Phase-Modulated Resonances Modeled as Self-Similar Processes With Application to Turbulent Sounds

Alexandros G. Dimakis, *Student Member, IEEE,* and Petros Maragos, *Fellow, IEEE*

***Abstract*—In this paper, we propose a nonlinear stochastic model for time-varying resonances where the instantaneous phase (and frequency) of a sinusoidal oscillation is allowed to vary proportionally to an α-stable self-similar stochastic processes. The main motivation of our work stems from previous experimental and theoretical evidence that speech resonances in fricative sounds can be modeled phenomenologically as AM–FM signals with randomly varying instantaneous frequencies and that several signal classes related to turbulent phenomena are self-similar 1/f processes. Our general approach is to model the instantaneous phase of an AM–FM resonance as a self-similar α-stable process. As a special case, this random phase model includes the class of random fractal signals known as fractional Brownian motion. We theoretically explore this random modulation model and analytically derive 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 real speech data and demonstrate that this model is suitable for resonances of fricative sounds.**

***Index Terms*—Alpha stable, fractal, fractional Brownian motion, modulation, 1/f process, power-law, self-similar, speech, turbulence.**

I. 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. Examples and application areas include spectrum estimation, speech and sound processing, waveform modulation, and general time-series analysis. The archetypal model for single-component oscillations is a linear harmonic oscillator whose motion is governed by a linear differential equation with constant coefficients

$$x''(t) + 2\beta x'(t) + \omega_0^2 x(t) = u(t) \tag{1}$$

where $x(t)$ may represent displacement in mechanical oscillators or an electric quantity (e.g., charge, current) in electrical oscillators, $x' = dx/dt$, β is a damping coefficient, ω_0 is the natural undamped oscillation frequency, and $u(t)$ is an excitation signal. We henceforth assume that $\beta < \omega_0$ so that the system can oscillate. The oscillator's zero-input response is

$$x(t) = A\exp(-\beta t)\cos(\omega_c t + \theta) \tag{2}$$

where $\omega_c = \sqrt{\omega_0^2 - \beta^2}$ is the damped oscillation frequency. This sinusoid is "stationary," i.e. the sine parameters are constant with the possible exception of an exponentially-decaying amplitude if damping is present. Such stationary exponentials and sines form the basis of linear models for many categories of real-world signals, e.g., the LPC model in speech and the ARMA models for general time-series.

Despite the mathematical tractability of the above models, the majority of significant problems in engineering and sciences involve 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 beats, and resonances in speech sounds. There are at least two ways in signal processing to approach nonstationarity in oscillation parameters: 1) Linear time-varying systems represented by differential or difference equations with time-varying coefficients, and 2) nonlinear signal models of sines with joint amplitude modulation (AM) and frequency modulation (FM) or phase modulation (PM)

$$\begin{aligned} x(t) &= a(t)\cos[\phi(t)], \quad \omega_i(t) = \phi'(t) \\ \phi(t) &= \omega_c t + p(t) = \int_0^t \omega_i(\tau)d\tau + \phi_0. \end{aligned} \tag{3}$$

This *AM–FM signal* is a sine with time-varying instantaneous amplitude $a(t)$, instantaneous frequency $\omega_i(t)$, and nonlinear instantaneous phase $p(t)$. Note that the two approaches are related. For example, if the damping and/or frequency related coefficients in the harmonic oscillator's equation (1) become time-varying functions, then it can be shown (see Van der Pol [48] for a classic analysis) that an *approximate solution* of the time-varying differential equation is of the AM–FM type. In general, given an AM–FM signal, we can construct a differential equation that can generate it. The converse, however, is not always possible. In the work presented herein, we focus on the

Manuscript received December 30, 2003; revised December 10, 2004. This work was supported in part by the European IST FP-6 under the Network of Excellence MUSCLE and the STREP HIWIRE. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Athina Petropulu.

A. G. Dimakis was with the National Technical University of Athens, Athens 15773, Greece. He is now with the Electrical Engineering and Computer Science Department, University of California, Berkeley, CA 94720 USA (e-mail: adim@eecs.berkeley.edu).

P. Maragos is with the School of Electrical and Computer Engineering, National Technical University of Athens, Athens 15773, Greece (e-mail: maragos@cs.ntua.gr).

Digital Object Identifier 10.1109/TSP.2005.857010

signal modulation approach because it allows us some mathematical tractability of its theoretical properties when the phase is a self-similar stochastic process.

The purpose of this paper is twofold since 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 1/f 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)* [43] and the *fractional Brownian motion (FBM)* [24] 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 [8]and [29] in the context of nonlinear speech modeling.

Our main motivations for this work include the following: 1) In the previous work of Maragos *et al.* [27] on AM–FM modeling of speech resonances, where instantaneous variations in formant frequencies were found in both vowel and fricative sounds, the conjecture was made based on experimental evidences that turbulent speech sounds accept such an AM–FM model with a random noise-like signal representing the frequency fluctuations. 2) FBMs, as popularized by Mandelbrot and coworkers [22], [24], [49], have become a versatile model for random self-similar processes used in physics, mathematics, and engineering. In particular, their additional property of being fractal may be helpful for modeling turbulent sounds since the geometry of turbulence has been linked with fractals [22]. 3) Random signals with 1/f spectrum (some of which are called "fractional noises") are ubiquitous in engineering and sciences. They have an extremely broad spectrum of applications including areas such as electronics, biology, acoustics, optics, communications, network traffic, and economics. Statistical self-similarity is a fundamental property often encountered in 1/f processes. It is related to long-range dependence in the data: an attribute that is also present in the time variation of the parameters of the speech resonances to which we wish to apply the model. A subclass of self-similar 1/f processes is FBM. A broader class is the family of α-stable processes, popularized by Taqqu and coworkers [39], [42], [43], for modeling self-similarity and impulsiveness in data with long-range dependence. For relatively recent expositions on stochastic modeling with 1/f self-similar processes for applications in signal processing and communications; see [1], [6], [33], [50], and the references therein. In particular, we deal with the FSLMs [9], [18], [43].

The thematic organization of the paper is as follows: In Section II, we briefly summarize the main concepts, models, and algorithms from previous work, which is needed for the analysis in this paper. The random phase modulation model is discussed in Section III, where the autocorrelation and its power spectrum are analytically derived under the assumption that the phase is any self-similar symmetric α-stable process of the FSLM or FBM type. Parameter estimation and testing of the model validity are discussed in Section IV, where an algorithm is proposed to fit this model to arbitrary data. Section V contains the application to turbulent sounds in speech.

II. Preliminaries

A. 1/f Self-Similar Processes

We begin with some definitions and basic properties of self-similar real processes mainly following [9].

The 1/f family of random signals is an important class of stochastic processes $X(t)$, which are generally defined as having measured power spectra obeying a power law decay of the form

$$S_x(\omega) = \frac{S_x(1)}{|\omega|^\gamma} \tag{4}$$

for some spectral exponent γ and for a broad frequency range. This spectral behavior is related to statistical self-similarity.

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 $r > 0$

$$X(rt) \overset{d}{=} r^H X(t) \tag{5}$$

where $\overset{d}{=}$ denotes equality of all finite-dimensional distributions. In 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 implies the following properties:

$$X(0) = 0 \text{ almost surely} \tag{6}$$
$$\mathsf{E}[|X(t)|^p] = \mathsf{E}[|X(1)|^p]\,|t|^{pH}, \quad p = 1, 2, \ldots \tag{7}$$
$$R_{xx}(t,s) = r^{-2H} R_{xx}(rt, rs), \quad r > 0 \tag{8}$$

where $\mathsf{E}[\cdot]$ denotes expectation, and $R_{xx}(t,s)$ is the autocorrelation of $X(t)$. The proof of the above properties is a straightforward corollary of the definition (5). In general, self-similar processes are *nonstationary*, given the time dependence of their moments in (7).

Self-similarity implies a 1/f spectrum, i.e., an H-ss process has a power spectrum of the type (4) with spectral exponent $\gamma = 2H + 1$. A proof of this important fact is given in [38] for FBMs, based on the Fourier pair between autocorrelation and power spectrum of stationary processes. Roughly, if $X(t)$ is *wide-sense stationary (WSS)*, then (8) implies that its autocorrelation $R_x(\tau)$ satisfies the relationship $R_x(r\tau) = r^{2H} R_x(\tau)$. Applying Fourier transform to this equation yields (4). A more general proof for nonstationary processes can be obtained based on Fourier analysis of the time-averaged autocorrelation; details are beyond the scope of this paper.

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 that are *self-similar* with exponent H and have *stationary increments*, which are denoted as H-sssi processes. If $X(t)$ is self-similar, then this property is inherited by its increment process. Namely, the following are easy to prove:

$$X(t+r\Delta t) - X(t) \stackrel{d}{=} r^H \left[X(t+\Delta t) - X(t)\right] \quad (9)$$

$$\mathsf{E}\left[|X(t+\Delta t) - X(t)|^2\right] = \mathsf{E}\left[|X(1)|^2\right] |\Delta t|^{2H}. \quad (10)$$

1) Fractional Brownian Motion (FBM): The most well-known example of an H-sssi process is the standard Brownian motion $B(t)$, which is a zero-mean Gaussian process with $B(0) = 0$ a.s., variance equal to t, and independent stationary increments.

Mandelbrot and van Ness [24] generalized this behavior by proposing the fractional Brownian motion (FBM) as a zero-mean Gaussian process $B_H(t), 0 < H < 1$ that has the property

$$\mathsf{E}\left[B_H(t)B_H(s)\right] = \frac{\mathsf{E}\left[B_H(1)^2\right]}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right). \quad (11)$$

They proved that FBM is an H-sssi process and can be obtained via a stochastic fractional integration of the standard Brownian motion $B(t)$. Actually, for $H = 1/2$, the FBM coincides with $B(t)$ modulo a multiplicative constant. Additionally, they proved that FBM is unique in the sense that it coincides with all Gaussian H-sssi processes, thus making this model very attractive for analytic manipulation.

2) Self-Similar α-Stable Processes: The main disadvantage of FBM is that due to its Gaussianity it fails to model impulsiveness. By impulsiveness, we are referring to very bursty behavior that cannot be effectively modeled with distributions of finite variance.

This behavior can be captured by using α-stable distributions that exhibit "heavy tails" that decay much slower than Gaussian distributions. Specifically, *symmetric α-stable* $(S\alpha S)$ distributions are defined by their characteristic function, which has the form

$$\Phi(\theta) = \exp\left(-|s\theta|^{\alpha}\right), \quad 0 < \alpha \leq 2. \quad (12)$$

where s is a scale parameter. Inverse Fourier transform yields the corresponding probability density functions. For $\alpha = 2$, we get the 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)* [43]. Specifically, a stochastic process $X(t)$ is a called a Levy process if it has independent stationary increments, enjoys certain forms of stochastic continuity, and $X(0) = 0$ a.s. [9]. In addition, if $X(1)$ has an α-stable distribution, then it is called an *α-stable Levy process*, which is denoted by $Z_\alpha(t)$ with $0 < \alpha \leq 2$. A generalization is the FSLM defined via stochastic fractional integration of $Z_\alpha(t)$:

$$L_H(t) = \int_{-\infty}^{+\infty} \left(|t-s|^{H-\frac{1}{\alpha}} - |s|^{H-\frac{1}{\alpha}}\right) dZ_\alpha(s) \quad (13)$$

where $t \geq 0$, $0 < H < 1$, and $H \neq 1/\alpha$. FLSM is an H-sssi process, which coincides with the FBM when $\alpha = 2$. If $\alpha < 2$, the FLSM yields H-sssi, SαS processes with non-Gaussian infinite-variance distributions. 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.

A related model is described in [7], where the authors present an *amplitude modulated* self-similar process. For recent expositions on 1/f self-similar processes and α-stable distributions with applications in signal processing and communications, see [1], [6], [18], [33], and the references therein.

B. Nonlinear Speech Modeling

For several decades, the traditional approach to speech modeling has been the linear (source-filter) model, where the true nonlinear physics of speech production is approximated via the standard assumptions of linear acoustics and 1–D plane wave propagation of the sound in the vocal tract. However, since the 1980s, there has been strong theoretical and experimental evidence [3], [17], [30], [44], [45] for the existence of important nonlinear aerodynamic phenomena during the speech production that cannot be accounted for by the linear model. In the 1990s, this motivated the development of nonlinear signal processing systems that were suitable to detect various such phenomena and extract related information. Two such types of nonlinear phenomena in speech are *modulations* and *turbulence*. Next, we briefly review some progress in these fields.

1) Speech Resonance Modulations: Although the linear model assumes that each speech resonance (formant) signal is an exponentially damped cosine with constant frequency within 10–30 ms, there is much experimental and theoretical evidence for the existence of amplitude modulation (AM) and frequency modulation (FM) in such signals. This motivated Maragos *et al.* [27] to propose to *model each speech resonance with an AM-FM signal* $x(t) = a(t)\cos[\phi(t)]$ and the total speech signal as a superposition of such AM-FM signals: one for each formant. Here, the instantaneous cyclic frequency signal of the time-varying formant equals $f_i(t) = (1/2\pi)d\phi/dt$. The short-time formant frequency average $f_c = (1/T)\int_0^T f_i(t)dt$, where T is on the order of a pitch period, is viewed as the carrier frequency of the AM-FM signal.

For demodulating a single resonance signal, Maragos *et al.* [27] used the nonlinear Teager-Kaiser energy-tracking operator $\Psi[x(t)] \triangleq [x'(t)]^2 - x(t)x''(t)$ to develop the following nonlinear algorithm:

$$\sqrt{\frac{\Psi\left[x'(t)\right]}{\Psi\left[x(t)\right]}} \approx 2\pi f_i(t), \qquad \frac{\Psi\left[x(t)\right]}{\sqrt{\Psi\left[x'(t)\right]}} \approx |a(t)|. \quad (14)$$

This is the *energy separation algorithm (ESA)* and provides AM–FM demodulation by tracking the physical energy implicit in the source, producing the observed acoustic resonance signal and separating it into its amplitude and frequency components. It yields very good estimates of the instantaneous frequency signal $f_i(t) \geq 0$ and of the amplitude envelope $|a(t)|$ of an AM–FM signal, assuming that $a(t)$, $f_i(t)$ do not vary too fast (small bandwidths) or too greatly compared with the carrier frequency f_c.

There is also a *discrete* version of the ESA, called *DESA* [27], which is obtained by using a discrete energy operator. The DESA is an effective approach for speech demodulation and outperforms other classical demodulation approaches such as the Hilbert transform in terms of complexity and time resolution. Its main disadvantage is a moderate sensitivity to noise. An effective remedy for this problem is to use *filterbanks* [5], [37], because the bandpass filtering reduces the noise and transforms the wideband speech signal into narrowband AM–FM components that the ESAs can efficiently demodulate.

2) Speech Turbulence: The airflow during the production of several speech sound classes (e.g., in fricative sounds or during loud speech) contains various amounts of *turbulence*, which is defined in [47] as a "state of continuous instability" characterized by broad-spectrum rapidly varying (in space and time) velocity and vorticity. In the linear speech model, this has been dealt with simply by having a white noise source exciting the vocal tract filter. Nowadays, turbulence can be explored from several nonlinear aspects. Mandelbrot [22] and others have conjectured that several geometrical aspects of turbulence (e.g., shapes of turbulent spots, boundaries of some vortex types found in turbulent flows, and shape of particle paths) are *fractal* in nature.

In fluid dynamics [47], the energy cascade theory attempts to understand turbulence by focusing on multiscale structures of vortices. It conjectures that energy produced by vortices with large size λ is transferred hierarchically to the small-size vortices, which actually dissipate this energy due to viscosity. A related result is the *Kolmogorov law*

$$E(k,r) \propto r^{\frac{2}{3}} k^{-\frac{5}{3}} \tag{15}$$

where $k = 2\pi/\lambda$ is the wavenumber, r is the energy dissipation rate, and $E(k,r)$ is the velocity wavenumber spectrum.

In the area of speech dynamics, motivated by Mandelbrot's conjecture that fractals can model multiscale structures in turbulence, Maragos [26] outlined several plausible mechanisms for speech turbulence and used the *short-time multiscale fractal dimension* of speech sounds as a feature to approximately quantify the degree of turbulence in them. This nonlinear feature has been found useful for speech sound classification segmentation and recognition [26], [28]. For example, related to the 5/3-law (15) is the fact that the variance between particle velocities at two spatial locations X and $X + \Delta X$ varies $\propto |\Delta X|^{2/3}$. By linking this to similar scaling laws in FBMs, it was concluded in [26] that speech turbulence leads to fractal dimension of 5/3, which was often approximately observed during experiments with fricatives.

Finally, there have been several previous approaches for analyzing speech turbulence from the viewpoint of nonlinear dynamics by using concepts from chaotic systems, either at the vocal fold level [4], [46] or at the speech signal level [31]. Further, for voiced fricatives, a source excitation model that is based on modulating frication noise by vocal fold vibration was studied in [15], and a related pitch-scaled harmonic filter was proposed in [16] to decompose the corresponding speech signal into a voiced and a turbulent component.

III. Model for Random Resonances

Motivated by the experimental evidence in [27] for the AM–FM structure of fricative sounds with random instantaneous modulating signals as well as by the theoretical background relating self-similar stochastic processes with turbulence, we extend the conjecture of Maragos *et al.* [27] that turbulent speech sounds accept an AM–FM model with a random noise-like signal representing the frequency fluctuations by proposing the following model for resonances of turbulent sounds: We assume the general phase modulation model

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

where ω_c is the center frequency of the resonance. For the subsequent analysis, we ignore the instantaneous amplitude and assume that it is constant.[1] Further, we assume that we know the statistics of the nonlinear randomly varying instantaneous phase modulation process $P(t)$. Specifically, we model $P(t)$ as an H-sssi SαS process. This process $P_H(t)$ has a similarity exponent $H > 0$, stationary increments and a symmetric α-stable distribution for each t. Thus, we will work with the *random phase modulated process*

$$X(t) = A\cos(\omega_c t + \lambda P_H(t) + \phi_0) \tag{16}$$

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 (averaged) power spectrum of $P_H(t)$ is proportional to $1/|\omega|^\gamma$, where $\gamma = 2H+1$. Thus, we are modeling the phase modulation signal $P(t)$ as a 1/f 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 $2H - 1$. Such a model was inspired by the special case where $P_H(t)$ is an FBM modeling the fractal aspects of speech turbulence [26], [28] and by the possibility of synthesisizing speech by exciting CELP coders with 1/f noises of the FBM type or their increments [25].

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 α-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.

A. Phase-Modulated H-sssi $S\alpha S$ Process

We will address the problem of finding the properties of a phase modulated SαS H-sssi process. Popular models for 1/f signals like the FBM [24] (which is Gaussian) or the FSLM [43] (which is α-stable) will be addressed as special cases.

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

$$\Phi(\theta, t) = \Phi(t^H \theta, 1). \tag{17}$$

[1]Note that in general, amplitude modulations could also occur, and the ESA algorithm we use can easily track them as well. However, this general case goes beyond the scope of this paper, and we plan to study it in future work.

Proof: Self similarity of $X(t)$ means that $X(rt) \stackrel{d}{=} r^H X(t)$, which implies $X(t) \stackrel{d}{=} t^H X(1)$. Therefore, the characteristic function of the r.v. $X(t)$ is

$$\Phi_{X(t)} = \mathsf{E}\left[e^{j\theta X(t)}\right] = \mathsf{E}\left[e^{j(t^H\theta)X(1)}\right] \tag{18}$$

which yields the result. ∎

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

Theorem 1: Consider the random process $X(t) = A\cos(\omega_c t + \lambda P_H(t) + \phi_0)$, 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

$$\Phi_H(\theta, t) = \exp\left(-|s(t)\theta|^\alpha\right), \quad 0 < \alpha \le 2 \tag{19}$$

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

$$R_{xx}(\tau) = \frac{1}{2}\cos(\omega_c\tau)\exp\left(-|s(1)\lambda|^\alpha |\tau|^{\alpha H}\right). \tag{20}$$

Proof: We define the complex processes

$$W(t) = \exp\left[j\lambda P_H(t)\right], \quad Z(t) = W(t)\exp\left[j(\omega_c t + \phi_0)\right] \tag{21}$$

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[4] between Z and its conjugate Z^*. First, for the mean

$$\mathsf{E}\left[Z(t)\right] = e^{j\omega_c t}\mathsf{E}\left[W(t)\right]\mathsf{E}[e^{j\phi_0}] = 0. \tag{22}$$

Second, for the correlations

$$R_{zz^*}(t+\tau, t) = \mathsf{E}\left[w(t+\tau)w(t)\right]\mathsf{E}[e^{2j\phi_0}] = 0 \tag{23}$$

$$R_{zz}(t+\tau, t) = e^{j\omega_c\tau}\mathsf{E}\left[e^{j\lambda(P_H(t+\tau) - P_H(t))}\right]. \tag{24}$$

Since the increments of $P_H(t)$ are stationary, the autocorrelation of $Z(t)$ and $W(t)$ can be written in the form

$$R_{ww}(\tau) = \mathsf{E}\left[e^{j\lambda P_H(\tau)}\right] = \mathsf{E}\left[W(\tau)\right] \tag{25}$$

$$R_{zz}(\tau) = e^{j\omega_c\tau}R_{ww}(\tau) = e^{j\omega_c\tau}\mathsf{E}\left[W(\tau)\right]. \tag{26}$$

Hence, $X(t)$ is a zero-mean WSS process.

Now, note that

$$\mathsf{E}\left[W(t)\right] = \mathsf{E}\left[e^{j\lambda P_H(t)}\right] = \Phi_H(\lambda, t) \tag{27}$$

where $\Phi_H(\lambda, t)$ is the time-dependent characteristic function (19) of the r.v. $P_H(t)$. Then, from the above results, it follows that the autocorrelation of $X(t)$ is

$$R_{xx}(\tau) = \frac{1}{2}\mathrm{Re}\left[R_{zz}(\tau)\right] = \frac{1}{2}\cos(\omega_c\tau)\Phi_H(\lambda, \tau). \tag{28}$$

Further, due to the self-similarity of $P_H(t)$, we use lemma (17) to obtain

$$\Phi_H(\lambda, t) = \exp\left(-|s(1)\lambda|^\alpha |t|^{\alpha H}\right). \tag{29}$$

Combining the last two equations yields the desired formula (20) for the autocorrelation. ∎

Having found the autocorrelation of $X(t)$, its 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 of the FBM type.

B. 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

$$\mathsf{Var}\left[P_H(t)\right] = \sigma_H^2|t|^{2H}, \quad \sigma_H^2 = \mathsf{E}\left[P_H(1)^2\right] \tag{30}$$

and characteristic function

$$\Phi_H(\theta, t) = \exp\left(-\frac{1}{2}\sigma_H^2\theta^2|t|^{2H}\right). \tag{31}$$

Therefore, if in the previous theorem we set $\alpha = 2$ and $s(1) = \sigma_H/\sqrt{2}$, we find the autocorrelation function of $X(t)$ to be

$$R_{xx}(\tau) = \frac{1}{2}\cos(\omega_c\tau)\exp\left(-\frac{1}{2}\sigma_H^2\lambda^2|\tau|^{2H}\right). \tag{32}$$

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:

$$S_{0.5}(\omega) = \frac{\sigma_H^2\lambda^2}{\sqrt{2\pi}}\left[\frac{1}{\lambda^4\sigma_H^4 + 4(\omega-\omega_c)^2} + \frac{1}{\lambda^4\sigma_H^4 + 4(\omega+\omega_c)^2}\right] \tag{33}$$

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 FBM), we obtain

$$S_1(\omega) = \frac{1}{4\sqrt{\lambda^2\sigma_H^2}}\left[\exp\left(-\frac{(\omega-\omega_c)^2}{2\lambda^2\sigma_H^2}\right) + \exp\left(-\frac{(\omega+\omega_c)^2}{2\lambda^2\sigma_H^2}\right)\right] \tag{34}$$

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.

[2]Some elementary assumptions and implications of the theorem are inspired from a random frequency modulation model analyzed by Papoulis [35], 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.

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

[4]The cross-correlation of two complex processes $X(t)$ and $Y(t)$ is defined as $R_{xy}(t_1, t_2) = \mathsf{E}[x(t_1)y^*(t_2)]$.

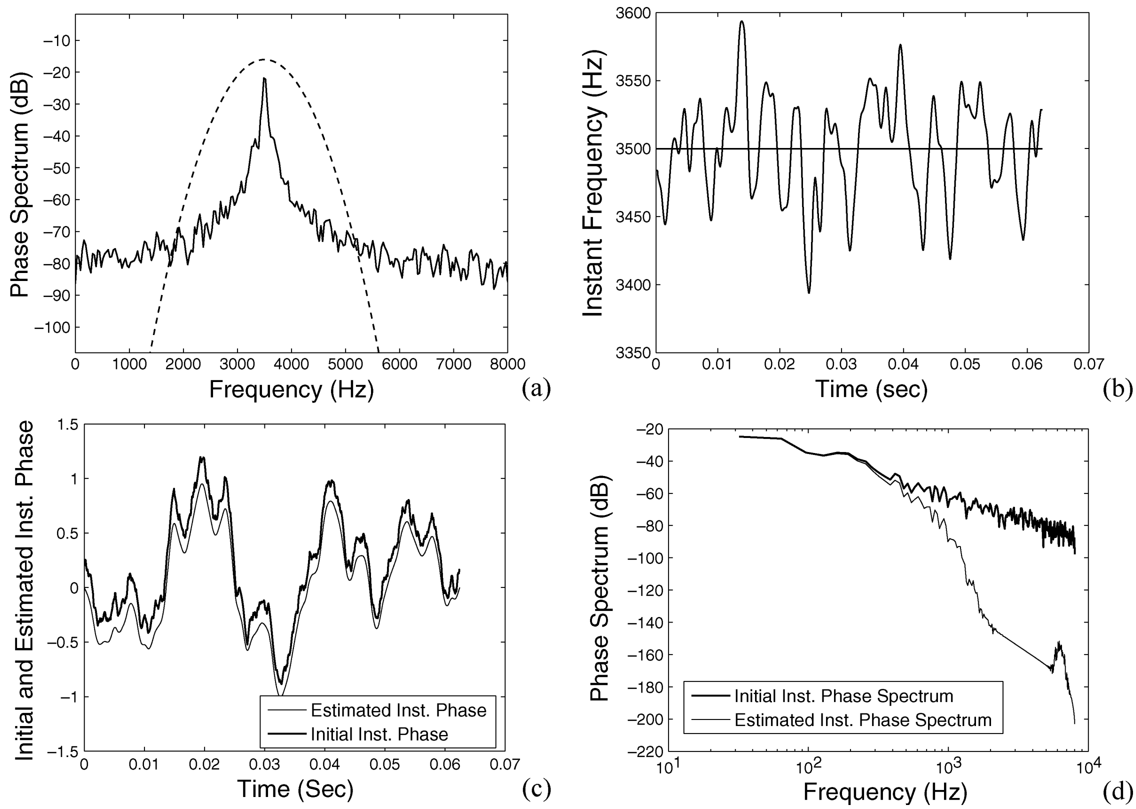

Fig. 1. (a) Artificial resonance with 1/f phase $P(t)(\alpha = 2, \gamma = 2.8)$. (b) Instant frequency. (c) Instant phase $P(t)$ and $\hat{P}(t)$ ($\hat{P}(t)$ is the low-pass signal). (d) Power spectrum of $P(t)$ and $\hat{P}(t)$.

C. Relation With α-Stable Distributions

One important observation is that the characteristic function (12) of an α-stable process is of the same form as the autocorrelation function (20) we derived before for the phase-modulated process by replacing the α exponent of the former with αH. The multiplication with the cosine function in (20) merely causes a shift centering the resonance at ω_c and $-\omega_c$. 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 [36].

It is well known that α-Stable distributions are invariant to convolution and therefore constitute an attractor under consecutive convolutions of other density functions (this is the generalized central limit theorem). Drawing an analogy with our results, the power spectrum of the modulated self-similar processes we proposed is also invariant and attracting with respect to convolution of power spectra. However, the latter is equivalent with multiplication of the autocorrelation functions. This important result implies that the *modulated self-similar processes we proposed constitute attracting processes with respect to multiplication*. This seems to be a mathematical link between the multiplicative models for turbulence developed by the Russian school [12] and our proposed modulation model. In our future work, we plan to explicitly formulate and investigate these relations.

IV. Parameter Estimation

After these theoretical derivations, the practical problems of testing the validity of the proposed model as well as fitting it to real data arise. We propose the following algorithm for parameter estimation, which consists of five steps.

1) Isolate the resonance by bandpass filtering the signal.
2) Use the ESA demodulation algorithm to estimate the AM and FM signals, $A(t)$ and $F(t)$.
3) Estimate the instant phase modulation signal $\hat{P}(t)$ 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[5] that best models the estimated phase modulation signal $\hat{P(t)}$ as a 1/f self-similar signal. See Fig. 1 for results of the algorithm on a test signal.

The last two steps are unquestionably the most complicated. 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 regression [19], [33] on the sample characteristic function to estimate the α parameter for the instant

[5]In this paper, γ refers to the spectral exponent appearing in the $1/f^{\gamma}$ (averaged) PSD and should not be confused with the dispersion of the stable distribution.

frequency process. (However, alternative methods can also be used.)

The problem of estimating the γ exponent has been approached from a number of different angles utilizing both time-domain, frequency-domain, and wavelet-domain approaches. Specifically, a variety of methods have been proposed, and they include, among others, least squares estimation of the slope of log-axes plots of sample periodograms, methods based on wavelets, aggregated variance, and maximum likelihood (ML) schemes. For detailed reviews, see [6], [34], [41], and the references therein. The ML estimators [21] are considered the most sophisticated because they are able to cope with measurement noise. However, these methods are based on parametric models, such as FBM. Further, the fact that the signal $\hat{P}(t)$ (from which γ will be estimated) is a *low-pass filtered* self-similar process creates difficulties for any estimator based on an exact model. Specifically, the Gabor filtering that is used to isolate the speech resonance induces a bias in the power spectrum of the instant phase for large frequencies. The wavelet EM approximation algorithm proposed in [50], which is an interesting approach not based on FBM, was recently shown in [34] to provide satisfactory estimation only when $0 < \gamma < 1$. After extensively testing all the above methods, the method that seemed to perform best in our specific case was the GPH local spectral estimator. (This seems to agree with Taqqu's empirical study [41] comparing a number of methods and finding the local spectral estimators to outperform the other methods for large γ.)

The GPH method was proposed in the early work of Geweke and Porter-Hudak [10] and is based on the fact that as $\omega \downarrow 0$, the log power spectrum becomes approximately

$$\log S(\omega) \simeq \log C - \gamma \log(\omega) \tag{35}$$

where C is a constant. The main advantage that makes this method suitable for speech resonances is that the GPH local spectral estimator tries to estimate the γ exponent in the spectral domain as the frequency approaches zero, thus imposing no restrictions on the behavior of the spectral density for large frequencies. The *tapered periodogram* is used to estimate the power spectrum:

$$S(\omega) = \left| \frac{1}{\sqrt{2\pi \sum_{t=1}^{n} |h_t^2|}} \sum_{t=1}^{n} h_t P(t) e^{it\omega} \right|^2. \tag{36}$$

The taper used is one proposed by Hurvich and Chen [14] and created specifically for this estimation process:

$$h_t = \left(1 - \exp\left(\frac{2\pi i t}{n}\right)\right)^p \tag{37}$$

where p is referred to as the order of the kernel. After estimating the power spectrum, the γ exponent is found according to (35) by linear regression. The problem is, however, that (35) is only valid in a neighborhood of the zero frequency, and thus, the regression line should be computed by using only a subset of the log-periodogram up to a given bandwidth. The choice of the bandwidth is known to be a difficult problem involving a tradeoff: If the bandwidth is small, the estimator is unbiased but has a large variance. On the other hand, a large bandwidth will yield small variance and large bias. Several techniques have been used to find the optimal bandwidth [6]. A simple approach is to draw a dynamic GPH plot (see Fig. 2 and next section for more details), that is, a plot of how the estimate of γ changes for different bandwidth values. To see how well this algorithm behaves, we performed simulations with artificial signals.

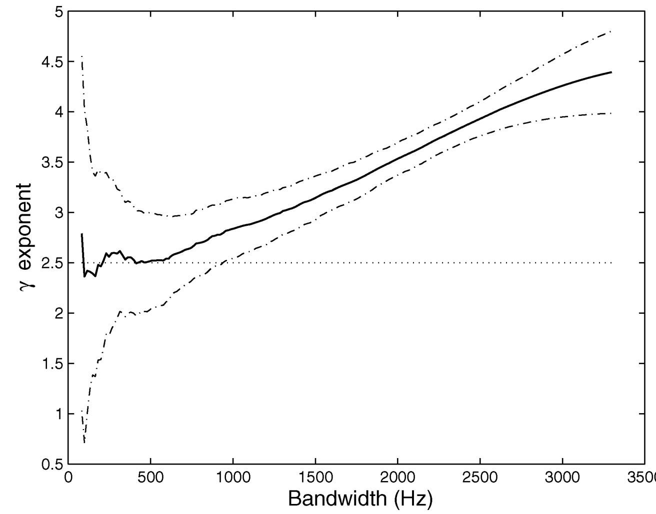

Fig. 2. GPH dynamic estimation of γ. Sample mean (solid line) and standard deviation averaging over 50 artificial 1/f modulated resonances. (True $\gamma = 2.5$).

A. Testing the Estimation Algorithm

Fig. 1 demonstrates the application of the estimation algorithm to an artificial phase-modulated resonance signal. For this case, the 1/f phase modulation signal was an FBM realization created by filtering white noise; however, any known method for 1/f noise synthesis can be used. Fig. 1(a) illustrates the power spectrum of the artificial modulated signal and the Gabor filter used to isolate the resonance. Fig. 1(b) is the instant phase that was extracted using the ESA algorithm. Fig. 1(c) is the original and reconstructed phase modulation signals, whereas Fig. 1(d) illustrates the power spectra of the original and reconstructed phase modulation signals. As seen in Fig. 1(d), the reconstructed phase modulation $\hat{P}(t)$ is a lowpass version of the original $P(t)$. This is not surprising because of the Gabor filtering and the inherent limit to the bandwidth that can be carried by a phase modulated signal, and this is the reason that local spectral estimators are the most efficient methods.

Fig. 2 illustrates the result of averaging dynamic GPH plots for artificial resonances. We plot the mean and the sample standard deviations. The estimation variance is decreasing for larger bandwidths but after some point, it starts to increase again (for very large bandwidths). For a true value of $\gamma = 2.5$, the estimation is most effective if we choose a bandwidth of about 500–1000 Hz, and it involves the aforementioned bias-variance tradeoff. The bandwidth is related with the Gabor filter used to isolate the resonance. Experimental evidence suggests that values up to one half of the rms bandwidth of the Gabor filter yield better results.

The concluding result from the simulations with artificial signals is that the algorithm is successful in recovering the instant

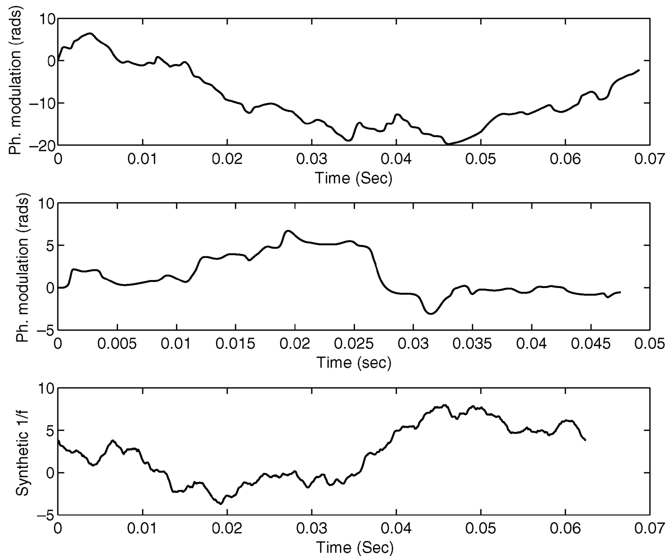

Fig. 3. Two actual speech phase modulation signals visually compared with an artificial self-similar process. (Top) Phase modulation signal extracted from a ZH phoneme (Male speaker, Western dialect). (Middle) Phase modulation signal extracted from a V phoneme (Male speaker, Northern Dialect) (Bottom) Artificial 1/f signal.

phase signal and estimating the γ exponent. However, the variance of the estimation is considerable, and more work remains to be done in this area.

V. Application to Speech Sounds

In this section, we present strong experimental evidence that turbulent speech signals have resonances that can be effectively modeled as phase modulated 1/f signals. We emphasize that this modeling is *phenomenological*, since it does not attempt to explain the acoustics of speech turbulence but only approximate the stochastic spectral structure of its resulting speech resonance signals. All the speech signals we use in this paper are segmented phonemes from the TIMIT database. Gender and dialect of the speaker of each phoneme used are given in the corresponding figure captions.

Fig. 3 offers a visual comparison between the phase modulation signals (extracted using the proposed algorithm) from two real speech signals and an artificial self-similar stochastic process. Plain observation implies that the statistical properties of the real signals and the proposed model are very similar. Notice that the artificial self-similar signal has more high-frequency energy. As mentioned in Section IV, this is due to the fact that the reconstructed phase modulation signal $\hat{P}(t)$ is a lowpass version of the original $P(t)$. This is due to the Gabor filter used to isolate the resonance.

Exact statistical tests to verify self-similarity follow in subsequent sections.

Fig. 4 illustrates the application of the described algorithm to two fricative phonemes. Subfigures (b1 and b2) illustrate the spectrum of the sound and the Gabor filters used to isolate the resonances. Subfigures (c1 and c2) illustrate the instant frequencies as estimated by the ESA algorithm, whereas subfigures (d1 and d2) are the instant phase modulation signals that are modeled as self-similar processes.

A. Testing the Validity and Experimental Results

After extracting the instant phase modulation signal, we want to verify the validity of the proposed model. Specifically, we want to establish that the instant phase modulation signals of fricative vowels can be effectively modeled as self-similar stochastic processes.

To show this, we use a number of popular statistical tests following an approach similar to the one found in [20] to rigorously establish the self-similar nature of Ethernet traffic. Before testing for self-similarity, we performed a number of statistical tests to examine more fundamental properties for the estimated instant phase signal $\hat{P(t)}$. Specifically, we tested our experimental signals for Gaussianity, linearity, and self-similarity.

A random signal $y(n)$ will be called *linear* if it can be represented by $y(n) = \sum_k u(n-k)h(k)$, where $u(n)$ are identically distributed random variables. If $u(n)$ are Gaussian, then $y(n)$ is linear Gaussian. We used a statistical test developed by Hinch [13] that jointly tests for Gaussianity and linearity. The test, which is documented and implemented in the Matlab Higher-Order Spectral Analysis Toolbox [40], is based on the fact that if a process is Gaussian, then its bispectrum is zero. Additionally, if a process is linear and non-Gaussian, its bicoherence is a nonzero constant. For a comprehensive text on higher spectral methods, see [32]. We applied the test to several instant phase signals we extracted from speech signals. As a result of these tests, the Gaussian hypothesis was rejected, and then, the linearity assumption was also rejected (with high confidence).

The tests we used for self-similarity are based on time-, frequency-, and wavelet-domain techniques. More specifically, we have the following.

- The (averaged) power spectral density of self similar processes obeys a power law behavior near the origin. We therefore plot the estimated spectral density and see that it clearly fits a straight line (which indicates a power law since the axes are logarithmic) for up to a given cut-off frequency. See Fig. 5(a1), (a2), and (a3).
- *The variance of the wavelet coefficients.* Following [50], if $\psi_n^m(t)$ is an Rth-order regular wavelet basis (R is determined by γ), then the process constructed via the expansion $p(t) = \sum_m \sum_n x_n^m \psi_n^m(t)$ is "nearly 1/f" when the wavelet coefficients have variances $\mathsf{Var}(x_n^m) = \sigma^2 2^{-\gamma m}$. See Fig. 5(b1), (b2), and (b3).
- Finally, for time domain analysis, we used the popular rescaled adjusted scale plot (R/S or Pox Diagram) [23]. The R/S statistic is defined as $R(n)/S(n) = (\max(0, W_1, W_2, \cdots W_n) - \min(0, W_1, W_2, \cdots W_n))/S(n)$ (where $S^2(n)$ is the sample variance $\bar{X}(n)$) is the sample mean, and $W_k = X_1 + X_2 + \cdots + X_k - k\bar{X}(n), k = 1, 2 \cdots n$) For self-similar processes (as first observed by Hurst), $\mathsf{E}(R(n)/S(n)) \propto n^H$ as $n \to \infty$. To create an R/S plot, we take logarithmically spaced values of n and plot $log(R(n)/S(n))$ versus $log(n)$. A typical R/S plot has a transient zone and asymptotic slope. When the asymptotic slope is larger than 1/2 and smaller than 1, this is an indication of self-similarity with parameter H that can be estimated from this slope. In all our tests, the

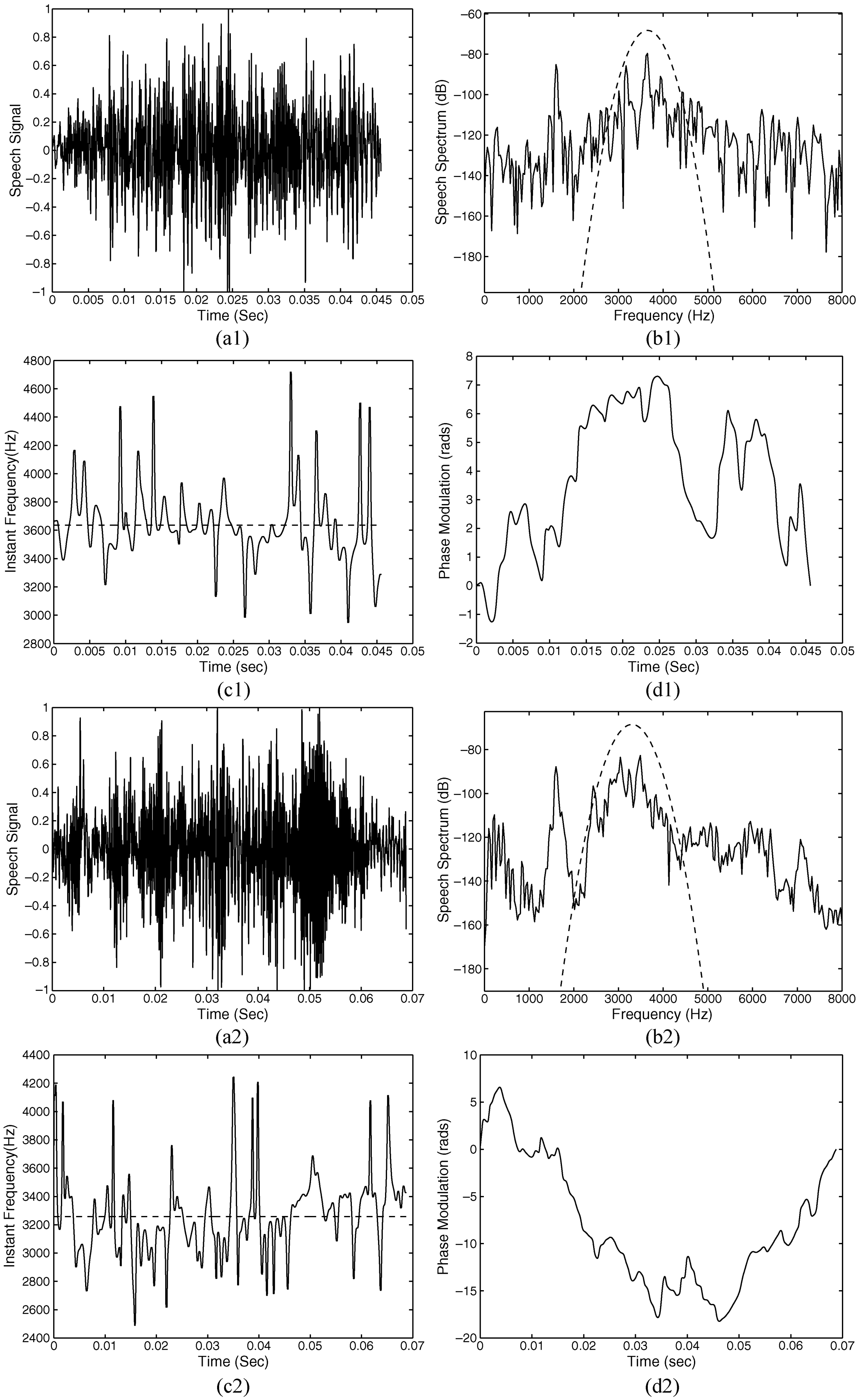

Fig. 4. Proposed algorithm using a phoneme S (Male speaker, South Midland dialect) and a phoneme ZH (Male speaker, Western dialect) a1-2) Speech signal $s(t)$. b1-2) Power spectrum of $s(t)$ and Gabor filter. c1-2) Instant frequency. d1-2) Phase modulation $\hat{P}(t)$.

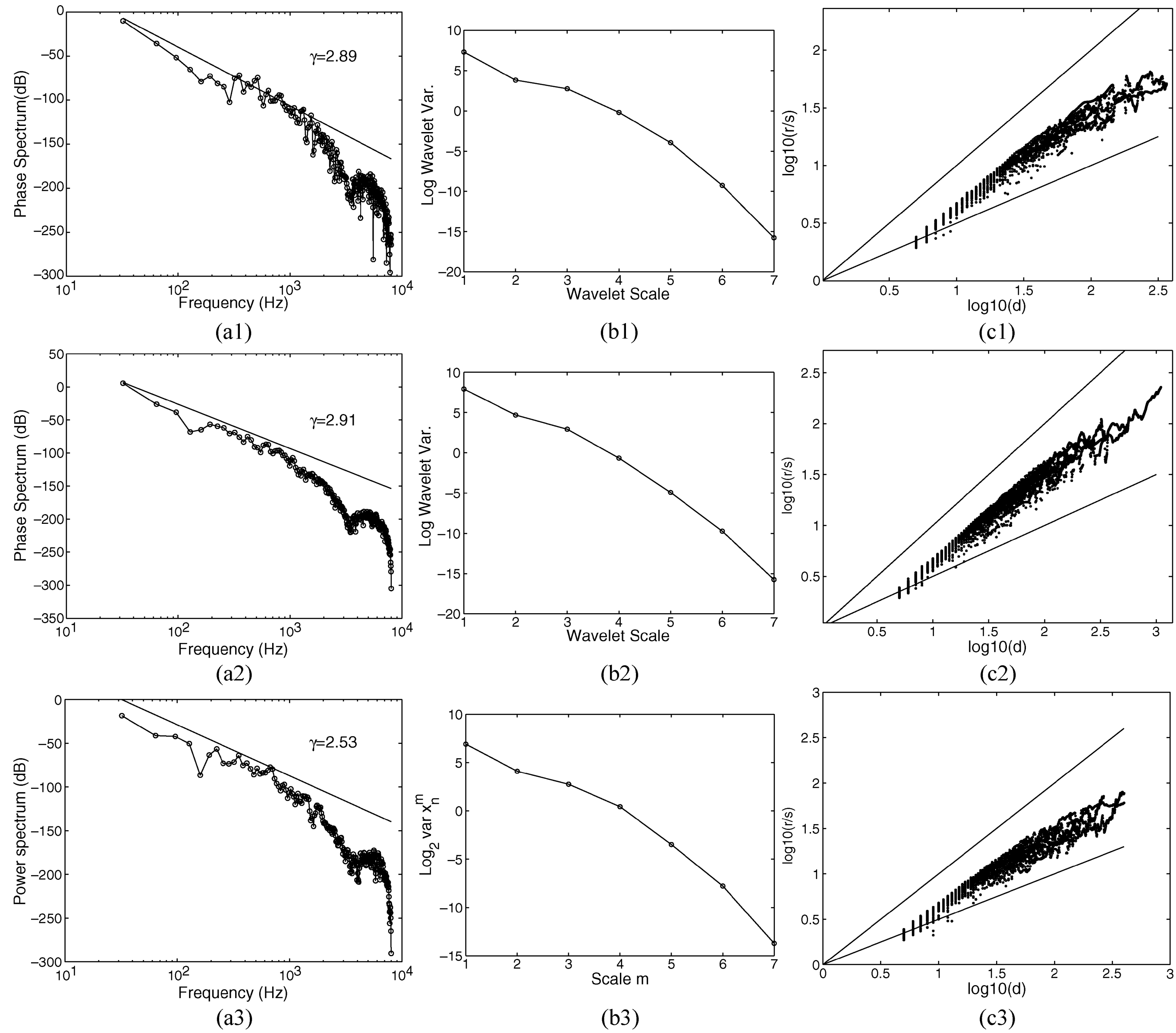

Fig. 5. Statistical tests for self-similarity. First row: Phoneme S (Male speaker, South Midland dialect). Second row: Phoneme ZH (Male speaker Western dialect). Third row: Phoneme F (Male speaker, Western dialect). Tests used: First column: logarithmic plot of (averaged) power spectral density and estimated slope. Second column: logarithm of Variance of the wavelet coefficients plotted as a function of scale. Third column: R/S or Pox diagram.

slope is clearly in the 1/2–1 range. See Fig. 5(c1), (c2), and (c3).

The tests we have performed are very similar to the tests used in [20] to verify the self-similarity of Ethernet traffic. We have not used the more statistically rigorous Whittle's approximate MLE approach because the speech signals have very few samples (less than 1000 in most cases since the sampling rate is 16 kHz), and more importantly, the bandpass filtering used to isolate the resonances limits the scales in which we can observe self-similarity (as already mentioned, we are essentially dealing with a *lowpass filtered self-similar signal*). Therefore, we cannot use the parametric methods that yield confidence intervals because our signals fit the parametric model only for a subset of scales. It is a very interesting direction for future work to rigorously analyze the effects of sampling and filtering and derive statistical tests and MLE estimators for signals that exhibit *self-similarity only for a range of scales*.

On the other hand, Ethernet data used in [20] had sequences with 360 000 observations. Testing and estimation problems are therefore much harder for speech data, but still, we believe that the self-similar character has been clearly demonstrated by the power spectra, wavelet variances, and R/S statistics, which approximate straight lines for broad ranges of scales.

We have performed similar tests on more than 100 phonemes from the TIMIT database, and in all these experiments, we found strong evidence that the phase modulation of speech resonances for fricative phonemes exhibits self-similarity.

The results of these experiments are summarized in Table I, where we give the measured γ and α exponents for a number of different phonemes. For each phoneme, 15 or more separate tokens were used (from the TIMIT database). After isolating a strong resonance using a broad Gabor filter (whose center frequency was manually determined), we estimate the γ and α exponents and tabulate their mean values and sample variances.

TABLE I
MEASURED EXPONENTS FOR DIFFERENT PHONEMES

Phoneme	$\overline{\gamma}$	Var(γ)	$\overline{\alpha}$	Var(α)	$F_c(Hz)$
F	2.59	0.06	1.92	0.0008	4400
S	2.67	0.10	1.92	0.0005	3850
V	2.85	0.05	1.94	0.0011	2450
ZH	2.78	0.10	1.94	0.0002	3100
A	3.02	0.15	1.95	0.0014	2200
IX	2.99	0.04	1.94	0.0012	2400
IY	3.01	0.09	1.93	0.0011	2200

†The rms bandwidth of the Gabor filters was 2000 Hz.

Our experiments indicate that their values are correlated with the nature of each phoneme. Namely, unvoiced fricatives usually yielded the smaller γ exponents.

Intuitively, γ and α are measuring the variability of a given resonance in time. Therefore, smaller γ exponents for unvoiced fricatives are not surprising since the resonances of unvoiced fricatives (like F or S) seem to be less smooth, and smaller γ exponents indicate more rapidly varying realizations. The α exponent measures the impulsiveness of the instant phase of the resonance. We observed that unvoiced fricatives have (slightly) smaller α exponents, which indicate more impulsive behavior relative to voiced fricatives and vowels.

An important thing to note is that the proposed model is suitable for modeling *broad resonances* that are isolated in the spectrum. In our experiments, we enforced the finding of such broad resonances by using an rms Gabor filter bandwidth of $B = 2000$ Hz. Phonemes that exhibit this type of resonance are typically voiced and unvoiced fricatives. Experiments on such broad frequency bands containing formants of smooth vowels (like A) indicate that their phase modulation signals have exponentially decaying spectra. Further, if there are no well-separated resonances (as is the case in vowel formants), the broad filter may allow "parasitic" modulations from neighbor formants [27]. In general, the applicability of the proposed model to the case of vowels is still under investigation.[6]

VI. CONCLUSIONS

In this paper, we have proposed a phenomenological random phase modulation model for resonances of turbulent sounds, where the instant phase modulation signal is an α-stable self-similar process. Our contribution includes a theoretical analysis of some statistical properties of the model, the development of an algorithm to estimate its parameters, and some experimental work on testing and fitting the model to real speech data.

The work herein is a continuation of previous work on modeling speech resonances with AM–FM signals and on modeling turbulence in fricative and other speech sounds with random fractal signals. Our ongoing work in this area includes research on 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. Another interesting extension is to study models that exhibit both amplitude and frequency modulation. 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 $1/f$ fluctuations in periodic phenomena seem to be ubiquitous in nature.

[6]When trying to fit the model to these sounds, one obtains very large γ exponents (even larger than 3), which vary considerably.

REFERENCES

[1] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch, "Multiscale nature of network traffic," *IEEE Signal Process. Mag.*, vol. 19, no. 5, pp. 28–46, May 2002.

[2] J. M. Bardet, "Un test d'autosimilarite pour les processus gaussien a accroissements stationnaires," *C.R. Acad Sci. Paris, Ser. I Math*, vol. 6, pp. 521–526, 1999.

[3] A. Barney, C. H. Shadle, and P. O. A. L. Davies, "Fluid flow in a dynamical mechanical model of the vocal folds and tract: part 1 & 2," *J. Acoust. Soc. Amer.*, vol. 105, no. 1, pp. 444–466, Jan. 1999.

[4] A. Behrman and R. J. Baken, "Correlation dimension of electroglottographic data from healthy and pathologic subjects," *J. Acoust. Soc. Amer.*, vol. 102, no. 4, pp. 2371–2379, Oct. 1997.

[5] A. C. Bovik, P. Maragos, and T. F. Quatieri, "AM-FM energy detection and separation in noise using multiband energy operators," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3245–3265, Dec. 1993.

[6] O. Cappe, E. Moulines, J. Pesquet, A. Petropulu, and X. Yang, "Long-range dependence and heavy-trail modeling for teletraffic data," *IEEE Signal Process. Mag.*, vol. 19, no. 5, pp. 14–27, May 2002.

[7] A. Coron, P. Flandrin, and M. Gache, "Modulated fractional gaussian noise: a process with non-Gaussian wavelet details," in *Proc. IEEE-SP Int. Symp. Time Freq. Time Scale Anal.*, Oct. 1998.

[8] A. G. Dimakis and P. Maragos, "Modeling resonances with phase modulated self-similar processes," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, May 2004.

[9] P. Embrechts and M. Maejima, *Selfsimilar Processes*. Princeton, NJ: Princeton Univ. Press, 2002.

[10] J. Geweke and S. Porter-Hudak, "The estimation and application of long memory time series models," *J. Time Series Anal.*, vol. 4, pp. 221–238, 1983.

[11] P. Flandrin, "On the spectrum of fractional Brownian motions," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 197–199, Jan. 1989.

[12] U. Frisch, *Turbulence: The Legacy of A.N. Kolmogorov*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

[13] M. J. Hinich, "Testing for Gaussianity and linearity of a stationary time series," *J. Time Series Anal.*, vol. 3, pp. 169–176, 1982.

[14] C. M. Hurvich and W. W. Chen, "An effective taper for potentially overdifferenced long-memory time series," *J. Time Series Anal.*, vol. 21, no. 2, pp. 155–180, 2000.

[15] P. J. Jackson and C. H. Shadle, "Frication noise modulated by voicing, as revealed by pitch-scaled decomposition," *J. Acoust. Soc. Amer.*, vol. 108, no. 4, pp. 1421–1434, Oct. 2000.

[16] ——, "Pitch-scaled estimation of simultaneous voiced and turbulence-noise components in speech," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 7, pp. 713–716, Oct. 2001.

[17] J. F. Kaiser, "Some observations on vocal tract operation from a fluid flow point of view," in *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control*, I. R. Titze and R. C. Scherer, Eds. Denver, CO, 1983.

[18] S. M. Kogon and D. G. Manolakis, "Signal modeling with self-similar α-stable processes: the fractional levy stable motion model," *IEEE Trans. Signal Process.*, vol. 44, no. 4, pp. 1006–1010, Apr. 1996.

[19] I. A. Koutrouvelis, "Regression type estimation of the parameters of stable laws," *J. Amer. Stat. Assoc.*, vol. 75, pp. 918–928, Dec. 1980.

[20] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic," *IEEE/ACM Trans. Networking*, vol. 2, no. 1, pp. 1–15, Feb. 1994.

[21] T. Lundahl, W. J. Ohley, S. M. Kay, and R. Siffert, "Fractional browninan motion: A maximum likelihood estimator and its application to image texture," *IEEE Trans. Med. Imag.*, vol. MI-5, pp. 152–161, Sep. 1986.

[22] B. B. Mandelbrot, *The Fractal Geometry of Nature*. New York: Freeman, 1982.

[23] ——, "Long-run linearity, locally Gaussian processes, H-spectra and infinite variances," *Int. Econom. Rev.*, vol. 10, pp. 82–113, 1969.

[24] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motion, fractional noises and applications," *SIAM Rev.*, vol. 10, no. 4, pp. 422–437, 1968.

[25] P. Maragos and K. L. Young, "Fractal excitation signals for CELP speech coders," in *Proc. ICASSP*, Albuquerque, NM, Apr. 1990, pp. 669–672.

[26] P. Maragos, "Fractal aspects of speech signals: dimension and interpolation," in *Proc. ICASSP*, Toronto, ON, Canada, May 1991, pp. 417–420.

[27] P. Maragos, J. F. Kaiser, and T. F. Quatier, "Energy separation in signal modulations with application to speech analysis," *IEEE Trans. Signal Process.*, vol. 41, no. 10, pp. 3024–3051, Oct. 1993.

[28] P. Maragos and A. Potamianos, "Fractal dimensions of speech sounds: computation and application to automatic speech recognition," *J. Acoust. Soc. Amer.*, vol. 105, no. 3, pp. 1925–1932, Mar. 1999.

[29] P. Maragos, A. Dimakis, and I. Kokkinos, "Some advances in nonlinear speech modeling using modulations, fractals, and chaos," in *Proc. Int. Conf. Digital Signal Process.*, Santorini, Greece, Jul. 2002.

[30] R. S. McGowan, "An aeroacoustics approach to phonation," *J. Acoust. Soc. Amer.*, vol. 83, no. 2, pp. 696–704, 1988.

[31] S. S. Naryanan and A. A. Alwan, "A nonlinear dynamical systems analysis of fricative consonants," *J. Acoust. Soc. Amer.*, vol. 97, no. 4, pp. 2511–2524, Apr. 1995.

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

[33] C. L. Nikias and M. Shao, *Signal Processing With Alpha-Stable Distributions*. New York: Wiley, 1995.

[34] B. Ninness, "Estimation of 1/f noise," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 32–46, Jan. 1998.

[35] A. Papoulis, "Random modulation: a review," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, no. 1, pp. 96–105, Feb. 1983.

[36] A. P. Petropulu, J. C. Pesquet, X. Yang, and J. Yin, "Power-law shot noise and its relationship to long-memory a-stable processes," *IEEE Trans. Signal Process.*, vol. 48, no. 7, pp. 1883–1892, Jul. 2000.

[37] A. Potamianos and P. Maragos, "Speech formant frequency and bandwidth tracking using multiband energy demodulation," *J. Acoust. Soc. Amer.*, vol. 99, no. 6, pp. 3795–3806, Jun. 1996.

[38] D. Saupe, "Algorithms for random fractals," in *The Science of Fractal Images*, H.-O. Peitgen and D. Saupe, Eds. New York: Springer-Verlag, 1988.

[39] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models With Infinite Variance*. New York: Chapman and Hall, 1994.

[40] A. Swami, J. M. Mendel, and C. L. Nikias, Higher-Order Spectral Analysis Matlab Toolbox Documentation, Natick, MA, 1998.

[41] M. S. Taqqu, V. Teverovsky, and W. Willinger, "Estimators for long-range dependence: an empirical study," *Fractals*, vol. 3, pp. 785–798, 1995.

[42] M. S. Taqqu, "Weak convergence to fractional Brownian motion and to the Rosenblatt process," *Zeitschr. fur Watrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 31, pp. 287–302, 1975.

[43] M. S. Taqqu and R. Wolpert, "Infinite variance self-similar processes subordinate to a poison measure," *Zeitschr. fur Watrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 62, pp. 53–72, 1983.

[44] H. M. Teager and S. M. Teager, "Evidence for nonlinear sound production mechanisms in the vocal tract," in *Speech Production Speech Modeling*, W. J. Hardcastle and A. Marchal, Eds., Jul. 1989, vol. 55, NATO Adv. Study Inst. Series D.

[45] T. J. Thomas, "A finite element model of fluid flow in the vocal tract," *Comput. Speech Language*, vol. 1, pp. 131–151, 1986.

[46] I. R. Titze, R. J. Baken, and H. Herzel, "Evidence of chaos in vocal fold vibration," in *Vocal Fold Physiology: Frontiers in Basic Science*, I. R. Titze, Ed. New York: Wiley, 1993, pp. 143–188.

[47] D. J. Tritton, *Physical Fluid Dynamics*, Second ed. New York: Oxford Univ. Press, 1988.

[48] B. van der Pol, "Frequency modulation," in *Proc. IRE*, vol. 18, Jul. 1930, pp. 1194–1205.

[49] R. F. Voss, "Fractals in nature: from characterization to simulation," in *The Science of Fractal Images*, H.-O. Peitgen and D. Saupe, Eds. New York: Springer-Verlag, 1988.

[50] G. Wornell, *Signal Processing With Fractals: A Wavelet-Based Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

Alexandros G. Dimakis (S'01) was born in Athens, Greece. He received the Diploma degree in electrical and computer engineering from the National Technical University of Athens in 2003 with highest honors. He is currently working toward the Ph.D. degree with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley.

His current research interests include coding theory, distributed signal processing and communications with applications to wireless networks and distributed systems.

Mr. Dimakis received the U.C. Berkeley Departmental Fellowship for the years 2003 to 2005.

Petros Maragos (S'81–M'85–SM'91–F'95) received the Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1980 and the M.Sc.E.E. and Ph.D. degrees from the Georgia Institute of Technology, Atlanta, in 1982 and 1985.

In 1985, he joined the faculty of the Division of Applied Sciences, Harvard University, Cambridge, where he worked for eight years as professor of electrical engineering, affiliated with the interdisciplinary Harvard Robotics Lab. He has also been a consultant to several industry research groups, including Xerox's research on image analysis. In 1993, he joined the faculty of the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta. During parts of 1996 to 1998, he was on academic leave, working as a senior researcher at the Institute for Language and Speech Processing, Athens. In 1998, he joined the faculty of the National Technical University of Athens, where he is currently a professor of electrical and computer engineering. His current research and teaching interests include the general areas of signal processing, systems, communications, pattern recognition, and their applications to image processing and computer vision and computer speech processing and recognition.

Dr. Maragos' research has received several awards, including a 1987 U.S. National Science Foundation Presidential Young Investigator Award; the 1988 IEEE Signal Processing Society's Paper Award for the paper "Morphological Filters"; the 1994 IEEE Signal Processing Society's Senior Award, and the 1995 IEEE Baker Award for the paper "Energy Separation in Signal Modulations with Application to Speech Analysis"; and the 1996 Pattern Recognition Society's Honorable Mention Award for the paper "Min-Max Classifiers." In 1995, he was elected Fellow of IEEE for his contributions to the theory and applications of nonlinear signal processing systems. He has served as associate editor for the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, editorial board member for the *Journal of Visual Communications and Image Representation*, and guest editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING; general chairman of the 1992 SPIE Conference on Visual Communications and Image Processing; co-chairman of the 1996 International Symposium on Mathematical Morphology; co-chairman of the 2001 International Workshop on VLBR Video Coding; and a member of two IEEE DSP committees.