Fractal dimensional analysis of voice fluctuation: normal and pathological cases.

Yuki Kakita and Hitoshi Okamoto

Dept. of Electronics, Kanazawa Inst. Tech. (KIT)
7-1 Ohgigaoka, Nonoichi-Machi, Kanazawa-Minami, 921 JAPAN

ABSTRACT - This paper reports on the values of fractal dimension of fundamental frequency (F_0) fluctuation and amplitude fluctuation for two kinds of voice, the normal and the pathological (perceived as "rough"). The values of fractal dimension were computed by the embedding method. The median of fractal dimension was 3.3 for the normal voice and 4.8 for the pathological, differing by 1 to 2 dimensions. Reconstructed phase portraits were also examined to characterize the difference between the two kinds of voices geometrically.

INTRODUCTION

The aim of the present study is to differentiate two kinds of voice, the normal and the pathological (perceived as "rough"), by using the fractal dimensions of fundamental frequency (F_0) fluctuation and amplitude fluctuation. The long term goal of the study is to establish a clinical method for detecting pathological changes in voice quality based on a fractal dimentional analysis.

Okamoto and Kakita (1988) previously reported on a fractal dimensional analysis of voice fluctuation in two subjects, one normal and one pathological. The median values of the dimension were 2.6 for the normal subject and 4.3 for the pathological, differing by 4.7 dimension. This paper presents the results of examining 6 normal cases and 6 pathological cases, thus providing more precise values of the fractal dimensions. The paper also discusses the characteristics of the generation model of the fluctuation based on quasi-phase-portraits.

METHODS

The time series of F_0 and amplitude were extracted from the speech waveform of sustained utterances of a Japanese vowel /e/. The speakers were instructed to produce /e/ for 2-3 seconds with comfortable pitch and loudness. The original waveform was sampled at 40kHz and was digitized with a 16 bit analog-to-digital converter. Fundamental periods and maximum amplitudes were obtained by peak detection method.

For both the normal and the pathological cases, 12 voice samples (2 repetitions x 6 subjects) were analyzed.

The value of the fractal dimension, or, strictly speaking, the phase dimension, was obtained as follows:

(1) Reconstruct phase portraits for d-fold ($=1, 2, 3, \ldots$) dimension by embedding;

(i). Suppose that $x_{(i)}$ $(i=1, 2, \ldots, M)$ is a time series.

(ii) Next, make a new series

$$X_j=\{x(j), x(j+\tau), x(j+2\tau), \ldots, x(j+(d-1)\tau)\} \ldots (1)$$
$$(j=1, 2, \ldots, N)\ (N\leq M-(d-1)),$$

where τ is an arbitrary constant of time interval.

(iii) Then plot X_j in a d-dimensional space.

(iv) X_j constructs a trajectory of the phase portrait, a sampled candidate of the original phase-portrait.

The phase space thus constructed is called a pseudo-phase-space, or embedding space.

(2) Calculate a correlation function for each pseudo-phase-portrait projected in d-dimensional space.

(i) Fix the value of d at a specific value.

(ii) The number of points to be plotted is N $(N \leq M-(d-1))$.

(iii) Compute the distance $l = |X_i - X_j|$ between a pair of points.

(iv) Count the number of pairs, n, where $l < r$.

(v) Calculate the rate $C(r) = \left(\frac{n}{N^2}\right)$.

The correlation function is defined as

$$C(r) = \left(\frac{1}{N^2}\right)(\text{number of pairs (i,j) with distance } s_{i,j} < r) \;...\; (2)$$

$C(r)$ represents the probability that the point X_j lies within a sphere of radius r. (Moon, 1987) If $C(r)$ is approximated as $C(r) \propto r^D$ ... (3), then D is called the correlation dimension. This definition of dimension was proposed by Grassberger and Procaccia (1983).

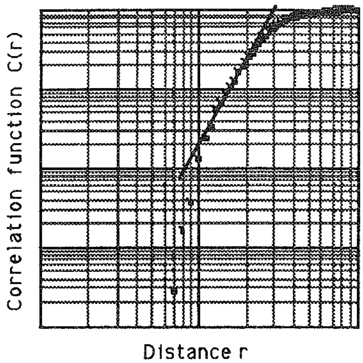

Figure 1. Correlation function $C(r)$

Figure 1 shows a log-log plot of an example of $C(r)$. Notice that, $C(r)$ shows a marked increase for a small r region, whereas $C(r)$ tends to saturate for a great r region.

According to Eq. 3, the linear portion of $C(r)$ is of interest. The value of the slope is specific for d value ($d=10$ in this data).

(3) Measure the maximum slope of the function:

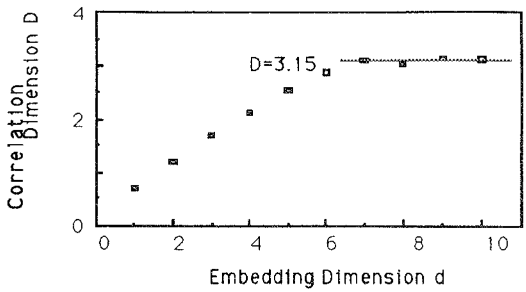

Figure 2. Correlation dimension D versus dimension of embedding pseudo-phase-space d for an amplitude fluctuation of normal voice.

Figure 2 shows a plot of D vs. d. Notice that D increases for small d region. While, D gets close to an asymptotic value for great d region. The asymptotic value of D is called the fractal dimension.

RESULTS

Figure 3 shows box & whisker plots for 12 normal and 12 pathological voice samples. Each data set consists of two repetitions each of 6 male subjects.

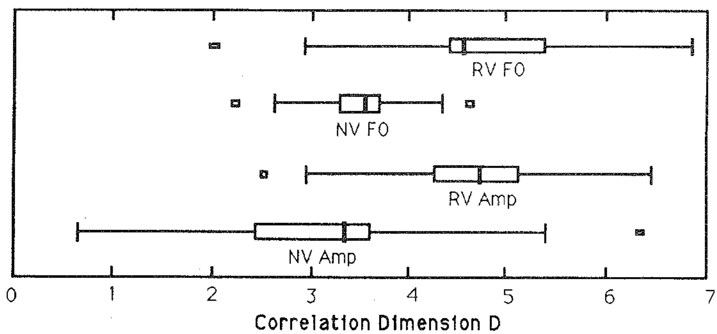

Figure 3. Box-and-whisker plots for the values of correlation dimensions

For both the F_0 series and the maximum amplitude series, the medians are about 3.3 for normal and about 4.8 for the "rough" voice. The median of the fractal dimension for the "rough" voice was greater than that for the normal voice. by approximately one.

H-spreads, or the interquartile ranges, are about 1 for all the cases except the F_0 series for normal voice, whose value was about 0.5.

DISCUSSION

According to the theory of fractal/chaos science, the value of fractal dimension can be considered as the number of independent parameters specifying the dynamical system of the phenomenon.

In order to find out what dynamical system yielded such a fluctuating time series, the phase portraits of attractors were examined for 12 normal cases.

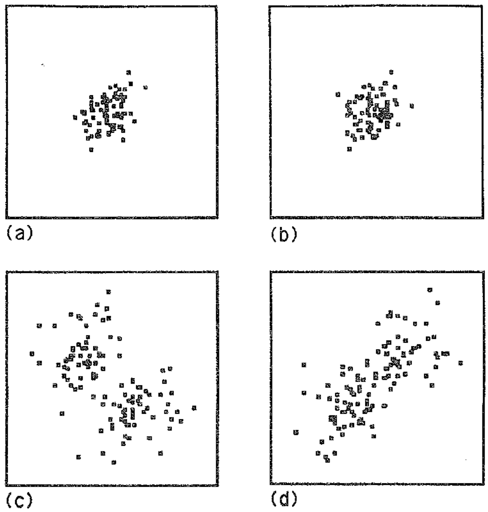

Figure 4 Reconstructed pseudo-phase-portraits for normal voice (a,b), and "rough" voice (c,d)

Figure 4 shows pseudo-phase-portraits reconstructed from the time series by embedding method. Plots (a, b) are for the normal voice and plots (c, d) are for the "rough" voice. For both cases, a sample whose fractal dimension was closest to the median was selected.

Plots (a) and (c) show the views of the portrait ($d = 10$, see Eq. 1) projected on the first and the second axes. Plots (b) and (d) show different views of the same portrait projected on the first and the third axes. Scale of the axes is the same for both cases. Each of the scattered points indicates a sampled point of the trajectory comprising the phase-portrait.

The data points in Plots (a) and (b) are compactly and evenly distributed. In contrast to this, the data points in plots (c) and (d) are widely spread as a whole and, moreover, form two locally dense clusters

For the normal voice, the true phase-portrait estimated from the "single-cluster" quasi-phase-portrait --- plots (a) and (b) --- may be expressed as an ellipsoid. The true trajectory turns around the surface of an ellipsoid in a rather random manner. For the "rough" voice, the true phase-portrait, estimated from the "double-cluster" quasi-phase-portrait --- plots (c) and (d) ---, may be expressed as a torus". The torus-like shape of the phase-portrait supports the fact that the time series manifests a two-fold (e.g. long-short) periodicity.

REFERENCES

Grassberger, P and Procaccia, I (1983) "Characterization of Strange Attractors", Phys. Rev. Lett. 50. 346-349.

Moon, F. C. (1987) "Chaotic Vibrations", Wiley, 216-235.

Okamoto, H and Kakita, Y (1988) "Analysis of fluctuations of fundamental frequency and maximum amplitude of voice based on the fractal dimension" Proc. Acoust. Soc. Jpn. Meeting (Fall),125-126. (In Japanese, with English title and figure captions.)

PHASE SPACE BEHAVIOUR OF SPEECH

Dr Leisa Condie

Dept of Mathematics, Statistics and Computing Science,
University of New England - Armidale

ABSTRACT - Characterisation of phase space behaviour of speech gives an alternate view of speech waveforms. Correlation dimension and false nearest neighbour measures are just two techniques for investigating the dynamical behaviour of speech. The recent false nearest neighbours technique for determining the embedding dimension for phase space reconstruction of a time series is used on both voiced and unvoiced speech. Results from false nearest neighbours are compared with the correlation dimension results.

THEORETICAL CONCEPTS

An n-dimensional manifold M is a space in which it is possible to set up a coordinate system near each point such that locally the space looks like a subset of the Euclidean space R^n. A dynamical system D is a smooth manifold M, together with a vector field v defined on M (Casti, 1989). An attractor is a subset of n-dimensional phase space which almost all sufficiently close trajectories approach asymptotically.

There are three major areas of interest in the characterisation of attractors: Lyapunov (or characteristic) exponents, entropy and dimensions. Lyapunov exponents give the average rate of expansion (if positive) or contraction (if negative) near a limit set. They generalise the linear stability criteria for fixed points and limit cycles. A non-chaotic system is asymptotically stable if all its Lyapunov exponents are negative. For any limit sets on continuous, time-dependent systems, except an equilibrium point, the exponent is zero. Although it is required that the sum of exponents is negative for dissipative systems, the presence of one or more positive exponents indicates a strange attractor (Condie, 1991).

Dimensions

The dimension of an attractor can be expressed in many ways, but the underlying idea is that there is a lower bound on the number of variables needed to describe the steady state behaviour of the system. A non-chaotic system has integer dimension, whilst a chaotic system almost always has fractal (non-integer) dimension.

Given an attractor in $\mathcal{R}^n$ covered by $N(r)$ n-dimensional hyperspheres of radius r then

- The fractal, or Hausdorff, dimension is calculated as the limit is taken. As $r \to 0$, $N(r) \propto r^{-D_H}$.
- The capacity dimension is defined as

$$D_{cap} = \lim_{r \to 0} \frac{\ln N(r)}{\ln(1/r)}$$

  or alternately, $N(r) \propto (1/r)^{D_{cap}}$ for some $k > 0$.

- Let P_i be the relative frequency of visitation of a typical trajectory to the ith hypersphere. The information dimension is then defined to be

$$D_I = \lim_{r \to 0} \frac{H(r)}{\ln(1/r)}$$

  where $H(r) = -\sum_{i=1}^{N(r)} P_i \ln P_i$, or alternately $H(r) = kr^{-D_I}$ for some $k > 0$.

- The correlation dimension can be expressed as

$$D_C = \lim_{r \to 0} \frac{\ln C(r)}{\ln r}$$

where

$$C(r) = \lim_{N\to\infty} \frac{1}{N^2} \sum_{i\neq j} \Theta(r - \|x_i - x_j\|)$$

$\Theta(x)$ is the Heaviside function and $\|\cdots\|$ represents the Euclidean norm. Alternately $C(r) = kr^{D_{cap}}$ for some $k > 0$.

- Given Lyapunov exponents $\lambda_1, \lambda_2, \cdots, \lambda_j$ where $\sum_{i=1}^{j} \lambda_i \geq 0$ then the Lyapunov dimension

$$D_L = j + \frac{\lambda_1 + \cdots + \lambda_j}{|\lambda_{j+1}|}$$

If no j exists to satisfy the condition, D_L is defined to be 0. The sum of positive exponents is essentially the Kolmogorov entropy.

False Nearest Neighbours

From the set of observations $x(n)$, multivariate vectors in d dimensional space $y(n) = [x(n), x(n+T), \cdots, x(n+(d-1)T)]$ are used to trace out the orbit of the system, where T is the time delay (usually chosen to be the first minimum of the average mutual information). The aim of the method proposed by (Kennel et al., 1992) is to find the embedding dimension d by examining the topology of the embedding. The purpose of time delay is to unfold the projection of multivariate state space onto the one dimensional time series back to a multivariate system representative of the original. Takens (1981) states that $d > 2d_A$ where d_A is the dimension of the attractor, but this is only a sufficient condition, and in experiments it is preferable to use d_E, the minimum embedding dimension.

The idea of false nearest neighbours is that in examining d and then $d+1$ one can distinguish true neighbours of a given point from false neighbours. False neighbours are points on the data set that are neighbours solely because we are viewing the attractor in too small an embedding space ($d < d_E$) and in a higher dimension those points will have moved apart and no longer be neighbours. The value for d_E is then obtained when there are no false nearest neighbours in successive embeddings. Other methods involve computation of some invariant on the attractor until it becomes independent of d, but such methods tend to be data intensive, and somewhat subjective in determining d_E.

Neighbours are determined by simply using the square of the Euclidean distance between points on the attractor. The rth nearest neighbour $y^{(r)}(n)$ of $y(n)$ has distance

$$R_d^2(n,r) = \sum_{k=0}^{d-1} (x(n+kT) - x^{(r)}(n+kT))^2$$

In $d+1$ dimensions $R_{d+1}^2(n,r) = R_d^2(n,r) + (x(n+dT) - x^{(r)}(n+dT))^2$. Thus a false nearest neighbour can be defined as any neighbour such that

$$\sqrt{\frac{R_{d+1}^2(n,r) - R_d^2(n,r)}{R_d^2(n,r)}} = \frac{|x(n+dT) - x^{(r)}(n+dT)|}{R_d(n,r)} > R_{tol}, \quad R_{tol} \geq 10$$

It is sufficient to use $r = 1$, for the nearest neighbour only, and interrogate on all points $n = 1, \cdots, N$ of the attractor. Unfortunately this condition is not sufficient for determining d_E: the nearest neighbour is not necessarily a *close* neighbour - indeed $R_d(n,1)$ can be comparable to the size of the attractor for a noise waveform, leading to erroneous results. Thus a second measure is used in conjunction with the first, and both must be satisfied for true neighbours.

$$\frac{R_{d+1}(n)}{R_A} > A_{tol}$$

where

$$R_A^2 = \frac{1}{N} \sum_{n=1}^{N} (x(n) - \frac{1}{N} \sum_{n-1}^{N} x(n))^2$$