Fractal and Chaotic Characteristics of Persian Speech Signal

R. Fazel-Rezai[1], *Sh. Oveisgharan*[2]

[1]Department of Electrical Engineering, University of Manitoba, Canada
[2]Department of Electrical Engineering, Sharif University of Technology, Iran

ABSTRACT

A chaotic and a fractal measures were calculated for Persian speech signal and their performances in speech classification were evaluated. The first measure was correlation dimension of each frame in speech signal which is based on its chaotic characteristics. The second measure was fractal dimension computed by fitting Hosking's ARMA filtered FdGn model [1] to speech signal and computing its Hurst parameter by Tewfik's iterative Maximum Likelihood approach [2]. Experimental results showed that, in Persian speech, better classification were obtained by Hosking's model because its ability to characterize short term dependencies of speech signal which is interpretable by ARMA model.

I. INTRODUCTION

*I*n the most commonly used model of speech production, speech signal is decomposed into a time varying filter component and an excitation component [3]. The excitation is represented by the superposition of two sources; periodic pulse train produced by vibration of the vocal cords and white Gaussian noise produced by forcing air past some constriction in the vocal tract. But this model lacks of the ability to observe the long term dependencies observed in speech signal because of its *ARMA* statistical model [2]. In other words, it has been observed that by assumption of white noise excitation we cannot interpret long term dependencies in speech signal.

Because of the ability of fractal models to observe the long term dependencies in signal and also the nonlinear dynamics of its source via chaotic Lyapanov exponent, fractal models come to play a significant role from 10 years ago [4]. Results of experiments have shown that the Hausdorf dimension of speech signal is strictly larger than its geometric dimension, especially in voiced and unvoiced fricatives [5].

Correlation dimension was applied to estimate and evaluate chaotic characteristics of speech signal. Experimental results showed strong chaotic characteristics in Persian speech signal. However this model seems to just observe the long term dependencies in speech signal and it cannot discover short term dependencies of speech signal which was found out by poles and zeros of ARMA model. Therefore, the existence of a model to catch both the long and short dependencies in speech signals appears to be necessary.

To solve this problem, Fractionally Brownian Motion *(FBM)* model is introduced by Mandelbrot and Ness [6]. In contrast with ARMA models which are characterized by correlation function that decay exponentially with the lag, FBM signals with $1/f$-type spectra have a correlation function that decreases hyperbolically fast with the lag k as k^{α} [2].

Because FBM is a non-stationary model, its derivative called Fractionally differenced Gaussian noise (FdGn) is applied to speech. But the FdGn model seems not to discover the short term dependencies of speech signal which could be found out by poles and zeros of ARMA model. Hosking [1]solved this problem by presenting the FdGn AR filtered model. In this model, the Gaussian excitation source is not assumed to be white at all, but it can also have weak dependencies between far samples. In this way, we also applied Hosking's model to speech signal in order to catch its short term dependencies. In order to compute model's parameters, we used the iterative method of Tewfik [2] which searches for a local optimal in log likelihood surface of model parameters.

Finally, we evaluate these two approaches in Persian speech classification and experimental results showed better classification of speech sounds by ARMA filtered FdGn Model than Correlation Fractal Dimension approach.

This paper is arranged as follows. In section II, we review correlation dimension. ARMA filtered FdGn model is discussed in section III. In section IV, the experimental results of applying these two approaches on Persian speech signal are explained.

II. CORRELATION DIMENSION

Turbulence in air pressure yields to a chaotic behavior and strange attractor trajectory of speech signal. One way to investigate the chaotic characteristics of vocal tract is to observe the trajectory of the system state variables. As we know in the case of stable systems the trajectory converges to a special point in state space, while in unstable systems the trajectory diverges to infinity and in quasi periodic systems the trajectory converges to a cycle. But in the case of a strange attractor the trajectory moves in a closed space

and doesn't converge to any point or cycle. Therefore, the trajectory seems to cover a thick and complex curve in space which means its Hausdorf dimension should be strictly greater than its geometric dimension. This system is fractal by Hausdorf's definition. So we can evaluate the characteristics of the chaotic system by evaluation of its trajectory's fractal dimension. But as we know, our input signal is a one dimensional signal which is yielded by a projection of trajectory curve and it is insufficient to get the trajectory curve. However, Takens' theorem [5]says that one dimensional time series can characterize a strange attractor having a degree of freedom F>1. The main requirement is that the time series must be sufficiently long according to the F (geometric dimension of state space).

The technique is based on construction of the state vector X_i using the elements of the given time series. Given a time series x_i, we can construct X_i according to

$$X_i = (x_i, x_{i+J}, ..., x_{i+(m-1)J}) \qquad (2.1)$$

where J is called lag or construction delay, and m is called embedding dimension. Takens suggested to use m>2F, however this estimation works for even smaller values [5]. However since state space variables are minimum number of variables describing a system, the value of J should be increased if cross correlation between entities of X_i is high. Having the vector X_i, we can find fractal dimension for any value of *m*. If the value of *m* is taken correctly, the value of computed fractal dimension is less than its embedding dimension (*m*). But in other cases the computed fractal dimension is greater than embedding dimension and so the value of *m*, embedding dimension, should be increased.

Now we present the correlation dimension of a trajectory. Let X_i be *i'th* state space vector, then we define correlation sum as follows:

$$C(r) = \lim \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \theta\left(r - \left|X_i - X_j\right|\right) \qquad (2.2)$$

where || is defined as the general norm function of vectors and function θ is defined as:

$$\theta(x) = \begin{cases} 1 \Leftarrow x > 0, \\ 0 \Leftarrow x \le 0, \end{cases} \qquad (2.3)$$

So $C(r)$ counts the number of vector pairs which have a distance less than *r*. As *r* decreases, the value of $C(r)$ is also decreased. But the behavior of $C(r)$ can be studied by using a quality:

$$C_v = \lim_{r \to 0} (C(r) r^{-v}) \qquad (2.4)$$

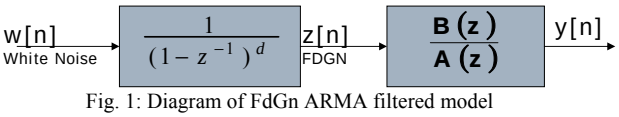

Fig. 1: Diagram of FdGn ARMA filtered model

If the value of *v* is chosen large, C_v tends to infinity, and if the value of *v* is chosen small, $C(r)r^{-v}$ tends to zero. However there exists a critical value of $v = D_C$ which we have:

$$C_v = \begin{cases} 0, & if\ v < D_c \\ const., & if\ v = D_C \\ \infty, & if\ v > D_c \end{cases} \qquad (2.5)$$

D_C is defined as correlation dimension and could be obtained using:

$$D_c = \lim_{r \to 0} \frac{\log(C(r))}{\log(r)} \qquad (2.6)$$

So in the case of a highly correlated signal, distance between vector pairs is small and hence we expect $C(r)$ to decrease slowly as *r* decreases and we obtain a small value of fractal dimension. But when the signal, the vector difference is not small and so we expect $C(r)$ to decrease faster as *r* decreases and thus we get a greater value of D_C. So according to this theorem, we should obtain a larger correlation fractal dimension for fricatives with more unvoiced characteristics.

Here, we applied an AR filter of degree *10* before computation of correlation dimension in order to eliminate the effect of linear part of vocal tract filter. For evaluation of the coefficients of AR filter, the **Levinson**'s Minimum Mean Square algorithm is used. In this algorithm a filter in form of

$$H(z) = \frac{1}{\sum_{i=1}^{K} a_i z^{-i}} \qquad (2.7)$$

is fitted to the input speech signal so that the power of the signal which is obtained by inverse filtering through $1/H(z)$ is set to minimum value.

III. FdGn ARMA FILTERED MODEL

III.A. Mathematical Background:

As shown in figure 1, the output of an ARMA filtered FdGn model is built of a composition of ARMA filtering and an FdGn filtering on white noise. The FdGn process can be defined as $(-d)^{th}$ fractional difference (or summation) of discrete time white Gaussian noise $(-0.5 \le d \le 0.5)$:

$$w_d[n] = \sum_{k=0}^{\infty} \binom{-d}{k} (-1)^k w[n-k] = \sum_{k=0}^{\infty} c[k] w[n-k] \qquad (3.1)$$

where $w[n]$ is white Gaussian noise. The continuous form of FdGn also can be defined as derivative of FBM:

$$x_{FdGn}(t) = \frac{d}{dt} x_{FBM}(t) \tag{3.2}$$

where FBM is defined as a zero mean Gaussian process with the following correlation property :

$$B_H(0) = 0, E\left\{\left(B_H(t) - B_H(s)\right)^2\right\} = \alpha(t-s)^{2H} \tag{3.3}$$

H is the Hurst parameter and is related to d with $d = H - 0.5$. In the case of $H = 0.5$ general Brownian motion would be obtained. Equation (3.3) implies the value of variance and autocorrelation matrix of FBM process:

$$\begin{aligned} \sigma^2_{B_H(t)} &= E\left\{\left(B_H(t) - B_H(0)\right)^2\right\} = \alpha t^{2H} \\ R_H(t,s) &= \frac{\alpha}{2}\left((t-s)^{2H} - t^{2H} - s^{2H}\right) \end{aligned} \tag{3.4}$$

According to figure 1, we can conclude that FdGn is a zero mean Gaussian Process with correlation function of

$$\begin{aligned} R_{w_H}(k) &= \frac{\sigma^2}{2}\left(|k+1|^{2H} + |k-1|^{2H} - 2|k|^{2H}\right) \\ &\cong C|k|^{2H-2} \end{aligned} \tag{3.5}$$

Note that approximation of equation 3.5 is true just for large values of *k*. The Power Spectrum of FdGn is also computable using figure 1 as follows:

$$\begin{aligned} S_{w_d}(\omega) &= \sigma_w^2 \left|H(f)\right|^2 = \sigma_w^2 \left|\left(1 - e^{-j2\pi f}\right)^{-d}\right|^2 \\ &= \frac{2^{-2d}\sigma_w^2}{\left(Sin\left(\frac{\omega}{2}\right)\right)^{2d}} \cong C|f|^{2H-2} \end{aligned} \tag{3.6}$$

As we observe in equation (3.5), correlation function of FdGn process decays hyperbolically and slowly, hence we can pursue the long term dependencies of any signal by this model. In other way, the ARMA Power Spectrum function is a summation of expressions in the form of $\frac{1}{1 - ae^{-j\omega}} = \sum_{n=0}^{\infty} a^n e^{-jn\varpi}$. So correlation function of an ARMA model decays exponentially and so fast and is unable to pursue long term dependencies of the signal.

In [6][2], Tewfik explained two iterative methods for estimation of parameters of ARMA filtered FdGn model: first algorithm is an EM (Expectation Maximum) algorithm which is based on an iterative method to find the Maximum Likelihood point of the probability space and the second algorithm is based on an iterative approximation method. The EM approach contains a more vast computational complexity and is not suitable for a long frame signal like speech therefore we just review and implement the second iterative method.

Before following on the Tewfik's algorithm, we first derive the log likelihood function of an observed data based on the Hurst parameter and variance of a simple FdGn process. After that we will discuss the problem of Maximum Likelihood Estimation (MLE) of general FdGn process according to observations and finally we will apply algorithm to problem of FdGn ARMA filtered parameters estimation.

Notice that according to the Gaussian property of FdGn, probability density function of an *N* member observed vector *x* of samples of an FdGn process is given by:

$$\begin{aligned} p(x;d,\sigma^2) &= \frac{1}{(2\pi)^{N/2}\left|R(d,\sigma^2)\right|^{1/2}} \cdot \\ &\cdot \exp\left\{-\frac{1}{2}x^T R^{-1}(d,\sigma^2)x\right\} \end{aligned} \tag{3.7}$$

where R can be obtained using equation 3.5. Now we define:

$$R(d) = \sigma^2 R_1(d) = \sigma^2 r(0) R_2(d) \tag{3.8}$$

Since the logarithm is a monotone function, so for maximum likelihood point search, it is sufficient to find the maximum logarithm likelihood point. So we have:

$$\begin{aligned} &L(x;d,\sigma^2) = \log(p(x;d,\sigma^2)) = \\ &-\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \\ &\frac{1}{2}\log(|R_1(d)|) - \frac{1}{2\sigma^2}x^T R_1^{-1}(d)x \end{aligned} \tag{3.9}$$

Now we explain Tewfik's algorithm [8]for obtaining log likelihood value by a complexity of $O(N^2)$.

III.B. Tewfik's Maximum Likelihood Estimator (MLE) of σ^2 and d of General FdGn

ML estimators are asymptotically unbiased and efficient. However in most cases it is difficult to find a closed form for the MLE in terms of a given data. Instead numerical methods such as Newton's method or Maximum Descent Algorithm should be applied to find the value of parameters which maximize the likelihood function [8]. By derivation with respect to sigma from equation (3.9), we obtain the value of variance for the optimum point as follows [8]:

$$\hat{\sigma}^2 = \frac{1}{N} x^T R_1^{-1}(d) x \tag{3.10}$$

Applying equation (3.10) into equation (3.9) yields equation (3.11) and (3.12) which should be maximized:

$$L_1(x;d) = -\frac{N}{2}\log(x^T R_1^{-1}(d)x) - \frac{1}{2}\log(|R_1(d)|) \tag{3.11}$$

$$L_2(x;d) = -\frac{N}{2}\log(x^T R_2^{-1}(d)x) - \frac{1}{2}\log(|R_2(d)|) \tag{3.12}$$

Now since the correlation matrices R_1 and R_2 are symmetric Toeplitz matrix, we can decompose them with respect to Levinson's algorithm and parameters of LPC for the coefficients of autocorrelation function which is defined as:

$$a_k(k) = \frac{-d}{k-d}, \qquad (3.13)$$
$$0 < i < k: \quad a_k(i) = \binom{k}{i} \frac{(i-d-1)!(k-d-i)!}{(-d-1)!(k-d)!}$$

For detailed proof of obtaining these values from Levinson's algorithm you can refer to [8].

Therefore, R_1 can be decomposed as $R_1^{-1}(d) = r(0)A^T P^{-1} A$ where A is a lower triangular matrix defined as

$$A(i,j) = \begin{cases} 1, & i = j \\ 0, & i < j \\ a_{i-1}(i-j), & oth.w. \end{cases} \qquad (3.14)$$

and P is in the form of a diagonal matrix $P = diag\{P_0, P_1, P_2, ..., P_{N-1}\}$ where P_k is defined as:

$$P_0 = 1,\ P_k = P_{k-1}\left(1 - \frac{d^2}{(k-d)^2}\right) \qquad (3.15)$$

So equation (3.12) could be computable according to equations (3.13) and (3.15):

$$\log\left(|R_2(d)|\right) = \log(P_0 P_1 \ldots P_{N-1}) = \qquad (3.16)$$
$$= \sum_{i=1}^{N-1} \sum_{j=1}^{i} \log\left(1 - \frac{d^2}{(d-j)^2}\right) =$$
$$= \sum_{j=1}^{N-1} (N-j) \log\left(1 - \frac{d^2}{(d-j)^2}\right)$$

$$x^T R_1^{-1} x = \frac{1}{P_{N-1}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \qquad (3.17)$$
$$(a_{N-1}(i-k)..a_{N-1}(j-k) -$$
$$a_{N-1}(N-i+k).a_{N-1}(N-j+k))\, x_i x_j$$
$$= \frac{1}{P_{N-1}} \sum_{k=1}^{N} \left\{\left[\sum_{i=1}^{N} a_{N-1}(i-k)\, x_i\right]^2 \right.$$
$$\left. + \left[\sum_{i=1}^{N} a_{N-1}(N-i+k)\, x_i\right]^2\right\}$$

As we observe in equation (3.16) and (3.17), the value of equation (3.12) is computable with a computational complexity of $O(N^2)$ which is so faster than other known algorithms with complexity of $O(N^3)$ [8][7][5]. So we can estimate *d* with any resolution by computation of log likelihood function for each value of *d* and applying a heuristic method like Maximum Descent Gradient (MDG) algorithm. After that we can estimate corresponding variance of FdGn using equation (3.10). Now we review Tewfik's iterative approach MLE of FdGn ARMA Filtered model's parameters estimation.

III.B. Tewfik's Approximate Iterative Algorithm for ARMA filtered model Parameters Estimation

In this algorithm, an initial value for ARMA filter's parameters, Hurst parameter and variance is selected. Then in *k'th* step, the observed vector, *y*[*n*], is filtered by inverse of ARMA filter whose parameters are $a_k(i), b_k(j);\ 1 \le i \le P, 1 \le j \le Q$ to obtain $z_k(n)$ and then according to ML algorithm described in *III.A*, we find the new values H_{k+1} and σ_{k+1}^2 for FdGn model. Then we apply the inverse of FdGn filter with respect to H_{k+1} and σ_{k+1}^2 to observe input vector, *y*[*n*], and to obtain $x_k[n]$. Then according to Levinson's algorithm, we find new values $a_{k+1}(i), b_{k+1}(j);\ 1 \le i \le P, 1 \le j \le Q$ of parameters of ARMA model. The algorithm stops at *k'th* step, if the difference norm of vector parameters between two consecutive steps becomes less than a predefined value, that is:

$$\left\|(a_{k+1}(i), b_{k+1}(j), H_{k+1}, \sigma_{k+1}) - (a_k(i), b_k(j), H_k, \sigma_k)\right\| < \varepsilon$$

No theoretical proof on computational complexity of this method has been found yet [2].

IV. EXPERIMENTAL RESULTS

For performance evaluation of two approaches explained in this paper, a database of *Persian* speech files recorded from a male speaker, sampled at *8KSample/Sec* and quantized by *16 bits* were used in speech classification. For implementing models, speech signal was divided into *20msec* frames in order to take the AR and ARMA filter on it. Numerical results proved the best window size for taking FdGn and correlation dimension is *60msec* (*480* samples).

Numerical results of applying correlation dimension are shown in figure 2. Because estimated correlation dimension (CD) of each Persian consonant is not a constant value at all and is a stochastic value changing from frame to frame, we estimated the mean value and variance of CD of each Persian consonant over all frames in our database and then we fitted a Gaussian function to our estimated mean value and variance of each consonant as shown in figure 2.

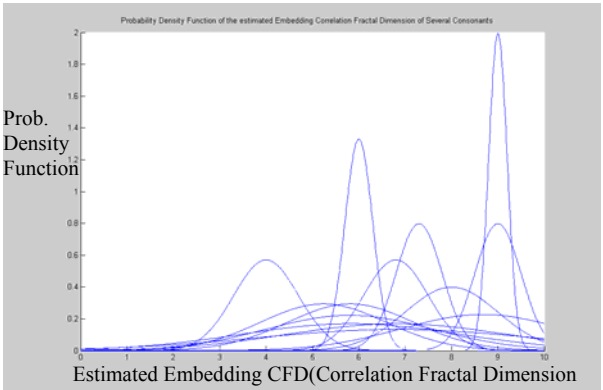

Fig.2: Probability density function of estimated embedding correlation dimension of several consonants

Also as we observe in figure 2, all the estimated CDs are less than 10 and CDs of several consonants are not distinguishable at all. However some of them like CD of "s" (س) which has a mean=9 and variance=*0.2* and CD of "z" (ز) which has a mean=*5.92* and variance=*0.3* are completely distinguishable. This phenomenon can be explained theoretically; "z" and "s" have an almost identical vocal tract filter but vocal cords in "z" have a more important role and hence "z" is a more voiced consonant. As we see, the algorithm fails in discriminating several consonants from each other.

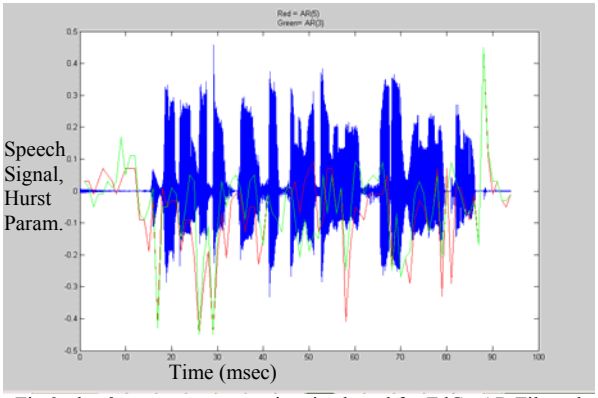

Fig.3: the *d* parameter versus time is plotted for FdGn AR Filtered model of degree5 (Red Curve) 3 (Green Curve) and Speech signal (Blue Curve) during the pronunciation of a sentence.

For applying FdGn AR filtered model, we chose several values for degree of AR filter and evaluated the *d* estimated curve versus time for several values of degree of AR filter, as figure 3 shows, and finally selected the minimum degree value which there is a little difference between its *d*-curve and *d*-curve of AR models with greater degree values as the optimal AR degree. As an example as figure 3 shows, AR(3) is not sufficient for modeling speech signal. Finally we concluded that AR(10) is totally sufficient for modeling the transient variations of speech signal.

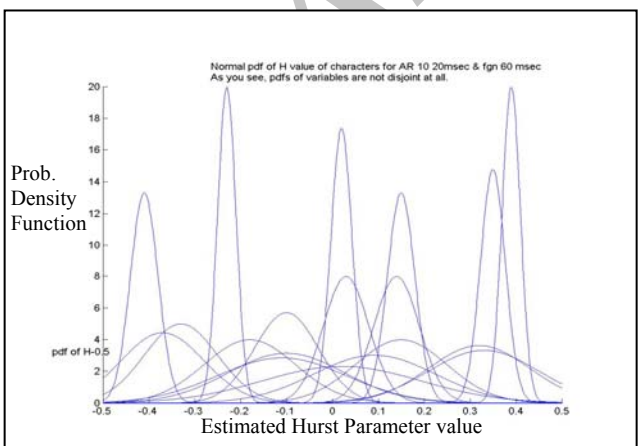

Fig. 4: pdf of Persian consonants according to their estimated mean and variance value and assumption of being Gaussian

Like estimated CFD, the estimated mean value and variance of *d* parameter of each consonant are used to fit a Gaussian function to probability density function of *d* parameter of each consonant. The results are plotted in figure 4. By evaluation of figures 2 and 4, we observe better discrimination between various consonants is obtained by applying AR Filtered FdGn model than CFD approach. The result is some how predictable. Because of using AR filter in AR Filtered FdGn model, it has the ability to pursue both the short term and long term varieties in speech signal and so is more perfect in speech modeling.

V. CONCLUSION

The chaotic and fractal characteristics of Persian speech signal are evaluated. Experimental results showed that the correlation dimension of speech signal is strictly greater than its geometric dimension which proves the strong chaotic characteristics in Persian speech signal. Finally, FdGn AR filtered approach and CFD approach in speech consonant recognition were evaluated against each other and as experimental results showed FdGn AR filtered approach yields better results in distinguishing Persian speech consonants.

REFERENCES

[1] J. R. M. Hosking, "*Fractional Differencing*", *Biometrica*, vol. 68, no. 1, pp. 165-176, 1981.

[2] A. H. Tewfik, et. Al., "*Signal Modeling with Filtered Discrete Fractional Noise Processes*", IEEE Trans. On Signal Processing,pp. 2839-2849, Vol. 41,Sept. 1993.

[3] S. E. Levinson and D. B. Roe, "*A Perspective on Speech Recognition*", IEEE Communications Magazine, pp. 28-34, Jan. 1990.

[4] T. J. Thomas, "*A Finite element model of Fluid Flow in the vocal tract*", Computer Speech and Language, pp. 131:151, 1986.

[5] A. Langi and W. Kinsner, "Consonant Characterization using Correlation Fractal Dimension for Speech Recognition", IEEE WESCANEX, pp. 208-213, 1995.

[6] B. B. Mandelbrot and W. Vanness, "*Fractional Brownian Motion , Fractional Noises and Applications*" *SIAM* Rev., vol. 10,no. 4, pp.422-437, Oct. 1968.

[7] P. Maragos, K. L. Young, "*Fractal Excitation Signals For CELP Speech Coders*", IEEE Trans., pp. 669-672, Feb. 1990.

[8] A. H. Tewfik, et. Al., "*Maximum Likelihood Estimation of the parameters of Discrete Fractionally Differenced Gaussian Noise Process*", IEEE Trans. On Signal Processing,pp. 2977-2989, Vol. 41,Oct. 1993.

[9] S. Fekkai, et. Al., "*Fractal Dimension Segmentation: Isolated Speech Recognition*", Inst. Of Electrical Engineers, pp. 1-4, 2000.

[10] D. J. Tritton, "Physical Fluid Dynamics", Oxf. U. P., 1988.

[11] A. Langi, et. Al., *"Multifractal Processing of Speech Signals"*, IEEE Conf. on Information, Communications and Signal Processing", pp. 527-531, Sept. 1997.

[12] P. Maragos, G. Kubin, et. Al., *"Nonlinear Speech Processing: Overview and Applications"*, Vienna Telecommunication Research.
[13] E. L. J. Bohez, et. Al., "*Fractal Dimension and Iterated Function System(IFS) For Speech Recognition*", Electronics Letters, pp. 1382-1384, Vol. 28, July 1992.
[14] S. E. Levinson and D. B. Roe, "*A Perspective on Speech Recognition*", IEEE Communications Magazine, pp. 28-34, Jan. 1990.