# Analysis of the Multifractal Nature of Speech Signals

Diana Cristina González, Lee Luan Ling, and Fábio Violaro

DECOM – FEEC, Universidade Estadual de Campinas (UNICAMP)
{dianigon,lee,fabio}@decom.fee.unicamp.br

**Abstract.** Frame duration is an essential parameter to ensure correct application of multifractal signal processing. This paper aims to identify the multifractal nature of speech signals through theoretical study and experimental verification. One important part of this pursuit is to select adequate ranges of frame duration that effectively display evidence of multifractal nature. An overview of multifractal theory is given, including definitions and methods for analyzing and estimating multifractal characteristics and behavior. Based on these methods, we evaluate the utterances from two different Portuguese speech databases by studying their singularity curves ($\tau(q)$ and $f(\alpha)$).We conclude that the frame duration between 50 and 100 ms is more suitable and useful for multifractal speech signal processing in terms of speaker recognition performance [11].

**Keywords:** Multifractal Spectrum, Hölder Exponent, Speech Signals, Scaling Analysis, Multifractal Characteristics.

## 1 Introduction

In recent years, the use of the multifractal theory as an alternative method for non-stationary signal modeling has considerably increased. Most traditional approaches for signal modeling and analysis are based on the use of short-time spectral approach performed by the DFT [1], [2], mainly focusing on the signal's stationary properties [3]. Those traditional methods fail largely to characterize non-stationary behavior in signals and therefore are unable to explore the information contained in most of their transient and non-stationary parts. In fact, most real world signals and processes, such as speech and video, can be better characterized by their non-stationary behavior [4]. In literature, there are some works reporting the use of multifractal techniques in speech processing. In [3] a multifractal-based approach was employed for characterization of speech consonants. In [5], [6], fractal parameters were extracted and used as new nonlinear feature of speech signals. In terms of analysis of the multifractal nature of speech signals, in [7] the geometry of speech turbulence was fractally modeled. In [8] [9], the authors concluded that multifractal methods of can be used for signal processing such as decomposition, representation and spectrum characterization. In [10] the multifractal nature of unvoiced speech signals was studied and demonstrated.

The current work arises from the necessity of determining the appropriate frame duration to perform the multifractal analysis of speech signals. The results of this study have provided a solid basis for the design and implementation of the speaker

L. Alvarez et al. (Eds.): CIARP 2012, LNCS 7441, pp. 740–748, 2012.

recognition system in [11]. More specifically, multifractal characteristics presented in speech signals are studied using multifractal curves including the multifractal spectrum $f(\alpha)$ and the scaling functions $\tau(q)$. These curves (also called singularity curves) are capable of providing some essential information for speech signal processing, such as signal decomposition, representation and characterization, similar to that performed by traditional Fourier approaches [8]. Two databases with different sampling rates were tested in order to observe and determine the multifractal nature of speech signals under different time scaling conditions.

## 2 Multifractal Processes

Multifractal signals, as well as multifractal processes, are usually characterized by their highly irregular behavior. In other words, time functions exhibit abrupt and varying levels of instantaneous transitions in time, also known as singular points, at which the time function is non-differentiable. This singularity level measure can be obtained through estimation of the Lipschitz exponent, which provides the so-called uniform measures of regularity, either evaluating it on small time intervals (neighborhood) or at isolated points (pointwise) [12]. In multifractal processes, the Lipschitz exponent, also known as the Hölder exponent $\alpha_t$, is a series of time dependent values. In literature, there are two widely adopted definitions for "multifractals" in terms of their nonlinear characteristics of statistical moments, observed under different time scales, measured locally or pointwisely:

*Definition 1.*
The first definition of multifractals can be viewed as a generalization of monofractals [13]. Thus, it is said that a process $X(t)$ is multifractal when it obeys the following scale relationship $X(ct) \overset{d}{=} c^{H(c)}X(t)$, where $c^{H(c)}$ represents the scaling factor with $0 < H(c) < 1$ and $c>0$. The equality operator "$\overset{d}{=}$" indicates equality in statistical distribution. For monofractal processes, $H(c) = H$ is a constant which can be characterized by a single scale factor, known as the Hurst parameter. For multifractal processes, the generalized Hurst parameter becomes a Hölder exponent.

*Definition 2.*
The second definition of multifractal processes is based on the analysis of local scaling properties of the random paths of the process $X(t)$, by way of its local Hölder exponent, which is roughly defined as follows:

$$|X(t) - P_n(t)| \leq C|t - t_0|^{h(t_0)} \tag{1}$$

where $P_n(t)$ is a Taylor polynomial of $X$ in $t$ of degree $n$, for $t$ sufficiently close to $t_0$, The degree $n$ of the polynomial indicates the number of times that the tion $X(t)$ is differentiable at $t_0$. Therefore, $h(t_0)$ provides a measure of the singularity (or regularity) level at $t_0$. A complete and more rigorous version of this definition can be found in [13].

## 3 Estimation of Multifractal Characteristics

This section presents two practical approaches to study the multifractal behavior of a time series. The first approach is based on the estimation of the partition function of the process using the method of moments, while the second relies on the analysis of regularity of the process through its "multifractal spectrum".

### 3.1 The Method of Moments

The method of moments assumes that the signal holds major characteristics of a multiplicative cascade process [14]. The basic idea of this approach consists in acquire knowledge of the Hölder exponent distribution, by analyzing singularity property of the cascade. A procedure widely used for this analysis is the partition function. Let $\{X_i\}_{i=1}^{2^N}$ be the time series that represents a level of the cascade with a measure on the interval [0, 1] on the scale $1/2^N$. The partition function for the moment of order $q$ is defined as [14]:

$$\mathcal{X}_m^X(q) \coloneqq \sum_{k=1}^{N/m} \left(\overline{\boldsymbol{X}_k^{(m)}}\right)^q \tag{2}$$

where,

$$\overline{\boldsymbol{X}_k^{(m)}} \coloneqq \sum_{i=1}^{m} \boldsymbol{X}_{(k-1)m+i}^m \tag{3}$$

where $m$ define the aggregation number for the construction of the cascade processes, for instance, process with dyadic partition $m = 2, 4, 8 \ldots 2^N$. the time series elements $\boldsymbol{X}_{(k-1)m+i}^m$ represent the aggregate data, generating the new interval of next cascade stage for a fixed value of $m$. The scaling nature of the partition function can be evaluated by using the scaling function $\tau(q)$ as follows,

$$log\ X_m^X(q) = \tau(q)\, log\, m + C \tag{4}$$

where $C$ is constant. For the special case of multifractal processes, $log\ X_m^X(q)$ exhibits linearity with $log\, m$ and $\tau(q)$ is not linear in terms of $q$.

### 3.2 Multifractal Spectrum

The multifractal spectrum $f(\alpha)$ is a representation of the distribution of its Hölder exponents. The spectral function can be determined using some techniques such as coarse graining spectrum, Hausdorff spectrum, and Legendre spectrum. Due to its simplicity of the technique, this study focuses on the Legendre spectrum [12] which can be obtained by means of the Legendre transform of $\tau(q)$ (scaling function) [14], as $f(\alpha) = \min_q\{q\alpha - \tau(q)\}$. Typically the spectrum of a multifractal process has negative concave shape, where the horizontal axis indicates the Hölder exponent values and the vertical axis the total amount of points with the same exponent value. In

particular, when a signal process is monofractal, the scaling function becomes as $\tau(q) = \beta q - 1$, which is linear in $q$ with a constant angular coefficient $\beta$.As a result, the Hölder exponent holds a unique value graphically represented by a single non-zero point or a straight line.

## 4 Tests, Results and Discussion

In this section, we use the theory and procedures described in the previous section to study and determine possible multifractal nature and behavior of speech signals. The main purpose is to verify the conditions under which a speech signal reliably reveals its multifractal behavior.

### 4.1 Description of Speech Signals

Two speech signal databases were used for this experimental investigation [15]. The speech signals of these two databases were collected via a high-quality microphone and recorded under a low noise, controlled environment. The speech signals of the first database are collected from 30 speakers under the 11.025 kHz sample rate. The utterances have an average duration of 2.5 s. The second database, contributed by 71 speakers, has their speech signal sampled at rate of 22.05 kHz. The average duration of each speech utterance is approximately 3 s. Before the speech signals were submitted to multifractal analysis, they underwent a pre-processing procedure which consisted of three operations in sequel: pre-emphasis [16], normalization and elimination of silence intervals.

### 4.2 Experimental Investigation

In this subsection, we graphically evaluate the multifractal behavior of the speech signals. First, applying the moment method we obtain the partition function and the scaling function $\tau(q)$. Then, via the Legendre spectrum, we observe the scaling behavior and any particular event appearing on each speech segment, namely consonants, vowel transitions, vowel-consonant pairs.

*Experimental Test 1:*The moment method determines graphically the behavior of the partition functions in terms of moment order $q$. In this experiment test we randomly selected 30 speech phrases recorded from different speakers (with varying genders and ages) of each database. For illustration purpose, Figs. 1.a and 1.c show the curves of the partition functions ($\log \mathcal{X}_m^X$ versus $\log m$) of two phrases arbitrarily selected from the two different databases. In fact, similar graphic behaviors are observed for most of the evaluated utterances. Notice that the partition functions exhibit linearity in relation to $\log m$, despite some soft inflection points, regardless of the sampling frequency and utterance duration. This suggests that speech signals may hold fractal behavior or characteristics, presenting different scaling properties that are monofractal or multifractal behaviors at different scales. In contrast, the curves of the scaling function $\tau(q)$, as illustrated by Figs. 1.b and 1.d, show some nonlinearity, what suggests

the existence of a multifractal spectrum. The vertical bars represent 95% confidence intervals of the estimated values of the scaling function $\tau(q)$ for each moment der $q$, all estimated confidence intervals are small and present similar dynamic shapes in $q$. This visual inspection alone, although suggesting the presence of different properties of scaling, may not be definitive or conclusive. Therefore, a complementary analysis approach was adopted, using a multifractal spectrum (spectrum of Legendre). This approach is usually more reliable and informative [14].

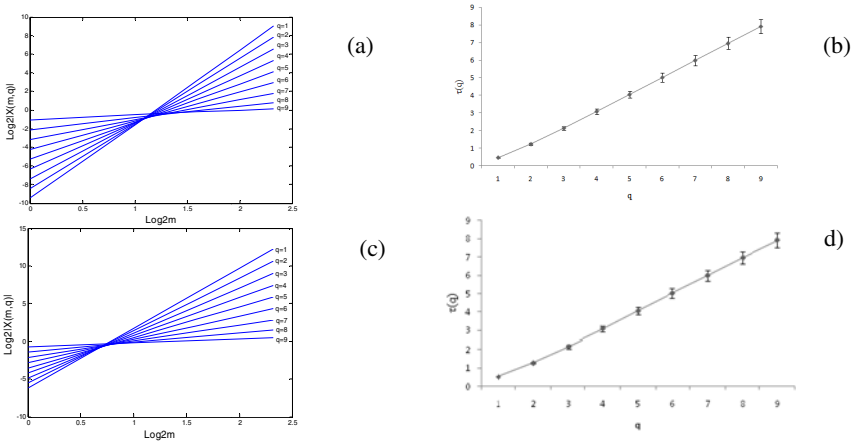


**Fig. 1.** Partition functions: (a) Database Ynoguti 1 (c) Database Ynoguti 2; scaling function $\tau(q)$ vs q; (b) Database Ynoguti 1, (d) Database Ynoguti 2

*Experimental Test 2:* In this experimental test, the speech signals are analyzed in terms of phonemes (the smallest sound units that conforms a word) as well as the relationship to their neighbor phonemes. As explained in Section 3-B, the multifractal spectrum provides information regarding the singularity degree of a time signal and, therefore, about the variation of its Hölder exponent function. Through this variation in time, the multifractal behavior of speech signals is determined. Here, the multifractal spectrum is obtained by applying Legendre Transform [1] (implemented in MATLAB) using the FRACLAB Tool [17].

**Table 1.** Phonetic classes

| Classes | Phonemes | Classes | Phonemes |
|---|---|---|---|
| Silences (s) | # | Fricatives (f) | f, s, ʃ, v, z, ʒ |
| Vowels (v) | a, e, ɛ, i, j, o, ɔ, u | Lateral (l) | l, ʎ |
| Nasal vowels (ns) | ɑ̃, ɛ̃, ĩ, õ, ũ | Nasals (n) | n, m, ɲ |
| Plosives (p) | p, t, tʃ, k, b, d, ʤ, g | Tap | r, r̄, R |

[1] More information on the Legendre Transform can be found in Appendix B from [14].

Since there are phoneme variations according to languages, this work focuses on 36 phones of the Brazilian Portuguese language. Table 1 shows to which class each phone belongs. After segmentation of speech, we analyzed the behavior of those segments. These segments were analyzed in different time scales (20-, 50-, 100- and 500-ms) in order to observe the dynamics of the speech in different scales and its singularities distribution. For smaller scales, the analysis focused on sub-segments of phonemes, implying practically the behavior analysis of isolated phonemes. For large time scales, the studied speech segments consist of a phoneme and its neighbor phonemes. The first phonetic class analyzed was "vowels". Vowels were chosen in different conditions, including start, end and middle of a word. For example, Fig. 2 illustrates the distribution of the Hölder exponents for the first vowel “a” in word “P**a**ra”. Due to space limitation, only predominating and informative results are presented.

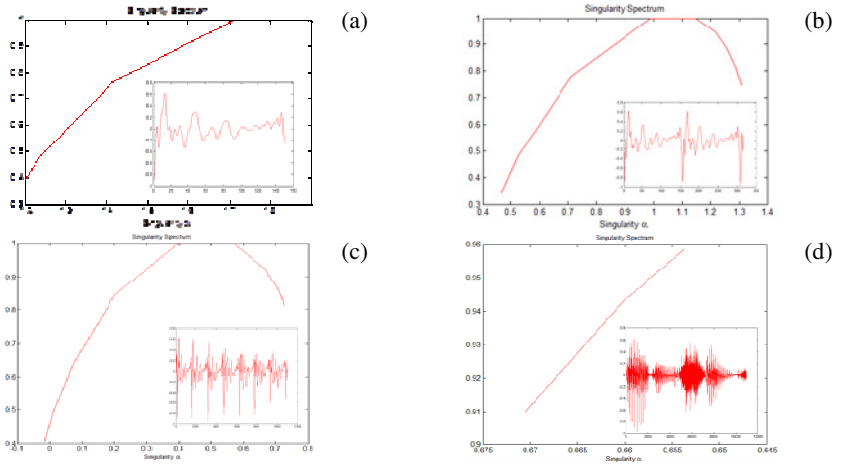


**Fig. 2.** Legendre spectrum of the phoneme "a". (a) One pitch period of the phoneme *“a”*, (b) Two pitch period of the phoneme *“a”*, (c)100 ms, (d) 500 ms.

The second studied phonetic class was "plosive consonants". Traditionally consonants can be characterized by their very short duration and are usually followed by vowels, and cause almost non-significant changes on the vowel sound. Therefore, in the vicinity of the plosive phoneme, the multifractal behavior is maintained in a similar manner to that given by the vowel phonemes. Fig. 3 shows two examples of such a phenomenon through the spectra of singularities of consonants “*t*” and “*b*”.

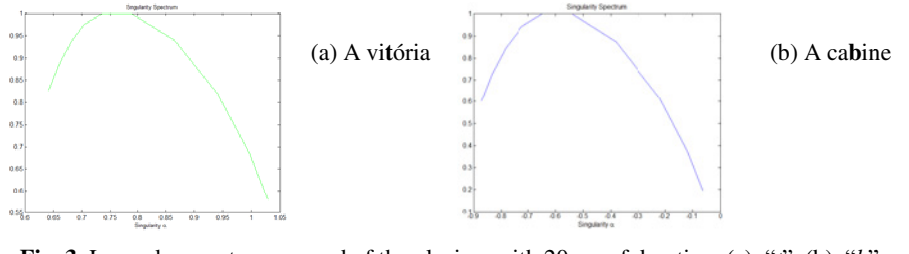


**Fig. 3.** Legendre spectrum around of the plosive, with 20 ms of duration. (a). “*t*”. (b). “*b*”.

The third phonetic class analyzed was "fricative consonants". Fricative sounds are consequences of the turbulence produced by lungs when the air is forced to pass through a constricted vocal tract [7]. During the analysis of fricatives, we observed two different behaviors which are exemplified by the “f” and “x” phonemes as shown in Fig. 4 and Fig. 5. Accordingly, the phoneme "f" of the word "foi", shown by Fig. 4, has random behavior similar to that of a random signal, which is usually characterized by monofractal processes. According to [7], the sound produced by one phone can be represented by one fractal dimension, and therefore can be better modeled by a monofractal process.

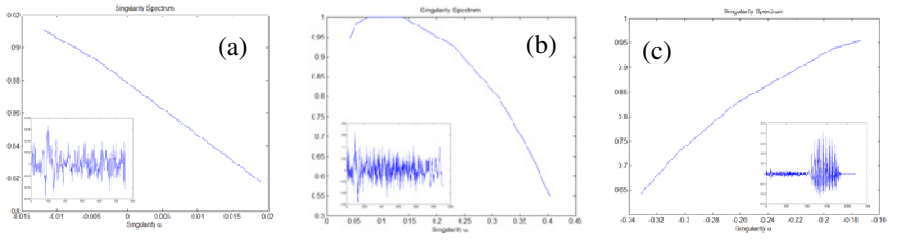


**Fig. 4.** Legendre spectrum of the phoneme "*f*". (a) 50 ms, (b) 100 ms, (c) 500 ms.

For spontaneous speech, fricative is usually accompanied by audible sounds in order to present multifractal characteristics at scales close to 100 ms. An example of this kind is illustrated by Fig. 4.b. We also observed that most of the fricative phones have similar behavior above mentioned. Another interesting phenomenon of fricatives is exemplified by the phoneme “s” in the word “próxima ['prɔsimɐ]”. The length of this fricative sound is very short; as a consequence, it rapidly connects to the followed vowel sound, as shown in Figs. 5.a and 5.b. As a result, the fricative component presents a multifractal behavior similar to that of vowel sounds, especially at low time scales (20- and100-ms).

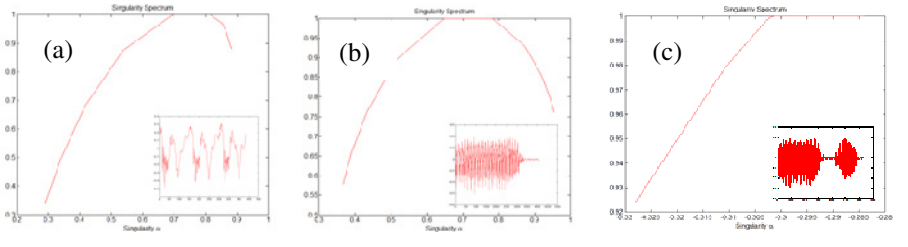


**Fig. 5.** Legendre spectrum of the phoneme "*x*". (a) 20 ms, (b) 100 ms, (c) 500 ms .

Generally speaking, every studied speech segment has demonstrated monofractal behavior on large time scales and multifractal behavior on small time scales. However, the range of small time scales on which the multifractal characteristics are observed varies. For instance, the phoneme “b” (Fig. 3) and the phoneme “f”

(Fig. 4.a show multifractal behavior on time scales shorter than 20 ms and greater than 50 ms, respectively. After detailed analysis and comparison, we found that spontaneous speech signals reliably present multifractal behavior on the range between 50m and 100ms time scales.

## 5 Conclusions

After extensive tests and evaluations performed on speech signals selected from two different speech databases, we summarize our conclusions as the following:

- Speech signals may present either monofractal or multifractal behavior depending on the time scales under which observation and analysis are performed. Experimental results show that speech signals composed of some phonetics classes (fricatives or taps) present monofractal behavior under short time interval analysis. This same behavior was found for long speech segment for all studied signals. However, definitely there is no rigid boundary for fractal behavior classification, due to fact that speech signal dynamics varies and is highly affected by both the speaker's speech rate and the signal's structure.
- It was found that, in general, all speech signals reveal multifractal behavior under a time frame analysis ranging from 50ms to 100ms. As this time interval includes a phone or a phone transition, we believe that this result is independent of the language.
- A new speaker recognition system in [11] that incorporates some multifractal features has reported a 3% increase in recognition rate with respect to one using only classical Mel-Cepstral features. This implies that multifractal features have increased and provided additional pattern discriminating capabilities.

## References

1. Campell, J.: Speaker Recognition: A Tutorial. Proceeding of the IEEE 85(9) (1998)
2. Reynolds, D.A., Rose, R.C.: Robust Text-Independent Speaker Identification Using Mixture Speaker Model. IEEE Trans. Speech Audio Processing 3(1), 72–82 (1995)
3. Langi, A., Kinsner, W.: Consonant Characterization Using Correlation Fractal Dimension for Speech Recognition. In: Proc. on IEEE Western Canada Conference on Communications, Computer, and Power in the Modem Environment, Winnipeg, Canada, vol. 1, pp. 208–213 (1995)
4. Jayant, N., Noll, P.: Digital Coding of Waveforms: Principles and Applications to Speech and Video, 688 p. Prentice-Hall, Englewood Cliffs (1984)
5. Sant'Ana, R., Coelho, R., Alcaim, A.: Text-Independent Speaker Recognition Based on the Hurst Parameter and the Multidimensional Fractional Brownian Motion Model. IEEE Trans. on Audio, Speech, and Language Processing 14(3), 931–940 (2006)
6. Zhou, Y., Wang, J., Zhang, X.: Research on Speaker Recognition Based on Multifractal Spectrum Feature. In: Second International Conference on Computer Modeling and Simulation, pp. 463–466 (2010)

7. Maragos, P.: Fractal Aspects of Speech Signals: Dimension and Interpolation. In: Proc. IEEE ICASSP, vol. 1, pp. 417–420 (1991)
8. Langitt, A., Soemintapurat, K., Kinsners, W.: Multifractal Processing of Speech Signals Information, Communications and Signal Processing. In: Han, Y., Quing, S. (eds.) ICICS 1997. LNCS, vol. 1334, pp. 527–531. Springer, Heidelberg (1997)
9. Kinsner, W., Grieder, W.: Speech Segmentation Using Multifractal Measures and Amplification of Signal Features. In: Proc. 7th IEEE Int. Conf. on Cognitive Informatics (ICCI 2008), pp. 351–357 (2008)
10. Adeyemi, O.A.: Multifractal Analysis of Unvoiced Speech Signals. ETD Collection for University of Rhode Island. Paper AAI9805227 (1997)
11. González, D.C., Lee, L.L., Violaro, F.: Use of Multifractal Parameters for Speaker Recognition. M. Eng. thesis, FEEC/UNCAMP, Campinas, Brazil (2011)
12. Sténico, J.W., Lee, L.L.: Estimation of Loss Probability and an Admission Control Scheme for Multifractal Network Traffic. M. Eng. thesis, FEEC/UNCAMP, Campinas, Brazil (2009)
13. Riedi, R.H., Crouse, M.S., Ribeiro, V.J., Baraniuk, R.G.: A Multifractal Wavelet Model with Application to Network Traffic. IEEE Trans. on Information Theory 45(3), 992–1018 (1999)
14. Krishna, M.P., Gadre, V.M., Dessay, U.B.: Multifractal Based Network Traffic Modeling. Kluwer Academic Publishers., Ed. Bombay (2003)
15. Ynoguti, C., Violaro, F.: Continuous Speech Recognition Using Hidden Markov Models. D. Eng. thesis, FEEC/UNCAMP, Campinas, Brazil (1999)
16. Holmes, J., Holmes, W.: Speech Synthesis and Recognition, 2nd edn. Tayor & Francis, London (2001)
17. Research Center INRIA Saclay, `http://fraclab.saclay.inria.fr/`