Neural-Network Architecture for Linear and Nonlinear Predictive Hidden Markov Models: Application to Speech Recognition

L. Deng, K. Hassanein, and M. Elmasry
Department of Electrical and Computer Engineering
University of Waterloo, Waterloo, Ont. Canada

Abstract

A speech recognizer is developed using a layered neural network to implement speech-frame prediction and using a Markov chain to modulate the network's weight parameters. We postulate that speech recognition accuracy is closely linked to the capability of the predictive model in representing long-term temporal correlations in data. Analytical expressions are obtained for the correlation functions for various types of predictive models (linear, nonlinear, and jointly linear and nonlinear) in order to determine the faithfulness of the models to the actual speech data. The analytical results, computer simulations, and speech recognition experiments suggest that when nonlinear and linear prediction are jointly performed within the same layer of the neural network, the model is better able to capture long-term data correlations and consequently improve speech recognition performance.

I. Introduction

Speech frames generated by speech preprocessors in automatic speech recognizers typically possess strong correlations over time [5, 8]. The correlations stem, to a large degree, from the complex interactions and overlap patterns among various articulators involved in the dynamic process of speech production [9]. Standard hidden Markov models (HMMs) [1], based on the state-conditioned IID (independent and identical distribution) assumption, are known to be weak in capturing such correlations. The strength of data correlations in the HMM source decays exponentially with time due to the Markov property, while the dependence among speech events does not follow such a fast and regular attenuation.

The linear predictive HMM proposed in [14] and [11] is intended to overcome this weakness but shows no clear evidence of superiority over the standard HMM in speech recognition experiments [11]. This can be understood because the correlation (or the envelop of the correlation function) introduced by the state-dependent linear prediction mechanism decays also in an exponential manner with time lag [2]. This makes the capability of the linear predictive HMM, in dealing with speech-frame correlations, essentially the same as that exhibited by a standard HMM having just a larger number of states.

Nonlinear time series models [15, 16] are believed to be capable of representing the temporal correlation structure of speech frames in a more general and realistic manner. In order to represent the well known nonstationary nature of speech frames, the parameters in the

0-7803-0118-8/91/0009-0411$01.00

time series models can be made to vary with time. One elegant way of achieving this is to assume that the evolution of the time series model parameters follows a Markov chain.

In this paper we describe an implementation of this idea where three-layered feed-forward neural networks are used as Markov-state-dependent nonlinear autoregressive-type *skeleton* functions (terminology borrowed from [16]) in a time series model. Layered neural networks are ideal tools for implementing mapping functions applicable to speech-frame prediction, an idea originally proposed in [12], for two main reasons. First, it has been proved that a network of just one hidden layer is sufficient to approximate arbitrarily well any continuous function [3, 10]. Thus prediction of highly dynamic and complex speech frames can be potentially made as accurate as possible. Second, the effective back-propagation algorithm is available for network parameter estimation. To understand the properties of predictive models, we carried out detailed analysis on the statistical correlation structures of various first-order predictive models. One principal conclusion drawn from the result of the analysis is that long-term temporal correlations in the modeled data cannot be achieved with only one single predictive term, either linear or nonlinear. However, combinations of linear and nonlinear terms are shown, analytically and by simulation, to be able to produce such signal correlations, which is a desirable property for a speech model. Speech recognition experiments conducted on a speaker-dependent discrete-utterance E-set task with various types of predictive HMMs demonstrate close relationships between the recognition accuracy and the capabilities of the models in handling temporal correlations of speech data.

II. Correlation Structure in Speech Data and Coarticulation in Speech Dynamics

Speech patterns are known to be highly dynamic and complex in nature [9]. One principal source of this complexity is coarticulation. In articulatory terms, coarticulation results from the fact that several articulators do not always move instantaneously and simultaneously from one targeted articulatory configuration to another. In acoustic terms, coarticulation is related to context dependence, whereby acoustic realization of a sound is strongly affected by the sounds just uttered and to be uttered next. This context dependence makes any IID source model, or the locally IID source model as is the case with the standard HMM, a poor choice for fitting speech data and is a major source of errors in speech recognition [6]. Good models should provide correlation structures rich enough to accommodate the context dependence and other types of temporal dependence in speech data.

III. Analysis of Correlation Structures for State Conditioned Predictive Models

In this section, we conduct analytical evaluation of the correlation functions for various types of predictive models (linear, nonlinear, and their combination) in order to assess their faithfulness as a speech model for use in speech recognition. For the sake of simplicity in exposition yet without apparent loss of generality, we assume first-order prediction and scalar observations.

III.1. Linear prediction

The state-conditioned linear predictive source model for speech data Y_t's is chosen to have the following form:

$$Y_{t+1} = \phi Y_t + \epsilon_{t+1}, \qquad t = 0, 1, ..., T. \tag{1}$$

where ϵ_t is an IID residual random variable with zero mean and variance σ^2 and the skeleton function is a linear function of the data.

It is well known [2] that when the predictive coefficient ϕ is less than one in absolute value, then the process (1) is stationary and its autocovariance function declines exponentially as the time lag τ with the time constant $-log\phi$.

III.2. Prediction with a single nonlinear term

The state-conditioned nonlinear predictive source model replaces the linear predictive term in (1) with a symmetric, continuously differentiable but otherwise arbitrary nonlinear skeleton function $f(\cdot)$:

$$Y_{t+1} = f(Y_t) + \epsilon_{t+1}, \qquad t = 0, 1, ..., T. \tag{2}$$

In this section, $f(\cdot)$ is restricted to contain only one single nonlinear term. In actual implementation of the predictive model, we chose $f(\cdot)$ to be a specific nonliner function, such as the tanh function. But otherwise $f(\cdot)$ is not restricted to any specific form when we study the statistical properties of model (2) in this section.

The method we use to derive the correlation function for (2) resembles the perturbation analysis for the study of nonlinear differential equations [13]. To proceed, we construct a family of models which is parameterized by α:

$$Y_{t+1}(\alpha) = \alpha f(Y_t(\alpha)) + \epsilon_{t+1}, \tag{3}$$

and model (2) is considered as one model in the family (3) whose statistical properties change continuously with the parameter α.

Once the model is parameterized, the autoregression on the data Y_t can be removed by performing power-series expansion of the nonlinear function $f(\cdot)$:

$$\begin{aligned}
Y_1(\alpha) &= \epsilon_1 + \alpha f(Y_0(\alpha)), \\
Y_2(\alpha) &= \epsilon_2 + \alpha f(Y_1(\alpha)) \\
&= \epsilon_2 + \alpha f(\epsilon_1) + \alpha^2 f(Y_0) f'(\epsilon_1) + \frac{1}{2}\alpha^3 f^2(Y_0) f''(\epsilon_1) + \cdots, \\
Y_3(\alpha) &= \epsilon_3 + \alpha f(Y_2(\alpha)) \\
&= \epsilon_3 + \alpha f(\epsilon_2) + \alpha^2 f(\epsilon_1) f'(\epsilon_2) + \alpha^3 f(Y_0) f'(\epsilon_1) f'(\epsilon_2) + \cdots, \\
&\vdots
\end{aligned}$$

and in general,

$$Y_t(\alpha) = \epsilon_t + \alpha f(\epsilon_{t-1}) + \alpha^2 f(\epsilon_{t-2}) f'(\epsilon_{t-1}) + \alpha^3 f(\epsilon_{t-3}) f'(\epsilon_{t-2}) f'(\epsilon_{t-1}) + \cdots. \tag{4}$$

(In the above, $f'(\cdot)$ denotes the derivative of $f(\cdot)$ with respect to its argument.)

From (4) the covariance function for model (3) is calculated to give

$$\begin{aligned}
&Cov[Y_t(\alpha), Y_{t+\tau}(\alpha)] \\
&\approx Cov(\epsilon_t, \epsilon_{t+\tau}) + \alpha Cov[f(\epsilon_t), \epsilon_{t+\tau}] + \alpha Cov[\epsilon_t, f(\epsilon_{t+\tau-1})] \\
&\quad + \alpha^2 Cov[f(\epsilon_{t-2}) f'(\epsilon_{t-1}), \epsilon_{t+\tau}] + \alpha^2 Cov[\epsilon_t, f(\epsilon_{t+\tau-2}) f'(\epsilon_{t+\tau-1})] \\
&\quad + \alpha^3 Cov[f(\epsilon_{t-2}) f'(\epsilon_{t-1}), f(\epsilon_{t+\tau-1})] \\
&\quad + \alpha^3 Cov[f(\epsilon_{t-1}), f(\epsilon_{t+\tau-2}) f'(\epsilon_{t+\tau-1})] \\
&\quad + \alpha^4 Cov[f(\epsilon_{t-2}) f'(\epsilon_{t-1}), f(\epsilon_{t+\tau-2}) f'(\epsilon_{t+\tau-1})].
\end{aligned} \tag{5}$$

Among the eight terms in (5), the first, second, fourth, and sixth terms are zero for $\tau \geq 0$. This is due to the IID assumption for ϵ_t and to the fact that $f(\cdot)$ is a static function containing no memory. The fifth term, $Cov[\epsilon_t, f(\epsilon_{t+\tau-2}) f'(\epsilon_{t+\tau-1})]$, is non-zero only for $\tau = 1$ and $\tau = 2$. The seventh and the eighth terms are non-zero only for $\tau = 1$. Likewise, any higher order terms of α in the covariance function which are omitted due to cutoff in the power-series expansion of $Y_t(\alpha)$ would contain non-zero values only for small time lags.

We conclude from the above analysis that prediction of a time series with a single nonlinear term alone does not produce long-term temporal correlations in the model's output.

III.3. Joint prediction with nonlinear and linear terms

In this section we investigate correlation properties of the data generated from the stationary time series model

$$Y_{t+1} = \phi Y_t + f(Y_t) + \epsilon_{t+1}, \qquad t = 1, 2, ..., T, \tag{6}$$

whose skeleton function has an additional linear predictive term to that of model (2) studied in Section III.2. Although a single nonlinear predictive term just by itself is unable to generate desirable long-term data correlations (Section III.2), we can expect the interaction of the nonlinear term with the additional linear predictive term to produce such desirable properties. The following analysis, and the simulation results shown in Section IV, confirm this expectation.

Following a similar approach to that of Section III.2, the family of models constructed for (6) appropriate for the ensuing perturbation analysis is

$$Y_{t+1}(\alpha) = \phi Y_t(\alpha) + \alpha f(Y_t(\alpha)) + \epsilon_{t+1}. \qquad t = 1, 2, ..., T. \tag{7}$$

We now decompose the stationary random process $Y_t(\alpha)$ into its stationary component processes by representing it as a power-series expansion on α

$$Y_{t+1}(\alpha) = Y_{t+1,0} + \alpha Y_{t+1,1} + \frac{1}{2!}\alpha^2 Y_{t+1,2} + \frac{1}{3!}\alpha^3 Y_{t+1,3} + \cdots. \tag{8}$$

In order to identify the component processes $Y_{t,i}, i = 0, 1, 2, ...$, we substitute (8) into (7) and approximate the nonlinear function $f(\cdot)$ by truncating its power-series expansion. This gives

$$\begin{aligned} Y_{t+1}(\alpha) &\approx \phi(Y_{t,0} + \alpha Y_{t,1} + \frac{1}{2!}\alpha^2 Y_{t,2} + \frac{1}{3!}\alpha^3 Y_{t,3}) \\ &\quad + \alpha[f(Y_{t,0}) + f'(Y_{t,0})(\alpha Y_{t,1} + \frac{1}{2!}\alpha^2 Y_{t,2} + \frac{1}{3!}\alpha^3 Y_{t,3})] + \epsilon_{t+1} \\ &= (\phi Y_{t,0} + \epsilon_{t+1}) + \alpha[\phi Y_{t,1} + f(Y_{t,0})] + \alpha^2[\frac{1}{2}\phi Y_{t,2} + f'(Y_{t,0})Y_{t,1}] \\ &\quad + \alpha^3[\frac{1}{6}\phi Y_{t,3} + \frac{1}{2}f'(Y_{t,0})Y_{t,2}] + \cdots. \end{aligned} \tag{9}$$

By equating the coefficients of α^i in (8) and in (9), we obtain the following recursive relations among the component processes $Y_{t,k}, k = 0, 1, 2, ...$:

$$\begin{aligned} Y_{t+1,0} &= \phi Y_{t,0} + \epsilon_{t+1}, \\ Y_{t+1,1} &= \phi Y_{t,1} + f(Y_{t,0}), \\ Y_{t+1,2} &= \phi Y_{t,2} + 2f'(Y_{t,0})Y_{t,1}, \\ Y_{t+1,3} &= \phi Y_{t,3} + 3f'(Y_{t,0})Y_{t,2}, \\ &\vdots \end{aligned} \tag{10}$$

According to (10), we can proceed to derive the autocovariance function for $Y_t(\alpha)$ denoted by

$$\gamma = Cov[Y_t(\alpha), Y_{t+\tau}(\alpha)].$$

Using (8) and truncating the expansion up to the first order, we have

$$\gamma \approx Cov[Y_{t,0} + \alpha Y_{t,1}, Y_{t+\tau,0} + \alpha Y_{t+\tau,1}]. \tag{11}$$

Use of the stationarity property of $Y_{t,0}$ and $Y_{t,1}$ leads to

$$\begin{aligned}\gamma &= \phi^2\gamma + \alpha^2 Cov[f(Y_{t-1,0}), f(Y_{t+\tau-1,0})] \\ &\quad + \phi\alpha Cov[Y_{t-1,0} + \alpha Y_{t-1,1}, f(Y_{t+\tau-1,0})] \\ &\quad + \phi\alpha Cov[Y_{t+\tau-1,0} + \alpha Y_{t+\tau-1,1}, f(Y_{t-1,0})].\end{aligned}$$

Re-arranging terms and using the stationarity property of $Y_{t,0}$ and $Y_{t,1}$ again give γ which is equal to

$$\frac{1}{(1-\phi^2)}\{\alpha^2 Cov[f(Y_{t-1,0}), f(Y_{t+\tau-1,0})] + 2\phi\alpha Cov[Y_{t,0} + \alpha Y_{t,1}, f(Y_{t+\tau,0})]\}. \tag{12}$$

$Y_{t,0}$, the zero-th order expansion of $Y_t(\alpha)$, is a linear process and its properties are well understood (Section III.1). To obtain the desired form for γ, we need an explicit expression for the component covariance in (12) involving nonlinear process $Y_{t,1}$. Repetitive use of the recursive relations in (10) gives

$$\begin{aligned}Cov[Y_{t,1}, f(Y_{t+\tau,0})] &= Cov[\phi Y_{t-1,1} + f(Y_{t-1,0}), f(Y_{t+\tau,0})] \\ &= \phi Cov[Y_{t-1,1}, f(Y_{t+\tau,0})] + Cov[f(Y_{t-1,0}), f(Y_{t+\tau,0})] \\ &= \phi Cov[\phi Y_{t-2,1} + f(Y_{t-2,0}, f(Y_{t+\tau,0})] + Cov[f(Y_{t-1,0}), f(Y_{t+\tau,0})] \\ &\;\;\vdots \\ &= \sum_{i=0}^{t-1} \phi^i Cov[f(Y_{t-i-1,0}), f(Y_{t+\tau,0})]\end{aligned}$$

Substitution of this result into (12) leads finally to

$$\begin{aligned}\gamma &= \frac{1}{(1-\phi^2)}\{\alpha^2 Cov[f(Y_{t-1,0}), f(Y_{t+\tau-1,0})] + 2\phi\alpha Cov[Y_{t,0}, f(Y_{t+\tau,0})] \\ &\quad + 2\phi\alpha^2 \sum_{i=0}^{t-1} \phi^i Cov[f(Y_{t-i-1,0}), f(Y_{t+\tau,0})]\}.\end{aligned}$$

The first two terms in the above expression are exponentially declining as a function of time lag τ because the component processes involved are just static functions of linear processes. The remaining summation, however, would in general decay more slowly because of the many contributing terms.

We conclude from the above result that in a model where linear and nonlinear terms are jointly used for prediction, the correlation function tends to decay more slowly than in models utilizing either linear or nonlinear predictive terms separately. In other words, if a model is to be constructed to represent natural data which is known to possess long-term inter-time correlation, such as speech, a model of joint linear and nonlinear predictive terms would be superior to (i.e. more faithful than) that of only one predictive term, either linear or nonlinear.

IV. Simulation Results on the Predictive Models

Computer simulations were carried out to check the analytical results obtained in Section III, where many approximations were employed to allow for the analysis to be carried out in a closed form. Using a random number generator which produces Gaussian IID residuals ϵ_t with a zero mean and unit variance, we created artificial "speech" data according to models (1), (2) and (6), respectively. The simulated data consisted of a total of 100,000 points, from which the sampled autocorrelation functions were computed for each model. The autocorrelation functions for models (1), (2) and (6) were superimposed on the same plot for comparison. Figs.1a,b,c,d correspond to four different forms of nonlinear functions $f(\cdot)$ in models (2) and (6): tanh, sigmoid, symmetric square root, and symmetric one-quarter power function, respectively. Parameter ϕ, interpreted as the neural network weight, is assumed a fixed value less than one (this guarantees stationarity of the modeled processes). It is apparent that regardless of the form of nonlinearities, joint use of linear and nonlinear prediction terms (model 6) produces significantly stronger correlations in the simulated data than the use of separate prediction terms at any $\tau > 0$. This conforms to the analytical results obtained in Section III.

V. Speech Recognition Experiments

The various predictive HMMs discussed so far were evaluated on a speech recognition task using a database of speaker-dependent speech recorded at the University of Waterloo [7]. The task domain of the recognizer was a total of six CV syllables where C encompasses six stop consonants /p/, /t/, /k/, /b/, /d/, /g/ and V is the vowel /i/. All the syllables were uttered with a short pause in between by native English speakers in a normal office environment. We chose this task for two reasons. First, stop confusion and E-set discrimination are known to be difficult tasks and are of fundamental significance for general speech recognition problems [4, 5]. Second, acoustic realization of stop-vowel syllables exhibits the typical nature of coarticulation and forms special sets of temporally correlated speech data. In order to faithfully represent such speech data, a model would have to be capable of handling long-term temporal correlations. The stop-E-set discrimination task allows us to perform comparative tests on the capability and the effectiveness of various types of predictive models and to assess the practical value of the analytical results obtained in Section III in speech recognition.

Sampled speech data were obtained using a DSP Sona-Graph workstation. Data were collected by digitally sampling the speech signal at 16 kHz. A Hamming window of a 25.6-ms duration was applied every 10 ms. Within each window, a 7-dimensional vector consisting of mel-frequency cepstral coefficients was computed as raw speech data to be fed to the predictive HMMs. These coefficients were appropriately scaled to accommodate the limited dynamic range in the neural network's operation. (We did not use delta cepstral coefficients over time as expanded feature sets since we felt this would violate the principle of consistency; models are considered good only if they can *generate* observations which are statistically consistent with raw data. Our intention was to develop predictive HMMs as data-generator models and we believe that the advantages of using delta coefficients can be coherently embedded in the predictive mechanisms of the models.)

In implementing the neural predictive HMMs, each syllable in the vocabulary was represented by a three-layered, feed-forward fully connected neural network. The network's weight parameters are modulated by a four state left-to-right Markov chain. The network consisted of seven input units, one accepting each scaled cepstral coefficient. Five hidden units were employed which were either all linear, all nonlinear, or two linear and three nonlinear depending on the type of the predictive model considered. Seven output units were all assumed linear, each having the desired value of a corresponding predicted cepstral coefficient one frame ahead. For comparison purposes, we also implemented the standard HMM, which is locally IID and is a degenerated case of the predictive HMM's when the skeleton is fixed at a state-dependent constant (i.e. Gaussian mean). To make a fair comparison, the standard HMM was implemented with an identical structure to the predictive HMM with the mere difference of replacing data prediction with locally IID data generation. The covariance matrices in the standard HMM were assumed identity matrices in keeping with the use of the unweighted least-mean-square error function in the training and testing for the predictive HMMs.

The segmental K-means algorithm was used for training the standard HMM and, in combination with the standard back-propagation algorithm, for training all three types of the predictive HMMs. Details of the training algorithm, as well as the classification algorithm, were described in [12] and omitted here. Training and testing data were obtained from three male speakers. For each speaker, the training set consisted of eight tokens for each of the six syllables in the vocabulary. The test set consisted of 14 tokens for each of the six syllables, giving a total of 84 test tokens for each speaker. Comparative recognition accuracy on the test data for the standard HMM recognizer (locally IID source with a fixed-valued "predictive" term) and for the HMM recognizers using various forms of speech-frame prediction is shown in Table 1. We draw particular attentions to the significantly higher recognition rate obtained with mixed linear and nonlinear hidden units in the neural network architecture compared with other types of recognizers.

VI. Conclusion

It is concluded from this work that the signal prediction mechanism implemented by carefully structured neural networks is a potentially effective scheme for high-accuracy speech recognition. In the specific task of stop-E-set recognition, use of nonlinear prediction in conjunction with linear prediction was demonstrated to be superior to linear or nonlinear prediction alone, as well as to the standard HMM. This superiority is believed to result from the higher capacity provided by this joint prediction mechanism in representing the inherent long-term correlations between successive speech frames. Analytical evaluation and computer simulations of the correlation functions for various types of simplified predictive models provide strong support for this postulation.

Speaker	Standard HMM	Linear Pred. HMM	Nonlinear Pred. HMM	Jointly Pred. HMM
1	80.9%	88.1%	89.3 %	89.3 %
2	85.7%	84.5%	92.8 %	97.6 %
3	91.7%	91.7%	96.4 %	100.0 %
Ave.	86.1%	88.1%	92.8 %	95.6%

Table 1: Comparative recognition accuracy on CV syllables for HMM recognizers using standard HMM and various forms of speech-frame prediction implemented with a layered neural network architecture.

References

[1] L.E. Baum. "An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes", *Inequalities*, vol. 3, pp. 1–8, 1972.

[2] G.E.P. Box and G.M. Jenkins. *Time Series Analysis—Forecasting and Control*, Holden-Day, San Francisco, CA, pp. 67–72, 1976.

[3] G. Cybenko. "Approximation by superpositions of a Sigmoidal function," *Mathematics of Control, Signals, and Systems,* Vol. 2, June, 1989, pp. 303-314.

[4] L. Deng, P. Kenny, M. Lennig, and P. Mermelstein. "Phonemic hidden Markov models with continuous mixture output densities for large vocabulary word recognition," *IEEE Trans. Signal Processing*, Vol. 39, No. 7, July, 1991, pp. 1677–1681.

[5] L. Deng, M. Lennig, and P. Mermelstein. "Modeling microsegments of stop consonants in a hidden Markov model based word recognizer," *J. Acoust. Soc. Am.*, Vol. 87, June, 1990, pp. 2738-2747.

[6] L. Deng, M. Lennig, F. Seitz and P. Mermelstein. "Large vocabulary word recognition using context-dependent allophonic hidden Markov models," *Computer Speech and Language*, Vol. 4, No. 4, 1990, pp. 345-357.

[7] L. Deng and K. Erler. "Microstructural speech units and their HMM representation for discrete utterance speech recognition," *Proc. IEEE Internat. Conf. Signal Processing*, Toronto, Ontario, Canada, May, 1991, pp. 193–196.

[8] L. Deng, P. Kenny, M. Lennig, and P. Mermelstein. "Modeling acoustic transitions in speech by state-interpolation hidden Markov models," *IEEE Trans. Signal Processing*, scheduled to appear in February, 1992.

[9] G.Fant. *Acoustic Theory of Speech Production*, Mouton, The Hague, 1960.

[10] K. Hornik, M. Stinchcombe, and H. White. "Multilayer feedforward networks are universal approximators," *Neural Networks*, Vol. 2, 1989, pp. 359-366.

[11] P. Kenny, M. Lennig, and P. Mermelstein. "A linear predictive HMM for vector-valued observations with applications to speech recognition," *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. 38, No. 2, February 1990, pp.220–225.

[12] E. Levin. "Word recognition using hidden control neural architecture," *Proc. IEEE Internat. Conf. Signal Processing*, Vol. 1, Albuquerque, NM, pp. 433–436, 1990.

[13] N. Minorsky *Nonlinear Oscillations*, Chapter 9, D. Van Nostrand Company Inc., Princeton, N.J., 1962.

[14] A.B. Poritz, "Hidden Markov models: A guided tour," *Proc. IEEE Internat. Conf. Signal Processing*, Vol.1, New York, New York, April 11–14, 1988, pp. 7–13.

[15] M.B. Priestley. *Non-Linear and Non-Stationary Time Series Analysis*, Academic Press, London, England, 1988, pp. 140–174.

[16] H. Tong. *Non-Linear Time Series — A Dynamical System Approach*, Oxford University Press, New York, 1990.

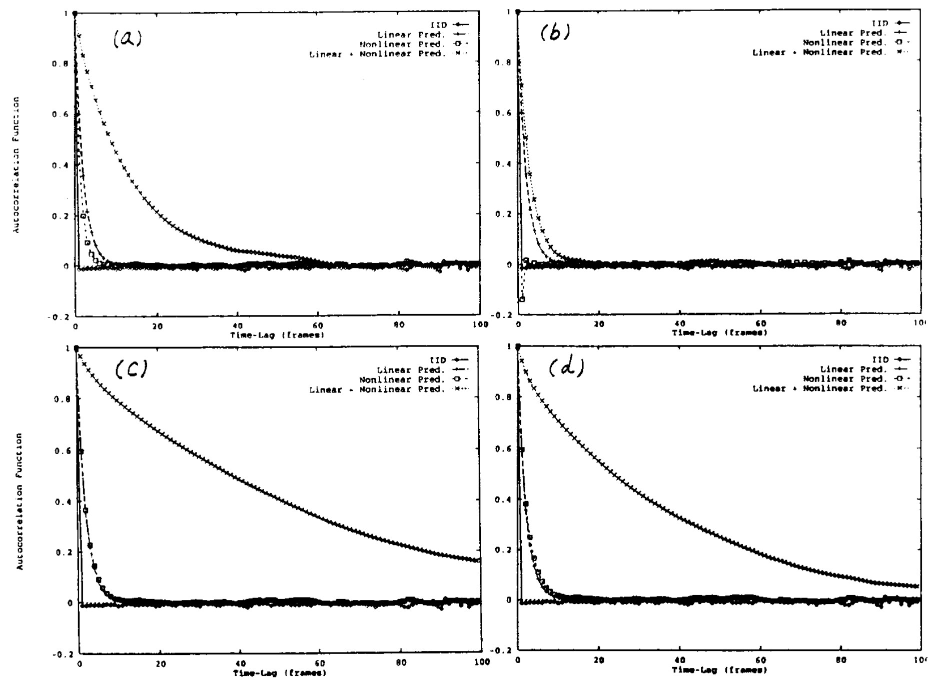

Figure 1: Comparison of autocorrelation functions for models (1), (2), (6) and an IID source. a,b,c and d are for four different forms of nonlinear functions $f(\cdot)$ in models (2) and (6): tanh, sigmoid, symmetric square root, and symmetric one-quarter power function, respectively. Parameter ϕ is fixed at 0.6.