International Journal of Bifurcation and Chaos, Vol. 2, No. 3 (1992) 715–719

UNIVERSAL MULTIFRACTAL CHARACTERIZATION AND SIMULATION OF SPEECH

CHRIS LARNDER, NICOLAS DESAULNIERS-SOUCY, SHAUN LOVEJOY, DANIEL SCHERTZER,* CLAUDE BRAUN,† D. LAVALLÉE‡

Department of Physics, McGill University, 3600 University St., Montreal, Que. H3A 2T8, Canada

Received October 12, 1991; Revised March 18, 1992

In the 1970's it was found that, for low frequencies (< 10 Hz), speech is scaling: it has no characteristic time scale. Now such scale invariance is associated with multiscaling statistics, and multifractal structures. Just as Gaussian noises frequently arise because they are generically produced by sums of many independent noise processes, scaling noises have an analogous universal behavior arising from nonlinear mixing of processes. We show that low frequency speech is consistent with these ideas, and use the measured parameters to produce stochastic speech simulations which are strikingly similar to real speech.

Although statistical methods such as spectrograms have frequently been applied to studying the properties of speech at audible frequencies, the study of low frequency components has been badly neglected. An important exception was Voss & Clarke [1975, 1978] (see also Pickover & Khorasani [1986]), who found that over the range ≈ 10 Hz to at least 10^{-3} Hz, the spectrum had the scaling form $f^{-\beta}$ with f the frequency, and $\beta < 1$. Scaling ideas now center on multifractals which involve an entire exponent function (rather than the single value β). Furthermore, attractive (stable) universality classes exist for multifractal processes [Schertzer & Lovejoy, 1987, 1989, 1991]. For the same reason that — irrespective of the detailed generating mechanism — Gaussians occur in a wide variety of noise processes, in scaling noises, special types of (universal) multifractals naturally occur. When present, Gaussian distributions greatly simplify statistical analysis since they require only two basic parameters (the mean and variance) to specify the entire probability law. Universal multifractals play the same role in scaling processes; with three parameters the entire (infinite hierarchy) of scaling exponents is specified. Indeed, universal multifractals have been empirically found in many scaling systems; turbulent velocity and temperature fields [Schertzer *et al.*, 1991; Schmitt *et al.*, 1992], cloud radiances [Gabriel *et al.*, 1988; Lovejoy & Schertzer, 1990; Tessier *et al.*, 1992], landscape topography [Lavallée *et al.*, 1992], ocean surfaces [Lavallée *et al.*, 1991a] and hadron jets [Brax & Peschansky, 1991; Ratti *et al.*, 1991].

To test these ideas on speech, recordings 890 seconds long were made using a standard eight-bit digitizer at 5.5 kHz. Seventeen samples of speech were studied including dictation and song, conversation, and subjects of both sexes, different ages, and speaking the French and English languages. Because of the physiology and anatomy of the phonatory apparatus, speech essentially consists of high frequency wave packets varying at much lower frequencies corresponding to phonemes, syllables, sentences and paragraphs. The low frequencies are

*Laboratoire de Metéorologie Dynamique, Université Pierre et Marie Curie, Tour 15-25, Boite 99, 4 place Jussieu, Paris F-75252 Cedex 05, France.

†Dept. of Psychology, Université du Québec à Montréal, Case Postale 8888, Succursale A, Montreal, Que. H3K 3P8.

‡Now at the Earth-Science Research Group, University of California, 5276 Hollister Ave. Suite 260, Santa Barbara, CA 93111.

essentially voluntary, involving significant muscucontraction and energy expenditure, whereas the high frequencies are primarily caused by small variations of the oscillations of the vocal chords and effects of minute valve action. At low frequencies, the only significant physiologically constrained time scale is at about 0.1 Hz associated with respiration. The cerebellum plays a critical role in the motor control of the low frequencies whereas the motor cortex does so for higher frequencies. The two regimes are therefore very different: using spectral analysis, we found that the transition (spectral minimum) occurs at about 10 Hz which separates the high frequency spectral bump (falling off quickly above 1 kHz) from the low frequency power law form discussed below. We concentrated on the low frequency part because it was apparently the simplest to analyze and model while having a greater neuropsychological significance due to its dependence upon higher order nonautomatic processing.

To isolate the low frequencies, we digitally lowpass filtered and resampled the signal at 9.8 Hz [Fig. 1(a)]. Visually, the signal is plausibly multifractal: it has structures at all scales as well as some very large values (the extreme singularities); for comparison, Fig. 1(b) shows a universal multifractal simulation. If the low frequency signal (A) is the outcome of a scale invariant speech process, the statistical properties of the fluctuations (ΔA) over various time intervals (Δt) will be related to each other by power laws of the scale ratios. Consider the fluctuations (ΔA_λ) at scale ratio λ obtained by degrading the original signal to resolution $\Delta t = \tau/\lambda$ (i.e. its resolution is λ times smaller than that of the entire signal, duration τ). Its average value will depend on the resolution. It is therefore convenient to introduce a resolution independent multifractal speech process (φ_λ) with the property $\langle\varphi_\lambda\rangle$ = constant ("$\langle\ \rangle$" indicates statistical averaging):

$$\Delta A_\lambda \approx \varphi_\lambda \lambda^{-H}\,, \qquad (1)$$

where H quantifies the resolution dependence of the mean: $\langle|\Delta A_\lambda|\rangle = \lambda^{-H}$. The scaling λ^{-H} is equivalent to a "fractional integration" of order H, (a power law filter by f^{-H}). Introducing $K_A(q)$, $K(q)$ to characterize the scaling of the qth statistical moment of A, φ respectively, we obtain

$$\langle|\Delta A_\lambda|^q\rangle = \lambda^{K_A(q)} = \langle\varphi_\lambda^q\rangle\lambda^{-qH} = \lambda^{K(q)-qH}\,. \qquad (2)$$

Resolution independence of $\langle\varphi_\lambda\rangle$ implies that $K(1) = 0$, hence $K_A(1) = -H$. The spectral exponent β mentioned above is given by $\beta = 1 - K_A(2)$ (the energy spectrum is a second-order statistic). "Mono" scaling behaviour occurs if φ_λ has no resolution dependence, $\langle\varphi_\lambda^q\rangle$ = constant for all q (not just $q = 1$), $K(q) = 0$ and $K_A(q) = -qH$, $\beta = 1 + 2H$; otherwise φ_λ is multiscaling and $K(q)$ is convex. If we consider a multiscaling time series, then regions that exceed an amplitude threshold will have fractal dimensions that decrease with increasing threshold; intense parts of the signal will be distributed more sparsely in time than weaker

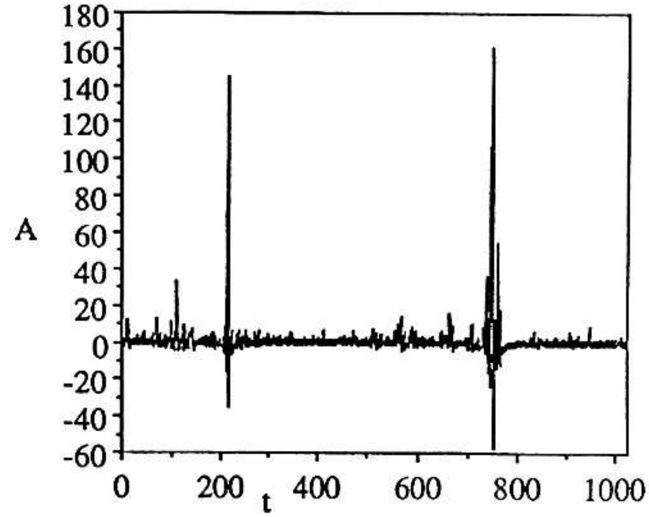

Fig. 1. (a) $A \approx 100$ seconds long section of "As You Like It" filtered and resampled at 9.8 Hz. Note the extreme intermittency and asymmetry.

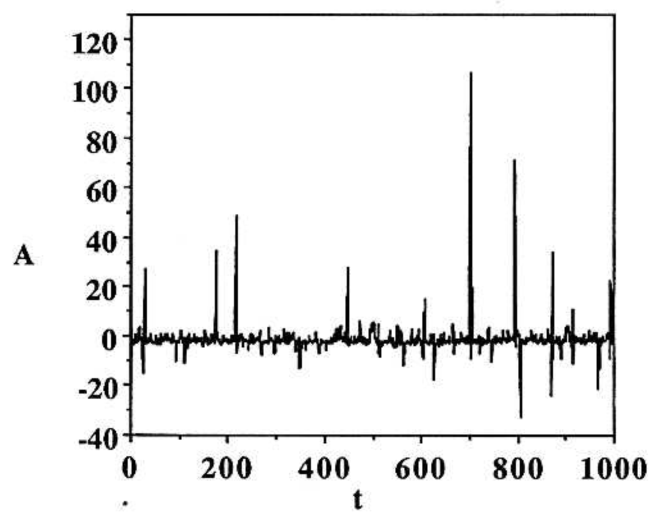

Fig. 1. (b) A stochastic universal multifractal simulation with $H = -0.37$, $C_1 = 0.1$, $\alpha = 2$, produced as described in Wilson *et al.* [1991]; the variability is plausibly of the same type as Fig. 1(a), changing the random seed in the Monte Carlo simulation produces a different mock speech record.

ones. In contrast, a monoscaling signal will have dimensions which are independent of the threshold. Figure 2 shows that the scaling (Eq. 2) is accurately followed over the entire range of available scales (with a slight bump possibly associated with the respiratory frequency at $\lambda \approx 200$), and Fig. 3 shows the $K(q)$ obtained by regression on Fig. 2, after removing the term qH ($H \approx -0.37 \pm 0.01$). It is clearly convex; low frequency speech is therefore multiscaling.

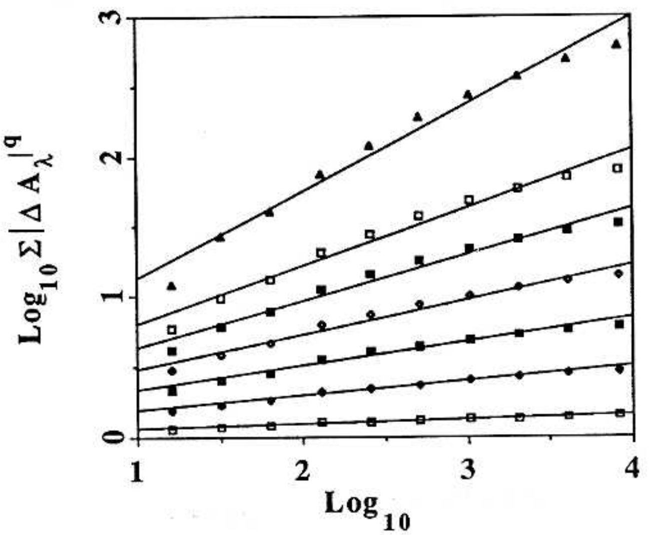

Fig. 2. The (log) sum over all disjoint intervals of the qth moment of the absolute speech fluctuation ($\Sigma|\Delta A_\lambda|^q \approx \lambda^{K_A(q)-(q-1)}$) as a function of the (log) scale ratio λ. From top to bottom q = 1.5, 1.1, 0.9, 0.7, 0.5, 0.3, 0.1. $K_A(q)$ is determined from the slope. Note that the multiple scaling is well followed over the entire range corresponding to ≈ 10 Hz to 10^{-3} Hz.

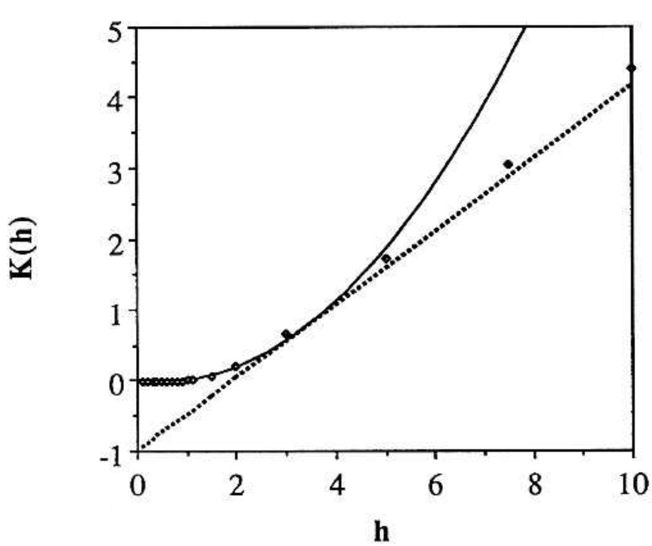

Fig. 3. The empirical $K(q)$ function obtained by subtracting $qK_A(1)$ from the slopes of Fig. 2, with the optimum piecewise quadratic/linear fit (Eq. 4) with $C_1 = 0.09$.

If φ is the result of nonlinear "mixing" (interaction) of many different processes [Schertzer & Lovejoy, 1987, 1989; Fan, 1989], most of the details of the process will be irrelevant and we may expect it to have the following universal form:

$$K(q) = \frac{C_1}{\alpha - 1}(q^\alpha - q)\,, \tag{3}$$

where C_1 is the codimension of the mean of φ_λ (in the one-dimensional time series here, $0 \le C_1 \le 1$ and the corresponding fractal dimension of the mean is $1 - C_1$). It represents the best monofractal approximation to the mean. α is the Levy exponent and characterizes the degree of multifractality.

It is fairly straightforward to estimate C_1; this can be done from, $K'(1) = C_1$ (yielding $C_1 \approx 0.09 \pm 0.01$). α is more difficult to estimate since it determines the concavity, and hence deviation from monofractality; $\alpha = 2$ is the maximum, $\alpha = 0$, the monofractal (minimum). A new method for estimating α, called the "Double Trace Moment" technique [Schertzer *et al.*, 1991; Lavallée, 1991], indicates that $\alpha = 1.95 \pm 0.1$, i.e. near its maximum: indeed, a quadratic ($\alpha = 2$) fit (Fig. 3) for the range $0 < q < 3$, yields an RMS error of only ± 0.02. The reason for this range restriction is that, above a critical value q_c, the empirical moments are no longer accurate estimates of the true (ensemble averaged) moments: we will expect a discontinuity in the slope and a linear behavior for the measured $K(q)$ for $q < q_c$ (as observed in Fig. 3 for $q > 3$). q_c can arise through two quite distinct causes (which mimic multifractal "phase transitions" [Schertzer *et al.*, 1992]):

(a) there is undersampling [Lavallée, 1991; Lavallée *et al.*, 1991]: any finite sample has a maximum order of singularity, hence all moments $q > q_s$ will be dominated by this value; this yields a second order phase transition;
(b) there is divergence of moments; for $q > q_D$, the theoretical $K(q)$ is infinite, and we detect a "spurious" scaling again characterized by a linear $K(q)$ for $q > q_D$; this yields a first order phase transition.

For a single sample in one dimension ($D = 1$), we obtain ($\alpha = 2$): $q_s = C_1^{-1/2}$, $q_1 = C_1^{-1}$ (hence

$q_s < q_1$). The overall empirical moments will follow:

$$K(q) = \begin{cases} C_1(q^2 - q)\,, & q < C_1^{-1/2}\,, \\ q(2C_1^{1/2} - C_1) - 1\,, & q > C_1^{-1/2}\,. \end{cases} \quad (4)$$

Using $C_1 = 0.09$, we find $q_c = q_s \approx 3.30$: Figure 3 shows that an excellent fit is obtained over the entire range $0 < q < 10$ (the maximum distance between the theoretical and empirical curves is 0.06).

For the other series we investigated, results were generally similar; α near 2, C_1 in the range 0.09 to 0.18 (a selection of opera singing yielding the largest). Just as many different nonscaling processes can give rise to Guassian noises, so many different scaling mechanisms for speech can give rise to universal multifractals with the same basic parameters. Traditional analysis methods have concentrated on the high frequency part of the spectrum using spectrograms and other methods sensitive to small differences in signal statistics: In contrast the universality parameters here are extremely insensitive to these differences. For this very reason, they quantify a very fundamental aspect of the speech process and help us rule out models of speech which break the scaling by introducing specific processes at well-defined time scales. Knowledge that low frequency speech fits into these broad universal categories will be important in improving our understanding of the high frequencies too. It also suggests other analyses — for example, the scaling of the distribution of tagged phomenes, syllables etc. that could be performed in the future. Finally, to demonstrate the usefulness of our characterization, we used continuous cascade processes [Schertzer & Lovejoy, 1987, 1989; Fan, 1989; Wilson *et al.*, 1991] calibrated with the measured parameters, to make simulated low frequency speech signals [see Fig. 2(b)]. These could provide the basis for full simulations by providing a signal to modulate the higher frequencies.

Acknowledgment

We thank E. Keller for kindly providing us with speech analysis software.

References

Brax, P. & Peschanski, A. [1991] "Levy stable law description of intermittent behaviour and quark-gluon phase transitions," *Phys. Lett.* **B253**, 225–230.

Clarke, J. & Voss, R. [1975] "$1/f$ noise in music and speech," *Nature* **258**, 317–318.

Fan, A. H. [1989] "Chaos additif et multiplicatif," *C. R. Acad. Sci. Paris* **I**, 308, 151–154.

Gabriel, P., Lovejoy, S., Schertzer, D. & Austin, G. L. [1988] "Multifractal analysis of satellite resolution dependence," *Geophys. Res. Lett.* **15**, 1373–1376.

Lavallée, D., Lovejoy, S., Schertzer, D. & Ladoy, P. [1992] "Nonlinear variability and landscape topography: analysis and simulation," in *Fractals in Geography*, eds. L. De Cola & N. Lam (in press, Prentice-Hall).

Lavallée, D., Lovejoy, S. & Schertzer, D. [1991a] "Universal multifractal theory and observations of land and ocean surfaces, and of clouds," *SPIE* **1558**, 60–75.

Lavallée, D., Schertzer, D. & Lovejoy, S. [1991b] "On the determination of the codimension function," in *Nonlinear Variability in Geophysics,: Scaling and Fractals*, eds. Schertzer, D. & Lovejoy, S., (Kluwer, Dordrecht, Holland) pp. 99–110.

Lavallée, D. [1991] "Multifractal analysis and simulation techniques and turbulent fields," PhD thesis, McGill University, pp. 132.

Lovejoy, S. & Schertzer, D. [1990] "Multifractals, universality classes, satellite and radar measurements of clouds and rain," *J. Geophys. Res.* **95**, 2021–2034.

Pickover C. & Khorasani, A. [1986] "Fractal characterization of speech waveform graphs," *Computers and Graphics* **10**, 51–61.

Ratti, S. P. *et al.* [1991] "Universal Multifractal analysis of multiparticle production in hadron-hadron collisions at $\sqrt{s} = 16.7$ GeV," *XXI International symposium on Multiparticle Production*, Wuhan, China, Sep. 23–27.

Schertzer, D., Lovejoy, S., Lavallée, D. [1992] "First and second order phase transitions in multifractal processes," submitted to *Phys. Rev. Lett.*

Schertzer, D., Lovejoy, S., Lavallée, D. & Schmitt, F. [1991] "Universal hard multifractal turbulence, theory and observations," in "*Nonlinear Dynamics of Structures*, eds. Sagdeev, R. Z. *et al.*, (World Scientific, Singapore) pp. 213–235.

Schertzer, D. & Lovejoy, S. [1991] "Nonlinear geodynamical variability: Multiple singularities, universality and observables," in *Scaling, Fractals and Nonlinear Variability in Geophysics*, eds. Schertzer, D., & Lovejoy, S., (Kluwer, Dordrecht, Holland) 41–82.

Schertzer, D. & Lovejoy, S. [1989] "Nonlinear variability in geophysics: multifractal analysis and simulations," in *Fractals: Their Physical Origins and Properties*, ed. Pietronero, L. (Plenum Press, New York) pp. 49–79.

Schertzer, D. & Lovejoy, S. [1987] "Physically based rain and cloud modeling by anisotropic, multiplicative turbulent cascades," *J. Geophys. Res.* **92**, 9692–9714.

Schmitt F., Lavallée, D., Schertzer, D. & Lovejoy, S. [1992] "Empirical determination of universal multifractal exponents in turbulent velocity fields," *Phys. Rev. Lett.* **68**, 305–308.

Tessier, Y., Lovejoy, S. & Schertzer, D. [1992] "Universal multifractals in rain and clouds: Theory and observations," in press *J. Appl. Meteor.*

Voss, R. & Clarke, J. [1978] "$1/f$ noise in music: music from $1/f$ noise," *J. Acoust. Soc. of Amer.* **63**, 258–263.

Wilson, J., Lovejoy, S., & Schertzer, D. [1991] "Physically based rain and cloud modeling," in *Scaling, Fractals and Nonlinear Variability in Geophysics*, eds. Schertzer, D. & Lovejoy, S. (Kluwer, Dordrecht, Holland) pp. 185–208.