Chapter 2
Review of Previous Work

2.1 Introduction

The last two decades have seen significant advances in human-machine interfaces. Speech and language technology, in particular - speech recognition is one among several areas, which have benefited enormously from these advances (Young, 2001). Initially, the research area of speech recognition was treated as a problem in statistical pattern recognition and classification, using small vocabularies of isolated words or digits recorded in low noise environments. Speech utterances were processed using traditional spectral techniques such as discrete Fourier transforms or filter banks, and direct classification techniques such as template matching were used to make a recognition decision. Some of the earliest systems used statistical pattern matching to do isolated digit recognition [Davis.K.H., Biddulph.R. et.al., 1952] or syllable recognition [Forgie.J.W. and Forgie.C.D, 1959],[Olson.H.F. and Belar.H.,1956].

Mathematical models for speech were developed as early as the 1940's [Dudley.H.W, 1940], based on linguistics research that viewed spoken language as the output of the filter system with the impulses from the larynx and vocal folds as the input to the system, the shape of the vocal tract representing the filter parameters, and the speech waveform as the system's output (Source filter Model) [Flanagan.J.L, 1965].
In the late 1960's and early 1970's, speech research became increasingly focused on a few key areas: feature selection and analysis, and template-based classification techniques targeted for speech data. Features such as cepstral coefficients and linear prediction coefficients [Thomas Parsons, 1987], [Rabiner.L and Juang.B.H, 1993] enabled the excitation portion of the speech waveform to be modeled and removed, leaving the vocal tract information relatively intact. Classification techniques such as Dynamic Time Warping (DTW) allowed for a temporally-motivated, non-linear mapping between speech inputs and templates, and resulted in an excellent method for measuring perceptual similarity [Itakura.F, 1975], [Vintsyuk.T.K, 1968]. From mid 1980 onwards, almost all speech research has involved using the Hidden Markov Model (HMM) technique. In the late 1980s, neural networks were also introduced to problems in speech recognition as a signal classification technique.

Source-filter models form the foundation of many speech processing applications such as speech coding, speech synthesis, speech recognition, and speaker recognition technology. Usually, the filter is linear and based on linear predictions. It neglects nonlinear structure known to be present in the speech production mechanism. While this approach has led to great advances in the last 30 years, a fully automatic speech-based interface to products, which would encompass real-time speech processing as well as language understanding, is still considered to be many years away. The replacement of the linear filter with nonlinear operators (models) should enable us to obtain
an accurate description of the speech. This in turn may lead to better performance of practical speech processing applications.

This chapter presents a review of previous works in the area of nonlinear speech processing and the applications of neural network for speech recognition and is organized as follows. Section 2.2 provides a summary of research findings in the area of nonlinear speech processing. Section 2.3 gives a review of previous works in the applications of neural network for speech recognition. Finally section 2.4 concludes this review.

2.2 Review of Previous works in Nonlinear Speech Processing

Nonlinear methods for speech processing are a rapidly growing area of research. Naturally, it is difficult to define a precise date for the origin of the field, but it is clear that there was a rapid growth in this area, which started in the mid-nineteen eighties. Since that time, numerous techniques were introduced for nonlinear time series analysis, which are ultimately aimed at engineering applications.

Among the nonlinear dynamics community, a budding interest has emerged in the application of theoretical results to experimental time series data analysis in 1980’s. One of the profound results established in chaos theory is the celebrated Takens’ embedding theorem. Takens’ theorem states that under certain assumptions, phase space of a dynamical system can be reconstructed through the use of time-delayed versions of the original scalar measurements. This new state space is commonly referred to in the
literature as Reconstructed Phase Space (RPS), and has been proven to be
topologically equivalent to the original phase space of the dynamical system.

the concept of phase space reconstruction in 1980. Soon after, Takens showed
that a delay-coordinate mapping from a generic state space to a space of
higher dimension preserves topology [Takens.F, 1980]. Sauer and Yorke have
modified Taken’s theorem to apply for experimental time series data analysis

Conventional linear digital signal processing techniques often utilize
the frequency domain as the primary processing space, which is obtained
through the Discrete Fourier Transform (DFT) of a time series. For a linear
dynamical system, structure appears in the frequency domain that takes the
form of sharp resonant peaks in the spectrum. However for a nonlinear or
chaotic system, structure does not appear in the frequency domain, because
the spectrum is usually broadband and resembles noise. In the RPS, a
structure emerges in the form of complex, dense orbits that form patterns
known as attractors. These attractors contain the information about the time
evolution of the system, which means that features derived from a RPS can
potentially contain more or different information.

The majority of literature that utilizes a RPS for signal processing
applications revolves around its use for control, prediction, and noise
reduction, reporting both positive and negative results. There is only scattered
research using RPS features for classification and/or recognition experiments.
In contrast to the linear source-filter model for speech production process, a large number of research works are reported in the literature to show the nonlinear effects in the physical process. Koizumi.T, Taniguchi.S, et.al. in 1985 showed that the vocal tract and the vocal folds do not function independently of each other, but that there is in fact some form of coupling between them when the glottis is open [Koizumi.T, Taniguchi.S, et.al., 1985]. This can cause significant changes in formant characteristics between open and closed glottis cycles. [Brookes.D.M and Naylor.P.A, 1988].

Teager and Teager [Teager.H.M and Teager.S.M, 1989] have claimed that voiced sounds are characterised by highly complex airflows in the vocal tract, rather than well behaved laminar flow. Turbulent flow of this nature is also accepted to occur during unvoiced speech, where the generation of sound is due to a constriction at some point in the vocal tract. In addition, the vocal folds will themselves be responsible for further nonlinear behaviour, since the muscle and cartilage, which comprise the larynx, have nonlinear stretching qualities.

Such nonlinearities are routinely included in attempts to model the physical process of vocal fold vibration, which have focussed on two or more mass models [Ishizaka.K and Flanagan.J.L, 1972], [Koizumi.T, Taniguchi.S, et.al., 1987] [Steinecke.I and Herzel.H, 1995] in which the movement of the vocal folds is modelled by masses connected by springs, with nonlinear coupling. Observations of the glottal waveform reinforce this evidence, where it has been shown that this waveform can change shape at different
amplitudes [Shoentgen.J, 1990]. Such a change would not be possible in a strictly linear system where the waveform shape is unaffected by amplitude changes.

Extraction of invariant parameters from speech signal has attracted researchers for designing speech and speaker recognition systems. In 1988, Narayanan.N.K. et.al. [Narayanan.N.K. and Sridhar.C.S, 1988] used the dynamical system technique mentioned in the nonlinear dynamics to extract invariant parameters from speech signal. The dynamics of speech signal is experimentally investigated by extracting the second order dimension of the attractor $D_2$ and the second order Kolmogorov entropy $K_2$ of speech signal. The fractal dimension of $D_2$ and non-zero value of $K_2$ confirms the contribution of deterministic chaos to the behavior of speech signal. The attractor dimension $D_2$ and Kolmogorov entropy $K_2$ are then used as a powerful tool for voiced / unvoiced classification of speech signals.

The dimension of the trajectories, or the dimension of the attractor is an important characteristic of the dynamic systems. The estimation of the dimension gives a lower bound of the number of parameters needed in order to model the system. The goal is to find if the system under study occupies all the state space or if it is most of the time in a subset of the space, called attractor. The correlation dimension [Tishby, 1990] is a practical method to estimate the dimension of an empirical temporal series.

There are a large variety of techniques found in the literature of nonlinear methods and it is difficult to predict which techniques ultimately
will be more successful in speech processing. However, commonly observed methods in the speech processing literature are various forms of oscillators and nonlinear predictors, the latter being part of the more general class of nonlinear autoregressive methods. The oscillator and autoregressive techniques themselves are also closely related since a nonlinear autoregressive model in its synthesis form forms a nonlinear oscillator if no input is applied. For the practical design of a nonlinear autoregressive model, various approximations have been proposed [Farmer.J.D and Sidorowich.J.D, 1988], [Casdagli.M, Des Jardins. D, et.al., 1992], [Abarbanel.H.D.I, Brown.R, et.al., 1993], [Kubin.G, 1995]. These can be split into two main categories: parametric and nonparametric methods.


Phase space reconstruction is usually the first step in the analysis of dynamical systems. An experimenter obtains a scalar time series from one observable of a multidimensional system. State-space reconstruction is then needed for the indirect measurement of the system’s invariant parameters like, dimension, Lyapunov exponent etc. Takens’ theorem gives little guidance, about practical considerations for reconstructing a good state space. It is silent on the choice of time delay (τ) to use in constructing m-dimensional data vectors. Indeed, it allows any time delay as long as one has an infinite amount of infinitely accurate data. However, for reconstructing state spaces from real-world, finite, noisy data, it gives no direction [Casdagli. M, Eubank.S, et. al., 1991]. Two heuristics have been developed in the literature for establishing a time lag [Kantz.H and Schreiber.T, 2003]. 1) The first zero of the
autocorrelation function and 2) the first minimum of the auto mutual information curve [Fraser.A.M and Swinney.H.L, 1986].

In their work, Andrew M Fraser and Harry L Swinney, the mutual information is examined for a model dynamical system and for chaotic data from an experiment on the Belousov-Zhabotinskii reaction. An N log N algorithm for calculating mutual information (I) is presented. A minimum in ‘I’ is found to be a good criterion for the choice of time delay in Phase Space Reconstruction from time series data. This criterion is shown to be far superior to choosing a zero of the autocorrelation function.

There have been many discussions on how to determine the optimal embedding dimension from a scalar time series based on Taken’s theorem or its extensions [Sauer.T, Yorke.J.A., and Casdagli. M, 1991]. Among different geometrical criteria, the most popular seems to be the method of False Nearest Neighbors [Kennel.M. B, Brown. R, and Abarbanel.H.D.I, 1992]. This criterion concerns the fundamental condition of no self-intersections of the reconstructed attractor.

Work by Banbrook, McLaughlin et. al.[Banbrook.M and McLaughlin.S, 1994], Kumar et. al. [Kumar.A and Mullick.S.K, 1996], and Narayanan et. al. [Narayanan.S.S and Alwan.A.A, 1995] has attempted to use nonlinear dynamical methods to answer the question: “Is speech chaotic?” These papers focused on calculating theoretical quantities such as Lyapunov exponents and Correlation dimension. Their results are largely inconclusive and even contradictory. A synthesis technique for voiced sounds is developed

In a work presented by Langi and Kinsner speech consonants are characterised by using a fractal model for speech recognition systems. [Langi.A and Kinsner.W, 1995] Characterization of consonants has been a difficult problem because consonant waveforms may be indistinguishable in time or frequency domain. The approach views consonant waveforms as coming from a turbulent constriction in a human speech production system, and thus exhibiting turbulent and noise like time domain appearance. However, it departs from the usual approach by modeling consonant excitation using chaotic dynamical systems capable of generating turbulent and noise-like excitations. The scheme employs correlation fractal dimension and Takens embedding theorem to measure fractal dimension from time series observation of the dynamical systems. It uses linear predictive coding (LPC) excitation of twenty-two consonant waveforms as the time series. Furthermore, the correlation fractal dimension is calculated using a fast Grassberger algorithm [Grassberger and Procaccia, 1983].

Wei Gang, Lu Yiqing et al. presented a new method for low bit rate speech coding method based on fractal code excited linear prediction [Wei Gang, Lu Yiqing and Quyang Jingzheng, 1996]. Based on the recently developed chaos and fractal theories they introduced new methods for speech signal processing. A novel phase space reconstruction algorithm is proposed for speech signal, the distributions of the maximum Lyapunov exponent and
the fractal dimension of speech signal are tested and analyzed statistically. The results of this study indicate that chaos and fractal theories have great potentials in the field of speech signal processing.

The criterion in the False Nearest Neighbor approach for determining optimal embedding dimension is subjective in some sense that, different values of parameters may lead to different results [Cao.L, 1997]). For realistic time series data, different optimal embedding dimensions are obtained if we use different values of the threshold value. Also with noisy data this method gives spurious results. [Kantz.H and Schreiber.T, 2003].

Liangyue Cao in 1997 proposed a practical method to determine the minimum embedding dimension from a scalar time series. It does not contain any subjective parameters except for the time delay for the embedding. It does not strongly depend on how many data points are available and it is computationally efficient. Several time series are tested to show the above advantages of the method [Cao.L, 1997]).

Petry et. al.[Petry.A, Augusto.D et.al., 2002] and Pitsikalis et. al. [Pitsikalis.V and Maragos.P, 2002] have used Lyapunov exponents and Correlation dimension in unison with traditional features (cepstral coefficients) and have shown minor improvements over baseline speech recognition systems. Central to both sets of these papers is the importance of Lyapunov exponents and Correlation dimension, because they are invariant metrics that are the same regardless of initial conditions in both the original and reconstructed phase space. Despite their significance, there are several
issues that exist in the measuring of these quantities on real experimental data. The most important issue is that these measurements are very sensitive to noise. Secondarily, the automatic computation of these quantities through a numerical algorithm is not well established and this can lead to drastically differing results. The overall performance of these quantities as salient features remains an open research question.

In addition to these speech analysis and recognition applications, nonlinear methods have also been applied to speech enhancement, speech coding etc. Papers by Hegger et al. [Hegger.R, Kantz.H, et.al. 2000], [Hegger.R, Kantz.H, et.al. 2001] demonstrated the successful application of what is known as local nonlinear noise reduction to sustained vowel utterances.

Michael T. Johnson et.al. proposed the implementation of two nonlinear noise reduction methods applied to speech enhancement [Michael T. Johnson, Andrew C. Lindgren, et.al, 2003]. The methods are based on embedding the noisy signal in a high-dimensional reconstructed phase space and applying singular value decomposition to project the signal into a lower dimension. The advantages of these nonlinear methods include that they do not require explicit models of noise spectra and do not have the typical ‘musical tone’ side effects associated with traditional linear speech enhancement methods.

In a work presented by Jinjin Ye et.al. Principal Component Analysis (PCA) is applied to feature vectors from the reconstructed phase space [Jinjin
Ye, Michael T. Johnson, et.al., 2003]. By using PCA projection, the basis of the feature space is orthogonalized. A Bayes classifier uses the transformed feature vectors to classify phonemes. The results show that the classification accuracy with PCA method surpasses the accuracy using only original features in most cases. PCA projection was implemented in three ways over the reconstructed phase space on both speaker-dependent and speaker-independent data.

Kevin M Lindrebo et.al. introduced a method for calculating speech features from third-order statistics of sub band filtered speech signals which are used for robust speech recognition [Kevin M. Indrebo, Richard J. Povinelli, et.al., 2005]. These features have the potential to capture nonlinear information not represented by cepstral coefficients. Also, because the features presented in this method are based on the third-order moments, they may be more immune to Gaussian noise than cepstrals, as Gaussian distributions have zero third-order moments.

Richard J Povinelli et.al. introduced a novel approach to the analysis and classification of time series signals using statistical models of reconstructed phase spaces [Povinelli.R.J, Michael T. Johnson, et.al., 2006]. With sufficient dimension, such reconstructed phase spaces are, with probability one, guaranteed to be topologically equivalent to the state dynamics of the generating system, and, therefore, may contain information that is absent in analysis and classification methods rooted in linear assumptions. Parametric and nonparametric distributions are introduced as
statistical representations over the multidimensional reconstructed phase space, with classification accomplished through methods such as Bayes maximum likelihood and artificial neural networks (ANNs). The technique is demonstrated on heart arrhythmia classification and speech recognition. This new approach is shown to be a viable and effective alternative to traditional signal classification approaches, particularly for signals with strong nonlinear characteristics.

In a recent study Marcos Faundez-Zanuy compared the identification rates of a speaker recognition system using several parameterizations, with special emphasis on the residual signal obtained from linear and nonlinear predictive analysis [Marcos Faundez-Zanuy, 2007]. It is found that the residual signal is still useful even when using a high dimensional linear predictive analysis. If instead of using the residual signal of a linear analysis a nonlinear analysis is used, both combined signals are more uncorrelated and although the discriminative power of the nonlinear residual signal is lower, the combined scheme outperforms the linear one for several analysis orders.

It is seen that the majority of literature that utilizes the nonlinear techniques for signal processing applications revolves around its use for control, prediction and noise reduction, reporting both positive and negative results. There is only scattered research using these methods for classification or recognition experiments. It is also important to notice that no work has been reported yet in nonlinear speech processing for Malayalam and other
Indian languages. The succeeding session of this chapter is focussed on the review of the applications of neural network for speech recognition.

2.3 Review of the applications of neural network for speech recognition

Artificial neural net (ANN) algorithms have been designed and implemented for speech pattern recognition by a number of researchers. ANNs are of interest because algorithms used in many speech recognizers can be implemented using highly parallel neural net architectures and also because new parallel algorithms are being developed making use of the newly acquired knowledge of the working of biological nervous systems. Hutton.L.V compares neural network and statistical pattern comparison method for pattern recognition purpose [Hutton.L.V 1992]. Neural network approaches to pattern classification problems complement and compete with statistical approaches. Each approach has unique strengths that can be exploited in the design and evaluation of classifier systems. Classical (statistical) techniques can be used to evaluate the performance of neural net classifiers, which often outperform them. Neural net classifiers may have advantages even when their ultimate performance on a training set can be shown to be no better than the classical. It is possible to be implemented in real time using special purpose hardware.

Personnaz L.and Dreyfus G presents an elementary introduction to networks of formal neurons [Personnaz L and. Dreyfus.G 1990]. The state of the art regarding basic research and the applications are presented in this work. First, the most usual models of formal neurons are described, together
with the most currently used network architectures: static (feedforward) nets and dynamic (feedback) nets. Secondly, the main potential applications of neural networks are reviewed: pattern recognition (vision, speech), signal processing and automatic control. Finally, the main achievements (simulation software, simulation machines, integrated circuits) are presented.

William Huang et. al. presents some neural net approaches for the problem of static pattern classification and time alignment [William Huang et. al., 1988]. For static pattern classification multi layer perceptron classifiers trained with back propagation can form arbitrary decision regions, are robust, and are trained rapidly for convex decision regions. For time alignment, the Viterbi net is a neural net implementation of the Viterbi decoder used very effectively in recognition systems based on Hidden Markov Models (HMMs).

Waibel.A et. al. proposed a time delay neural network (TDNN) approach to phoneme recognition, which is characterized by two important properties [Waibel.A et. al., 1988]. Using a three level arrangement of simple computing units, it can represent arbitrary non-linear decision surface. The TDNN learns these decision surfaces automatically using error back propagation. The time delay arrangement enables the network to discover acoustic phonetic features and temporal relationships between them independent of position in time and hence not blurred by temporal shifts in the input. For comparison, several discrete Hidden Markov Models (HMM) were trained to perform the same task, i.e. the speaker dependent recognition of the phonemes “B”, “D” and “G” extracted from varying phonetic contexts.
The TDNN achieved a recognition rate of 98.5% correct compared to 93.7% for the best of HMMs. They show that the TDNN has well known acoustic–phonetic features (e.g., F2-rise, F2-fall, vowel-onset) as useful abstractions. It also developed alternate internal representations to link different acoustic realizations to the same concept.

Yoshua Bengio and Renato De Mori used the Boltzmann machine algorithm and the error back propagation algorithm to learn to recognize the place of articulation of vowels (front, center or back), represented by a static description of spectral lines [Yoshua Bengio and Renato De Mori, 1988]. The error rate is shown to depend on the coding. Results are comparable or better than those obtained by them on the same data using hidden Markov Models. They also show a fault tolerant property of the neural nets, i.e. that the error on the test set increases slowly and gradually when an increasing number of nodes fail.

Moore. K.L discussed different types of neural network in his paper entitled “Artificial neural networks” [Moore. K.L, 1992]. Three different tasks for which they are suitable are discussed. They are pattern classification and associative memory, self-organization and feature extraction, and optimization.

relate the number of input, output and hidden nodes to the problem features and parameters. In particular, each hidden neuron corresponds to a discriminant in the input space. They point out that the interactions between number of discriminant, the size and distribution of the training set, and numerical magnitudes make it very difficult to provide precise guidelines. They found that the shape of the threshold function plays a major role in both pattern recognition, and quantitative prediction and interpolation. Tuning the sharpness parameter could have a significant effect on neural network performance. This feature is currently under-utilized in many applications. For some applications linear discriminant is a poor choice.

Janssen. R.D.T, Fanty. M and Cole. R.A developed a phonetic front-end for speaker-independent recognition of continuous letter strings [Janssen. R.D.T, Fanty.M and Cole.R.A, 1991]. A feedforward neural network is trained to classify 3 msec speech frames as one of the 30 phonemes in the English alphabet. Phonetic context is used in two ways: first, by providing spectral and waveform information before and after the frame to be classified, and second, by a second-pass network that uses both acoustic features and the phonetic outputs of the first-pass network. This use of context reduced the error rate by 50%. The effectiveness of the DFT and the more compact PLP (perceptual linear predictive) analysis is compared, and several other features, such as zerocrossing rate, are investigated. A frame-based phonetic classification performance of 75.7% was achieved.
Ki-Seok-Kim and Hee-Yeung-Hwang present the result of the study on the speech recognition of Korean phonemes using recurrent neural network models conducted by them [Ki-Seok-Kim and Hee-Yeung-Hwang, 1991]. The results of applying the recurrent multi layer perceptron model for learning temporal characteristics of speech phoneme recognition, is presented. The test data consist of 144 vowel+consonant+vowel (VCV) speech chains made up of 4 Korean monothongs and 9 Korean plosive consonants. The input parameters of the artificial neural network model used are the FFT coefficients, residual error and zero crossing rates. The baseline model showed a recognition rate of 91% for vowels and 71% for plosive consonants of one male speaker. The authors obtained better recognition rates from various other experiments compared to the existing multilayer perceptron model, thus showing the recurrent model to be better suited to speech recognition. The possibility of using the recurrent models for speech recognition was experimented upon by changing the configuration of this baseline model.

Ahn.R and Holmes.W.H propose a voiced / unvoiced / silence classification algorithm of speech using 2-stage neural networks with delayed decision input [Ahn .R and Holmes.W.H, 1996]. This feed forward neural network classifier is capable of determining voiced, unvoiced and silence in the first stage and refining unvoiced and silence decisions in the second stage. Delayed decision from the previous frame's classification along with preliminary decision by the first stage network, zero crossing ratio and
energy ratio enable the second stage to correct the mistakes made by the first stage in classifying unvoiced and silence frames. Comparisons with a single stage classifier demonstrate the necessity of two-stage classification techniques. It also shows that the proposed classifier performs excellently.

Sunilkumar.R.K and Narayanan.N.K investigated the potential use of zero-crossing based information of the signal for Malayalam vowel recognition [Sunilkumar.R.K. 2002]. A vowel recognition system using artificial neural network is developed. The highest recognition accuracy obtained for normal speech is 90.62%.

Dhananjaya.N, Guruprasad.S, et.al. proposed a method for detecting speaker changes in a multi-speaker speech signal [Dhananjaya.N, Guruprasad.S, et.al., 2004]. The statistical approach to a point phenomenon (speaker change) fails when the given conversation involves short speaker turns (< 5 sec duration). They used auto associative neural network (AANN) models to capture the characteristics of the excitation source that present in the linear prediction (LP) residual of speech signal. The AANN models are then used to detect the speaker changes.

Xavier Domont, Martin Heckmann, et.al. proposed a feed forward neural network for syllable recognition [Xavier Domont, Martin Heckmann, et.al., 2007]. The core of the recognition system is based on a hierarchical architecture initially developed for visual object recognition. In this work, they showed that, given the similarities between the primary auditory and visual cortices, such a system can successfully be used for speech
recognition. Syllables are used as basic units for the recognition. Their spectrograms, computed using a Gammatone filter bank, are interpreted as images and subsequently feed into the neural network after a preprocessing step that enhances the formant frequencies and normalizes the length of the syllables.

In a recent work by Ana I. Garcia Moral, Ruben Solera Urena, et.al., the solutions provided in the past for Artificial Neural Network are recalled and applied them to Support Vector Machines (SVM), performing a comparison between them [Ana I. Garcia Moral, Ruben Solera Urena, et.al., 2007]. Support Vector Machines are state-of-the-art methods for machine learning but share with more classical ANN the difficulty of their application to temporally variable input patterns. Preliminary results are encouraging.

2.4 Conclusion

This chapter provides a summary about the recent advances, new trends and important contributions in the area of nonlinear speech processing and the applications of neural network for speech recognition. Nonlinear signal processing techniques have several potential advantages over traditional linear signal processing methodologies. They are capable of recovering the nonlinear dynamics of the signal of interest possibly preserving natural information. They are not constrained by strong linearity assumptions. Despite these facts, the use of nonlinear signal processing techniques also have disadvantages as well, which is why they have not been widely used in the past. Primarily, they are not as well understood as conventional linear
methods. Salient features of Reconstructed Phase Space for classification or recognition have yet to be firmly established, and this work is clearly in the early stages, which is what motivates this research pursuit. Moreover, researchers have just begun to study nonlinear signal processing techniques for a variety of engineering tasks with mixed success. The use of nonlinear methodologies especially as it applies to speech recognition is truly in its infancy with very little work published, most of which has been in the last five years.