4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), April 2003, Nara, Japan

MONAURAL ICA OF WHITE NOISE MIXTURES IS HARD

Lars Kai Hansen and Kaare Brandt Petersen

Informatics and Mathematical Modeling,
Technical University of Denmark,
DK-2800 Lyngby, DENMARK.
*e*mail: lkh,kbp@imm.dtu.dk

ABSTRACT

Separation of monaural linear mixtures of 'white' source signals is fundamentally ill-posed. In some situations it is not possible to find the mixing coefficients for the full 'blind' problem. If the mixing coefficients are known, the structure of the source prior distribution determines the source reconstruction error. If the prior is strongly multi-modal source reconstruction is possible with low error, while source signals from the typical 'long tailed' distributions used in many ICA settings can not be reconstructed. We provide a qualitative discussion of the limits of monaural blind separation of white noise signals and give a set of *no go* cases, finally, we use a so-called Mean Field approach to derive an algorithm for ICA of noisy monaural mixtures with a bi-modal source prior and demonstrate that low error source reconstructions are possible when the bi-modal source is close to binary. This is the first demonstration of blind source separation in noisy monaural mixtures without invoking temporal correlation information.

1. INTRODUCTION

Blind separation of linear signal mixtures is a fascinating problem with numerous applications. The over-complete separation problem consists in separating mixtures of more sources than measurements [8]. The extreme situation arises when the aim is to separate two or more signals mixed linearly into only one channel, and is referred to as *monaural* or single channel separation.

This work was supported by Oticon Fonden and EU project MAPAWAMO

Blind signal separation problem consist two subproblems: 1) Learning the mixing coefficients from a sample of the mixture and 2) Estimating the sources. These two aspects of the problem are often considered together. In fact, in many EM-like algorithms one iterates between solving 1) and 2). But it turns out that in the monaural problem it is sometimes possible to learn the mixing coefficients without being able to fully recover the source signals.

A number of monaural separation schemes have been proposed, all based on the use of temporal structure [4, 13]. In this contribution we will focus on the harder problem of over-complete mixtures of sources with no temporal correlation as in [8].

We show that for source signals with densities belonging to the so-called symmetric stable family it is not possible to learn the mixing coefficients. For heavy tailed signals, in general, it is not possible to recover the sources even if the mixing coefficients are known.

Finally, we can provide a bit of good news: For certain multi-modal source priors (e.g., binary sources) one can in fact easily learn the mixing coefficients and subsequently recover the source signals. In fact we show how to use a Mean Field approach to derive an algorithm for ICA of noisy monaural mixtures with a bi-modal source prior and demonstrate that low error source reconstructions are possible when the bi-modal source is close to binary. This is the first demonstration of blind source separation in noisy monaural mixtures without invoking temporal correlation information. Separation of bi-modal sources is of practical relevance in telecommunications and in functional neuroimaging. In fMRI activation studies the observation is a mixture of hemodynamic fluctuations, including the response of neural tissue to on-off stimuli. Such stimulus induced components can have bi-modal histograms see [12].

2. LEARNING THE MIXING COEFFICIENTS

Let us consider monaural separation of white noise signals, i.e., signals without temporal dependencies and without measurement noise. In particular we study the model,

$$x^n = \mathbf{a} \cdot \mathbf{s}^n, \tag{1}$$

where $\mathbf{s}$ is a vector of D source signals $s_k, k = 1, ..., D$, while $\mathbf{a}$ is a vector of mixing coefficients, and $x^n, n = 1, ..., N$ is the univariate mixture. The sources are independent in time and space,

$$P(\mathbf{s}^1, \mathbf{s}^2, ..., \mathbf{s}^N) = \prod_{n=1}^{N} \prod_{k=1}^{D} P(s_k^n) \tag{2}$$

The blind source separation problem consist of estimating the mixing coefficients and the sources. The posterior distribution of mixing coefficients is given by

$$\begin{aligned} &P(\mathbf{a}|\{x^n\}) \propto P(\mathbf{a}) P(\{x^n\}|\mathbf{a}) \\ &P(\{x^n\}|\mathbf{a}) = \\ &\prod_{n=1}^{N} \int D\mathbf{s}^n \delta\left(x^n - \mathbf{a} \cdot \mathbf{s}^n\right) \prod_{k=1}^{D} P(s_k^n) \end{aligned} \tag{3}$$

In general the likelihood function will be a function of the components of $\mathbf{a}$. If the source distributions are symmetric and identical, the posterior will 'inherit' reflection and permutation symmetries as in conventional ICA problems [7]. Hence, in the case of a one-dimensional mixture of two such sources there will be four degenerate 'peaks' in the likelihood (and in the posterior if the prior $P(\mathbf{a})$ is symmetric).

However, under certain source distributions the posterior can be more degenerate.

2.1. Symmetric stable signals

The class of symmetric α-stable ($S\alpha S$) distributions are heavy-tailed distributions investigated as models of impulsive noise and speech [5, 6]. The densities are denoted $S\alpha S(\mu, \alpha, \gamma)$ where μ is a location parameter, γ is a scale parameter and α is a shape parameter and they are defined by their closedness under addition

$$\begin{aligned} &s_1 \sim S\alpha S(\mu_1, \gamma_1 \alpha,) \wedge \; s_2 \sim S\alpha S(\mu_2, \alpha, \gamma_2) \\ &\Rightarrow \\ &s_1 + s_2 \sim S\alpha S\left(\mu_1 + \mu_2, \alpha, (|\gamma_1|^\alpha + |\gamma_2|^\alpha)^{1/\alpha}\right) \end{aligned}$$

The density functions for $S\alpha S$-variables can only be written in closed form for $\alpha = 2$ and $\alpha = 1$, where we obtain the normal distribution and the Cauchy distribution:

$$P(s|\gamma, \mu) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (s - \mu)^2}. \tag{4}$$

The addition rule implies that information about the mixing coefficients is lost in stable mixtures: The likelihood function for a mixture of two unit-scale stable variables (with mixing coefficients a_1, a_2) is a function of the combination $(|a_1|^\alpha + |a_2|^\alpha)^{1/\alpha}$ only.

2.2. Temporally correlate d gaussian sources

In [1, 4, 13, 14] monaural separation was analyzed in the case of temporally correlated signals, typically speech signals. For temporally correlated and *stationary* signals the matrix

$$\mathbf{C}(\tau) = \langle \mathbf{s}(t) \mathbf{s}^\top (t + \tau) \rangle \tag{5}$$

is a diagonal matrix with elements describing the autocorrelation of the sources at lag τ. Now compute the autocorrelation function for the mixture signal

$$\begin{aligned} \gamma(\tau) &= \langle x(t) x(t + \tau) \rangle \\ &= \mathbf{a}^\top \mathbf{C}(\tau) \mathbf{a} \\ &= \sum_{k=1}^{D} (a_k)^2 C_{k,k}(\tau). \end{aligned} \tag{6}$$

We note that from a measurement of $\gamma(\tau)$ is it *not* possible to identify both $\mathbf{a}$ and the D source functions $C_{k,k}(\tau)$. Hence, we conclude that an ICA approach like the one of Molgedey and Schuster [9] is infeasible for monaural separation. Other schemes that either makes use of higher order moments or non-stationarity may still succeed in separating monaural mixtures. Since our main objective here is to discuss white noise mixtures, we will not pursue this topic further, instead we refer to [1, 4, 13, 14] for examples.

3. GEOMETRY OF MONAURAL WHITE NOISE SOURCE RECONSTRUCTION

The above degeneracy is specific for stable distributions and in general we expect that the mixing coefficients are learnable from samples of a monaural mixture. However, then the second problem arises: How accurate can the source signals be recovered?.

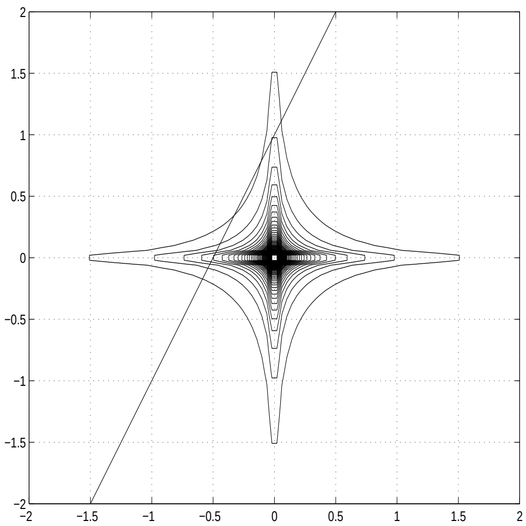

Fig. 1. Contours of the joint distribution of two sparse source signals. An observation $x = a_1 s_1 + a_2 s_2$ constrains the possible sources to a line. The 'true' source are drawn from the prior, hence, the line passes trough at least one arm of the cross formed by a contour line. In the two-dimensional case we see that a line determined by mixing coefficients in general position always intersects the cross in *two places*. The actual values of the mixing coefficients determine which of the two peaks in along the line is maximal, not the measurement x, hence, we conclude that any source reconstruction will have a large error for such sparse source mixing problems.

To illustrate the ill-posed nature of the source identification problem we first note that if the source are i.i.d. we are faced with a sequence of independent problems of the form: Find $\mathbf{s}$ from a single measurement of x, given $\mathbf{a}$. The constraint

$$x = \mathbf{a} \cdot \mathbf{s}, \tag{7}$$

defines a $D-1$ dimensional hyperplane of solutions in the D-dimensional source space. The posterior distribution of the source signal is given by

$$P(\mathbf{s}|x, \mathbf{a}) = \frac{\delta\left(x - \mathbf{a} \cdot \mathbf{s}\right) P(\mathbf{s})}{P(x|\mathbf{a})} \tag{8}$$

Let us be specific and consider first two signals mixed in one receiving channel. In Figure 1 we show the geometry for a sparse (Cauchy) prior. The density contours form a 'cross' strongly peaked along the axes. For two independent sparse signals, there is low probability that they are both large at the same time. The constraint $x = a_1 s_1 + a_2 s_2$ forms a single linear relation between the component of $\mathbf{s}$, hence, the source prior determines which solution we choose, say in a maximum posteriory approach. Clearly, for a sparse prior we see that for all $\mathbf{a}$ there are two 'peaks', along the line. Since the source signals are drawn from the prior we face a situation where we need to choose between two competing solutions: The 'true' solution and a 'false' solution. Note, it is the coefficients $\mathbf{a}$ that determines which of the two peaks have the maximum density value, not x, hence we can expect a high reconstruction error rate for sparse priors.

Now why does the same argument not prevent the solution of the mixing of three signals in the binaural (two channels) case, as e.g., considered in [8, 7]? The scenario is depicted in Figure 2, the two measurements form two linear contraints leading to a 1D linear manifold of solutions within which the prior determines the posterior. Since the true sources are drawn from the prior, the line will typically intersect one of the six 'arms' of the density, but unlike the previous case, there is limited probability of a second "false" solution, corresponding to the line also intersecting another arm! It is easy to generalize this argument to higher dimensional mixtures in two channels. Source reconstruction with sparse sources is fundamentally ill-posed in one-dimensional mixtures and well-posed in typical two or higher dimensional channel case.

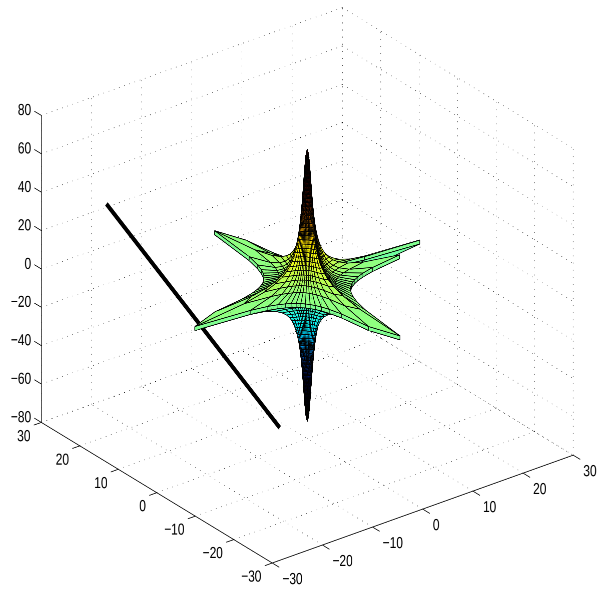

Fig. 2. Iso-surface for the joint prior distribution of three sparse sources. If we have access to two observation channels (microphones) each observation $x_j = a_{j1}s_1 + a_{j2}s_2 + a_{j3}s_3,\ j = 1, 2$ will constrain the possible sources to a plane, hence, the likelihood function contrains the posterior to the intersection of the two planes forming a linear manifold of possible sources. Since the 'true' sources are drawn from the prior the line will pass trough at least one arm of the star shaped iso-surface. However, for mixing coefficients in general position it is not likely that the line passes through two arms. Hence, the posterior is strongly peaked in the vicinity of the 'true' sources.

4. RECONSTRUCTION OF BI-MODAL SOURCE SIGNALS

The geometric origin of the sparse source reconstruction problem suggests that sources with a multi-modal prior will reconstruct better. Figure 3 illustrates the situation for a quasi-binary prior formed by a mixture of two gaussians. The prior has high density values in the neighborhood of the four source signal combinations $\{(1, 1), (1, -1), (-1, 1), (-1, -1)\}$. For a typical set of mixing coefficients $\mathbf{a}$ the linear constraint imposed by the likelihood leads to a line of solutions intersecting one and only one of the four 'prior peaks', hence strongly selecting the correct solution.

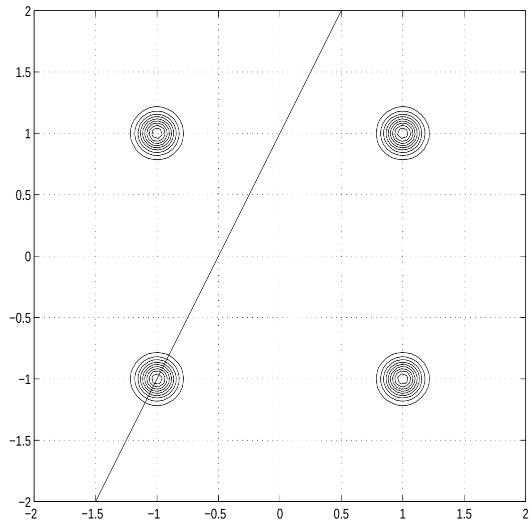

Fig. 3. Contours of the joint density of two source signals with a 'mixture of two gaussians' priors. The narrow widths create a 'quasi-binary' scenario. The observation of $x = a_1s_1 + a_2s_2$ again constrains the posterior to the intersection of the prior and the 'likelihood' line. The line only intersects one of the gaussians, hence the posterior is stongly peaked a the 'true' source configuration $(s_1, s_2) = (1, -1)$. If we observe a noisy mixture $x = a_1s_1 + a_2s_2 + \nu$, the line is broadened to to standard deviation of the noise process ν.

4.1. Mean Field approach for noisy ICA

Using the mean-field approach of [3] we have investigated the reconstruction of source signals from a *noisy* mixture $x = a_1s_1 + a_2s_2 + \nu$, where the sources both have a bi-modal prior density composed by a mixture of two gaussians as in Figure 3. The two gaussians are centered at $+/-1$ and have width σ^2

We assume that the additive noise is gaussian, $\nu \sim \mathcal{N}(0, \eta^2)$. The probability density of the measured data conditioned on the parameters, i.e., the likelihood function, is obtained by noting that $\nu^n = x^n - \mathbf{a} \cdot \mathbf{s}^n$, hence, $P(x^n|\mathbf{a}, \mathbf{s}) = \mathcal{N}(\mathbf{a}\cdot\mathbf{s}^n, \eta^2)$. Bayes theorem then relates the source priors and the source posterior as

$$P(\mathbf{S}|\mathbf{a}, \eta^2, \mathbf{x}) = \frac{P(\mathbf{x}|\mathbf{a}, \eta^2, \mathbf{S})P(\mathbf{S})}{P(\mathbf{x}|\mathbf{a}, \eta^2)}$$

where capital $\mathbf{S}$ is defined by $\mathbf{S}_i^n \equiv (\mathbf{s}^n)_i$. Maximizing the source posterior with respect to $\mathbf{a}$ and η^2 leads to the following estimators

$$\begin{aligned}\hat{\mathbf{a}} &= X\langle\mathbf{S}\rangle^T\langle\mathbf{SS}^T\rangle^{-1}\\ \widehat{\eta^2} &= \frac{1}{N}\langle(x-\mathbf{aS})(x-\mathbf{aS})^T\rangle\end{aligned}$$

The bracket $< \ldots >$ denotes expectation with respect to the source posterior. These expressions depend on $\langle\mathbf{S}\rangle$ and $\langle\mathbf{SS}^T\rangle$ approximately determined through Mean Field Theory (MFT) using **a** and η^2 [3]. In MFT the true source posterior is approximated by a factorized density function. The interaction not represented directly in the factorized model is compensated by a linear term - the so-called 'external field', in particular we introduce two auxiliary variables: The interaction $\mathbf{J} = \eta^{-2}\mathbf{a}^T\mathbf{a}$ and the external field $\mathbf{h} = \eta^{-2}\mathbf{a}^T\mathbf{x}$. From these we define:

$$\lambda_i^n = \lambda_i = J_{ii} \qquad (9)$$

$$\gamma_i^n = (\mathbf{h} - (\mathbf{J} - \mathrm{diag}\mathbf{J})\langle\mathbf{s}\rangle)_i^n\,, \qquad (10)$$

and the implicit non-linear equations for the sufficient statistics then read,

$$\begin{aligned}&\langle\mathbf{S}\rangle_{in} = \frac{1}{\lambda_i\sigma^2+1}\Big(\gamma_i^n\sigma^2 + \tanh\Big[\frac{\gamma_i^n}{\lambda_i^n\sigma^2+1}\Big]\Big)\\ &\langle\mathbf{SS}^T\rangle_{ij} =\\ &\sum_n\Big[\frac{\delta_{ij}}{(\lambda_i\sigma^2+1)^2}\Big\{1+\sigma^2+\sigma^4(\lambda_i+(\gamma_i^n)^2)\\ &+2\gamma\sigma^2\tanh\left(\frac{\gamma_i^n}{\lambda_i\sigma^2+1}\right)\Big\} + (1-\delta_{ij})\langle\mathbf{S}\rangle_i^n\langle\mathbf{S}\rangle_j^n\Big]\end{aligned}$$

This way we arrive at a two-step algorithm. The first step in which **a** and η^2 are estimated using $\langle\mathbf{S}\rangle$ and $\langle\mathbf{SS}^T\rangle$ and a second step in which $\langle\mathbf{S}\rangle$ and $\langle\mathbf{SS}^T\rangle$ are estimated using **a** and η^2. We use the posterior means as our posterior source estimate, i.e. $\widehat{\mathbf{S}} = \langle\mathbf{S}\rangle$. In the limit $\sigma^2 \to 0$, i.e. the binary case, the posterior mean values minimize the mean bit error rate.

We set up a simulation to demonstrate that bi-modal monaural mixtures can be separated. The sources are bi-gaussian with a variable width, and furthermore we vary the additive noise level. The reconstruction error of the recovered source signals is measured as

$$E = \frac{1}{4\langle s_*^2\rangle}\langle(s_* - \widehat{s})^2\rangle \qquad (11)$$

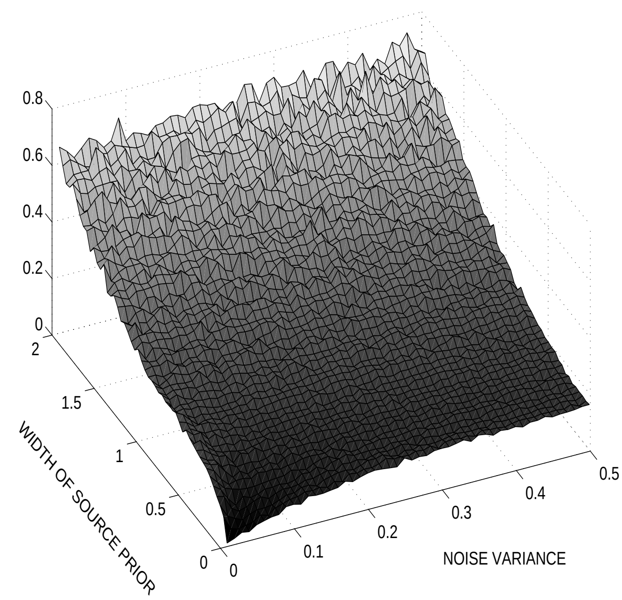

Fig. 4. The scaled mean square source reconstruction error vs. the width of the bi-gaussian source distributions and the noise variance. The gaussians were centered at $+/-1$ and the widths were identical and varied from zero (binary sources) to $\sigma = 2$. The mixing coefficients were $(a_1, a_2) = (\cos(\pi/8), sin(\pi/8))$. The zero error rate in the limit of low noise and narrow source priors is achieved for mixing coefficients in 'general position'. The sources were estimated using the mean field approach presented [3]

where $s_*, \widehat{s}$ are the 'true' and reconstructed sources respectively. The scaling assures that the error becomes equivalent to the *bit error rate* for binary sources. The reconstruction error is a function of the width of the source distribution and the noise variance as shown in Figure 4. For mixing coefficients in 'general position' we achieve perfect reconstruction in the binary, noise free limit.

5. CONCLUSION

In this paper we have discussed the problem of blind source separation based on a single white noise mixture signal. We have shown that in some situations it is impossible to identify the mixing coefficients, namely for α-stable mixtures. In general it is possible to identify the mixing coefficients, however, it may still be infeasible to reconstruct the source signals. A qualitative analysis lead us to conclude that heavy tail signals can only be reconstructed with a high error rate, hence, monaural mixtures are harder than two or more dimensional mixtures for simple topological reasons.

For signals with a multi-modal source distribution, however, source reconstruction is feasible and in the noise free limit, and in the specific case of binary sources the reconstruction can take place without errors.

Acknowledgment

We thank Preben Kidmose for discussions on stable signals, and the ISP-group at DTU for discussion of ICA problems.

6. REFERENCES

[1] R. Balan & and J.A. Rosca: *AR processes and sources can be reconstructed from degenerate mixtures*. Proceedings of the First International Conference on Independent Component Analysis and Blind Source Separation ICA'99, Aussois, France. pages 467-47 (1999).

[2] A. Bell & T.J. Sejnowski: *An Information-Maximization Approach to Blind Separation and Blind Deconvolution* Neural Computation **7**, 1129-59 (1995).

[3] P. Højen-Sørensen, O. Winther, and Lars Kai Hansen: Mean Field Approaches to Independent Component Analysis. Neural Computation **14** 889-918 (2002).

[4] G-J. Jang, T-W. Lee, Y-H. Oh: *Blind separation of single channel mixture using ICA basis functions*. 3rd International Conference on Independent Component Analysis and Blind Signal Separation. Dec. 9-12, 2001, San Diego, California, USA Eds: T-W. Lee et al. pp. 595-600, (2001).

[5] G.B. Georgiou, P. Tsakalides, and C. Kyriakis: *Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise*. IEEE Transactions on Multimedia **1** 291-301 (1999).

[6] P. Kidmose: *Alpha-stable distributions in signal processing of audio signals*. In proceedings of the 41'st Conference on Simulation and Modelling, Scandinavian Simulation Society 87-94 (2000).

[7] T.-W. Lee *Independent Component Analysis: Theory and Applications*. Kluwer Academic Publ. Bostin (1998).

[8] M.S. Lewicki and T.J. Sejnowski: *Learning overcomplete representations* Neural Computation **12** 337-65 (2000).

[9] L. Molgedey & H. Schuster: *Separation of Independent Signals using Time-Delayed Correlations* Physical Review Letters **72** 3634-37 (1994).

[10] E. Moulines, J.-F. Cardoso, E. Gassiat. *Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models* Proc. ICASSP'97 Munich, vol. 5, pp. 3617-20 (1997).

[11] B.A. Olshausen: *Learning linear, sparse, factorial codes* A.I. Memo 1580, Massachusetts Institute of Technology (1996).

[12] K. Petersen, L.K. Hansen, T. Kolenda, E. Rostrup, S. Strother. *On the Independent Components of Functional Neuroimages*. In Proceedings of ICA-2000. Eds. P. Pajunen and J Karhunen, 615-620 (2000).

[13] S.T. Roweis: *One microphone source separation* in Advances in Neural Information Processing Systems **13** 793-799 (2001).

[14] E. Wan and A. Nelson: *Neural dual extended Kalman filtering: applications in speech enhancement and monaural blind signal separation* In Proceedings Neural Networks for Signal Processing Workshop. IEEE, 1997.