Digital Signal Processing 11, 204–221 (2001)
doi:10.1006/dspr.2001.0398, available online at http://www.idealibrary.com on IDEAL®

Hyper Radial Basis Function Neural Networks for Interference Cancellation with Nonlinear Processing of Reference Signal

Sergiy A. Vorobyov and Andrzej Cichocki[1]

Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN (The Institute of Physical and Chemical Research), 2-1, Hirosawa, Wako-shi, Saitama, 351-0198, Japan
E-mail: cia@brain.riken.go.jp

Vorobyov, S. A., and Cichocki, A., Hyper Radial Basis Function Neural Networks for Interference Cancellation with Nonlinear Processing of Reference Signal, *Digital Signal Processing* **11** (2001) 204–221.

Efficient interference cancellation often requires nonlinear processing of a reference signal. In this paper, hyper radial basis function (HRBF) neural networks for adaptive interference cancellation is developed. We show that the HRBF networks, with an appropriate learning algorithm, is able to approximate the interference signal more efficiently than standard radial basis function (RBF) networks. The HRBF network-based canceller achieves better results for interference cancellation. This is due to the capabilities of the HRBF networks to approximate arbitrary multidimensional nonlinear functions and better flexibility in comparison to RBF networks. Simulation examples and comparisons to the FIR-based linear canceller and the RBFN-based canceller demonstrate the usefulness and effectiveness of the HRBFN based canceller. © 2001 Academic Press

Key Words: interference and noise cancellation; hyper radial basis functions; Green's functions; nonlinear mapping; neural networks; learning; Manhattan algorithm.

1. INTRODUCTION AND PROBLEM FORMULATION

Interference cancellation refers to the minimization or cancellation of an interference in an observed waveform, based on an estimate of the interference signal that is a mapping of a separate signal called the reference signal. Linear adaptive filters with the mean squared error criterion have been usually used for interference and noise cancellation [1, 2]. However, for many applications such as communication, geophysics, and especially biomedical signal process-

[1] A. Cichocki is on leave from the Warsaw University of Technology, Poland.

1051-2004/01 $35.00

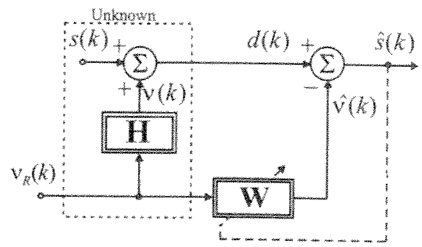

FIG. 1. Interference cancellation principle.

ing, linear filter structures cannot in general implement optimum interference and noise cancellers in a minimum mean squared error sense [3]. Using linear finite impulse response (FIR) or infinite impulse response filters we often cannot achieve acceptable levels of interference and noise cancellation because these techniques are not able to satisfactorily approximate an unknown deterministic, but nonlinear mapping between the available reference and unknown interference signals. For many applications, the reference and interference signals are usually primarily related through a deterministic nonlinear dynamic mapping (nonlinear filter) that is shown in Fig. 1 corresponding to block **H**. Thus, it is reasonable to try to find an optimal cancellation system using nonlinear adaptive processing models and associated training methods. Methods based on nonlinear processing of the interference signal could be called nonlinear noise cancellers.

In [3–5] Cha *et al.* and Lu *et al.* have demonstrated that accurate channel equalization and cancellation of interference and noise can be achieved with the use of radial basis function (RBF) neural networks and generalized radial basis function networks. However, we are going to show that it is possible to achieve better results using neural networks with more general and flexible architectures. Alternative approaches for interference cancellation problem have been studied in [6, 7]. In [6] an adaptive nonlinear filter modeled by a single layer perceptron and a Pth-order Volterra filter have been used for estimation of the nonlinear mapping between reference and interference signals. In [7], third-order moments were used for the same purpose. However, it was shown in [3] that RBF network is more accurate for the problem of interference cancellation than multilayer perceptron and some other standard methods including linear filters.

In this paper we study hyper radial basis function (HRBF) neural networks with all parameters fully adjusted by using an associated adaptive online learning algorithm for interference cancellation.

The basic problem is illustrated in detail in Fig. 1. The model of a corrupted signal $d(k)$ is

$$d(k) = s(k) + \nu(k), \tag{1}$$

where $s(k)$ is an unknown primary signal and $\nu(k)$ is the undesired interference or noise signal and $k = 1, 2, \ldots, n$ is discrete time. The reference noise $\nu_R(k)$ is assumed to be available. Reference noise is related to the interference signal $\nu(k)$ via an unknown nonlinear operator **H** (i.e., unknown nonlinear feedforward filter). Signals $s(k)$, $\nu(k)$, and $\nu_R(k)$ are all assumed to be stationary random processes with zero mean. In practice they can be also deterministic. The solution consists of identifying the unknown nonlinear operator **H** by a nonlinear filter (neural network) **W**. The canceller output $\hat{s}(k)$, which is an estimate of the desired signal $s(k)$, is then given by

$$\hat{s}(k) = d(k) - \hat{\nu}(k) = s(k) + \nu(k) - \hat{\nu}(k). \tag{2}$$

2. MAIN RESULTS

2.1. Fundamentals

A hyper basis function (HBF) neural network has been first introduced by Poggio and Girosi [8]. The main idea of a HBF network is to consider the mapping to be approximated (in our problem it is nonlinear dynamic mapping **W**) by a sum of several functions, each one with its own prior. The corresponding regularization principle then yields a superposition of Green's functions, in particular, generalized Gaussian basis functions with different widths. In detail, the operator **W** is regarded as the sum of M components f_m, $m = 1, 2, \ldots, M$, each having a different prior probability. The functional $L(\nu)$ to minimize is defined as

$$L(\nu) = \sum_{k=1}^{n} \left(\sum_{m=1}^{M} f_m(\nu_R(k)) - \nu(k) \right)^2 + \sum_{m=1}^{M} \gamma_m \| P_m f_m \|^2, \tag{3}$$

where P_m, $m = 1, 2, \ldots, M$, are constraint operators (stabilizers) in Tikhonov's regularization theory [9] and γ_m, $m = 1, 2, \ldots, M$, are positive real numbers (regularization parameters).

The approximate solution to the minimization problem defined by (3) is

$$\hat{\nu} = \sum_{m=1}^{M} \sum_{j=1}^{n} w_j^m G_j^m(\overline{\mathbf{v}}_R, \mathbf{q}_j^m), \tag{4}$$

where w_j^m are the weight parameters and $G_j^m(\overline{\mathbf{v}}_R, \mathbf{q}_j^m)$, $j = 1, 2, \ldots, n$, $m = 1, 2, \ldots, M$, are Green's functions. See [8] for details.

2.2. Practical Realization

To make the model (4) practical and as simple as possible we are interested in the special case of Green's functions. In particular, the appropriate choice of stabilizers, which are rotationally and translationally invariant, (4) leads to radial basis functions of multiple scales. Choosing, for example, a set of M stabilizers whose Green's functions are Gaussian ($G_j^m(\overline{\mathbf{v}}_R, \mathbf{q}_j^m) := \Phi_j[\rho_j(\overline{\mathbf{v}}_R, \mathbf{q}_j^m)] = \Phi_j(\rho_j)$) and setting $w_j^1 = \cdots = w_j^M = w_j$ we can write a simple form of hyper radial basis function network as a two-layer neural network in which the hidden

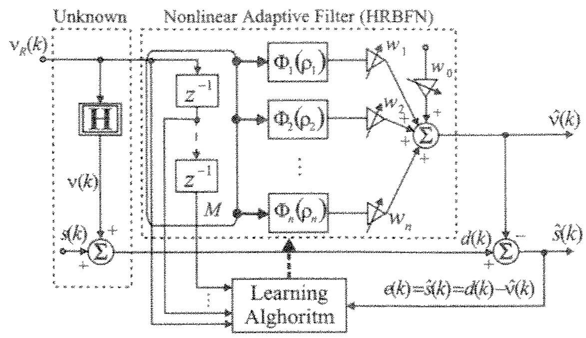

FIG. 2. Model of HRBF network-based, nonlinear, interference cancellation system.

layer performs an adaptive nonlinear transformation with adjustable weight and center parameters. This allows an M dimensional input space $\bar{\boldsymbol{v}}_R(k) = (v_R(k), v_R(k-1), \dots, v_R(k-M+1))^\top$ to be mapped to the one-dimensional output space $\hat{v}(k)$: $R^M \to R$ as

$$\hat{v}(k) = \sum_{j=1}^{n} w_j(k)\Phi_j[\rho_j(\bar{\boldsymbol{v}}_R, \mathbf{q}_j^m)] = \sum_{j=1}^{n} w_j(k)\Phi_j(\rho_j) = \mathbf{w}^\top \boldsymbol{\Phi}(\rho), \tag{5}$$

where $\mathbf{w}(k) = (w_1(k), w_2(k), \dots, w_n(k))^\top \in R^n$ is the vector of weights, $\boldsymbol{\Phi}(\rho) = (\Phi_1(\rho_1), \Phi_2(\rho_2), \dots, \Phi_n(\rho_n))^\top$ is the vector with elements

$$\Phi_j(\rho_j) = \frac{1}{2}\exp\left(-\frac{1}{2}\rho_j^2\right) \tag{6}$$

and

$$\begin{aligned} \rho_j^2 &= \rho_j^2(\bar{\boldsymbol{v}}_R(k), \mathbf{Q}_j) = \rho_j^2(\bar{\boldsymbol{v}}_R(k), \mathbf{q}_j^m) = \|\mathbf{Q}_j(\bar{\boldsymbol{v}}_R(k) - \mathbf{c}_j)\|^2 \\ &= (\bar{\boldsymbol{v}}_R(k) - \mathbf{c}_j)^\top \mathbf{Q}_j^\top \mathbf{Q}_j(\bar{\boldsymbol{v}}_R(k) - \mathbf{c}_j) \\ &= \sum_{m=1}^{M}\left[\sum_{\varsigma=1}^{M} q_j^{m\varsigma}(v_R(k-\varsigma+1) - c_{j\varsigma})\right]^2, \end{aligned} \tag{7}$$

$j = 1, 2, \dots, N$ (N is a number of neurons in hidden layer), with adaptive centers $\mathbf{c}_j = (c_{j1}, \dots, c_{jM})^\top$ and matrices of widths $\mathbf{Q}_j$ (see also Fig. 2).

Let us note that if $(M \times M)$ positive definite matrix $\mathbf{Q}_j^\top \mathbf{Q}_j$ is reduced to a diagonal matrix $\text{diag}(\sigma_{j1}^{-2}, \dots, \sigma_{jM}^{-2})$ then the HRBF network is simplified (degenerated) to the standard RBF neural network.

2.3. Learning Procedure

To adjust the HRBF network we must estimate all free parameters w_j, $\mathbf{c}_j$, and $\mathbf{Q}_j$, $j = 1, 2, \dots, n$, in (5) and (7).

There is no explicit training set of input–output examples for learning of the HRBF network in reference to the interference cancellation problem. Hence, the objective becomes the minimization of the cancellation system output power, which can be written as

$$\hat{s}(k) = e(k) = d(k) - \hat{\nu}(k) = d(k) - \frac{1}{2}\sum_{j=1}^{n} w_j \exp\left(-\frac{1}{2}\|\mathbf{Q}_j(\bar{\boldsymbol{\nu}}_R(k) - \mathbf{c}_j)\|^2\right). \quad (8)$$

This is equivalent to the minimization of mean squared error between $\nu(k)$ and $\hat{\nu}(k)$ under assumption that $s(k)$ is uncorrelated with both $\nu(k)$ and $\hat{\nu}(k)$. Thus, we can use $d(k) = s(k) + \nu(k)$ as the "desired output" and $\hat{\nu}(k)$ as the actual output for learning of the HRBF network. The "error" $e(k)$, whose power is to be minimized, is $s(k) + \nu(k) - \hat{\nu}(k)$, i.e., output of the canceller.

Now, our objective is to estimate the set of all free parameters $\Theta = \{\mathbf{w}, \{\mathbf{c}_j\}, \{\mathbf{Q}_j\}\}$ of the HRBF network using the standard power function

$$J(\Theta) = \frac{1}{2}e^2(k) = \frac{1}{2}\hat{s}^2(k), \quad (9)$$

where the error $e(k)$ is calculated using (8).

Applying the gradient descent approach, we obtain the following algorithm for online learning of the HRBF network

$$\begin{aligned} \mathbf{w}(k) &= \mathbf{w}(k-1) - \eta_{\mathbf{w}}(k)\nabla_{\mathbf{w}} J(\Theta), \\ \mathbf{Q}_j(k) &= \mathbf{Q}_j(k-1) - \eta_{\mathbf{Q}_j}(k)\frac{\partial J(\Theta)}{\partial \mathbf{Q}_j}, \\ \mathbf{c}_j(k) &= \mathbf{c}_j(k-1) - \eta_{\mathbf{c}_j}(k)\nabla_{\mathbf{c}_j} J(\Theta), \end{aligned} \quad (10)$$

where $J(\Theta) = J(\mathbf{w}, \{\mathbf{c}_j\}, \{\mathbf{Q}_j\})$, $j = 1, 2, \ldots, N$, and $\eta_{\mathbf{w}}(k)$, $\eta_{\mathbf{Q}_j}(k)$, $\eta_{\mathbf{c}_j}(k)$ are learning rates, which can be fixed for the simplest case.

The results of the gradient components calculation for all free parameters $\Theta = \{\mathbf{w}, \{\mathbf{c}_j\}, \{\mathbf{Q}_j\}\}$ are

$$\begin{aligned} \nabla_{\mathbf{w}} J(\Theta) &= -\boldsymbol{\Phi}(\rho)(d(k) - \hat{\nu}(k)), \\ \frac{\partial J(\Theta)}{\partial \mathbf{Q}_j} &= \delta_j(k)\mathbf{Q}_j(\bar{\boldsymbol{\nu}}_R(k) - \mathbf{c}_j)(\bar{\boldsymbol{\nu}}_R(k) - \mathbf{c}_j)^\top, \\ \nabla_{\mathbf{c}_j} J(\Theta) &= -\delta_j(k)\mathbf{Q}_j^\top \mathbf{Q}_j(\bar{\boldsymbol{\nu}}_R(k) - \mathbf{c}_j), \end{aligned} \quad (11)$$

where

$$\begin{aligned} \delta_j(k) &= \frac{1}{2}w_j \exp\left(-\frac{1}{2}(\bar{\boldsymbol{\nu}}_R(k) - \mathbf{c}_j)^\top \mathbf{Q}_j^\top \mathbf{Q}_j(\bar{\boldsymbol{\nu}}_R(k) - \mathbf{c}_j)\right)(d(k) - \hat{\nu}(k)) \\ &= w_j \boldsymbol{\Phi}_j(\rho_j)\hat{s}(k). \end{aligned} \quad (12)$$

Substituting (11) into (10) results in a full online learning procedure for the HRBF network-based canceller.

However, it should be noted that the power function (9) may have many local minima. To avoid stacking in local minima, we apply the Manhattan learning algorithm of the form [10, 11]. Hence, procedure (10) can be changed to

$$\mathbf{w}(k) = \mathbf{w}(k-1) - \begin{cases} \eta_{\mathbf{w}}(k)\,\mathrm{sign}(\nabla_{\mathbf{w}} J(\Theta)), & \text{if} \quad \beta_{\mathbf{w}}(k) \geq \mathbf{0}_{1\times n}, \\ r_{\mathbf{w}}(\mathbf{w}(k-1) - \mathbf{w}(k-2)), & \text{otherwise}, \end{cases}$$

$$\mathbf{Q}_j(k) = \mathbf{Q}_j(k-1) - \begin{cases} \eta_{\mathbf{Q}_j}(k)\,\mathrm{sign}\left(\dfrac{\partial J(\Theta)}{\partial \mathbf{Q}_j}\right), & \text{if} \quad \beta_{\mathbf{Q}_j}(k) \geq \mathbf{0}_{M\times M}, \\ r_{\mathbf{Q}_j}(\mathbf{Q}_j(k-1) - \mathbf{Q}_j(k-2)), & \text{otherwise}, \end{cases} \tag{13}$$

$$\mathbf{c}_j(k) = \mathbf{c}_j(k-1) - \begin{cases} \eta_{\mathbf{c}_j}(k)\,\mathrm{sign}(\nabla_{\mathbf{c}_j} J(\Theta)), & \text{if} \quad \beta_{\mathbf{c}_j}(k) \geq \mathbf{0}_{1\times M}, \\ r_{\mathbf{c}_j}(\mathbf{c}_j(k-1) - \mathbf{c}_j(k-2)), & \text{otherwise}. \end{cases}$$

where $j = 1, 2, \ldots, N$, $\beta_{\mathbf{w}}(k) = [\nabla_{\mathbf{w}} J(\Theta(k))] \circ [\nabla_{\mathbf{w}} J(\Theta(k-1))]$, $\beta_{\mathbf{Q}_j}(k) = [\partial J(\Theta(k))/\partial \mathbf{Q}_j] \circ [\partial J(\Theta(k-1))/\partial \mathbf{Q}_j]$, $\beta_{\mathbf{c}_j}(k) = [\nabla_{\mathbf{c}_j} J(\Theta(k))] \circ [\nabla_{\mathbf{c}_j} J(\Theta(k-1))]$, $\circ$ is Hadamard product ($A \circ B = [a_{ij} b_{ij}]$) symbol, $\mathbf{0}_{m\times n}$ is $(m \times n)$ zeros matrix, $0 \leq r_{\mathbf{w}} \leq 1$, $0 \leq r_{\mathbf{Q}_j} \leq 1$, $0 \leq r_{\mathbf{c}_j} \leq 1$ are some values (typically equal 0.2 to 0.5) and the gradient components are calculated using (11). This adaptive procedure can work also in a slowly time-varying environment.

2.4. Learning Rates Adaptation

In order to improve the convergence speed of the procedure (13), the learning rates $\eta_{\mathbf{w}}(k)$, $\eta_{\mathbf{Q}_j}(k)$, $\eta_{\mathbf{c}_j}(k)$ should also be optimally tuned. For this purpose we may use heuristics, proposed by Cichocki and Unbehauen [10] and Mazurek *et al.* [11] and summarized as follows. Each weight has its own learning rate. The learning rates are adaptively adjusted during the learning process on the basis of gradient information of the power function. When the gradient component has the same sign for several iterations, the corresponding learning rate is increased by some constant. When the gradient component alternates (flips sign) for several consecutive time steps, the corresponding learning rate is decreased exponentially to allow rapid decay when necessary. Generally speaking, it is possible to show that some of these heuristics have rigorous fundamentals and are simply modifications of nonparametric Mann–Whitney and Kolmogorov–Smirnov criteria [13, 14].

It is possible to design a number of slightly different procedures for learning rate adaptation using modifications of the previously mentioned nonparametric criteria. However, in order to make learning of the HRBF network-based canceller as simple as possible it is desirable and enough to construct some very simple procedure. The simplest one is if we take into account only the signs of the two last gradient components. Jacobs' procedure, based on this simple principle and can be written (for learning rate $\eta_{\mathbf{w}}(k)$) in the form

$$\eta_{\mathbf{w}}(k) = \begin{cases} \eta_{\mathbf{w}}(k-1) + \Delta\eta_{\mathbf{w}}^{\mathrm{add}}, & \text{if } \max\{[\overline{\beta}_{\mathbf{w}}(k)]_j\} > 0, \\ \Delta\eta_{\mathbf{w}}^{\mathrm{mul}}\eta_{\mathbf{w}}(k-1), & \text{if } \max\{[\overline{\beta}_{\mathbf{w}}(k)]_j\} < 0, \\ 0, & \text{otherwise}, \end{cases} \tag{14}$$

where $\Delta\eta_{\mathbf{w}}^{\text{add}}$ is a parameter for an additive increase and $\Delta\eta_{\mathbf{w}}^{\text{mul}}$ is a parameter for a multiplicative (exponential) decrease of the learning rate. Typically $10^{-4} \le \Delta\eta_{\mathbf{w}}^{\text{add}} \le 10^{-1}$ and $0.5 \le \Delta\eta_{\mathbf{w}}^{\text{mul}} \le 0.9$, $\overline{\beta}_{\mathbf{w}}(k) = [\nabla_{\mathbf{w}} J(\Theta(k))] \circ [(1-\upsilon)\nabla_{\mathbf{w}} J(\Theta(k)) + \upsilon\nabla_{\mathbf{w}} J(\Theta(k-1))]$, υ is the momentum parameter $(0 \le \upsilon \le 1)$, $[\overline{\beta}_{\mathbf{w}}(k)]_j$ is jth element of the vector $\overline{\beta}_{\mathbf{w}}(k)$, and $\max\{[\overline{\beta}_{\mathbf{w}}(k)]_j\}$ is the maximum element of the vector $\overline{\beta}_{\mathbf{w}}(k)$.

The procedures for the adaptation of learning rates $\eta_{\mathbf{Q}_j}(k)$ and $\eta_{\mathbf{c}_j}(k)$ are the same and we will not write them explicitly here.

In procedure (13) we have $2n+1$ different learning rates. We can expect that adaptation properties of the learning algorithm (13) might be improved if we would introduce a learning rate for each element of the vectors $\mathbf{w}(k)$, $\mathbf{c}_j(k)$ and the matrix $\mathbf{Q}_j(k)$, $j = 1, 2, \ldots, n$. However, then we would have $n + nM + nM^2$ learning rates to adjust. This results in an increase of computational complexity, which is not attractive. We will show in the simulation section that excellent results can be achieved using only $2n+1$ learning rates.

3. SIMULATION AND COMPARISON RESULTS

To demonstrate the performance of the HRBF network-based canceller we present some simulation results. In order to make some comparison to the previously proposed cancellers we use the following well-known models:

(i) the linear, FIR filter-based canceller with recursive least squares learning algorithm [2];

(ii) the Gaussian, RBF network-based canceller that described as

$$\hat{\upsilon}(k) = \sum_{i=1}^{n} w_i \exp\left(-\frac{\|\overline{\mathbf{v}}_R(k) - \mathbf{c}_i\|^2}{\sigma_i^2}\right) \tag{15}$$

(here σ_i are width parameters) with stochastic gradient learning algorithm [3].

Performance of the interference cancellers is measured by the normalized mean squared error (NMSE) defined as

$$\text{NMSE} = \frac{E\{(\hat{s}(k) - s(k))^2\}}{E\{\upsilon^2(k)\}} = \frac{\sum_{k=1}^{N}(\hat{s}(k) - s(k))^2}{\sum_{k=1}^{N}\upsilon^2(k)},$$

where N is the number of iterations.

The NMSE performance is evaluated on a separate test set of 5000 samples measured at intervals of 500 samples during training.

Following the traditional style of notation, we will denote by HRBFN(M,n) the HRBF network-based canceller with input dimension M and number of hidden units by n. Likewise, RBFN(M,n) and FIR(M) will refer to RBF, network-based interference canceller and a linear, FIR filter-based canceller with M taps, respectively.

Initialization conditions are as follows. The weights $\{w_j\}$, $j = 1, 2, \ldots, n$, are all initialized to small random values in the rang $[-0.1, 0.1]$. The initial centers of the RBF network are determined by a K-means clustering using the first 100 samples of the input data [15]. The initial width parameters σ_j are set to the average of the M nearest-neighbor distances among the initialized centers. It was observed before [3] and we have found in our simulations that the performance of the RBF network-based canceller was rather robust to variations over a significant range of values for the initial spread parameters. For the HRBF network, the initial conditions are the same as for the RBF network; i.e., we use the same initial centers. Initial elements of matrix $\mathbf{Q}_j$ are $q_j^{mm} = \sigma_j$, and $q_j^{m\varsigma} = 0$ for all $m \neq \varsigma$. Then the RBF and HRBF networks are equivalent before learning.

It is well known that feedforward networks are universal approximators [16–18]. Hence, the study of different classes of nonlinearity between the reference signal and interference in our application seems not to be necessary. However, theoretical results of the paper [18] are valid if the network has an infinite number of neurons. Thus, the performance of such methods, depending on the number of neurons in the hidden layer, is highly desirable. Next, we follow this scenario in our simulation study. We give examples for signals of a different nature (deterministic and random) and a linear relationship between the reference signal and interference. Each example is given by different levels of details for results presentation.

EXAMPLE 1. In this example, both the signal $s(k)$ and the interference $\nu(k)$ are deterministic sinusoids while the reference signal $\nu_R(k)$ is a random process. The signal $s(k)$ and the reference signal are given by

$$s(k) = 0.7\cos(2\pi 0.1k) + 0.3\sin(2\pi 0.25k),$$
$$\nu_R(k) = \sin(2\pi 0.06k + \chi(k)),$$

where $\chi(k)$ is an iid super-Gaussian random process with mean zero and variance σ_χ. The interference $\nu(k)$ is

$$\nu(k) = \cos(2\pi 0.06k) - 0.5\sin^2(2\pi 0.06(k-1)) + (\cos(2\pi 0.06k) - 0.5\sin^2(2\pi 0.06(k-1)))^2.$$

The interference canceller has to estimate the deterministic interference $\nu(k)$ with samples of a random reference signal $\nu_R(k)$. Obviously, the relationship between $\nu_R(k)$ and $\nu(k)$ is highly nonlinear, even without considering the effect of the phase noise $\chi(k)$ in the reference signal (see also Fig. 3). Figures 3 and 4 show the time waveforms and magnitude spectra of the corrupted signal $d(k) = s(k) + \nu(k)$, the desired signal $s(k)$, the interference $\nu(k)$, and the reference signal $\nu_R(k)$, respectively. Figure 4 shows that the two frequency components (one at $f_1 = 100$ and the other at $f_2 = 250$) of the signal $s(k)$ are heavily masked by the frequency components of the interference.

The NMSE performance is evaluated on a separate test set of 5000 samples measured at intervals of 500 samples during training. Figures 5–7 show the NMSE learning curves of the FIR(M), RBFN(5,n), and HRBFN(5,n) interference cancellers, respectively. The number of neurons in the hidden layer is n and is varied from 5 to 40. The input dimension for the RBFN and HRBFN

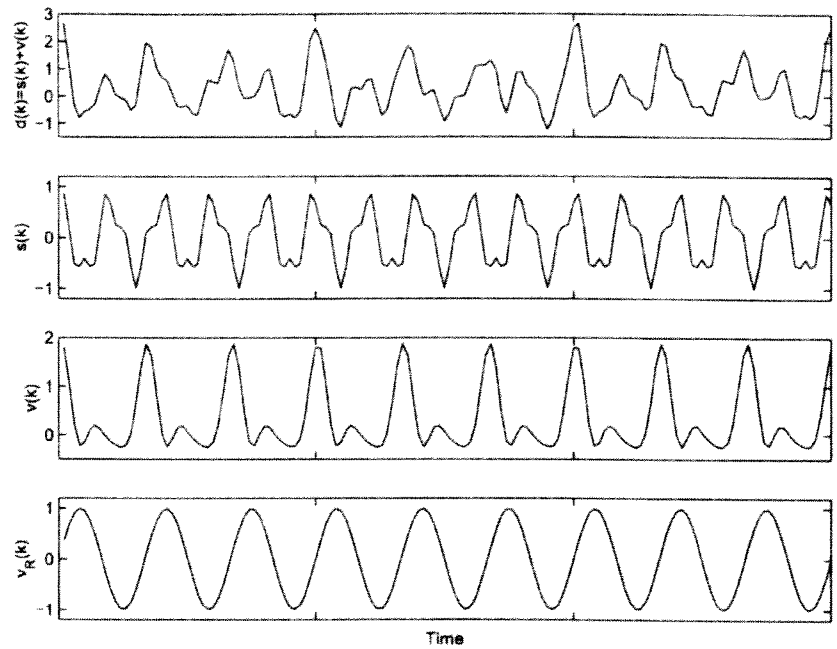

FIG. 3. Time waveforms of the corrupted signal $d(k) = s(k) + v(k)$, the desired signal $s(k)$, the interference $v(k)$, and the reference signal $v_R(k)$, respectively (Example 1).

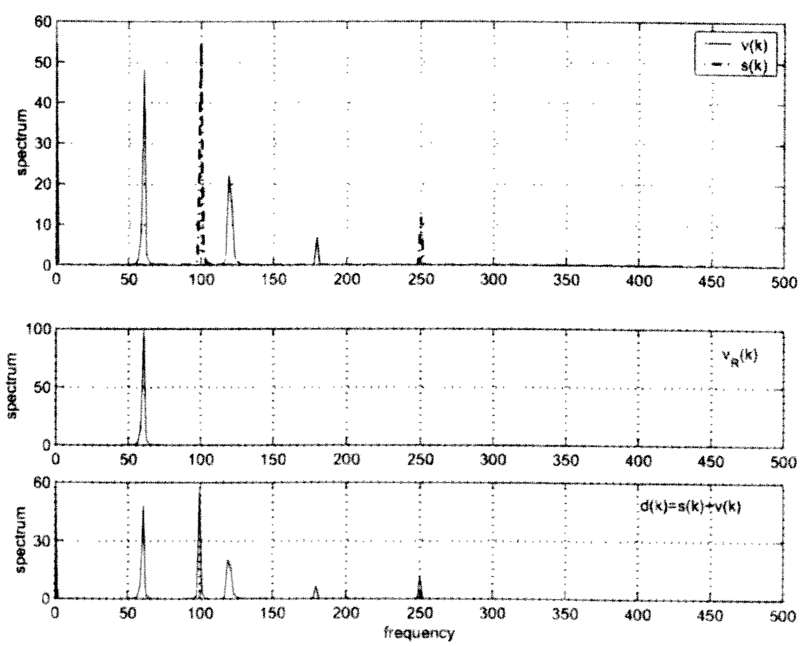

FIG. 4. Magnitude spectra of the desired signal $s(k)$, the interference $v(k)$, the reference signal $v_R(k)$, and the corrupted signal $d(k) = s(k) + v(k)$ (Example 1).

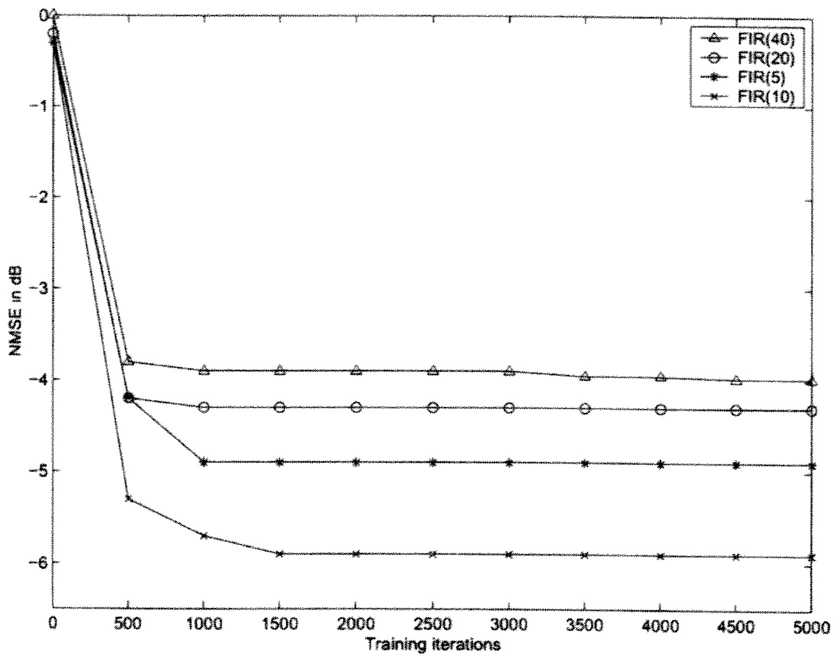

FIG. 5. Normalized mean square error for the linear FIR-based canceller: FIR(M), $M = 5, 10, 20, 40$ (phase noise power $\sigma_\chi^2 = 0.001$) (Example 1).

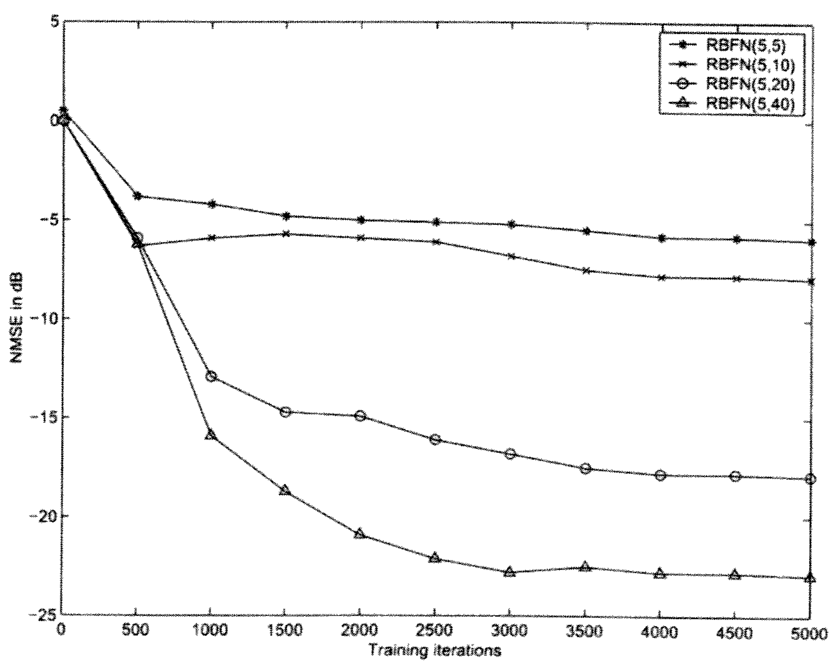

FIG. 6. Normalized mean square error for the RBF network-based canceller: RBFN(5,n), $n = 5, 10, 20, 40$ (phase noise power $\sigma_\chi^2 = 0.001$) (Example 1).

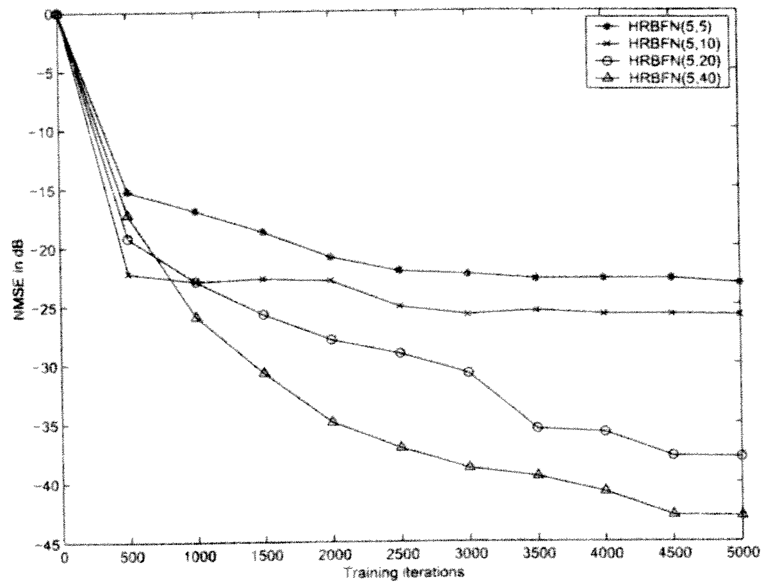

FIG. 7. Normalized mean square error for the HRBF network-based canceller: HRBFN(5,n), $n = 5, 10, 20, 40$ (phase noise power $\sigma_\chi^2 = 0.001$) (Example 1).

cancellers is fixed to 5. The input dimension for the FIR canceller is varied from 5 to 40. The power σ_χ^2 of the phase noise $\chi(k)$ is 0.001 in this simulation, corresponding to -30 dB in logarithmic scale. Here we focus on the study of performance of the various cancellers for approximating the nonlinear mapping between the reference signal and interference. The influence of phase noise on the performance we will discuss in detail below.

The FIR canceller performed very poorly. The performance of the FIR canceller depends on the size of the filter. The best performance was achieved for the HRBF network-based canceller. The HRBF network-based canceller was about 15–20 dB better than the RBF network based canceller of the same size. However, performance of the HRBFN canceller of a smaller size can, of course, be worse than the performance of the RBFN canceller a of larger size. As expected, the performance generally improved with increased size for both RBFN and HRBFN cancellers. Let us note that the stochastic gradient algorithm was used for learning of the RBFN canceller. A possible improvement in the performance of the RBFN canceller could be achieved using some other learning algorithm. However, the results of this simulation seems to be general. Better performance of the HRBFN canceller is related to the highly flexible structure of the network. However, as we will show below, it can lead to less robustness in comparison to the RBFN canceller.

The Figs. 8 and 9 show the waveforms and magnitude spectra of the test set outputs of the FIR(10), RBFN(5,20), HRBFN(5,20) cancellers, respectively. Cancellers with nonlinear processing of the reference signal with moderate size are obviously effective in contaminating interferences that are related to the reference signal in a nonlinear way. The HRBFN canceller clearly achieves better performance than the RBFN canceller of the same size.

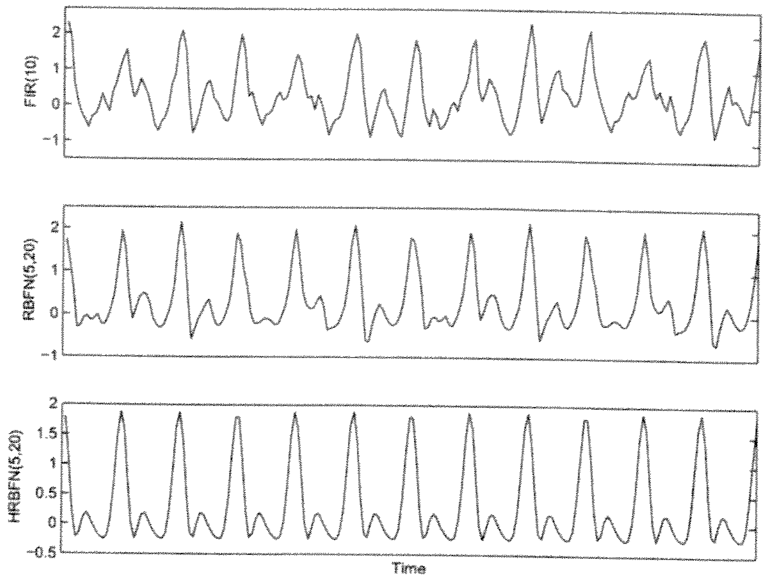

FIG. 8. Estimated waveforms of the interference signal $\hat{v}(k)$ obtained by the FIR(10), RBFN(5,20), and HRBFN(5,20) cancellers, respectively (Example 1).

We observed the effect the phase noise on the performance of the various cancellers. Figure 10 shows the performance of the FIR(10), RBFN(5,20), and HRBFN(5,20) cancellers, when the phase noise power is varied. In this figure, the NMSEs for various cancellers, after training of 10,000 samples, were plotted versus the value of σ_{χ}^2 on a logarithmic scale. We can see from the graphs that the cancellers with nonlinear processing of the reference signal performed much

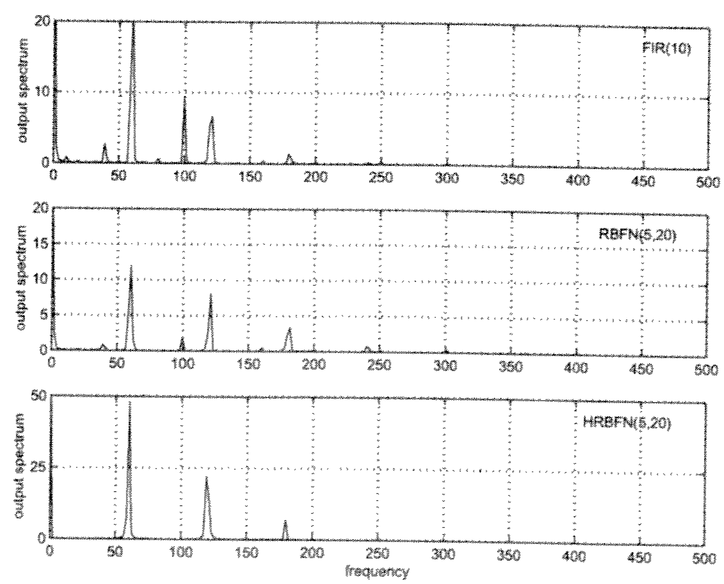

FIG. 9. Magnitude spectra of the interference signal estimations $\hat{v}(k)$ obtained by the FIR(10), RBFN(5,20), and HRBFN(5,20) cancellers, respectively (Example 1).

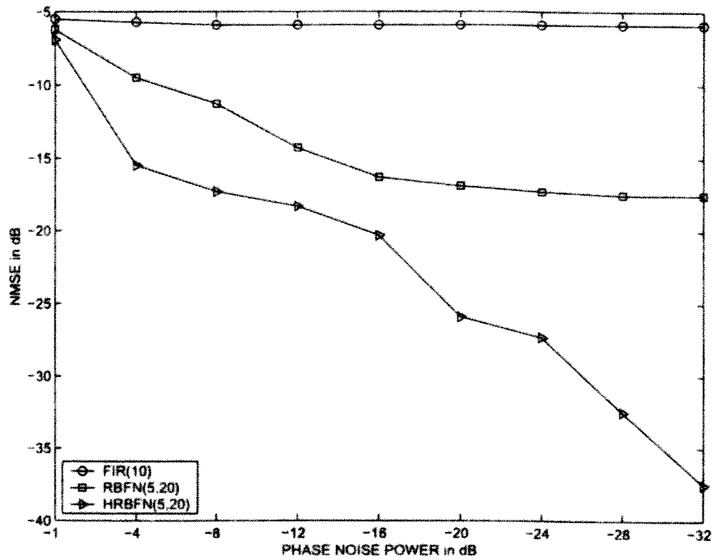

FIG. 10. Normalized mean square error learning curves for the FIR(10), RBFN(5,20), and HRBFN(5,20) cancellers, respectively, versus the phase noise power $\lg(\sigma_\zeta^2)$ (Example 1).

better than the linear canceller, even when the phase noise power is significant. Moreover, the performance of the HRBFN canceller is always better than the performance of the RBFN canceller. However, we can observe that the RBFN canceller is more robust to the power of the phase noise than the HRBFN canceller. The performance of all the cancellers is almost the same for the power of the phase noise around 0 dB.

EXAMPLE 2. This example shows interference cancellation when both the interference and reference signal are autoregressive, moving-average (ARMA) random processes and the relationship between the interference and reference signal is highly nonlinear. Thus, in this example we will consider the case of random interference and reference signals. The reference signal is given by

$$\nu_R(k) = o(k) - 0.5o(k-1) + 0.3o(k-2),$$
$$o(k) = 1.1443o(k-1) - 0.9801o(k-2) + w(k), \qquad w(k) \sim N(0, 0.05).$$

The interference is related to the reference signal is nonlinear way and is given by

$$\nu(k) = 0.3\nu(k-1) + 0.6\nu(k-2) + f(u(k)),$$
$$f(u(k)) = u^3(k) + 0.3u^2(k) - 0.4u(k) + v(k), \qquad v(k) \sim N(0, 0.05),$$
$$u(k) = 0.34\nu_R(k) + 0.87\nu_R(k-1) + 0.34\nu_R(k-3).$$

The desired signal $s(k)$ is the two-component sinusoid as used in the previous example, but with modified amplitudes; i.e.,

$$s(k) = 0.2\cos(2\pi 0.1k) + 0.1\sin(2\pi 0.25k).$$

The waveforms of all of the signals are shown in Fig. 11.

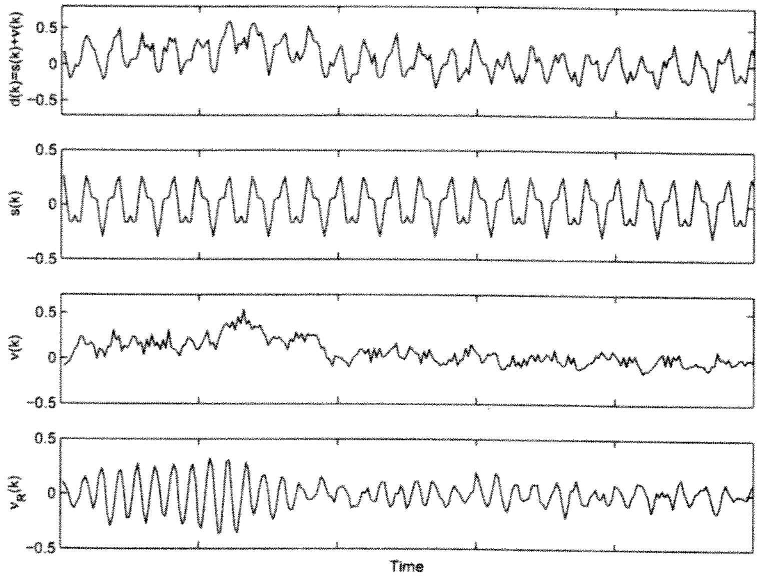

FIG. 11. Time waveforms of the corrupted signal $d(k) = s(k) + \nu(k)$, the desired signal $s(k)$, the interference $\nu(k)$, and the reference signal $\nu_R(k)$, respectively (Example 2).

Similar to previous example, the NMSE performance is evaluated on a separate test set of 5000 samples measured at intervals of 500 samples during training. Figures 12–14 show, respectively, the NMSE learning curves of the FIR(M), RBFN(15,n), and HRBFN(15,n) cancellers, where n is varied from 15 to 50 and the input dimension M for both RBFN and HRBFN cancellers is fixed

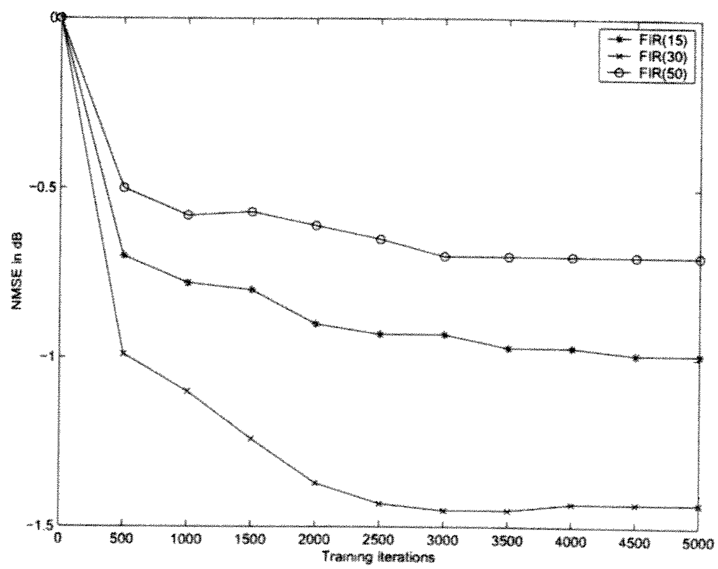

FIG. 12. Normalized mean square error for the linear FIR filter-based canceller: FIR(M), $M = 15, 30, 50$ (Example 2).

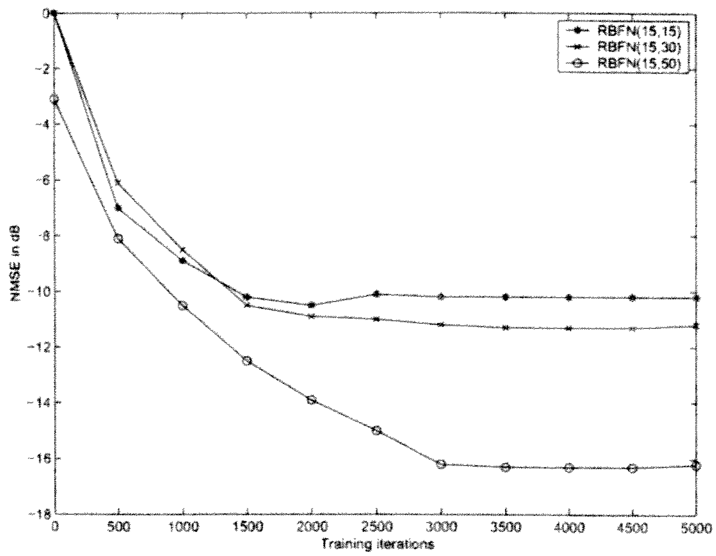

FIG. 13. Normalized mean square error for the RBF network-based canceller: RBFN(15,n), $n = 15, 30, 50$ (Example 2).

and equal to 15. The input dimension M for the FIR canceller is varied from 15 to 50.

The FIR canceller failed in this situation. The best performance was achieved for the HRBFN canceller provided that the RBFN and HRBFN cancellers have the same size. However, an RBFN canceller of bigger size may demonstrate

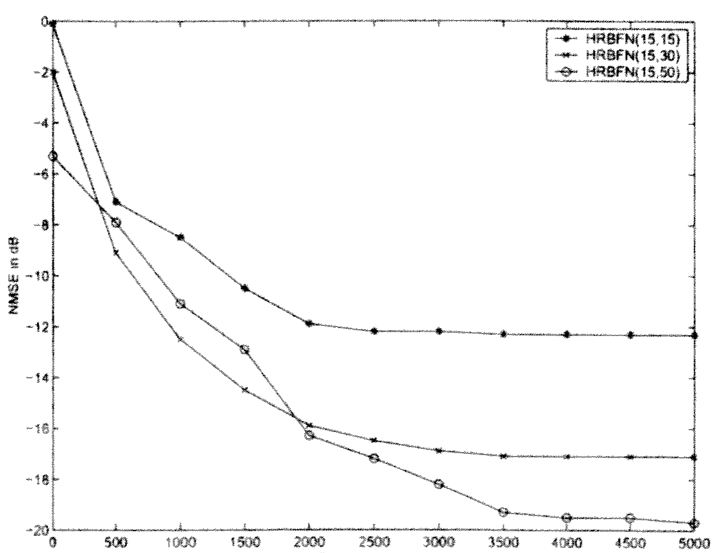

FIG. 14. Normalized mean square error for the HRBF network-based canceller: HRBFN(15,n), $n = 15, 30, 50$ (Example 2).

better performance than the HRBFN canceller of much smaller size. Generally, the results of interference cancellation for all cancellers were worse than the results of previous example for the deterministic interference signal.

EXAMPLE 3. In this example, the interference and reference signals are jointly Gaussian ARMA random processes. They are generated according to

$$\nu_R(k) = o(k) - 0.5o(k-1) + 0.3o(k-2),$$
$$o(k) = 1.1443o(k-1) - 0.9801o(k-2) + w(k);$$

and

$$\nu(k) = u(k) + 0.6u(k-1) - 0.3u(k-2),$$
$$u(k) = 1.1443u(k-1) - 0.9801u(k-1) + v(k).$$

Here, the two correlated Gaussian noise processes $w(k)$ and $v(k)$ are given by

$$w(k) = 0.3a(k) + 0.9\xi(k),$$
$$v(k) = 0.3b(k) + 0.9\xi(k),$$

where $a(k)$, $b(k)$ and $\xi(k)$ are iid zero-mean Gaussian processes with variance 1, all independent of each other. The desired signal $s(k)$ is the same as in previous examples.

Since $\nu(k)$ and $\nu_R(k)$ are jointly Gaussian, it may be expected that the FIR canceller would outperform both the RBFN and the HRBFN cancellers. Figure 15 depicts the NMSE learning curves for FIR(5), RBFN(5,30), and HRBFN(5,25) and illustrates that, indeed, all cancellers of appropriate size can give approximately the same results.

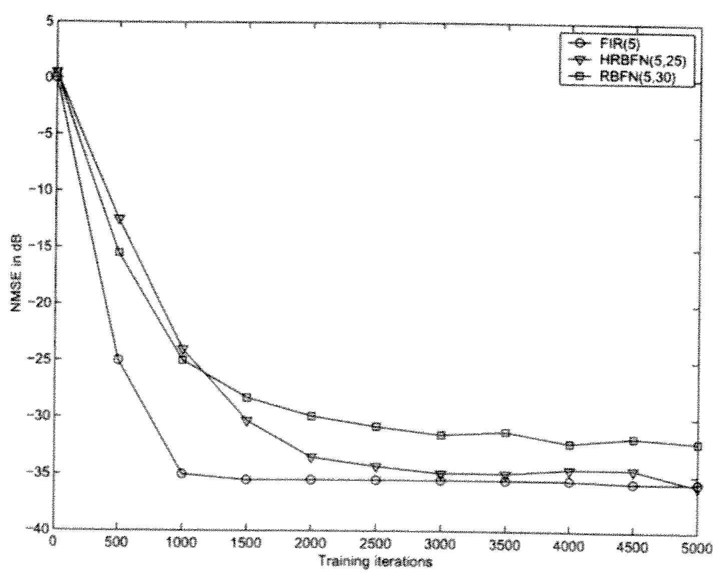

FIG. 15. Normalized mean square error learning curves for the FIR(5), RBFN(5,30), and HRBFN(5,25) cancellers, respectively (Example 3).

4. CONCLUSION

In this paper, the HRBF network-based canceller and corresponding new learning algorithm have been proposed. The HRBF network-based cancellers achieve better approximation of the interference signal in comparison to the standard RBF network-based cancellers. The reason lies in the use of a nonlinear mapping between reference and interference signals, approximated by a linear combination of specialized Green's functions called Gaussian hyper radial basis functions. The fully adaptive online learning algorithm was developed for the proposed flexible and simple form of HRBF network. Simulation study has shown applicability and efficiency of the proposed canceller. It is also important that the HRBF network-based canceller is robust to initial conditions. However, the HRBFN canceller may be less robust to the value the phase noise power than the RBNF canceller, even though it always achieves better performance.

ACKNOWLEDGMENTS

The authors thank the referees and Mr. Nicolangelo Iannella (Lab. for Visual Neurocomputing, Brain Science Institute, RIKEN, Japan) for their constructive comments and suggestions.

REFERENCES

1. Widrow, B., Glower, J. R. (Jr.), McCool, J. M., Kaunitz, J., Williams, C. S., Heart, R. H., Zeldler, J. R., Dong, E. (Jr.), and Goodlin, R. C., Adaptive noise cancelling: Principles and applications. *Proc. of the IEEE* **63** (1975), 1692–1716.
2. Widrow, B. and Walach, E., *Adaptive Inverse Control.* S.S. Series. Prentice Hall International, Englewood Cliffs, NJ, 1996.
3. Cha, I. and Kassam, S. A., Interference cancellation using radial basis function networks. *Signal Processing* **47** (1995), 247–268.
4. Cha, I. and Kassam, S. A., Channel equalization using adaptive complex radial basis function networks. *IEEE J. Selected Areas Comm.* **13** (1995), 122–131.
5. Lu, Y., Sundararajan, N., and Saratchandran, P., Performance evaluation of a sequential minimal radial basis function (RBF) neural network learning algorithm. *IEEE Trans. Neural Networks* **9** (1998), 308–318.
6. Knecht, W. G., Schenkel, M. E., and Moschytz, G. S., Nonlinear filters for noise reduction. In *Proc. of Conf. on Signal Processing VII: Theories and Applications.* 1994, pp. 1500–1503.
7. Serviere, C., Baudois, D., and Silvent, A., Noise cancelling with signal and reference correlated, using third order moments. In *Higher Order Statistics.* 1992, pp. 279–282.
8. Poggio, T. and Girosi, F., Network for approximation and learning. *Proc. of the IEEE* **78** (1990), 1481–1497.
9. Tikhonov, A. N. and Arsenin, V. Y., *Solution of Ill-Posed Problems.* Winston, Washington, DC, 1977.
10. Cichocki, A. and Unbehauen, R., *Neural Networks for Optimization and Signal Processing.* Wiley, New York, 1994.
11. Mazurek, J., Krzyzak, A., and Cichocki, A., Rates of convergence of the recursive radial basis function networks. In *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing,* 1997, pp. 3317–3320.
12. Jacobs, R. A., Increased rates of convergence through learning rate adaptation. *Neural Networks* **1** (1988), 295–307.

13. Brodsky, B. and Darkovsky, B., *Non-Parametric Methods in Change-Point Problems.* Kluwer Academic, The Netherlands, 1993.

14. Vorobyov, S. A. and Bodyanskiy, Ye. V., On one non-parametric algorithm for smoothing parameter control in adaptive filtering. *Eng. Simulation* **16** (1999), 341–350.

15. Moody, J. and Darken, D., Fast learning in networks of locally-tuned processing units. *Neural Comput.* **1** (1989), 281–294.

16. Kolmogorov, A. N., On the representation of continuous functions of several variables by superposition of continuous functions of one variable and addition. *Dokl. Akad. Nauk SSSR* **114** (1957), 953–956.

17. Cybenko, G., Approximation by superposition of a sigmodal function. *Math. Control Systems Signals* **2** (1989), 303–314.

18. Hornik, K., Multilayer feedforward networks are universal approximators. *Neural Networks* **2** (1989), 359–366.

SERGIY A. VOROBYOV was born in Ukraine, on May 24, 1972. He received the M.Sc. (with honors) in mathematical engineering, and Ph.D. in control engineering from Kharkiv State Technical University of Radioelectronics in 1994, and 1997, respectively.

In 1997 he joined the Control Systems Research Laboratory in Kharkiv State Technical University of Radioelectronics as researcher and later as senior researcher. In 1996 he visited the Forschungszentrum Karlsruhe (Germany), Institut fuer Angewandte Informatik. In 1999 he came to Japan for research activity in Brain Science Institute, RIKEN, in Laboratory for Advanced Brain Signal Processing. His current research interests include statistical signal and array processing, theory and applications of neural networks, blind signal separation, deconvolution and identification, adaptation and learning theory, and adaptive control. Dr. Vorobyov is a reviewer of *IEEE Transactions on Neural Networks* and some IEEE conferences in signal processing.

ANDRZEJ CICHOCKI was born in Poland. He received the M.Sc. (with honors), Ph.D., and Habilitate Doctorate (Dr.Sc.), all in electrical engineering and computer science, from Warsaw University of Technology, Poland, in 1972, 1975, and 1982, respectively.

Since 1972 he has been with the Institute of Theory of Electrical Engineering and Electrical Measurements at the Warsaw University of Technology, where he became a full professor in 1991. He spent a few years at the University Erlangen-Nuernberg. From 1995 to 1996 he worked as a team leader of the Laboratory for Artificial Brain Systems at the Frontier Research Program, RIKEN, Saitama, Japan, in the Brain Information Processing Group. Currently he is Head of the Laboratory for Advanced Brain Signal Processing at the Brain Science Institute. His current research interests include signal and image processing (especially biomedical signal/image processing), neural networks and their electronic implementation, learning theory and algorithms, dynamic independent and principal component analysis, optimization problems, and nonlinear circuits and systems theory and applications. He is the co-author of two books: *MOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems* (Springer-Verlag, Berlin/New York, 1998) and *Neural Networks for Optimization and Signal Processing* (Wiley, New York, 1993) and author or co-author of more than 200 scientific papers. Dr. Cichocki is Associate Editor of *IEEE Transactions on Neural Networks* and a reviewer of *IEEE Transactions on Signal Processing, IEEE Transactions on Circuits and Systems, Biological Cybernetics, Electronics Letters, Neurocomputing,* and *Neural Computation.* He is a member of several international scientific committees.