Control and Intelligent Systems, Vol. 30, No. 1, 2002

NONLINEAR SPEECH PROCESSING: OVERVIEW AND APPLICATIONS

M. Faúndez-Zanuy,* S. McLaughlin,** A. Esposito,*** A. Hussain,**** J. Schoentgen,***** G. Kubin,****** W.B. Kleijn,******* and P. Maragos********

Abstract

This article presents an overview of various nonlinear processing techniques applied to speech signals. Eevidence relating to the existence of nonlinearities in speech is presented, and the main differences between linear and nonlinear analysis are summarized. A brief review is given of the important nonlinear speech processing techniques reported to date, and their applications to speech coding, speech synthesis, speech and speaker recognition, voice analysis and enhancement, and analyses and simulation of dysphonic voices.

Key Words

Nonlinear processing, speech coding, recognition, synthesis, analysis and enhancement

1. Introduction

Source-filter models form the foundation of many speech-processing applications such as speech coding, speech synthesis, speech recognition, and speaker recognition. Usually, the filter is linear and based on linear prediction. The excitation of the linear filter is either left undefined, modelled as noise, described by a simple pulse train, or described by an entry from a large codebook. Although this approach has enabled progress over the last 30 years, it neglects structure known to be present in the speech signal. In practical applications, this neglect manifests itself as an increase in the bit rate, less natural speech synthesis, and an inferior ability to discriminate between speech sounds.

Nonlinear techniques are indeed potentially useful in the framework of the modelling or analyses of the following:

- Nonlinearities of the systems that generate the signal and/or noise
- Nonlinearities of the signal acquisition system
- Nonlinearities of the transmission channel
- Nonlinearities of the human perception mechanism

* Escuela Universitaria Politécnica de Mataró, Spain; e-mail: faundez@eupmt.es, http://www.eupmt.es/veu
** University of Edinburgh, U.K.
*** IIASS, Italy
**** University of Stirling, UK
***** Université Libre de Bruxelles, Belgium
****** Vienna Telecommunications Research, Austria
******* KTH Royal Institute of Technology Stockholm, Sweden
******** National Technical University of Athens, Greece

(paper no. 201-1222)

In addition, some problems are difficult to solve with linear techniques and are more tractable with nonlinear ones. On the other hand, there are drawbacks when dealing with nonlinear techniques. The main ones are:

- A lack of a unifying theory of the different nonlinear processing tools (neural nets, homomorphic, polynomial, morphological, and ordered statistics filters, and so on)
- The computational burden is usually greater than with linear techniques
- Nonlinear systems are difficult to analyze, because well-known analysis tools are not applicable, especially frequency domain analysis. Attempts have been made to improve nonlinear filters analysis. One example is the time slope transform domain for morphological systems [1–3]. A second example is the obtaining of the root-signal set of median filters [4], which try to emulate the eigenfunctions of a linear system
- Sometimes, closed-form formulations of nonlinear models do not exist. This means that an iterative solving procedure must be used, and local minima problems exist.

The conventional linear approaches to speech signal modelling are based on approximations of the physics of speech production via linear acoustics and one-dimensional plane wave propagation of the sound in the vocal tract. Other assumptions are that the excitation signal and the vocal-tract filter model are mutually independent and the airflow through the vocal tract is laminar. A summary of experimental evidence that demonstrates limitations of conventional linear analyses follows.

1.1 Residue Comparison

Consider the following empirical argument: a linear and a nonlinear predictor (with the same prediction order) are applied to the same speech signal. One observes that the residual signal of the nonlinear predictor has smaller energy. A nonlinear predictor can even remove the glottal cycle periodicity without the use of a long-term predictor.

In [5] linear predictive coding analysis of order 10 was applied several times to the same speech segment. It was noted that the prediction gain after the fifth iteration was 0 dB. Thus, the remaining redundancies could not be linear. The next step was to apply a quadratic Volterra filter to the residual signal obtained by means of the linear filter. The nonlinear predictor further reduced the energy of the residual signal. The conclusion was that the speech signal contained nonlinearities. The authors of [6] report similar experiments.

1.2 Correlation Dimension

An important characteristic of a dynamical system is the dimension of its attractors, that is, the subspace of the state-space towards which a time history evolves after transients die out. The estimation of this dimension gives a lower bound of the number of parameters needed in order to model a dynamical system. One goal is to find out whether the attractor is high dimensional or low dimensional. The correlation dimension ([7, 8]) is a practical method to estimate this dimension via an empirical temporal series.

The author of [9] presented an experiment that performed an LPC analysis of a speech signal, replaced the residual for each speech frame with Gaussian white noise of the same energy, and re-synthesized the speech signal. This new signal presented the property that the determinism was limited to the linear components, and the best predictor was linear. When comparing the correlation dimension obtained via the artificial signal to the original signal, one observed that the dimension obtained via the artificial signal was greater. The conclusion was that the residual of the original signal contained a deterministic nonlinear component, which could possibly be removed by a nonlinear predictor.

1.3 Higher Order Statistics

Higher order statistics (HOS) [10] go beyond second-order statistics such as the correlation. The problem with HOS is that they require a large number of records to obtain accurate estimates. This requirement is in conflict with the nonstationary nature of speech.

HOS can be used to test whether a signal can be fitted by a linear model. By definition, a linear stochastic process can be represented as the output of a linear filter excited by a sequence of independent random variables. For linear signal models, the magnitude of the bicoherence is equal to one. This enables testing the adequacy of linear models while representing non-Gaussian signals. For Gaussian signals, the bicoherence vanishes. The author of [11] presented the following conclusions regarding speech segments.

- Unvoiced fricative sounds have a bi-spectrum close to zero, so they can be well modelled via a Gaussian process.
- Voiced sounds have a bi-spectrum significantly different from zero. Thus, they cannot be considered as Gaussian.

Other research, [12–14] also examined the bi-spectrum of vowel signals and found quadratic nonlinearities, which were revealed by a phase coupling between harmonic spectral components. This phase coupling was due to a nonlinear process. A nonlinear predictor could use the phase coupling to eliminate most of the cycle periodicity, even by means of a short-term prediction [14]. It is indeed possible to consider the cycle periodicity as the outcome of some nonlinear short-term interaction, instead as a long-term dependency. These experiments do not, however, exclude the possibility of the existence of nonlinearities of an order greater than two.

1.4 Probability Density Functions

The probability density function of a speech signal may be asymmetric, especially when the lower frequencies (<300Hz) are included. A summarizes of work regarding this observation is given in [15]. Although the unvoiced sounds can be considered to be Gaussian, this is not generally valid for voiced sounds. The shape of the probability density function may vary strongly with the category of speech segment. Thus, the hypothesis of universal Gaussianity of [16, 32] is not exact, and the linear predictor that is optimal (in the mean square error sense) for Gaussian signals is no longer optimal in the case of speech signals.

2. Brief Review of Nonlinear Techniques in 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 rapid growth in this area starting in the mid-1980s. Since that time, numerous techniques, which are ultimately aimed at engineering applications, have been described. An excellent recent overview of these techniques is given in [15].

Inherent in the broad scope of nonlinear methods is the large variety of methods found in the literature and the difficulty in classifying the techniques. Moreover, it is difficult to predict which techniques ultimately will be more successful. However, commonly observed 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, as a nonlinear autoregressive model in its synthesis form forms a nonlinear oscillator if no input is applied. For this reason we focus here on nonlinear autoregressive models. For the practical design of a nonlinear autoregressive model, various approximations have been proposed [15, 17–19]. These can be split into two main categories: parametric and nonparametric methods.

Parametric Methods. Parametric methods are perhaps best exemplified by the polynomial approximation (truncated Volterra series with the special case of quadratic filters; [5, 6, 20, 21], locally linear models [9, 22–25], including threshold autoregressive models [26], and state-dependent models [27]. Another important group of parametric methods is based on neural nets: radial basis functions approximations [28–34], multilayer perceptrons [5, 7, 35–41], and recurrent neural nets [42–45].

Nonparametric Methods. Nonparametric nonlinear autoregressive methods also play an important role in nonlinear speech processing. Examples are Lorenz's method of analogues [28, 46–48], perhaps the simplest of various nearest neighbour methods [17, 49] which also include nonlinear predictive vector quantization [43, 50–52] or codebook prediction [24, 53]. Another nonparametric approach [54] is based on kernel-density estimates [55–56] of the conditional expectation.

Speech Fluid Dynamics, Modulation and Fractal Methods. Another class of nonlinear speech processing methods includes algorithms proposed to analyze nonlinear phenomena of the fluid dynamics type in the speech airflow during speech production. Such nonlinear phenomena are described in [57, 58]. The investigation of the speech airflow nonlinearities can proceed in at least two directions:

- Numerical simulations of the nonlinear differential (Navier-Stokes) equations governing the 3D dynamics of the speech airflow in the vocal tract
- Development of nonlinear signal processing systems suitable to detect such phenomena and extract related information. The second direction has been followed by Maragos and his co-workers to model and detect modulations in speech resonances of the AM-FM type [59], to model and measure the effects of turbulence in speech sounds using fractals [60], and to apply related nonlinear speech features to problems of speech recognition and speech vocoders [60–63].

3. Applications of Nonlinear Speech Processing

Although nonlinear speech processing is applicable to almost all the fields of speech processing, the main contributions are in the following fields.

3.1 Speech Coding

The bit rate available for speech signals must be strictly limited in order to accommodate the constraints of the channel resource. For example, new low-rate speech-coding algorithms are needed for interactive multimedia services on packet-switched networks such as the evolving mobile radio networks or the Internet, and nonlinear speech processing offers a good alternative to conventional techniques. Voice transmission will have to compete with other services such as data/image/video transmission for the limited bandwidth resources allocated to an ever-growing, mobile network user base, and very low bit rate coding at consumer quality will see increasing demand in future systems.

In speech coding, it is possible to obtain good results using models based on linear predictive coding, as the residual can be coded with sufficient accuracy given a high enough bit rate. However, it is also evident that some of the best results in terms of optimizing both quality and bit rate are obtained from codec structures that contain some form of nonlinearity. Analysis-by-synthesis coders fall into this category. For example, in CELP coders the closed-loop selection of the vector from the codebook can be seen as a data-dependent, nonlinear mechanism.

With a nonlinear predictor, it may be possible to improve the long-term pitch predictor in analysis-by-synthesis coders. In [64], it was reported that this long-term prediction contributed around 75% to the overall SNR of a typical CELP coder. Therefore, it is reasonable to expect that the nonlinear predictor, will contribute to improving this long-term prediction and hence the performance of the coder.

3.2 Speech Synthesis

New telecommunication services include the capability of a machine to speak with a human in a "natural way"; to this end, a lot of work must be done in order to improve the actual voice quality of text-to-speech and concept-to-speech systems. The richness of the output signals of self-excited nonlinear feedback oscillators will allow synthetic voices to be better matched to human voices.

Speech synthesis technology plays an important role in many aspects of human-machine interaction, particularly in telephony applications. Improvement can focus on new techniques for the speech signal generation stage in a speech synthesizer based on concepts from nonlinear dynamical theory.

To model the nonlinear dynamics of speech, the one-dimensional speech time-domain signal is embedded into an appropriate higher dimensional space. This reconstructed state-space representation has approximately the same dynamical properties as the original speech generating system and is thus an effective model for speech synthesis.

Improvement can also focus on systems that will reproduce the natural dynamics of speech. This will involve constructing models that operate in the state-space domain, such as neural network architectures. The speech synthesized by these methods will be more natural sounding than linear concatenation techniques because the low-dimensional dynamics of the original signal are learnt, which means that phenomena such as inter-pitch jitter are automatically included into the model. In addition to generating high-quality speech, other associated tasks will also be addressed. The most important of these is to examine techniques for natural pitch modification that can be linked into the nonlinear model.

3.3 Speech and Speaker Recognition

Speech recognition plays an increasingly important role in modern society. nonlinear techniques allow one to merge feature extraction and the classification problem, and to include the dynamics of the speech signal in the model. This is likely to lead to significant improvements over current methods, which are inherently static.

Security in transactions, information access, and the like is another important question to be addressed in the future, and speaker identification/verification is perhaps one of the most important biometric systems, because of its feasibility for remote (telephonic) recognition without additional hardware requirements.

Many problems remain in continuous speech recognition. Much of this may be due to the static nature of the hidden Markov models (HMM) used: they are unable to follow the dynamics of the speech between individual states. In addition, it is typically a series of mel-frequency cepstral coefficients that form the acoustic feature vector used for classification, with the inclusion of first- and second-order differentials to try to provide some continuity between frames. However, the cepstrum itself is fundamentally based on a linear speech model, and the inclusion of the differential terms is an unsatisfactory method to attempt to include the speech dynamics into the inherently static HMMs.

The nonlinear predictor may be able to perform the tasks of front-end feature extractor and low-level classifier simultaneously. A simple recognition system can be envisaged where each class (which could be of phones, diphones, etc.) can be characterized by a nonlinear model. Then, given an input frame of speech, it will be possible to use the sum of the error residual from the predictor over the frame to decide to which class the input speech belongs. Thus the feature extraction and the classification problem are merged together and solved by one unit. Further, the dynamics of the speech signal may be included in the nonlinear model, unlike the HMM structure, which is inherently static. This has been highlighted previously as a promising area to pursue for continuous speech recognition.

For speaker recognition applications it has been shown that the residual signal of a linear analysis contains enough information to enable human beings to identify people. Thus, there is relevant information that is ignored with a linear analysis. Several articles [65, 66] have shown that it is possible to improve the identification rates with a combination of linear and nonlinear predictive models.

Further, for both speech and speaker recognition there is growing experimental evidence that using nonlinear aero acoustic features of the modulation or fractal type as input to HMM-based classifiers (in addition to the standard cepstrum linear features) leads to better recognition performance than using only linear features. Thus, work on detecting such features and using them in recognition systems is very promising.

3.4 Voice Analysis and Enhancement

3.4.1 Voice Analysis

The underlying groundwork of actual speech systems is the use of some parametric representations of the speech signal, which reflect our understanding of speech production and speech perception mechanisms. However, as prior research focused on the analysis of read or laboratory speech and written text, our knowledge of running speech perception and speech production mechanisms is limited. To formulate a better representation of the speech signal the first step is to analyze it in a very detailed manner, using corpora based on spontaneous speech. The use of spontaneous speech will allow one to account for phonological phenomena, such as assimilation, disfluencies, speaking styles, and emotional state [78, 80].

The actual speech technology is a new approach to speech problems, which overcomes the classical source filter theory and is able to quantify the nonlinear features of speech time series data, and to embed these new features in automatic speech-based systems. New ideas and new algorithms can be generated only after a detailed acoustic analysis of spontaneous speech.

The properties that are found to be significant at the acoustic level may or may not be so at the perceptual level. For example, an acoustic parameter such as the amplitude of the peaks in the spectrum may vary significantly from one consonant to the other, and therefore exhibit values that are peculiar for a consonant [76], although perceptually the amplitude information might or might not be (it is still an open problem) integrated with the information contained in the frequency range. The preprocessing processes performed in the inner ear are highly nonlinear, with the result that perception of speech sounds undergoes a series of phenomena that are still not completely understood (see [77] for details). Because of these processes, the relevance of the acoustic attributes must be supported and integrated by a perceptual analysis in order to evaluate their perceptual significance.

The need to acquire deeper knowledge of the acoustic and perceptual properties of running speech justifies the set-up of acoustic and perceptual experiments for a variety of speech sounds. It is necessary a multidisciplinary research, for the development of shared speech corpora (extracted from running speech) from several languages. Accurate analyses carried out on both consonants and vowels should be performed. The aim is to find a combination of acoustic properties that prove to be speaker independent (normalization problem), context independent (co-articulation problem), and robust to the noise effects.

Speaker dependence of the speech signal leads to speaker dependence of the speech recognizer. Normalization is the process of finding acoustic properties of speech sounds that prove to be independent of physiological differences, linguistic differences, or even the physical and emotional state of the speaker. Two approaches can be found in the literature. The first is based on the assumption that in some projections of acoustic space there must be an invariant representation of the speech sounds (radical invariance) [82]. The second is based on the hypothesis that a speaker-independent representation comes out from a certain number of direct (such as F3 formant frequency) and indirect (such as mode of vocal cord vibration) acoustic cues based on a perceptual estimate of the speaker's vocal tract length. Both the methods are strategies to tackle the variations from speaker, channel, and environment. Both can bring useful improvements when embedded in automatic speech-based systems. An accurate acoustic and perceptual analysis can suggest new features for improving these normalization methods.

Co-articulation produces changes in the articulation and the acoustic of a speech sound, due to its phonetic context [83]. Co-articulation causes significant problems for automatic speech synthesis and recognition systems. An acoustical and perceptual analysis of spontaneous speech can help in identifying these features and give a quantitative assessment of their practical use. Moreover, a fitting procedure for these physical parameters (or eventually the development of richer language-based models) can constitute an additional approach to solving the problems of automatic speech-based systems, overcoming the current paradigms based on the analysis of statistical properties of recorded utterances (through LPC, computation of cepstrum coefficients, etc.). The experiments suggested constitute a first step towards the definition of new properties of speech segments that can play a fundamental role in the design of preprocessing algorithms, classification methods, and system structure for speech applications.

3.4.2 Nonlinear Speech Enhancement

Contamination of speech signals with background noise reduces the signal-to-noise ratio (SNR) of for example, portable phones, hands-free telephones, and security screens. In particular, speech recognition equipment experiences problems due to noisy environments that are quite acceptable to human listeners. Speech enhancement is motivated by the need to improve the performance of voice communications systems in noisy conditions. Applications range from front-ends for speech recognition systems to enhancement of telecommunications in aviation, military, teleconferencing, and cellular environments. The goal is either to improve the perceived *quality* of the speech or to increase its *intelligibility*.

Recently, new nonlinear speech-processing methods using artificial neural networks (ANN) have been investigated [67–69] that are shown to be more able to take into account nonlinearities in the acoustics or electro-acoustic transmission systems [70] and the non-Gaussian nature of speech. Several other distinct neural-network-based frameworks for speech enhancement have also emerged in the literature. These include time-domain filtering, transform-domain mapping, state-dependent model switching, and on-line iterative approaches. A good overview of these can be found in [71].

A novel ANN-based multisensor sub-band adaptive signal-processing scheme has been described for enhancing acoustic-speech corrupted by real noise and reverberation in [67, 68]. Numerically robust adaptation-algorithms were employed for the single-layered linear-in-parameters ANN sub-band filters; simulation experiments using real-reverberant automobile data were used to demonstrate that nonlinear speech-enhancement schemes are capable of outperforming conventional linear filtering-based wideband and multiband noise-cancellation implementations.

Currently, the binaural ANN based sub-band processing schemes are being further developed to incorporate additional, more complex cross-band and cross-channel interactions. Both cross-band and cross-channel interactions are major influences on human hearing abilities; cross-band effects support "spectral sharpening" operations and cross-channel (binaural) effects invoke lateralization and noise cancellation to separate desired from undesired signals [72]. The developed schemes will be assessed both quantitatively and qualitatively using formal listening and intelligibility tests with human subjects.

In related work [87, 88], it has recently recently shown that under certain circumstances, *adding* the right amount of noise can actually enhance rather than diminish the detection of a weak periodic signal. This phenomenon is known as stochastic resonance (SR). Currently, SR is being investigated in a leaky integrate-and-fire (LIF) neuron model, and in networks of such neuron models. LIF neuron models are widely used in neuronal modelling as they capture most of the important sub-threshold dynamics of real neurons. SR has been assessed by applying a sub-threshold periodic signal plus noise to an LIF neuron and examining its output SNR (with a peak indicating SR at a certain noise strength). In a new study [88], SR has been demonstrated in both continuous (floating point) and low-resolution discrete (FPGA-based) LIF neuron models. Future work will investigate SR in LIF neuron networks using mixed tones and eventually real speech signals.

3.5 Analyses and Simulation of Dysphonic Voices

Voiced speech signals are produced by filtering the acoustic source signal by means of the vocal tract transfer function. The source signal is generated by pulsatile airflow through the glottis. Pulsatile glottal airflow is the outcome of the vibrating vocal folds, the vibratory patterns of which are controlled via the positions of the laryngeal cartilages and tensions of the intrinsic and extrinsic laryngeal muscle pairs. The vibration of the vocal folds is self-sustained, and the equations that describe the vibrations are nonlinear. The lack of linearity is due to the contact between the left and right vocal folds, the nonlinear relations between the stresses and strains of the vocal folds, and the nonlinear relations between glottal shape and airflow rate or pressure. In normal phonation, the periodicity of the acoustic voice source signal is thought to be due to the synchronization of a small number of modes of vibration of the vocal folds [75]. Although periodicity is the norm in healthy voices, observations of the acoustic source signal generated by in vivo, in vitro, or simulated vocal fold vibrations have revealed patterns that are suggestive of bifurcations from periodic regimes of vibration to subharmonic or quasiperiodic regimes, as well as chaos [86].

These observations have motivated the application of analysis techniques inspired by nonlinear system dynamics, especially those based on state-space reconstruction. These techniques enable the computation of dynamical invariants, that is, quantities like the global embedding dimension, the local dynamic dimension, and Lyapunov exponents. The global embedding dimension indicates the number of geometric dimensions that are needed to completely unfold the attractor; the local dynamic dimension indicates the number of the dynamical degrees of freedom that specify the evolution of the dynamical system along local areas of the attractor [74]; and Lyapunov exponents characterize the stability of the trajectories in the state-space of the dynamical system. Positive exponents would suggest that the trajectories are unstable, that is, sensitive to small perturbations [79].

Ideally, these analyses are performed not only to represent the speech signal in a more economical or more revealing manner than would be possible by means of linear techniques, but also to draw conclusions regarding the dynamics of the glottal vibrator. The down side, therefore, is that these techniques require the glottal vibrator to be in a steady-state and disturbance free, assumptions that are difficult to check and are presumably never exactly true, especially in the case of dysphonic voices. Analysis techniques inspired by nonlinear system dynamics have therefore mainly been used to infer qualitatively the dynamics of simulated or in vitro vibrations [75, 81]. Such nonlinear analyses of speech signals sustained by healthy or dysphonic speakers have so far produced few incontrovertible results. The most up-to-date study is [84]. The article of Kumar and Mullick [79] is interesting insofar as it explicitly discusses the relative merit of linear and some nonlinear techniques in the light of an economical representation of the speech signal rather than of vocal fold dynamics.

It is important to note that even in sustained speech sounds, the timing of the glottal cycles is determined not only by the dynamic regime of the fold vibration, but also by external disturbances, which are known as vocal jitter and shimmer, vocal tremor, vibrato, and glissandi. These perturbations of the dynamic regimes are best understood as modulation noise. A source of additive noise (as opposed to modulation noise) is acoustic noise generated by turbulent airflow that is accelerated or decelerated by obstacles in the glottis or vocal tract. It has been shown that modulation noise of the speech signal is stochastic rather than deterministic [85]. Additive and modulation noise is therefore likely to interfere with methods of analysis that are aimed at describing the dynamics rather than the noise [73].

4. Conclusion

It seems evident, after 50 years of research in the field, that phonetic percepts cannot be seen as "knowledge-innocent" records of a raw speech signal. Therefore, all the attempts to improve automatic speech systems (for both synthesis and recognition) seem to require a step forward, towards the assumption of nonlinearity. Speech production and speech perception are adaptively organized. The simulation of these processes by a machine requires one to include in the model the dynamic of this adaptation, and it is clear that this dynamic is nonlinear. Implementing techniques of nonlinear speech processing may help to improve the applications and to understand this dynamic and how it evolves, both perceptually and physically, as speech is a learned skill affected by the constraints of the production and perception systems. The techniques discussed above are a first step in this direction. What is further needed is that experts, from both speech science and speech technology, make an active effort to try to share their knowledge with the aim of embedding these nonlinear aspects in a mathematical and computational theory, which might allow the development of better speech models and consequently improve current and future speech applications.

References

[1] P. Maragos, Max-min difference equations and recursive morphological systems, *Proc. 1st Int. Workshop on Mat. Morphology and Its Applications to Signal Processing*, Barcelona, May 1993.

[2] P. Maragos, A time slope domain theory of morphological systems: Slope transforms and max-min dynamics, *EUSIPCO-94*, 1994, 971–974.

[3] L. Dorst & R. Boomgaard, An analytical theory of mathematical morphology, *Proc. 1st Int. Workshop on Mathematical Morphology and Its Applications to Signal Processing*, Barcelona, May 1993.

[4] G. Arce & N. Gallagher, State description for the root-signal set of median filters, *IEEE Trans. on ASSP, 30*(6), December 1982.

[5] J. Thyssen, H. Nielsen, & S.D. Hansen, Non-linear short term prediction in speech coding, *Proc. Int. Conf. Acoustics, Speech & Signal Processing*, Australia, April 1994, I-185–I-188.

[6] E. Mumolo & D. Francescato, Adaptive predictive coding of speech by means of volterra predictors, *Workshop on Nonlinear Digital Signal Processing*, Tampere 1993, 2.1-4.1–2.1-4.4.

[7] N. Tishby, A dynamical systems pproach to speech processing, *Proc. Int. Conf. Acoustics, Speech & Signal Processing*, Albuquerque, April 1990.

[8] S. McLaughlin & A. Lowry, Non-linear dynamical systems concepts in speech analysis, *Proc. EUROSPEECH 1993*, 1993.

[9] B. Townshend, Non-linear prediction of speech, *Proc. Int. Conf. Acoust. Speech, Sign. Proc.*, Toronto, May 1991, 425–428.

[10] C.L. Nikias & J.M. Mendel, Signal processing with higher-order spectra, *IEEE Signal Processing Magazine*, July 1993, 10–37.

[11] B.B. Wells, Voiced/unvoiced decision based on the bispectrum, *Proc. Int. Conf. Acoust. Speech, Sign. Proc.*, 1985, 1589–1592.

[12] S. McLaughlin, S. Hovell, & A. Lowry, Identification of nonlinearities in vowel generation, *Proc. EUSIPCO'94, 2*, 1133–1136.

[13] J.W.A. Fackrell, Bispectral analysis of speech, doctoral diss., University of Edinburgh, 1996 (http://www.ee.ed.ac.uk/ jwaf/ Work/thesis.html)

[14] J. Thyssen, H. Nielsen, & S.D. Hansen, Non-linearities in speech, *Proc. IEEE Workshop Non-linear Signal and Image Processing, NSIP'95*, June 1995.

[15] G. Kubin, Non-linear processing of speech, W.B. Kleijn & K.K. Paliwal eds, *Speech coding and synthesis*, Chap. 16,(Elsevier 1995).

[16] N.S. Jayant & P. Noll, Digital coding of waveforms: Principles & Applications (Englewood Cliffs: Prentice Hall, 1984).

[17] J.D. Farmer & J.D. Sidorowich, Exploiting chaos to predict the future and reduce noise, in *Evolution, learning and cognition*, ed. Y.Lee, (Singapore: World Scientific, 1988).

[18] M. Casdagli, D. Des Jardins, S. Eubank, J.D. Farmer, J. Gibson, J. Theiler, & N. Hunter, Non-linear modelling of chaotic time series: Theory and Applications, J. Kim & J. Stringer, eds., *Applied Chaos* (New York: Wiley 1992).

[19] H.D.I. Abarbanel, R. Brown, J.J. Sidrorowich, & L.S. Tsimring, The analysis of observed chaotic data in physical systems, *Tec. Mod. Phys., 65*, Oct. 1993.

[20] G.L. Sicuranza, Quadratic filters for signal processing, *Proc. IEEE, 80*, Aug. 1992, 1263–1285.

[21] E. Mumolo, A. Carini, & D. Francescato, ADPCM with nonlinear predictors, *Signal Processing 7: Theories and Applications,* Vol. 1, (Elsevier, 1994).

[22] B. Townshend, Non-linear prediction of speech signals, in *Nonlinear modeling and forecasting, SFI Studies in the Sciences of Complexity*, Vol. 12 (Addison-Wesley, 1992).

[23] A.C. Singer, G.W. Wornell, & A.V. Oppenheim, Non-linear autoregresive modelling and estimation in the presence of noise, *Digital Signal Processing, 4*, 1994, 207–221.

[24] A. Kumar & A. Gersho, LD-CELP speech coding with nonlinear prediction, *IEEE Signal Processing Letters, 4*(4), 1997, 89–91.

[25] N. Ma & G. Wei, Speech coding with non-linear local prediction model, *ICASSP'98, Vol. 2*, 1101–1104.

[26] H. Tong, *Non-linear time series: A dynamical system approach* (Oxford: Oxford University Press, 1990).

[27] M.B. Priestley, *Non-linear and non-stationary time series analysis* (Academic Press, 1988).

[28] M. Casdagli, Non-linear prediction of chaotic time series, *Physica D, 35D*, 1989, 335–356.

[29] M. Birgmeier, A fully Kalman-trained radial basis function network for nonlinear speech modeling, *Proc. 1995 IEEE Int. Conf. Neural Networks, ICNN'95*, Perth, November 1995.

[30] M. Birgmeier, Non-linear prediction of speech signals using radial basis function networks, *EUSIPCO Vol. 1*, 1996, 459–462.

[31] F. Diaz & A. Figueiras, Non-linear prediction for speech coding using radial basis functions, *Proc. Int. Conf. Acoust. Speech, Sign. Proc.*, 1995, 788–791.

[32] F. Diaz de Maria & A.R. Figueiras-Vidal, Radial basis functions for nonlinear prediction of speech in analysis-by-synthesis coders, *Proc. IEEE Workshop on Non-linear Signal and Image Processing NSIP'95*, Haldiki, June 1995.

[33] P. Yee & S. Haykin, A dynamic regularized Gaussian radial basis function network for non-linear, nonstationary time series prediction, *Proc. Int. Conf. Acoust. Speech, Sign. Proc.*, 1995.

[34] I. Mann & S. McLaughlin, Stable speech synthesis using recurrent radial basis functions, *Proc. 6th European Conference on Speech Communication and Technology*, Vol. 5, September 1999, 2315–2318.

[35] A. Lapedes & R. Farber, How neural nets work, in *Evolution, learning and cognition*, (World scientific, 1988).

[36] H. Reininger & D. Wolf, Non-linear prediction of stochastic processes using neural networks, in *Signal processing 5: Theories and applications* (Elsevier, 1990).

[37] M. Faúndez, E. Monte, & F. Vallverdú, A comparative study between linear and non-linear speech prediction, *Biological and Artificial Computation: from Neuroscience to Technology*, (IWANN'97), (Springer Verlag, 1997).

[38] M. Faúndez, F. Vallverdú, E. Monte, Non-linear prediction with neural nets in ADPCM, *Proc. Int. Conf. Acoust. Speech, Sign. Proc.*, Vol. 1, Seattle, 1998, 345–348.

[39] M. Faúndez & O. Oliva, ADPCM with non-linear prediction, *EUSIPCO'98*, Rhodes, 1205–1208.

[40] M. Faúndez, Adaptive hybrid speech coding with a MLP/LPC structure, in Alicente, (ed.), *IWANN'99*, LNCS Vol. 2, (Springer-Verlag), 814–823.

[41] M. Faúndez, Non-linear predictive models computation in adpcm schemes, *European Signal Processing Conf.*, EUSIPCO'2000, Vol. 2, Tampere, September, 2000, 813–816.

[42] L. Wu & M. Niranjan, On the design of non-linear speech predictors with recurrent nets, *Proc. Int. Conf. Acoustics, Speech & Signal Processing*, Adelaide, April 1994, II-529–II-532.

[43] L. Wu, M. Niranjan, & F. Fallside, Fully vector quantized neural network-based code-excited non-linear predictive speech coding, *IEEE trans. Speech and Audio Processing, 2*(4), 1994.

[44] S. Haykin & L. Li, Non-linear adaptive prediction of nonstationary signals, *Signal Process., 43*, 1995, 526–535.

[45] A. Hussain, Novel artificial neural-network architectures and algorithms for non-linear dynamical system modelling and digital communications applications, doctoral diss., University of Strathclyde, Glasgow, UK, 1996.

[46] E.N. Lorenz, Atmospheric predictability as revealed by naturally occurring analogues, *J. Atmospheric Sciences, 26*, 1969, 636–646.

[47] R.E. Bogner, "Signal prediction by pattern search, *J. Institution Electron. Telecomm. Engineers, 34*(1), 1988, 43–49.

[48] R.E. Bogner & T. Li, Pattern search prediction of speech, *Proc. Int. Conf. Acoustics, Speech & Signal Processing*, Vol. 1, Glasgow, May 1989, 180–183.

[49] S. Yakowitz, Nearest neighbour methods for time series analysis, *J. Times Series Anal., 8*(2), 1987, 235–247.

[50] A. Gersho, Optimal non-linear interpolative vector quantization, *IEEE Trans. Comm., 38*(9), 1989, 1285–1287.

[51] S. Wang, E. Paksoy, & E. Gersho, Performance of non-linear prediction of speech, *Proc. Int. Conf. on Spoken Language Processing*, Kobe, November 1990, 29–32.

[52] A. Gersho & R.M. Gray, Vector quantization and signal compression (Kluwer, 1992).

[53] A.C. Singer, G.W. Wornwell, & A.V. Oppenheim, Codebook prediction: A non-linear signal model paradigm, *Proc. Int. Conf. Acoustics, Speech & Signal Processing, 5*, San Francisco, March 1992, 325–328.

[54] Y. Lee & D. Johnson, Nonparametric prediction of non-gaussian time series, *Proc. Int. Conf. Acoustics, Speech & Signal Processing, 4*, Minneapolis, 1993, 480–483.

[55] B.L.S. Prakasa Rao, Nonparametric functional estimation, (Academic Press, 1983).

[56] L. Devroye & L. Gyorfi, Nonparametric density estimation—The L1 view (Wiley, 1985).

[57] H.M. Teager & S.M. Teager, Evidence for non-linear sound production mechanisms in the vocal tract, W.J. Hardcastle & A. Marchal, eds., *Speech Production and Speech Modelling*, NATO Advanced Study Institute Series D, Vol. 55, (Bonas, France: NATO, 1989).

[58] J.F. Kaiser, Some observations on vocal tract operation from a fluid flow point of view, in I.R. Titze & R.C. Scherer (eds.), *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control*, (Denver, CO: Denver Center for the Performing Arts, 1983).

[59] P. Maragos, J.F. Kaiser, & T.F. Quatieri, Energy separation in signal modulations with application to speech analysis, *IEEE Trans. Signal Processing, 41*, October 1993, 3024–3051.

[60] P. Maragos & A. Potamianos, Fractal dimensions of speech sounds: Computation and application to automatic speech recognition, *J. of Acoustical Society of America, 105*(3), 1999, 1925–1932.

[61] P. Maragos, Modulation and fractal models for speech analysis and recognition, *Proc. COST249 European Meeting*, Porto, Portugal, February 1998.

[62] A. Potamianos & P. Maragos, Speech processing applications using an AM–FM modulation model, *Speech Communication, 28*, 1999, 195–209.

[63] A. Potamianos & P. Maragos, Time-frequency distributions for automatic speech recognition, forthcoming in *IEEE Trans. on Speech and Audio Processing*.

[64] R. Soheili, A.M. Kondoz, & B. Evans, An 8kb/s LD-CELP with improved excitation and perceptual modelling, *Proc. Int. Conf. Acoust. Speech Sign. Process, Vol. 2*, Minneapolis, April 1993, 616–619.

[65] M. Faúndez & D. Rodriguez, Speaker recognition using residual signal of linear and non-linear prediction models, *5th Int. Conf. on Spoken Language Processing*, ICSLP'98, Vol. 2, 1998, 121–124.

[66] M. Faúndez, Speaker recognition by means of a combination of linear and non-linear predictive models, *EUROSPEECH'99*, Vol. 2, Budapest, 763–766.

[67] A. Hussain & D.R. Campbell, Binaural sub-band adaptive speech enhancement using artificial neural networks, *J. Speech Communication, Special Issue on Robust Speech Recognition for Unknown Communication Channels, 25*, 1998, 177–186.

[68] A. Hussain, Multi-sensor neural network processing of noisy speech, *Int. J. Neural Systems, 9*(5), 1999, 467–472.

[69] W.G. Knecht, M.E. Schenkel, & G.S. Moschytz, Neural network filters for speech enhancement, *IEEE Trans. Speech & Audio Proc., 3*(6), 1995, 433–438.

[70] F. Hienle, R. Rabenstein, & A. Stenger, Measuring the linear and non-linear properties of electro-acoustic transmission systems, *Proc. IEEE Int. Workshop on Acoustics, Echo & Noise Cancellation (IWAENC'97)*, London, September 11-12, 1997, 33–36.

[71] E.A. Wan & A.T. Nelson, Networks for speech enhancement, in S. Katagiri (ed.), *Handbook of neural networks for speech processing* (Boston: Artech House, 1998).

[72] A. Hussain & D.R. Campbell, Intelligibility improvements using binaural diverse sub-band processing applied to speech corrupted with automobile noise, *IEE Proc. Vision, Image and Signal Processing, 148*(2), 2001, 127–132.

[73] A. Behrman & R.J. Baken, Correlation dimension of electroglottographic data from healthy and pathologic subjects, *J. Acoust. Soc. Am., 102*(4), 1997, 2371–2381.

[74] A. Behrman, Global and local dimensions of vocal dynamics, *J. Acoust. Soc. Am., 105*(1), 1999, 432–443.

[75] D.A. Berry, H. Herzel, I.R. Titze, & K. Krischer, Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions, *J. Acoust. Soc. Am., 95*(6), 1994, 3595–3604.

[76] A. Esposito, The amplitudes of the peaks in the spectrum as acoustic attributes of the place of articulation, *Proc. ICPhS 95*, Vol. 1, 1995, 132–135.

[77] W.M. Hartmann, Signals, sounds, and sensations (New York: Springer, 1996).

[78] M. Kenstowicz, Phonology in generative grammar (Blackwell, 1994).

[79] A. Kumar & S.K. Mullick, Non-linear dynamical analysis of speech, *J. Acoust. Soc. Am., 100*(1), 1996, 615–629.

[80] E.E. Shriberg, Acoustic properties of disfluent repetitions, *Proc. ICPhS 1995*, Vol. 4, 1995, 384–387.

[81] I. Steinecke & H. Herzel, Bifurcations in an asymmetric vocal fold model, *J. Acoust. Soc. Am., 97*, 1995, 1874–1884.

[82] A.J. Watkins & S.J. Makin, Perceptual compensation for speaker differences and for spectral envelope distortion, *J. Acoustic Society of America, 96*, 1994, 1263–1282.

[83] S.A.J. Wood, Assimilation or coarticulation? Evidence from the temporal co-ordination of tongue gestures for the palatization of Bulfarian alveolar stops, *J. Phonetics, 24*, 1996, 139–164.

[84] A. Behrman, Global and local dimensions of vocal dynamics, *J. Acoust. Soc. Am., 105*(1), 1999, 432–443.

[85] J. Schoentgen, Stochastic models of jitter, *J. Acoust. Soc. Am., 109*(4) 2001.

[86] H. Herzel, A.D. Berry, I.R. Titze, & M. Saleh, Analysis of vocal disorders with methods from non-linear dynamics, *J. Speech Hear. Res., 37*, 1994, 1008–1019.

[87] N. Mtetwa & L.S. Smith, An FPGA implementation of a leaky integrate-and-fire neuron model to investigate stochastic resonance, *3rd Conf. on Postgraduate Research in Electronics, Photonics, Communications and Software*, UK, April 2001.

[88] N. Mtetwa, L.S. Smith, & A. Hussain, Stochastic resonance and finite resolutions in a leaky integrate-and-fire neuron model, paper submitted to the European Symposium on Artificial Neural Networks (ESANN), 2002.

Biographies

M. Faúndez-Zanuy received his degrees?

S. McLaughlin received his degrees?

A. Esposito received degrees?

A. Hussain received degrees?

J. Schoentgen received degrees?

G. Kubin received degrees?

W.B. Kleijn received degrees?

P. Maragos received degrees?