FRACTAL DIMENSION SEGMENTATION: ISOLATED SPEECH RECOGNITION

S. Fekkai, M. Al-Akaidi, J.Blackledge
Faculty of Computing Sciences & Engineering
De Montfort University,
The Gateway, Leicester,
LE1 9BH. UK.

Email: mma@dmu.ac.uk

Abstract:

This paper investigates the use of fractal geometry for segmenting digital signals. A method of texture segmentation is introduced which is based on the Fractal Dimension. Using this approach, variations in texture across a signal or image can be characterized in terms of variations in the fractal dimension.
By analyzing the spatial fluctuations in fractal dimension obtained using a conventional moving window approach, a digital signal or image can be texture segmented; this is the principle of Fractal Dimension Segmentation (FDS).
In this paper, we apply this form of texture segmentation to isolated speech signals.
An overview of methods for computing the fractal dimension is presented focusing on an approach that makes use of the characteristic Power Density Function (PSDF) of a Random Scaling Fractal Signal.
FDS is applied to a number of different speech signals and the results discussed for isolated words and the components (e.g. fricatives) from which these words are composed. In particular, it is shown that by pre-filtering speech signals with a low-pass filter of the form 1/k. This provides confidence in the approach to speech segmentation considered in this paper and in principle, allows a template matching scheme to be designed that is based exclusively on FDS.

Introduction:

Speech recognition introduces a new range of communication services that extend man's capabilities, serve his social needs and increase his productivity.
Isolated word recognition for example, is adequate for logging freight destinations in warehouses or identifying and counting items for inventory control. It also requires a short pause before and after utterances that are to be recognized as entity [1]. In other term the words are spoken in isolation. Pauses between words simplify recognition because they make it relatively easy to identify endpoints (i.e., the start and end of each word), and they minimize co-articulation effects between words. In addition, isolated words tend to be pronounced somewhat more carefully, since the need to pause between words impedes fluency, which would otherwise tend to encourage a more natural and hence more careless pronunciation.

Techniques of Computing Fractal Dimension:

There are many possible ways for computing the *fractal dimension* of speech signals.
In this paper four methods are used and cited as below:

The box counting method:

Box counting or box dimension is one of the most widely used dimensions. Its popularity is largely due to its relative ease of mathematical calculation and empirical estimation [2].
Box counting in general involves covering a fractal with a grid of n-dimensional boxes with side length δ and counting the number of non-empty boxes $N(\delta)$. If a speech signal with N elements (N being a power of 2) is used as input signal then the slope β obtained in a bilogarithmic plot of the number of boxes used against their size gives the fractal dimension D [3]. The box counting applied is a regular grid where successive divisions by a factor of 2 are used for the box size to give a regular spacing in the bilogarithmic plot. The fractal dimension is determined as follow:

$$D = -\frac{\ln N(\delta)}{\delta}$$

Where $N(\delta)$ is the number of non-empty boxes and δ the size of the box.
The word "zone "has been segmented into phonemes and the fractal dimension of each phoneme has been calculated using this method. The results obtained are shown below.

Phonemes	Fractal dimension
/z/.	1.1
/o/	1.02
/ne/	1.01

In general Box counting algorithms behave well and produce accurate estimates for fractal dimensions between 1 and 1.5 for digital signals and between 2 and 2.5 for digital images and are easy to code and fast to compute [4].

The Continuous Box Counting method (CBCM):

It is similar to the box counting method. The CBCM looks not only at the points within each column along the curve, as the dimension is calculated but also to the relationship between columns [5].
The continuous box counting should give a more accurate value for the fractal dimension compared with the previous cited method.
The fractal dimensions for each phoneme of the word "zone " computed using the CBCM are given in table below.

Phonemes	Fractal dimension
/z/.	1.3
/o/	1.15
/ne/	1.12

The Line Divider method:

In this method use of a chord length (*step*) and measures the number of chord lengths (*length*) needed to cover a fractal curve. This technique is based on the principle of taking smaller and smaller rulers of size *step* to cover the curve and counting the number of rulers length required in each case [3]. The process is repeated until there are enough points to reasonably find the line of the best fit for the relationship:

Log [total length] Vs Log [Step size]

The fractal dimension is then given by:

D = - (Log [total length] Vs Log [Step size])

The values of D obtained for the phonemes /z/,

/o/ and /ne/ of the word "zone" pronounced by a male speaker are given in table below.

Phonemes	Fractal dimension
/z/.	1.53
/o/	1.47
/ne/	1.34

The Power Spectrum method (PSM):

The PSM is implemented by applying the FFT to the speech signal in order to obtain a spectral representation of the phoneme. A pre-filter step is then used to adjust the estimated values of the fractal dimension to fit within the range 1 and 2. The power spectrum of the pre-filtered signal is computed and the least square approach is applied for the calculation of the power exponent β and the fractal dimension D of the used phoneme.
The least square method used to calculate the spectral exponent β yields to the following equation [6]:

$$\beta = \frac{N\sum_{i=1}^{N}(\ln P_i)(\ln|k_i|) - (\sum_{i=1}^{N}\ln P_i)(\sum_{i=1}^{N}\ln|k_i|)}{(\sum_{i=1}^{N}\ln|k_i|)^2 - N\sum_{i=1}^{N}(\ln|k_i|)^2} \qquad (1)$$

where P_i is the measured power spectrum of the speech signal and k_i its spatial frequency.
The relationship:

$$D = \frac{5-\beta}{2}$$

provides a simple formula for computing the fractal dimension from the power spectrum of a signal [7].
The results obtained for the same phonemes when we used the PSM are shown below.

Phonemes	Fractal dimension
/z/.	1.32
/o/	1.09
/ne/	1.1

What is Texture?

Texture is a word that is commonly used in a variety of contexts but is at best a qualitative description of a sensation. Visual texture can be associated with a wide range of scenes and images but the term cannot be taken to quantify any particular characteristic.
Typically one or more 'measures' for texture can be defined and a moving window (usually square) passed over a signal or an image [8].

Fractal dimension segmentation:

Within a predetermined window, the power spectrum algorithm has been applied to different phonemes of N elements. Then moving this window one element at a time over the waveform, a set of values for the fractal dimension D is obtained as a function of the window position. The values of D provide a characteristic profile of the fractal dimension, which, from the definition of fractals should vary between 1 and 2 for fractal speech. The data can therefore be segmented into regions of similarity, which depend on the fractal dimension [9].
The fractal dimension obtained for the fricative /f/ as pronounced by a male and a female speaker are shown in figure 1.

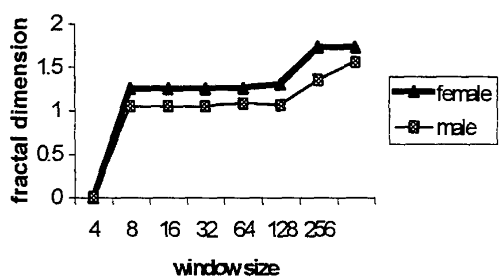


Figure 1: Fractal Segmentation

It is apparent from figure1 that the fractal dimensions obtained for the female speaker are higher than the male ones, which in term of irregularity means that the female speech signals possess a higher irregularity in their representation. However they still apply to the definition of fractal signals where the values of the fractal dimension lie within the range 1 and 2.

Random Scaling Fractals:

A random scaling fractal (RSF) is one that exhibits statistical self –affinity [10].
There are many approaches working with fractional differentials but they all rely on a generalization of results associated with differentials of integer order [11, 12].

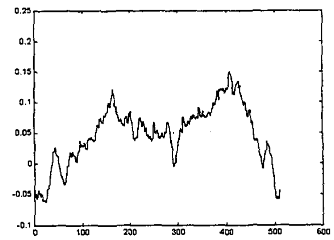
The relationship between white noise and fractal noise can be considered in term of fractional differential and fractional integration and is given by the stochastic fractional differential (SFD) equation:

$$\frac{d^{\beta}(x)}{d^{\beta}x} = n(x)$$

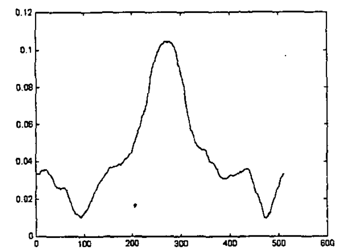
where $n(x)$ is a white Gaussian noise whose power density function (PSDF) is constant and $f(x)$ the fractal noise. This allow the forward problem given β find f and the inverse problem given f find β associated with fractal signals to be suited in term of stochastic fractional differential equation.

***Forward algorithm** (given β find f)*

Step1: compute a random white Gaussian noise $n_i \ i \in (1,2,3,4,....,N)$.
Step2: Fourier transform n_i to create $N_i(w)$.
Step3: Filter in complex space to get $F(w) = (iw)^{-\beta} \times N_i(w)$.
Step4: Inverse Fourier transform $F(w)$ to create $f(x) = \mathrm{Re}(F^{-1}\{F(w)\})$.
The results obtained for different values of β and D are plotted in figure 2.



D=1.9 ; β =1.2



D=1.4 ; $\beta = 1.8$

Figure 2: Fractal Signals

The fractal dimension in general measures the degree of irregularity of a signal or image. So that, the rougher the texture of an image or the shape of a speech waveform is, the higher the fractal dimension will be. In fact, figure 2 confirm that conclusion as the fractal signal for D=1.9 looks more irregular that the one for which D=1.4.

***Inverse algorithm** (given f find β)*

This algorithm is based on the least square method to estimate the value of the spectral component β. It uses the following four steps.
Step1: Compute the power spectrum of the fractal signal.
Step2: Extract the positive half space data (excluding the DC level).
Step3: Use the least square fit approach to estimate the spectral component β (Eq.1).
Step4: The relationship $\beta = 5 - 2D$, provides a non-iterative formulae for computing the fractal dimension from the power spectrum of a signal.

The values of the fractal dimension obtained are shown in figure 3 where D0 is the original (real) fractal dimension and D1 the calculated ones.

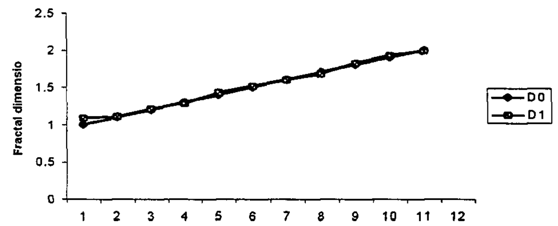


Figure 3: Fractal dimension evaluation

It can be seen from figure 3 that the two dimensions are matching perfectly well. That is, the power spectrum algorithm gives a very accurate estimation of the fractal dimension for random fractals.

Conclusion:

The application of the fractal dimension as a method of segmenting speech signals/image texture seems to be an adequate approach.
In fact the use of a moving window for fricative component (fig.1) is very useful in the process of template creation.
The RSF model is appropriate to measure the fractal dimension of a random white noise and the results obtained in this paper (fig. 3) are very satisfactory, compared to the real fractal dimension values.

Bibliography:

[1] Huang X., Y. Ariki and M. Jack, "Hidden markov models for speech recognition ", Edinburgh University Press, Edinburgh, UK, 1990.

[2] Falconer K., "fractal geometry: mathematical foundation and applications", John Wiley & sons Ltd, UK, 1990.

[3] Turner M.J., Blackledge J.M. and Andrew P.R. "Fractal geometry in imaging", Academic Press, San Fransisco, 1998.

[4] Al-Akaidi M., Blackledge J. M and Mikhailov S. "Random Fractal Coding Techniques", Euromedia 98, DeMontfort University, Leicester, Uk, pp.14-18.

[5] Boshoff H.F.V, " A Fast box counting algorithm for determining the fractal dimension of sampled continuous functions ", Proceedings, south african sympossium on communication and signals (compsig 92), capetown, south africa, pp.43-48, 1992.

[6] Fekkai S. and Al-Akaidi M. "A Novel approach to measure fractal dimension of speech phonemes" Euromedia, Belgium, 1999. (submitted for publication).

[7] Pickover C.A. and Khorassani A. " Fractal characterisation of speech waveform graphs", Comput. and Graphics, vol. 10, No1, pp. 51-61, 1986.

[8] Blackledge J.M. "On the Synthesis and Processing of Fractal Signals of Images", Springer-Verlag, Berlin, 1993.

[9] Rosenfeld A. and Troy E B, "Visual texture analysis", University of Maryland, technical report, TR-116, June 1970.

[10] Reuff M., " Scale space filtering and the scalling regions of fractals, in from pixels to features", J C simon (Ed.), North Holland, 49-60, 1989.

[11] Oldham K.B.and Spanier J, "The fractional calculus, Mathematics in Science and Engineering, Vol 3, Academic Press, 1974.

[12] Samko S.G., Kilbas A.A and Marichev O.I, "Fractional integrals and derivatives: theory and applications, Gordon and Breach scientific Publishers, 1993.