MODELING RESONANCES WITH PHASE MODULATED SELF-SIMILAR PROCESSES

Alexandros G. Dimakis

Dept. of EECS
University of California, Berkeley*
adim@eecs.berkeley.edu

Petros Maragos

School of ECE
National Technical University of Athens
maragos@cs.ntua.gr

ABSTRACT

In this paper we propose a nonlinear model for time-varying random resonances where the instantaneous phase (and frequency) of a sinusoidal oscillation is allowed to vary proportionally to a random process that belongs to the class of α-stable self-similar stochastic processes. This is a general model that includes phase modulations by fractional Brownian motion or fractional stable Levy motion as special cases. We explore theoretically this random modulation model and derive analytically its autocorrelation and power spectrum. We also propose an algorithm to fit this model to arbitrary resonances with random phase modulation. Further, we apply the above ideas to some speech data and demonstrate that the model is suitable for fricative sounds.

1. INTRODUCTION

Oscillations and resonances are phenomena of great importance in physical systems. Their modeling and detection in signals emanating from such systems are significant problems in signal processing and communications. Despite the mathematical tractability of the linear models, the majority of significant problems in engineering and sciences involves nonstationary signals. Thus, real-world oscillations may usually have a time-varying frequency, and this time variation may be of a random nature. Examples include frequency fluctuations in quartz crystals, atomic clocks, heart beat variation and resonances in speech sounds. In all these phenomena, it is the *frequency and phase fluctuations* that have been observed to have long range correlations and 1/f spectra (evidence of self-similarity). A related model is described in [4] where the authors present an *amplitude modulated* self-similar process. For relatively recent expositions on 1/f self-similar processes for applications in signal processing and communications see [1, 3, 15, 20] and the references therein.

In this paper we advance two main ideas, which we explore both theoretically and experimentally: First, we propose a random phase modulation model for arbitrary oscillations where fluctuations in their instantaneous frequency and phase are represented by self-similar signals. Second, we apply this model to explore the structure of resonances in turbulent speech sounds.

Our contributions in theory consist of using the class of *self-similar α-stable processes* as stochastic representations of the random instantaneous phase in our model and in deriving analytically the autocorrelation and power spectrum of the modulated process. This theoretical framework is quite general. For instance, popular models such as the *fractional stable Levy motion (FSLM)* [18] and the *fractional Brownian motion (FBM)* [10] are special cases of stable self-similar processes. From the algorithmic and experimental side, our contributions consist of developing an algorithm to estimate the model parameters, applying it to real speech sounds, and testing its validity. A summary of our results has also been presented in [13] in the context of nonlinear speech modeling.

The thematic organization of the paper is as follows: In section 2 we briefly summarize the main concepts and models needed for the analysis in this paper. The random phase modulation model is discussed in Section 3 where the autocorrelation and its power spectrum are analytically derived. Parameter estimation and application to turbulent speech sounds are discussed in Section 4

2. PRELIMINARIES

We begin with some definitions and basic properties of self-similar real processes following mainly [5].

A stochastic process $X(t)$, $t \geq 0$, is called (strict-sense) *self-similar* if there exists a parameter $H > 0$, called similarity exponent, such that, for any scale $r > 0$, $X(rt) \stackrel{d}{=} r^H X(t)$, where $\stackrel{d}{=}$ denotes equality of all finite-dimensional distributions. For short, this structure is denoted by H-ss. The above strict sense can become a wide-sense self-similarity if we restrict it only to the mean and correlation. Self-similarity often implies a 1/f spectrum [7]; i.e., an H-ss process has a (generalized) power spectrum of the form $S_x(\omega) \propto 1/|\omega|^\gamma$ for some spectral exponent $\gamma = 2H + 1$ for a wide range of frequencies.

A process $X(t)$ is said to have *stationary increments* $Y_s(t) = X(t+s) - X(t)$ if all finite-dimensional distributions of $Y_s(t)$ are independent of t. Throughout this paper we will be interested in stochastic processes which are *self-similar* with exponent H and have *stationary increments*, denoted as H-sssi processes.

Mandelbrot and van Ness [10] proposed the most popular model for self-similar processes, called fractional Brownian Motion (FBM). FBM is the only Gaussian H-sssi process and can be obtained via a stochastic fractional integration of the standard Brownian motion. The main disadvantage of FBM is that due to its Gaussianity it fails to model impulsiveness. By impulsiveness, we are referring to non-negligible probability of observing values extremely distant from the mean. This feature can be captured by using α-stable distributions that exhibit "heavy tails" that decay much slower than Gaussian distributions. Specifically, *symmetric a-stable (SαS)* distributions are defined by their characteristic function which has the form

$$\Phi(\theta) = \exp(-|s\theta|^\alpha), \quad 0 < \alpha \leq 2. \tag{1}$$

where s is a scale parameter. Inverse Fourier transform yields the corresponding probability density functions. For $\alpha = 2$ we get the

*This work was performed at NTUA and was partially supported by the European network of excellence MUSCLE and the STREP HIWIRE.

Gaussian, whereas $\alpha = 1$ yields the Cauchy density. There exist no closed-form expressions for the density function for α different than 1 and 2. A real-valued stochastic process $X(t)$ is said to be SαS if any linear combination $\sum_k a_k X(t_k)$ has a SαS distribution.

A popular model for H-sssi SαS processes is the *fractional stable Levy motion (FSLM)* [18], defined via stochastic fractional integration of the Levy process. Note that the class of H-sssi SαS processes is very general. It includes FBM, FLSM, and a number of other processes used to model impulsiveness and long range dependence as special cases. Applications of signal processing with α-stable distributions can be found in [1, 3, 14, 15].

3. A MODEL FOR RANDOM RESONANCES

We assume the general phase modulation model[1]

$$X(t) = A(t)\cos[\phi(t)], \quad \phi(t) = \omega_c t + P(t) + \phi_0$$

where ω_c is the center frequency of the resonance. For the subsequent analysis we ignore the instantaneous amplitude and assume it constant. Further, we model the nonlinear instant phase $P(t)$ as an H-sssi SαS process. This process $P_H(t)$ has a similarity exponent $H > 0$, stationary increments and an symmetric α-stable distribution for each t. Thus we shall work with the *random phase modulation process*

$$X(t) = A\cos(\omega_c t + \lambda P_H(t) + \phi_0) \tag{2}$$

where the center frequency $\omega_c > 0$ is assumed a known constant, $\lambda > 0$ is the modulation index, ϕ_0 is the phase offset at $t = 0$, and $P_H(t)$ is an H-sssi process. The power spectrum of $P_H(t)$ is proportional to $1/|\omega|^\gamma$, where $\gamma = 2H + 1$. Thus, we are modeling the modulating signal $P(t)$ as an H-sssi stochastic process. The increments process and hence the instantaneous frequency $\omega_i(t) = \omega_c + \lambda P'_H(t)$ is a stationary process with a 1/f spectrum whose spectral exponent is $2H-1$. In this section we analytically derive the autocorrelation function and power spectrum of this phase modulated process and demonstrate a mathematical relation linking these processes with a-stable processes. The problems of testing the validity of the proposed model as well as fitting it to real data arise in the following sections.

3.1. Phase-Modulated H-sssi SαS Process

Lemma 1 *If $X(t)$, $t \geq 0$, is self-similar process with similarity exponent $H > 0$, then for a given t the r.v. $X(t)$ has a characteristic function $\Phi_{X(t)}(\theta, t)$ with the property:*

$$\Phi_{X(t)}(\theta, t) = \Phi_{X(1)}(t^H\theta, 1) \tag{3}$$

Proof: $X(rt) \stackrel{d}{=} r^H X(t)$, which implies $X(t) \stackrel{d}{=} t^H X(1)$. Therefore the characteristic function of the r.v. $X(t)$ is $\Phi_{X(t)} = \mathsf{E}[e^{j\theta X(t)}] = \mathsf{E}[e^{j(t^H\theta)X(1)}]$ which yields the result. ■

We can now present the basic theorem which determines the autocorrelation function of the phase modulated process.[2]

[1] This model was motivated by the experimental evidence in [11] for the AM-FM structure of fricative sounds with random noise-like instantaneous modulating signals representing the frequency fluctuations.

[2] Some elementary assumptions and implications of the theorem are inspired from a random frequency modulation model analyzed by Papoulis [16] where the nonlinear instant phase $P(t)$ was equal to $\int_0^t F(\tau)d\tau$ and the instant frequency $F(t)$ was a strict-sense stationary process.

Theorem 1 *Consider the random process $X(t)$ of (2) where A, ω_c and λ are real constants, ϕ_0 is a random variable uniformly distributed[3] over $[0, 2\pi)$ and independent of $P_H(t)$, and $P_H(t)$ is an α-stable H-sssi process with characteristic function at each t*

$$\Phi_{P_H(t)}(\theta, t) = \exp(-|s(t)\theta|^\alpha), \quad 0 < \alpha \leq 2. \tag{4}$$

where $s(t)$ is a positive scale parameter. Then:
(a) $X(t)$ is a wide sense stationary process with zero mean.
(b) Its autocorrelation function is given by

$$R_{xx}(\tau) = \frac{1}{2}\cos(\omega_c\tau)\exp(-|s(1)\lambda|^\alpha|\tau|^{\alpha H}) \ . \tag{5}$$

Proof: We define the complex processes

$$W(t) = \exp[j\lambda P_H(t)], \quad Z(t) = W(t)\exp[j(\omega_c t + \phi_0)] \tag{6}$$

Then, since $X(t) = [Z(t) + Z^*(t)]/2$, to check whether $X(t)$ is WSS it suffices to check the constancy of the mean of Z and the stationarity of the autocorrelation of Z and of the cross-correlation between Z and its conjugate Z^*. First, for the mean, $\mathsf{E}[Z(t)] = e^{j\omega_c t}\mathsf{E}[W(t)]\mathsf{E}[e^{j\phi_0}] = 0$. Second, for the correlations,

$$R_{zz^*}(t+\tau, t) = \mathsf{E}[w(t+\tau)w(t)]\mathsf{E}[e^{2j\phi_0}] = 0 \tag{7}$$

$$R_{zz}(t+\tau, t) = e^{j\omega_c\tau}\mathsf{E}[e^{j\lambda(P_H(t+\tau)-P_H(t))}] \tag{8}$$

Since the increments of $P_H(t)$ are stationary, the autocorrelation of $Z(t)$ and $W(t)$ can be written $R_{ww}(\tau) = \mathsf{E}[e^{j\lambda P_H(\tau)}] = \mathsf{E}[W(\tau)]$ and $R_{zz}(\tau) = e^{j\omega_c\tau}R_{ww}(\tau) = e^{j\omega_c\tau}\mathsf{E}[W(\tau)]$. Hence, $X(t)$ is a zero-mean WSS process. Now note that $\mathsf{E}[W(t)] = \mathsf{E}[e^{j\lambda P_H(t)}] = \Phi_{P_H(t)}(\lambda, t)$ where $\Phi_{P_H(t)}(\lambda, t)$ is the time dependent characteristic function (4) of the r.v. $P_H(t)$. Then, from the above results, it follows that the autocorrelation of $X(t)$ is $R_{xx}(\tau) = \frac{1}{2}\mathrm{Re}[R_{zz}(\tau)] = \frac{1}{2}\cos(\omega_c\tau)\Phi_{P_H(t)}(\lambda, \tau)$ Further, due to the self-similarity of $P_H(t)$, we use lemma (3) to obtain $\Phi_{P_H(t)}(\lambda, t) = \exp(-|s(1)\lambda|^\alpha|t|^{\alpha H})$. Combining the last two equations yields the desired formula (5). ■

The power spectrum can be found as the Fourier transform of $R_{xx}(\tau)$. This spectrum has a closed formula only for the special cases when $\alpha H = 1$ or $\alpha H = 2$, which correspond respectively to Cauchy or Gaussian resonances centered around $\pm\omega_c$. Next we analyze these special cases when the phase is FBM.

3.2. Phase is FBM

For the special case when the nonlinear instant phase $P_H(t)$ is an FBM with $0 < H < 1$, at each t $P_H(t)$ is a Gaussian r.v. with variance $\mathsf{Var}[P_H(t)] = \sigma_H^2|t|^{2H}$, $\sigma_H^2 = \mathsf{E}[P_H(1)^2]$ and characteristic function $\Phi_{P_H(t)}(\theta, t) = \exp(-\frac{1}{2}\sigma_H^2\theta^2|t|^{2H})$ Therefore, if in the previous theorem we set α=2 and $s(1) = \sigma_H/\sqrt{2}$, we determine the autocorrelation function of $X(t)$. The power spectrum $S_H(\omega)$ of $X(t)$ can be found as the Fourier transform of $R_{xx}(\tau)$. However, there is no closed formula for arbitrary H. There are only two special cases where the Fourier transform can be analytically derived. Specifically for $H = 0.5$ we obtain the following power spectrum: $S_{H=0.5}(\omega) = \frac{\sigma_H^2}{\sqrt{2\pi}}[\frac{1}{\lambda^4\sigma_H^4+4(\omega-\omega_c)^2} + \frac{1}{\lambda^4\sigma_H^4+4(\omega+\omega_c)^2}]$ which is a sum of two Cauchy resonances centered at ω_c and $-\omega_c$. For $H = 1$ (which is only a limit case for

[3] A more general assumption for ϕ_0 (for which the theorem is valid) is $\mathsf{E}[e^{j\phi_0}] = \mathsf{E}[e^{j2\phi_0}] = 0$.

FBM) we obtain: $S_{H=1}(\omega) = \frac{1}{2\sqrt{\lambda^2\sigma_H^2}}[\exp(-\frac{(\omega-\omega_c)^2}{2\lambda^2\sigma_H^2}) + \exp(-\frac{(\omega+\omega_c)^2}{2\lambda^2\sigma_H^2})]$ which is a sum of two Gaussian resonances centered at ω_c and $-\omega_c$. In this case, the resonance spectrum has the same form as the frequency response of a Gabor filter.

3.3. Relation with a-Stable Distributions

The characteristic function (1) of an α-stable process is of the same power form[4] as the autocorrelation function (5) we derived for the phase-modulated process by replacing the α exponent of the former with αH. This analogy leads to the following interesting conclusion: The power spectrum of a phase-modulated H-sssi α-stable process has an analytic form that is equivalent with the probability density function of an αH-stable distribution centered at the carrier frequency ω_c. This result seems to establish an underlying relation between phase-modulated self-similar processes and α-stable distributions which is interesting for further exploration. A related result connecting power law shot noise with α-stable distributions can be found in [17].

It is well known that α-stable distributions are invariant to convolution and constitute an attractor under consecutive convolutions of other density functions. Drawing an analogy with our results, the power spectrum of the processes we propose is also invariant and attracting with respect to convolution of power spectra, which is multiplication of the autocorrelation functions in the time domain. This important result implies that the modulated self-similar processes we proposed constitute attracting processes with respect to multiplication of autocorrelations for independent stochastic processes. This seems to be a mathematical link between the multiplicative models for turbulence developed by the Russian school[8] and our proposed modulation model and is interesting for further exploration.

4. PARAMETER ESTIMATION AND APPLICATION TO SPEECH SOUNDS

Motivated by experimental evidence in [19, 11, 12] for the existence of important nonlinear aerodynamic phenomena in speech sounds as well as by the theoretical background relating self-similar stochastic processes with turbulence, we advance the conjecture made in [11] that turbulent speech sounds accept a modulation model with a random noise-like signal representing phase fluctuations. To fit and estimate the proposed model we need to isolate speech resonances and estimate their instant phase modulation which can be efficiently achieved by using the energy separation algorithm (ESA) [11]. Therefore, the proposed algorithm consists of following five steps:
(1) Isolate the resonance by bandpass filtering the speech signal.
(2) Use the ESA demodulation algorithm to estimate the AM and FM signals, $A(t)$ and $F(t)$.
(3) Determine the instant phase modulation signal $P(t)$ by integrating the instant frequency: $\hat{P}(t) = 2\pi\int_0^t(F(\tau)-F_c)d\tau$, where F_c is the short-time average of $F(t)$.
(4) Estimate the α exponent that best models the instant phase modulation signal as a realization of a SαS process.
(5) Estimate the γ (or equivalently, H) exponent that best models the phase modulation signal $\hat{P(t)}$ as a 1/f self-similar signal.

[4] The multiplication with the cosine function in (5) merely causes a shift centering the resonance at ω_c and $-\omega_c$.

The last two steps are unquestionably the most perplexed. To estimate the α exponent we exploit the fact that the increments of the instant phase (namely instant frequency) are *stationary* SαS random variables with the same α. Therefore, we used the Koutrouvelis [15, 9] regression on the sample characteristic function to estimate the α parameter for the instant frequency process (other methods can also be used however). For fricative speech phonemes the estimated α values were roughly in the $[1.6, 2]$ interval.

The problem of estimating the γ exponent has been approached from a number of different angles utilizing time, frequency and wavelet-domain approaches. See [3] for an excellent review. After extensively testing several techniques, the method which performed best in our experiments was the GPH local spectral estimator [6]. The estimated γ exponents varied in the $[2.6, 3]$ interval.

The estimation algorithm was tested by creating various artificial resonances. The reconstructed phase modulation signal $\hat{P(t)}$ was a good approximation of the original artificial $P(t)$, the variances of the estimated exponents however, where significant and more work is required to reach exact conclusions.

Our experiments indicate that the measured exponents are correlated with the nature of each phoneme. Namely, voiced fricatives usually generated larger γ exponents than unvoiced. This is a rational thing to expect since unvoiced fricatives (like /f/ or /s/) seem to be less smooth and phase fluctuations with larger γ exponents indicate smoother paths. The α exponent measures the impulsiveness of the instant phase. We observed that the unvoiced fricatives had smaller exponents which indicate longer trails and thus more impulsive behavior. Note that the model is suitable for modeling *broad resonances* that are well isolated in the spectrum. Experiments on formants of smooth vowels (like /a/) indicate that the phase modulation signals have exponentially decaying spectra thus exhibiting no long range correlations or self-similarity.

Figure (1) demonstrates the application of the described algorithm to a /z/ phoneme (from TIMIT database). Fig(1b) illustrates the spectrum of the sound and the Gabor filter used to isolate the resonance. Fig(1c) illustrates the instant frequency as estimated by ESA algorithm while Fig(1d) is the instant phase modulation signal which is modeled as a self-similar process. Fig(1e) is the power spectrum of the phase modulation and the estimated slope. Fig(1f) illustrates the variance of the the wavelet detail coefficients. Following [20], if the wavelet detail coefficients have variances that behave like $\mathsf{Var}(x_n^m) \propto 2^{-\gamma m}$ as a function of scale m, then the process is "nearly 1/f". The power spectrum (1e) and the wavelet variance (1f) approximate straight lines for a broad range of scales and that constitutes evidence that the proposed model is suitable for fricative sounds. We have also performed numerous other similar experiments on real speech signals [13]. In all these experiments we have found evidence that the phase modulation of speech resonances for fricative phonemes exhibits self-similarity.

5. CONCLUSIONS AND DISCUSSION

In this paper we have proposed a random phase modulation model for resonances where the instant phase modulation signal is an α-stable self-similar process. We analytically derived the second order statics and proved that the power spectrum has the same form as an α-stable density. We further proposed an algorithm to estimate its parameters, and some experimental evidence of validity for fricative speech.

The work herein is a continuation of previous work [11, 12, 13] on modeling resonances with AM-FM signals and on model-

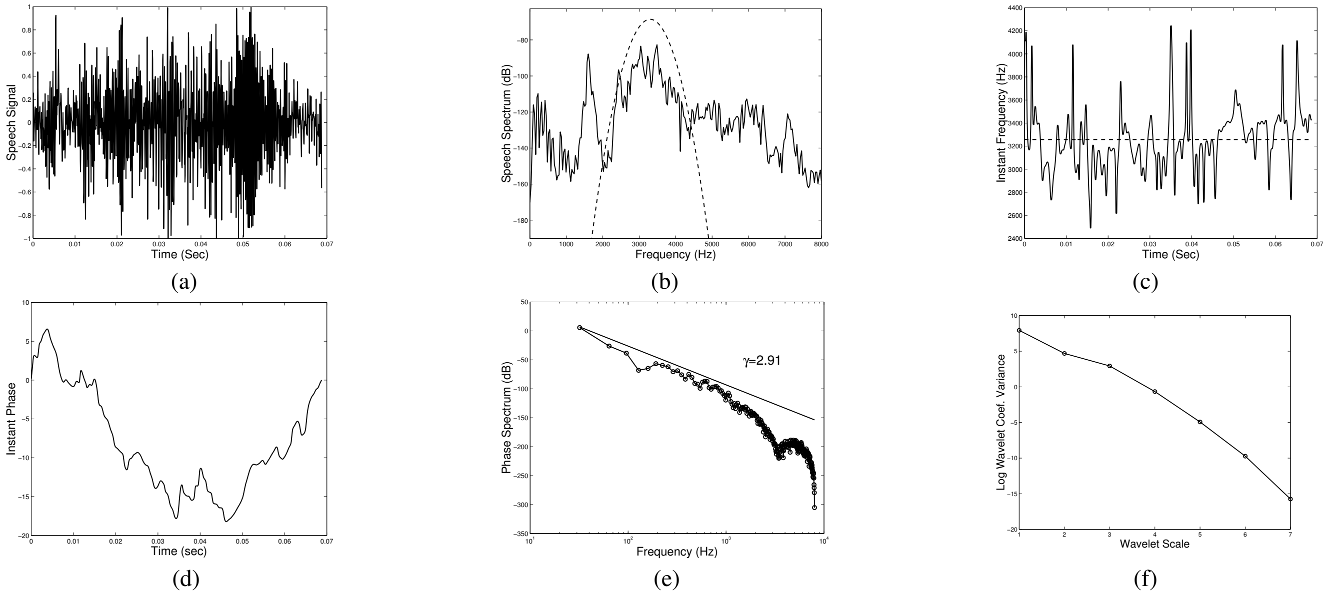

Fig. 1. Experiment with /z/. a) Speech signal $s(t)$. b) Power Spectrum of $s(t)$ and Gabor filter. c) Instant Frequency. d) Phase modulation $\hat{P}(t)$. e) Power Spectrum of $\hat{P}(t)$ and estimated slope. f) Variance of the wavelet coefficients.

ing turbulence in fricative and other speech sounds with random fractal signals. Our on-going work in this area includes better estimation algorithms as well as a statistical study relating estimated exponents with types of sounds. Such relations can be used in speech recognition applications. Relating our model with turbulence and multifractals is another promising research direction. Finally, we believe that our model can be used in the study of other time-varying oscillating physical systems since self-similar fluctuations in periodic phenomena seem to be ubiquitous in nature.

6. REFERENCES

[1] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch, "Multiscale Nature of Network Traffic", *IEEE Sig. Proc. Mag.*,vol.19, pp.28-46, May 2002.

[2] A. C. Bovik, P. Maragos, and T.F. Quatieri, "AM-FM Energy Detection and Separation in Noise Using Multiband Energy Operators", *IEEE Trans. Sig. Proc.*, Dec. 1993.

[3] O. Cappe, E. Moulines, J. Pesquet, A. Petropulu, and X. Yang, "Long-Range Dependence and Heavy-Trail Modeling for Teletraffic Data", *IEEE Sig. Proc. Mag.*, May 2002.

[4] A. Coron, P. Flandrin and M. Gache, "Modulated Fractional Gaussian Noise: A process with non-Gaussian wavelet details", Proc. IEEE-SP Int. Symp. on Time freq. and Time Scale An. Oct 1998.

[5] P. Embrechts and M. Maejima, *Selfsimilar Processes*, Princeton University Press, 2002.

[6] J. Geweke and S. Porter-Hudak, "Estimation and application of long memory time series models", *J. Time Ser. An.*, 1983.

[7] P. Flandrin, "On the spectrum of fractional Brownian motions", IEEE Tran. Info. Theory, Jan. 1989.

[8] U. Frisch, *Turbulence: The Legacy of A.N. Kolmogorov*, Cambridge Univ. Pr. 1995.

[9] I. A. Koutrouvelis "Regression type Estimation of the Parameters of Stable Laws", J. Amer. Stat. Assoc., Vol. 75(Dec.).

[10] B. Mandelbrot and J. Van Ness, "Fractional Brownian motion, fractional noises and applications", *SIAM Rev.*, 1968.

[11] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy Separation in Signal Modulations with Application to Speech Analysis", *IEEE Trans. Sig. Proc.*, Oct.1993.

[12] P. Maragos and A. Potamianos, "Fractal Dimensions of Speech Sounds: Computation and Application to Automatic Speech Recognition", *J. Acoust. Soc. Amer.*, March 1999.

[13] P. Maragos, A. Dimakis and I. Kokkinos, "Some advances in nonlinear speech modeling using Modulations, Fractals and Chaos". *Proc. Int. Conf. DSP-02*, Greece, July 2002.

[14] C.L. Nikias and A.P. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*, Prentice Hall, 1993.

[15] C. L. Nikias and M.Shao, *Signal Processing with Alpha-Stable Distributions*, Wiley, 1995.

[16] A. Papoulis, "Random Modulation: A Review", *IEEE Trans. Acoust., Speech & Sig. Proc.*, pp.96-105, Feb. 1983.

[17] A.P. Petropulu, J.C. Pesquet, X. Yang and J. Yin "Power-law shot noise and its relationship to long-memory α-stable processes", *IEEE Trans. Sig. Proc.*, July 2000.

[18] G. Samorodnitsky and M.S. Taqqu, *Stable Non-Gaussian Random Processes*, Chapman and Hall, 1994.

[19] H. M. Teager and S. M. Teager, "Evidence for Nonlinear Sound Production Mechanisms in the Vocal Tract", in *Speech Production and Speech Modelling*, NATO Adv. Study Inst. Series D, vol. 55, France, July 1989.

[20] G. Wornell, *Signal Processing with Fractals: A Wavelet-Based Approach*, Prentice Hall, 1995.