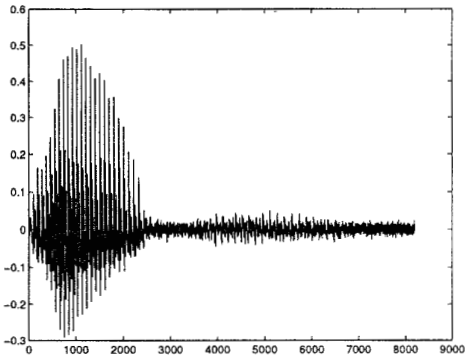


(b)

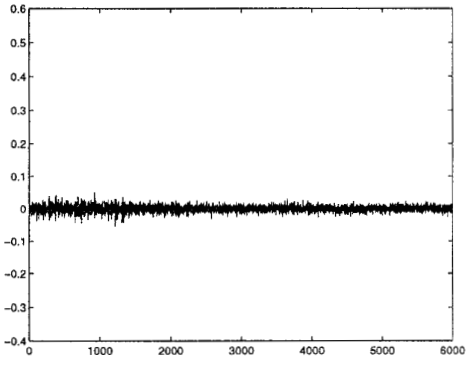
Figure 5: Example of the application of the method: Speech plus White Gaussian Noise. a: original signal, b: contaminated signal

The graphs clearly show a definite improvement of the reproduction of the different signals in each case. The outputs are acoustically close to their original versions. The improvement in SNR versus input SNR is given in figure 4. It can be seen that the proposed method gives good results in very adverse conditions. Note that the SNR displayed is a segmental SNR.

Tests for removing sinusoidal interference from speech were performed. For an input SNR of -3.7 dB, the output SNR was 16.12 dB for a fixed frequency of the sine



(a)



(b)

Figure 6: Example of the application of the method: Speech plus White Gaussian Noise. a: reconstructed signal ,b: reconstructed noise

wave (improvement 19.81 dB) and 12.2 db for a slowly varying one (improvement 15.9 db).

6. Conclusions

A new method to enhance signals, based on source separation techniques is presented. The initial results obtained are quite promising. Several improvements are possible in a variety of directions, for example in using different filters and different objective functions. The method is potentially useful in many applications to other signal processing problems, such as for example Voice Activity Detection. Research is currently under way to explore the fundamental parameters that influence this approach in a decisive manner and to determine the limits of its applicability. Further development of this work is reported in [2].

7. References

[1] A. Cichoki and R. Unbehauen. *Neural Networks for Optimization and Signal Processing.* Wiley, 1993.

[2] N. Doukas, P. Naylor, and T. Stathaki. A single sensor source separation approach to noise reduction. In *CESA 96 IMACS Multiconference*, 1996.

[3] F. Ehrmann, R. Le Bouquin-Jeannes, and G. Faucon. Optimisation of a two-sensor noise reduction technique. *IEEE Signal Processing Letters*, 2(6):108–110, June 1995.

[4] E. Ertin. Lagrange programming neural networks for visual reconstruction. Master's thesis, Imperial College, University of London, September 1993.

[5] C. Jutten and J. Herault. Blind separation of sources, part 1: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1):1–10, July 1991.

[6] K. Matsuoka and M. Kawamoto. A neural net for blind separation of nonstationary signal sources. In *IEEE International Conference on Neural Networks - IEEE World Congress on Artificial Intelligence*, volume 1, pages 221–226, 1994.

[7] P.Common, C. Jutten, and J. Herault. Blind separation of sources, part 2: Problems statement. *Signal Processing*, 24(1):11–20, July 1991.

[8] E. Sorouchyari. Blind separation of sources, part 3: Stability analysis. *Signal Processing*, 24(1):21–30, July 1991.

[9] E. Weinstein, M. Feder, and A. Oppenheim. Multi-channel signal separation by decorrelation. *IEEE Transactions on Speech and Audio Processing*, 1(4):405–413, October 1993.

[10] D. Yellin and E. Weinstein. Criteria for multichannel signal separation. *IEEE Transactions on Signal Processing*, 42(8):2158–68, August 1994.

[11] S. Zhang and A. G. Constantinides. Lagrange programming neural networks. *IEEE Transactions on Circuits and Systems*, 39(7):441–52, July 1992.