Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 72 (2015) 1–12

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Wavelet leader multifractal analysis of period and amplitude sequences from sustained vowels

Roberto F. Leonarduzzi *, Gabriel A. Alzamendi, Gastón Schlotthauer, María E. Torres

Lab. de Señales y Dinámicas No Lineales, Fac. de Ingeniería, Universidad Nacional de Entre Ríos, Argentina
Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

Received 4 December 2014; received in revised form 29 March 2015; accepted 20 April 2015
Available online 29 April 2015

Abstract

Irregularities in the amplitude and period are characteristic of both normal and pathological sustained vowels; they are a product of perturbations inherent in the phonation process. Their analysis provides useful diagnostic information for several vocal pathologies, and their accurate modelling has been shown to improve the quality of synthesized voice. In this work, we propose the application of multifractal analysis for the characterization of amplitude and period fluctuations in sustained vowels. Using a combination of high order statistics, this signal processing tool generalizes previous approaches and provides a rich description of the fluctuation in the regularity of the data. Our results suggest that both amplitude and period fluctuations show a multifractal behavior, independent of the gender of the speaker. We also analyze the problem of classification between healthy and nonhealthy speakers as an example to show the usefulness of multifractal attributes. We conclude that amplitude and period sequences of sustained vowels should be analyzed and modelled by the multifractal paradigm.

Keywords: Multifractal analysis; Wavelet leaders; Sustained vowels; Amplitude and period sequences

1. Introduction

Irregularities in the amplitude and period are characteristic of the normal voice as a product of perturbations inherent in the phonation process. Despite the fact that they are more evident in the presence of voice disorders, some degree of irregularity is seen even in nonpathological stable conditions (Titze, 1995; Bonilha and Deliyski, 2008). These perturbations occur in an unpredictable fashion and are usually concealed or modulated in the speech records (Leong et al., 2013). It is currently believed that perturbations arise from a combination of neurological, biomechanical, aerodynamic and acoustic sources, throughout the speech production system (Titze, 2000). Additionally, it has been argued that perturbations have different dynamics in pathological and nonpathological voices (Baken and Orlikoff, 2000; Velasco García et al., 2011). Therefore, acoustical parameters quantifying these disturbances have become useful instruments in voice screening, pathology detection and vocal therapy.

Amplitude and period sequences (denoted as AS and PS, respectively, from now on) are generally composed of identifiable structures presenting short- and long-term behavior (Silva et al., 2009; Fraj et al., 2012). In particular, short-term structures carry the complete information related to period and amplitude perturbations, consisting of random cycle-to-cycle fluctuations known as jitter and

* Corresponding author at: Facultad de Ingeniería, UNER, C.C. 47, Suc. 3, 3100 Paraná, Entre Ríos, Argentina. Tel.: +54 0343 4975100x122.
E-mail addresses: rleonarduzzi@bioingenieria.edu.ar (R.F. Leonarduzzi), galzamendi@bioingenieria.edu.ar (G.A. Alzamendi), gschlott@bioingenieria.edu.ar (G. Schlotthauer), metorres@santafe-conicet.gov.ar (M.E. Torres).

http://dx.doi.org/10.1016/j.specom.2015.04.004
0167-6393/

shimmer, respectively (Titze and Liang, 1993; Titze, 1995). It has been recently shown that both AS and PS provide valuable information about the speaker itself – e.g. identity, gender or mood (Low et al., 2011, 2010) – or related to the physiology of the speech production system – e.g. vocal folds dynamics (Fraile et al., 2012). As a consequence, theoretical models and biomedical tools have been developed and successfully used in a wide variety of applications, like enhancement of the perceptual quality of synthesized voices in communication systems (Ruinskiy and Lavner, 2008), synthesis of expressive voices in human–computer interfaces and entertainment technology (Govind and Prasanna, 2013), identification of speakers in security systems (Farrus and Hernando, 2009), and simulation of pathological voices under controlled conditions (Schlotthauer et al., 2010), among many others.

Many different models have been used for AS and PS. Gaussian models with an adjustable jitter or shimmer factor have recently been used for neutral voice transformation in emotional speech synthesis (Cabral and Oliveira, 2006), fundamental frequency estimation (Schlotthauer et al., 2010) and for voice synthesis with high perceptual quality (Alzamendi et al., 2013). Further, Gaussian mixture models have been considered for jitter analysis in emotional speech (Wang et al., 2006). More complex strategies, which consider the correlation of the sequences, have been based on auto-regressive (AR) and auto-regressive moving average (ARMA) models (Schoentgen and De Guchteneere, 1995; Endo and Kasuya, 1996; Schoentgen and De Guchteneere, 1997). A jitter-bank based approach to characterize the relative amplitude and correlation in PS and a stochastic model of shimmer, suitable for naturally hoarse voice synthesis, has been proposed in Ruinskiy and Lavner (2008). Also, models based on stochastic difference equations have been proposed in Schoentgen (2001), and used for hoarse voice synthesis (Fraj et al., 2009; Fraj et al., 2012) and for training voice pathologists (Dejonckere et al., 2011; Manfredi et al., 2011). Recently, we have proposed a new method for modelling PS extracted from real voices, combining state-space methods and structural time series analysis (Alzamendi et al., 2015). Non-linear models have also been considered in Zhang et al. (2005). Finally, scale-invariant statistically self-similar processes have also been used for the synthesis of sustained vowels with an improved quality (Aoki and Ifukube, 1999).

All the techniques that were mentioned above amount to a *monofractal* analysis of the data. This means that realizations of those stochastic models are characterized by a local regularity exponent which is constant in time. The dynamics of this kind of data can be fully characterized by second order statistics. However, this simple framework does not necessarily capture the full complexity of AS and PS. It is now known that a better description of the singular behavior of the data can be achieved by the joint analysis of all statistical moments in the framework of *multifractal analysis* (Jaffard et al., 2014). This framework allows the description and modelling of data for which a local regularity exponent changes drastically with time, thus providing richer possibilities for analysis and modelling. As far as the authors know, there have been no previous analysis to determine which kind of framework should be used to model AS and PS data.

Multifractal analysis allows to measure information encoded in the fluctuations of the data. It is traditionally based on a measure of the local regularity provided by the Hölder exponent. Rather than focusing on the value that the Hölder exponent takes on each time instant, multifractal analysis provides a global description of the distribution of the Hölder exponent throughout the data by means of a *multifractal spectrum*. Its practical estimation is tied to the use of a specific multiresolution quantity through a *multifractal formalism* (Riedi, 2003). Among the many multiresolution quantities that have been proposed in the literature, the wavelet leaders (Jaffard, 2004; Jaffard et al., 2007, 2014) benefit from a solid theoretical background and excellent practical performance. Multifractal analysis has been widely used in a large variety of fields, including but not limited to turbulence (Lashermes et al., 2008), finance (Calvet and Fisher, 2001), climatology (Lovejoy and Schertzer, 2013), art investigation (Abry et al., 2013), and biomedical data such as heart beat (Ivanov et al., 1999; Leonarduzzi et al., 2010; Doret et al., 2011), fMRI (Ciuciu et al., 2012) and thermography for cancer detection (Gerasimova et al., 2014).

In light of the previous discussion, the goal of the present contribution is to investigate the application of multifractal analysis to AS and PS of sustained vowels. We concentrate on this vocal emissions because they are usually considered in clinical practice for speech assessment. We hypothesize that the use of models which better take into account the rich complexity present in both sequences, might improve the quality of voice synthesis and provide more detailed information to voice analysts. In particular, the aim of this contribution is twofold. First, to assess whether AS and PS data actually show scale-invariant and multifractal characteristics. Second, to provide an example of the benefits that the incorporation of multifractal characteristics could provide over the use of only monofractal attributes. To that end, we provide an example of discrimination between healthy and nonhealthy voices.

We start with a brief review of wavelet leader multifractal analysis in Section 2, followed by a description of the database we used and the experimental setup in Section 3. Then, in Section 4 we discuss the results that we obtained. First, we analyze records of normal voices from a large database to determine if AS and PS dynamics indeed show a scale invariant (Section 4.1) and a multifractal behavior (Section 4.2). Then, in Section 4.3, we provide a particular example of how the information provided by multifractal analysis can be used to classify between healthy and nonhealthy voices. Finally, in Section 5, we present the conclusions of this work.

2. Multifractal analysis and wavelet leaders

In this section, we provide a brief review of the main concepts of wavelet leader multifractal analysis. For further details, we refer the interested reader to e.g. Jaffard (2004) and Wendt et al. (2007, 2009).

2.1. Local regularity and multifractal spectrum

Let $X(t), t \in \mathbb{R}$, be a locally bounded function or the sample path of a stochastic process. Then, X is said to locally belong to $C^\alpha(t)$, with $\alpha \in \mathbb{R}$ and $\alpha \geqslant 0$, if there exists a constant $K > 0$ and a polynomial $P_t(t)$ such that $|X(t+a) - P_t(t+a)| \leqslant K|a|^\alpha$. The *Hölder exponent* is defined as $h(t) = \sup\{\alpha : X \in C^\alpha(t)\}$ and measures the local regularity of X. Indeed, the closer $h(t)$ is to 0, the more irregular $X(t)$ is.

Typically the function $h(t)$ is extremely erratic and is not of much use itself. Therefore, rather than analyzing the fluctuations of $h(t)$ as a function of time, multifractal analysis is concerned with the global distribution of the values that the Hölder exponent can take on. This notion is formalized in terms of the *multifractal spectrum* $D(h)$, defined as $D(h) = \mathrm{Dim_H}\{t \in \mathbb{R} : h(t) = h\}$, where $\mathrm{Dim_H}$ denotes the Hausdorff dimension. Loosely speaking, $D(h)$ provides information on the *proportion* of points in X with Hölder exponent h. For further details, the reader is referred to e.g. Riedi (2003), Jaffard (2004), and Jaffard et al. (2014) and references therein.

In practice, however, the multifractal spectrum cannot be reliably computed from its definition. On the contrary, a procedure that allows its estimation from easily measurable quantities must be used. This procedure is called a *multifractal formalism*. A particular version, the wavelet leader multifractal formalism, is described in the following sections.

2.2. Wavelet coefficients and wavelet leaders

Let $\psi_0(t)$ denote a compact support mother wavelet, characterized by its number of *vanishing moments*, a positive integer N_ψ such that $\int_{\mathbb{R}} t^k \psi_0(t)dt = 0$, $\forall k = 0, 1, \ldots, N_\psi - 1$, and $\int_{\mathbb{R}} t^{N_\psi}\psi_0(t)dt \neq 0$. The mother wavelet is chosen such that the set $\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k),\ j \in \mathbb{N},\ k \in \mathbb{N}\}$ forms an orthonormal basis of $L^2(\mathbb{R})$. The wavelet coefficients of a signal X are defined as $d_X(j,k) = 2^{-j/2}\int_{\mathbb{R}} X(t)\psi_{j,k}(t)dt$ (note the use of the non conventional L^1 norm that better matches multifractal analysis). In practice, these inner products are computed efficiently with the classical piramidal recursive algorithm, using a suitable filter bank. For a detailed introduction to wavelet analysis, the reader is referred to e.g. Mallat (2009).

Let us denote the *dyadic intervals* with $\lambda_{j,k} = [k2^j, (k+1)2^j)$, the concatenation of 3 such intervals as $3\lambda_{j,k} = \bigcup_{m\in\{-1,0,1\}}\lambda_{j,k+m}$, and $d(j,k) \equiv d_\lambda$. The wavelet leader $L_X(j,k)$ is defined from wavelet coefficients as a local supremum, taken within a narrow time neighborhood of $t = 2^j k$ at any finer scales $2^{j'} < 2^j$ (Jaffard, 2004; Wendt et al., 2007): $L_X(j,k) = \sup_{\lambda' \subset 3\lambda} |d_{\lambda'}|$.

2.3. Wavelet leader multifractal formalism

The wavelet leader multifractal formalism allows to estimate $D(h)$ from easily computable global quantities termed *structure functions* $S(q,j)$, defined as

$$S(q,j) = \frac{1}{n_j}\sum_{k=1}^{n_j} L_X^q(j,k), \tag{1}$$

where n_j is the number of leaders available at scale 2^j. That is, $S(q,j)$ is the qth sample moment of $L_X(j,k)$ at scale 2^j.

Assuming that the structure functions follow a power-law behavior across the scales 2^j:

$$S(q,j) \simeq C_L(q)2^{j\zeta(q)}, \tag{2}$$

it is well known that a Legendre transform of the *scaling exponents* $\zeta(q)$ yields an (upper bound) estimate of the multifractal spectrum $\mathcal{L}(h) = \inf_q(1 + qh - \zeta(q)) \geqslant D(h)$, cf. e.g. Riedi (2003) and Jaffard (2004).

It is important to emphasize that this estimation procedure relies on the assumption that the model in Eq. (2) holds. From a practical perspective, this means that the data must be shown to exhibit such a behavior before this analysis procedure can be used.

2.4. Log-cumulants

To avoid the computation of the function $\zeta(q)$ for all qs, it has been proposed to make use of a polynomial expansion (Delour et al., 2001; Wendt et al., 2007): $\zeta(q) = \sum_{p\geqslant 1} c_p \frac{q^p}{p!}$. Interestingly, it was shown that the coefficients c_p can be related to the cumulants of order p, $C_p(j)$, of the log-leader $\ln L_X(j,k)$ according to the linear behaviors:

$$C_p(j) = c_{0,p} + c_p \ln 2^j, \quad \forall p \geqslant 1. \tag{3}$$

where the intercepts $c_{0,p}$ play no role in multifractal analysis. Often, using only the first two cumulants yields a satisfactory approximation for both $\zeta(q) \approx c_1 q + c_2 q^2/2$ and $D(h) \approx 1 + (h - c_1)^2/(2c_2)$. The definition of cumulants implies that if $c_p = 0$, then $c_{p'} = 0$ for all $p' > p$. For further details, see e.g. Wendt et al. (2009) and Jaffard et al. (2014).

2.5. Monofractal and multifractal models

Monofractal models are defined as those where the Hölder exponent takes on a constant value, i. e. $h(t) \equiv H$, where H is the so-called *Hurst exponent*. Therefore, it is clear that the multifractal spectrum of monofractal data collapses to a single point: $D(H) = 1$ and $D(h) = -\infty$ for $h \neq H$. The Legendre transform

indicates that the scaling function of monofractal data is linear: $\zeta(q) = qH$; moreover, the log-cumulants satisfy $c_1 = H$ and $c_p = 0$ for $p \geqslant 2$.

On the contrary, the Hölder exponent of multifractal models is an extremely erratic function of time, which takes on many different values. Therefore, its multifractal spectrum has a broad support and its scaling function is not linear. This issue is further discussed in Section 4.2, and is illustrated in Fig. 4 for real voice data.

2.6. Practical estimation

As studied in details in Wendt et al. (2007, 2009), the estimation of $\zeta(q)$ (resp., c_p) can be performed by linear regression of $\log_2 S(j,k)$ (resp., $(\log_2 e)C_p(j)$) against $\log_2 2^j = j$:

$$\zeta(q) = \sum_{j=j_1}^{j_2} w_j S(q,j), \tag{4}$$

$$c_p = (\log_2 e)\sum_{j=j_1}^{j_2} w_j C_p(j), \tag{5}$$

where $C_p(j)$ are the standard sample cumulant estimators. The weights w_j are computed using the standard formula $w_j = b_j(Y_0 j - Y_1)/\left(Y_0 Y_2 - Y_1^2\right)$, with $Y_i = \sum_{j\in\underline{j}} j^i b_j$, $i = 0,1,2$. The coefficients b_j reflect the confidence granted to each scale in the regression.

In the case of monofractal processes, the traditional estimator for the Hurst exponent is based on the scaling function computed from wavelet coefficients. Let

$$S(j) = \frac{1}{n_j}\sum_{k=1}^{n_j} |d_{j,k}|^2.$$

Then the Hurst exponent can be estimated by means of the linear regression:

$$H = \sum_{j=j_1}^{j_2} w_j S(j), \tag{6}$$

where the weights w_j satisfy the same constraints than before. For more details on this classical estimator, the reader is referred to e.g. Abry et al. (2003, 2000).

2.7. Uniform Hölder exponent and fractional integration

Both the definition of the Hölder exponent (Section 2.1) and of the wavelet leaders (Section 2.2) require the data X to be locally bounded. This requirement is equivalent to the condition that the minimum regularity $h_m = \inf_t h(t)$ must be non-negative (Wendt et al., 2009; Jaffard et al., 2014), which can be easily verified in practice by using an estimate based on wavelet coefficients (cf. Wendt et al., 2009; Jaffard et al., 2014):

$$h_m = \lim_{j\to-\infty} \frac{\sup_k |d(j,k)|}{j}. \tag{7}$$

Therefore, if $h_m \geqslant 0$, then X is locally bounded and the multifractal formalism can be applied.

If this condition is not met, the problem can be circumvented by performing a fractional integration of order $\gamma \geqslant h_m$ on the data before performing the analysis. The effect of this procedure is to shift the multifractal spectrum of the integrated data: $D_\gamma(h) = D(h-\gamma)$ (Wendt et al., 2009; Jaffard et al., 2014; Abry et al., 2013). Therefore, all estimations can be safely performed on the integrated data and then the spectrum can be shifted back.

To avoid the painful process of fractional integration, it has been proposed in Wendt et al. (2009) to use a *pseudo fractional integration in the wavelet domain*, which amounts to computing the wavelet leaders from the modified wavelet coefficients $d_\gamma(j,k) = 2^{j\gamma} d(j,k)$. Contrary to fractional integration, this procedure is straightforward to implement in practice, and has been shown to provide equivalent results (Wendt et al., 2009; Abry et al., 2013; Jaffard et al., 2011).

3. Experimental setup

3.1. Database

Our study was performed using the Disordered Voice database, recorded at the Massachusets Eye and Ear Infirmary, and provided by KayElemetrics Corp (Massachusetts Eye and Ear Infirmary MEEI Voice and Speech Lab, 2009). The database contains records of the sustained vowel /a/ from 53 normal speakers and 653 talkers suffering from a variety of organic, neurological, psychogenic or traumatic vocal disorders. Clinical information for each record is provided, gathering the results of clinical trials and the opinion of experts. Voice samples from nonhealthy subjects were recorded in a soundproof booth, at a sampling rate of 25 kHz and 16 bit quantization (Massachusetts Eye and Ear Infirmary MEEI Voice and Speech Lab, 2009). The subjects were asked to produce a sustained phonation of the vowel /a/ at a comfortable pitch and loudness (Parsa and Jamieson, 2000). Samples from normal subjects were recorded at KayElemetrics in equivalent acoustic conditions (Massachusetts Eye and Ear Infirmary MEEI Voice and Speech Lab, 2009). Normal speakers were not examined for voice disorders, but none of them had a history nor complaints of such disorders (Parsa and Jamieson, 2000). We made use of the subset of 175 nonhealthy speakers listed in Parsa and Jamieson (2000). This subset was chosen to include subjects with similar age distribution, and a balanced male and female ratio. Moreover, it has been widely used in several studies, cf. e.g. Markaki and Stylianou (2011), Henríquez et al. (2009), and Arias-Londono et al. (2011). Descriptive statistical information about the

analyzed subjects can be found in Parsa and Jamieson (2000), and is reproduced in Table 1 for convenience.

3.2. Amplitude and period sequence extraction

We obtained PS from the voice records by means of *Praat* (Boersma and Weenink, 2013), a widely used software for objective voice analysis. *Praat* uses a short-term analysis procedure, where pitch periods are obtained from waveform-matching methods. This technique consists in the estimation of the location of fixed points in the glottal cycle where two consecutive waveforms look maximally similar by means of autoregressive analysis (Boersma, 1993, 2009). These locations are called *pitch marks*. Then, we computed pitch periods as the difference between pitch marks. Finally, we computed the amplitudes as the difference between the maximum and minimum values on each period, following the guidelines in Titze (2000). Fig. 1 illustrates the definition of pitch marks, AS and PS.

3.3. Analysis parameters

We performed the wavelet leader multifractal analysis using the MATLAB toolbox `WLBMF`, developed by members of the SiSyPhe group, ENS-Lyon, France (Wendt et al., 2007), freely available at www.irit.fr/Herwig.Wendt/software.html. We used a compactly supported Daubechies mother wavelet. Given that the analyzed voice data is short, we chose the number of vanishing moments $N_\psi = 2$ in order to minimize the support of the wavelet and, therefore, the number of octaves polluted by border effects. This choice is perfectly legitimate since the analyzed data never showed regularity smoother than $h = 1$. Furthermore, we observed no high-order trends superimposed to the data, rendering the detrending power provided by the wavelet acceptable. We also checked that results were consistent when N_ψ was increased.

An a priori analysis of the data, using the estimator in Section 2.7, showed very disparate values of the uniform Hölder regularity h_m, in particular between healthy and nonhealthy data (this is further discussed in Section 4.3). It was found that most representative records (male and female, AS and PS healthy and nonhealthy) are characterized by $h_m > -1$. Records which did not satisfy this condition showed poor fits for the linear model in Eq. (7); they were considered as outliers and not influential in the selection of the integration order. Therefore, following the guidelines in Jaffard et al. (2014), we performed a pseudo-fractional integration of the same order $\gamma = 1$ to all data before carrying out the multifractal analysis.

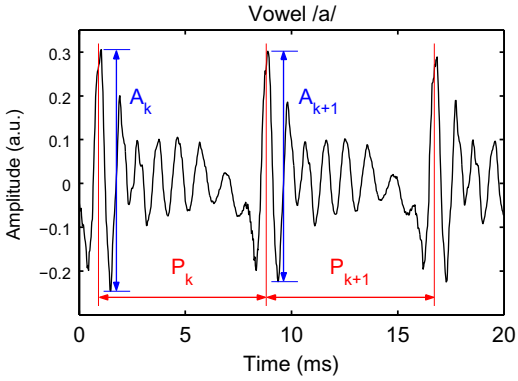

Fig. 1. Obtention of amplitude and period sequences. Sample voice record and illustration of the definition of analyzed sequences. A_k and P_k stand for the kth sample of AS and PS respectively.

We performed all linear regressions setting the weights $b_j \equiv 1$, that is, using ordinary unweighted regressions.

4. Results and discussion

4.1. Evidence of scaling

As a first step in our analysis, we verified the existence of a scaling behavior in both AS and PS. As we mentioned in Section 2.3, the presence of scaling is the main hypothesis the data must be shown to meet in order to enable a valid application of the multifractal formalism.

Fig. 2 shows (black lines) $\log_2 S(q,j)$ and $\log_2(e)C_p(j)$ (Eqs. (1) and (3)), for selected choices of the orders q and p, for the AS of a single selected speaker (record GPC1NAL). 95% bootstrap-based confidence intervals are shown in red lines (cf. e.g. Wendt et al. (2007) for details on how this confidence intervals are computed.) There is evidence of a very clear linear behavior in a wide range of scales, $j \in [2,5]$, for all cases. In fact, this behavior comprises almost the entire range of scales available from the wavelet decomposition. Blue lines show the linear fits obtained in each case, displaying a perfect fit with the data in the selected scaling range and suggesting an excellent agreement between the data and the theoretical model of scale invariance. Fig. 3 shows equivalent results for the PS of the same speaker. Again, the data display an excellent scaling behavior.

A theoretical requirement on multifractal models is that the range of scales where the scale invariance is observed must be the same for all statistical orders, i. e., q in the case

Table 1
Analyzed subjects: descriptive statistics. Statistics for number, age and length of analyzed records, for healthy and nonhealthy subject, and for each gender.

Group	Number		Age (years) mean ± std		Length (samples) mean ± std	
	Male	Female	Male	Female	Male	Female
Healthy	20	33	38.8 ± 8.3	34.2 ± 7.7	433 ± 69	710 ± 106
Nonhealthy	70	103	41.8 ± 9.3	37.4 ± 8.1	116 ± 25	185 ± 46

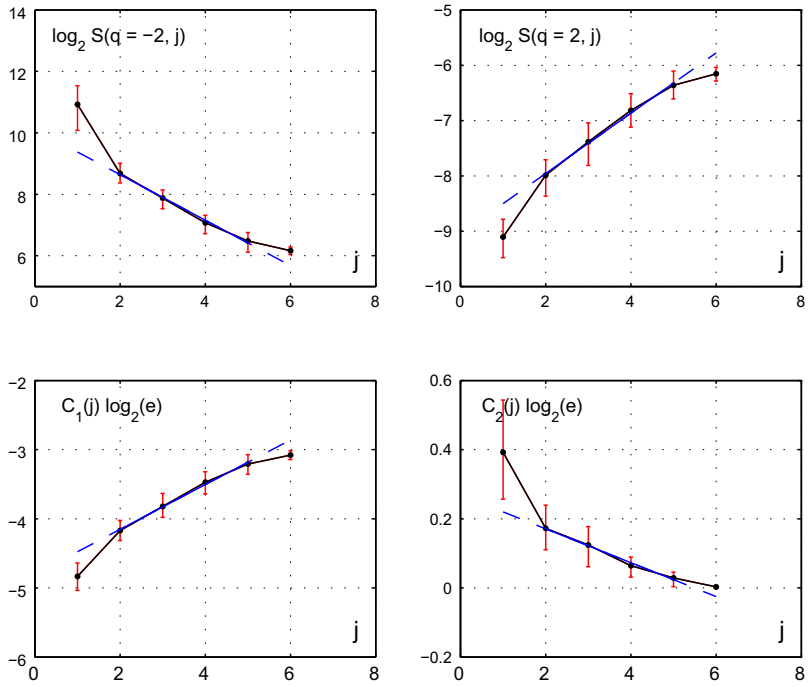

Fig. 2. AS: structure functions and cumulants. $\log_2 S(q,j)$ (top row) and $\log_2(e)C_p(j)$ (bottom row), for several q and p (black lines), bootstrap-based 95% confidence intervals (red lines) and least square linear fit (blue line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

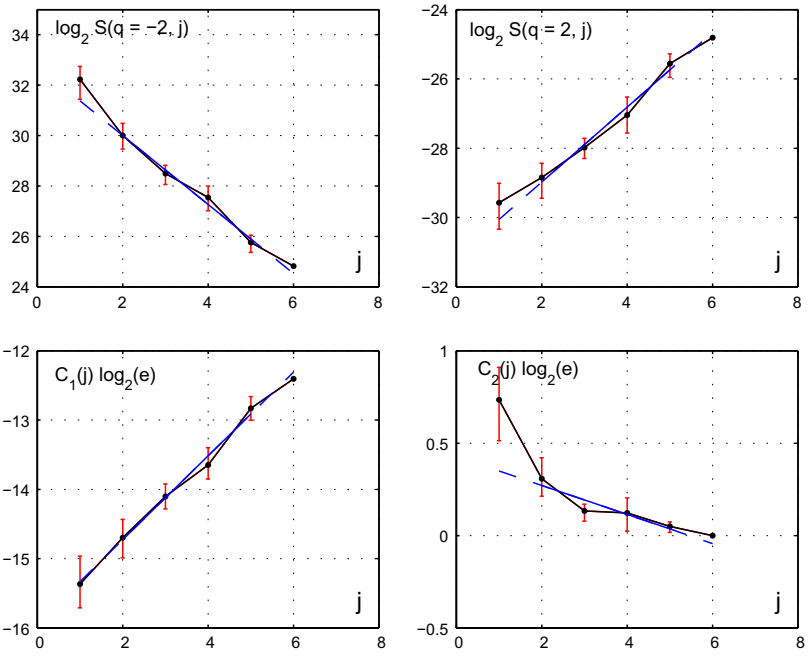

Fig. 3. PS: structure functions and cumulants. $\log_2 S(q,j)$ (top row) and $\log_2(e)\,C_p(j)$ (bottom row), for several q and p (black lines), bootstrap-based 95% confidence intervals (red lines) and least square linear fit (blue line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of $S(q,j)$ or p in the case of $C_p(j)$. This is indeed the case in Figs. 2 and 3, where the scaling range $j \in [2,5]$ is appropriate for both $S(q,j)$ and $C_p(j)$, for all statistical orders, and for both sequences. This last remark might have an interesting physiological interpretation, since it seems to suggest that the physiological mechanisms responsible for the scale invariance operate on the same time scales to control both the amplitude and the pitch of the voice.

We checked that the scaling range $j \in [2, 5]$ was a good choice in all the records in the database. Therefore, we adopted it for the estimation procedures used to obtain the results that follow.

4.2. Evidence of multifractal behavior

Having studied in Section 4.1 the evidence of scaling behavior in AS and PS, and having showed that both sequences are amenable to the application of the multifractal formalism, we proceed in this section to answer the question of whether multifractal analysis is a relevant tool to model and/or analyze this kind of data.

Fig. 4 shows $\zeta(q)$ and $D(h)$ for the AS (top row) and PS (bottom row) of a selected healthy subject. Red lines indicate bootstrap-based 95% confidence intervals. It can be easily seen that $\zeta(q)$ deviates from the linear behavior of an "equivalent" monofractal process (that is, with $H = c_1$), illustrated with blue dashed lines. In both cases $\zeta(q)$ shows a downward concavity characteristic of multifractal processes. Equivalently, $D(h)$ shows a wide support that does not collapse to a single point, as indicated by the confidence intervals of the extreme points which, despite being wide, do not overlap. Blue crosses show the spectra for the equivalent monofractal process. The differences between the "∩-shaped" estimated spectra and the monofractal one are evident. These results allow to see how a monofractal process would fail to capture the full complexity of AS and PS for this speaker, and suggest that the analysis should be based on the multifractal paradigm instead.

Fig. 5 reports the results of the estimation of multifractal attributes on all healthy subjects in the database. It shows boxplots of c_1 (left) and c_2 (right) for AS (top) and PS (bottom), separately for male subjects, female subjects and all subjects grouped together. Points outside the 99.3% normal coverage were considered outliers, and are reported with red '+' signs. Table 2 summarizes the statistics of those estimations. It can be seen that boxes for c_2 are all below the value $c_2 = 0$, suggesting that, in fact, $c_2 \neq 0$ for both AS and PS, and therefore multifractal models should be preferred over monofractal ones to study these sequences. To further support this claim, we performed, for all cases, Wilcoxon sign-rank tests with the null hypothesis $H_0 : c_2 = 0$ against the double sided alternative $H_1 : c_2 \neq 0$. This nonparametric procedure tests the null hypothesis that the sample comes from a population with a symmetric distribution with a given median. This test was chosen because an elementary analysis of probability plots, not reported here, indicated that the distributions of the estimators were not normal. The results of the tests are reported in Table 3, and clearly state that, with a high confidence, both AS and PS show a multifractal behavior, for both male and female subjects, as well as for all subjects grouped together.

Further analysis of Fig. 5 and Table 2 reveals that c_1 and c_2 take on remarkably similar values for AS and PS. Table 4 shows the p-values of Wilcoxon sign-rank tests

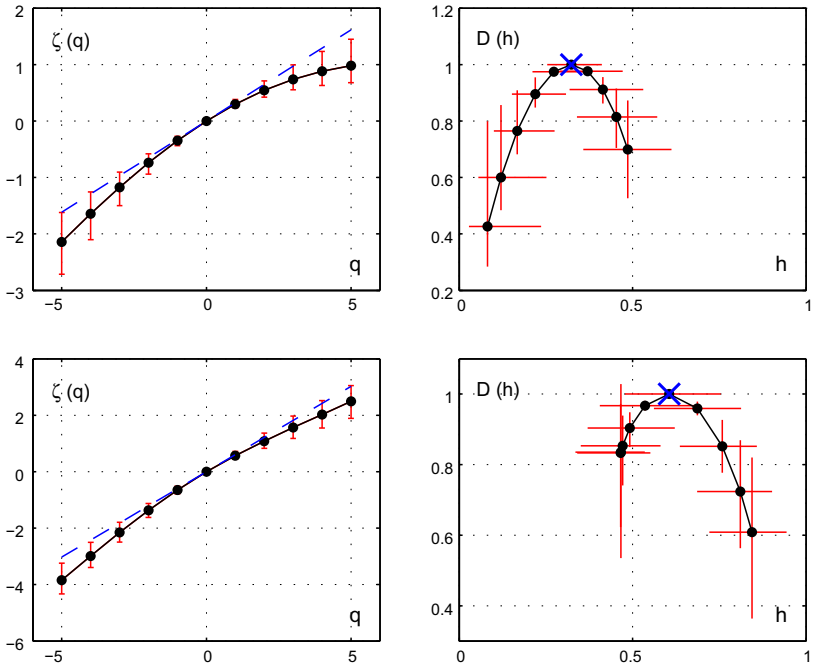

Fig. 4. Scaling exponents and multifractal spectrum. $\zeta(q)$ (left column) and $D(h)$ (right column) corresponding to the AS (top row) and PS (bottom row) of a healthy subject. Red lines indicate bootstrap-based 95% confidence intervals. Blue dashed line and cross indicate $\zeta(q)$ and $D(h)$ of an "equivalent" monofractal process where $H = c_1$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

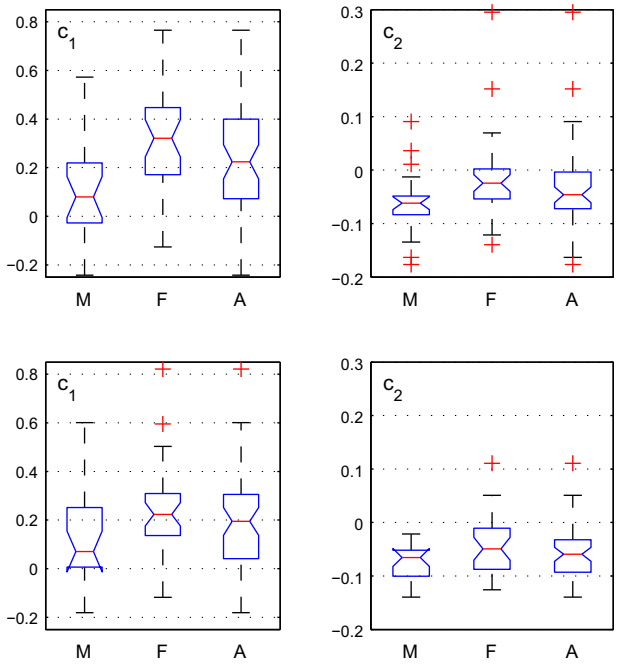

Fig. 5. Healthy subjects: log-cumulant estimations. Boxplots of the estimations of c_1 (left column) and c_2 (right column) for AS (top row) and PS (bottom row), for all healthy subjects of the database. Results are grouped as male (M), female (F) and all (A) subjects. Red '+' signs indicate points that where considered as outliers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Healthy subjects: statistics. Statistics for estimates of c_1 and c_2, for male (M) and female (F) subjects.

			Mean	Std	Min	Max
M	AS	c_1	0.0917	0.204	−0.242	0.572
		c_2	−0.0609	0.0607	−0.176	0.0907
	PS	c_1	0.127	0.193	−0.181	0.600
		c_2	−0.0770	0.0348	−0.140	−0.0215
F	AS	c_1	0.316	0.204	−0.126	0.766
		c_2	−0.0141	0.0807	−0.139	0.295
	PS	c_1	0.253	0.187	−0.118	0.821
		c_2	−0.0480	0.0618	−0.217	0.111

Table 3
Healthy subjects: tests. p-values for Wilcoxon sign-rank tests with the null hypothesis $H_0 : c_2 = 0$ against the double sided alternative $H_1 : c_2 \neq 0$, for male (M), female (F), and all (A) subjects.

	M	F	A
AS	9.02E−4	3.79E−2	1.65E−4
PS	5.96E−5	2.66E−4	3.76E−8

with the null hypothesis $H_0 : c_p^{AS} - c_p^{PS} = 0$ against the two sided alternative $H_0 : c_p^{AS} - c_p^{PS} \neq 0$, with $p = 1, 2$. The tests fail to reject, at a 5% confidence level, the null hypothesis for c_1 on both genders. However, the null hypothesis is rejected for c_2 on female subjects. These results indicate that the dominant regularity, as measured by c_1 and by classical monofractal models, is similar on both sequences. On the contrary, the *multifractality*, measured by c_2, seems to be different for the sequences of female subjects. This evidence seems to suggest that the dominant scale-invariant physiological phenomena that control the vocal folds have the same manifestation in amplitude and period fluctuations, but that secondary contributions might be different. This result emphasizes the need to replace monofractal models by a multifractal analysis that can provide a full characterization of the singular behavior of data.

Table 4
Healthy subjects: tests. p-values for the double sided sign-rank test with $H_0 : c_p^{AS} - c_p^{PS} = 0$, where $p = 1, 2$.

	M	F
c_1	0.958	0.304
c_2	0.614	0.0261

Moreover, both c_1 and c_2 seem to have larger values for females than for males, suggesting that, in the studied database, female voices are more regular and have less variability than male ones. To formalize this claim, we performed Wilcoxon rank-sum tests to assess whether there was a statistically significant difference between multifractal attributes of male and female subjects. This non-parametric procedure is used with two related variables to test the null hypothesis that they have the same distribution. Results are shown in Table 5. It can be seen that in all cases the null hypothesis is rejected at a 5% confidence level, confirming that the multifractal behavior of both sequences is dependent on the gender of the speaker. Moreover, the p-value for c_1 computed from the amplitude sequences is remarkably small, suggesting that this multifractal attribute might be particularly useful for the discrimination between male and female voices.

Since all analyzed groups are homogeneous in age, cf. Table 1, no dependence on age has been observed in the results. However, it would be interesting to perform the analysis on a larger database, with a wider age range, to asses whether regularity of AS and PS remains constant with age or not.

Results in this section clearly state that multifractal paradigm should be preferred over the monofractal one for modelling AS and PS. This would allow, for instance, the synthesis of such sequences for their use in voice synthesizers. We hypothesize that the use of multifractal sequences, with changes in regularity over time, would

Table 5
Healthy subjects: tests. p-values for the double sided rank-sum test with $H_0 : c_p^{M} = c_p^{F}$, where $p = 1, 2$.

	c_1	c_2
AS	4.33E−4	1.24E−2
PS	2.84E−2	3.57E−2

allow the synthesis of more natural-sounding voices. Also, from the analysis perspective, multifractal analysis provides a collection of parameters $\{\zeta(q), c_p, h_m\}$ that could be used for diagnostic purposes. This issue is further explored in Section 4.3.

4.3. Application to nonhealthy voices

In this section, we provide some preliminary results intended to illustrate the possibilities that multifractal analysis could bring to the automatic detection of pathologies from voice records.

We studied records from healthy subjects as well as from patients suffering from different pathological conditions, as described in Section 3. For the sake of space we only report results for PS of female subjects, but we have verified a similar performance for male subjects and both sequences. The database included 32 healthy and 86 nonhealthy female speakers.

Fig. 6 shows the scaling exponents (left) and the multifractal spectra (right) computed from a selected healthy record (record DFP1NAL, blue crosses) and a selected nonhealthy record (record PLW14AN, red circles), each one representative of its class. In both cases, estimations were performed using the same scaling range $j \in [3, 6]$, and a pseudo fractional integration of order $\gamma = 1$. Both spectra are clearly different, with the healthy spectrum being slightly narrower and shifted towards higher regularity than the nonhealthy case. Moreover, the nonhealthy spectrum's support is completely included in $(-\infty, 0]$, indicating that all singularities have *negative regularity exponents*. On the contrary, all singularities for the healthy subject have positive regularity exponents only. This behavior was consistently observed in all nonhealthy subjects that we considered, and is analyzed in greater depth later on.

In order to study the classification power provided by multifractal features, Fig. 7 shows c_1 vs c_2 (left), c_1 vs h_m (middle) and c_2 vs h_m (right) for healthy (blue crosses) and nonhealthy (red circles) speakers. Two clusters, corresponding to each one of the categories can be easily observed in all cases. In particular, attributes c_1 and h_m seem to provide the greatest separation between the classes (middle plot). On the contrary, c_2 does not seem to provide useful information and seems to be unable to distinguish between the two classes; this is not surprising considering that the estimation of c_2 is known to be difficult on short signals.

To further assess whether there exist statistically significant differences between the multifractal indices for healthy and nonhealthy records, we have used the Wilcoxon rank-sum test. Results are shown in Table 6. Clearly, all three indices show significant differences, meaning that the presence of the pathology under analysis

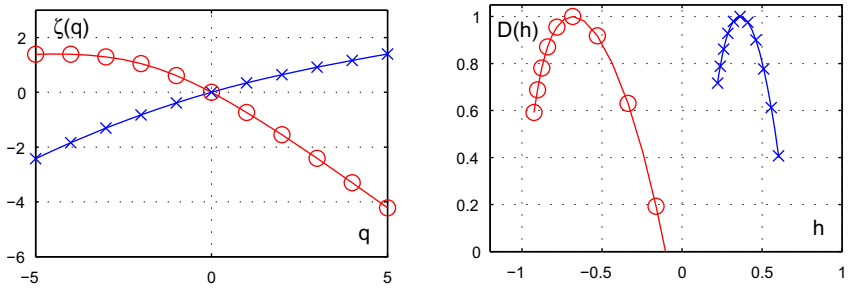

Fig. 6. Healthy and nonhealthy subjects. Scaling functions $\zeta(q)$ (left) and multifractal spectra $D(h)$ (right) for healthy (blue crosses) and nonhealthy (red circles) PS for a selected representative speaker. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

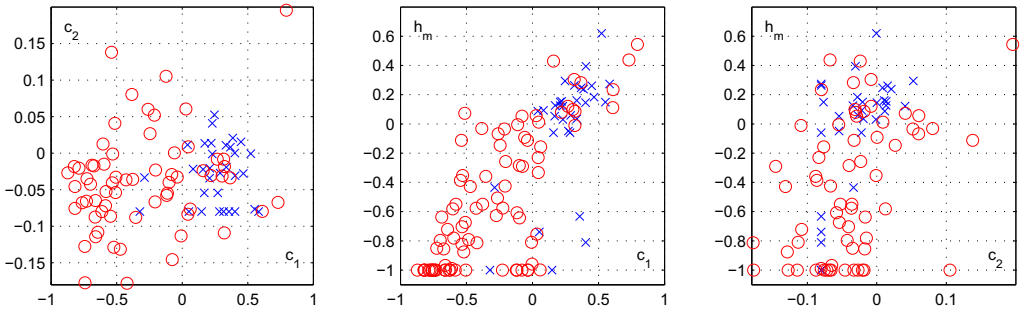

Fig. 7. Healthy and nonhealthy subjects: multifractal attributes. c_1 vs c_2 (left), c_1 vs h_m (middle) and c_2 vs h_m (right) for healthy speakers (blue crosses) and non healthy speakers (red circles). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
Healthy and nonhealthy subjects: tests. p-values for the non-parametric Wilcoxon rank-sum tests with the null hypothesis $H_0 : \theta^H = \theta^{NH}$ against the alternative $H_1 : \theta^H \neq \theta^{NH}$, for $\theta \in \{c_1, c_2, h_m\}$.

	c_1	c_2	h_m
p	2.03E−10	1.99E−2	3.80E−7

modifies the distribution of all attributes. c_2 appears to be the least significant attribute, in concordance with the behavior we have observed in Fig. 7.

Furthermore, to evaluate the possibility of classifying between healthy and nonhealthy subjects, we performed linear discriminant analysis (Bishop, 2006) on the triplet (c_1, c_2, h_m). This procedure generates a linear discriminant function, using a linear combination of the analyzed variables, that provides the best discrimination between the two groups. First, this discriminant function is obtained from a training sample of cases for which the membership to each group is known. Then, it is applied to new samples of unknown membership. We decided to use a simple classifier because we wanted to assess the power of the multifractal features for classification. The use of more sophisticated classification schemes, in combination with other features, will be considered in a forthcoming work. Moreover, we used the classical leave-one-out method to perform the validation of the classifier: each one of the available records was used as a validation sample, and for each case the remaining records were used as the training set. In other words, each record was classified with a classifier that was trained using all the other available records. Results are reported in Table 7. It can be seen that an excellent performance is achieved in the classification of healthy subjects, where almost all cases are correctly considered to belong to the proper group. In the case of the pathological group, one quarter of the nonhealthy samples were incorrectly considered to be healthy. Further analysis remains to be done to determine why these subjects show such similar dynamics with respect to healthy ones. Table 7 also reports several performance metrics computed from the confusion matrix, further indicating the overall acceptable performance of the classification procedure.

To assess the benefits provided by multifractal features to discrimination between healthy and nonhealthy speakers, we also performed the classification under a monofractal model setting. We used the Hurst exponent H, estimated with Eq. (6), as a feature. Results are reported in Table 8. Comparing with Table 7, it can be seen that the use of features from a multifractal model improves the number of true positives and negatives around a 10%, with similar gains in sensitivity, specificity and accuracy, and a 5% improvement in precision.

From this preliminary analysis we can conclude that the information provided by multifractal analysis shows great promise for the automatic detection of some voice pathologies: an acceptable performance is achieved using only three parameters and a simple classifier. Moreover, classification performance is significantly improved with respect to the one achieved using just a monofractal model. It remains to be studied if the addition of multifractal attributes improves state-of-the-art classification schemes, such as the one proposed in Arias-Londono et al. (2011).

Results in Fig. 7 also suggest that h_m is negative for nonhealthy records and positive for healthy ones. This distinction is very important from a modelling point of view. This amounts to the fact that healthy sequences would need to be modelled by a *function* whereas nonhealthy ones would need to be modelled by a *measure* (Jaffard et al., 2007; Arneodo et al., 1998). For instance, considering the

Table 7
Healthy and nonhealthy subjects: multifractal classification. Discriminatory analysis carried out using all variables c_1, c_2 and h_m, for healthy (H) and nonhealthy (NH) subjects.

	Group	Predicted group		
		H	NH	
Count	H	30	2	Sensitivity: 0.756
	NH	21	65	Specificity: 0.938
%	H	93.75	6.250	Precision: 0.970
	NH	22.1	77.90	Accuracy: 0.805

Table 8
Healthy and nonhealthy subjects: monofractal classification. Discriminatory analysis carried out using the Hurst exponent H, for healthy (H) and nonhealthy (NH) subjects.

	Group	Predicted group		
		H	NH	
Count	H	27	5	Sensitivity: 0.686
	NH	27	59	Specificity: 0.844
%	H	84.38	15.62	Precision: 0.922
	NH	31.40	68.68	Accuracy: 0.729

Table 9
Healthy and nonhealthy subjects: tests. p-values for sign tests with $H_0 : h_m \geqslant 0$ against $H_1 : h_m < 0$.

	H	NH
p	0.996	3.54E−10

monofractal and Gaussian setting that was used in Aoki and Ifukube (1999), this would mean that healthy sequences would need to be modelled by fBm (as in Aoki and Ifukube (1999)) and nonhealthy ones would need to be modelled by fractional Gaussian noise (fGn). To further analyze this point, we conducted a sign test on the values of h_m for both classes, with the null hypothesis $H_0 : h_m \geqslant 0$ against the alternative $H_1 : h_m < 0$. p-values are shown in Table 9. They suggest that, in fact, the data support the hypothesis that $h_m < 0$ for nonhealthy records and $h_m > 0$ for healthy ones. This result provides valuable insight for the selection of the multifractal models that should be used on each case.

5. Conclusions

In the present contribution we have proposed the use of wavelet leader multifractal analysis to study voice period and amplitude sequences obtained from real sustained vowels. We have shown that both sequences indeed show scaling, and that they have a multifractal behavior. This expands on previous work where analysis was limited to *monofractal* models, where a single parameter is enough to characterize the regularity of the sample path. Instead, we have successfully shown that both sequences should be considered on the light of *multifractal* models. This provides a richer description of the sample's regularity in terms of a set of parameters, and therefore allows a better characterization.

This change of framework could provide potential benefits to the modelling scenario. It is known that the characteristics of both sequences are related to the quality of synthesized voices. Therefore, the use of multifractal models that accurately describe the changes in regularity of the data could enhance the quality of synthesizers. This line is currently under investigation.

We have also illustrated, by means of an example, that the information provided by multifractal analysis could be extremely useful for the discrimination between healthy and nonhealthy voices. In fact, we showed that three multifractal attributes exhibited significant differences between the two classes. The preliminary analysis reported here needs to be continued by a more thorough study, and this is also currently under investigation.

We must mention, however, that there is an important caveat that limits the impact that this methodology could have: the data length. Amplitude and period sequences are typically short, since the length is limited by the amount of time the speaker can sustain the phonation of a vowel. The situation is even worse in the case of nonhealthy voices. This limitation derives in estimators with a large variance and, in consequence, in a decrease of discrimination power.

Acknowledgements

This work was supported by the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) under Grant PICT-2012-2954 and Universidad Nacional de Entre Ríos under Grant PID-UNER-6136, Argentina.

References

Abry, P., Flandrin, P., Taqqu, M., Veitch, D., 2000. Wavelets for the analysis, estimation, and synthesis of scaling data. In: Park, K., Willinger, W. (Eds.), Self-Similar Network Traffic and Performance Evaluation. Wiley, pp. 39–88.

Abry, P., Flandrin, P., Taqqu, M.S., Veitch, D., 2003. Self-similarity and long-range dependence through the wavelet lens. Theory and Applications of Long-range Dependence, 527–556.

Abry, P., Jaffard, S., Wendt, H., 2013. When Van Gogh meets Mandelbrot: multifractal classification of painting's texture. Signal Process. 93, 554–572.

Alzamendi, G.A., Schlotthauer, G., Rufiner, H.L., Torres, M.E., 2013. Evaluation of a new model for vowels synthesis with perturbations in acoustic parameters. Lat. Am. Appl. Res. 43, 225–230.

Alzamendi, G.A., Schlotthauer, G., Torres, M.E., 2015. State space approach to structural representation of perturbed pitch period sequences in voice signals. J. Voice. http://dx.doi.org/10.1016/j.jvoice.2014.11.007, ISSN: 0892-1997.

Aoki, N., Ifukube, T., 1999. Analysis and perception of spectral $1/f$ characteristics of amplitude and period fluctuations in normal sustained vowels. J. Acoust. Soc. Am. 106, 423–433.

Arias-Londono, J.D., Godino-Llorente, J.I., Sáenz-Lechón, N., Osma-Ruiz, V., Castellanos-Dominguez, G., 2011. Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients. IEEE Trans. Biomed. Eng. 58, 370–379.

Arneodo, A., Bacry, E., Muzy, J.F., 1998. Random cascades on wavelet dyadic trees. J. Math. Phys. 39, 4142–4164.

Baken, R.J., Orlikoff, R.F., 2000. Clinical Measurement of Speech and Voice. Singular Thomson Learning, San Diego, USA.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer.

Boersma, P., 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In: Proceedings of the Institute of Phonetic Sciences, Amsterdam, pp. 97–110.

Boersma, P., 2009. Should jitter be measured by peak picking or by waveform matching?. Folia Phoniatr. Logo. 61 305–308.

Boersma, P., Weenink, D., 2013. Praat: Doing Phonetics by Computer. <http://www.praat.org/>.

Bonilha, H.S., Deliyski, D.D., 2008. Period and glottal width irregularities in vocally normal speakers. J. Voice 22, 699–708.

Cabral, J.P., Oliveira, L.C., 2006. Emovoice: a system to generate emotions in speech. In: INTERSPEECH 2006, pp. 1798–1801.

Calvet, L., Fisher, A., 2001. Forecasting multifractal volatility. J. Econom. 105, 27–58.

Ciuciu, P., Varoquaux, G., Abry, P., Sadaghiani, S., Kleinschmidt, A., 2012. Scale-free and multifractal time dynamics of fMRI signals during rest and task. Front. Physiol. 3.

Dejonckere, P., Schoentgen, J., Giordano, A., Fraj, S., Bocchi, L., Manfredi, C., 2011. Validity of jitter measures in non-quasi-periodic voices. Part I: Perceptual and computer performances in cycle pattern recognition. Logopedics Phoniatr. Vocol. 36, 70–77.

Delour, J., Muzy, J., Arneodo, A., 2001. Intermittency of 1D velocity spatial profiles in turbulence: a magnitude cumulant analysis. Euro. Phys. J. B 23, 243–248.

Doret, M., Helgason, H., Abry, P., Goncalves, P., Gharib, C., Gaucherand, P., 2011. Multifractal analysis of fetal heart rate variability in fetuses with and without severe acidosis during labor. Am. J. Perinatol. 28, 259.

Endo, Y., Kasuya, H., 1996. A stochastic model of fundamental period perturbation and its application to perception of pathological voice quality. In: Proc. of Fourth Int. Conference on Spoken Language ICSLP, 1996, pp. 772–775.

Farrus, M., Hernando, J., 2009. Using jitter and shimmer in speaker verification. IET Signal Proc. 3, 247–257.

Fraile, R., Kob, M., Godino-Llorente, J.I., Sáenz-Lechón, N., Osma-Ruiz, V.J., Gutiérrez-Arriola, J.M., 2012. Physical simulation of laryngeal disorders using a multiple-mass vocal fold model. Biomed. Signal Process. Contr. 7, 65–78.

Fraj, S., Grenez, F., Schoentgen, J., 2009. Synthetic hoarse voices: a perceptual evaluation. In: Int. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications – MAVEBA, pp. 95–98.

Fraj, S., Schoentgen, J., Grenez, F., 2012. Development and perceptual assessment of a synthesizer of disordered voices. J. Acoust. Soc. Am. 132, 2603–2615.

Gerasimova, E., Audit, B., Roux, S.G., Khalil, A., Gileva, O., Argoul, F., Naimark, O., Arneodo, A., 2014. A wavelet-based method for multifractal analysis of medical signals: application to dynamic infrared thermograms of breast cancer. Nonlinear Dynamics of Electronic Systems. Springer, pp. 288–300.

Govind, D., Prasanna, S.R. Mahadeva, 2013. Expressive speech synthesis: a review. Int. J. Speech Technol. 16, 237–260.

Henríquez, P., Alonso, J.B., Ferrer, M.A., Travieso, C.M., Godino-Llorente, J.I., Díaz-de María, F., 2009. Characterization of healthy and pathological voice through measures based on nonlinear dynamics. IEEE Trans. Audio Speech Lang. Process. 17, 1186–1195.

Ivanov, P.C., Amaral, L.A.N., Goldberger, A.L., Havlin, S., Rosenblum, M.G., Struzik, Z.R., Stanley, H.E., 1999. Multifractality in human heartbeat dynamics. Nature 399, 461–465.

Jaffard, S., 2004. Wavelet techniques in multifractal analysis. In: Fractal geometry and applications: multifractals, probability and statistical mechanics, applications, Proceedings of Symposia in Pure Mathematics (AMS), vol. 72, pp. 91–132.

Jaffard, S., Lashermes, B., Abry, P., 2007. Wavelet leaders in multifractal analysis. In: Wavelet Analysis and Applications. Birkhäuser. Applied and Numerical Harmonic Analysis, pp. 201–246.

Jaffard, S., Abry, P., Roux, S., 2011. Function spaces vs. scaling functions: tools for image classification. Mathematical Image Processing. Springer, pp. 1–39.

Jaffard, S., Abry, P., Wendt, H., 2014. Irregularities and Scaling in Signal and Image Processing: Multifractal Analysis. World Scientific Publishing, pp. 1–79. arXiv:1210.0482v1 [math.FA].

Lashermes, B., Roux, S.G., Abry, P., Jaffard, S., 2008. Comprehensive multifractal analysis of turbulent velocity using the wavelet leaders. Eur. Phys. J. B – Condens. Matter Complex Syst. 61, 201–215.

Leonarduzzi, R.F., Schlotthauer, G., Torres, M.E., 2010. Wavelet leader based multifractal analysis of heart rate variability during myocardial ischaemia. Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE. IEEE, pp. 110–113.

Leong, K., Hawkshaw, M.J., Dentchev, D., Gupta, R., Lurie, D., Sataloff, R.T., 2013. Reliability of objective voice measures of normal speaking voices. J. Voice 27, 170–176.

Lovejoy, S., Schertzer, D., 2013. The Weather and Climate: Emergent Laws and Multifractal Cascades. Cambridge University Press.

Low, L.S., Maddage, M., Lech, M., Sheeber, L., Allen, N., 2010. Influence of acoustic low-level descriptors in the detection of clinical depression in adolescents. In: IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 5154–5157.

Low, L.S., Maddage, M., Lech, M., Sheeber, L., Allen, N., 2011. Detection of clinical depression in adolescents' speech during family interactions. IEEE Trans. Biomed. Eng. 58, 574–586.

Mallat, S., 2009. A Wavelet Tour of Signal Processing. third ed.. Academic Press, Cambridge.

Manfredi, C., Giordano, A., Schoentgen, J., Fraj, S., Bocchi, L., Dejonckere, P., 2011. Validity of jitter measures in non-quasi-periodic voices. Part II: The effect of noise. Logopedics Phoniatr. Vocol. 36, 78–89.

Markaki, M., Stylianou, Y., 2011. Voice pathology detection and discrimination based on modulation spectral features. IEEE Trans. Audio Speech Lang. Process. 19, 1938–1948.

Massachusetts Eye and Ear Infirmary MEEI Voice and Speech Lab, 2009. Disordered Voice Database, Model 4337. <http://www.kayelemetrics.com>.

Parsa, V., Jamieson, D.G., 2000. Identification of pathological voices using glottal noise measures. J. Speech Lang. Hear. Res. 43, 469–485.

Riedi, R., 2003. Multifractal processes. In: Doukhan, P., Oppenheim, G., Taqqu, M. (Eds.), Theory and Applications of Long Range Dependence. Birkhäuser, pp. 625–717.

Ruinskiy, D., Lavner, Y., 2008. Stochastic models of pitch jitter and amplitude shimmer for voice modification. In: Proc. IEEEI 2008, pp. 489–493.

Schlotthauer, G., Torres, M.E., Rufiner, H.L., 2010. Pathological voice analysis and classification based on empirical mode decomposition. In: Esposito, A., Campbell, N., Vogel, C., Hussain, A., Nijholt, A. (Eds.), Development of Multimodal Interfaces: Active Listening and Synchrony, . In: Lecture Notes in Computer Science, vol. 5967. Springer, Berlin Heidelberg, pp. 364–381.

Schoentgen, J., 2001. Stochastic models of jitter. J. Acoust. Soc. Am. 109, 1631–1650.

Schoentgen, J., De Guchteneere, R., 1995. Time series analysis of jitter. J. Phonet. 23, 189–201.

Schoentgen, J., De Guchteneere, R., 1997. Predictable and random components of jitter. Speech Commun. 21, 255–272.

Silva, D.G., Oliveira, L.C., Andrea, M., 2009. Jitter estimation algorithms for detection of pathological voices. EURASIP J. Adv. Signal Process. 2009, 1–9.

Titze, I.R., 1995. Workshop on Acoustic Voice Analysis: Summary Statement. Technical Report. National Center for Voice and Speech. Denver, USA.

Titze, I.R., 2000. Principles of Voice Production, second ed. National Center for Voice and Speech, Iowa, USA.

Titze, I., Liang, H., 1993. Comparison of f_0 extraction methods for high-precision voice perturbation measurements. J. Speech Hear. Res. 36, 1120–1133.

Velasco García, M.J., Cobeta, I., Martín, G., Alonso-Navarro, H., Jimenez-Jimenez, F.J., 2011. Acoustic analysis of voice in Huntington's disease patients. J. Voice 25, 208–217.

Wang, L., Li, A., Fang, Q., 2006. A method for decomposing and modeling jitter in expressive speech in chinese. In: Proc. of Speech Prosody.

Wendt, H., Abry, P., Jaffard, S., 2007. Bootstrap for empirical multifractal analysis. IEEE Signal Process. Mag. 24, 38–48.

Wendt, H., Roux, S., Jaffard, S., Abry, P., 2009. Wavelet leaders and bootstrap for multifractal analysis of images. Signal Process. 89, 1100–1114.

Zhang, Y., Jiang, J.J., Biazzo, L., Jorgensen, M., 2005. Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis. J. Voice 19, 519–528.