Exploiting Nonlinearity in Adaptive Signal Processing

Phebe Vayanos, Mo Chen, Beth Jelfs, and Danilo P. Mandic

Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK
{foivi.vayanos,mo.chen,beth.jelfs,d.mandic}@ic.ac.uk

Abstract. *Quantitative* performance criteria for the analysis of machine learning architectures and algorithms have been long established. However, the *qualitative* performance criteria, *e.g.*, nonlinearity assessment, are still emerging. To that end, we employ some recent developments in signal characterisation and derive criteria for the assessment of the changes in the nature of the processed signal. In addition, we also propose a novel online method for tracking the system nonlinearity. A comprehensive set of simulations in both the linear and nonlinear settings and their combination supports the analysis.

1 Introduction

Real–world processes are typically mixtures of linear and nonlinear signal components (which can be either deterministic or stochastic) and noise, yet it is a common practice to process them using linear, mathematically tractable, models. To illustrate the need to asses the *nature of a real world signal* prior to choosing the actual computational model, Figure 1 (modified from [1]), shows the range spanned by the fundamental signal properties of "nonlinear" and "stochastic". Despite the fact that real-world processes, due to nonlinearity, uncertainty and noise, are located in areas such as those denoted by (a), (b), (c) and '?', in terms of computational models, only the very specialised cases such as the linear-stochastic autoregressive moving average (ARMA), and chaotic (nonlinear-deterministic) models are well understood. It is therefore necessary to verify the presence of an underlying linear or nonlinear signal generation system, *before* the actual filters or models are constructed. Indeed, in the absence of nonlinearity within a signal in hand, it is not advantageous to use nonlinear models since these are more difficult to train than their linear counterparts, due to issues such as overfitting and computational complexity.

Research on "signal modality characterisation" started in physics in the mid 1990s and its applications in machine learning and signal processing application are just beginning to emerge. It is essential that during processing of such signals we not only optimise for the "best" performance in terms of a certain quantitative performance criterion, but also that the processing preserves the desired fundamental properties of the signal, for instance, the nonlinear and deterministic nature (qualitative performance). If the desired signal property has

M. Chetouani et al. (Eds.): NOLISP 2007, LNAI 4885, pp. 57–77, 2007.

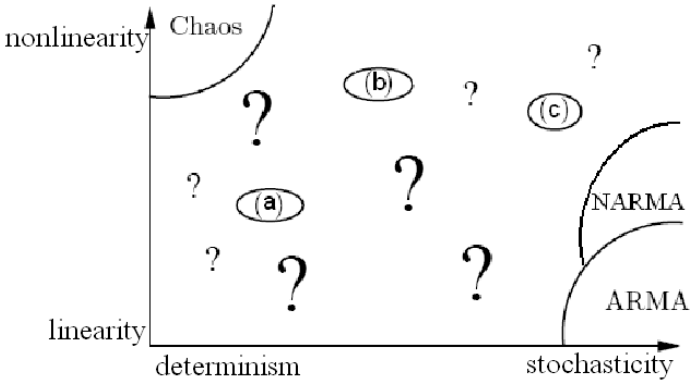


Fig. 1. Sketch of the variety of signals spanned by the properties "nonlinearity" and "stochasticity". Areas where theoretical knowledge and technology for the analysis of time series are available are outlined, such as "Chaos" and "ARMA". (Modified from (Schreiber, 1999)).

significantly changed after processing (*e.g.* prediction within compression algorithms and denoising), the application of such filters will be greatly limited. As a consequence, such architectures and algorithms will not be suitable in situations where the signal nature is of critical importance, for instance, in speech processing.

Notice that the very core of adaptive learning is the change in the shapes of signal spectrum, this reflects only the performance in terms of second order statistics, and no account of other signal characteristics is provided. To that cause, we propose a new framework for the assessment of *qualitative performance* in machine learning, and set out to investigate whether an improvement in the quantitative performance is necessarily followed by the improvement in the qualitative performance. For generality and to illustrate this trade-off, this is achieved for both the linear and nonlinear filters.

On the other hand, the existing signal modality characterisation algorithms in this area are typically based on hypothesis testing [2,3,4] and describe the signal changes in a statistical manner. However, there are very few online algorithms which are suitable for this purpose. Therefore, in this chapter, we will also propose to demonstrate the possibility of online algorithms which can be used not only to identify the nature of the signal, but also to track changes in the nature of the signal (signal modality detection).

2 Background Theory

Before introducing new criteria for the analysis of qualitative performance in machine learning and the online algorithm for tracking system nonlinearity, we set out to provide some necessary background focusing on some recent results on signal characterisation.

2.1 "Nature" of a Signal

By the signal 'nature' [5][6], we adhere to a number of signal properties such as:-

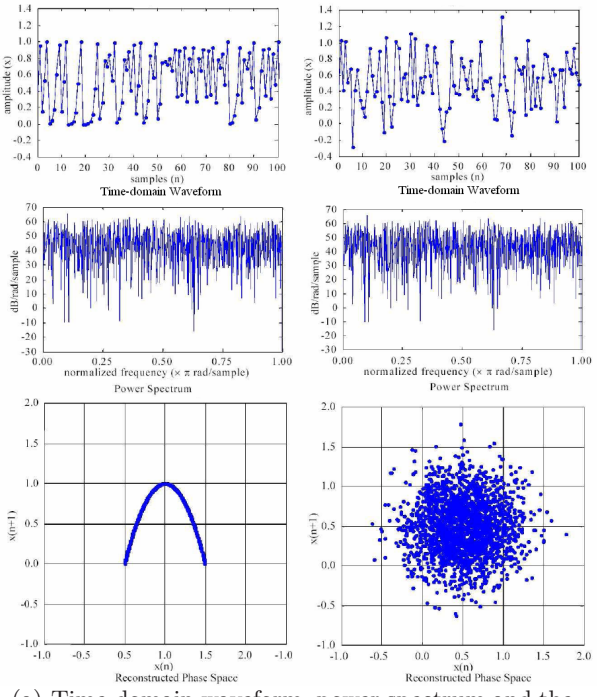
i) Linear (strict definition) – A linear signal is generated by a linear time-invariant system, driven by white Gaussian noise;
ii) Linear (commonly adopted) – Definition i) is relaxed somewhat by allowing the distribution of the signal to deviate from the Gaussian one, which can be interpreted as a linear signal from i), measured by a static (possibly nonlinear) observation function;
iii) Nonlinear – A signal that cannot be generated in the above way is considered nonlinear;
iv) Deterministic (predictable) – A signal is considered deterministic if it can be precisely described by a set of equations;
v) Stochastic – A signal that is not deterministic.

2.2 Method of Surrogate Data and the Concept of "Phase Space"

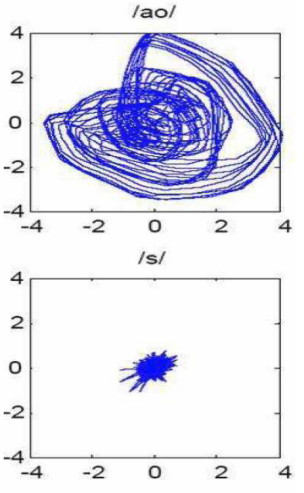
Research on signal nonlinearity detection started in physics in the 1990s, and out of the several proposed methods, the so-called 'surrogate data' method, introduced by Theiler *et al.* [7], has been extensively used in the context of statistical nonlinearity testing. A surrogate time series, or 'surrogate' for short, is a realisation of a 'composite' null hypothesis. In our case this null hypothesis is that the original signal is linear, *i.e.*, generated by a linear stochastic system driven by white Gaussian noise, measured by a static, monotonic and possibly nonlinear observation function. Then, a discriminating statistic is calculated for both the original time series and a set of surrogate data. If the statistics for the original time series do not lie in the range of those for the surrogate data, the null hypothesis is rejected, and the original data is judged to be nonlinear, otherwise, it is judged to be linear. There exist many discriminating statistics, the commonly used ones include the so-called third-order auto-covariance (C3) [4] and the asymmetry due to time reversal (REV) [4]. In order to increase the power of the surrogate test and decrease the spurious rejections of the null hypothesis, several modified methods for the generation of surrogate data have been proposed. In this chapter, we adopt the iterative amplitude adjusted Fourier Transform (iAAFT) surrogate method [8]. The iAAFT surrogate data have their amplitude spectra similar and their amplitude distribution identical to those of the original time series.

Techniques described in this chapter rest upon the method of time delay embedding for representing a time series in so-called 'phase space', *i.e.*, by a set of delay vectors (DVs) $\mathbf{x}(k)$ of a given embedding dimension m, that is $\mathbf{x}(k) = [x_{k-m\tau}, \ldots, x_{k-\tau}]^T$, where τ is a time lag, which for simplicity is set to unity in all simulations. In other words, $\mathbf{x}(k)$ is a vector containing m consecutive time samples.

From Figure 2(a), although the wave form and the power spectrum of the two signals are similar to one another, distinct difference can be observed in two



(a) Time-domain waveform, power spectrum and the phase space scatter plot of a chaotic signal (left column) and its surrogate (right column)



(b) Phase space scatter plot for phonemes /ao/ and /s/

Fig. 2. Phase space scatter plot

phase space scatter plots. There is some sort of structure in the scatter plot for the chaotic signal, whereas the surrogate displays randomness in the scatter plot. The reason for this is that during the surrogate-generation process the temporal and spatial correlations were completely destroyed due to the randomisation scheme. Figure 2(b) illustrates the attractors for vowel /ao/ and consonant /s/ in phase space scatter plot. From the Figure, it is clear that these two phonemes differ from one another in nature.

2.3 Signal Characterisation Tool: "Delay Vector Variance" (DVV) Method

Many methods for detecting the nonlinear structure within a signal have been proposed, such as the above mentioned surrogate data with different choices of discriminating statistics, *"Deterministic versus Stochastic"* (DVS) plot [9], δ-ε Method [10]. For our purpose, it is desirable to have a method which is straightforward to visualise, and which makes use of some notions from nonlinear dynamics and chaos theory, *e.g.*, embedding dimension and phase space. One such method is our recently proposed "Delay Vector Variance" (DVV) method

[2], which is based upon examining the predictability of a signal in the phase space, and examines simultaneously the determinism and nonlinearity within a signal.

The DVV algorithm can be summarised in the following way:- For a given optimal embedding dimension[1] m:

- Map the original time series from time domain to a set of delay vectors (DVs) in phase space, $\mathbf{x}(k) = [x_{k-\tau m}, \ldots, x_{k-\tau}]^T$, where τ is the time lag which for convenience is set to unity in all the simulations and the corresponding target x_k;
- The mean μ_d and standard deviation σ_d are computed over all pairwise Euclidean distances between DVs, $\|\mathbf{x}(i) - \mathbf{x}(j)\| (i \neq j)$;
- The sets $\Omega_k(r_d)$ are generated such that $\Omega_k(r_d) = \{\mathbf{x}(i) | \|\mathbf{x}(k) - \mathbf{x}(i)\| \leq r_d\}$, *i.e.*, sets which consist of all DVs that lie closer to $\mathbf{x}(k)$ than a certain distance

$$r_d(n) = \mu_d - n_d\sigma_d + (n-1)\frac{2n_d\sigma_d}{N_{tv}-1}, \quad n = 1, \ldots, N_{tv} \tag{1}$$

  where N_{tv} denotes how fine the standardised distance is uniformly spaced, and n_d is a parameter controlling the span over which to perform the DVV analysis;
- For every set $\Omega_k(r_d)$, the variance of the corresponding targets, $\sigma_k^2(r_d)$, is computed. The average over all sets $\Omega_k(r_d)$, normalised by the variance of the time series, σ_x^2, yields the 'target variance', $\sigma^{*2}(r_d)$:

$$\sigma^{*2}(r_d) = \frac{\frac{1}{N}\sum_{k=1}^{N}\sigma_k^2(r_d)}{\sigma_x^2} \tag{2}$$

  We only consider a variance measurement *valid*, if the set $\Omega_k(r_d)$ contains at least $N_0 = 30$ DVs, since too few points for computing a sample variance yields unreliable estimates of the true variance. For more details, please refer to [2] [5].

For a predictable signal, the idea behind the DVV method is:- if two DVs lie close to one another in terms of their Euclidean distance, they should also have similar targets. The smaller the Euclidean distance between them, the more similar targets they have. Therefore, the presence of a strong deterministic component within a signal will lead to small target variances for small spans r_d. The minimal target variance, $\sigma_{min}^{*2} = min_{r_d}[\sigma^{*2}(r_d)]$, is a measure for the amount of noise present within the time series. Besides, the target variance σ_{min}^{*2} has an upper bound which is unity. This is because, when r_d becomes large enough, all the DVs belong to the same set $\Omega_k(r_d)$. Thus, the variance of the corresponding target of those DVs will be almost identical to that of the original time series.

In the following step, the linear or nonlinear nature of the time series is examined by performing the DVV test on both the original and a number of

[1] In this chapter, the optimal embedding dimension is calculated by Cao's method [11], since this method is demonstrated to yield robust results on various signals.

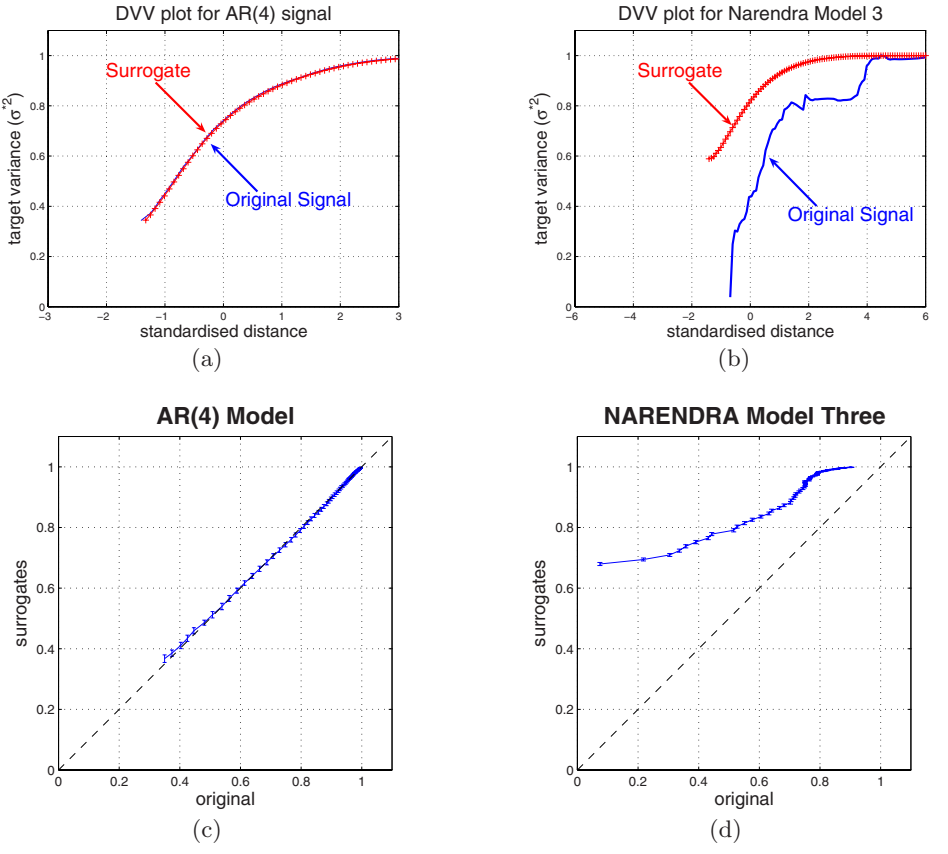


Fig. 3. Nonlinear and deterministic nature of signals. The upper two diagrams (DVV plot) are obtained by plotting the target variance as a function of standardised distance. The lower two diagrams (DVV scatter diagram) are obtained by plotting the target variance of the original data against the mean of the target variances of the surrogate data.

surrogate time series[2] [7], using the optimal embedding dimension of the original time series. Due to the standardisation of the distances, the DVV plots can be conveniently combined[3] within a *scatter diagram*, where the horizontal axis corresponds to the DVV plot of the original time series, and the vertical axis to that of the surrogate time series. If the surrogate time series yield DVV plots similar to that of the original time series, the 'DVV scatter diagram' coincides with the bisector line, and the original time series is *judged to be linear*, as shown

[2] In this chapter, all the DVV tests are performed using 19 surrogate data realisations. The reason for that is that with the increase in the number of surrogate data, DVV test does not yield a much better result whereas the computational complexity is much increased.

[3] In fact, target variance (σ^{*2}) of the original data is plotted against the mean of the target variance of 19 surrogate data, for all corresponding distances ($\frac{r_d - \mu_d}{\sigma_d}$).

in Figure 3(c). If not, the original time series is *judged to be nonlinear*, as depicted in Figure 3(d). Since the minimal target variance indicates a strong deterministic component within the signal, we conclude that in DVV scatter diagrams, the more the curve approaches the vertical axis, the more deterministic the nature of the signal. This can be employed as a convenient criterion for estimating the level of noise within a signal.

3 Qualitative and Quantitative Performance Analysis

To assess the *quantitative* performance of learning algorithms, it is convenient to use the standard one-step forward prediction gain [12]:-

$$R_p = 10 \log_{10} \left(\frac{\hat{\sigma}_s^2}{\hat{\sigma}_e^2} \right) [dB] \tag{3}$$

which is a logarithmic ratio between the estimated signal variance $\hat{\sigma}_s^2$ and estimated prediction error variance $\hat{\sigma}_e^2$. On the other hand, to assess the *qualitative* performance, that is, *a possible change in the signal nature introduced by a filter*, we proposed to compare DVV scatter diagrams of the output signals with those of the original signal. In the prediction setting, the target variances Eq. (2) for the predicted signal and its surrogates are obtained by performing the DVV test on the predicted signal. For robustness, these steps are repeated 100 times.

If the considered filters yield high prediction gain (R_p), the quantitative performance of the filters is judged to be "good". As for the qualitative performance, the more similar the DVV scatter diagram for the filtered signal is to that for the original signal, the better the qualitative performance of the considered prediction architecture.

For generality, we illustrate the usefulness of the proposed methodology for both linear and nonlinear (neural networks) adaptive filters and their combinations.

4 Experimental Settings

To illustrate the effect of the chosen mode of processing (linear, nonlinear, etc.), we have chosen a general hybrid architecture, which is shown to be able to improve the overall quantitative performance, as compared to the performance of single modules. In particular, it has been suggested that a cascaded combination of a recurrent neural network (RNN) and finite infinitive response (FIR) filter can simultaneously model the nonlinear and linear component of a signal [12]. The nonlinear neural filter can model the nonlinearity and a portion of the linearity within a signal, while the subsequent linear FIR filter models the remaining linear part of the signal.

The nonlinear neural filters used in simulations were the dynamical perceptron (nonlinear FIR filter) trained by the nonlinear gradient descent algorithm (NGD) and a recurrent perceptron, trained by the real time recursive learning (RTRL)

[13] algorithm. The linear filters considered were standard FIR filters trained by least mean square (LMS) and recursive least squares (RLS) algorithms.

The inputs were a benchmark linear AR(4) signal, given by

$$\begin{aligned} x(k) &= 1.79x(k-1) - 1.85x(k-2) + 1.27x(k-3) \\ &\quad - 0.41x(k-4) + n(k) \end{aligned} \tag{4}$$

where $n(k) \sim \mathcal{N}(0,1)$ and a benchmark nonlinear signal, the Narendra Model Three, given by [14]

$$\begin{aligned} z(k) &= \frac{z(k-1)}{1+z^2(k-1)} + r^3(k) \\ r(k) &= 1.79\, r(k-1) - 1.85\, r(k-2) + 1.27\, r(k-3) \\ &\quad - 0.41\, r(k-4) + n(k) \end{aligned} \tag{5}$$

where $n(k) \sim \mathcal{N}(0,1)$. For these signals their DVV scatter diagrams are shown as Figure 3(c) and 3(d), clearly indicating the linear nature of (4) (DVV scatter diagram on the bisector line), and nonlinear nature of (5) (DVV scatter diagram deviating from the bisector line).

4.1 Simulations

The first experiment was conducted for prediction of the linear benchmark signal (4). The DVV scatter diagrams show the nonlinearity information about the output of such filters. From Figure 4, in terms of preserving the nature of the signal (linear in this case), both the nonlinear filters and hybrid filters performed well on a linear AR(4) signal, indicated by the fact that all the DVV scatter diagrams in Figure 4 lie on the bisector line. In terms of the prediction gain R_p, the NGD and RTRL performed similarly, and as expected, the hybrid filters performed better than single nonlinear filters. The hybrid filter realised as a cascaded combination of a dynamical perceptron trained by RTRL and FIR filter trained by RLS gave the best performance, as illustrated in bottom right diagram of Figure 4.

Figure 5 illustrates a similar experiment performed on prediction of a much more complex benchmark nonlinear signal (5). From Figure 5, both nonlinear filters trained by NGD and RTRL performed poorly on their own in terms of the prediction gain. However, from the change of the nature of the original signal, seen in Figure 3(d), they preserved the nature of the benchmark nonlinear signal better than the hybrid filters, even though the quantitative gain R_p for hybrid filters was higher. For instance, the recurrent perceptron trained by the RTRL exhibited worse quantitative performance but better qualitative performance. A hybrid filter consisting of a combination of a dynamical perceptron trained by NGD and an FIR filter trained by LMS, showed a considerable increase in gain, however, the signal was considerably linearised as illustrated by the DVV scatter diagram approaching the bisector line. The bottom right diagram in Figure 5 shows the performance of a hybrid filter consisting of a recurrent perceptron

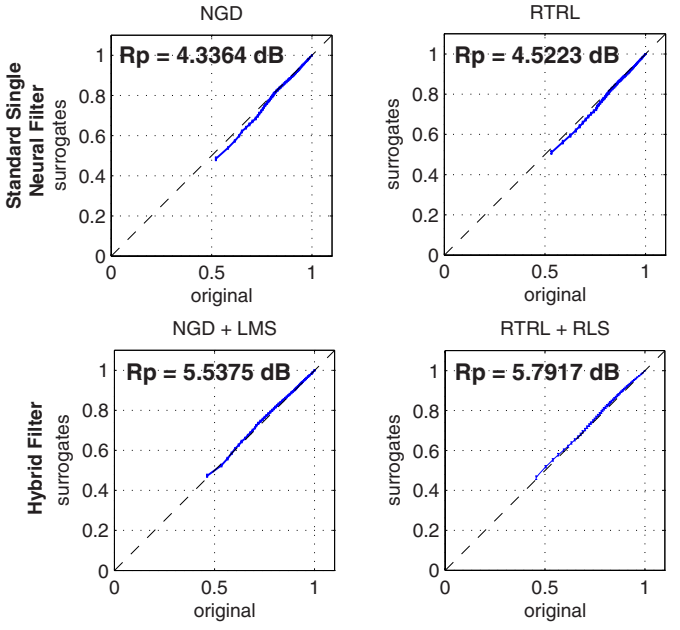


Fig. 4. Qualitative and quantitative comparison of the performance between nonlinear neural and hybrid filters for a linear benchmark signal (4). The top panels denote the DVV scatter diagrams for single neural filters (feedforward and feedback), trained by NGD and RTRL algorithm respectively. The bottom panel diagrams relate to hybrid filters.

trained by RTRL followed by a FIR filter trained by the RLS algorithm. This case gave best balance between the quantitative and qualitative performance out of all the combinations of hybrid filters considered. The quantitative performance gain for this combination was the second best of all the combinations, whereas the nature of the signal was preserved reasonably well.

We now investigate whether exchanging the order of filters within a hybrid filter will affect the overall performance. Given the highly nonlinear nature of the problem, it is expected that the performances will be significantly different. To this end we re-ran the experiments for the nonlinear benchmark signal. The results of the experiments are shown in Figure 6.

Figure 6 confirms that exchanging the order of the modules within a hybrid architecture does not provide the same performance, both quantitatively and qualitatively. Indeed, the quantitative performance are considerably worse and also the nature of the predicted signal changed significantly towards a linear one. This can be explained in the following way. When a linear filter is placed at the first stage of the hybrid filter, it linearise the input signal significantly and the subsequent nonlinear filter will not be able to recover the lost information as the system is not aware of presence of a nonlinear signal. However, if a nonlinear

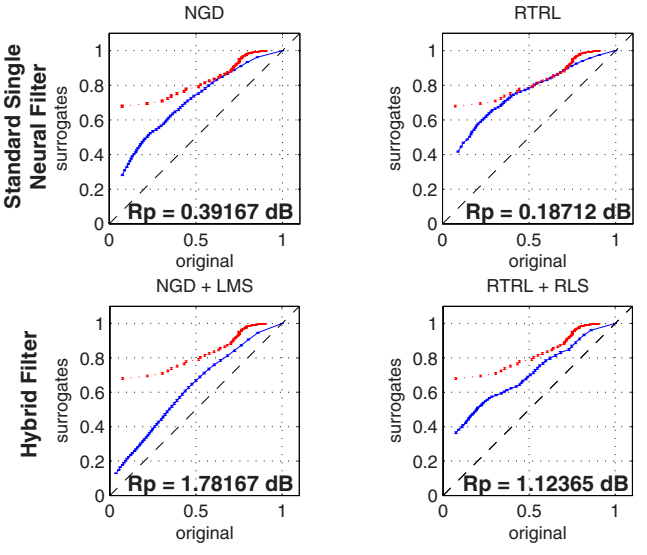


Fig. 5. Qualitative and quantitative performance comparison of the performance between nonlinear neural and linear filters for a nonlinear benchmark signal (5). The dotted line denotes the DVV scatter diagram for the original nonlinear benchmark signal Eq.(5), whereas the solid line denotes that for the output of the filters.

filter is placed as the first module, it will capture the input signal nonlinearity and inform the system that the upcoming signal is nonlinear in nature so that the subsequent linear filter is able to refine the output.

5 Online Nonlinearity Tracking Using Hybrid Filters

We have shown a novel framework of evaluating the qualitative performance of adaptive filters. This is achieved based upon examining the change in signal nature in terms of nonlinearity and determinism, which is considered an offline method. It is natural to ask whether it is possible to track the system nonlinearity online.

In [15] one such 'online' approach is considered which relies on parametric modeling to effectively "identify" the signal in hand. Figure 7 shows an implementation of this method which uses a third order Volterra filter (nonlinear subfilter) and a linear subfilter trained by the normalised LMS (NLMS) algorithm with a step size $\mu = 0.008$ to update the system parameters. The system was fed with the signal $y[k]$ obtained from

$$u[k] = \sum_{i=0}^{I} a_i x[k-i] \text{ where } I = 2 \text{ and } a_0 = 0.5, a_1 = 0.25, a_2 = 0.125 \quad (6)$$

$$y[k] = F(u[k]; k) + \eta[k] \quad (7)$$

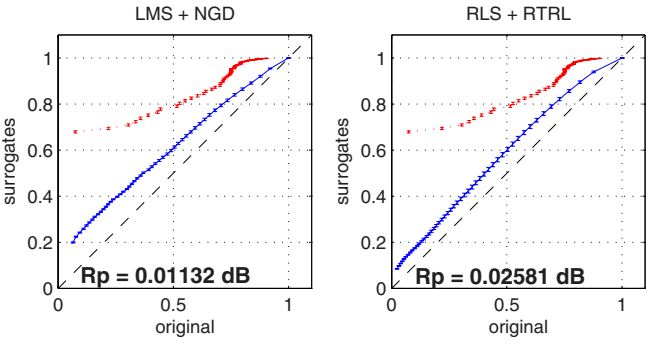


Fig. 6. Qualitative and quantitative comparison of the performance between "inverse order" hybrid filters for a nonlinear benchmark signal (5). The filter order is interchanged from the one in previous experiments. The dotted line denotes the DVV scatter diagram for the original nonlinear benchmark signal (5), whereas the solid line denotes that for the output of the filters.

where $x[k]$ are independent identically distributed and uniformly distributed over the range $[-0.5, 0.5]$ and $\eta[k] \sim \mathcal{N}(0, 0.0026)$. The function $F(u[k]; k)$ varies with the range of k as follows

$$F(u[k];k) = \begin{cases} u^3[k], \text{ for } 10000 < k \leq 20000 \\ u^2[k], \text{ for } 30000 < k \leq 40000 \\ u[k], \text{ elsewhere} \end{cases} \tag{8}$$

The signal $y[k]$ can be seen in the first trace of Figure 7. The second and third traces show the residual estimation errors of the optimal linear system and Volterra system respectively. The final trace is the estimated degree of system nonlinearity. Whilst these results show that this approach can detect changes in nonlinearity and is not affected by the presence of noise, this may be largely due to nature of the input signal being particularly suited to the Volterra model.

In this chapter, we propose to overcome this problem be making use of the concept of convexity. A convex combination can be described as [16]

$$\lambda x + (1 - \lambda)y \text{ where } \lambda \in [0, 1] \tag{9}$$

as illustrated on Figure 8. The point resulting from the convex combination of x and y will lie somewhere on the line defined by x and y, between the two. The benefits of using convex optimisation are threefold:

- The existence of the solution is guaranteed;
- The solution is unique;
- This facilitates the collaborative adaptive filtering approach

Intuitively, a convex combination of the output of two adaptive filters with different dynamical characteristics ought to be able to "follow" the subfilter with better performance, provided a suitable adaptation of λ. Indeed, such a

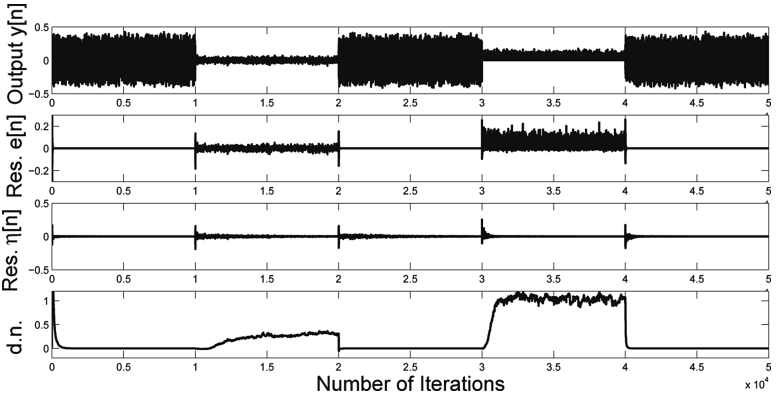


Fig. 7. NLMS with Volterra series

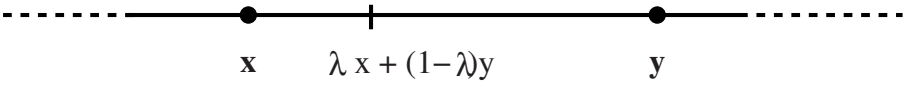


Fig. 8. Convexity

hybrid filter has been proposed in [17,18] in a form that adaptively combines the outputs of the subfilters based on their instantaneous output error. In [19], this approach has demonstrated to yield considerable improvement in the steady state and convergence capabilities of the resultant filter.

While previous applications of hybrid filters have focused on the improved performance they can offer over the individual constituent filters, our approach relies on the observation of the evolution of the so-called mixing parameter λ over time. For example, in the standard setting, λ would vary so as to initially favour the faster subfilter (learning) and finally, the filter with the best steady state properties[4].

In this section, we consider hybrid combinations of filters. The analysis of λ then provides insight into the nature of the signal under consideration. In particular, we focus on quantifying the degree of "nonlinearity" in a signal. As a subset of nonlinearity we also consider the degree of "sparsity", as sparse signals occur naturally in many real world applications. The benchmark signals considered are linear ($AR(4)$) or nonlinear by design, whereas the real world signals considered in this case are speech data.

5.1 Hybrid Adaptive Filter for Signal Modality Characterisation

Figure 9 shows a block diagram of a hybrid adaptive filter aimed at signal modality characterisation. In this Chapter, we focus on tracking the degree of

[4] This differs from the traditional "search then converge" approach, since it caters for potentially nonstationary data.

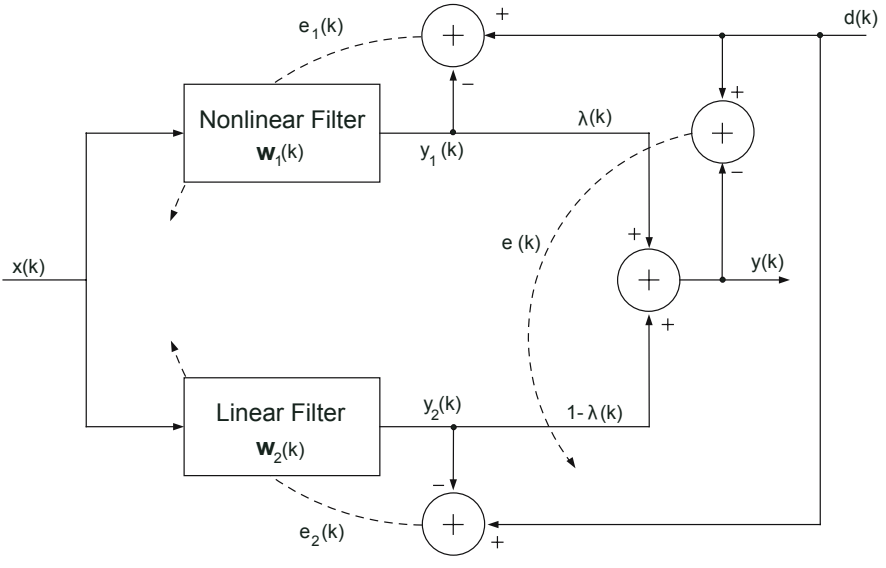


Fig. 9. Block-diagram of a hybrid adaptive filter for nonlinearity tracking

nonlinearity in a signal by combining the outputs of a linear and a nonlinear subfilter in a hybrid fashion. Following the approach from [17,19], the output of such a structure was obtained as

$$y(k) = \lambda(k)y_1(k) + [1 - \lambda(k)]\, y_2(k) \tag{10}$$

where $y_1(k) = \mathbf{x}^T(k)\mathbf{w}_1(k)$ and $y_2(k) = \mathbf{x}^T(k)\mathbf{w}_2(k)$ are the outputs of the two subfilters, with respective weight vectors $\mathbf{w}_1^T(k)$ and $\mathbf{w}_2^T(k)$ and where $\mathbf{x}(k)$ is the common input vector. For simplicity, $\mathbf{w}_1(k)$ and $\mathbf{w}_2(k)$ are assumed to be of equal length $L = 10$ and are adapted independently, using their own design rules and depending on the property we aim at tracking. Parameter $\lambda(k)$ is a mixing scalar parameter, which is adapted using a stochastic gradient rule that minimises the quadratic cost function $J(k) = e^2(k)$ of the overall filter, where $e(k)$ is the output error given by $e(k) = d(k) - y(k)$. Using LMS type adaptation to minimise the error of the overall filter, the generic form the the λ update can be written as

$$\lambda(k+1) = \lambda(k) - \mu_\lambda \, \nabla_\lambda J(k)|_{\lambda=\lambda(k)} \tag{11}$$

where μ_λ is the adaptation step-size of the hybrid filter. Using (10) and the expression for the output error, the partial derivative of the cost function with respect to $\lambda(k)$ can be written as

$$\nabla_\lambda J(k)|_{\lambda=\lambda(k)} = -e(k)[y_1(k) - y_2(k)] \tag{12}$$

Then equation (11) can be rewritten as

$$\lambda(k+1) = \lambda(k) + \mu_\lambda e(k)[y_1(k) - y_2(k)] \tag{13}$$

To ensure the combination of adaptive filters remains a convex function, λ was kept in the range $0 \leq \lambda(k) \leq 1$. For this purpose, in [17] the authors used a sigmoid function to bound $\lambda(k)$ in the range $[0, 1]$. Since, in order to determine the changes in the modality of a signal, we are not interested in the overall performance of the filter but in the variable λ, the use of a sigmoid function would interfere with the true values of $\lambda(k)$ and was therefore not used. Instead, a hard limit on the set of allowed values for λ was implemented.

5.2 Performance of the Combination on Benchmark Signals

In order to illustrate the operation of the convex combination aimed at signal modality tracking, simulations were initially performed on a set of synthetic signals made by alternating between blocks of linear and nonlinear data. 100 runs of independent trials were performed and averaged, in the one-step ahead prediction setting. The linear signal used was a stable $AR(4)$ process given by

$$x(k) = 1.79x(k-1) - 1.85x(k-2) + 1.27x(k-3) - 0.41x(k-4) + n(k) \quad (14)$$

where $n(k) \sim \mathcal{N}(0, 1)$ is white Gaussian noise (WGN). The benchmark nonlinear input signal was [14]

$$x(k) = \frac{x^2(k-1)(x(k-1)+2.5)}{1 + x(k-1)^2 + x(k-2)^2} + n(k-1) \quad (15)$$

For the experiments, the linear adaptive filter was the ϵ-NLMS (Normalised Least Mean Square) while the nonlinear filter was the Normalised Nonlinear Gradient Descent (NNGD) [20]. NNGD and NLMS were used as opposed to the standard NGD and LMS algorithms in order to overcome the issue of high dependence of the convergence of the individual subfilters and hence of the combination on input signal statistics. Furthermore, since these subfilters exhibit a rate of convergence that is potentially faster, this alternative also increased the speed of adaptation of λ.

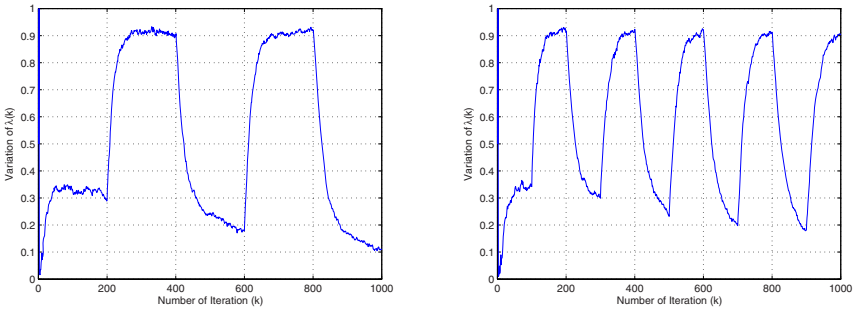
The nonlinearity at the output of the nonlinear filter was the logistic sigmoid function, given by

$$\Phi(z) = \frac{1}{1 + e^{-\beta z}} \ , \ z \in \mathbb{R} \quad (16)$$

with a slope of $\beta = 1$. Intuitively, we expect the linear filter to take over (i.e. $\lambda \to 0$) when the modality of the input signal is more linear while the output is expected to follow the more nonlinear filter (i.e. $\lambda \to 1$) when the input is nonlinear [see Fig. 9].

Figure 10 shows the evolution of λ at the output of the hybrid combination from Figure 9 for a signal alternating between linear and nonlinear every 200 and 100 samples respectively. The combination proved robust to changes in step-sizes within the combination and was always capable of tracking the degree of nonlinearity in the input signal, provided $\mu_{NLMS} = \mu_{NNGD}$ and provided the step-size values were such that both subfilters converged.

Having demonstrated the ability of the combination at tracking the degree of nonlinearity in synthetically generated data, we next perform simulations on real-world speech data.



(a) Input signal nature alternating every 200 samples

(b) Input signal nature alternating every 100 samples

Fig. 10. Mixing parameter λ at the output of the hybrid combination from Figure 9, with $\mu_\lambda = 20$, for input signal nature alternating between linear (14) and nonlinear (15)

5.3 Tracking the Degree of Nonlinearity in Speech Data

In this section, we aim at giving a flavour of the potential of the hybrid adaptive filtering approach on speech data. The area of speech modality characterisation is only emerging and in fact, only little is known about the nature of speech. Recently, much effort has been devised in developing accurate models for the speech production system and for characterising the modality of speech. It is believed that the accurate knowledge of speech characteristics will lead significant advances in several areas of speech processing, including speech coding and speech synthesis.

Typically, the vocal tract[5] is modeled as an all-pole filter, i.e. using a linear difference equation. This is mainly due to the solid theory underlying linear systems and to the corresponding decrease in computational complexity. However, the physical nature of the vocal tract is itself an indication of the potentially nonlinear nature of the radiated speech. In fact, several studies have suggested that linear models do not sufficiently model the human vocal tract [21,22].

Due to its nonstationary nature, the characterisation of speech is a complex task and much research has been done recently to study the nonlinear properties of speech and to find an efficient model for the speech signal. These studies have typically been based on a classification between vowels and consonants or between voiced and unvoiced sounds[6]. It is known that all vowels and certain consonants are voiced, i.e. highly periodic in nature with a periodic excitation source. In the case of unvoiced consonants, the folds may be completely open (e.g. for the /s/, /sh/ and /f/ sounds) or partially open (e.g. for /h/ sound), resulting in a noise like waveform [23,24,25].

[5] The vocal tract is the cavity where sound that is produced at the sound source is filtered. It consist of the laryngeal cavity, the pharynx, the oral and nasal cavities; it starts at the vocal folds (vocal cords).

[6] A sound is referred to as being voiced when the vocal folds are vibrating, whereas it is voiceless (or unvoiced) in a contrary case.

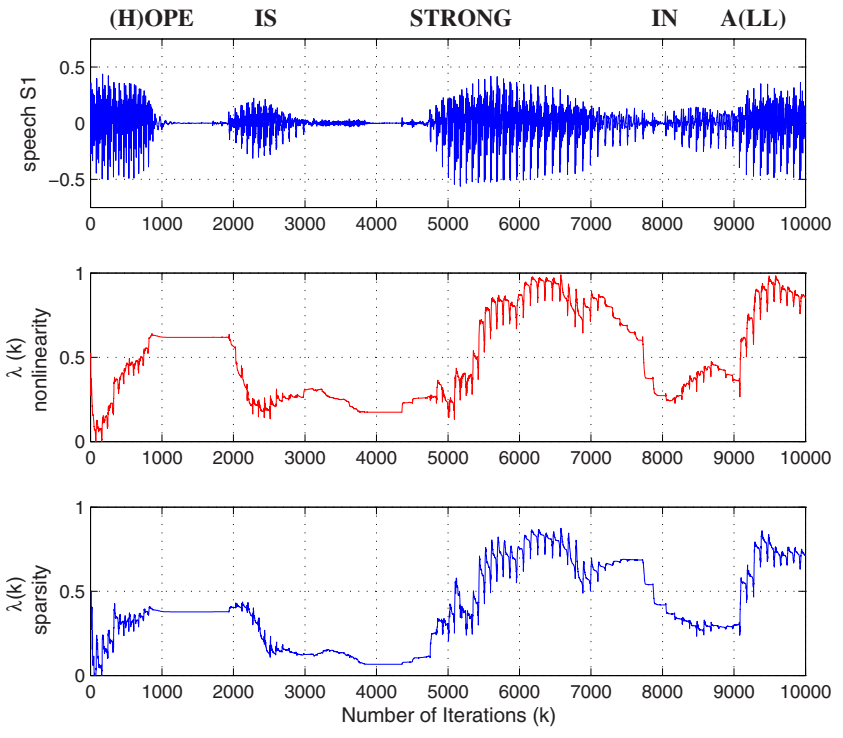


Fig. 11. Speech signal S1 and corresponding variation of λ for determining the degree of "nonlinearity" and "sparsity"

In [26], Kubin shows that there are several nonlinearities in the human vocal tract, whereas he demonstrates that linear autoregressive models are fully adequate for unvoiced speech. In [27,28,29] chaotic behaviour is found in voiced sounds such as vowels and nasals like /n/ and /m/. In [30], the speech signal is modeled as a chaotic process. Finally, hybrid methods combining linear and nonlinear structures have previously been applied to speech processing [31,32,33].

While the majority of the studies so far have suggested a nonlinear nature of voiced speech, the form of fundamental nonlinearity is still unknown. In [34], it is suggested that speech may contain different types of linear/nonlinear characteristics, and that for example, vowels may be modeled by either chaotic features or types of higher order nonlinear features, while consonants may be modeled by random processes.

For the simulations, speech signals S1 and S3 from [35] were first analysed. Finally, a randomly selected recording from the APLAWD database [36] was considered, together with the corresponding laryngograph[7] signal. All amplitude

[7] A laryngograph monitors vocal fold closure by measuring variations in the conductance between a transmitting electrode delivering a high frequency signal to the neck on one side of the larynx and a receiving electrode on the other side of the larynx.

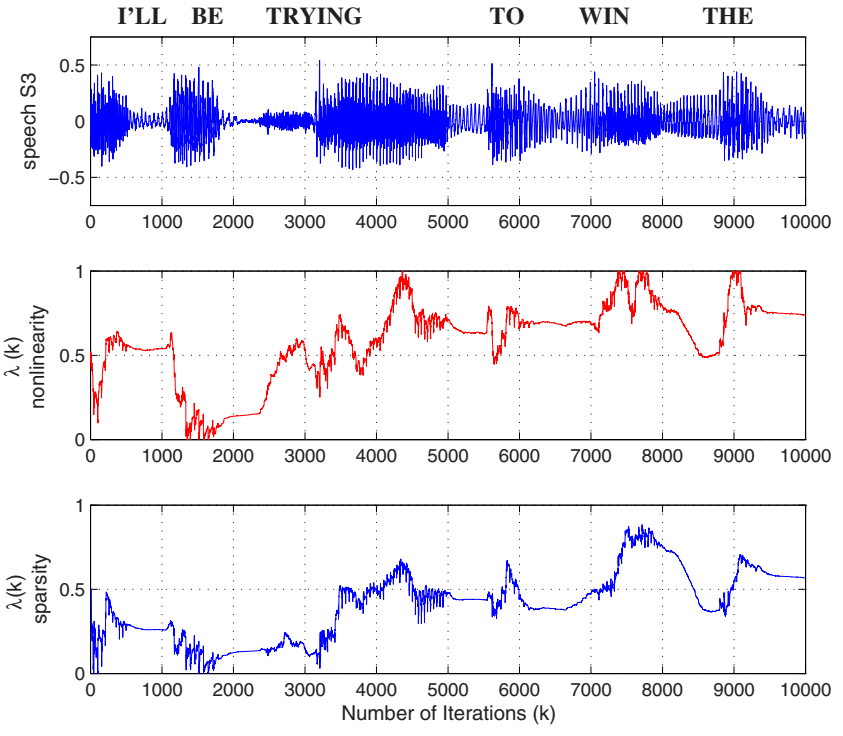


Fig. 12. Speech signal S3 and corresponding variation of λ for "nonlinearity" and "sparsity" tracking

signals were standardised so that the amplitude range was between $[-0.5, 0.5]$. For generality, the values of step-sizes were kept as in the simulations on stationary data, namely $\mu_{NLMS} = \mu_{NNGD} = 0.4$ and μ_λ was varied according to the aim of the experiment (larger μ_λ used for demonstrating the correlation between the laryngograph signal and the evolution of λ). The nonlinearity used in the complex NNGD (CNNGD) algorithm was the hyperbolic tanh function given by

$$\Phi(x) = \tanh(x) = \frac{\sinh x}{\cosh x} = \frac{\exp^x - \exp^{-x}}{\exp^x + \exp^{-x}} \ , \ x \in \mathbb{R} \tag{17}$$

Prediction was performed in the one-step ahead setting (short-term prediction). One may in the future perform simulations using long-term prediction, i.e. using a prediction delay of one pitch period, as in [37].

In order to investigate the potential of using hybrid filters for the purpose of determining the degree of nonlinearity and sparsity in a speech waveform, the combination of CNLMS and CNNGD (complex linear and nonlinear subfilters) and NLMS and SSLMS (signed sparse LMS) (following the approach from [38]) were both fed with the speech waveforms S1 and S3 in turn. The first trace from Figures 11 and 12 shows the speech waveform while the second and third traces respectively show the corresponding variations of λ for tracking the degree of

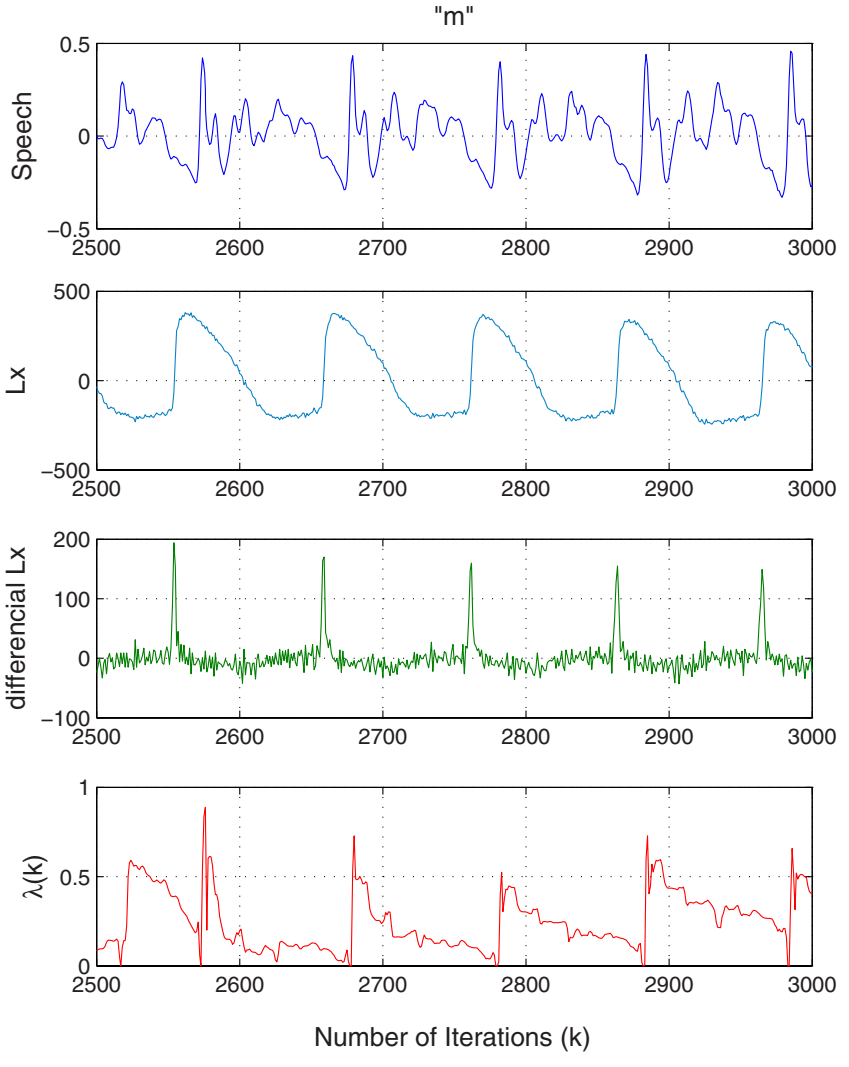


Fig. 13. Speech waveform for letter "m"; corresponding Lx waveform and differential Lx; variation of λ

nonlinearity and sparsity in the waveform. In the second trace, and as above, a value of λ close to 1 indicated the predominantly nonlinear nature of speech and vice versa for $\lambda \to 0$. Finally, in the third trace, and for consistency, a $\lambda \to 1$ showed the predominantly sparse nature of the waveform.

From the figures, the expected correlation between nonlinearity and sparsity is confirmed. We note that certain parts of a speech signal are better modeled using nonlinear structures ($\lambda \to 1$), while for others, linear structures

are sufficient ($\lambda \rightarrow 0$). Furthermore, voiced speech appears to be indicated by regions where λ exhibits a "spiky" behaviour. From Figure 11, it can be noticed that the noise like sounds /z/ (around samples 2800-3200) and /s/ (around samples 4100-4200) are linear which agree with previous findings in the field. From Figure 12, it can be inferred that highly voiced sounds such as /a/ in "trying" is more nonlinear.

5.4 Correlation Between Laryngograph Signal and the Variation of λ

In this section, we aim at exploring the relationship between the variation of λ and the laryngograph waveform (Lx). For this purpose, simulations are performed on the randomly selected speech waveform from the APLAWD database [36]: the letter "m" read by a male speaker. Figure 13 shows the speech and corresponding laryngograph waveforms and the evolution of λ at the output of the hybrid combination of CNLMS and CNNGD. From this figure, it is clear that there is some correlation between the two waveforms during certain periods of voiced speech. In particular, it appears that sharp transitions in λ and in the derivative of the Lx waveform (indicating glottal opening instant) occur simultaneously (the delay between the two waveforms is due to the larynx-to-microphone delay and estimated in [39] to be of approximately 0.95ms, i.e. 20 kHz $\times$ 0.95ms = 19samples). This does not necessarily imply that the hybrid filter is capable of detecting glottal opening instants, but that there is a clear relationship between the two signals which requires further investigation.

References

1. Schreiber, T.: Interdisciplinary application of nonlinear time series methods. Phys. Rep. 308(1), 1–64 (1999)
2. Gautama, T., Mandic, D.P., Van Hulle, M.M.: The delay vector variance method for detecting determinism and nonlinearity in time series. Physica D 190(3–4), 167–176 (2004)
3. Gautama, T., Mandic, D., Van Hulle, M.: Signal nonlinearity in fMRI: A comparison between BOLD and MION. IEEE Trans. Med. Imaging 22(5), 636–644 (2003)
4. Schreiber, T., Schmitz, A.: On the discrimination power of measures for nonlinearity in a time series. Phys. Rev. E 55(5), 5443–5447 (1997)
5. Gautama, T., Mandic, D.P., Van Hulle, M.M.: On the characterisaion of the deterministic/stochastic and linear/nonlinear nature of time series. Technical Report DPM-04-5, Imperial College London (2004)
6. Schreiber, T., Schmitz, A.: Surrogate time series. Physica D 142, 346–382 (2000)
7. Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., Farmer, J.: Testing for nonlinearity in time series: The method of surrogate data. Physica D 58, 77–94 (1992)
8. Schreiber, T., Schmitz, A.: Improved surrogate data for nonlinearity tests. Phys. Rev. Lett., 635–638 (1996)
9. Weigend, A.S., Gershenfeld, N.A.: Time Series Prediction: Forecasting the Future and Understanding the Past. Addison-Wesley, Reading, MA (1993)

10. Kaplan, D.: Exceptional events as evidence for determinism. Physica D 73(1), 38–48 (1994)
11. Cao, L.: Practical method for determining the minimum embedding dimension of a scalar time series. Physica D: Nonlinear Phenomena 110(1-2), 43–50 (1997)
12. Haykin, S., Li, L.: Nonlinear adaptive prediction of nonstationary signals. IEEE Transactions on Signal Processing 43(2), 526–535 (1995)
13. Mandic, D.P., Chambers, J.A.: Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability. John Wiley & Sons, Chichester (2001)
14. Narendra, K.S., Parthasarathy, K.: Identification and control of dynamical systems using neural networks. IEEE Transactions on Neural Networks 1(1), 4–27 (1990)
15. Mizuta, H., Jibu, M., Yana, K.: Adaptive estimation of the degree of system nonlinearity. In: IEEE Adaptive Systems for Signal Processing and Control Symposium (AS-SPCC), pp. 352–356 (2000)
16. Cichocki, A., Unbehauen, R.: Neural networks for optimization and signal processing. Wiley, Chichester (1993)
17. Figueras-Vidal, A.R., Arenas-Garcia, J., Sayed, A.H.: Steady state performance of convex combinations of adaptive filters. In: ICASSP 2005. Proceedings of the International Conference on Acoustis, Speech and Signal Processing, pp. 33–36 (2005)
18. Kozat, S.S., Singer, A.C.: Multi-stage adaptive signal processing algorithms. In: Proceedings of the 2000 IEEE Sensor Array and Multichannel Signal Processing Workshop, pp. 380–384 (2000)
19. Mandic, D.P., Vayanos, P., Boukis, C., Goh, S.L., Jelfs, B., Gautama, T., Rutkowski, T.: Collaborative adaptive learning using hybrid filters. In: Proceedings of ICASSP 2007, vol. III, pp. 921–924 (2007)
20. Mandic, D.P.: NNGD algorithm for neural adaptive filters. Electronics Letters 36(9), 845–846 (2000)
21. Schroeter, J., Sondhi, M.: Speech coding based on physiological models of speech production. In: Furui, S., Sondhi, M. (eds.) Advances in speech Signal Processing, pp. 231–268. Marcel Dekker, New York, NY, USA (1992)
22. Thyssen, J., Nielsen, H., Hansen, S.: Non-linear short-term prediction in speech coding. In: ICASSP 1994. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, vol. 1, pp. 185–188 (1994)
23. Deller, J.R., Proakis, J.G., Hansen, H.L.: Discrete Time Processing of speech Signals. Prentice-Hall, Englewood Cliffs (1987)
24. Rabiner, L., Juang, B.H.: Fundamentals of speech recognition. Prentice-Hall, Englewood Cliffs (1993)
25. Rabiner, L., Schafer, R.W.: Digital Processing of Speech Signals. Prentice-Hall, Englewood Cliffs (1978)
26. Kubin, G.: Nonlinear processing of speech. In: Kleijn, W., Paliwal, K. (eds.) Speech coding and synthesis, pp. 557–610. Elsevier Science B.V., Amsterdam (1995)
27. Banbrook, M., McLaughlin, S., Mann, I.: Speech characterisation and synthesis by nonlinear methods. In: Proceedings of the International Conference on Speech and Audio Processing, vol. 7, pp. 1–17 (1999)
28. Martinez, F., Guillamon, A., Alcaraz, J., Alcaraz, M.: Detection of chaotic behaviour in speech signals using the largest lyapunov exponent. In: DSP 2002. IEEE International Conference on Digital Signal Processing, pp. 317–320 (2002)
29. Miyano, T., Nagami, A., Tokuda, I., Aihara, K.: Detecting nonlinear determinism in voiced sounds of japanese vowel /a/. International Journal of Bifurcation and Chaos 10(8), 1973–1979 (2000)

30. Townshend, B.: Nonlinear prediction of speech. In: ICASSP 1991. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, pp. 425–428 (1991)
31. Hansen, J., Gavidia-Ceballos, L., Kaiser, J.: A nonlinear operator-based speech feature analysis method with applications to vocal fold pathology assessment. IEEE Transactions on Biomedical Engineering 45(3), 300–313 (1998)
32. Maragos, P., Quatieri, T., Kaiser, J.: Speech nonlinearities, modulations, and energy operators. In: ICASSP 1991. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, pp. 421–424 (1991)
33. Wokurek, W.: Time-frequency analysis of the glottal opening. In: ICASSP 1997. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, pp. 1435–1438 (1997)
34. Turunen, J., Tanttu, J.T., Loula, P.: Hammerstein model for speech coding. EURASIP Journal on Applied Signal Processing 2003(12), 1238–1249 (2003)
35. Mandic, D.P., Baltersee, J., Chambers, J.A.: Nonlinear prediction of speech with a pipelined recurrent neural network and advanced learning algorithms. In: Prochazka, A., Uhlir, J., Rayner, P.J.W., Kingsbury, N.G. (eds.) Signal Analysis and Prediction, pp. 291–309. Birkhauser, Boston (1998)
36. Lindsey, G., Breen, A., Nevard, S.: Spar's archivable actual-word databases (1987)
37. Birgmeier, M., Bernhard, H.P., Kubin, G.: Nonlinear long-term prediction of speech signals. In: ICASSP 1997. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp. 1283–1286 (1997)
38. Jelfs, B., Mandic, D.P.: Toward online monitoring of the changes in signal modality: The degree of sparsity. In: Proceedings of the 7th IMA International Conference on Mathematics for Signal Processing, pp. 29–32 (2006)
39. Brookes, M., Gudnasonand, J., Kounoudes, A., Naylor, P.: Estimation of glottal closure instants in voiced speech using the DYPSA algorithm. IEEE Transactions on Audio, Speech and Language Processing 15(1), 34–43 (2007)