Environmental Interference Cancellation of Speech with the Radial Basis Function Networks: An Experimental Comparison

Nima Hatami

Abstract—In this paper, we use Radial Basis Function Networks (RBFN) for solving the problem of environmental interference cancellation of speech signal. We show that the Second Order Thin-Plate Spline (SOTPS) kernel cancels the interferences effectively. For make comparison, we test our experiments on two conventional most used RBFN kernels: the Gaussian and First order TPS (FOTPS) basis functions. The speech signals used here were taken from the OGI Multi-Language Telephone Speech Corpus database and were corrupted with six type of environmental noise from NOISEX-92 database. Experimental results show that the SOTPS kernel can considerably outperform the Gaussian and FOTPS functions on speech interference cancellation problem.

Keywords—Environmental interference, interference cancellation of speech, Radial Basis Function networks, Gaussian and TPS kernels.

I. INTRODUCTION

THE need for robustness in adaptive interference and noise cancellation of the speech corrupted in the real environments such as airports, automobiles, aircraft and car-cockpits, offices, and factory floors is a very important problem in robust speech and speaker recognition and verification. Many researchers have considered the noise reduction of corrupted speeches in noise-distorted environments previously. Their approaches include adaptive noise cancelling using linear and nonlinear techniques [1-4], linear and nonlinear spectral subtraction [5-7], suppression of nonharmonic frequencies [8, 9] and hidden Markov models [10].

Speech interference cancellation refers to the minimization or cancellation of interference in an observed speech, based on an estimate of the interference signal that is a function of a separate signal called the reference signal. Linear adaptive filters trained with (Least Mean Square) LMS and similar algorithms have been usually used for interference and noise cancellation [11, 12]. However, for many applications linear filter structures cannot in general implement optimum interference and noise cancellation task [13]. Using linear FIR or IIR filters, we often cannot achieve acceptable levels of

Nima Hatami is with the Department of Electrical Engineering, Shahed University, Tehran, Iran. (phone: +98-914-423 6549; e-mail: nimahatami@gmail.com; hatami@ shahed.ac.ir).

interference and noise cancellation because these techniques are not able to effectively approximate an unknown deterministic nonlinear function between the available reference and unknown interference signals. Thus, it is reasonable to try to find an optimal cancellation system using nonlinear adaptive processing models.

With the development of Neural Networks (NN) [15], new approaches to design adaptive nonlinear filters for the purpose of interference and noise cancellation have been proposed. Neural network architectures, such as multilayer perceptron (MLP) and radial basis function (RBF), have been successfully used as nonlinear tools for interference and noise cancellation [4]. Long training times and multiple parameters that need careful adjustment make it hard to apply the MLP in comparison with RBF networks. In addition, it was shown in [13] that RBF network is more accurate for the problem of interference cancellation than MLP and some other standard methods including linear filters. In Ref. [13, 14] it was demonstrated that accurate cancellation of interference and noise can be achieved with the use of RBFNs and generalized radial basis function networks, however the RBF networks are not without shortage. Depending on the architecture, the type of RBF kernels, and the training methods used, their performance varies.

In this paper, we compare the effect of different type of RBF kernels, on speech interference cancellation problem. Experimental results show the robustness of RBF network-based cancellers with SOTPS kernel.

This paper organized as follows. In section 2, we consider the problem formulation. The RBF network architecture and learning rules are described in section 3. Experimental results and discussion are presented in Section 4. Finally, conclusions are presented in Section 5.

II. PROBLEM FORMULATION

Adaptive interference and noise cancellation is based upon the availability of a corrupted signal and a reference noise signal (the environmental interferences such as voice babble or destroyer operation noise). In Fig.1, the corrupted signal contains the desired speech signal s, which is corrupted by the

undesired interference or noise signal v generated from the reference noise signal v_R. The received signal is thus given by
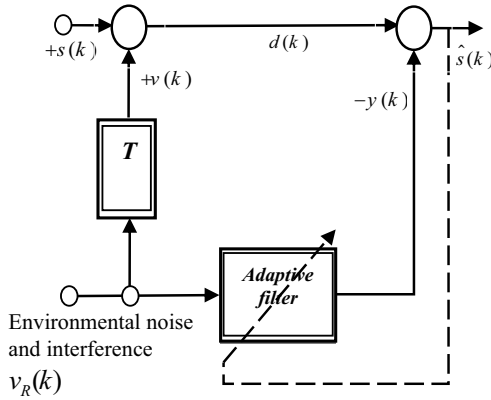
$$d(k) = s(k) + v(k) \tag{1}$$

Fig .1. The principle of adaptive interference cancellation system

The function T represents the nonlinear dynamics of the channel mapping between v and v_R. The principle of adaptive interference and noise cancellation is to adaptively process (by adjusting the filter's weights) the reference noise signal v_R to approximate the noise signal and then subtract it from the corrupted signal d to recover the desired signal s. The adaptive filter output is the noise canceller signal $y(k)$. We assume that s, v and v_R are stationary random processes with zero means, s is uncorrelated with v and v_R, and v and v_R are correlated. From Fig.1, we have

$$e(k) = s(k) + v(k) - y(k) \tag{2}$$

Squaring and taking expectation on both sides gives

$$E[e^2(k)] = E[s^2(k)] + E[(v(k) - y(k))^2] \tag{3}$$

The objective of adaptive interference cancellation is to minimize $E[(v(k) - y(k))^2]$ term. From equation (3), it is obvious that the objective is equivalent to minimizing $E[e^2(k)]$ and

when $E[(v(k) - y(k))^2] = E[(v(k) - F(v_R(k)))^2]$ approaches zero, the remaining error $e(k)$ is, in fact, the desired signal $s(k)$ where F represents the dynamics of the nonlinear adaptive filter.

III. RADIAL BASIS FUNCTION NETWORKS

A. Network structure

RBFNs are two layer networks comprising a hidden layer and an output layer. The hidden layer contains nodes which calculate the Euclidean distance r, between a center c_j and an input vector x. The result is passed through a non-linear function to generate the node output, h_j, which can be written

$$h_j = \Phi(\|x - c_j\|) \tag{4}$$

where $\Phi(.)$ is the stimulation function of the RBFs hidden layer which is also called the kernel. There are some popular choices for the kernels type, e.g. the multiquadric, thin-plate spline, Gaussian function or any other suitable functions [15]. Gaussian is the radial basis function most commonly used in the neural network community and each neuron in the RBF layer is identified by the two parameters center c_j and width δ_j. Its profile function is

$$\Phi(r) = e^{(-r^2/\delta^2)} \tag{5}$$

Thin-plate spline function (TPS) is an example of a smoothing spline, as popularized by Wahba [16]. They are usually supplemented by low-order polynomial terms. An mth order TPS is defined as

$$\Phi(r) = r^{2m} \log(r) \quad m = 1, 2, 3, ... \tag{6}$$

The thin-plate spline functions are chosen here, for their non-localized response which accommodates the rapidly changing environmental interference at the input vector. According to Eq. 6, we don't need to determine width for TPS functions. It means the network needs fewer parameters to determine and lead to lower computational cost with respect to the Gaussian kernels. Furthermore, the RBF network's ability to function approximation can depend on the distribution of the training samples. Uneven distribution of training data, however, may contribute to imprecise approximation. In that regard, it is shown in [17] that the TPS kernels are more effective than the Gaussian kernels.

In our experiment we used Gaussian and first and second order ($m = 1, 2$) TPS kernels to investigate their effect on RBFN's performance in speech interference cancellation problem.The output layer comprises a linear combiner which calculate the weighted sum of hidden layer nodes, giving an output of

$$y_i = \sum_{j=1}^{n} w_{ij} h_j \tag{7}$$

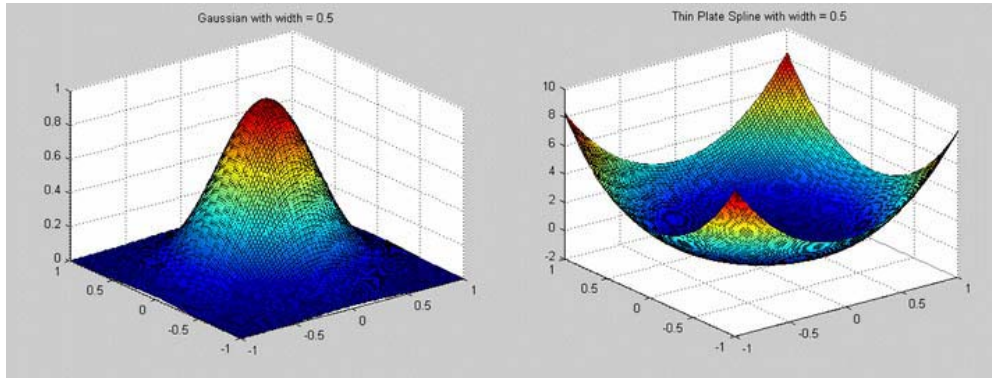where w_{ij} are the node weights of RBFN with n hidden nodes.

Fig. 2. Two conventional kernels for RBFN. Left: the Gaussian, right: TPS kernel

B. Training rules

To train the RBFN we used two stage training algorithm to set the optimal weights. Firstly, the centers are determined by fitting a Gaussian mixture model (GMM) with circular covariance using the Expectation Maximization, EM, algorithm. For more detail refer to [18, 19]. Note that the mixture model is initialized using a small number of iterations of the K-means algorithm. If the Gaussian kernels are used, the basis function widths are then set to the maximum inter-centre squared distance. Then for hidden to output weights, the least squares solution can be determined using the pseudo-inverse:

$$W^T = \Phi^+ T \qquad (8)$$

Where T is the desired output matrix and Φ^+ is the pseudo-inverse matrix of Φ

$$\Phi^+ = (\Phi^T \Phi)^{-1} \Phi^T \qquad (9)$$

This method provides a computationally simple but efficient method for the weights so that it can be computed very quickly and can be employed for real time applications.

IV. EXPERIMENTAL RESULTS

In this section, we compare the effect of three above mentioned kernels in the RBFN on noise cancellation of corrupted speeches. The clean speeches used here, are selected from the OGI Multi-language Telephone Speech Corpus database [20] where prerecorded speech signals are sampled at 8 kHz. The OGI Multi-language database consists of telephone speech from eleven languages such as English, Farsi and French. We selected four female speeches that are spoken in English and last about 3.5 min.

The noises then added to the speech signals at the SNR of +10 dB. Six different noise signals from NOISEX-92 database [21] used in this experiment: 1- White Noise acquired by sampling high-quality analog noise generator; 2- Voice

Babble where its source is 100 people speaking in a canteen that individual voices are slightly audible. 3- Destroyer engine room noise; 4- Destroyer operations room background noise. 5- Factory floor noise that recorded near plate-cutting and electrical welding equipment. 6- And finally, Vehicle interior noise signal recorded in a Volvo car running at 120 km/h, in 4th gear, on an asphalt road, in rainy conditions.

We train our networks with white noise corrupted speeches and test their generalization performance with another noise types such that our test data not been seen during training process. Since the neural networks are stochastic methods, the results are averages of ten repetitions on data set.

The networks initial conditions are as follows: The initial centers of the RBFN are determined by 50 iterations of the K-means clustering algorithm. If the activation functions are Gaussians, then the basis function widths are then set to the maximum inter-centre squared distance. The weights initialized to small random values in the range of [−0.1, 0.1]. The input dimension for the RBFNs is fixed to 2.
Performance of the noise and interference cancellers is measured by the *normalized mean squared error* (NMSE):

$$NMSE = \frac{E\{(\hat{s}(k) - s(k))^2\}}{E\{v^2(k)\}} \qquad (10)$$

where $\hat{s}(k)$ is the simulated signal.

Because of universal approximation feature of feed forward neural networks [22], it is not necessary to study the different class on nonlinearity mapping between reference and interference signal. We assume that the relation between the reference noise and the interference signal is a nonlinear function given by

$$v(k) = 0.6 \times (v_R(k))^3 \qquad (11)$$

The results of our experiments are shown in Figs. 3 and 4. Since the performance of such methods, highly depend on the number of neurons in the hidden layer, we measure performance of different networks versus the number of

hidden neurons as shown in figure 3. Figures 3a, b and c compare the NMSE generalization curves of the three RBFN with the conventional Gaussian, FOTPS and SOTPS kernels, respectively. The number of neurons in the hidden layer is varied from 3 to 40. We train our networks with 100 iterations of EM algorithm.
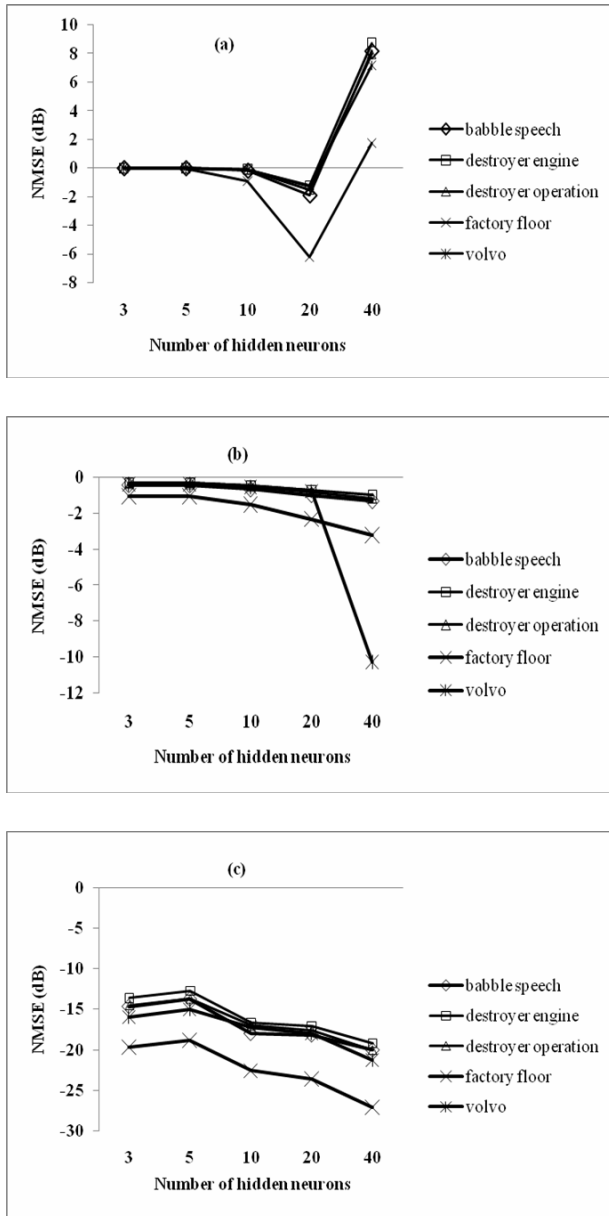

Fig. 3. Generalization error (NMSE) versus the number of hidden neurons in RBFN: (a) Gaussian kernel, (b) FOTPS kernel and (c) SOTPS kernel.

These results indicate that the RBFN canceller with Gaussian kernel performed very poorly especially when number of hidden neurons reaches to 40. The best performance was achieved for the RBFN with SOTPS kernel. It was about 13–23 dB better than the RBFN with FOTPS kernel and 13-30 dB better than RBFN with conventional Gaussian kernel of the same size. An important point is that the performance of the SOTPS RBFN canceller of a smaller size was better than the performance of the RBFN canceller with both FOTPS and Gaussian kernel of larger size. We conjectured that this was due to their non-localized response that accommodates the rapidly changing environmental interference state-space at the input vector.

As expected, the performance generally improved with increased size for RBFN cancellers. As shown above, for RBFN cancellers with 20 or fewer size, performance of interference cancellation increased with increasing the number of hidden neurons. For the size of 40 the Gaussian kernel shows different treatment because performance of both RBFN with FOTPS and SOTPS kernels increased but performance of RBFN with conventional Gaussian kernel decreased. Furthermore, performance of cancellers on factory floor noise for all size was considerably better than others.

In second experiment, we fixed all hidden neurons to 20. Other experiment conditions were not changed. As shown in figure 4 the RBFN with Gaussian kernel have poor generalization on the another types of environmental interference because of its better performance on speech corrupted with white noise and poorly cancellation of other noise types. But RBFN cancellers with SOTPS kernel's performance is very satisfactory on all type of unseen noises however it's cancellation ability on noise type which used in training phase (white noise) was slightly lower than the Gaussian kernel.

V. CONCLUSION

In this paper, we investigated the effect of different kernels in the radial basis function networks for environmental interference and noise cancellation of speech signals. The RBF network-based cancellers with SOTPS kernel achieve better approximation of the interference signal in comparison to the standard RBF network-based cancellers. Simulation study has shown generalization ability and efficiency of the SOTPS based interference canceller on distorted speech signals corrupted with unseen noise types regardless of network size.
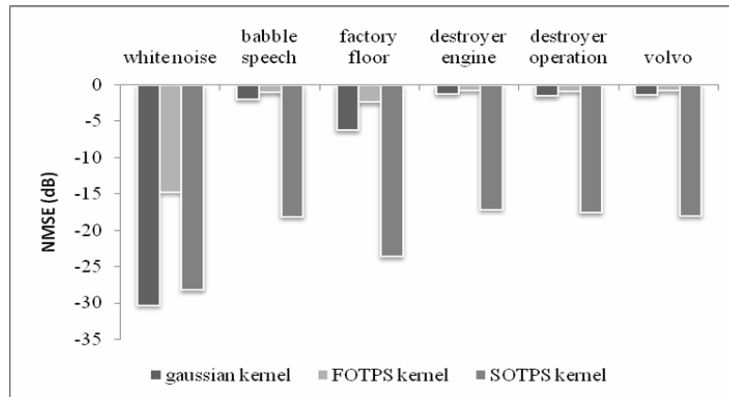
Fig. 4. Experimental comparison of RBFN canceller kernels effect on different interference.

REFERENCES

[1] R.W. Jones, B.L. Olsen, B.R. Mace, "Comparison of convergence characteristics of adaptive IIR and FIR filters for active noise control in a duct", Applied Acoustics, vol. 68, pp. 729–738, 2007.

[2] J. Elliott Stephen, M. Stothers Ian, and A. Nelson Philip, "A Multiple Error LMS Algorithm and Its Application to the Active Control of Sound and Vibration", IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-35, No. 10, pp. 1423-1434, 1987.

[3] M. Feder, A. V. Oppenheim, and E. Weinstein, "Maximum Likelihood Noise Cancellation Using the EM Algorithm", IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-37, No. 2, pp. 204-216, February 1989.

[4] C. K. Chen and T. D. Chiueh, "Multilayer Perceptron Neural Networks for Active Noise Cancellation", in Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS), Atlanta, GA, May 1996.

[5] J. S. Lim, "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-26, pp. 471–472, Oct. 1978.

[6] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-27, pp. 113–120, Apr. 1979.

[7] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech", Proc. IEEE, vol. 67, pp. 1586–1604, Dec. 1979.

[8] J. S. Lim, A. V. Oppenheim, and L. D. Braida, "Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-26, pp. 354–358, Aug. 1978.

[9] T. W. Parsons, "Separation of speech from interfering of speech by means of harmonic selection", J. Acoust. Soc. Amer., vol. 60, pp.911–918, Oct. 1976.

[10] H. Sameti, H. Sheikhzadeh, L.Deng, and R. L. Brennan, "HMM-Based Strategies for Enhancement of Speech Signals Embedded in Nonstationary Noise", IEEE Trans. Speech and Audio Processing, Vol. 6, No. 5, September 1998.

[11] J. Chen, J. Benesty, Y. Huang, "On the optimal linear filtering techniques for noise reduction", Speech Communication, vol.49, pp. 305–316, 2007.

[12] B. Widrow and E. Walach , "Adaptive Inverse Control". S.S. Series. Prentice Hall International, Englewood Cliffs, NJ, 1996.

[13] I. Cha and S. A. Kassam, "Interference cancellation using radial basis function networks". Signal Processing, vol. 47, pp. 247–268, 1995.

[14] Y. Lu, N. Sundararajan and P. Saratchandran, "Performance evaluation of a sequential minimal radial basis function neural network learning algorithm". IEEE Trans. Neural Networks, vol. 9, pp. 308–318, 1998.

[15] S. Haykin, "Neural Networks: A Comprehensive Foundation", Prentice Hall International, 1999.

[16] G. Wahba, http://www.stat.wisc.edu/~wahba/

[17] K. Mike Tao, "A Closer Look at the Radial Basis Function Networks", conference record of the 27th asilomar conference on signals, systems and computers, vol. 1, pp. 401-405, 1993.

[18] C. M. Bishop, "Pattern recognition and machine learning", Springer, 2006.

[19] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum Likelihood from Incomplete Data via de EM Algorithm", in Journal of the Royal Statistical Society, B 39(1) 1-38, 1976.

[20] Y. K. Muthusamy, R. A. Cole, and B. T. Oshika, "The OGI multi-language telephone speech corpus", Proceedings of the International Conference on Spoken Language Proceedings, Banff, Alberta, Canada, pp 895-898, October 1992.

[21] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems", Speech Communication, Vol. 12, No. 3, pp. 247 - 251, 1993.

[22] K. Hornik, "Multilayer feedforward networks are universal approximators". Neural Networks, vol. 2, pp. 359–366, 1989.

Nima Hatami received the B.S. degrees in Electrical Engineering from Shahid Rajaee University in 2006 and the M.S. degree in the Department of Electrical Engineering, Shahed University, Tehran, Iran, in 2009.

His research interests include Pattern recognition, Machine vision, Applications of Machine Learning in pattern recognition, neural network ensembles, multiple classifier systems and applications of these areas to Biometrics, Visual object recognition and Handwriting recognition. He is a member of the IEEE, IAPR, IAENG and Scientific and Technical Committee of WASET and G-Conferences.