1997

Multifractal Analysis of Unvoiced Speech Signals

Olufemi A. Adeyemi
University of Rhode Island

Follow this and additional works at: http://digitalcommons.uri.edu/oa_diss

Terms of Use
All rights reserved under copyright.

Recommended Citation
http://digitalcommons.uri.edu/oa_diss/468

This Dissertation is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.
MULTIFRACTAL ANALYSIS OF UNVOICED SPEECH SIGNALS
BY
OLUFEMI A. ADEYEMI

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN ELECTRICAL ENGINEERING

UNIVERSITY OF RHODE ISLAND
1997
DOCTOR OF PHILOSOPHY DISSERTATION
OF
OLUFEMI A. ADEYEMI

APPROVED:
Dissertation Committee
Major Professor

[Signatures]

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
1997
Abstract

In this thesis, we analyze the complexity involved in the production of unvoiced speech signals with measures from nonlinear dynamics and chaos theory. Previous research successfully characterized some speech signals as chaotic. However, in this dissertation, we use multifractal measures to postulate the presence of various fractal regimes present in the attractors of unvoiced speech signals. We extend prior work which used only correlation dimension $D_2$ and Lyapunov Exponents to analyze some speech sounds. We capture the chaotic properties of unvoiced speech signals in the embedded vector space more succinctly by not only estimating the correlation dimension $D_2$, but by also estimating the generalized dimension $D_q$. The (non-constant) generalized dimensions were estimated from phase space reconstructed vectors of single scalar variable realization of unvoiced speech signals. The largest of those dimensions is an indicator of the minimum dimension required in the phase space of any realistic dynamical model of speech signals.

Results of the generalized dimension estimation support the hypothesis that unvoiced speech signals indeed have multifractal measures. The multifractal analysis also reveals that unvoiced speech signals exhibit low-dimensional chaos as well as "soft" turbulence. This is in contrast to the opinion that unvoiced speech signals are generated from what is technically known as "hard" turbulent flow, in which the dimension of a dynamical model is very high. Unvoiced speech signals may actually be generated from "soft" turbulent flow.

In this dissertation, we also explore the relationship between the estimated generalized dimensions $D_q$ and the singularity spectrum $f(\alpha)$. Existing algorithms for accurately estimating the resulting singularity spectrum $f(\alpha)$ from the samples of generalized dimensions $D_q$ of a multifractal chaotic time series use either (a) linear
interpolation of the known, coarsely sampled, $D_q$ values or (b) a finely sampled $D_q$ curve obtained at great computational/experimental expense. Also, in conventional techniques the derivative in the expression for Legendre transform necessary to go from $D_q$ to $f(\alpha)$ is approximated using first order centered difference equation. Finely sampling the $D_q$ is computationally intensive and the simple linear approximations to interpolation and differentiation give erroneous end points in the $f(\alpha)$ curve. We propose using standard min-max filter design methods to more accurately interpolate between known samples of the $D_q$ values and compute the differentiation needed to evaluate the Legendre transform. We use optimum (min-max) interpolators and differentiators designed with the Parks-McClellan algorithm. We have computed the generalized dimensions and singularity spectrum of 20 unvoiced speech sounds from the ISOLET database. The results not only indicate multifractality of certain unvoiced speech sounds, but also may lead to nonlinear maps that may be useful in improving the nonlinear dynamical modeling of speech sounds.

This new approach to $f(\alpha)$ singularity spectrum calculation exhibits computational reduction and improved accuracy. The proposed method also provides estimates of the generalized dimensions at $D_\infty$ and $D_{-\infty}$ which are almost impossible to obtain from real data with limited number of data samples. Also, the asymmetric spread of $\alpha$ values with the corresponding $f(\alpha)$ around the maximum of $f(\alpha)$ reveal the inhomogeneity in the attractors of unvoiced speech signals just like the variations in the $D_q$ values. The asymmetric spread of $\alpha$ values may also be an indication that the turbulent energy fields generated during unvoiced speech production are made of non-homogeneous fractals.
Acknowledgments

It is indeed with pleasure and admiration that I want to acknowledge the expert guidance, consistent support and helpful suggestions of my major professor, Prof. G. Faye Boudreaux-Bartels in guiding and being patient with me through every stage of this dissertation. Without her effort and financial support, there are times when I would have given up. Her quickness of thought, ability to teach and willingness to learn about unrelated fields have helped me tremendously in appreciating the meaning of a great teacher, researcher, administrator and friend. I would also like to acknowledge the technical assistance of Professors Ken Hartt and Deborah Pence who served on my thesis committee in various capacities.

I want to thank the Graduate School for awarding me the Patricia Robert Harris Fellowship from 1991 through 1994, without this fellowship, I would not have dreamed of being in the doctoral program. I want to thank the Graduate Fellowship committee and Dr. Harold Bibb for awarding me a special URI Graduate Fellowship in 1995 without which the completion of my thesis would have been disrupted. I also want to thank Professors Boudreaux-Bartels and Bill Ohley for supporting me in the Summers of 1992 and 1993, respectively.

It has been a cherished experience to work with and study under world renown experts in the areas of digital and statistical signal processing. I have come to appreciate the power of signal processing through the lectures of Professors Faye Boudreaux-Bartels, Steven Kay, Peter Swaszek, Leland Jackson and Ramdas Kumaresan.

One of the best aspects of my stay at URI has been the friendships developed with fellow graduate students and particularly, the Time-Frequency Representations (TFR) group. To Antonia and Robin, many thanks for your warm kindness, support
and love. I will always cherish your words of encouragement and help during difficult times.

I would like to thank Phyllis Golden and Donna Gray for their help during the course of my doctoral program.

Finally, I want to thank members of my family and friends for their patience and support during my doctoral program. I am particularly indebted to my brother Bayo Adeyemi, my wife Dana Akins-Adeyemi and my little angel Atinuke.

Thank you !!!

Femi.
Dedicated to four Adeyemi ladies:
Alice, Bunmi, Dana and Atinuke
Contents

Abstract ii
Acknowledgments iv

Table of Contents vii
List of Tables xi
List of Figures xii

1 Introduction and Overview 5
  1.1 Organization of Dissertation 6

2 Chaos and Dynamical Systems 8
  2.1 Introduction 8
    2.1.1 What is Chaos ? 9
  2.2 Lorenz System 9
  2.3 Characteristics of the Logistic Map 12
    2.3.1 Importance of fixed points 13
    2.3.2 Bifurcation 18
    2.3.3 Lyapunov Exponent 20
    2.3.4 Invariant Density 20
    2.3.5 Ergodicity 22
  2.4 Spatial Properties 26
2.5 Measures of Chaos ........................................ 26
   2.5.1 Dimensions ........................................ 28
2.6 Correlation dimension, $D_2$ .......................... 31
   2.6.1 Correlation Integral .............................. 31
2.7 Generalized Dimension, $D_q$ ......................... 34
   2.7.1 Box-counting Dimension ......................... 35
   2.7.2 Information Dimension ............................ 35
2.8 Taken’s Embedding Theorem ............................ 36
   2.8.1 Mutual Information ............................... 38
2.9 Numerical Implementation of the Generalized Correlation Integrals GCI 41
   2.9.1 Numerical Problems ............................... 43
   2.9.2 Other GCI-GD Algorithms ......................... 45
2.10 Lyapunov Exponents .................................. 47
   2.10.1 Logistic Map LE ................................ 48
2.11 Entropies ............................................. 53
   2.11.1 Entropy $K_2$ .................................. 54
   2.11.2 Numerical Implementation of $K_2$ .............. 55

3 Background on speech production and processing 57
   3.1 Introduction ........................................ 57
      3.1.1 Importance of Speech .......................... 57
   3.2 Classification of Speech ............................. 58
      3.2.1 Voiced Speech ................................ 60
      3.2.2 Unvoiced Speech ............................... 60
      3.2.3 Mixed Voiced Speech ........................... 62
   3.3 Speech Modeling ..................................... 62
      3.3.1 Linear Models ................................ 64
      3.3.2 Nonlinear Models .............................. 68
      3.3.3 Nonlinear Dynamical Models .................... 71

4 $f(\alpha)$ Singularity spectrum formalism from Generalized Dimensions $D_q$ 75

viii
4.1 Introduction .................................................. 75
4.2 Singularity Spectrum $f(\alpha)$ ............................... 76
4.3 Multifractal characterization of the Koch curve ................. 77
4.4 Characteristics of $f(\alpha)$ .................................... 83
4.5 Relationship between $f(\alpha)$ and $D_q$ ......................... 83
4.6 Evaluating $f(\alpha)$ from $D_q$ ................................ 87
  4.6.1 Cantor Set ............................................... 87
  4.6.2 Problems of evaluating $f(\alpha)$ from $D_q$ ................ 88
4.7 Existing Algorithms for evaluating $f(\alpha)$ ...................... 89
  4.7.1 Analysis of existing methods of evaluating $f(\alpha)$ from $D_q$ 92
  4.7.2 Example - Effects of Coarse Sampling ..................... 96
4.8 Improved Accuracy in the Singularity Spectrum .................. 99
4.9 Min-Max Digital Filter Design .................................. 99
4.10 Interpolation Filter .......................................... 102

5 Multifractal analysis of unvoiced speech signals .................. 109
  5.1 Speech Data ............................................... 110
  5.2 Noise in speech signals ...................................... 111
  5.3 Concerns in chaos analysis of experimental data ............... 114
  5.4 Limitations of chaos analysis due to the properties of speech data 115
    5.4.1 Finite record length effect on $D_q$ ...................... 116
    5.4.2 Noise level effect on $D_q$ .............................. 118
  5.5 Procedure for estimating the generalized dimensions from speech data 118
  5.6 Inadequacy of the $D_2$ measure in characterizing the attractors of speech signals ........................................ 120
  5.7 Effect of high embedding dimensions on real data .............. 123
  5.8 Low dimensional turbulence in unvoiced speech signals ........ 141
  5.9 Generalized dimension $D_q$ of unvoiced speech signals .......... 146
    5.9.1 Result of $D_q$ estimation for unvoiced speech signals .. 146
  5.10 $f(\alpha)$ singularity spectrum for unvoiced speech .......... 151
  5.11 Discussion of $D_q$ and $f(\alpha)$ results .................... 169
6 Conclusions and Future work

6.0.1 Future Research

A Numerical Computation of the Generalized Dimensions $D_q$

B Frequency domain implementation of $f(\alpha)$ singularity spectrum

C Plot of the Generalized Correlation Integrals

D Plots of various measures of chaos in unvoiced speech signals

List of References

Bibliography
List of Tables

5.1 Shows a summary of the generalized dimension $D_q$ results for unvoiced speech signals \"s\" and \"z\". .................................................. 172

5.2 Shows a summary of the values of $\alpha_{min}$ and $\alpha_{max}$ for the unvoiced fricative speech sound \"s\" and \"z\". The values correspond to the estimates of $D_\infty$ and $D_{-\infty}$ obtained by the min-max method described in chapter 4. .................................................. 172

5.3 Shows a summary of the multifractal results for the unvoiced speech sound \"s\" from the ISOLET database, spoken by female (f) and male (m) speakers and located in the file whose name is given in the third column. The $D_0$ data points were obtained from the $D_q$ estimates at the index $q = 0$. The $D_{-\infty}$ and $D_\infty$ values were estimated from the singularity spectrum $f(\alpha)$ estimates to correspond to the points $\alpha_{min}$ and $\alpha_{max}$ respectively at which $f(\alpha) = 0$. $D_{-\infty}$ and $D_\infty$ were not estimated for sounds with invalid (non-monotonic) $D_q$ values in the range $|q| < 3$. .................................................. 173

5.4 Shows a summary of the multifractal results for the unvoiced speech sound \"z\" from the ISOLET database, spoken by female (f) and male (m) speakers and located in the file whose name is given in the third column. The $D_0$ data points were obtained from the $D_q$ estimates at the index $q = 0$. The $D_{-\infty}$ and $D_\infty$ values were estimated from the singularity spectrum $f(\alpha)$ estimates to correspond to the points $\alpha_{min}$ and $\alpha_{max}$ respectively at which $f(\alpha) = 0$. $D_{-\infty}$ and $D_\infty$ were not estimated for sounds with invalid (non-monotonic) $D_q$ values in the range $|q| < 3$. .................................................. 174
List of Figures

2.1 A schematic diagram of the Rayleigh-Bernard cell. It assumes that the cell extends infinitely from left to right and out of the page (adopted from [1]) ................................................. 11

2.2 This figure illustrates the behavior of two trajectories of the logistic map when \( r = 2.8 \). With two different initial conditions \( x[0] = 0.2 \) (solid line) and \( x[0] = 0.9 \) (dashed line) that are far apart from each other, the map settles to a stable period 1 cycle about the fixed point \( x^* = 1 - \frac{1}{2.8} = 0.6428 \). ................................. 16

2.3 Behavior of two trajectories of the Logistic map with \( r = 3.5 \) in equation (2.4). Notice that with two initial conditions \( (x[0] = 0.2 \) (solid line) and \( x[0] = 0.9 \) (dashed line)) that are not close to each other, the trajectories of the map still converge to a stable cycle of period 4 about the fixed points \( x^* \) of equation (2.13). ................................. 17

2.4 This figure illustrates some of the stable and unstable fixed points of various periods that arise from a bifurcation process in equation (2.4) taken from [2, 3]. Basically, as \( r \) increases, stable fixed points of period \( 2^{n-1} \) become unstable and give rise by bifurcation to new stable fixed points of period \( 2^n, n = 0, 1, \ldots \) ................................. 19

2.5 A plot of the behavior of two trajectories of the Logistic map in the chaotic region with \( r = 4 \) in equation (2.4). The trajectories start at initial conditions \( (x[0] = 0.0903000 \) (solid line) and \( x[0] = 0.0903001 \) (dashed line)) that are very close. The divergence in the orbits becomes apparent at about sample number 16. ................................. 21
2.6 The normalized invariant density $p(x) = \frac{1}{\pi(x(1-x))^{\frac{1}{2}}}$ versus $x$ for the Logistic map at $r = 4$. Notice the singularities at $x = 0$ and $x = 1$. 

2.7 (a) depicts the first 100 samples of the length 1500 ensemble time series of the Logistic map at $r = 4$ generated using 1500 uniformly distributed initial conditions. The ensemble of samples at time instant $n = 300$ were recorded and the first 100 samples is displayed. (b) shows the corresponding normalized histogram of the 1500 ensemble time series sorted into 10 equally spaced bins along the x-axis.

2.8 (a) is a plot of the last 100 out of 1500 samples of a single realization of the Logistic map with initial condition $x[0] = 0.345$ and $r = 4$. (b) shows the corresponding normalized histogram of the 1500 points sorted into 10 equally spaced bins along the x-axis.

2.9 This is a phase-space plot of the famous Lorenz Attractor. This figure illustrates the projection of the phase-space orbit of equation (2.3) on the x-z plane. That is, the pair $(x(t), z(t))$ in equation (2.3) form the co-ordinates of a point along the horizontal and vertical axis, respectively. The curve shows the trajectory of this pair as time, $t$, evolves.

2.10 Shows an illustration of the covering of an object (a set of points indicated by the dots) by cubes of width $\epsilon$ (adopted from [4]).

2.11 Shows the normalized mutual information as a function of time lag $\tau$ for the scalar $x(t)$ variable of Lorenz data in (2.3). The plot is normalized by the mutual information at lag zero. The mutual information at lag zero is 2.411492 bits.

2.12 A schematic diagram of the correlation dimension $D_2$ plotted as function of the embedding dimension. As the embedding dimension is increased, a value of $d_E$ is reached where the correlation dimension converges and is independent of the embedding dimension. The correlation dimension of the attractor in this case is 2.01 (adopted from [1]).
2.13 A schematic sketch of the local slope $\frac{\partial}{\partial \log \varepsilon} \log C_{2,d_E}$ as a function of $\log \varepsilon$ showing the distinct regions that are typical for chaotic attractors (adopted from [5]).

2.14 (a) Shows the log of correlation integral plotted as a function of $\log \varepsilon$ for different embedding dimensions $d_E$ from 5 to 15. Fig. 2.14: (b) depicts the corresponding local slopes which are estimates of the correlation dimension $D_2$. The dimension estimation was performed on the x-component of the Lorenz data and the resulting estimated correlation dimension $D_2$ in the plateau region between $6 < \log \varepsilon < 7.5$ is 2.06.

2.15 Shows the Lyapunov exponent as a function of the tuning parameter $r$ for the Logistic map. When $r > r_c = 3.57$, the Lyapunov exponent becomes positive indicating chaotic behavior.

2.16 A schematic diagram of the evolution and replacement procedure for estimating the largest Lyapunov exponent from a time series. The fiducial trajectory is shown by the thick solid line with the fiducial points $X_i$ encircled. The dashed lines shows the evolution of the non-fiducial points after each time step. The largest Lyapunov exponent is computed from the growth of length elements (adopted from [6]).

2.17 Values of Grassberger-Procaccia Entropy $K_{2,d_E}$ for the Lorenz Attractor (adopted from [7]).

3.1 A cross-section of the human head, showing the vocal and nasal tracts (adopted from [8]).

3.2 A cross-section of the human larynx (adopted from [8]).

3.3 Digital processing model for production of speech signals (adopted from [9]).

4.1 (a) The first three iterations of the triadic Koch curve. Notice the increase in the number of segments as well as the decrease in the length of each segment as $k$ increases (adopted from [10]). (b) shows the associated uniform distribution after the first three iterations.
4.2 (a) The first three iterations of the non-uniform triadic Koch curve. 

(b) Shows the associated geometric mass distribution (pdf) of the non-uniform triadic Koch curve. Here $p_1 = 0.35$ and $p_2 = 0.15$. The original uniform pdf is displayed in the upper-left panel. The resulting pdf after three iterations of the Koch curve is shown column-wise from top to bottom. The vertical-axis corresponds to the scaled pdf of the length segments and therefore the area below the density curve provides the probability in any subinterval (adopted from [11]). The pdf has been scaled to clearly show the resulting intermittency as the non-uniform triadic Koch curve is iterated. 

4.3 Shows the associated scaled pdf of the non-uniform triadic Koch curve at the eighth iteration, i.e. $k = 7$. Notice the intermittent and self-similar structure (adopted from [11]). Also notice the points in the pdf that tend to increase beyond bound while others shrink towards zero. The pdf has been purposely scaled to show the intermittent nature of the resulting pdf.

4.4 Shows the Singularity spectrum $f(\alpha)$ for the triadic Koch curve with $p_1 = 0.35$. Notice that the maximum of $f(\alpha)$ corresponds to the box-counting dimension $D_0 \approx 1.26$.

4.5 (a) $D_q$ curve for a Cantor set, finely sampled from the closed form expression in equation (4.20). Figure 4.5: (b) shows the corresponding $f(\alpha)$ spectrum using the min-max filter design method [12]. Note that the maximum value of $f(\alpha)$ corresponds to $D_0 = 0.538$ and that $\alpha_{min} = D_{\infty} = 0.387$, $\alpha_{max} = D_{-\infty} = 0.773$ at $f(\alpha) = 0$.

4.6 Magnitude of the Fourier transform of the Cantor set $D_q$ curve in Fig. 4.5 (a). The frequency axis is normalized by $\frac{2\pi}{\Delta q}$, where $\Delta q = 1$, such that the Nyquist frequency corresponds to $F = \frac{1}{2}$. 

xv
4.7 Shows a schematic diagram of the frequency response of an ideal low-pass filter needed for interpolation and the sinc-squared type filter which corresponds to linear interpolation. The dashed lines shows the frequency response of an ideal lowpass filter whereas the solid line represents the sinc-squared frequency response of linear interpolation. $F_c$ is the passband cut-off frequency that separates the passband from the stopband: $F_s$ is the stopband frequency at the edge of the stopband. $\delta_p$ and $\delta_s$ are the maximum deviation or errors to be tolerated in the passband and stopband respectively. Here, $F = \frac{1}{2}$ corresponds to the Nyquist frequency, $f_N = \frac{1}{2\Delta q}$ Hz.

4.8 Shows a schematic diagram of the magnitude response of an ideal differentiator $j\omega$ and that of the sinusoidal approximation produced by a centered first order difference operation. Notice that the centered first order approximation (in solid line) is only close to the ideal differentiator (in dashed line) for frequencies close to zero. The centered first order approximation becomes a very poor approximation as normalized frequency $F$ becomes close to the normalized Nyquist frequency $F_N = \frac{1}{2}$.

4.9 (a) Coarsely sampled (i.e. decimated using $\Delta q = 4$) of the generalized dimensions $D_q$ of the Cantor set. (b) The corresponding $f(\alpha)$ of the decimated $D_q$ computed using the Centered Difference Equation approximation to differentiation. This is not a valid $f(\alpha)$ singularity spectrum.

4.10 (a) Linearly interpolated generalized dimensions of the Cantor set computed by linearly interpolation between the decimated $D_q$ in Fig. 4.9 (a). * indicates the original decimated samples in Fig. 4.9 (a); + indicates the linearly interpolated samples. Fig. 4.10: (b) $f(\alpha)$ of Linearly interpolated generalized dimensions $D_q$ in Fig. 4.9 (a) using the Centered Difference Equation to approximate differentiation. Notice the spurious erroneous points near $f(\alpha) \approx 0$ at $\alpha = 0.387$ and $\alpha = 0.773$. 

xvi
4.11 (a) Shows the original signal spectrum, (b) "Stretched" signal spectrum resulting from inserting 3 zeros in-between the $D_q$ samples. The ideal interpolation filter frequency response, indicated by the dashed line, removes the copies of the original spectrum created by stretching, (c) the spectrum of the interpolated signal.

4.12 (a) Shows the magnitude response of a $20^{th}$ order min-max interpolation filter used in evaluating the $f(\alpha)$ singularity spectrum for the Cantor set. Filter specifications are $F_p = 0.003$, $F_s = 0.246$, $\delta_p = 0.1$, $\delta_s = 1 \times 10^{-6}$, $\beta_p = 10$ and $\beta_s = 1$. (b) shows the magnitude response of a $6^{th}$ order min-max differentiator designed using the Park-McClellan program for the Cantor set example. Filter specifications are $F_p = 0.1$, $F_s = 0.25$, $\delta_p = 0.1$, $\delta_s = 1 \times 10^{-6}$, $\beta_p = 10$ and $\beta_s = 1$.

4.13 (a) Shows the absolute value in dB of the Fourier transform of the decimated $D_q$ Cantor set example. In this case, the decimation factor is 4. The frequency axis is normalized by $\frac{2\pi}{\Delta q'}$, for $\Delta q' = 4$. (b) shows the absolute value in dB of the Fourier transform of the "stretched" $D_q$ of Cantor set example plotted in dB. The frequency axis is normalized by $\frac{2\pi}{\Delta q}$, for $\Delta q = 1$. $\Delta q' = 4 \Delta q$.

4.14 (a) Min-Max interpolated generalized dimensions of the Cantor set using the coarsely sampled $D_q$ in Fig. 4.9 (a). * = decimated $D_q$ samples in Fig. 4.9 (a) with $\Delta q = 4$; + = min-max interpolated samples with $L = 4$ and $\Delta q' = 1$. (b) $f(\alpha)$ of the optimally interpolated generalized dimensions $D_q$ in Fig. 4.14 (a) using min-max interpolator and differentiator.

5.1 Shows an unvoiced fricative "s" from the ISOLET database. Notice the distinct regions of silence, quasi-periodicity, noise-like region and silence. ($T = 6.25 \times 10^{-5}$ sec.)

5.2 A length 1000 sample segment of the unvoiced fricative "s" starting at sample number 6001 in Fig 5.1. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.)
5.3 Shows the normalized mutual information as a function of time lag \( \tau \) for the unvoiced speech sound \"s\" from the ISOLET database. The plot is normalized by the mutual information at lag zero. The mutual information at lag zero is 2.406866 bits.

5.4 "Normal" example. (a) Shows the log of the correlation integral plotted as a function of \( \log \epsilon \) for different embedding dimensions \( d_E \) from 5 to 15. (b) depicts the corresponding local slopes curves which are estimates of the correlation dimension \( D_2 \) as a function of \( d_E \). These estimates were made using the unvoiced speech data \"z\" for a female speaker, taken from the ISOLET (/ISOLET/isolet1/fec0) database. The number of speech samples \( N \) is 7184 starting at sample number 1316. Using the plateau region between \( \log \epsilon \in [6.9, 7.8] \), the estimated correlation dimension \( D_2 \) is 1.1825.

5.5 "Normal" example. (a) Shows the log of the correlation integral plotted as a function of \( \log \epsilon \) for different embedding dimensions \( d_E \) from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension \( D_2 \) as a function of \( d_E \). These estimates were computed using the unvoiced speech data \"z\" for a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjc1) database. The number of speech samples \( N \) is 5040 starting at sample number 2182. Using the plateau region between \( \log \epsilon \in [7.1, 8.0] \), the estimated correlation dimension \( D_2 \) is 1.4484.

5.6 "Normal" example. (a) Shows the log of the correlation integral plotted as a function of \( \log \epsilon \) for different embedding dimensions \( d_E \) from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension \( D_2 \) as a function of \( d_E \). These estimates were computed using the unvoiced speech data \"s\" for a female speaker, taken from the ISOLET (/ISOLET/isolet1/fme0) database. The number of speech samples \( N \) is 4324 starting at sample number 4849. Using the plateau region between \( \log \epsilon \in [6.5, 7.5] \), the estimated correlation dimension \( D_2 \) is 3.1363.
5.7 "Normal" example. (a) Shows the log of the correlation integral plotted as a function of $\log \epsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension $D_2$ as a function of $d_E$. These estimates were made using the unvoiced speech data "s" for a male speaker, taken from the ISOLET (/ISOLET/isolet1/mnjh0) database. The number of speech samples $N$ is 7275 starting at sample number 4429. Using the plateau region between $\log \epsilon \in [6.5, 7.2]$, the estimated correlation dimension $D_2$ is 4.4251.

5.8 "Multiscale" example. (a) Shows the log of the correlation integral plotted as a function of $\log \epsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension $D_2$ as a function of $d_E$. These estimates were made using the unvoiced speech data "z" for a female speaker, taken from the ISOLET (/ISOLET/isolet1/fmbd0) database. The number of speech samples $N$ is 8841 starting at sample number 1441. Using the plateau region between $\log \epsilon \in [7.0, 7.8]$, the estimated correlation dimension $D_2$ is 1.0867. Notice the other potential plateau region at $\log \epsilon \in [4.8, 6.0]$ typical of "multiscale examples".

5.9 "Multiscale" example. (a) Shows the log of the correlation integral plotted as a function of $\log \epsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension $D_2$ as a function of $d_E$. These estimates were made using the unvoiced speech data "s" for a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjrs0) database. The number of speech samples $N$ is 6014 starting at sample number 4720. Using the plateau region between $\log \epsilon \in [6.8, 7.5]$, the estimated correlation dimension $D_2$ is 2.583. Notice the other potential plateau region at $\log \epsilon \in [5.7, 6.0]$ typical of "multiscale examples".
5.10 "Multiscale" example. (a) Shows the log of the correlation integral plotted as a function of log $\epsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension $D_2$ as a function of $d_E$. These estimates were made using the unvoiced speech data \"s\" for a male speaker, taken from the ISOLET (/ISOLET/isolet1/mrmh1) database. The number of speech samples $N$ is 3713 starting at sample number 4628. Using the plateau region between log $\epsilon \in [7.0, 7.5]$, the estimated correlation dimension $D_2$ is 2.9269. Notice the other potential plateau region at log $\epsilon \in [6.0, 6.5]$ typical of "multiscale" examples.

5.11 "Lacunarity" example. (a) Shows the log of the correlation integral plotted as a function of log $\epsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension $D_2$ as a function of $d_E$. These estimates were made using the unvoiced speech data \"z\" for a female speaker, taken from the ISOLET (/ISOLET/isolet1/fet0) database. The number of speech samples $N$ is 7400 starting at sample number 2882. Using the plateau region between log $\epsilon \in [5.5, 6.0]$, the estimated correlation dimension $D_2$ is 4.1646.

5.12 "Lacunarity" example. (a) Shows the log of the correlation integral plotted as a function of log $\epsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension $D_2$ as a function of $d_E$. These estimates were made using the unvoiced speech data \"s\" for a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjc1) database. The number of speech samples $N$ is 2723 starting at sample number 4364. Using the plateau region between log $\epsilon \in [6.9, 7.5]$, the estimated correlation dimension $D_2$ is 3.31.
5.13 "Lacunarity" example. (a) Shows the log of the correlation integral plotted as a function of \( \log \epsilon \) for different embedding dimensions \( d_E \) from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension \( D_2 \) as a function of \( d_E \). These estimates were made using the unvoiced speech data \( \text{"z"} \) for a female speaker, taken from the ISOLET (/ISOLET/isolet1/fdcf0) database. The number of speech samples \( N \) is 7667 starting at sample number 785. Using the plateau region between \( \log \epsilon \in [5.5, 7.0) \), the estimated correlation dimension \( D_2 \) is 3.3562.

5.14 "Mixed" example. (a) Shows the log of the correlation integral plotted as a function of \( \log \epsilon \) for different embedding dimensions \( d_E \) from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension \( D_2 \) as a function \( d_E \). These estimates were made using the unvoiced speech data \( \text{"z"} \) for a female speaker, taken from the ISOLET (/ISOLET/isolet1/fmbd0) database. The number of speech samples \( N \) is 8841 starting at sample number 1441. Using the plateau region between \( \log \epsilon \in [7.0, 7.6) \), the estimated correlation dimension \( D_2 \) is 1.0867. This is a "mixed" example because of the lacunarity exhibited between \( \log \epsilon \in [4.5, 5.5) \) in addition to the normal plateau region \( \log \epsilon \in [7.0, 7.6] \).

5.15 "Mixed" example. (a) Shows the log of the correlation integral plotted as a function of \( \log \epsilon \) for different embedding dimensions \( d_E \) from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension \( D_2 \) as a function \( d_E \). These estimates were made using the unvoiced speech data \( \text{"z"} \) for a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjfv0) database. The number of speech samples \( N \) is 3210 starting at sample number 2665. Using the plateau region between \( \log \epsilon \in [7.0, 7.9) \), the estimated correlation dimension \( D_2 \) is 0.4749. This is a "mixed" example because of the lacunarity exhibited between \( \log \epsilon \in [4.5, 5.3) \) in addition to the normal plateau region \( \log \epsilon \in [7.0, 7.9] \).
5.16 (a) Shows the log of the correlation integral plotted as a function of $\log \varepsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension, $D_2$ as a function of $d_E$. The estimates are computed for the unvoiced speech data "s" for a female speaker, taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. The number of speech samples $N$ is 4158 starting at 3810. Observe the ambiguity (artifacts) in the value of $D_2$ created by the overlapping or crossing of the local slope curves at higher embedding dimensions near the range of $\log \varepsilon \in [6.8, 7.8]$. 

5.17 (a) Shows the log of the correlation integral plotted as a function of $\log \varepsilon$ for different embedding dimensions $d_E$ from 5 to 8 for the same speaker as in Fig. 5.16. (b) depicts the corresponding local slope curves also plotted as a function of $\log \varepsilon$ and $d_E$. Notice the four distinct regions discussed in section 2.13 and a better resolution of the ambiguity in Fig. 5.16 in the plateau region located near $\log \varepsilon \in [6.6, 7.4]$. For this sound, the number of speech samples $N$ is 2665 starting at sample number 3210. The estimated correlation dimension $D_2$ is 2.5542. 

5.18 (a) Shows the log of the correlation integral plotted as a function of $\log \varepsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension, $D_2$ as a function of $d_E$. The estimates are computed for the unvoiced speech data "s" for a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjfv0) database. The number of samples $N$ is 4820 starting at sample number 858. Observe the ambiguity (artifacts) in the interval $\log \varepsilon \in [6.0, 6.8]$ created by overlapping local slope curves at higher embedding dimensions.
5.19 (a) Shows the log of correlation integral plotted as a function of log $\epsilon$ for different embedding dimensions $d_E$ from 5 to 8 for the same speaker as in Fig. 5.18. (b) depicts the corresponding local slope curves also plotted as a function of log $\epsilon$ and $d_E$. Notice the four distinct regions discussed in section 2.13 and a better resolution of the ambiguity in Fig. 5.18 (b) in the plateau region located near log $\epsilon \in [6.0, 6.5]$. For this sound, the number of speech samples $N$ is 4820 starting at sample number 858. The estimated correlation dimension $D_2$ is 2.2468.

5.20 (a)-(d) Shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. The speech data are from both male and female speakers taken from the ISOLET database (/ISOLET/isolet1/mtdw0,fcmg0,mrmh1,fec0, respectively). Notice that in all cases, there exists regions of soft turbulence between the noisy and plateau regions, marked by the vertical dash lines. In each subfigure, the vertical-axis corresponds to the correlation dimension $D_2$ whereas the horizontal-axis corresponds to log $\epsilon$.

5.21 (a)-(d) Show the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in Fig. 5.20 (a)-(d), respectively. In each figure, the vertical-axis corresponds to the correlation dimension $D_2$, whereas the horizontal-axis corresponds to embedding dimension $d_E$.

5.22 (a) shows the local slope curves of the correlation integral as a function of log $\epsilon$ and $d_E$ of the speech sound "z" spoken by a female speaker from the ISOLET (/ISOLET/isolet1/fec0) database. (b) Cross-sectional slices of the local slope curves through the plateau region, the soft-turbulent region and the noise region are displayed, aa corresponds to the plateau region in Fig. 5.22 (a) located at log $\epsilon = 7.0$, bb corresponds to the soft-turbulent region at log $\epsilon = 6.0$ and cc is the noisy region at log $\epsilon = 4.5$. 
5.23 Shows the "normal" generalized dimensions $D_q$ versus $q$ curve for an unvoiced speech sound \"s\" uttered by a female speaker. The speech data are from the ISOLET (/ISOLET/isolet1/fcmc0) database.

5.24 Shows the "normal" generalized dimensions $D_q$ versus $q$ curve for an unvoiced speech sound \"z\" uttered by a female speaker. The speech data are from the ISOLET (/ISOLET/isolet1/fdcf0) database.

5.25 Shows the generalized dimensions $D_q$ versus $q$ curve for an unvoiced speech sound \"z\" uttered by a male speaker. The speech data are from the ISOLET (/ISOLET/isolet1/mtdw0) database.

5.26 Schematic plot of the generalized dimensions $D_q$ versus $q$ showing the phase-transition for of Logistic Map [3]. The discontinuity in the derivative of the $D_q$ occurs at $q = 2$.

5.27 Shows a "phase transition" in the estimated values of the generalized dimension $D_q$ versus $q$ for an unvoiced speech sound \"z\" uttered by a male speaker. The speech data are taken from the ISOLET (/ISOLET/isolet1/mrhm1) database. Notice the change in the $D_q$ curve around $q = 0$.

5.28 Shows a "phase transition" in the estimated values of the generalized dimension $D_q$ versus $q$ for an unvoiced speech sound \"s\" uttered by a male speaker. The speech data are taken from the ISOLET (/ISOLET/isolet1/mrhm1) database. Notice the change in the $D_q$ curve around $q = 1$.

5.29 Shows an "abnormal" curve of estimated generalized dimensions $D_q$ versus $q$ values for an unvoiced speech \"s\" uttered by a male speaker. The speech data are taken from the ISOLET (/ISOLET/isolet1/mrs0) database. This is an invalid $D_q$ curve. The expression in equation (5.6) that requires $D_q$ to be monotonic decreasing has been violated around $q < -3$. 

xxiv
5.30 Shows an "abnormal" curve of estimated generalized dimensions $D_q$ versus $q$ values for an unvoiced speech \"z\" uttered by a female speaker. The speech data are taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. This is an invalid $D_q$ curve. The expression in equation (5.6) that requires monotonic decreasing $D_q$ has been violated around $q < -2$ and $q < -1$. ..................... 156

5.31 Shows an "abnormal" curve of estimated generalized dimensions $D_q$ versus $q$ values for an unvoiced speech \"z\" uttered by a male speaker. The speech data are taken from the ISOLET (/ISOLET/isolet1/mrs0) database. This is an invalid $D_q$ curve. The expression in equation (5.6) that requires monotonic decreasing $D_q$ has been violated around $|q| < 1$. ..................... 157

5.32 (a) Shows the optimally interpolated generalized dimensions (GD) $D_q, |q| = 3$ of an unvoiced speech signal \"s\". The speech data for a female speaker are taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. Fig. 5.32(b) shows the corresponding $f(\alpha)$ singularity spectrum using the min-max filter design method for both interpolation and differentiation filters. From the $f(\alpha)$ curve, $D_0 = 6.1$ and the estimated $D_{-\infty}$ and $D_\infty$ values are 9.25 and 1.4, respectively. ........ 159

5.33 Shows the $f(\alpha)$ singularity spectrum using the min-max filter design method on the generalized dimensions (GD) $D_q, |q| = 5$ of an unvoiced speech signal \"z\". The speech data for a male speaker are taken from the ISOLET (/ISOLET/isolet1/mdtw0) database. ........ 160

5.34 Shows the $f(\alpha)$ singularity spectrum using the min-max filter design method on the generalized dimensions (GD) $D_q, |q| = 3$ of the same speech data as the one used in estimating the $D_q$ in Fig. 5.33. The speech data for a male speaker are taken from the ISOLET (/ISOLET/isolet1/mdtw0) database. ..................... 161
5.35 (a) Shows the $f(\alpha)$ singularity spectrum using the conventional techniques which uses linear approximation and a first difference approximation to differentiation. (b) shows the $f(\alpha)$ singularity spectrum using the min-max filter design method on the generalized dimensions (GD) $D_q$, in the range $|q| = 3$. The speech data \"s\" for a female speaker are taken from the ISOLET (/ISOLET/isolet1/mdtw0) database.

5.36 (a) Shows the optimally interpolated generalized dimensions $D_q$, $|q| = 3$ of an unvoiced speech signal \"z\". The speech data for a female speaker are taken from the ISOLET (/ISOLET/isolet1/fdcf0) database. (b) shows the corresponding $f(\alpha)$ singularity spectrum using the min-max filter design method. From the $f(\alpha)$ curve, we obtain $D_0 = 5.03$ and the estimated $D_{-\infty}$ and $D_{\infty}$ values are 9.6 and 1.8, respectively.

5.37 (a) Shows the optimally interpolated generalized dimensions $D_q$, $|q| = 3$ of an unvoiced speech signal \"z\". The speech data for a male speaker are taken from the ISOLET (/ISOLET/isolet1/mdtw0) database. (b) shows the corresponding $f(\alpha)$ singularity spectrum using the min-max filter design method. From the $f(\alpha)$ curve, we obtain $D_0 = 4.25$ and the estimated $D_{-\infty}$ and $D_{\infty}$ values are 6.35 and 1.85, respectively.

5.38 (a) Shows the optimally interpolated generalized dimensions $D_q$, $|q| = 3$ of an unvoiced speech signal \"z\". The speech data for a male speaker are taken from the ISOLET (/ISOLET/isolet1/mrmh1) database. Fig. (b) shows the corresponding $f(\alpha)$ singularity spectrum using the min-max filter design method. From the $f(\alpha)$ curve, we obtain $D_0 = 3.73$ and the estimated $D_{-\infty}$ and $D_{\infty}$ values are 8.7 and 0.9, respectively.
5.39 (a) Shows the optimally interpolated generalized dimensions $D_q$, \( q \neq 3 \) of an unvoiced speech signal \( \text{"s"} \). The speech data for a male speaker are taken from the ISOLET (/ISOLET/isolet1/mrmh1) database. Fig. 5.39(b) shows the corresponding $f(\alpha)$ singularity spectrum using the min-max filter design method. From the $f(\alpha)$ curve, we obtain $D_0 = 4.45$ and the estimated $D_{-\infty}$ and $D_{\infty}$ values are 8.5 and 1.5, respectively. ................. 168

C.1 (a) Shows the log-log plot of the generalized correlation integrals $C_{q,d\varepsilon}(\varepsilon)$ as a function of the volume element $\varepsilon$ for embedding dimension $d\varepsilon$ ranging from 5 to 15. In this plot, $q = -5$. The unvoiced speech \( \text{"s"} \) spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. .................. 184

C.2 (a) Shows the log-log plot of the generalized correlation integrals $C_{q,d\varepsilon}(\varepsilon)$ as a function of the volume element $\varepsilon$ for embedding dimension $d\varepsilon$ ranging from 5 to 15. In this plot, $q = -4$. The unvoiced speech \( \text{"s"} \) spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. .................. 185

C.3 (a) Shows the log-log plot of the generalized correlation integrals $C_{q,d\varepsilon}(\varepsilon)$ as a function of the volume element $\varepsilon$ for embedding dimension $d\varepsilon$ ranging from 5 to 15. In this plot, $q = -3$. The unvoiced speech \( \text{"s"} \) spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. .................. 186

C.4 (a) Shows the log-log plot of the generalized correlation integrals $C_{q,d\varepsilon}(\varepsilon)$ as a function of the volume element $\varepsilon$ for embedding dimension $d\varepsilon$ ranging from 5 to 15. In this plot, $q = -2$. The unvoiced speech \( \text{"s"} \) spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. .................. 187
C.5 (a) Shows the log-log plot of the generalized correlation integrals
\( C_{q,dE}(\epsilon) \) as a function of the volume element \( \epsilon \) for embedding dimension \( d_E \) ranging from 5 to 15. In this plot, \( q = -1 \). The unvoiced speech "s" spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. .......................... 188

C.6 (a) Shows the log-log plot of the generalized correlation integrals
\( C_{q,dE}(\epsilon) \) as a function of the volume element \( \epsilon \) for embedding dimension \( d_E \) ranging from 5 to 15. In this plot, \( q = 0 \). The unvoiced speech "s" spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. .......................... 189

C.7 (a) Shows the log-log plot of the generalized correlation integrals
\( C_{q,dE}(\epsilon) \) as a function of the volume element \( \epsilon \) for embedding dimension \( d_E \) ranging from 5 to 15. In this plot, \( q = 1 \). The unvoiced speech "s" spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. .......................... 190

C.8 (a) Shows the log-log plot of the generalized correlation integrals
\( C_{q,dE}(\epsilon) \) as a function of the volume element \( \epsilon \) for embedding dimension \( d_E \) ranging from 5 to 15. In this plot, \( q = 2 \). The unvoiced speech "s" spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. .......................... 191

C.9 (a) Shows the log-log plot of the generalized correlation integrals
\( C_{q,dE}(\epsilon) \) as a function of the volume element \( \epsilon \) for embedding dimension \( d_E \) ranging from 5 to 15. In this plot, \( q = 3 \). The unvoiced speech "s" spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. .......................... 192

C.10 (a) Shows the log-log plot of the generalized correlation integrals
\( C_{q,dE}(\epsilon) \) as a function of the volume element \( \epsilon \) for embedding dimension \( d_E \) ranging from 5 to 15. In this plot, \( q = 4 \). The unvoiced speech "s" spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. .......................... 193
C.11 (a) Shows the log-log plot of the generalized correlation integrals $C_{q,d_E}(\epsilon)$ as a function of the volume element $\epsilon$ for embedding dimension $d_E$ ranging from 5 to 15. In this plot, $q = 5$. The unvoiced speech "s" spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. ................. 194

D.1 (a) Shows an unvoiced fricative "s" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. (b) shows an isolated, truncated length 4158 sample segment of the unvoiced fricative "s" starting at sample number 3810. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.) ... 196

D.2 (a) Shows an unvoiced fricative "s" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fcmg0) database. (b) shows an isolated, truncated length 4520 sample segment of the unvoiced fricative "s" starting at sample number 4380. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.) ... 198

D.3 (a) Shows an unvoiced fricative "s" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fdcf0) database. (b) shows an isolated, truncated length 4711 sample segment of the unvoiced fricative "s" starting at sample number 5210. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.) ... 200

D.4 (a) Shows an unvoiced fricative "s" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fec0) database. (b) shows an isolated, truncated length 3352 sample segment of the unvoiced fricative "s" starting at sample number 4100. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.) ... 202

D.5 (a) Shows an unvoiced fricative "s" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fet0) database. (b) shows an isolated, truncated length 4990 sample segment of the unvoiced fricative "s" starting at sample number 4323. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.) ... 204
D.6 (a) Shows an unvoiced fricative \"s\" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fkh0) database. (b) shows an isolated, truncated length 4300 sample segment of the unvoiced fricative \"s\" starting at sample number 4318. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5}\) sec. ..... 206

D.7 (a) Shows an unvoiced fricative \"s\" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fmb0) database. (b) shows an isolated, truncated length 4934 sample segment of the unvoiced fricative \"s\" starting at sample number 3879. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5}\) sec. ..... 208

D.8 (a) Shows an unvoiced fricative \"s\" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fmb0) database. (b) shows an isolated, truncated length 3908 sample segment of the unvoiced fricative \"s\" starting at sample number 4849. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5}\) sec. ..... 210

D.9 (a) Shows an unvoiced fricative \"s\" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fme0) database. (b) shows an isolated, truncated length 4324 sample segment of the unvoiced fricative \"s\" starting at sample number 4849. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5}\) sec. ..... 212

D.10 (a) Shows an unvoiced fricative \"s\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjc1) database. (b) shows an isolated, truncated length 2723 sample segment of the unvoiced fricative \"s\" starting at sample number 4364. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5}\) sec. ..... 214

D.11 (a) Shows an unvoiced fricative \"s\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjfv0) database. (b) shows an isolated, truncated length 4434 sample segment of the unvoiced fricative \"s\" starting at sample number 3658. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5}\) sec. ..... 216

xxx
D.12  (a) Shows an unvoiced fricative \“s\” spoken by a male speaker, taken from the ISOLET (/ISOLET/iselet1/mjp0) database. (b) shows an isolated, truncated length 4739 sample segment of the unvoiced fricative \“s\” starting at sample number 4323. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5} \text{ sec.}\) ............... 218

D.13  (a) Shows an unvoiced fricative \“s\” spoken by a male speaker, taken from the ISOLET (/ISOLET/iselet1/mnjh0) database. (b) shows an isolated, truncated length 7275 sample segment of the unvoiced fricative \“s\” starting at sample number 4429. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5} \text{ sec.}\) ......... 220

D.14  (a) Shows an unvoiced fricative \“s\” spoken by a male speaker, taken from the ISOLET (/ISOLET/iselet1/mnre0) database. (b) shows an isolated, truncated length 4601 sample segment of the unvoiced fricative \“s\” starting at sample number 4655. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5} \text{ sec.}\) ............... 222

D.15  (a) Shows an unvoiced fricative \“s\” spoken by a male speaker, taken from the ISOLET (/ISOLET/iselet1/mrmh1) database. (b) shows an isolated, truncated length 3713 sample segment of the unvoiced fricative \“s\” starting at sample number 4628. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5} \text{ sec.}\) ............... 224

D.16  (a) Shows an unvoiced fricative \“s\” spoken by a male speaker, taken from the ISOLET (/ISOLET/iselet1/mrs0) database. (b) shows an isolated, truncated length 3215 sample segment of the unvoiced fricative \“s\” starting at sample number 4460. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5} \text{ sec.}\) ............... 226

D.17  (a) Shows an unvoiced fricative \“s\” spoken by a male speaker, taken from the ISOLET (/ISOLET/iselet1/msa0) database. (b) shows an isolated, truncated length 3927 sample segment of the unvoiced fricative \“s\” starting at sample number 4133. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5} \text{ sec.}\) ............... 228
D.18 (a) Shows an unvoiced fricative "s" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mtdw0) database. (b) shows an isolated, truncated length 7247 sample segment of the unvoiced fricative "s" starting at sample number 4572. Notice the noise-like appearance of the speech signal. (T = 6.25 x 10^{-5} sec.) . . . . . . . . . . . . 230

D.19 (a) Shows voiced fricative "z" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. (b) shows an isolated, truncated length 9587 sample segment of the voiced fricative "z" starting at sample number 1552. Notice the noise-like appearance of the speech signal. (T = 6.25 x 10^{-5} sec.) . . . . . . . . . . . . 232

D.20 (a) Shows voiced fricative "z" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fcmg0) database. (b) shows an isolated, truncated length 4050 sample segment of the voiced fricative "z" starting at sample number 3950. Notice the noise-like appearance of the speech signal. (T = 6.25 x 10^{-5} sec.) . . . . . . . . . . . . 234

D.21 (a) Shows voiced fricative "z" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fdcf0) database. (b) shows an isolated, truncated length 7667 sample segment of the voiced fricative "z" starting at sample number 785. Notice the noise-like appearance of the speech signal. (T = 6.25 x 10^{-5} sec.) . . . . . . . . . . . . 236

D.22 (a) Shows voiced fricative "z" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fec0) database. (b) shows an isolated, truncated length 7184 sample segment of the voiced fricative "z" starting at sample number 1316. Notice the noise-like appearance of the speech signal. (T = 6.25 x 10^{-5} sec.) . . . . . . . . . . . . 238

D.23 (a) Shows voiced fricative "z" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fet0) database. (b) shows an isolated, truncated length 7400 sample segment of the voiced fricative "z" starting at sample number 2882. Notice the noise-like appearance of the speech signal. (T := 6.25 x 10^{-5} sec.) . . . . . . . . . . . . 240
D.24 (a) Shows voiced fricative \"z\" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fews0) database. (b) shows an isolated, truncated length 7677 sample segment of the voiced fricative \"z\" starting at sample number 1662. Notice the noise-like appearance of the speech signal. \( T = 6.25 \times 10^{-5} \) sec. . . . . . . . . . . . . 242

D.25 (a) Shows voiced fricative \"z\" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fkh0) database. (b) shows an isolated, truncated length 8592 sample segment of the voiced fricative \"z\" starting at sample number 1468. Notice the noise-like appearance of the speech signal. \( T = 6.25 \times 10^{-5} \) sec. . . . . . . . . . . . . 244

D.26 (a) Shows voiced fricative \"z\" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. (b) shows an isolated, truncated length 7843 sample segment of the voiced fricative \"z\" starting at sample number 1468. Notice the noise-like appearance of the speech signal. \( T = 6.25 \times 10^{-5} \) sec. . . . . . . . . . . . . 246

D.27 (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjc1) database. (b) shows an isolated, truncated 5404 length sample segment of the voiced fricative \"z\" starting at sample number 2182. \( T = 6.25 \times 10^{-5} \) sec. . . . . 248

D.28 (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjfv0) database. (b) shows an isolated, truncated 3210 length sample segment of the voiced fricative \"z\" starting at sample number 2665. \( T = 6.25 \times 10^{-5} \) sec. . . . . 250

D.29 (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjp0) database. (b) shows an isolated, truncated 8092 length sample segment of the voiced fricative \"z\" starting at sample number 1358. \( T = 6.25 \times 10^{-5} \) sec. . . . . 252

D.30 (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjrs0) database. (b) shows an isolated, truncated 8591 length sample segment of the voiced fricative \"z\" starting at sample number 1108. \( T = 6.25 \times 10^{-5} \) sec. . . . . 254
D.31 (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mnjh0) database. (b) shows an isolated, truncated 8601 sample length segment of the voiced fricative \"z\" starting at sample number 1544. \(T = 6.25 \times 10^{-5}\) sec. . . . 256

D.32 (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mnre0) database. (b) shows an isolated, truncated 6758 length sample segment of the voiced fricative \"z\" starting at sample number 1690. \(T = 6.25 \times 10^{-5}\) sec. . . . 258

D.33 (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mrmh1) database. (b) shows an isolated, truncated 5705 length sample segment of the voiced fricative \"z\" starting at sample number 1616. \(T = 6.25 \times 10^{-5}\) sec. . . . 260

D.34 (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mrs0) database. (b) shows an isolated, truncated 7538 length sample segment of the voiced fricative \"z\" starting at sample number 1524. \(T = 6.25 \times 10^{-5}\) sec. . . . 262

D.35 (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/msa0) database. (b) shows an isolated, truncated 7205 length sample segment of the voiced fricative \"z\" starting at sample number 1718. \(T = 6.25 \times 10^{-5}\) sec. . . . 264

D.36 (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mtdw0) database. (b) shows an isolated, truncated 10359 length sample segment of the voiced fricative \"z\" starting at sample number 1385. \(T = 6.25 \times 10^{-5}\) sec. . . . 266
# LIST OF SYMBOLS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_c$</td>
<td>Area of Constriction in the vocal tract</td>
</tr>
<tr>
<td>$a[k]$</td>
<td>Filter coefficients, coefficients of Remez-Exchange algorithm</td>
</tr>
<tr>
<td>$b$</td>
<td>Constant</td>
</tr>
<tr>
<td>$c$</td>
<td>Periodic cycles</td>
</tr>
<tr>
<td>$d$</td>
<td>Hausdorff Dimension</td>
</tr>
<tr>
<td>$dd(t)$</td>
<td>Separation between two trajectories after a time $t$</td>
</tr>
<tr>
<td>$d_E$</td>
<td>Embedding Dimension</td>
</tr>
<tr>
<td>$f(\alpha)$</td>
<td>Singularity Spectrum</td>
</tr>
<tr>
<td>$g$</td>
<td>Functional relationship between vectors</td>
</tr>
<tr>
<td>$gw[n]$</td>
<td>Gaussian Noise</td>
</tr>
<tr>
<td>$h$</td>
<td>Converged embedding dimension</td>
</tr>
<tr>
<td>$h[k]$</td>
<td>Impulse response of a lowpass filter</td>
</tr>
<tr>
<td>$i, j, k, l, n$</td>
<td>Discrete time index</td>
</tr>
<tr>
<td>$m$</td>
<td>Order of Remez-Exchange filter</td>
</tr>
<tr>
<td>$p$</td>
<td>Order of AR process</td>
</tr>
<tr>
<td>$p_i$ or $p_i()$</td>
<td>probability associated with either a random variable or volume element in phase-space (pdf)</td>
</tr>
<tr>
<td>$q$</td>
<td>Order of the generalized correlation, dimension, or entropy</td>
</tr>
<tr>
<td>$r$</td>
<td>Tuning parameter of the logistic map</td>
</tr>
<tr>
<td>$r_c$</td>
<td>Value of $r$ where chaotic behavior starts in the Logistic Map</td>
</tr>
<tr>
<td>$s[n]$</td>
<td>AR Process</td>
</tr>
<tr>
<td>$t$</td>
<td>Continuous time</td>
</tr>
<tr>
<td>$u_c$</td>
<td>Volume Velocity</td>
</tr>
<tr>
<td>$u(t), v(t)$</td>
<td>Continuous time signal</td>
</tr>
<tr>
<td>$w, w_0$</td>
<td>Adjustable weights</td>
</tr>
<tr>
<td>$x^*$</td>
<td>Fixed point of the logistic map</td>
</tr>
<tr>
<td>$x, y, z$</td>
<td>Cartesian Coordinates</td>
</tr>
<tr>
<td>$x[n]$</td>
<td>Discrete Time Signal</td>
</tr>
<tr>
<td>$\dot{x}, \dot{y}, \dot{z}$</td>
<td>Derivative with respect to time</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>(A)</td>
<td>Attractor of a chaotic system</td>
</tr>
<tr>
<td>(A_t)</td>
<td>Amplitude of a Signal</td>
</tr>
<tr>
<td>(C_2(\varepsilon))</td>
<td>Correlation Integral</td>
</tr>
<tr>
<td>(C_{q}(\varepsilon))</td>
<td>Generalized Correlation Integrals</td>
</tr>
<tr>
<td>(D)</td>
<td>Manifold Dimension on which an attractor is generated or projected</td>
</tr>
<tr>
<td>(D_2)</td>
<td>Correlation Dimension</td>
</tr>
<tr>
<td>(D_{q})</td>
<td>Generalized (q^{th}) Correlation Dimension</td>
</tr>
<tr>
<td>(E(w))</td>
<td>Mean Squared Error</td>
</tr>
<tr>
<td>(F_0)</td>
<td>Fundamental Frequency</td>
</tr>
<tr>
<td>(F_p, F_s, F_t)</td>
<td>Edges of normalized Frequency band in a lowpass filter</td>
</tr>
<tr>
<td>(G)</td>
<td>Discrete Random variable</td>
</tr>
<tr>
<td>(G_h)</td>
<td>Model Gain Predictive Filter</td>
</tr>
<tr>
<td>(H_U())</td>
<td>Uncertainty of the random variable (U)</td>
</tr>
<tr>
<td>(H(Z))</td>
<td>Vocal Tract Transfer Function</td>
</tr>
<tr>
<td>(H(\omega))</td>
<td>Frequency response</td>
</tr>
<tr>
<td>(I(), I_r())</td>
<td>Mutual information between two random variables</td>
</tr>
<tr>
<td>(J(\alpha))</td>
<td>The number of volume elements with scaling exponent in the range from (\alpha) to (\alpha + d\alpha)</td>
</tr>
<tr>
<td>(K_2)</td>
<td>Entropy</td>
</tr>
<tr>
<td>(K)</td>
<td>Kolmogorov Entropy</td>
</tr>
<tr>
<td>(L)</td>
<td>Length between a fiducial point and its neighbors</td>
</tr>
<tr>
<td>(L^1)</td>
<td>Manhattan norm</td>
</tr>
<tr>
<td>(L^2)</td>
<td>Euclidean norm</td>
</tr>
<tr>
<td>(L^\infty)</td>
<td>Maximum norm</td>
</tr>
<tr>
<td>(M, M_E)</td>
<td>Arbitrary constant</td>
</tr>
<tr>
<td>(N)</td>
<td>Arbitrary Constant</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>(N(\varepsilon))</td>
<td>Number of non-empty volume elements covering an attractor</td>
</tr>
<tr>
<td>(P_I)</td>
<td>Ideal frequency response</td>
</tr>
<tr>
<td>(P_d)</td>
<td>Desired frequency response</td>
</tr>
<tr>
<td>(Q)</td>
<td>Arbitrary constant</td>
</tr>
<tr>
<td>(\mathbb{R}^N)</td>
<td>N-dimensional Euclidean space</td>
</tr>
<tr>
<td>(R)</td>
<td>Tuning parameter of the Lorenz attractor</td>
</tr>
<tr>
<td>(Re)</td>
<td>Reynolds number</td>
</tr>
<tr>
<td>(Re_{\text{critical}})</td>
<td>Critical Reynolds number</td>
</tr>
<tr>
<td>(S)</td>
<td>Shannon Entropy</td>
</tr>
<tr>
<td>(S_{ij})</td>
<td>(\sum_{j=1}^{M} \Theta(\varepsilon - |X_i - X_j|)) inter-point distances less than (\varepsilon)</td>
</tr>
<tr>
<td>(T)</td>
<td>Sampling interval (seconds)</td>
</tr>
<tr>
<td>(T_0)</td>
<td>Temperature</td>
</tr>
<tr>
<td>(U, V)</td>
<td>Discrete random variables</td>
</tr>
<tr>
<td>(W(w))</td>
<td>Weight function</td>
</tr>
<tr>
<td>(X_0)</td>
<td>Vector points or Embedded vectors in phase-space</td>
</tr>
<tr>
<td>(Z)</td>
<td>Z-transform variable</td>
</tr>
<tr>
<td>(\alpha)</td>
<td>Singularity strength or scaling exponents</td>
</tr>
<tr>
<td>(\beta_0)</td>
<td>Passband and stopband weights in filter design</td>
</tr>
<tr>
<td>(\delta_T)</td>
<td>Temperature difference</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td>Diameter or width of volume elements such as spheres, hyper-spheres, boxes, etc.</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>Spectral coefficient</td>
</tr>
<tr>
<td>(\lambda)</td>
<td>Lyapunov exponent</td>
</tr>
<tr>
<td>(\mu)</td>
<td>Fluid Viscosity</td>
</tr>
<tr>
<td>(\omega)</td>
<td>Angular frequency</td>
</tr>
<tr>
<td>(\omega_n)</td>
<td>Nyquist frequency</td>
</tr>
<tr>
<td>(\omega_c)</td>
<td>Cut-off frequency</td>
</tr>
</tbody>
</table>
LIST OF SYMBOLS cont.

\( \phi \)     Phase of a signal
\( \rho \)     Fluid Density
\( \sigma \)   Arbitrary Constant in Lorenz system
\( \sigma^2 \) Signal power
\( \sigma^2_{n\nu} \) Noise power
\( \tau \)     Time delay
\( \tau_{opt} \) Optimum delay time
\( \xi \)     Angular separation
\( \Theta() \) Heaviside or unit step function
\( | \ldots | \) Magnitude or absolute value
\( || \ldots || \) Norm of a vector
\( \Delta q \)  Increments of \( q \) in \( D_q \)
\( A(f) \)     Fourier transform of scaling exponent \( a(q) \)
\( D(f) \)     Fourier transform of the generalized dimension \( D_q \)
\( \dot{D}(f) \) Derivative of \( D(f) \) with respect to frequency
\( \ddot{D}(f) \) Second derivative of \( D(f) \) with respect to frequency
\( T(f) \)     Fourier transform of \( \tau(q) \)
Chapter 1

Introduction and Overview

I would therefore urge that people be introduced to the logistic equation earlier in their mathematical education. This equation can be studied phenomenologically by iterating it on a calculator, or even by hand. Its study does not involve as much conceptual sophistication as does elementary calculus. Such study would greatly enrich the student's intuition about nonlinear systems. Not only in research but also in everyday world of politics and economics, we would all be better off if more people realized that simple nonlinear systems do not necessarily possess simple dynamical properties.

Robert M. May

In this dissertation, a refinement is made to the nonlinear dynamical analysis of human speech signals. Previous research in this area has successfully characterized certain speech signals as chaotic. Evidence of low-dimensional chaos, which includes the existence of strange attractors, positive Lyapunov exponents, bifurcations and finite entropies, has been found in speech [14, 15, 16, 17, 18, 19, 20, 21, 22, 23].

Due to the complexity involved in speech production, in terms of fluid flow and nonlinear interaction within the vocal apparatus, we analyze unvoiced speech signals using multifractal measures to determine the presence of various fractal regimes in speech attractors. These fractal regimes (multifractal structure) are not revealed by

---

\footnote{The quote is from a review article on simple mathematical models [13].}
the estimation of only the correlation dimensions and Lyapunov exponents as done by Narayanan et al [14].

We therefore propose the estimation of the spectra of generalized dimensions to explain the multifractal nature of voiced and unvoiced fricative sounds. This method is based on the estimation of the generalized dimensions $D_q$ from the phase-space reconstructed vectors of single scalar realization of speech signals. We also propose an improved method of evaluating the singularity spectrum, $f(\alpha)$, from the estimated generalized dimensions $D_q$. The new method is based on the best min-max digital filter that approximates the interpolation and differentiation operations needed to implement the Legendre transformation relating the generalized dimensions $D_q$ to the singularity spectrum $f(\alpha)$. We later use this technique to determine the $f(\alpha)$ singularity spectrum for unvoiced speech signals.

1.1 Organization of Dissertation

This dissertation is organized as follows:

In chapter 2, the Logistic Map [13] and the Lorenz equations [24] are used to introduce chaotic parameters, concepts, and properties such as phase-space, invariant density, bifurcations, ergodicity, etc. [4]. We also discuss the generalization of the correlation dimension $D_2$ to an infinite hierarchy of generalized dimensions $D_q$. This is important as prior work by Narayanan et al [14] used only $D_2$ in the analysis of speech signals whereas we will evaluate unvoiced speech signals using all $D_q$. We later quantitatively describe chaos in terms of measures such as dimension, entropy and Lyapunov exponents. In section 2.5, systematic steps relevant to the estimation of generalized dimensions from a single component speech signal are introduced. These include the concepts of time delay embedding, correlation integral and local slopes.

Chapter 3 contains a brief overview of speech production and processing. In the chapter, we discuss the importance of speech in human beings as well as the different classes of sounds produced by human beings. In section 3.3, we describe the different methods that have been used in speech modeling in relation to their relative advantages and disadvantages. In particular, we describe why the popular Linear
Predictive Coding model is inadequate to explain the production of certain sounds, and how newer models using non-linearities and chaos concepts are intriguing. Towards the end of the section, we hypothesize the presence of different fractal regimes in unvoiced speech sounds. This is based in part on the success and inadequacies of previous nonlinear and chaos models.

In chapter 4, we examine another measure, the $f(\alpha)$ singularity spectrum, to characterize multifractal behavior on chaotic attractors. The triadic Koch curve will be used to illustrate how a multifractal may be generated and quantified using the invariant density measure of chaotic attractors. We also explore the relationship between the generalized dimensions $D_q$ and the $f(\alpha)$ singularity spectrum. We discuss some of the existing methods of using the Legendre transform of $D_q$ to compute $f(\alpha)$. We focus on computational complexity and introduction of erroneous artifacts. We therefore propose a new signal processing method of evaluating the Legendre transform in a more accurate fashion than popular techniques [12].

In chapter 5, we present the results of applying multifractal analysis to unvoiced speech sounds from the ISOLETE database [25]. This database consist of letters of the English alphabet spoken in isolation. There are about 7800 spoken letters; 2 productions of each letter is spoken by 150 speakers. We selected 40 male speakers and female speakers making the unvoiced speech sounds, “s” and “z”. The selection criteria used were availability of unvoiced speech samples and adequate number of speakers. The unvoiced speech sounds “s” and “z” were chosen so as to include both voiced and unvoiced fricatives. The broad speech selection is to ensure that the use of generalized dimensions $D_q$ to analyze unvoiced speech signals is statistically conclusive. For each speech sample, an estimation of the generalized dimensions $D_q$ as well as the singularity spectrum $f(\alpha)$ will be performed.

In chapter 6, we discuss the results of the multifractal analysis of chapter 5. The chapter ends with conclusions from the analysis and a brief description of future work in this area of research.
Chapter 2

Chaos and Dynamical Systems

2.1 Introduction

In the last 15 years, the study of chaos and nonlinear dynamical systems has provided theories and concepts to explain some of the strange behavior exhibited in large and complex systems such as weather prediction and biological population growth. This has been accomplished by the introduction of simple system of equations which mimic the strange behaviors found in complex systems. Chaos has provided scientists with tools to understand, visualize, and analyze behaviors using simple system of equations which may provide insight into some of the unexplained characteristics of larger systems. An example of such a complex system is speech which is characterized by many unanswered questions. Speech is a nonlinear phenomena that has been quantitatively described in terms of fluid flow and turbulence [26]. Recently, however, speech has been found to exhibit low-dimensional chaos [27, 23, 28, 14].

This dissertation seeks to obtain a more complete, low-dimensional chaotic description of speech signals. However, before springing into this large task, we focus on qualitatively defining chaos in terms of its properties and characteristics. We provide examples from both the Lorenz system and Logistic map. Later in this chapter, the problem of quantitatively describing chaos in terms of chaotic measures such as dimension, Lyapunov exponent and entropy is addressed.
2.1.1 What is Chaos?

Chaos is a "strange behavior" observed in either very simple, nonlinear or complex dynamical systems. The term chaos is used to describe the apparently random, aperiodic, but bounded, behavior of a deterministic system. A system is deterministic if its behavior over a long period of time is completely determined by the knowledge of the time evolution equations, the parameters describing the system and the initial conditions. With such knowledge, all values of the system over a long period of time or the steady state values can be computed or "determined" a priori. The fascination with chaos therefore arises from reconciling the determinism in its generation with the apparent randomness that is produced. Therefore, the formal definition of a chaotic system is a deterministic system that appears to exhibit random behaviors [2, 1, 29, 30, 4].

Chaotic systems exhibit sensitive dependence on initial conditions. This property is explained by the exponential divergence of the steady states (chaotic orbits) of chaotic system generated from two closely spaced initial conditions. That is, no matter how close are the initial conditions of separate implementations of a chaotic system, the trajectory of those implementations will eventually diverge exponentially, making them appear totally dissimilar. Chaotic signals are also characterized by broadband spectra and, therefore, rapidly decaying autocorrelation functions. These characteristics are reminiscent of random noise which makes it non-trivial for non-experts in chaos to distinguish chaos from apparent randomness [2].

However, it is fascinating and intriguing that chaos sometimes arises from very simple systems (with only a few degrees of freedom) which are noise free. In the next section, we describe an example of a relatively simple system of coupled nonlinear equations that has been used in explaining some complex behaviors found in weather patterns.

2.2 Lorenz System

One of the earliest example of chaos in a simple system that is capable of explaining complex behavior in a larger system was by Lorenz. In 1963, Edward Lorenz, the
father of modern chaos, gave one of the earliest descriptions of a chaotic process [24]. He developed a system of coupled-nonlinear differential equations to model weather patterns. The standard form of the Lorenz coupled equations is given as

\[ \begin{align*}
\dot{x} &= \sigma(y - x) \\
\dot{y} &= -y - xz + Rx \\
\dot{z} &= xy - bz
\end{align*} \]  

(2.1) 
(2.2) 
(2.3)

where \( R, \sigma, b \) are constants and \( \dot{x} \) indicates the derivative of \( x(t) \) with respect to time. \( x(t) \) is related to the time dependence of the fluid motion whereas \( y(t) \) and \( z(t) \) are related to the time dependence of the temperature deviations away from linear temperature drop from bottom to top plate. The Lorenz equation is generated on the 3-D Cartesian coordinates\(^1\). Lorenz used the simple setup shown in Fig. 2.1 to describe a physical system of weather patterns. The fluid system (atmosphere) was modeled as a fluid layer between two plates that are heated from the bottom (due to the effect of the sun rays on the earth’s surface) and cooled from the top. The bottom (warm) plate is kept constant at temperature \( T_w \) and the top (cool) plate temperature is maintained at \( T_c \). The temperature difference between the two plates \( \delta_T = T_w - T_c \).

The parameter \( b \) is related to the onset of convection in the fluid layer. In particular, when \( b = \frac{3}{2} \), convection begins for the smallest value of temperature difference \( \delta_T \). The parameter \( \sigma \) compares the rate of energy loss from the fluid due to viscosity to the rate of energy loss due to thermal conduction. Finally, \( R \) is proportional to the temperature difference between the bottom and top fluid layer, i.e., \( R \propto \delta_T = T_w - T_c \).

These kinds of experiments were first studied in 1900 by Bernard and a theoretical foundation for the experiment was later provided by Rayleigh in 1916, hence the famous Rayleigh-Bernard convection [1, 4]. Depending on the magnitude of \( \delta_T \), the model is capable of predicting various weather patterns. For example, when \( \delta_T \) is not too large, heat is transferred from the bottom plate to the top plate by thermal conduction\(^2\) whereas when \( \delta_T \) is very large, heat is transferred to the top plate by convection\(^3\). Other fluid circulation patterns could also be generated by tweaking

\(^1\)This is also referred to as the 3-dimensional manifold of the attractor.
\(^2\)Conduction is the transfer of heat due to the temperature gradient of a fluid at rest.
\(^3\)Convection is the transfer of heat by fluid motion between regions of unequal density that may result from non-uniform heating.
the parameter $R$. The phenomena experienced in the Rayleigh-Bernard experiments were also noticed in the earth’s atmospheric patterns [24]. Since the equations in (2.1-2.3) are sensitive to initial conditions, a meteorologist cannot predict the weather accurately for long periods of time.

Unfortunately for the science of chaos, the unprecedented work of Lorenz was published in a prestigious but obscure journal where it languished until the mid-70's when the interest in the study of chaos exploded. Since then, the study of chaos has attracted the attention of biologists [11], mathematicians, physicists, economists [31, 32], scientists and engineers [33, 34]. In recent years, due to availability of high speed computers and refined experimental techniques, it has become clear that examples of chaotic phenomena are abundant in nature and that they have far-reaching consequences in many branches of science and engineering [1, 35].

In the next section, we use the so called "Logistic Map"\(^4\), which is a first order, nonlinear difference equation, to introduce various concepts and terms in chaos. Robert May, a population biologist, first investigated some of the complex behavior exhibited by this simple model [13]. The map will also be used to introduce some of

\(^4\) Also called the Quadratic Map.
the properties that are typical to chaotic systems. These properties can be loosely classified into two category of properties: (1) those, such as invariant density, that describe the temporal evolution of a chaotic system and (2) those that characterize the spatial distribution of points in the phase-space. Although the Logistic Map is used in the introduction of the concepts and properties of chaos, it is important to realize that these concepts are not particular to the Logistic Map, but rather, are common to almost all chaotic systems regardless of their underlying equation or mode of generation.

2.3 Characteristics of the Logistic Map

One of the simplest 1-dimensional nonlinear systems is the Logistic Map. It is an iterative map that relates the present sample $x[n]$ with the next sample $x[n + 1]$ by a simple map

$$x[n + 1] = rx[n](1 - x[n])$$

(2.4)

In functional form, this could be written as

$$x[n + 1] = g^n(x),$$

(2.5)

where $g(x) = rx(1 - x)$. Let $g^n$ denote the $n^{th}$ iteration of $g$, so that $g^0(x) = x$, $g^1(x) = g(x)$, $g^2(x) = g(g(x))$, ..., $g^n(x) = g(g^{n-1}(x))$. This is called a sequence of iterations and the iterative function $g^n$ is sometimes called the iterated map function [36]. Each iteration of the function $g$ produces an iterate and the sequence of values generated by this procedure, i.e., $\{g^n\}_{n=1}^{\infty}$ will be called the trajectory or orbit.

The logistic equation in equation (2.4) maps the unit interval onto itself, i.e., $g : [0,1] \rightarrow [0,1]$. If the magnitude of the $n^{th}$ iterate, $|x[n + 1]|$ in equation (2.5) is greater than 1, then the expression diverges to infinity; hence, to observe the rich dynamical behavior of the logistic map, the magnitude $|x[n]|$ must be less than or equal to 1.

The logistic map attains a maximum value equal to $\frac{r}{2}$ at $x[n] = \frac{1}{2}$. Plugging these values into equation (2.4) we see that the map only possess interesting properties
when \( r \leq 4 \). On the other hand, when \( r < 1 \), the iterates of the logistic map converge to zero. Thus, a rich dynamical behavior is only observed when \( r \) is between 1 and 4.

The dynamical properties of any chaotic system, and in particular the logistic map, can be well understood when we consider the behavior of the map at fixed points, \( x^* \). By fixed points, we mean

\[
x^* = x[n + 1] = x[n].
\]  

or equivalently,

\[
x^* = g(x^*). 
\]  

That is, a fixed point is a convergence point of the map; repeated iterations of the map do not change the outcome. If we substitute the fixed point \( x^* \) into (2.4) and solve for \( x^* \), we obtain

\[
rx^*(x^* - 1 + \frac{1}{r}) = 0,
\]  

with two potential solutions,

\[
x^* = 1 - \frac{1}{r}
\]  

or \( x^* = 0 \).

Although the logistic equation has been determined to have rich dynamical behavior when \( r \) is between 1 and 4, the dynamical behaviors are different as \( r \) scans the range \([1,4)\). The various dynamical behaviors unfold when the fixed points are examined. Therefore, in the next section, we will carefully consider the fixed points \( x^* \) as they affect the dynamical behavior of the logistic map for different values of \( r \) in the range from 1 to 4.

### 2.3.1 Importance of fixed points

The stability of a fixed point is determined by the slope of the function \( g(x) \) at that fixed point. If the magnitude of the slope of the equation evaluated at a given fixed point, is less than 1, then that fixed point is called a stable fixed point, i.e., the trajectory of any initial condition approaches \( x^* \) as the iteration proceeds. If the
magnitude of the slope when evaluated at the fixed point is greater than 1, then we have an unstable or repelling fixed point.

In functional form, the magnitude of the slope of the logistic map is

$$| r(1 - 2x[n]) | .$$

(2.10)

Evaluating the slope at the fixed point $x^* = 1 - \frac{1}{r}$, we obtain

$$| 2 - r | .$$

(2.11)

When $r$ is between 1 and 3, i.e., $1 < r < 3$, the fixed point is stable since the magnitude of the slope in equation (2.11) is less than 1. Evaluating the magnitude of the slope in equation (2.10) at fixed point $x^* = 0$, we obtain $| r |$, which means that the fixed point $x^* = 0$ is unstable for $1 < r < 3$. Thus, in the region where $r$ is between 1 and 3, i.e., $1 < r < 3$ the logistic map has only one stable fixed point. Given any initial condition in this range of $r$, the iteration of the map converges to that one fixed point, i.e., $x^* = 1 - \frac{1}{r}$. Hence, the orbit of the logistic map from any initial condition with this range of $r$ settles to a point. It is therefore referred to as a **stable cycle of period 1**. The behavior of the logistic map in this range of $r$ is depicted in Fig. 2.2 with $r = 2.8$. With any initial condition $0 < x[0] < 1$, the trajectory or orbit of the logistic map when $1 < r < 3$, converges to a stable cycle of period 1 orbit, after the initial transients have passed.

As soon as $r > 3$, the magnitude of the slope in equation (2.11) at the fixed point $x^* = 1 - \frac{1}{r}$ is greater than 1. Therefore, from our definition of stability, this fixed point is now unstable and two new stable fixed points appear [37, 38]. The new stable fixed points are determined by solving for $x^*$ in the expression

$$g(g(x^*)) = g^2(x^*) = x^*$$

(2.12)

and evaluating these points when the magnitude of the slope of $g(g(x)) < 1$. The phenomenon by which one stable fixed point becomes unstable and yields two new stable fixed points is referred to as **period doubling** [2, 3, 1, 4]. When $r$ is slightly greater than 3, the trajectory of any initial condition settles into a **stable cycle of period 2** around the two stable fixed points calculated from equation (2.12).
The two stable points of \( g(g(x)) \) continue to be stable fixed points until \( r = 3.449 \). When \( r \) is slightly greater than \( r = 3.449 \), then the magnitude of the slope of \( g(g(x)) \) at these two stable fixed points is greater than 1. Hence, the two new fixed points become unstable and yield four additional stable fixed points. The four new stable fixed points can again be determined by solving for \( x^* \) in the expression

\[
g(g(g(x^*)))) = g^4(x^*) = x^*
\]  

(2.13)

Again, in this range of \( r \), the trajectory of any initial condition settles into a stable cycle of period 4 around the four stable fixed points of equation (2.13). Fig. 2.3 shows the trajectory of the map for \( r = 3.5 \). With any two initial conditions, the map still settles down to a period 4 orbit around the four stable fixed points. Up to this point in the range of \( r \), the behavior of the trajectory can be well predicted since it is periodic.

As \( r \) increases beyond 3, successive period doubling of period \( 2^0, 2^1, 2^2, 2^3, \ldots \) takes place as shown in Fig. 2.4. This is the famous bifurcation diagram [39, 3, 40, 1]. The figure shows the plot of accumulation point \( x[n] \) versus \( r \). By accumulation, we mean that on the vertical-axis, we plot 200 successive iterates of the logistic map as a point after the first few hundred points have been discarded while on the horizontal-axis, we plot the corresponding value of \( r \). The initial points were discarded to ensure that the steady state behavior of the system has been reached before points were plotted. We see from Fig. 2.4 that as \( r \) scans the interval \( 1 < r < 4 \) and periodic doubling progresses, a critical value of \( r \) denoted by \( r_c \) is attained. Beyond this value, there are an infinite number of fixed points with different periodicities and an infinite number of unstable periodic cycles. There are also an infinite number of initial conditions \( x[0] \) that could lead to such phenomena. We are now at the onset of chaos [39, 4, 1].

In contrast to the behavior of the trajectory in the interval, \( 1 < r < 3 \), the trajectory in the region \( r > r_c \) never repeats; it is totally unpredictable without infinite precision knowledge of \( r \) and the initial conditions. It is the existence of an infinite number of different orbits which never repeat and the lack of apparent predictability that is termed chaos [4, 29, 2, 3]. Due to the dependence of the behavior of the logistic map on \( r \), the parameter \( r \) is aptly called the tuning parameter. For the
Figure 2.2: This figure illustrates the behavior of two trajectories of the logistic map when \( r = 2.8 \). With two different initial conditions \( x[0] = 0.2 \) (solid line) and \( x[0] = 0.9 \) (dashed line) that are far apart from each other, the map settles to a stable period 1 cycle about the fixed point \( x^* = 1 - \frac{1}{2.8} = 0.6428 \).
Figure 2.3: Behavior of two trajectories of the Logistic map with \( r = 3.5 \) in equation (2.4). Notice that with two initial conditions \( x[0] = 0.2 \) (solid line) and \( x[0] = 0.9 \) (dashed line) that are not close to each other, the trajectories of the map still converge to a stable cycle of period 4 about the fixed points \( x^* \) of equation (2.13).
logistic map, the erratic (chaotic) behavior marked by infinite periodicities starts when \( r = r_c = 3.57 \) and ends at \( r = 4 \).

In the next section, we formally define bifurcation and how it is produced. We will also comment on periodic windows observed in the bifurcation diagram when \( r > r_c \). The bifurcation diagram for the Logistic map is shown in Fig. 2.4.

### 2.3.2 Bifurcation

The sudden change in the behavior of the logistic map as \( r \) is varied in the range, \( 3 < r < 4 \), is called bifurcation. It is the splitting that occurs when one stable fixed point become unstable and yields two new stable fixed points. The process of bifurcating from stable periodic cycles to an infinite number of unstable ones is called the bifurcation route to chaos.

The bifurcation route to chaos is not exclusive to the logistic map or 1-dimensional systems [41]. It is common to all non-linear first order difference equations of the form

\[
x[n + 1] = g(x[n])
\]

where \( g(x[n]) \) has only a single maximum in the interval \( 0 < x[n] < 1 \). Recall that the logistic map has a single maximum at \( x[n] = \frac{1}{2} \). Bifurcation is also possible in higher dimensional systems [3, 40]. It is important to notice the presence of various periodic windows in the bifurcation diagram of Fig. 2.4 around \( r = 3.68, 3.74, \text{etc.} \). These \( r \)-windows are characterized \( c \)-cycles \( (c = 3, 5, 6, \ldots) \) with successive bifurcations \( c \times 2^0, c \times 2^1, c \times 2^2, c \times 2^3, \ldots \) [40]. For the Logistic map, the period 3 window around \( r = 3.83 \) has been proven by Li and Yorke [42] to imply chaos. The bands of periodic windows in the bifurcation diagram are sometimes called crises [3, 40]. For a full discussion of periodic windows and crises, refer to [43].

The bifurcation route to chaos is certainly not the only route leading to chaos. Two other different routes to chaos have been discussed extensively in chaos literature. These are the intermittency route to chaos and the transition from quasi-periodicity to chaos. For a full discussion of these routes to chaos, the reader is referred to Schuster [2] and Newhouse [29]. They are not discussed here because these chaotic concepts are beyond the scope of this dissertation.
Figure 2.4: This figure illustrates some of the stable and unstable fixed points of various periods that arise from a bifurcation process in equation (2.4) taken from [2, 3]. Basically, as $r$ increases, stable fixed points of period $2^{n-1}$ become unstable and give rise by bifurcation to new stable fixed points of period $2^n, n = 0, 1, \ldots$. 
2.3.3 Lyapunov Exponent

One of the hallmarks of chaos is depicted in Fig. 2.5. In the chaotic region of the Logistic map when $r = 4$, with two very close initial conditions $x[0] = 0.0903000$ and $x[0] = 0.0903001$, the plot shows the quick divergence that occurs as the trajectories unfold. The rate at which the trajectories diverge from each other is called the Lyapunov exponent [2, 3, 40, 1]. When the observed process is chaotic, the Lyapunov exponent (LE) will be positive whereas the LE will be negative when the process is not chaotic. The LE is not peculiar to the Logistic map or 1-dimensional maps. In fact, for higher dimensional system, there are as many LE's as there are degrees of freedom in the system [2].

2.3.4 Invariant Density

Before discussing Lyapunov exponents quantitatively in section 2.10, let us first define the invariant density. If $x[n]$ is a random variable drawn from a certain distribution or density, then the next iterate $x[n + 1]$, will in general, be distributed differently. However, when both $x[n]$ and $x[n + 1]$ are distributed identically, then the density for which this holds is called the invariant density. In the case of the Logistic map, the invariant density is dependent upon the tuning parameter $r$ in equation (2.4). From the bifurcation diagram in Fig. 2.4, which shows the different transitions from periodicity to chaos in the logistic map as a function of $r$, we can get a glimpse of the computational difficulty involved in determining the invariant density of the Logistic map. Beyond $r = r_c$, notice the existence of periodic windows in the bifurcation plot. This implies impulsive components in the probability density function (pdf) [40]. The continuous invariant density becomes impulsive for some value of $r > r_c$, making it difficult to have a general formula for the invariant density for all values of $r$ where the map is defined. However, computer simulation can be used to obtain a histogram which is representative of the nature of the underlying distribution. A closed form expression for the invariant density of the Logistic map only exists for the chaotic state when $r = 4$. It is independent of $r$ (since $r$ is fixed.
Figure 2.5: A plot of the behavior of two trajectories of the Logistic map in the chaotic region with $r = 4$ in equation (2.4). The trajectories start at initial conditions $x[0] = 0.0903000$ (solid line) and $x[0] = 0.0903001$ (dashed line) that are very close. The divergence in the orbits becomes apparent at about sample number 16.
at 4) and is given as

\[ p(x) = \frac{1}{\pi \sqrt{x(1-x)}} \]  

(2.15)

which is depicted in Fig. 2.6. In section 2.10, the invariant density will be used to estimate the exponential divergence (Lyapunov exponents) of the orbits of two close initial conditions.

### 2.3.5 Ergodicity

The time series generated by the Logistic map in the most chaotic state has many more properties. The most important one that will be discussed here is ergodicity. To demonstrate this, we compute 1500 samples of a single realization of the map with a random initial condition. Next, we compare the single realization time series with an ensemble time series generated as follows: 1500 initial conditions \( x_i[0] \) were drawn from a uniform distribution. Each initial condition is used to generate a corresponding time series \( x_i[n] \). But, for each initial condition, only the 300th iterate \( x_i[300] \) is kept. The collection of all the 1500 samples \( x_i(300), i = 1, 2, \ldots, 1500 \) is referred to as an ensemble time series. The first 100 of the 1500 samples of the ensemble time series is plotted in Fig. 2.7 (a). The histogram of the 1500 sample ensemble time series is shown in Fig. 2.7 (b). Next, we c Fig. 2.8 shows the last 100 samples of the single realization time series. The corresponding histogram is shown in Fig. 2.8 (b). The last 100 samples of the single realization time series is chosen so as to avoid initial transient. Clearly, the histogram in Fig. 2.8 (b) is similar to the one shown in Fig. 2.7 (b) which in turn is close to the plot of the closed form expression for the invariant density in Fig. 2.6. Thus, regardless of whether or not the time series comes from a single realization or across an ensemble of realizations, the time series values are distributed identically. A system that exhibits such a property is said to be \textit{ergodic}. Up until now, we have only examined the temporal evolution of the trajectory of a chaotic system. We will now investigate some of the spatial properties of chaotic systems.
Figure 2.6: The normalized invariant density $p(x) = \frac{1}{\pi(x(1-x))^{\frac{1}{2}}}$ versus $x$ for the Logistic map at $r = 4$. Notice the singularities at $x = 0$ and $x = 1$. 
Figure 2.7: (a) depicts the first 100 samples of the length 1500 ensemble time series of the Logistic map at $r = 4$ generated using 1500 uniformly distributed initial conditions. The ensemble of samples at time instant $n = 300$ were recorded and the first 100 samples is displayed. (b) shows the corresponding normalized histogram of the 1500 ensemble time series sorted into 10 equally spaced bins along the x-axis.
Figure 2.8: (a) is a plot of the last 100 out of 1500 samples of a single realization of the Logistic map with initial condition $x[0] = 0.345$ and $r = 4$. (b) shows the corresponding normalized histogram of the 1500 points sorted into 10 equally spaced bins along the $x$-axis.
2.4 Spatial Properties

Although the trajectories of chaotic systems tend to move exponentially apart with time, the orbits remain within some bounded region by wrapping around each other without intersecting or repeating [1]. The geometry just described is strange and may be fractal in nature. Fractals are characterized by self similarity. That is, they exhibit the same intrinsic form no matter how many times they are magnified. This fractal set is referred to as the strange attractor and the set of initial conditions that evolve to an attractor is called its basin of attraction.

A space which is useful for conceptually visualizing the strange attractor of a chaotic dynamical system is the phase-space or state-space. In chaos literature, phase-space and state-space are used interchangeably. It is an abstract space whose coordinates (which completely specify the state of the system) are the degrees of freedom of the system’s motion. Shown in Fig. 2.9 is a phase-space plot of the Lorenz attractor in equation (2.3) in the chaotic state with initial condition \( x[0] = 33.3, y[0] = -12.5, z[0] = -14.4 \). This Figure is a 2-dimensional projection of the co-ordinates \((x(t), z(t))\) in equations (2.1)-(2.3) as time, \( t \), evolves. The attractor is an infinite loop of the trajectory, which never repeats or crosses itself. The chaotic attractor of the Lorenz attractor is only attracted to a certain lower dimensional region of the phase-space. This lower dimension is also referred to as the attractor dimension, correlation dimension or \( D_2 \) [1].

In the last 15 years, extensive study of chaos has provided new concepts and tools to understand and categorize the complex chaotic behavior that were previously described as random noise. In the next section, we describe some of the concepts and theories needed to quantify chaos. Some of these measures and their extensions are later used in this dissertation to quantify chaos in the acoustic waveforms generated during the production of unvoiced speech signals.

2.5 Measures of Chaos

In this section, we introduce measures such as dimension, entropy, Lyapunov exponents and generalized correlation function as quantitative measures to characterize
Figure 2.9: This is a phase-space plot of the famous Lorenz Attractor. This figure illustrates the projection of the phase-space orbit of equation (2.3) on the x-z plane. That is, the pair \((x(t), z(t))\) in equation (2.3) form the co-ordinates of a point along the horizontal and vertical axis, respectively. The curve shows the trajectory of this pair as time, \(t\), evolves.
either the temporal evolution or spatial distribution of points in the chaotic motion generated by dynamical systems.

2.5.1 Dimensions

Each chaotic system can be characterized by its strange attractor as defined in section 2.4. The dimension of a strange attractor is the dimension of the fractal set to which the long term behavior of the corresponding chaotic system is attracted (settles down) to in phase space. Dimension has been one of the most studied quantifiers of dynamical systems [3, 4]. This was spawned by the realization that the attractors associated with chaotic dynamics have fractional\(^5\) dimension \(d\), in contrast to non-chaotic systems, which always have integer dimensions. Therefore, a strange attractor may, for the purpose of simplicity, be thought of as an attractor in phase-space with dimension \(d < D\), where \(d\) is non-integer. \(D\) is the number of degrees of freedom of the system’s motion. In the example of the Lorenz system in section 2.2, \(D\) is equal to 3, since the system’s motion is generated from 3 Cartesian co-ordinates \(x, y\) and \(z\). However, its attractor dimension \(d\) is equal to 2.06 [1].

For a chaotic system, it is important to emphasize the existence of these two different dimensional spaces in the definition of an attractor. The first dimensional space \(D\), also referred to as the Euclidean dimension, is the space onto which the attractor is projected. The other dimension, \(d\), is a lower-dimensional space to which the trajectory, or orbits of the dynamical system, is attracted. For chaotic attractors, \(d\) is always non-integer and less than \(D\) [1, 3, 4].

In the next section, we define a theoretical method of evaluating the dimension \(d\). We also discuss some of the problems of implementing this theoretical method.

Hausdorff dimension, \(d\)

The relevant definition of dimension \(d\) which is due to Hausdorff [44, 45] is as follows:
Cover the attractor (\(A\)) of a \(d\)-dimensional object in phase space with a total number \(N(\epsilon)\) of \(d\)-dimensional volume elements such as spheres, cubes, hyper-cubes etc., where \(\epsilon\) is the diameter or width of the volume element. The process of covering

\(^5\)The idea of fractional dimension was first suggested in the literature by Mandelbrot [28].
the attractor of a chaotic system with volume elements is sometimes referred to as “partitioning” [1]. Let \(N(\epsilon)\) be the number of non-empty volume elements needed to cover \(A\). As \(\epsilon\) is made smaller and smaller, the sum of the occupied volume elements approaches the volume of \(A\). Being a purely geometric measure, \(d\) should be independent of the width of the volume elements used in covering the attractor. Fig. 2.10 illustrates the covering of a set of points in \(d\)-dimensional space with cubes of width \(\epsilon\). As \(\epsilon\) get smaller in Fig. 2.10, the relationship in equation (2.16) is used to estimate \(d\).

If \(A\) lies on a \(d\)-dimensional space, then for small \(\epsilon\), the number of volume elements needed to cover \(A\) is inversely proportional to \(\epsilon^{-d}\),

\[
N(\epsilon) \propto \epsilon^{-d},
\]

as \(\epsilon \to 0\), where \(d\) is the Hausdorff dimension of the strange attractor. The dimension \(d\) is obtained by solving equation (2.16) in the limit as \(\epsilon \to 0\) and \(N(\epsilon) \to \infty\), i.e.,

\[
d = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{N(\epsilon) \to \infty \log \left(\frac{1}{\epsilon}\right)}.
\]

If the limit does not exist, then \(d\) is undefined. Theoretically, as \(\epsilon \to 0\) and \(N(\epsilon) \to \infty\), a log-log plot of \(\log N(\epsilon)\) versus \(\log \left(\frac{1}{\epsilon}\right)\) is a straight line whose slope is the Hausdorff dimension \(d\). The Hausdorff dimension measures the “intertwined” and “wrapped around” structure of the fractal attractor [4].

The calculation of the Hausdorff dimension \(d\) by numerically partitioning the phase-space with volume elements and counting the number of points in each volume element is exceedingly hard, cumbersome and impractical for higher dimensional systems [46, 47]. In the next section, we will discuss the correlation dimension \(D_2\) introduced by Grassberger and Procaccia in 1983 [44]. \(D_2\) is also a useful measure of the intricate structure of the strange attractor and its calculation yields an estimate of the Hausdorff dimension \(d\). However, in contrast to the Hausdorff dimension \(d\), the correlation dimension \(D_2\) is relatively easy to estimate.
Figure 2.10: Shows an illustration of the covering of an object (a set of points indicated by the dots) by cubes of width $\varepsilon$ (adopted from [4]).
2.6 Correlation dimension, $D_2$

If the chaotic attractor is analyzed from a probabilistic context, we can define the probability of finding a point on the attractor in a certain volume element. Assume the trajectory of a dynamical system consists of sequence of $M$-points. If the points are partitioned in phase space by volume elements, then the probability $p_i$ of finding a point of the attractor in the volume element number $i$, $(i = 1, 2, 3, \ldots, N(\epsilon))$, is given by

$$p_i = \lim_{M \to \infty} \left( \frac{M_i}{M} \right)$$

(2.18)

where $M_i$ is the number of points of the attractor in a particular volume element number $i$. A graph of this probability $p_i$ of the $i^{th}$ volume element as a function of $i$ gives the probability distribution for the attractor. In many chaotic systems, the probabilities $p_i$ do not depend on any particular starting point on the attractor as long as the trajectory of the chaotic system is allowed to evolve for a long time. The limit of the $p_i$'s as $i$ tends to infinity is called the invariant density or invariant distribution that was discussed earlier in section 2.3.4.

Grassberger and Procaccia (GP) [44] in 1983 introduced a simple and efficient means of estimating the correlation dimension $D_2$ from experimental time series. The GP method of estimating correlation dimension $D_2$ is related to the spatial correlation between different points on the attractor. In the next section, we will relate (1) the probability, $\sum_{i=1}^{N(\epsilon)} p_i^2$, that two points on the attractor lie within a volume element of diameter $\epsilon^d$ to (2) the spatial correlation between points on the chaotic attractor.

2.6.1 Correlation Integral

Consider an $M$-point time series of a dynamical systems generated from $D$ dimensional coordinates, i.e.

\[
\begin{align*}
x_1 & \quad x_2 & \quad x_3 & \quad \cdots & \quad x_M \\
y_1 & \quad y_2 & \quad y_3 & \quad \cdots & \quad y_M \\
\vdots & \quad \vdots & \quad \vdots & \quad \ddots & \quad \vdots \\
z_1 & \quad z_2 & \quad z_3 & \quad \cdots & \quad z_M
\end{align*}
\]

(2.19)
where each row of equation (2.19), which is $M$-points long, represents a single coordinate of a $D$-dimensional system. The length each column is $D$. The discrete time series of equation (2.19) could also be sampled versions of a continuous $D$-dimensional time series sampled in time using a fixed sample rate $T$. If we now form a vector time series such that

\[
X_1 = \begin{bmatrix} x_1 \\ y_1 \\ \vdots \\ z_1 \end{bmatrix}, \\
X_2 = \begin{bmatrix} x_2 \\ y_2 \\ \vdots \\ z_2 \end{bmatrix}, \\
\vdots \\
X_M = \begin{bmatrix} x_M \\ y_M \\ \vdots \\ z_M \end{bmatrix},
\]

then the $D$-dimensional vectors $X_1, X_2, \ldots, X_M$ are now vector points on the attractor of a dynamical system in the phase-space. These vector points are spatially correlated since they lie on the same attractor in phase-space [2].

The spatial correlation of a time series can be measured by the expression

\[
\sum_{k=1}^{N(E)} p_k^2 \approx \frac{1}{M} \sum_{k=1}^{M} p_k^2
\]

which is approximately equal to the probability that two vector points of the attractor lie within a volume element of diameter $\epsilon^D$. For large $M$ in equation (2.20) and small volume width $\epsilon$,

\[
\sum_{k=1}^{N(E)} p_k^2 \approx \frac{1}{M} \sum_{k=1}^{M} p_k.
\]

This is the probability that two vector points $X_i$ and $X_j$, $i \neq j$, on the attractor are separated by a distance smaller than $\epsilon$. In functional form, this can be written as approximately

\[
\lim_{M \to \infty} \frac{1}{M^2} \left\{ \text{number of pairs, } X_i \text{ and } X_j, \text{ whose distance } ||X_i - X_j|| \text{ is less than } \epsilon \right\}
= \lim_{M \to \infty} \frac{1}{M^2} \sum_{i=1}^{M} \sum_{j=1}^{M} \Theta(\epsilon - ||X_i - X_j||)
\]

where $\Theta$ is the Heaviside function\footnote{The Heaviside or unit step function is defined as $\Theta(\epsilon) = \begin{cases} 1 & \text{if } \epsilon > 0 \\ 0 & \text{if } \epsilon < 0 \end{cases}.$}, $\epsilon$ is a distance measure limit between vectors and $|| . ||$ is a normed distance measure between two vectors. This corresponds to the correlation integral $C_2(\epsilon)$ [40, 2]. We emphasize that this result is rigorously true only when both $M$ and $\epsilon^{-1}$ approach infinity. For a full discussion of this derivation, see Grassberger et al [44] and Ding et al [48].
When the distance between any vector point with itself, i.e. the identical pair \( X_i = X_j \), is discarded in order to avoid statistical error \([49]\), the correlation integral is given as

\[
C_2(\epsilon) = \frac{1}{M(M-1)} \sum_{i=1}^{M} \sum_{j=1, j \neq i}^{M} \Theta(\epsilon - \| X_i - X_j \|).
\] (2.22)

The correlation integral \( C_2(\epsilon) \) counts the number of point pair \((X_i, X_j)\) on the attractor whose distance is smaller than \( \epsilon \). \( \|X_i - X_j\| \) is sometimes referred to as the inter-point distance between two vector points \( X_i \) and \( X_j \) on the attractor.

There are three norm metrics commonly used in calculating the correlation integral \( C_2(\epsilon) \). They include: the Manhattan or \( L^1 \) norm

\[
\|x\|_1 = \sum_i (|x_i|),
\] (2.23)

the Euclidean or \( L^2 \) norm

\[
\|x\|_2 = \left[ \sum_i (|x_i|^2) \right]^{\frac{1}{2}},
\] (2.24)

and the max or \( L^\infty \) norm

\[
\|x\|_\infty = \left[ \sum_i (|x_i|^\infty) \right]^{\frac{1}{\infty}} = \max_i(|x_i|)
\] (2.25)

[50]. In this dissertation, the max or \( L^\infty \) norm is used in the chaos analysis of unvoiced speech so as to reduce computational complexity. However, it has been shown historically that the choice of the norm metric does not severely affect the results of the estimated correlation dimension \([50, 51]\).

The correlation integral can be used to determine the correlation dimension. The correlation integral \( C_2(\epsilon) \) has been shown by Grassberger and Procaccia \([44]\) to be proportional to \( \epsilon^{D_2} \), i.e.,

\[
C_2(\epsilon) \sim \epsilon^{D_2}
\] (2.26)

or equivalently, the correlation dimension

\[
D_2 = \lim_{\epsilon \to 0} \frac{1}{\log \epsilon} \log C_2(\epsilon)
\] (2.27)

\[
\approx \lim_{\epsilon \to 0} \frac{1}{\log \epsilon} \log \sum_{i=1}^{N(\epsilon)} p_i^2.
\] (2.28)
$D_2$ is now relatively easy to estimate using equation (2.26) since it only involves the counting of inter-point distances that are less than $\epsilon$.

In the next section, we review a generalization of correlation integral $C_2(\epsilon)$ and correlation dimension $D_2$ to a hierarchy of generalized correlations and dimensions, $C_q(\epsilon)$ and $D_q$ respectively.

### 2.7 Generalized Dimension, $D_q$

The expression for the correlation dimension $D_2$ in (2.28) can now be generalized to a hierarchy of generalized dimensions indexed by the parameter $q$,

$$D_q = \frac{1}{q-1} \lim_{\epsilon \to 0} \frac{1}{\log \epsilon} \log \sum_{i=1}^{N(\epsilon)} p_i^q - \infty < q < \infty$$

(2.29)

or

$$D_q = \lim_{\epsilon \to 0} \frac{1}{\log \epsilon} \log C_q(\epsilon) - \infty < q < \infty$$

(2.30)

where

$$C_q(\epsilon) = \lim_{M \to \infty} \left\{ \frac{1}{M} \sum_{i=1}^{M} \left( \frac{1}{M-1} \sum_{j=1}^{M} \Theta(\epsilon - \| X_i - X_j \|) \right) \right\}^{\frac{1}{q-1}}.$$  

(2.31)

$C_q(\epsilon)$ is the generalized correlation integral; for $q = 2$, it is equal to the correlation integral in equation (2.22). When $q > 0$, the $i^{th}$ volume elements with large probability $p_i$ have a greater influence in the determination of $D_q$ in equation (2.29) whereas when $q < 0$, those volume elements with small $p_i$ are accentuated in their contribution towards $D_q$.

Now by substituting (2.31) into (2.30), the generalized dimensions, $D_q$, can be written as [3, 52, 53]

$$D_q = \lim_{M \to \infty} \lim_{\epsilon \to 0} \frac{1}{\log \epsilon} \left( \frac{1}{q-1} \right) \log \left\{ \frac{1}{M} \sum_{i=1}^{M} \left( \frac{1}{M-1} \sum_{j=1}^{M} \Theta(\epsilon - \| X_i - X_j \|) \right) \right\}^{\frac{1}{q-1}}.$$  

(2.32)

In contrast to the correlation dimension $D_2$, the generalized $q^{th}$ dimension $D_q$ determines the various fractal regimes present in the attractor when the attractor...
is non-uniform. The generalized dimensions $D_q$ in equation (2.32) are the order-$q$ Renyi generalized dimensions (GD) which were formulated in the context of chaotic dynamics by Hentschel and Procaccia in 1983 [54]. $D_q$ is a measures of the inhomogeneity or multi-fractality of an attractor whenever the $D_q$ values are not equal for all $q$. However, when the $D_q$ are the same, the distribution of vector points on the attractor is fairly homogeneous. A chaotic attractor with non-constant values of the generalized dimension is called a multifractal.

### 2.7.1 Box-counting Dimension

Note that when $q = 0$ in equation (2.29), we have

$$D_0 = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{\log \epsilon}. \quad (2.33)$$

This is called the box-counting dimension [3].

### 2.7.2 Information Dimension

We can also take the limit as $q$ goes to 1 in equation (2.29) to define

$$D_1 = \lim_{q \to 1} D_q. \quad (2.34)$$

By applying L'Hospital's rule to equation (2.29), one can obtain [3]

$$D_1 = \lim_{\epsilon \to 0} \frac{\sum_{i=1}^{N(\epsilon)} p_i \log p_i}{\log \epsilon}. \quad (2.35)$$

Due to the form $p_i \log p_i$ in equation (2.35), which has been used as a measure of information capacity in information theory [55], $D_1$ is aptly referred to as the information dimension.

Another property of the generalized dimensions $D_q$ is that they decrease with increasing $q$. In general, it has been shown in [54] that

$$D_q \leq D_{q'}, \quad q > q' \quad (2.36)$$
Thus, for example, $D_2$ provides a lower bound for $D_1$, $D_1$ provides a lower bound for $D_0$ and so on.

In the next section, we systematically describe how the generalized dimensions can be estimated given an experimental time series. These steps are very crucial since in most experimental settings, it may be too expensive or almost impossible to obtain all the $D$-dimensional components of the experimental data.

2.8 Taken's Embedding Theorem

Our analysis of the generalized dimensions up to this point is based on the assumption that we have all of the actual $D$-dimensional components in equations (2.19-2.20) in the manifold of the chaotic system. However, it is possible to analyze from a chaos point of view a scalar or 1-dimensional component of this signal. In general, the chaotic properties of any dynamical system can be reconstructed from a single component by using Takens embedding theorem [56]. If the embedding space is generated properly, the behavior of the trajectories in this embedding space will have the same geometrical and dynamical properties as those that characterize the actual trajectories in the full $D$-dimensional state space.

Given an $N$-point scalar component time series, $\{x[1], x[2], x[3] \ldots x[N]\}$, the reconstructed trajectory $X_i$, $i = 1, 2, \ldots, ME$ from a single component can be expressed as a matrix where each column is a phase-space vector, i.e.,

$$
\begin{bmatrix}
X_1 & X_2 & \ldots & X_{ME}
\end{bmatrix}
$$

where $X_i$ is the state of the system corresponding to the discrete time index $i$. $X_i$ can now be regarded as an embedded vector point in phase-space. Each $X_i$ is given by

$$
X_i = (x[1] \ x[i+\tau] \ \ldots \ x[i+(dE-1)\tau])
$$

where $\tau$ is a chosen time delay, $dE$ is the embedding dimension and $ME = N - dE + 1$ is the total number of the embedded vectors. $ME$ is equivalent to $M$ in equations
The generalized correlation integral is now given as

\[
C_{q,d_E}(\epsilon) = \lim_{M \to \infty} \left\{ \frac{1}{M} \sum_{i=1}^{M} \left\{ \frac{1}{M-1} \sum_{j=1}^{M} \Theta(\epsilon - \| X_i - X_j \|) \right\}^{q-1} \right\}^{1/(q-1)}, \tag{2.39}
\]

where the additional subscript \(d_E\) is used to denote the level of embedding used in equation (2.38).

In theory, for noise free data, the particular choice of the delay time \(\tau\) plays no role in the reconstruction of the attractor. However, in the absence of noise free data, which is always the case in experimental settings, it is important to optimize the time delay parameter \(\tau\) and then find a practical embedding dimension \(d_E\) smaller than \(D\) which has been proven to be sufficient for proper embedding [56]. Information theory has been of practical value in finding the optimum \(\tau\) [57]. An optimum time delay often chosen, due to Fraser and Swinney, is the first minimum of the mutual information measure [57].

Often, in chaotic time series analysis of experimental data, only one component of the \(D\)-dimensional system described in equation (2.19) is available. This is particularly crucial in speech production where any attempt to insert transducers in the vocal cavity for the purpose of recording more than one single component of speech will alter or impair the speech produced [58, 26]. For most speech analysis, the single component that is available, i.e., that can be measured, is the acoustic pressure waveform recorded by a microphone in front of the mouth. Therefore, in order to have a chaotic analysis of speech or other experimental data, a method of reconstructing the multi-dimensional dynamics of the system from a scalar or 1-dimensional representation is very crucial.

The chaos analysis of speech in this dissertation is performed on experimental data which requires reconstructing the nonlinear dynamics from a single representation or component of a speech signal. Therefore, we shall henceforth parameterize all expressions for estimating the generalized correlation integrals of speech signals as well as the generalized dimensions by using the optimum Taken’s [56] embedding dimension \(d_E\). In the next section, an important method of determining the proper delay time needed for the reconstruction of experimental data is explained.
2.8.1 Mutual Information

If we consider a time series \( u(t) \) with its delayed version \( u(t+\tau) \), Fraser and Swinney [57] demonstrated that the optimum delay time \( \tau_{opt} \) corresponds to the delay equal to the occurrence of the first minimum of the mutual information between the time series and its delayed version. If we assume that the time series can be sampled, then we can treat the time series and its delayed version as discrete random variables \( U \) and \( V \).

Mutual information is a measure of the amount of information about a random variable \( V \) given by the random variable \( U \). If we now define discrete probabilities

\[
p_i(u) = \text{Prob}\{(i+1)\varepsilon < u < i\varepsilon\}
\]

and

\[
\hat{p}_i(v) = \text{Prob}\{(i+1)\varepsilon < v < i\varepsilon\},
\]

then associated with these probabilities are a measure of uncertainties \( H_U(\varepsilon) \) and \( \hat{H}_V(\varepsilon) \) where

\[
H_U(\varepsilon) = - \sum p_i(u) \log p_i(u)
\]

and

\[
\hat{H}_V(\varepsilon) = - \sum \hat{p}_i(v) \log \hat{p}_i(v)
\]

Similarly, we can define joint probabilities when the discrete time sequence is embedded using (2.38) with \( d_E = 2 \). i.e.,

\[
\hat{p}_i(u,v) = \text{Prob}\{(i+1)\varepsilon < u < i\varepsilon \text{ and } (i+1)\varepsilon < v < i\varepsilon\}.
\]

The corresponding joint uncertainty \( \hat{H}_{UV}(\varepsilon) \) between the discrete random variables \( U \) and \( V \) is then defined as

\[
\hat{H}_{UV}(\varepsilon) = - \sum \hat{p}_i(u,v) \log \hat{p}_i(u,v)
\]

where the sum in equation (2.45) may now be interpreted as the sum taken over all the non-empty 2-dimensional volume elements of width \( \varepsilon^2 \) used in covering the discrete time sequence in the embedded space.
The mutual information between the two random variables is now defined as

$$I(U, V) = \lim_{\varepsilon \to 0} H_U(\varepsilon) + \hat{H}_V(\varepsilon) - \hat{H}_{UV}(\varepsilon)$$

in the limit as \( \varepsilon \) get smaller and smaller. This mutual information gives the average amount of information about the delayed version \( V \sim u(t + \tau) \) when \( U \sim u(t) \) is observed. In essence, this method can be used to calculate the uncertainties for a series of possible delay times.

We can now parameterize \( I(U, V) \) by \( \tau \) where \( I_\tau(U, V) \) is a function of the series of time lags \( \tau \). When \( \tau \) is too small, the information given about the delayed version is too large and when \( \tau \) is too large, the information given about the delay version is almost nothing. The prescription given by Fraser and Swinney is to choose the value of \( \tau \) where the first minimum of \( I_\tau(U, V) \) occurs with respect to a series of delay times. This value of \( \tau \), called \( \tau_{\text{opt}} \), is the optimum delay time for reconstructing the dynamics of the system from its scalar measurements [57]. Fig. 2.11 shows the mutual information as a function of the delay time \( \tau \) for the x-component of the Lorenz data defined in equation (2.3). Notice that the first minimum occurs after a delay time of 5. When the base 10 logarithm in the expressions for the uncertainties \( H_U \) and \( \hat{H}_V \) in equations (2.42-2.43) is replaced by the base 2 logarithm, the mutual information will have units of bits.

Other methods for determining the optimum delay time have been proposed; they are variations of the one discussed above. In many cases, these methods give comparable results and are only attractive because some are easier to implement than others. The methods include: the variations of the Fraser-Swinney algorithm by Liebert-Schuster [59] and Pineda-Sommerer [47]. In this dissertation, however, we use the Fraser-Swinney method to determine the optimum delay time \( \tau \) because we found that the method gives accurate results for conventional attractors.
Figure 2.11: Shows the normalized mutual information as a function of time lag $\tau$ for the scalar $x(t)$ variable of Lorenz data in (2.3). The plot is normalized by the mutual information at lag zero. The mutual information at lag zero is 2.411492 bits.
2.9 Numerical Implementation of the Generalized Correlation Integrals GCI

A straightforward implementation of the generalized correlation integral in equation (2.39) for each \( q \) is extremely time consuming since it involves counting all the pairs of vector points \((X_i, X_j)\) on the attractor. This is a tedious process which may be on the order of \( O(M^2) \), where \( M \), which is usually very high, is the number of reconstructed vector points of the attractor in phase-space. However, it has been noted in [60, 49] that for large widths, \( \epsilon \), a small fraction of the pairs \((X_i, X_j)\) yields sufficient statistics, so it is not absolutely necessary to compute all \( M^2 \) pairs. This observation leads to a reduction in the search over all neighbors of a vector point. The search is now limited to a few \( M_{seed} \) subset of pairs which are either randomly or uniformly selected from \( M \). Algorithmically, the expression in equation (2.39) can now be rewritten as

\[
C_{q,d_E}(\epsilon) \approx \left\{ \frac{1}{M_{seed}} \sum_{i=1}^{M_{seed}} \left\{ \frac{1}{M - 1} \sum_{j=1, j \neq n(i)}^{M} \Theta(\epsilon - \| X_{n(i)} - X_j \|) \right\} \right\}^{q-1} \left( \frac{1}{(q-1)} \right)
\]

(2.47)

where \( j \) runs over all the \( M \) embedded vector points, \( i \) only runs over \( M_{seed} \) and \( n(i) \) is the random or uniform index. \( M_{seed} \) is typically \( \frac{M}{5} \). With this computational reduction, the estimation of the generalized dimensions \( D_q \) from equation (2.47) is at least practical.

The standard procedure for estimating the generalized dimensions using \( C_{q,d_E}(\epsilon) \) is to increase the embedding dimension \( d_E \) incrementally until the calculated generalized dimensions appear to have converged. An example of this method is depicted in Fig. 2.12 for the case when \( q = 2 \). The embedding dimension \( d_E \) is increased until it reaches a value that is greater than or equal to twice the dimension \( D \) of the attractor space. At this value of \( d_E \), the correlation dimension \( D_2 \) becomes stable and independent of the embedding dimension.

In practice, one computes the generalized dimension as a function of embedding dimension \( d_E = 1, 2, 3, \ldots \). For an infinite amount of data and small \( \epsilon \), we expect

\[
C_{q,d_E}(\epsilon) \propto \epsilon^{D_q}.
\]

(2.48)
Figure 2.12: A schematic diagram of the correlation dimension $D_2$ plotted as function of the embedding dimension. As the embedding dimension is increased, a value of $d_E$ is reached where the correlation dimension converges and is independent of the embedding dimension. The correlation dimension of the attractor in this case is 2.01 (adopted from [1]).
The generalized dimensions $D_q$ are then extracted from the generalized correlation integral by computing logarithm of equation (2.48) and estimating the local slopes of the $\log C_{q,d_E}(\epsilon)$ versus $\log \epsilon$ plot, i.e.

$$D_q = \frac{\partial}{\partial \log \epsilon} [\log C_{q,d_E}(\epsilon)].$$

(2.49)

### 2.9.1 Numerical Problems

Numerically, with infinite data, we expect the expression for the local slopes in equation (2.49) to be constant. However, with finite data constraints encountered in experiments, the plot of the local slopes as function of $\log \epsilon$ reveals a great deal about the distribution of points on the attractor at different values of $\epsilon$. As shown in Fig. 2.13, every plot of local slopes of local slopes of local slopes of $\log C_{q,d_E}(\epsilon)$ versus $\log \epsilon$ for finite data, contains four distinct regions [5, 49]. In this section, we describe each region and their significance in dimension estimation. In region I, the reference distance, $\epsilon$, is too small compared to the average inter-point spacing of vector points on the attractor. Therefore, there is no information about the attractor in this region and $C_{q,d_E}(\epsilon)$ as well as the local slopes are zero\(^7\). In region II, which is also called the noise region [49], the local slopes are erratic because noise in the data is smeared out in the phase-space, i.e., the noise is not confined to a fractal self-similar attractor. The noise level in the experimental data can even be estimated since the local slope in this region increases and reaches the noise level at the value of the embedding dimension [5]. Region III, is the "linear region" of the generalized correlation integral and it is marked by a plateau-like structure in the local slopes plot [5, 49, 48, 1]. In this scaling range, the true fractal nature of the underlying attractor is revealed in the sense that the local slopes are constant for increasing embedding dimension $d_E$. The average of the local slopes as a function of the embedding dimension $d_E$ at this plateau are the estimate of the generalized

\(^7\)In order to avoid estimating the $\log (0)$ which is mathematically undefined for this region, a method which scales the inter-point distance between vector points on the attractor such that the minimum inter-point space $\epsilon_{\text{min}} = 1$ and the maximum inter-point spacing $\epsilon_{\text{max}} = 10000$ is adopted. In essence, $\epsilon$ in equation (2.49) can only be integer values between 1 and 10,000. A variation of this method of converting the inter-point spacing into integers may be found in [47]. Also in Appendix A, a completely new derivation of the numerical computation of $\log C_{q,d_E}(\epsilon)$ which avoids taking the logarithm of zero is provided.
Figure 2.13: A schematic sketch of the local slope \( \frac{\partial}{\partial \log \epsilon} \log C_{2,d_{H}} \) as a function of \( \log \epsilon \) showing the distinct regions that are typical for chaotic attractors (adopted from [5]).
dimension of the attractor. In region IV, we experience the opposite of the lack of inter-point distance of region I. However, in this case, the reference distance is now too large compared with the inter-point distance between vector points on the attractor. As such, the plateau-like structure in region III is lost due to finite data size.

The “plateau” technique above is used for estimating all the generalized dimensions $D_q$. When $q < 0$, the duration of the plateau-like structure in region III is reduced to almost a point convergence whereas the duration is increased when $q > 0$. Fig. 2.14(a) shows a typical plot of the correlation integral $C_{q,d_E}(\epsilon)$ and Fig. 2.14(b) shows the corresponding local slopes $\frac{\partial}{\partial \log \epsilon} \log C_{q,d_E}$, for different embedding dimensions $d_E = 5, 6, \ldots, 15$. The data used in generating this plot are from the x-component of the Lorenz system defined in equation (2.1). The calculation of the $C_{q,d_E}$ for each $d_E$ was performed on 10000 data points. In Fig. 2.14(b) the estimated correlation dimension $D_2$ is the average of the local slopes for different embedding dimension at a certain point $\log \epsilon$ in the plateau region. For the Lorenz data, the estimated dimension is 2.06, which is identical to the known estimated value in [44]. Also notice the distinct regions of the local slopes as discussed related to Fig. (2.13) of section 2.9.1. In this dissertation, we adopt the method described above to estimate the generalized dimensions of unvoiced speech signals.

### 2.9.2 Other GCI-GD Algorithms

In the last five years, a number of algorithms for efficiently calculating generalized correlation integrals and estimating generalized dimensions $D_q$ have been developed [61, 47, 60]. Most of these methods suitably structure the vector points on the attractor so that optimal neighbor-search algorithm such as K-tree [62], minimal spanning tree [61] and efficient box-counting method [47] may be used to count the inter-point distances between a vector point on the attractor and its neighbors. In some cases, automated procedures for extracting the generalized dimensions have also been proposed [60, 63].

In the next section, we discuss one of the major quantitative measures of chaotic behavior. The existence of a positive Lyapunov exponent qualifies any dynamical
Figure 2.14: (a) Shows the log of correlation integral plotted as a function of log $\epsilon$ for different embedding dimensions $d_E$ from 5 to 15. Fig. 2.14: (b) depicts the corresponding local slopes which are estimates of the correlation dimension $D_2$. The dimension estimation was performed on the x-component of the Lorenz data and the resulting estimated correlation dimension $D_2$ in the plateau region between $6 < \log \epsilon < 7.5$ is 2.06.
Lyapunov Exponents

While the generalized dimensions $D_q$ in equation (2.30) characterize the spatial distribution of points on an attractor in phase-space, the Lyapunov Exponents (LE) describe the dynamics of the temporal evolution of the trajectories of chaotic systems [6, 64]. The notion of exponential divergence or convergence of two closely spaced orbits of the chaotic system is made formal in the definition of the Lyapunov exponent.

Let two closely spaced trajectories on an attractor start with an initial separation $dd(0)$. Then, the trajectories diverge (or converge) so that their separation after a time $t$, denoted by $dd(t)$ is given by

$$dd(t) = dd(0) \exp^{\lambda t}. \quad (2.50)$$

The parameter $\lambda$ is the Lyapunov exponent. From equation (2.50), we see that the two orbits will diverge when $\lambda$ is positive and converge when $\lambda$ is negative. This exponential separation is a function of the slope of the nonlinear mapping function (such as $g$ in equation (2.14)) that generated the initially closely spaced trajectory points that are now being exponentially separated as a function of time on the attractor.

The average exponential separation is given as

$$\lambda = \int_{-\infty}^{\infty} p(x) \log | \frac{dg(x(t))}{dx(t)} | \, dx(t). \quad (2.51)$$

A more practical formula for the average Lyapunov exponent is

$$\lambda \approx \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \log | \dot{g}(x[n]) | \quad (2.52)$$

where $N$ is the number of samples in the discrete time series.

In general, for a $D$-dimensional chaotic system, there will be $D$ LE's governing the behavior of the system. The set of these Lyapunov exponents $\lambda_i$, $i = 1, 2, \ldots, D$, constitutes the Lyapunov spectrum. Lyapunov exponents are usually ordered in a
decreasing fashion: \( \lambda_1 \geq \lambda_2 \geq \lambda_3 \ldots \geq \lambda_D \). Any dynamical system containing at least one positive LE is defined as being chaotic. For the system to be well behaved in steady state, it must have one LE that is strictly negative. Positive LE are a hallmark of chaotic behavior indicating a strong instability within the attractor [3]. The presence of one positive Lyapunov exponent indicates simple chaos whereas the presence of more than one positive Lyapunov exponent indicates hyper chaos [3].

In the next section, we provide an example of Lyapunov exponents calculation for the Logistic Map. The result will show that, indeed, the Logistic map is chaotic when \( r = 4 \).

### 2.10.1 Logistic Map LE

The Lyapunov exponent for the logistic equation is a function of the tuning parameter \( r \). Fig. 2.15 shows numerically calculated values of the LE as \( r \) in the Logistic Map increases from 2.8 to 4 using equation (2.52). Later, we mathematically determine the value of the Lyapunov exponent at \( r = 4 \) to illustrate Lyapunov exponent calculation using the invariant density.

From Fig. 2.15, when \( r < r_c = 3.57 \), the LE is negative, but when \( r \geq r_c \), the Lyapunov exponent becomes positive. This is an indication that the system is now chaotic. For \( r = 4 \), the value of the Lyapunov exponent is \( \log 2 \).

It can be shown that the average Lyapunov exponent in equation (2.51) for the Logistic map is positive when \( r = 4 \), indicating a chaotic system. This produces the divergence depicted in Fig. 2.5 when the trajectories of two chaotic sequences of the \( r = 4 \) logistic map generated with closely spaced initial conditions diverge from one another as a function of time. Recall that for the case of \( r = 4 \), the iterate will be generated from

\[
x[n+1] = 4x[n](1 - x[n]).
\]  

Therefore, the functional map, \( g(x) = 4x(1 - x) \), has the derivative

\[
\dot{g}(x) := 4(1 - 2x).
\]  

We need this derivative to compute the average Lyapunov exponent by making use
Figure 2.15: Shows the Lyapunov exponent as a function of the tuning parameter $r$ for the Logistic map. When $r > r_c = 3.57$, the Lyapunov exponent becomes positive indicating chaotic behavior.
of the invariant density of the logistic map at $r = 4$, i.e.,

$$\lambda_{AVG} = \int_0^1 p(x) \log | j(x) | \, dx$$

$$= \frac{1}{\pi} \int_0^1 \log \left| \frac{4(1-2x)}{\sqrt{x(1-x)}} \right| \, dx,$$

where $p(x)$ is defined in equation (2.15). If we make the substitution $x = \sin^2(\frac{\pi y}{2})$, we obtain [1]

$$\lambda = \int_0^1 \log(4 \cos(\frac{\pi y}{2})) \, dy$$

$$= \log 2.$$  

Since the value of the Lyapunov exponent is positive, i.e., $\log 2 > 0$, the Logistic map at $r = 4$ is indeed chaotic [1].

An algorithm to compute only the maximum Lyapunov exponent was developed by Wolf in 1983 [6]. Since then, other algorithms that estimate all the Lyapunov exponents have been developed. New algorithms that greatly reduce computational complexity have been proposed [65, 66, 67, 68, 69, 70]. In this dissertation, we implement the Wolf algorithm\(^8\) to prove that unvoiced speech possesses positive Lyapunov exponents and, hence, is chaotic. For a complete characterization of unvoiced speech in terms of all Lyapunov exponents, the reader is referred to Narayanan et al [14]. A brief summary of the Wolf algorithm for estimating the largest Lyapunov exponent from experimental data is as follows:

**Wolf Algorithm [6]**

Given a discrete time sequence $x[n]$, where $n = 1, 2, \ldots N$, one can reconstruct the trajectory using Takens embedding theorem of section 2.8, provided one has

$$X_1, X_2, \ldots, X_{M_E}$$

where each $X_i$ is an embedded vector point in phase-space defined in equation (2.38). To estimate the largest Lyapunov exponent, the Wolf algorithm monitors the long-term evolution of a single pair of nearby orbits.

\(^8\)Our special thanks to Prof. Wolf for giving us a copy of the compiled version of his algorithm for estimating the largest Lyapunov exponent.
Fig. 2.16 shows a schematic diagram of the process of monitoring the orbits. The figure will be used to explain the LE estimation. The thick solid line connecting the embedded vector points $X_i$ is referred to as the fudicial trajectory. Assume that the length $N$ of the discrete time series $x[n]$ is long enough such that each embedded vector point $X_i$ on the fudicial trajectory is surrounded by neighboring embedded vector points. We can define an hyper-sphere enclosing the neighbors within a certain Euclidean distance, i.e.,

$$L(k) < \epsilon \quad k = 0, 1, 2, \ldots, M_E$$

(2.60)

where $L$ is the maximum initial separation between a vector point on the fudicial trajectory and its $N$ neighbors, $\epsilon$ is the radius of hyper-sphere in $\mathbb{R}^{d_E}$, and $d_E$ is the embedding dimension. If we define $L(1)$ as the initial separation between $X_1$ and any neighbor within $\epsilon$ radius, then after a certain time $t_1$, $L(1)$ will have evolved to $L'(2)$ which is a distance from $X_2$ larger than $\epsilon$. We then replace this non-fudicial point with a point closer to the fudicial point $X_2$. The replacement point which is at a distance $L(2)$ from $X_2$ is chosen such that it is closer to the nearest fudicial point $X_2$ and the angular separation $\xi$ between $L'(2)$ and $L(2)$ is small. Again, after a certain time $t_2$, this new separation $L(2)$ between $X_2$ and any neighbor will have evolve to $L'(3)$ which is the distance between the fudicial point $X_3$ and the new non-fudicial point. If the distance $L'(3)$ is less than $\epsilon$ as shown in Fig. 2.16, the non-fudicial point is not replaced. This point is now re-named $L(3)$ and allowed to evolve at the next time instance. If, however, the distance $L'(3)$ is larger than $\epsilon$, the point is replaced as defined earlier. The evolution and replacement procedure is repeated until all the embedded vector points along the fudicial trajectory have been exhausted.

The largest Lyapunov exponent is then estimated by

$$\lambda_{\text{max}} = \sum_{l=2}^{N} \log \frac{L'(l)}{L(l-1)},$$

(2.61)

where $N$ is the total number of replacement steps required in a certain time series. The important point of this algorithm is that, through a simple replacement procedure that attempts to preserve orientation and minimize the size of the replacement vectors, it is possible to monitor the long-term behavior of a single principal axis vector.
Figure 2.16: A schematic diagram of the evolution and replacement procedure for estimating the largest Lyapunov exponent from a time series. The fudicial trajectory is shown by the thick solid line with the fudicial points $X_i$ encircled. The dashed lines show the evolution of the non-fudicial points after each time step. The largest Lyapunov exponent is computed from the growth of length elements (adopted from [6]).
2.11 Entropies

Entropy is another important measure in chaos analysis; it is similar to generalized dimensions in the way it is estimated in practice. In information theory [71, 55], the Shannon entropy $S$ is proportional to the uncertainty $H$ in equation (2.42) earlier used in the derivation of the mutual information between two random variables. The Shannon entropy can be expressed as

$$S \propto - \sum p_i \log p_i,$$

where $p_i$ is the probability of finding the system in the $i^{th}$ state [72]. $S$ is a measure of the amount of information required on average to describe the system.

The Kolmogorov Entropy $K$ is a measure by which chaotic motion can be characterized in phase-space [1]. When related to Shannon's entropy $S$, the Kolmogorov Entropy $K$ is proportional to the rate at which information about a dynamical system is lost in the course of time. For a chaotic system, $K$ is equivalent to $S$ when the probability $p_i$ is related to the likelihood of finding a vector point of the attractor in a certain volume element in phase space as defined in section 2.5.1. The Kolmogorov entropy is then

$$K = - \lim_{T \to \infty} \lim_{\epsilon \to 0} \lim_{d \to \infty} \frac{1}{d T} \sum_{i_1, i_2, \ldots, i_d} p(X_{i_1}, X_{i_2}, \ldots, X_{i_d}) \log p(X_{i_1}, X_{i_2}, \ldots, X_{i_d}).$$

When $K$ equals zero, we have an ordered system whereas when $K$ is infinite, we have a random system. When $K$ is a finite constant that is not equal to zero, we have a chaotic deterministic system [7].

In the next section, we define another entropy measure $K_2$ which is close to $K$, but easier to estimate because it can be formulated in terms of the correlation integral of section 2.6.1.
2.11.1 Entropy $K_2$

In 1983, Grassberger and Procaccia [7] defined a new quantity $K_2$ which has the properties that it is less than the Kolmogorov Entropy $K$ for a chaotic system, infinite for a random system, but finite for a chaotic system, i.e.,

$$K_2 = \begin{cases} 0 \leq K_2 \leq K < \infty & \text{for a chaotic system} \\ \text{infinite} & \text{for a random system} \\ 0 & \text{for an ordered system} \end{cases}$$  \hspace{1cm} (2.64)

Further, $K_2 > 0$ is a sufficient condition for chaos [7].

If we now define a generalization of $K$ to the set of order-$q$ Renyi entropies [73] as

$$K_q = - \lim_{T \to \infty} \lim_{\epsilon \to 0} \lim_{d_E \to \infty} \frac{1}{d_E T} \log \sum_{i_1, i_2, \ldots, i_{d_E}} p^q(X_{i_1}, X_{i_2}, \ldots, X_{i_{d_E}})$$

then with $q = 2$ and $d_E = 1$, the joint probabilities $p(X_{i_1}, X_{i_2}, \ldots, X_{i_{d_E}})$ are reduced to $\sum p_i^2$ which is equivalent to the correlation integral $C_{2,d_E}(\epsilon)$ as defined in section 2.6.1. The correlation integral in this case is parameterized by the embedding dimension $d_E$ because the expression in equation (2.65), which is equivalent to generalized entropies, is a function of the increasing embedding dimension $d_E$. Also, as previously discussed in equation (2.26), $C_{2,d_E}(\epsilon)$ is proportional to $\epsilon^{D_2}$. Consequently, combining equations (2.26) and (2.65) leads to [7]

$$\lim_{d_E \to \infty} C_{2,d_E}(\epsilon) \sim \epsilon^{D_2} \exp(-d_E T K_2),$$  \hspace{1cm} (2.66)

from which it is fairly straightforward to extract $K_2$. The most important property of $K_2$, however, is that it can be written in terms of the spatial correlation $C_2(\epsilon)$ between vector points on the attractor, as defined earlier in equation (2.22) in the calculation of the generalized correlation integrals. Therefore, an algorithm used for estimating the correlation integral and dimension will also yield the entropy $K_2$ with minor modifications. In the next section, we discuss a method of estimating $K_2$ from the correlation integrals at different embedding dimension.
2.11.2 Numerical Implementation of $K_2$

If we plot $\log C_{2,d_E}(\epsilon)$ in equation (2.66) as a function of $\log \epsilon$ for a series of embedding dimensions, we obtain straight lines in the linear regions which are displaced from each other by the factor $\exp(-d_E T K_2)$. Recall that the linear region corresponds to the value of $\epsilon$ bounding the plateau section of Fig. 2.13. For increasing embedding dimensions, the slope of these lines converges to the correlation dimension $D_2$. From the displacement factor described above, it is fairly easy to estimate $K_2$. Due to the dependence of $K_2$ on $C_{2,d_E}(\epsilon)$ in equation (2.66), we now parameterized $K_2$ as a function of the embedding dimensions $d_E$, i.e., $K_{2,d_E}$. We also estimate equation (2.66) only in the linear region (region III) of Fig. 2.13. The expression in equation (2.66) is also simultaneously evaluated at successive embedding dimensions $d_E$ and $d_E + Q$, where $Q$ is a positive constant selected to increase the embedding dimension. We then obtain

$$\log C_{2,d_E}(\epsilon) = -d_E T K_{2,d_E} + D_2 \log \epsilon$$

(2.67)

and

$$\log C_{2,d_E+Q}(\epsilon) = -(d_E + Q) T K_{2,d_E} + D_2 \log \epsilon.$$  

(2.68)

If equation (2.68) is subtracted from (2.67), we obtain

$$K_{2,d_E} = \frac{1}{TQ} \log \frac{C_{2,d_E}(\epsilon)}{C_{2,d_E+Q}(\epsilon)}.$$  

(2.69)

A plot of $K_{2,d_E}$ as a function of the embedding dimension $d_E$ converges to $K_2$, i.e.,

$$\lim_{d_E \to \infty} \lim_{\epsilon \to 0} K_{2,d_E} \sim K_2.$$  

(2.70)

Fig. 2.17 shows the values of $K_{2,d_E}$ for the Lorenz attractor. This same method is also used to estimate the remaining generalized entropies.
Figure 2.17: Values of Grassberger-Procaccia Entropy $K_{2,d_E}$ for the Lorenz Attractor (adopted from [72]).
Chapter 3

Background on speech production and processing

3.1 Introduction

3.1.1 Importance of Speech

The importance of speech communication in the life of human beings cannot be overemphasized. It undoubtedly constitutes one of the fastest and most convenient ways of communicating intents and emotions. The modes of speech production and communication in human beings have fascinated scientists for over fifty years. It has actually been documented that speech communication is preferred over all other human modes of interaction [74, 75]. Hence, speech is such an integral part of human life that many areas of technological advancement have been built around it. These technological advancements include speech synthesis, speech recognition, speech enhancement, speech coding, etc.

The main features of the human body that are primarily responsible for speech production are the lobes of the lungs, larynx, pharynx, trachea, nose, and various parts of the mouth [8], which are depicted in Fig. 3.1 and Fig. 3.2. Speech production involves the combination of different sounds from many sources within the speech production system shown in Fig. 3.1. These sounds, and the transitions between them, serve as the symbolic representation of the information content of any spoken message. By varying the source of the excitation and the nonlinear interaction of
the vocal apparatus (larynx, pharynx, etc.) in Fig. 3.1, humans have the ability to produce different sounds [76]. The vocal tract is an acoustic tube with non-uniform cross-sectional area and about 17 cm in length. It extends from the top of the trachea to the lips. During laminar or turbulent flow of air through the vocal tract, speech is generated in the form of acoustic waves. In electronic speech processing, the acoustic waves are recorded by a microphone in front of the mouth.

Speech communication in humans can be broadly classified into two parts. The first part involves the conversion of intentions or emotions from the speaker into neural activities in the brain. The neurons in the brain cells then send a message to the part of the human body that controls the articulatory mechanisms, which produce acoustic waves that contain encoded information present in the original intent. The second part involves a decoding of the received message by the listener. The received acoustic signals are converted into neural impulses that are transported to the brain. This suggests that a lot of inherent signal processing is involved in speech communication.

Speech communication in humans can be intrinsically described as discrete in nature by a concatenation from a finite set of symbols [77]. Every sound can be classified in symbolic form by phonemes. Each language has its own peculiar set of phonemes, usually numbering between 30 to 50. The English language, for example, can be represented by a set of about 42 phonemes.

In the next section, we will examine the different classifications of speech and their mode of production. The classifications will highlight the direction of our work and the class of speech we intend to focus upon in our analysis.

### 3.2 Classification of Speech

The different sounds produced by humans are classified as voiced, unvoiced, and mixed voiced or unvoiced plosives. In the subsequent three subsections, a presentation of these different classifications of speech with particular attention to air flow in the specific organ of the human body involved in speech production is given.
Figure 3.1: A cross-section of the human head, showing the vocal and nasal tracts (adopted from [7])
3.2.1 Voiced Speech

Voiced sounds, which normally include consonants and vowels such as "e" (the sound of the underlined letter in "ven"), are generally produced when air from the lungs is forced through the trachea in Fig. 3.1. The air is then modulated by the vibration of the vocal folds (also referred to as the vocal cords) located in the larynx. The vocal folds are two folds of tissue stretched across the opening in the larynx in Fig. 3.2 such that the front end of the folds are joined to the thyroid cartilage and the rear end to the arytenoid cartilages. Under muscular control, the arytenoids can move far apart, leaving openings in the vocal fold for activities such as breathing [77]. They can also be tightly closed together when one holds one's breath or during swallowing. It is also possible to hold the arytenoids, such that the vocal folds are almost touching. If air is now forced through this small openings, the vocal folds will start to vibrate, resulting in a modulation of the air. The rate of the vocal fold vibration is commonly termed the pitch or fundamental frequency, $F_0$ [77].

3.2.2 Unvoiced Speech

Unvoiced sounds, which normally include sustained consonants, are the result of air turbulence that is produced when air is forced through a constriction in the vocal tract at high velocity [77]. Such constrictions may be formed in the region of the larynx, as in the case of the "h" sound [8] or at many other vocal tract locations, such as the roof of the mouth, between the different parts of the tongue, or between the teeth and lips. Sustained consonant sounds that are produced due to air turbulence excitation, such as "s, f, z", are called fricatives [77].

When fricatives are being produced, the air becomes turbulent at a critical Reynolds number\(^1\), $Re_{\text{critical}}$. The square of the Reynolds number can be expressed as

$$Re^2 = \frac{4\rho^2 u_c^2}{\pi \mu^2 a_c}$$

(3.1)

where $u_c$ is the volume velocity at the constriction, $a_c$ is the area of constriction, $\rho$ is

\(^1\)Reynolds number is a dimensionless measure from fluid mechanics that describes the various types of fluid flow that may result when air is forced through a tube at high velocity. When $Re \leq Re_{\text{critical}}$, the flow is laminar whereas the flow is turbulent when $Re > Re_{\text{critical}}$. 

60
Figure 3.2: A cross-section of the human larynx (adopted from [7])
the fluid density and \( \mu \) is the viscosity [78]. For air flow in tubes with rough surfaces, \( Re_{\text{critical}} \) values are around 2000 [79]. Typical flow rates during fricative production are in the range of 200–500 \( cm^3 \) per second and the supraglottal constriction areas, i.e., the trachea, bronchi, lobes of the lung and diaphragm range between 0.075–0.4 \( cm^2 \). The flow rate and area data in the average human suggests that \( Re \) values for speech production ranges from the number 2700 to 12625 [80].

3.2.3 Mixed Voiced Speech

Mixed voiced and unvoiced plosives, such as "b", "v", are produced primarily from the build-up of pressure that occurs when the vocal tract is closed at some point for stop consonants. The release of this pressure produces a transient excitation of the vocal tract. Voiced plosives are a result of transient excitation in conjunction with vocal cords vibrations, whereas if the vocal cords are not vibrating during closure, unvoiced sounds are produced.

3.3 Speech Modeling

The use of speech offers many great advantages in the areas of man-machine interactions, such as telephone, communications and artificial intelligence. However, human speech production processes are so complicated, as alluded to earlier in section 3.2, that it is almost impossible to fully mechanically emulate human speech. Therefore, an ideal man-machine interaction cannot be fully realized. In order to advance man-machine communications, speech models have been developed [77]. Applications of digital signal processing techniques and processors have accelerated speech model developments. The digital representation of speech signals (in terms of speech models) in communication systems offers many advantages in speech data compression, quality, reliability and intelligibility in communication. Consequently, many mathematical models have been proposed to represent the speech production system [78].

Speech production models and their analysis can be generally classified into three main categories. They include:
- **Linear Models:** This is the linear representation of speech in terms of a source-system model. The vocal tract is approximated by a straight tube of varying cross-sectional areas. The linearity of the model is inherent in the assumption that the vocal tract can be modeled over a short time interval by a linear system consisting of an input signal, a transfer function and an output signal.

Over the years, many variations of the source-system model have been proposed. They include the source-system model introduced by Fant [81] almost 30 years ago, the canonical autoregressive decomposition model introduced by Kay et al [82, 83], the fractal-excited speech coders by Maragos [84], the linear combination of harmonic sinewaves of Kumaresan et al [85], and many more [86].

- **Nonlinear Models:** This is the category of other speech models that completely discard the source-system model. Instead, speech is modeled either in terms of nonlinear interactions between the excitation (vocal cord vibrations and turbulence) and the vocal tract or by phonemic description of sounds. These include the physiological model of Schroeter et al and the fluid flow model by Kaiser and Teager [58, 87].

- **Nonlinear Dynamical Models:** These models analyze and/or describe speech in terms of nonlinear dynamical system (chaos). They include the dynamical approach to speech processing by Tishby [27], analysis of vocal disorder with chaos [17], chaos analysis of fricative consonants [14] and multifractal description of unvoiced speech [63].

We shall now discuss some of these existing speech modeling techniques in relation to their relative advantages, disadvantages, and limitations. We will also introduce the dynamical systems approach to speech processing.
3.3.1 Linear Models

Source-System Model

In this section, we describe the source-system model. We examine each section of the block diagram of Fig. 3.3 to highlight various digital processing algorithms that have been developed to represent different parts and functions of the human speech production mechanism. We discuss the relative advantages as well as problems of the source-system model. We also examine some other models that were later proposed to avert some of the problems encountered in the source-system model.

The block diagram describing the source-system model of a speech production system is shown in Fig. 3.3 [77]. The system consists of a model of the vocal tract and the nasal tract shown in Fig. 3.1. The vocal tract is an acoustic tube that extends from the top of the trachea to the lips. The nasal tract is a path for the transmission of sound that extends from the velum to the nostrils.

The source-system model, also referred to as the Linear Predictive Coding (LPC) method, models speech as the output of a quasi-stationary linear system consisting of an all-pole filter in Fig. 3.3 excited either by a periodic train of pulses (spaced at a distance of a pitch period apart) for voiced speech or by broadband noise for unvoiced speech. In the digital processing model for speech production depicted in Fig. 3.3, the periodic train of pulses excitation into the time varying all-pole filter (for voiced speech) are generated by an impulse train generator while the broadband excitation (for unvoiced speech) is produced by a random number generator.

LPC is an analysis by synthesis technique. In the analysis stage, the predictor coefficients are extracted from the speech signal. These coefficients are then used to synthesize speech. In most cases, the synthesized speech is close to the original speech signal from which the predictor coefficients were extracted.

The recursive model used in the source-system model assumes that the vocal tract can be modeled over a short time interval by a linear system described by the transfer function

\[ H(Z) = \frac{G_h}{1 - \sum_{k=1}^{p} a[k]Z^{-k}}, \]

(3.2)
Figure 3.3: Digital processing model for production of speech signals (adopted from [9])
which relates the air flowing into the glottis to the speech sound emanating from 
the mouth. $G_h$ is the model gain and the short-term predictor coefficients $a[k]$ are 
obtained by linear prediction of the present speech sample from the past $p$ speech 
samples, hence, the term LPC. The predictor coefficients are updated every 10-20 
milliseconds or every two to three pitch periods. The poles in equation (3.2), i.e., the 
roots of the denominator polynomial, are used to represent the three to five resonant 
frequencies which characterize most speech sounds [88].

Source-system models have been used most extensively for speech analysis and 
synthesis because of the model's mathematically elegant linear representation of a 
rather complex natural system. LPC is generally regarded as capable of produc­
ing synthetic speech of high quality. Yet, in close listening, it has been found that 
with LPC, speech loses some of its naturalness during the analysis and synthesis 
procedures. This occurs even when the speech from which the predictor coefficients 
are obtained is of high quality [88]. The performance of LPC for speech synthesis 
becomes poor when some parts of the modeled speech signal contain unvoiced seg­
ments or when a speech segment was produced from the nasal cavity [89]. Most 
importantly, the performance of the LPC is also closely related to some of the sup­
porting algorithms that are needed in conjunction with the LPC. For example, the 
periodic pulses that excite the vocal tract transfer function in Fig. 3.3 are spaced 
at one pitch period apart. This implies that in the voiced section of a speech sig­
nal, a pitch period extraction algorithm correctly identifies the location as well as 
the correct distance of separation between the pulses. This can become very tricky 
and could involve synchronous or asynchronous speech synthesis where the train of 
pulses are specially placed to excite the voice speech segment. Also, a voicing deter­
mination algorithm (VDA) (which could range from a simple thresholding analysis 
to more complex algorithms based on pattern recognition) [90] is essential for LPC 
speech synthesis. A voicing determination algorithm VDA is primarily used to seg­
ment the speech signal into sections that may require either voice excitation (impulse 
train) or unvoiced excitation (broadband noise) during LPC synthesis. Therefore, 
any degradation in the performance of the pitch extraction or voicing determination 
algorithms ultimately affects the overall performance of the LPC.
In order to avoid some of the problems of LPC highlighted above, many improvements have been developed and new algorithms are being proposed. There have been several attempts to improve the naturalness of LPC synthesis. They include:

- The algorithm of Rosenberg et al. [89] attributed some of the inadequacies of the LPC method to the high peak factor of the synthetic speech waveform and proposed a time-varying non-impulse source for exciting the LPC to produce voiced sounds. The non-impulse source reduces the peak factor of the signal and at the same time introduces a certain amount of controlled irregularity in the excitation of the voiced speech.

- The recently introduced method of Evangelista [91] proposes replacing the train of pulses in Fig. 3.3 with waveforms generated from pseudo-periodic \( \frac{1}{f} \) processes. These waveforms have approximate \( \frac{1}{f} \) spectral behavior around the harmonics of a certain assumed fundamental frequency \( F_0 \). The generated, self-similar, pseudo-periodic signals have fluctuations or low level amplitude signals between each approximate pitch period which is typical of natural sound [91].

Canonical Autoregressive Model

In order to avoid all the rigorous tasks of pitch period determination, speech segmentation and speech synchronization explained in section 3.3.1, Kay and Nagesh [82] proposed to model speech (voiced and unvoiced) as the sum of sinusoids plus an AR process, i.e.,

\[
x[n] = \sum_{i=1}^{M} A_i \cos(2\pi f_0 n + \phi_i) + s[n]
\]  

for \( n = 0, 1, \ldots, N - 1 \) and \( s[n] \) is an AR process of order \( p \), i.e.,

\[
s[n] = -\sum_{k=1}^{p} a[k]s[n - k] + gw[n].
\]

Here, \( gw[n] \) is white Gaussian noise with variance \( \sigma^2_{nw} \). The parameters to be estimated are \( \{A_i, \phi_i, F_0, a[1], \ldots, a[p], \sigma^2_{nw}\} \). For voiced speech, the waveform of the prediction residual obtained by inverse filtering the speech signal by the predictor coefficient is quasi-periodic with the same pitch period as the speech wave. This method
therefore modeled voiced speech as a sum of harmonically related sinusoids with slowly varying fundamental frequency $F_0$ and arbitrary amplitudes $A_i$ and phases $\phi_i$. For unvoiced speech, the prediction residual from LPC resembles white noise, hence this method estimates the parameters of a mixed filter that jointly produces voiced or unvoiced speech given white noise and periodic sinusoids as input. The parameters in equations (3.3) and (3.4) are estimated using the maximum likelihood estimation (MLE) formulation coupled with the minimum description length (MDL) criterion [92].

A comparative performance test between the LPC and the canonical autoregressive model (CARM) was conducted in [83]. In a listening test, albeit not with many speech examples to be statistically conclusive, CARM was found to perform better than the LPC, i.e., the speech synthesis from CARM sounds very close to the original speech data. Also, CARM does not require additional algorithms such as pitch estimation and voicing determination algorithm (VDA). In that sense, the CARM was easier to implement. However, CARM involves the estimation of many parameters which makes it computationally intensive and, hence, inappropriate for real-time implementation.

### 3.3.2 Nonlinear Models

#### Fluid Flow Model

In the discussion that follows, the fluid flow approach of Teager and Kaiser [58, 87] which addresses the anomalies of the source-system model will be examined. The discussion centers on the inability of the source-system to fully explain other types of sounds that human beings are capable of producing. Examples of the anomalies of the source-system model are given and a model that is capable of resolving the anomalies of the source-system model is proposed.

Teager et al [58, 76, 87] in their work over the years have discarded the simplified source-system description of speech production. They contend that with our vocal tract, human beings are capable of producing a variety of sounds such as clicks, whistles, snores, barnyard sounds and other noises that do not fit the source-system description, because the source of these sounds are not glottal [58]. A major part
of their disregard for the source-system model comes from the model's inability to explain some of the nonlinear interactions between the walls of the vocal cavity. These nonlinearities, which are largely responsible for other sounds disregarded in the source-system model, are obviously more subtly difficult to physiologically observe or record since in electronic speech analysis, any attempt to insert transducers into the vocal cavity for the purpose of recording speech will alter or impair the speech production.

An example of some the anomalies that the general source-system model cannot explain is the Myna bird which has the ability to mimic repeated noises as well as to produce long passages of speech in a voice that is almost indistinguishable from that of its trainer. Yet, the bird has an anatomically short cochlea that is not suitable for frequency selection. The bird also lacks the conventional vocal tract as described in humans. Teager and Teager [58], also question the linear speech model with respect to:

- The inability of humans to produce sound when the mouth and throat are parched or after eating a dry cracker.
- The change in the human voice due to illness or loss of teeth that has absolutely nothing substantial to do with the vocal tract.

In an attempt to provide an explanation for the questions raised by their quest for the ultimate speech production model, Teager and Teager did not propose an overall model. An overall model would be incapable of representing all the phonemic subgroups encountered in human speech vocalization at once just like the source-system model. Instead, they pose a phonemenological explanation that can be separately applied to each phoneme. This is because each sound is generated at various locations of the vocal cavity by taking advantage of the local shapes, flow instability due to the movement of the larynx, tongue, and lips, as well as the mental state of the speaker.

The proposed Teager model is capable of providing a complete description of the speech production system. However, it is rather cumbersome to implement, since even a simple sentence in the English may be comprised of a fairly large number of phonemes.
Physiological Model

In contrast to the simplified LPC, where the properties of the recorded speech signal are being modeled, Schroeter and Sondhi proposed a model for speech that is dependent on the properties of the vocal apparatus i.e., the pressure in the lungs, tension in the vocal cord, the coupling of the nasal cavity, etc. They proposed a three-part model that separately considers the geometry of the vocal and nasal tracts, wave propagation in the tracts as well as the sound sources (vocal cord vibration and turbulent air flow) and their interactions with the tract. Based on X-ray measurements of the articulatory system, a vocal tract geometry that maps the positions of the elements of the articulatory system, i.e., tongue, lips, glottis, etc. was defined. The nasal tract was modeled by a modification of the method of Maede [93] in which a side cavity, to simulate the nasal sinuses, was attached to an acoustic tube representing the nasal cavity. This model was efficiently implemented in [94] by coupling a Helmholtz resonator to the nasal tract [78]. With this model, the acoustic properties of the vocal and nasal tracts were computed under the assumptions that:

- the acoustic wave propagation inside the vocal tract can be simulated as plane waves.
- the vocal tract is flexible; it can be straightened out and approximated by a variable-area tube.

The turbulent excitation of the vocal tract is handled by automatically injecting noise of correct amplitude into the appropriate place within the vocal tract.

This Schroeter/Sondhi model has the advantage of automatically adapting to sudden changes in pitch and glottal waveforms since it is derived from a physiological point of view. The extracted model parameters are also suitable for low-bit rate speech coding. The unfortunate part is that although this model discards the linear source-system model, linearity is inherently built into the model. This is due to the assumption that the acoustic wave tube (a model of the vocal tract) can be approximated by a straight variable-area tube in order to solve the otherwise nonlinear equations which would have resulted. Therefore, the assumption in this model contradicts the highly controversial argument that a complete description of
speech requires the solution of nonlinear, partial differential equations with varying boundary conditions [76, 27, 58].

3.3.3 Nonlinear Dynamical Models

Teager Model

The evidence of nonlinearity in speech signals has been in debate for over thirty years. The early models of the vocal tract by Ishizaka and Flanagan [95] have nonlinearities incorporated into them. Teager and Teager in their earlier work described the human voice as a "complex wind instrument that has nonlinear regenerative oscillators whose amplification methodology can be related to the interaction separated flows (along the vocal tract walls) with frequency selective cavities" [58]. This suggests that sound is generated throughout the vocal tract, from the lobes of the lungs to the lips, acting as nonlinear coupled system. Recently, evidence of nonlinearity in speech, in support of Teager's work, has been discussed by Maragos, Quatieri and Kaiser [84, 18, 96].

Dynamic Model

Speech as a nonlinear phenomena has been qualitatively described in terms of fluid flow and turbulence [58, 76]. In speech modeling literature, the word turbulence is often use in a qualitative sense to mean non-laminar flow. However, in dynamical modeling of what is technically defined as turbulence, the dimension of the chaos model is expected to be so high as to discourage such an approach. Yet, the turbulent velocity field of fluid flow has been extensively investigated to reveal a low-dimensional chaotic structure [15, 23, 97].

Hence, there has been considerable interest towards applying nonlinear dynamical system theory and principles to speech analysis and modeling since speech production involves the turbulent flow of air through the vocal tract. In fact, evidence of low-dimensioned chaos such as strange attractors, positive Lyapunov exponents, and bifurcations, have been found in newborn infant cries [20]. Voice fluctuation and vocal tract instabilities have also been modeled with chaos [16, 23].

Dynamical speech modeling and analysis can be broadly classified into two groups.
They include the following:

- global or local approximation of the chaotic attractors of speech signals [27, 98]. This, in most cases, ends up being a neural network-like model with chaotic input parameters.

- models or analysis techniques that are only well suited to specific types of speech, i.e., fricatives [14].

We will now introduce some of these nonlinear dynamical systems approach to speech modeling and analysis.

**Neural network model**

Tishby [27] suggested the possibility of modeling speech as an output of a chaotic dynamical system. He proposed an analysis-synthesis method of speech signals based on a neural network-like nonlinear signal prediction. Given a discrete time signal $x[1], x[2], \ldots, x[k]$, the correlation dimension $D_2$ is estimated as a function of the embedding dimension $d_E$ until $D_2$ appears to have converged as discussed in the last chapter. If we denote the value of the embedding dimension at the convergence by $h$, then one can construct embedded vectors $X_i$'s like those in equation (2.38) of length $h$ which according to Takens' Theorem [56] will have the same geometrical and dynamical properties as the attractor from the time series with all its components present.

A neural network-like model was then used to estimate an output sample $x[k+1]$ by minimizing the mean squared prediction error

$$E(w) = (x[k+1] - \sum_{j=1}^{k} w_j f(X_j))^2$$

where $w$ are the weights in a neural network that are adjusted as the network minimizes the error in equation (3.5). $f(X_j)$ is a nonlinear $h$-dimensional function. The error is minimized by adjusting the parameters $w_j$ in the network using a standard multidimensional optimization method. The result of such a nonlinear predictor reveals a model that better captures the nonlinearity inherent in speech sounds [27, 99].
A 3 dB improvement in the root mean squared (RMS) residual error was achieved when the nonlinear predictor method was compared with LPC of the same order [27].

The calculations in equation (3.5) were made from a global approximation of the speech attractor. Unfortunately, the number of speech samples $x[n]$ needed for these calculations is on the order of thousands of time series samples. An improved nonlinear predictor of speech was later developed by Sidorowich [98] which was based on local approximation. The computation time coupled with a speaker dependent training requirement of the nonlinear predictor has seriously curtailed the use of this method of speech modeling. When the neural network is trained with a set of embedded point vectors $X_i$ from one set of speech data and then used to model new speech, the network has to be trained all over again before it can be used to meaningfully model another set of speech data.

Other Models

Other studies have used nonlinear deterministic techniques in the analysis and characterization of the vocal fold vibration in the normal and pathological voiced speech [21, 23, 16]. Many more studies have also performed chaos analysis on special speech sounds such as fricatives [14, 16, 19]. Recently, Narayanan et al [14] estimated the correlation dimension $D_2$ as well as positive Lyapunov exponents and finite entropy $K_2$ for several fricative sounds. These chaos measures are important in nonlinear speech modeling because the existence of low correlation dimension, positive Lyapunov exponents and finite entropy are indicative of chaotic behavior in speech sounds.

Due to the complexity involved in speech production, in terms of fluid flow and the nonlinear interaction within the vocal apparatus, we hypothesize the presence of different fractal regimes in the chaotic attractors of unvoiced speech sounds. These fractal regimes (multifractal structures) are not revealed by the estimation of only the correlation dimension as done by Narayanan et al [14]. The focus of this thesis is motivated by a desire to find a more complete low-dimensional characterization of the dynamical description of speech signals. Chaotic signals typically have a
multifractal orbit. We therefore propose the estimation of the spectra of dimensions to explain the multifractal nature of fricative sounds. This method is based on the estimation of the generalized dimensions from a phase-space reconstructed vectors of a single scalar variable of speech signal [63, 56].

The goal of the thesis is to ultimately model speech sounds as a chaotic multifractal and quantitatively compare the performance of such a model with the classical all-pole model described in section 3.3.1. The estimation of the multifractal parameters is a first step in the direction of the chaotic multifractal model.
Chapter 4

\( f(\alpha) \) Singularity spectrum formalism from Generalized Dimensions \( D_q \)

4.1 Introduction

In chapter 2, we reviewed how the correlation dimension \( D_2 \) was generalized to an infinite hierarchy of generalized dimensions, \( D_q \) [100]. Any chaotic attractor with non-constant, monotonic decreasing generalized dimensions, i.e. \( D_q \geq D_{q'} \) for \( q < q' \) is a multifractal [101, 102, 61, 11, 3, 1]. In this chapter, we examine an alternative characterization of multifractal behavior in chaotic dynamical systems called the singularity spectrum \( f(\alpha) \) [103, 101]. We use a generalization of the Koch curve [104] to introduce various concepts and terms related to the scaling behavior of multifractals. Finally, we shall discuss some of the problems encountered in evaluating the \( f(\alpha) \) singularity spectrum from the generalized dimensions \( D_q \) based upon the fact that both \( f(\alpha) \) and \( D_q \) are related by a Legendre transformation [100].

The generalized dimensions \( D_q, \ -\infty < q < \infty \), characterize the different fractal regimes present in a chaotic attractor [54]. Dimensions beyond the traditional correlation dimension, \( D_2 \), are needed since two attractors with the same correlation dimension can be quite different visually. Multifractal attractors consist of many intricately intertwined fractals, each with a different fractal dimension. The strength or denseness of each fractal set is measured by the singularity strength or scaling exponent \( \alpha \). The fractal dimension of the fractal sets with singularity
strength between $\alpha$ and $\alpha + d\alpha$ is denoted by $f(\alpha)$, the singularity spectrum.

### 4.2 Singularity Spectrum $f(\alpha)$

The generalized dimensions $D_q$ constitute a set of order $q$ Renyi's dimensions. Hentschel and Procaccia in 1983 [54] defined this information theoretic concept in the context of a chaotic dynamical system. The partitioning of the attractor of a chaotic system in phase-space can be described in two ways. Assume we are given an $N$-point chaotic sequence and that the phase space is partitioned by volume elements of diameter $\epsilon$. The probability of finding a vector point on the attractor in the volume element number $i$, $(i = 1, 2, 3, N(\epsilon))$, is given in equation (2.18), i.e.,

$$p_i(\epsilon) = \lim_{M \to \infty} \left( \frac{M_i}{M} \right), \quad (4.1)$$

where $M_i$ is the number of vector points that fall in a particular volume element $i$ and $M$ is the total number of vector points on the attractor of the $N$-point chaotic sequence. Alternatively, if we consider the attractor itself in phase space as a geometric object and cover it with volume elements of diameter $\epsilon$, then we can define the fraction of the mass of the geometric object called $p_i(\epsilon)$ that falls within volume element $i$. Let us assume that the mass $p_i(\epsilon)$ in the $i^{th}$ volume element is proportional to $\epsilon^{\alpha_i}$, i.e.,

$$p_i(\epsilon) \propto \epsilon^{\alpha_i} \quad (4.2)$$

where $\epsilon$ is the diameter of volume element $i$ and $\alpha_i$ is the scaling exponent associated with the volume element $i$. As the diameter of the volume element decreases, the number of volume elements $N(\epsilon)$ needed to cover the attractor increases. The number of volume elements with scaling exponents in the range from $\alpha$ to $\alpha + d\alpha$ is now denoted by $J(\alpha)$. It is proportional to $\epsilon^{-f(\alpha)}$, i.e.,

$$J(\alpha) \propto \epsilon^{-f(\alpha)}. \quad (4.3)$$

$f(\alpha)$ is called the singularity spectrum. Due to the similarity between equations (2.16) and (4.3), $f(\alpha)$ is sometimes referred to as the fractal dimension of volume elements with scaling exponent $\alpha$. 

76
In the next section, we will expand on the geometric distribution of points on chaotic attractors by using the Koch curve to describe multifractal scaling properties. It is important to note that the \( f(\alpha) \) singularity spectrum, albeit explained with the aid of the Koch curve, is not unique to this curve. In fact, the singularity spectrum method of characterizing strange attractors has been used extensively with experimental data. Examples include the study of turbulent energy dissipation in fluid flow by Meneveau et al. [101], the hydrological characterization of river basins by Veneziano et al. [105], and zooplankton biomass variability by Pascual et al. [11].

4.3 Multifractal characterization of the Koch curve

In this section, we introduce the concept of multifractals in terms of the scaling exponent \( \alpha \) and singularity spectrum \( f(\alpha) \). We examine how multifractality is produced with the aid of the triadic Koch curve, and discuss the relationship between the \( f(\alpha) \) singularity spectrum and and the generalized dimensions \( D_q \) that were reviewed in chapter 2.

Consider the uniform triadic Koch curve [104] whose ends are at 0 and 1, i.e., \( k = 0 \) the top left curve in Fig. 4.1 (a). Let us associate a uniform probability density function (pdf) with this Koch curve at \( k = 0 \) as shown in Fig. 4.1 (b). The uniform pdf associated with \( k = 0 \) is shown at the uppermost right column of Fig. 4.1. As we iterate the triadic Koch curve, we will be examining Figures 4.1(a) and 4.1 (b) simultaneously. We construct the (triadic) Koch curve shown in Fig. 4.1, where \( k \) is a discrete index that represents the stages of the construction.

At \( k = 0 \), we start with a straight line segment of unit length. The construction methodology is such that at each iteration \( k \), the number of straight line segments is increased by a factor of 4 and the length of each segment \( \epsilon \) is decreased by \( \frac{1}{3} \) as shown in Fig. 4.1 with \( k = 1, 2 \). By considering the first three iterations of the Koch curve shown in Fig. 4.1(a), we find that after \( k \) iterations, the number of line segments \( N_k(\epsilon) \) is equal to \( 4^k \) and the length of each segment \( \epsilon_k = (\frac{1}{3})^k \) decreases as
Figure 4.1: (a) The first three iterations of the triadic Koch curve. Notice the increase in the number of segments as well as the decrease in the length of each segment as \( k \) increases (adopted from [9]). (b) shows the associated uniform distribution after the first three iterations.
Figure 4.2: (a) The first three iterations of the non-uniform triadic Koch curve. (b) Shows the associated geometric mass distribution (pdf) of the non-uniform triadic Koch curve. Here $p_1 = 0.35$ and $p_2 = 0.15$. The original uniform pdf is displayed in the upper-left panel. The resulting pdf after three iterations of the Koch curve is shown column-wise from top to bottom. The vertical-axis corresponds to the scaled pdf of the length segments and therefore the area below the density curve provides the probability in any subinterval (adopted from [10]). The pdf has been scaled to clearly show the resulting intermittency as the non-uniform triadic Koch curve is iterated.
increases.

Recall that associated with the line segment at iteration \( k = 0 \) is a uniform pdf. Also, if we now associate the same uniform pdf with each new line segment generated during the \( k \) iterations, then all the line segments at the \( k \)th iteration will be identical in length, and, therefore, we still have a uniform pdf after each iteration. Notice that the pdfs associated with the iterations \( k = 1, 2 \) in Fig. 4.1(b) are all uniform. This is the uniform triadic Koch curve [104]. If we now invoke the expression in equation (2.16) for the calculation of the box-counting or fractal dimension \( D_0 \) for the Koch construction with uniform pdf, then

\[
D_0 = \lim_{\varepsilon \to 0} \frac{\log N_k(\varepsilon)}{\log \left( \frac{1}{\varepsilon} \right)} = \frac{\log 4}{\log 3} \approx 1.26. \tag{4.4}
\]

If, on the other hand, we generalize the Koch construction [106] to two unequal line segments of length \( \varepsilon_1 \) and \( \varepsilon_2 \) such that \( \varepsilon_1 > \varepsilon_2 \) and associate a fraction (\( p_1 \) and \( p_2 \), respectively) of the uniform pdf at \( k = 0 \) in Fig. 4.2 (b) with each of the two line segments such that \( 2p_1 + 2p_2 = 1 \), then the line segment length as well as the overall pdf of this Koch construction will be non-uniform as the iteration progresses. Fig. 4.2 (a) shows the first three iterations of the non-uniform triadic Koch construction.

In particular, for \( k = 1 \), there are four line segments; two with segment length \( \varepsilon_1 \) and probability \( p_1 \) and the other two have segment length \( \varepsilon_2 \) with probability \( p_2 \). When \( k = 2 \), the length and associated pdf of each of the new line segments from the previous iteration are rescaled again by \( \varepsilon_1 \) or \( \varepsilon_2 \) and \( p_1 \) or \( p_2 \). There are now 16 line segments; four with probability \( p_1^2 \), four with probability \( p_2^2 \) and eight with probability \( p_1p_2 \). The process is then repeated ad infinitum. In this example, each set of line segments with the same probability constitutes a fractal and the entire collection is a multifractal. Fig. 4.2 (b) shows the resulting pdf or geometric mass distribution associated with the generalized Koch curve after the first three iterations with \( p_1 = 0.35 \). Notice the intermittent and self-similar structure of the non-uniform pdf. Since it is fractal in nature, a magnification and stretching of a small section of the pdf at the eighth iteration in Fig. 4.3 will resemble the entire pdf structure.

Another important property of the pdf considered is the existence of singular points. From Fig. 4.2 (b), as the iteration progresses, i.e., \( k \to \infty \), some points
in the pdf increase beyond bounds while others shrink towards zero. These points are the singularities in the pdf and are characterized in strength by \( \alpha_i, i = 1, 2, \ldots \), where the index \( i \) refers to each set of line segments with the same length. Fig. 4.3 shows the non-uniform pdf associated with the generalized triadic Koch curve at the eighth iteration.

An example of the singularities just described is found in the invariant density of the Logistic map in Fig. 2.6. The singularities in the invariant density of the Logistic map occur at \( x = 0 \) and \( x = 1 \). The scaling exponents for the Logistic map are \( \alpha = \frac{1}{2} \) and 1, respectively [3].

In the case of non-uniform triadic Koch curve, the probability \( p_{ik}(\epsilon) \) of the number of points on the attractor that falls in volume element \( i \) after the \( k^{th} \) iteration is related to the pdf assigned to the different length segments by \( p_1^i p_2^{k-i} \) and \( \epsilon_k = \left( \frac{1}{3} \right)^k \) [106]. If we now substitute \( p_{ik}(\epsilon) \) and \( \epsilon_k \) into equation (4.2), one obtains [1]

\[
p_1^i p_2^{k-i} \propto \left( \frac{1}{3} \right)^{\alpha_i k},
\]

which gives

\[
\alpha_{ik} = \frac{i \log p_1 + (k - i) \log p_2}{k \log \frac{1}{3}}.
\]

(4.6)

where \( \alpha_{ik} \) is the scaling exponent for the \( i^{th} \) volume element at the \( k^{th} \) iteration. Also, the number of volume elements with probability \( p_{ik}(\epsilon) \) after \( k \) iterations is \( J_k(\alpha_i) = 2^k \binom{k}{i} \), where \( \binom{k}{i} \) is the binomial coefficient. If we substitute \( J_k(\alpha_i) \) into equation (4.3) and use Stirling’s approximation \( \log k = k \log k - k \) [107] in evaluating the factorial terms, one obtains

\[
f(\alpha_{ik}) = \frac{-[k \log k - (k - i) \log(k - i) - i \log i + k \log 2]}{k \log \frac{1}{3}}.
\]

(4.7)

A plot of \( \alpha_{ik} \) versus \( f(\alpha_{ik}) \) for a large number of iterations \( k \) will produce the \( f(\alpha) \) singularity spectrum for the Koch curve. A plot of \( \alpha \) versus \( f(\alpha) \) in equations (4.6) and (4.7) with \( p_1 = 0.35 \) and \( k = 200 \) is shown in Fig. 4.4.

In the next section, we will discuss some of the characteristics of the singularity spectrum \( f(\alpha) \) using the plot of \( \alpha \) versus \( f(\alpha) \) for speech data.
Figure 4.3: Shows the associated scaled pdf of the non-uniform triadic Koch curve at the eighth iteration, i.e. $k = 7$. Notice the intermittent and self-similar structure (adopted from [10]). Also notice the points in the pdf that tend to increase beyond bound while others shrink towards zero. The pdf has been purposely scaled to show the intermittent nature of the resulting pdf.
4.4 Characteristics of $f(\alpha)$

The plot in Fig. 4.4 is typical for the $f(\alpha)$ singularity spectrum. The $f(\alpha)$ curve is always convex with a single maximum. The extrema $\alpha$ values, i.e. where $f(\alpha) = 0$, are termed $\alpha_{\text{min}}$ and $\alpha_{\text{max}}$. Both $\alpha_{\text{min}} = D_{\infty}$ and $\alpha_{\text{max}} = D_{-\infty}$ are scaling exponents for the densest and the most rarefied areas of the multifractal attractor, respectively [101, 108, 2]. When $\alpha$ is less than 1, the corresponding singularity point in the invariant density of the attractor grows beyond bound whereas when $\alpha$ is greater than 1, the corresponding singularity point in the invariant density tends to zero. In essence, the variable $\alpha$ reflect spikes and zeros in the invariant density of the multifractal attractor. The maximum of $f(\alpha)$ corresponds to the box-counting dimension. From Fig. 4.4, we see that for the Koch curve, $\max_{\alpha} f(\alpha) = D_0 = 1.26$, which agrees with equation (4.4).

In the next section, we consider the relationship between the $f(\alpha)$ singularity spectrum and the rest of the generalized dimensions $D_q$. We also describe how the generalized dimension $D_q$ values are estimated from real data. This will become important as we need to compute the transformation from $D_q$ to $f(\alpha)$.

4.5 Relationship between $f(\alpha)$ and $D_q$

Covering the attractor of a multifractal chaotic system with volume elements of diameter $\epsilon$, can be used to evaluate the generalized dimensions $D_q$ in equation (2.29) as

$$D_q = \frac{1}{q-1} \lim_{\epsilon \to 0} \frac{1}{\log \epsilon} \log B_q(\epsilon) \quad -\infty < q < \infty \quad (4.8)$$

where

$$B_q(\epsilon) = \sum_{i=1}^{N(\epsilon)} p^q_i, \quad (4.9)$$

and $p_i$ is the probability of finding vector points on the attractor in the $i^{th}$ volume element. The sum in equation (4.9) is over all $N(\epsilon)$ volume elements of diameter $\epsilon$ needed to cover the attractor.

If we now substitute equations (4.2) and (4.3) into equation (4.9), one obtains in
Figure 4.4: Shows the Singularity spectrum $f(\alpha)$ for the triadic Koch curve with $p_1 = 0.35$. Notice that the maximum of $f(\alpha)$ corresponds to the box-counting dimension $D_0 \approx 1.26$. 
the limit \([3, 1]\) for small \(\epsilon\),

\[
B_q(\epsilon) = \int e^{-f(\alpha)} \epsilon^{q\alpha} d\alpha. \tag{4.10}
\]

The first factor, \(e^{-f(\alpha)}\), in the integrand in equation (4.10) gives the number of volume elements with scaling index \(\alpha\) whereas the second factor, \(\epsilon^{q\alpha}\), is the \(q\)th power of the probability associated with the scaling index \(\alpha\).

In evaluating the integral in equation (4.10), we follow the argument in [100] that since \(\epsilon\) is supposed to be very small, the numerical value of the integrand is largest when the combined exponent of \(\epsilon\), \(q\alpha - f(\alpha)\), is smallest. That is, for each \(q\), there is a corresponding \(\alpha\), termed \(\alpha^* = \alpha(q)\) for which

\[
\left[ \frac{d}{d\alpha} (q\alpha - f(\alpha)) \right]_{\alpha = \alpha^*} = 0, \tag{4.11}
\]

and

\[
\left[ \frac{d^2}{d\alpha^2} (q\alpha - f(\alpha)) \right]_{\alpha = \alpha^*} > 0. \tag{4.12}
\]

From equations (4.11) and (4.12), it follows that

\[
q = \left[ \frac{df(\alpha)}{d\alpha} \right]_{\alpha = \alpha^*}, \tag{4.13}
\]

and

\[
\left[ \frac{d^2f(\alpha)}{d\alpha^2} \right]_{\alpha = \alpha^*} < 0, \tag{4.14}
\]

which reveals that the \(f(\alpha)\) singularity spectrum curve must be convex.

If we now substitute the value \(\alpha = \alpha^* = \alpha(q)\) into equation (4.10) to evaluate the integral, then one obtains [100]

\[
B_q(\epsilon) = \epsilon^{q\alpha^* - f(\alpha^*)}. \tag{4.15}
\]

If we also substitute this \(B_q(\epsilon)\) into equation (4.8), one obtains [100]

\[
(q - 1)D_q = q\alpha - f(\alpha) \tag{4.16}
\]

where the subscript \(*\) has been dropped. This change from the variables \(q\) and \(D_q\) on the left hand side of equation (4.16) to \(\alpha\) and \(f(\alpha)\) on the right hand side is an
example of the Legendre transformation [1]. For a rigorous derivation of equation (4.16), see Benzi et al. [97], Meneveau et al. [101] and Halsey et al. [100].

The $f(\alpha)$ singularity spectrum and the generalized dimension $D_q$ are theoretically smooth functions of $\alpha$ and $q$. The $D_q$ in equation (4.8) are related to the singularity spectrum $f(\alpha)$ by the Legendre transformation [100, 3] given from equation (4.16) in the following steps:

\[
\begin{align*}
\text{Define} & \quad \tau(q) = (q - 1)D_q, \\
\text{then} & \quad \alpha(q) = \frac{d}{dq}\tau(q) = D_q + (q - 1)\frac{d}{dq}D_q \\
\text{and} & \quad f(\alpha) = f(\alpha(q)) = q\alpha(q) - \tau(q).
\end{align*}
\]

(4.17) (4.18) (4.19)

Thus, if we know $D_q$, we can find $\alpha(q)$ and $f(\alpha)$ from equations (4.18)-(4.19) and vice versa.

Calculating the $D_q$ and $f(\alpha)$ from real data can be problematic. For example, $\alpha_{\min}$ and $\alpha_{\max}$ defined in the previous section correspond to the asymptotic values of the generalized dimensions, i.e., $\alpha_{\min} = D_{\infty}$ and $\alpha_{\max} = D_{-\infty}$. Typically, $D_{-\infty}$ and $D_{\infty}$ are almost impossible to extract using equation (2.30) on real data because they require infinite data records as explained in equation (2.29). Although theoretically, the $D_q$ in equations (2.29) and (4.19) are known for all values of $q$, in practice, they are only computed for a finite range $q \in [q_{\min}, q_{\max}], q_{\min}, q_{\max} \ll \infty$ and along a sampled grid with spacing $\Delta q$, i.e., $q = q_{\min}, q_{\min} + \Delta q, q_{\min} + 2 \Delta q, \ldots, q_{\max}$. Large record lengths are needed to compute $D_q$ when $|q| \gg 0$. Even when the recorded data lengths are long enough to give valid $D_q$, the most efficient box-counting algorithm [47] for computing each value of $D_q$ has a computational complexity of $O(N \log N)$, where $N$ is the (large) record length of the time series.

Hence, generally, the number of computed samples of $D_q$, $\frac{|q_{\max} - q_{\min}|}{\Delta q}$, is kept small by making the range $|q_{\max} - q_{\min}|$ small or by coarsely sampling $D_q$ using large $\Delta q$.

In the next section, we examine some of the problems of published procedures for evaluating the singularity spectrum $f(\alpha)$ given sampled generalized dimensions $D_q$.  

86
4.6 Evaluating $f(\alpha)$ from $D_q$

In this section, we discuss the importance of evaluating the $f(\alpha)$ singularity spectrum for real systems. We also examine some of the problems of existing algorithms using the Legendre transform of $D_q$ to compute $f(\alpha)$ in equation (4.19). We propose a new method that is based on digital filter design theory [109].

Based on the work of Grassberger et al. [44] and others [100], and the development of faster search algorithms mentioned in [47, 61], it has become increasingly attractive to evaluate $D_q$ from attractors arising from physical systems. Therefore, $f(\alpha)$ curves are usually determined by the Legendre transform [101] of equation (4.19). However, such an operation involves either finely sampling $D_q$ using small $\Delta q$ defined in the previous section or interpolating the $D_q$ curve since, in most cases, the $D_q$ are empirically calculated or coarsely sampled values along a smooth $D_q$ curve. The Legendre transform also involves the evaluation of the derivative $\frac{d}{dq}$ in the expression for $\alpha(q)$ and $f(\alpha(q))$ in equations (4.18) and (4.19). Accurately approximating the derivative and Legendre transform in equations (4.17) and (4.19) using a finite range of sampled values of $D_q$ is non-trivial. It becomes especially troublesome when the $D_q$ versus $q$ curve has discontinuous derivatives, as can happen when there are "phase transitions" in the attractor [110, 111]. Phase transitions in a chaotic attractor introduce a small amplitude, high frequency content into the otherwise smooth, low frequency $D_q$ curve. In order to illustrate some of the problems of the existing methods of computing $f(\alpha)$ from samples of $D_q$, we introduce the classical Cantor set in the next paragraph.

4.6.1 Cantor Set

The Cantor set starts with a unit interval as shown in Fig. 4.1 with $k = 0$. We make the same assumption that a uniform pdf is associated with the unit interval just as in section 4.3. In the first iteration, the unit interval is replaced with two intervals, $\epsilon_1$ and $\epsilon_2$, each of length $\frac{1}{3}$. Each of these interval is assigned equal fraction of the uniform pdf, i.e., $p_1 = \frac{1}{2}$. At the next iteration, each of the two intervals from the previous iteration is rescaled. Each interval is replaced by another two intervals of
length $\frac{1}{5}$ with equal fraction of the uniform pdf. At the next stage of the construction, the rescaling procedure is repeated on each of the newly created sub intervals. The process is repeated ad infinitum. If $\epsilon_1 = 0.408$ and $\epsilon_2 = \epsilon_1^2$ with $p_1 = \frac{1}{2}$ for each of the intervals, then the generalized dimensions $D_q$ of the Cantor set has been derived in closed form [10] to be

$$D_q = \begin{cases} \frac{\log[0.5(\sqrt{(1+2q+q^2)}-1)]}{(1-q)\log \epsilon_1} & q \neq 1 \\ 0.515 & q = 1 \end{cases} \quad (4.20)$$

These generalized dimensions are plotted in Fig. 4.5(a).

In the next section, we use this closed form expression of the generalized dimensions of the Cantor set to describe some of the problems that may be encountered when using the Legendre transform of the generalized dimensions $D_q$ to compute the singularity spectrum $f(\alpha)$. We sample the closed form expression of $D_q$ in equation (4.20), for $q = q_{\min}, q_{\min} + \Delta q, q_{\min} + 2\Delta q, \ldots q_{\max}$, using different values of sample spacing $\Delta q$. We compute the singularity spectrum $f(\alpha)$ to simulate the results of existing $D_q \rightarrow f(\alpha)$ algorithms.

### 4.6.2 Problems of evaluating $f(\alpha)$ from $D_q$

Ideally, we want the Legendre transformation of the generalized dimension $D_q$ curve to produce a clean singularity spectrum $f(\alpha)$ curve with stable endpoints as shown in Fig. 4.4. However, the Legendre transform of coarsely sampled values of $D_q$ produces errors in $f(\alpha)$. Fig. 4.5 (a) shows the $D_q$ curve for the Cantor set computed by sampling the closed form expression in equation (4.20). To represent an ideal case, the closed form expression for the $D_q$ curve [10] was finely sampled using $\Delta q = 1$ for $q$ in the relatively large range of $q \in [-36, 36]$, i.e. $q = -36, -35, \ldots, 35, 36$. The corresponding singularity spectrum $f(\alpha)$ plotted in Fig. 4.5 (b) was evaluated using a 5th order min-max differentiator\(^1\) [109]. Notice also that the maximum value

\(^1\)The differentiator was designed assuming passband cut-off frequency $F_p = 0.10$ and stopband $F_s$ between 0.25 and 0.5, and error tolerances $\delta_p = 0.1$ and $\delta_s = 1 \times 10^{-6}$. The design of the min-max differentiator with the above specifications will be discussed later in section 4.9. It is important, however, to notice that the passband of the differentiator is greater than the bandwidth of the region of concentration of most of the energy in the spectrum of the $D_q$ in Fig. 4.6.
of the \( f(\alpha) \) in Fig. 4.5 (b) corresponds to \( D_0 = 0.538 \) and that the two extrema \( \alpha_{\text{min}} \) and \( \alpha_{\text{max}} \) (corresponding to the values of \( \alpha \) where \( f(\alpha) = 0 \)) give the correct asymptotic values of \( D_{\infty} = \alpha_{\text{min}} = 0.387 \) and \( D_{-\infty} = \alpha_{\text{max}} = 0.773 \) [10] in equation (4.20). Fig. 4.5 (b) is very close to the ideal singularity spectrum, and will be used for comparison purposes in later examples. In the next section, we discuss some of the existing algorithms used in the determination of \( f(\alpha) \) in relation to their relative advantages and limitations.

### 4.7 Existing Algorithms for evaluating \( f(\alpha) \)

A number of algorithms have been proposed to address the problem of accurate evaluation of \( f(\alpha) \). These algorithms can be broadly classified into two categories.

- The first set of algorithms directly evaluate the \( f(\alpha) \) singularity spectrum from real data without resorting to the intermediate Legendre transform [103, 112, 113]. These algorithms are based on direct computation of \( \alpha \) and \( f(\alpha) \) from the curve of log-log plots of \( \log(p) \) versus \( \log(J(\alpha)) \) and \( \log(J) \) versus \( \log(J) \) in equations (4.2) and (4.3).

Recall that due to the similarity between equations (2.16) and (4.3), \( f(\alpha) \) is sometimes referred to as the fractal dimension. This leads to an unfortunate mathematical ambiguity in the application of the direct \( f(\alpha) \) evaluation methods to experimental data, i.e., is \( f(\alpha) \) the Hausdorff dimension or the box counting dimension? Also, the direct \( f(\alpha) \) evaluation algorithms suffer from large errors because of logarithmic corrections which arise from the dependence of equation (4.3) on \( \alpha \) [114].

Other direct \( f(\alpha) \) computation algorithms that were proposed recently include the algorithms due to Chhabra-Jensen [103], Arneodo et al. [112] and Veneziano et al. [105]. The algorithms by Chhabra-Jensen and Arneodo et al. are only very accurate around \( D_0 \) whereas the method of Veneziano et al. is computationally expensive since it is a repeated application of the Chhabra-Jensen [103] algorithm using an analysis window on the experimental data.
Figure 4.5: (a) $D_q$ curve for a Cantor set, finely sampled from the closed form expression in equation (4.20). Figure 4.5: (b) shows the corresponding $f(\alpha)$ spectrum using the min-max filter design method [12]. Note that the maximum value of $f(\alpha)$ corresponds to $D_0 = 0.538$ and that $\alpha_{\text{min}} = D_\infty = 0.387$, $\alpha_{\text{max}} = D_\infty = 0.773$ at $f(\alpha) = 0$. 

90
Figure 4.6: Magnitude of the Fourier transform of the Cantor set $D_q$ curve in Fig. 4.5 (a). The frequency axis is normalized by $\frac{2\pi}{\Delta q}$, where $\Delta q = 1$, such that the Nyquist frequency corresponds to $F = \frac{1}{2}$. 
Although these methods provide qualitative statistical information about the attractor of the experimental data, one still needs to perform the inverse Legendre transform in order to obtain the generalized dimensions $D_q$ from $f(\alpha)$.

- The second category of algorithms evaluates the $f(\alpha)$ singularity spectrum from samples of the generalized dimensions $D_q$. They include:

  - the development of a minimal spanning tree algorithm [61]. This is an efficient algorithm for finding the nearest neighbors of any arbitrary point on the attractor. The method was proposed by Dominguez-Tenreiro et al. as an alternative to the generalized correlation integral (GCI) methods (equation (2.30)) of estimating the generalized dimensions such that the Legendre transform of the estimated $D_q$ will give $f(\alpha)$ with no spurious end-points. However, the method is only well suited for the Lorenz attractor data due to the structure of points on the attractor [61].

  - other methods [101, 11] have used the GCI to compute $D_q$ except that in order to obtain the continuous $D_q$ curve needed for the Legendre transform, either (1) the known $D_q$ values were linearly interpolated or (2) the $D_q$ values were computed at finely sampled points along the $D_q$ curve, i.e., very small $\Delta q$. Also, derivatives in the expressions for the Legendre transform were approximated by first order centered difference equations.

These methods are computationally intensive, give spurious end-points, are sometimes window length dependent, and are sometimes only well suited to certain chaotic attractors [61].

In the next section, we analyze some of the methods of existing algorithms to point out where the erroneous spurious endpoints in the $f(\alpha)$ curve could have originated.

### 4.7.1 Analysis of existing methods of evaluating $f(\alpha)$ from $D_q$

In this dissertation, we shall concentrate on the second category of algorithms discussed in the previous section, i.e., algorithms that use $D_q$ to compute $f(\alpha)$. Some
of the major drawbacks of these methods are the following: (1) the introduction of spurious/interior points in \( f(\alpha) \) caused by errors introduced in the linear interpolation of known or sampled \( D_q \) values and (2) the approximation of a derivative using a first order centered difference equation in the expressions for \( \alpha(q) \) and \( f(\alpha) \).

One source of error is the use of linear interpolation methods. Linear interpolation corresponds to sending the \( D_q \) data through a linear interpolation algorithm, or equivalently, an interpolation filter whose frequency response is a sinc-squared type filter [115]. This can be a very poor approximation to the ideal lowpass filter needed for interpolation if the \( D_q \) are not highly oversampled.

An ideal interpolation filter will pass the input signal spectra components whose frequency is between 0 and the cut-off frequency \( F_c \) while completely attenuating signal components in the range from \( F_c \) to \( \frac{1}{2} \) where \( \frac{1}{2} \) is the normalized frequency corresponding to the Nyquist frequency. \( F_c \) is aptly called the cut-off frequency. Fig. 4.7 shows a schematic diagram of the difference between the ideal lowpass needed for interpolation and the approximation that is provided by a sinc-squared type filter corresponding to linear interpolation. Notice that the sinc-squared frequency response of linear interpolation is only a good approximation to that of the ideal interpolation filter when the cut-off \( F_c \) is very small. The sinc-squared filter also has high lobes in the stopband region (\( F_s \leq F < \frac{1}{2} \)) which will not properly attenuate high amplitude signal components in that region.

A second source of error is the use of a first order difference equation to approximate differentiation in equation (4.17), \( \frac{\Delta}{\Delta q} D_q \approx D_q^{q+1} - D_{(q-1)\Delta q} \). The centered first order difference operation on the \( D_q \) is equivalent to a filtering operation that produces a poor sinusoidal approximation to the ideal \( j\omega \) frequency response of a differentiator. Fig. 4.8 also shows the distinction between the frequency response of the ideal \( j\omega \) differentiator and that of the approximation produced by the centered first order difference operation. In this case also, the sinusoidal approximation is only good at very low frequencies, or equivalently for generalized dimension sequences that are highly oversampled. At high frequencies, the sinusoidal approximation deviates significantly from the ideal differentiator.

The Legendre transform in equation (4.19) needs accurate, closely spaced samples
Figure 4.7: Shows a schematic diagram of the frequency response of an ideal lowpass filter needed for interpolation and the sinc-squared type filter which corresponds to linear interpolation. The dashed lines shows the frequency response of an ideal lowpass filter whereas the solid line represents the sinc-squared frequency response of linear interpolation. $F_c$ is the passband cut-off frequency that separates the passband from the stopband: $F_s$ is the stopband frequency at the edge of the stopband. $\delta_p$ and $\delta_s$ are the maximum deviation or errors to be tolerated in the passband and stopband respectively. Here, $F = \frac{1}{2}$ corresponds to the Nyquist frequency, $f_N = \frac{1}{2\Delta q}$ Hz.
Figure 4.8: Shows a schematic diagram of the magnitude response of an ideal differentiator $j\omega$ and that of the sinusoidal approximation produced by a centered first order difference operation. Notice that the centered first order approximation (in solid line) is only close to the ideal differentiator (in dashed line) for frequencies close to zero. The centered first order approximation becomes a very poor approximation as normalized frequency $F$ becomes close to the normalized Nyquist frequency $F_N = \frac{1}{2}$. 
of \( D_q \) and \( \frac{d}{dq} D_q \). When the differentiation and interpolation of \( D_q \) needed for \( \alpha \) and \( f(\alpha) \) are not properly implemented, i.e., not close to the ideal cases shown in Fig. 4.7 and 4.8, the Legendre transform gives rise to an \( f(\alpha) \) curve with spurious endpoints \([12]\). This may lead to ambiguous determination of \( D_\infty \) and \( D_{-\infty} \) \([61]\).

### 4.7.2 Example - Effects of Coarse Sampling

In order to show the effects of coarse sampling on the performance of some of the existing algorithms, a test case was simulated. The known \( D_q \) samples of the Cantor set, shown in Fig. 4.5 (a), were downsampled or decimated by a factor of 4, that is, they were evaluated at only 19 relatively coarsely spaced (\( \Delta q = 4 \)) values of \( q \) as shown in Fig. 4.9 (a), \( q = -36, -32, -28 \ldots, 32, 36 \). The corresponding \( f(\alpha) \) spectrum computed using the existing centered difference method of \([11, 101]\) on the coarsely sampled \( D_q \) is depicted in Fig. 4.9 (b). This is an invalid spectrum; it is not convex and it has more than one maximum point. Figures 4.10 (a) and (b) show the result of linear interpolation of the coarsely spaced \( D_q \) and the corresponding \( f(\alpha) \) spectrum that results using the centered difference approximation to differentiation.

The decimated \( D_q \) curve in Fig. 4.9 are linearly interpolated by a factor of 4 to give a first order approximation to the original samples in Fig. 4.5 (a). It is important to notice that although the \( f(\alpha) \) curve in Fig. 4.10 (b) looks more like the “correct” convex singularity spectrum curve in Fig. 4.5 (b) than did Fig. 4.9 (b), there are nonetheless, spurious, invalid end points in \( f(\alpha) \) in Fig. 4.10 (b) near \( \alpha = 0.38 \) and \( \alpha = 0.778 \). These spurious endpoints make it difficult to estimate \( D_\infty \) and \( D_{-\infty} \) from \( \alpha_{\min} \) and \( \alpha_{\max} \).

The importance of proper determination of \( D_\infty \) and \( D_{-\infty} \) from \( f(\alpha) \) cannot be overemphasized. For any chaotic attractor, the estimation of \( D_q \) is an asymptotic result that theoretically requires an infinite number of data points. For real systems (with limited number of data points), it is almost impossible to estimate the generalized dimensions as \( |q| \to \infty \). The inadequate length of the time series shows up when the estimated \( D_q \) violate the expression in equation (2.36) i.e., \( D_q < D_{q'} \) for \( q > q' \). Therefore a careful implementation of the Legendre transformation that

\[ \text{In contrast to the expression in equation}(2.36), \text{the monotonic expression here is mathematically} \]
Figure 4.9: (a) Coarsely sampled (i.e. decimated using $\Delta q = 4$) of the generalized dimensions $D_q$ of the Cantor set. (b) The corresponding $f(\alpha)$ of the decimated $D_q$ computed using the Centered Difference Equation approximation to differentiation. This is not a valid $f(\alpha)$ singularity spectrum.
Figure 4.10: (a) Linearly interpolated generalized dimensions of the Cantor set computed by linearly interpolation between the decimated $D_q$ in Fig. 4.9 (a). * indicates the original decimated samples in Fig. 4.9 (a); + indicates the linearly interpolated samples. Fig. 4.10: (b) $f(\alpha)$ of Linearly interpolated generalized dimensions $D_q$ in Fig. 4.9 (a) using the Centered Difference Equation to approximate differentiation. Notice the spurious erroneous points near $f(\alpha) \approx 0$ at $\alpha = 0.387$ and $\alpha = 0.773$. 

98
gives unambiguous estimates of $D_q$ at $q = \pm \infty$ is desirable for real systems.

In the next section, we consider a signal processing method for evaluating $f(\alpha)$ from $D_q$. We applied this method to the generalized dimensions of the Cantor set introduced in section 4.6 to show that it works well, even when some of the existing algorithms fail.

### 4.8 Improved Accuracy in the Singularity Spectrum

In this section, we propose an algorithm that more accurately estimates the $f(\alpha)$ singularity spectrum using only coarsely sampled $D_q$. We interpolate between the coarsely sampled $D_q$ points and evaluate the necessary derivatives in equation (4.17-4.19) by using the Parks-McClellan algorithm [109] to design the best min-max approximation to an ideal interpolator and differentiator, respectively. This is in sharp contrast to prior algorithms which either linearly interpolate coarsely sampled $D_q$, using a sinc$^2$ interpolation filter or compute, at great expense, finely sampled $D_q$ using equation (2.29) [11]. These existing methods can be computationally intensive and may give spurious results.

In the next section, we discuss the min-max approximation to the ideal interpolator and differentiator. We will show examples of the frequency response of lowpass interpolation filters and differentiators designed using this method. We compare the frequency response of the interpolation filter and differentiation approximation with the ideal cases. Finally, we apply the min-max interpolator and differentiator to the coarsely sampled values of the $D_q$ of the Cantor set ideal “test case” to show that our proposed method works well when compared with existing methods.

### 4.9 Min-Max Digital Filter Design

The ideal frequency response $H_I(\omega)$ of a lowpass filter is shown by the dashed line in Fig. 4.7. It is impossible to realize an ideal lowpass filter in practice. This is because restricted to allow for a one-to-one transformation from $D_q$ to $f(\alpha)$.
an ideal lowpass filter is non-causal and has an infinite duration impulse response. For an indepth study of causality and stability in digital filter design, see Oppenheim et al. [116], Proakis et al. [109] and Jackson [117].

In digital filter design, a transition region is allowed so that there is no sharp discontinuity between the passband and stopband region as depicted in the schematic diagram for an ideal lowpass filter in Fig. 4.7. Two frequency bands are indicated the plot, i.e., passband and the stopband bounded by the passband edge normalized frequency, \( F_c \), and the stopband edge frequency \( F_s \). The passband is now between 0 and \( F_p \) (0 < \( F \) ≤ \( F_p \)) and the stopband is between \( F_s \) and \( \frac{1}{2} (F_s \leq F < \frac{1}{2}) \). The frequency band between \( F_p \) and \( F_s \) is known as the transition region. Notice that for the lowpass filter, the full frequency band being processed in Hertz is between 0 and the Nyquist frequency, \( \frac{1}{2} f \). In normalized digital frequency \( F \), this also corresponds to 0 and 0.5, respectively.

The min-max Parks/McClellan filter design method [116] provides an optimum approximation to the ideal frequency response where it is possible to specify error tolerances (ripples) \( \delta_p \) and \( \delta_s \) in both the passband and stopband, respectively. If we define a weight function \( W(\omega) \) to be \( \beta_p \) in the passband and \( \beta_s \) in the stopband region such that we can choose the relative size of errors in each band \( r \) then a weighted approximation error function in both bands can be written as

\[
E(\omega) = W(\omega) [H_I(\omega) - H_d(\omega)],
\]

where \( H_d(\omega) \) is the desired frequency response of the filter we design and \( H_I(\omega) \) is the frequency response of an ideal filter. The relative error deviations \( \delta_p \) and \( \delta_s \) from the ideal response would be allowed in both the passband and stopband, respectively. With this error function definition, we seek a frequency response \( H_d(\omega) \) that minimizes the maximum weighted approximation error of equation (4.21), i.e.,

\[
E(\omega) = \min_{a[k]} \left( \max_{\omega} \left[ \left| W(\omega) \left( H_I(\omega) - \sum_{k=0}^{m} a[k] \cos \omega k \right) \right| \right] \right)
\]

(4.22)

where

\[
H_d(\omega) = \sum_{k=0}^{m} a[k] \cos \omega k.
\]

(4.23)
Due to the minimization of the maximum weighted error $E(\omega)$ in the passband and stopband, the method is termed min-max. From equation (4.23), $H_d(\omega)$ can be reformulated into a polynomial of degree $m$. It implies that $H_d(\omega)$ can have at least $m + 1$ and at most $m + 3$ local maxima and minima in the interval $0 < \omega < \frac{\pi}{\Delta_\omega}$. Hence, the filter order can be at least $m + 1$. The minimization problem in equation (4.22) is a Chebyshev approximation problem to determine the $m + 1$ coefficients $a[k], k = 0, 1, \ldots, m$ that minimize the error. Here, $m$ is the filter order. A unique solution of the minimization is guaranteed [109, 118].

The coefficients of the filter are readily obtained from the well known Parks/McClellan Remez Exchange filter design algorithm [119]. After the coefficients $a[k]$ are found in equation (4.23), the impulse response $h_d[n]$, $k = 0, 1, \ldots, m$ of the linear-phase finite impulse response (FIR) filter can be determined from simple expressions [109, 118] leading to four types of impulse responses, i.e., even or odd symmetric and even or odd anti-symmetric. The program is available for designing linear phase finite impulse response (FIR) filters based on the Chebyshev approximation criterion. The Remez exchange algorithm finds the filter which minimizes the maximum weighted error in the passband and stopband. The program may be use to design lowpass, highpass or bandpass filters, and differentiators [118]. The maximum errors in the passband and stopband are user-specified as $\delta_p$ and $\delta_s$, respectively. The Parks/McClellan algorithm gives the optimum (min-max) $m^{th}$ order filter. If the initial frequency response of the filter does not satisfy the above specifications, then the filter order $m$ is increased until the desired frequency response is attained. Filters with larger filter order $m$ give better filter responses, but require more computation to implement. Therefore, a trade-off between computation time and accuracy of frequency response must be made in the selecting the filter order $m$.

In the next section, we describe the design of a typical interpolation filter used in this dissertation. We also describe how the various frequency bands are determined. 

---

3the impulse response $h_d[n]$, is the inverse Fourier transform of the frequency response, $H_d(\omega)$
4.10 Interpolation Filter

Given any multifractal time series and coarsely sampled values of $D_q$ estimated from equation (2.29) in the finite range $q \in [q_{min}, q_{max}]$, we interpolate between these points by a factor $L$. This is accomplished by “stretching the signal”, i.e. inserting $L - 1$ zeros between each pair of given $D_q$ samples [117]. The effect of “stretching” the given $D_q$ samples is to create $L - 1$ mirror images of the Fourier transform of the original $D_q$ samples. To remove the mirror images, we pass the stretched signal spectrum through a lowpass filter whose passband edge is a function of $L$ and the bandwidth of the original Fourier transform of the $D_q$.

Assuming that the original spectrum of Fig. 4.11 (a) is bandlimited between 0 and $F_0$ in normalized frequency, the effect of “stretching” for $L = 4$ is the creation of mirror images of the original spectrum at $\frac{F_N}{2}$, $F_N$ and $\frac{3F_N}{2}$, where $F_N = \frac{1}{2}$ is the normalized Nyquist frequency, corresponding to $f_N = \frac{1}{2\Delta q}$ Hz. To remove the mirror images, we pass the stretched $D_q$ spectrum through a lowpass filter whose passband edge is $F_p = \frac{F_N}{4}$, whose transition bands between $F_p$ and $\frac{2F_N + F_N}{4}$ and whose stopband edge is between $\frac{2F_N - F_N}{4}$, and $F_N$. Thus, zero insertion between $D_q$ samples followed by lowpass filtering reduces to computing the convolution of the lowpass filter impulse response coefficients $h[n]$ with the $D_q$ curve with zeros inserted [117].

In particular, $\Delta q = 4$ and $L = 4$ for the coarsely sampled generalized dimensions $D_q$ curve of the Cantor set in Fig. 4.9 (a). To restore the $D_q$ curve back to the original finely sampled ($\Delta q = 4$) $D_q$ curve of Fig. 4.5 (a), we need to interpolate 3 samples in-between each existing sample in Fig. 4.9 (a). The original and stretched spectra of the Cantor set are shown in 4.13.

Fig. 4.12 (a) gives the frequency response of a 20th order min-max interpolator and 6th order min-max differentiator. The lowpass interpolation filter has the specifications $F_p = 0.003$, $F_s = 0.246$, $\delta_p = 0.1$ and $\delta_s = 1 \times 10^{-6}$. The interpolator was designed to approximate an ideal lowpass filter frequency response with passband $F_p = 0.003$ and stopband $F_s = 0.246$. As long as the sampled $D_q$ curve is essentially bandlimited to the passband of the interpolation filter, the filter design prescription in Fig. 4.11 (b) will provide accurate interpolation of the $D_q$ curve. Figures 4.12 (a) and (b) show the magnitude of the frequency responses of the interpolation filter and
Figure 4.11: (a) Shows the original signal spectrum, (b) “Stretched” signal spectrum resulting from inserting 3 zeros in-between the $D_q$ samples. The ideal interpolation filter frequency response, indicated by the dashed line, removes the copies of the original spectrum created by stretching, (c) the spectrum of the interpolated signal.
the differentiator used in the evaluating the \( f(\alpha) \) singularity spectrum for the Cantor set. The differentiator was designed to approximate an ideal \( j\omega \) frequency response between the normalized frequencies 0 and 0.1. The passband for the differentiator is \( F_p = 0.1 \) and for stopband is \( F_s = 0.25 \).

The Park-McClellan Remez Exchange filter design program is used in this dissertation to design the optimum (min-max) lowpass interpolation filters and differentiators used on the coarsely sampled unvoiced speech \( D_q \) curves. Depending on the bandwidth where most of the energy is concentrated in the spectrum of the unvoiced speech \( D_q \), lowpass interpolation filters and differentiators with different passbands and stopbands as well as filter order \( m \) were designed.

Figures 4.14 (a) and (b) show the results of the min-max method with a 20th order min-max interpolator and 6th min-max differentiator [12] applied to interpolate the coarsely sampled \( D_q \) of the Cantor set described in section 4.6.1. The sparse \( D_q \) in Fig. 4.9 (a) with \( \Delta q = 4 \) are well interpolated in Fig. 4.14 (a) and the spurious/noisy end-points in the singularity spectrum \( f(\alpha) \) in Fig. 4.14 (b) have been greatly reduced when compared with Fig. 4.10 (b). Note that no spurious endpoints appear near \( \alpha = 0.78 \) in Fig. 4.14 (b) as compared to those of Fig. 4.10 (b). The result of the min-max filter design method in Fig. 4.14 (b) and the ideal case depicted in Fig. 4.5 (b) are very close except at the ends of the curve, where \( f(\alpha) < 0 \). In general, any deviation from the ideal interpolation filter would result in some form of distortion when a sequence is passed through it. Since the min-max method is an approximation to the ideal interpolation filter, the erroneous endpoints in Fig. 4.14 (b) may be due to the approximations in the filter design. Nonetheless, these errors are even worse in the conventional approach in Fig. 4.10.

The importance of the new min-max approach is increased accuracy and its computational reduction due to the fact that the \( D_q \) curve can be sparingly sampled at various integer values of \( q \) and later optimally interpolated to obtain the smooth curve that is needed for the singularity spectrum \( f(\alpha) \). The method works well for attractors that exhibit multifractal behavior and satisfy the expression in equation (2.36).
Figure 4.12: (a) Shows the magnitude response of a 20th order min-max interpolation filter used in evaluating the $f(\alpha)$ singularity spectrum for the Cantor set. Filter specifications are $F_p = 0.003$, $F_s = 0.246$, $\delta_p = 0.1$, $\delta_s = 1 \times 10^{-6}$, $\beta_p = 10$ and $\beta_s = 1$ (b) shows the magnitude response of a 6th order min-max differentiator designed using the Park-McClellan program for the Cantor set example. Filter specifications are $F_p = 0.1$, $F_s = 0.25$, $\delta_p = 0.1$, $\delta_s = 1 \times 10^{-6}$, $\beta_p = 10$ and $\beta_s = 1$. 
Figure 4.13: (a) Shows the absolute value in dB of the Fourier transform of the decimated $D_q$ Cantor set example. In this case, the decimation factor is 4. The frequency axis is normalized by $\frac{2\pi}{\Delta q}$, for $\Delta q' = 4$. (b) shows the absolute value in dB of the Fourier transform of the “stretched” $D_q$ of Cantor set example plotted in dB. The frequency axis is normalized by $\frac{2\pi}{\Delta q}$, for $\Delta q = 1$. $\Delta q' = 4 \Delta q$. 
Figure 4.14: (a) Min-Max interpolated generalized dimensions of the Cantor set using the coarsely sampled $D_q$ in Fig. 4.9 (a). * = decimated $D_q$ samples in Fig. 4.9 (a) with $\Delta q = 4$; + = min-max interpolated samples with $L = 4$ and $\Delta q' = 1$. (b) $f(\alpha)$ of the optimally interpolated generalized dimensions $D_q$ in Fig. 4.14 (a) using min-max interpolator and differentiator.
In the next chapter on multifractal speech analysis, the min-max method of evaluating the $f(\alpha)$ singularity spectrum from the generalized dimension $D_q$ is used to determine the $f(\alpha)$ singularity of unvoiced speech signals. For each speech signal, the corresponding generalized dimensions $D_q$ are evaluated. We examine the spectrum of the $D_q$ to determine the passband of the lowpass filter needed for interpolation as well as the passband of the differentiator. The lowpass filter coefficients (obtained from the Parks-McClellan algorithm) are appropriately convolved with the sparsely sampled $D_q$ curve “stretched” by inserting zeros samples between the original $D_q$ samples. The derivatives in the Legendre transform of equations (4.18) and (4.19) are then implemented by convolving the impulse response coefficients of the min-max differentiator with the interpolated $D_q$ sequence. This is a “q”-domain implementation of the min-max method of evaluating the $f(\alpha)$ singularity spectrum. In appendix B, we provide the frequency domain derivation of the Legendre transform which may be useful for frequency-domain implementation.
Chapter 5

Multifractal analysis of unvoiced speech signals

In this chapter, we apply some of the previously defined measures of chaos and their estimation techniques to unvoiced speech signals. We address issues that concern the validity of estimating chaos parameters from experimental data. Some of these issues include: (1) the presence of measurement noise in the experimental data and how it affects multifractal analysis and (2) the constraint of finite record length $N$ of an experimental data sequence $x[n]$ and how this affects estimates of the generalized dimension $D_q$. We discuss the inadequacy of the correlation dimension $D_2$ when trying to quantify the complexity in unvoiced speech signals. We then use the embedding theorem [56] as well as the generalized correlation integral method of equation (2.30) to estimate the generalized dimensions $D_q$ from the phase-space reconstructed vectors of single scalar realization of unvoiced speech signals. Finally, we use the Parks/McClellan min-max filter design method to approximate the interpolation and differentiation filters needed to evaluate the $f(\alpha)$ singularity spectrum from the generalized dimensions $D_q$ of unvoiced speech signals using the Legendre transform of equation (4.19).

However, before discussing these measures of chaos, we start the next section with a description of some of the properties of the analyzed speech signals such as the speech recording method and the determination of the speech signal to noise ratio (SNR).
5.1 Speech Data

In this dissertation, the time series used in the multifractal analysis of speech consists of acoustic pressure waveforms of speech from the ISOLET (standard) database [25]. The ISOLET database consists of letters of the English alphabet spoken in isolation by American males and females. The database has 7800 spoken letters with two productions of each letter given by 150 speakers. Each speaker is identified by a character string specifying his/her gender, and initials followed by a number for uniqueness, i.e., “mbjt0” is the label of a sound pronounced by a male with initials “bjt”. The number “0” is uniquely assigned to distinguish between speakers with the same initials.

The 150 speakers are organized into five subgroups ISOLET-1, ..., ISOLET-5. Each subgroup consists of speech utterances by 30 speakers. Each data file in the subgroups has a header which contains information about the file, e.g., sampling rate, number of samples, etc., followed by a series of 16 bit integers corresponding to uniformly spaced samples of the speech sound. The sampling rate for speech speech data in the ISOLET database is 16 kHz; equivalently, the sampling period \( T = 6.25 \times 10^{-5} \) sec/sample.

Fig. 5.1 shows a 0.606 second segment of the unvoiced fricative “s”. The unvoiced fricative “s” in Fig. 5.1 consists of two distinct components. One quasi-periodic part (from approximately sample number 1500 to 3500) is due to the vowel “e” at the onset of “es” in the pronunciation of the unvoiced fricative “s”, whereas the other noise-like part (from approximately sample number 3500 to 7960) is due to turbulent air flow in the neighborhood of the constriction in the vocal tract [77]. Also notice the silence region of the first 1300 samples and the last 1700 samples.

In the following speech analysis, we work with only the segment of speech signal that corresponds to the noise-like or turbulent regions of the unvoiced fricatives. We isolated these noise-like sections by looking at the speech waveforms and removing regions of silence in voiced fricatives as well as the regions of silence and quasi-periodic portions in the unvoiced fricatives. Therefore, for each speech signal analyzed, we give the following: (1) The sample number indicating the onset of the segment of speech data to be retained for chaos analysis; all earlier samples have been truncated.
to remove the initial silence and or the quasi-periodic region. (2) The total number of speech samples that are retained for analysis. Fig. 5.2 shows a 0.0625 second segment extracted from the unvoiced fricative “s” in Fig. 5.1; the noise-like characteristics are evident. This extracted segment starts at sample number 6001 in Fig. 5.1 and is of duration 1000 samples.

From ISOLET-1, we selected a total of 40 speech signals to analyze. The set consists of 10 male speakers and 10 female speakers, each pronouncing both the “s” and the “z” sounds. It is important to perform the multifractal analysis both on male speakers and female speakers because of the differences in their speech bandwidth and average pitch period.

In the next section, we describe the data acquisition method for speech signals in the ISOLET database. We also describe a procedure for estimating the signal-to-noise ratio SNR.

5.2 Noise in speech signals

Due to errors from microphone movement and human agitation which are likely to occur in the acquisition of experimental speech data, the speech signals in the ISOLET database are assumed to be noisy. However, since the speech signals in the database were recorded under a controlled environment in a speech recognition laboratory, the noise level is very low. The data were recorded with a Sennheiser HMD 224 noise-cancelling microphone. The data acquisition was performed on an AT&T DSP 32 board interfaced to a Sun workstation [25].

The signal-to-noise ratio, SNR, of the speech data is defined as $10 \log_{10} \left( \frac{\sigma_s^2}{\sigma_{nv}^2} \right)$ where $\sigma_s^2$ is the signal power and $\sigma_{nv}^2$ is the noise power. It was estimated after subtracting the mean signal value from each sample of the digitized speech signal so that the resulting speech signals have zero mean. The mean SNR for the ISOLET database is 31.5dB with a standard deviation of 5.6dB. This indicates a relatively noise free recording. The SNR for the ISOLET database was calculated as follows: Consider the “t” utterance from each speaker. The silence before the “t” burst is assumed to be relatively free of breath noise. Therefore, it reflects the typical
Figure 5.1: Shows an unvoiced fricative \"s\" from the ISOLET database. Notice the distinct regions of silence, quasi periodicity, noise-like region and silence. \( T = 6.25 \times 10^{-5} \text{ sec.} \)
Figure 5.2: A length 1000 sample segment of the unvoiced fricative \"s\" starting at sample number 6001 in Fig 5.1. Notice the noise-like appearance of the speech signal. \( T = 6.25 \times 10^{-5} \) sec.)
background noise in the database very well. The noise variance $\sigma_{nu}^2$ is estimated from the first $Y_n$ samples of the "t" utterance from each speaker as

$$\sigma_{nu}^2 = \frac{\sum_{i=1}^{Y_n} t_i^2}{Y_n}.$$  \hspace{1cm} (5.1)

$Y_n$ is chosen to make sure that only the samples in the silent region are used in equation (5.1). The notation $t_i$ refers to the $i^{th}$ sample of the sound "t". The mean signal amplitude squared,

$$\sigma_s^2 = \frac{\sum_{i=\Delta Y_m+1}^{\Delta Y_m+Y_m} t_i^2}{Y_m},$$  \hspace{1cm} (5.2)

is calculated from the middle $Y_m$ samples, where $Y_m$ is determined by removing the beginning $\Delta Y_m$ samples and the ending samples of each speech segment such that only the middle portion of each "t" utterance from every speaker is retained. This minimizes the transitions from starting and finishing each pronunciation. The values of $\sigma_s^2$ and $\sigma_{nu}^2$, and hence the SNR are estimated for all the speakers in the database. The SNR value for the database is the average of the SNR's for all speakers. There is nothing special about the "t" utterance. Other speech utterances that would reflect the background noise such as "p" and "b" would have been appropriate.

In the next section, we highlight some of the questions and concerns that should be addressed in successful chaos analysis of experimental data. Satisfactory answers to these questions may be a measure of the reliability of the results provided by the multifractal analysis.

### 5.3 Concerns in chaos analysis of experimental data

In the chaos analysis of experimental data, there are some questions to be answered in order for one to be satisfied that the results of such an analysis are not erroneous and misleading. These questions may include:
• Is the data sequence length, \( N \), long enough? Recall the mathematical definitions of dimension, entropy and Lyapunov exponent require theoretically infinite data length? How long a data sequence \( N \) is enough?

• What is the noise level in the experimental data and how does this affect the estimated chaos parameters?

• How many dimensions should we use in the embedding space?

• Is the number of time series samples, \( N \), used in estimating the correlation dimension \( D_2 \) also valid for estimating the generalized dimensions \( D_q \) as \(| q | \to \infty\)?

Unfortunately, it is difficult to answer some of these questions definitively, partly because the theoretical foundations of the Taken’s embedding technique in section 2.8 in relation to the estimation of generalized dimensions \( D_q \) in section 2.7 are not presently fully developed and partly because the answers to some of these questions depend to some extent on the experimental data being analyzed. Therefore, the relevant established criteria we will use as guidelines in the multifractal analysis of unvoiced speech signals are the choice of delay time \([57]\), the optimum embedding dimension \([120]\), the noise levels in chaotic signals \([60]\) and the length of the data sequence required for chaotic analysis \([49, 50]\).

In the next section, we examine a few of the criteria listed above to show that the results obtained from the multifractal analysis of unvoiced speech signals are reliable since the sample size and signal-to-noise SNR of speech signals from the ISOLET database are within the tolerable range of values suggested to be used in literature \([49]\).

5.4 Limitations of chaos analysis due to the properties of speech data

In order to answer some of the questions posed in the last section, we examine in this section the relationship between the length of the data record, \( N \), and the validity of the estimation of the correlation dimension. We also discuss the effects of noise
levels on the generalized dimension $D_q$ using the method of generalized correlation integrals (GCI) in section 2.7.

5.4.1 Finite record length effect on $D_q$

In the chaos analysis of experimental data, if the length $N$ of the time series is too small compared to the dimension of the time series, then the estimated value of the correlation dimension $D_2$ may be wrong [49]. Several authors have suggested different upper bounds on the correlation dimension $D_2$ that an attractor may have, given a limited amount of data [50, 49, 121, 122]. In the multifractal analysis of unvoiced speech signals, we use the method of Kantz et al. [49] as a guide in the determination of the appropriate number of time samples needed for chaos analysis.

Consider the schematic diagram of local slopes versus log $\epsilon$ in Fig. 2.13. At a point between regions II and III, the noise region ends and the scaling or plateau region begins. Let the value of $\epsilon$ at this point of transition be called $\epsilon_{\text{min}}$. Assume that one needs about 100 pair of vector points [49] on the attractor; their inter-point distance is less than or equal to $\epsilon_{\text{min}}$. If we also assume that $\epsilon_{\text{min}}$ is about 10% of the maximum possible inter-point distance on the attractor, then the correlation integral at $\epsilon_{\text{min}}$ may be approximated as [49]

\[
C_2(\epsilon_{\text{min}}) \approx \frac{100}{N^2}. \tag{5.3}
\]

From equation (2.26), $C_2(\epsilon) \sim \epsilon^{D_2}$, from which one can conclude that we need at least $N_{\text{min}}$ signal samples to be able to accurately estimate $D_2$ [49]. Combining equations (2.26) and (5.3) yields

\[
N_{\text{min}} \approx 10^{0.5(2+D_2)}. \tag{5.4}
\]

Using the Kantz et al. criterion of equation (5.4) and the estimated range of the correlation dimension $D_2$ for speech to be between 3 and 7 as given by Narayanan et al and Tishby [27, 14], the minimum number of samples required for $D_2$ estimation in speech signals is between 400 and 30,000 samples. This implies that the number of data samples needed to estimate dimensions grows exponentially with the $D_2$ [122]. The number of data samples requirement may be trivial to some applications in
digital signal processing (DSP), but in the next paragraph we will discuss the futile effort of trying to increase the number of data samples $N$ in DSP applications with short data lengths.

**Problems with increasing the sampling rate to increase the record length**

From a signal processing perspective, longer length data sets can be generated by increasing the sampling rate of the original time series, i.e., reducing the sampling time $T$ or interpolating between the given samples of a time series. In nonlinear dynamics, however, increasing the number of samples in this way has generated no new information since no new phase space trajectories have been created. Therefore, the estimates of the spatial measures of chaos such as Lyapunov exponents, entropy and generalized dimensions $D_q$, will remain essentially unchanged.

Alternatively, sustained fricative utterances may be used to generate considerably longer data sets [123]. This is the situation where the speaker tries to hold his/her breath as long as possible to sustain the pronounced fricative. It should be pointed out that sustained fricatives are produced by a relatively static vocal tract shape and do not involve any apparent laminar to turbulent flow transitions, such as those found in isolated or vocalic fricatives [14]. A better alternative which may be the object of future research, is to append as many isolated or vocalic fricatives of the same type from the same speaker together to generate a relatively longer time series.

In this dissertation, however, the multifractal analysis was performed on speech data with record length $N$ ranging from 3500 to 11,000 points. Although these values of $N$ fall within the range suggested by the Kantz criterion, we only use the estimates as a guide since it was designed for $D_2$ (i.e. $D_q$ for $q = 2$ only) and, to the best of our knowledge, there is no criterion relating the sample size $N$ to the generalized dimensions $D_q$ as $|q| \rightarrow \infty$. We, therefore, estimate the generalized dimensions $D_q$ of unvoiced speech signals with $|q| \leq 5$ under the assumption that the upper bound on the minimum number of samples required for $D_2$ will also give accurate $D_q$ estimates for $|q| \leq 5$. The range of $q$ values in the estimated generalized dimension

\[ q \leq 5 \]

this value is selected based on our experience with unvoiced speech samples from the ISOLET database where when $q < 5$, it is almost impossible to estimate the generalized dimension $D_q$. 

117
$D_q$ of unvoiced speech will be examined to verify the validity of this assumption.

### 5.4.2 Noise level effect on $D_q$

Unlike the analysis of adequate data sequence length $N$, noise analysis has been performed on chaotic attractors where the generalized dimensions $D_q$ were estimated as the noise level is increased [60, 124, 49]. Several authors have analyzed different chaos attractors and suggest various noise levels. In particular, Meisel and Johnson [60] reported generalized dimensions $D_q$ which are within 1% of their analytical values when noise of variance 0.5% of the numerical range was added to the time series produced from the asymmetric snow-flake fractal [104].

From section 5.2, the average signal to noise ratio $SNR$ in the speech data from the ISOLET database is about 31.5dB. This is a noise level with variance of about 0.005% of the numerical range, i.e., high signal-to-noise ratio $SNR$. The high signal-to-noise $SNR$ and reasonable record length $N$ are good indicators that the estimated generalized dimensions for speech signals are fairly accurate.

In the next section, we describe the procedure for estimating the generalized dimensions $D_q$ for unvoiced speech signals. We apply the procedure to all 40 speech data signals manually selected from the ISOLET database. The standard procedure is to increase the embedding dimension $d_E$ incrementally until the calculated dimensions appear to have converged [49, 110].

### 5.5 Procedure for estimating the generalized dimensions from speech data

In this dissertation, the generalized dimensions [3],

$$D_q = \lim_{\epsilon \to 0} \frac{1}{\log \epsilon} \log C_q(\epsilon), \quad -\infty < q < \infty$$

defined in section 2.7 were evaluated for different speech sounds in the ISOLET database using the generalized correlation integral $C_q(\epsilon)$ as defined in equation (2.47) [60]. Note that, theoretically, in (5.5) if in the limit as $\epsilon \to 0$, we plot $\log C_q(\epsilon)$ as
a function of $\log(\epsilon)$, then this log-log plot is linear with slope $D_q$. However, for real data such as speech signals (with a limited number of data points), only a portion of this curve is linear because the expression in equation (5.5) is an asymptotic result which requires an infinite data record length. This portion of the log-log plot where the curve is approximately linear is referred to as the “linear region” as described in section 2.9.1. The linear region can also be determined from the plot of the local slopes $\frac{\partial}{\partial \log \epsilon} [\log C_{q,d\epsilon}(\epsilon)]$ versus $\log \epsilon$. From the plot, the linear region corresponds to a plateau region as depicted in Fig. 2.13. For various values of $q$, we determine a spectra of dimensions $D_q$ in equation (5.5) from the plateau region of the plot of local slopes $\frac{\partial}{\partial \log \epsilon} [\log C_{q,d\epsilon}(\epsilon)]$ versus $\log \epsilon$.

A numerical approximation to the generalized correlation integral is given in equation (2.47) [60]. Also, a numerical implementation algorithm for the generalized correlation integrals, GCI, which avoids taking the logarithm of zero is derived in Appendix A.

A procedure for extracting the generalized dimensions $D_q$ is as follows:

For each speech signal data record:

- Multiply each sample of the speech time series by a scaling factor such that the inter-point spacing, $\epsilon_{ij} = ||X_i - X_j||$ of the embedded vectors, $X_i$, in equation 2.38 using the highest embedding dimension $d_{Emax}$, has a minimum

$$\epsilon_{min} \approx 1 = \min_{ij} \epsilon_{ij}$$

and a maximum inter-point spacing

$$\epsilon_{max} \approx 10000 = \max_{ij} \epsilon_{ij}$$

as proposed in [125].

- Estimate the optimum delay time $\tau_{opt}$ using the mutual information method [57] described in section 2.8.1. An example of the optimum delay time estimation using mutual information on unvoiced speech sound \"s\"\ from the ISOLET database is shown in Fig. 5.3.

- For each value of $q$: 119
For each embedding dimension from 5 to the highest embedding dimension $d_{E_{max}}$:

- Evaluate the generalized correlation integral (GCI) in equation (2.31).
- Evaluate and plot the local slopes $\frac{\partial}{\partial \log \varepsilon} \left[ \log C_{q,d_E}(\varepsilon) \right]$ versus $\log \varepsilon$. Determine the corresponding dimension $D_q$ by taking a slice of the local slopes plot along $\log \varepsilon$ in the plateau region shown in Fig. 2.13. This method correctly identifies the linear region of the log-log plot of $\log C_q(\varepsilon)$ as a function of $\log(\varepsilon)$ in equation (5.5).
- Average the value of the local slopes in the plateau region to obtain the estimate of the dimension $D_q$ in equation (5.5).

In the next section, we show examples of the correlation integrals used in the estimation of the correlation dimension $D_2$ for unvoiced speech signals. We also examine the various correlation integral and local slope structures that may be encountered in the multifractal analysis of unvoiced speech signals. This is to draw attention to the inadequacy of estimating only the correlation dimension $D_2$ when the system under analysis is as complex as speech production. Examples of the plot showing such inadequacies are provided in the next section.

5.6 Inadequacy of the $D_2$ measure in characterizing the attractors of speech signals

In the numerical implementation of the generalized correlation integral, GCI, in equation (2.47), there are a number of factors that affect the length and quality of the plateau region (described in section 2.9.1) obtained from the generalized correlation integral by the method described in section 2.13. These factors become more important in the chaos analysis of experimental data. They include: the number of data points $N$, the number of reference points $M_{seed}$, the incremental increase in the width of the volume elements $\varepsilon$, the increase in the embedding dimensions $d_E$, etc. The effect of all of these factors can be understood from a careful study of equation (2.47).
Figure 5.3: Shows the normalized mutual information as a function of time lag $\tau$ for the unvoiced speech sound \"s\" from the ISOLET database. The plot is normalized by the mutual information at lag zero. The mutual information at lag zero is 2.406866 bits.
In this section, we show four distinct correlation integral and local slope structures we encountered in estimating the correlation dimension $D_2$ estimation for unvoiced speech signals. It should be pointed out that the estimated correlation dimensions $D_2$ from these correlation structures may not be precise due to the relatively small plateau regions. This is indicative of the limited length of the data record available from unvoiced speech. However, the different correlation integral and the corresponding local slope structures are indications of non-constant $D_q$ and hence of chaotic complexity involved in the speech production process. These structures are common to both male speakers and female speakers as well as to the different types of unvoiced speech analyzed. Unless stated otherwise, all the speech data are evaluated for integer embedding dimensions $d_E$ ranging from 5 to 15. The correlation integral and local slope structures fall into the following categories:

- **Normal**: By "normal", we mean that the four distinct regions described in section 2.13 are clearly shown and that the plateau region can be identified without any ambiguities. Examples of typical "normal" correlation integrals and the corresponding dimensions estimates from the local slopes plot are shown from Fig. 5.4 to Fig. 5.7.

- **Multiscale**: By "multiscale", we mean that the correlation integral, and hence the local slope, plotted as a function of $\log \epsilon$ and increasing embedding dimension $d_E$ has more than one plateau region. This may lead to ambiguities in the determination of the correlation dimension $D_2$ and raises interesting questions as to which of the plateau regions is valid? Examples of multiscale structure found in the analysis of chaos in unvoiced speech signals are depicted in Fig. 5.8 to Fig. 5.10.

- **Lacunarity**: For many systems, the plateau region is flat and the correlation dimension $D_2$ can be easily determined from the plot of the local slopes versus $\log \epsilon$. However, for some systems, a well defined plateau region is replaced by a region of ragged and almost periodic-like structure referred to as lacunarity. This occurs as a result of the fractal nature of the time series. Lacunarity has been studied extensively in the triadic Cantor sets [126, 5]. Fig. 5.11 to Fig. 5.13 are examples of local slope plots with lacunarities found in the chaos

122
analysis of unvoiced speech. The raggedness of the plateau region in these cases makes it difficult to determine the correlation dimension $D_2$ from the plot of the local slopes versus log $\epsilon$. However, a least square linear fit to the corresponding region in the log-log plot of the correlation integral versus log $\epsilon$ will give a reasonable estimate of the correlation dimension $D_2$ [5, 127].

- Mixed: By "mixed", we mean that more than one type of local slope structure can be identified for certain speech signals. In most cases for unvoiced speech signals, we have a “mixed” combination of the “normal” and of the “lacunarity” structures. Just as in the case of “multiscale” structure, the mixed structure also presents some interesting, but difficult problems. Fig. 5.14 and Fig. 5.15 show correlation integral structure that is referred to as mixed structure.

In Appendix D, more examples of plots of the correlation integral and local slopes showing the various structures are given for the rest of the 40 unvoiced speech samples analyzed.

Since the variability in the correlation integral structures may not be fully explained by estimating only the correlation dimensions $D_2$, we propose multifractal analysis of unvoiced speech signals. In the next section, we describe another property of experimental data which may adversely affect the estimation of correlation dimension $D_2$ from experimental data.

### 5.7 Effect of high embedding dimensions on real data

In section 5.4.1, we examined the relationship between the estimated correlation dimension $D_2$ and the number of available data samples $N$. In practice, one computes the generalized dimensions as a function of the embedding dimension until convergence it reached. However, care must be taken to ensure that the local slopes in the plateau region are not corrupted by artifacts generated when the embedding dimension $d_E$ is too high for a fixed number of data points.

Consider the numerical implementation of the generalized correlation integral for
Figure 5.4: "Normal" example. (a) Shows the log of the correlation integral plotted as a function of log $\epsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slopes curves which are estimates of the correlation dimension $D_2$ as a function of $d_E$. These estimates were made using the unvoiced speech data \"z\" for a female speaker, taken from the ISOLET (/ISOLET/isolet1/frc0) database. The number of speech samples $N$ is 7184 starting at sample number 1316. Using the plateau region between log $\epsilon \in [6.9, 7.8]$, the estimated correlation dimension $D_2$ is 1.1825.
Figure 5.5: "Normal" example. (a) Shows the log of the correlation integral plotted as a function of $\log \varepsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension $D_2$ as a function of $d_E$. These estimates were computed using the unvoiced speech data \"z\" for a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjc1) database. The number of speech samples $N$ is 5040 starting at sample number 2182. Using the plateau region between $\log \varepsilon \in [7.1, 8.0]$, the estimated correlation dimension $D_2$ is 1.4484.
Figure 5.6: "Normal" example. (a) Shows the log of the correlation integral plotted as a function of $\log \epsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension $D_2$ as a function of $d_E$. These estimates were computed using the unvoiced speech data "s" for a female speaker, taken from the ISOLET (/ISOLET/isoleit1/fme0) database. The number of speech samples $N$ is 4324 starting at sample number 4849. Using the plateau region between $\log \epsilon \in [6.5, 7.5]$, the estimated correlation dimension $D_2$ is 3.1363.
Figure 5.7: “Normal” example. (a) Shows the log of the correlation integral plotted as a function of log $\epsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension $D_2$ as a function of $d_E$. These estimated were made using the unvoiced speech data “s” for a male speaker, taken from the ISOLET (/ISOLET/isolet1/mnjh0) database. The number of speech samples $N$ is 7275 starting at sample number 4429. Using the plateau region between log $\epsilon \in [6.5, 7.2]$, the estimated correlation dimension $D_2$ is 4.4251.
Figure 5.8: "Multiscale" example. (a) Shows the log of the correlation integral plotted as a function of $\log \epsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension $D_2$ as a function of $d_E$. These estimates were made using the unvoiced speech data "z" for a female speaker, taken from the ISOLET (/ISOLET/isolet1/fmbd0) database. The number of speech samples $N$ is 8841 starting at sample number 1441. Using the plateau region between $\log \epsilon \in [7.0, 7.8]$, the estimated correlation dimension $D_2$ is 1.0867. Notice the other potential plateau region at $\log \epsilon \in [4.8, 6.0]$ typical of "multiscale" examples.
Figure 5.9: "Multiscale" example. (a) Shows the log of the correlation integral plotted as a function of log $\epsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension $D_2$ as a function of $d_E$. These estimates were made using the unvoiced speech data "s" for a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjrs0) database. The number of speech samples $N$ is 6014 starting at sample number 4720. Using the plateau region between log $\epsilon \in [6.8, 7.5]$, the estimated correlation dimension $D_2$ is 2.583. Notice the other potential plateau region at log $\epsilon \in [5.7, 6.0]$ typical of "multiscale examples".
Figure 5.10: “Multiscale” example. (a) Shows the log of the correlation integral plotted as a function of log $\varepsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension $D_2$ as a function of $d_E$. These estimates were made using the unvoiced speech data for a male speaker, taken from the ISOLET (/ISOLET/isolet1/mrmh1) database. The number of speech samples $N$ is 3713 starting at sample number 4628. Using the plateau region between $\log \varepsilon \in [7.0, 7.5]$, the estimated correlation dimension $D_2$ is 2.9269. Notice the other potential plateau region at $\log \varepsilon \in [6.0, 6.5]$ typical of “multiscale” examples.
Figure 5.11: "Lacunarity" example. (a) Shows the log of the correlation integral plotted as a function of log \(\varepsilon\) for different embedding dimensions \(d_E\) from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension \(D_2\) as a function of \(d_E\). These estimates were made using the unvoiced speech data \"z\" for a female speaker, taken from the ISOLET (/ISOLET/isolet1/fet0) database. The number of speech samples \(N\) is 7400 starting at sample number 2882. Using the plateau region between log \(\varepsilon\) \(\in [5.5, 6.0]\), the estimated correlation dimension \(D_2\) is 4.1646.
Figure 5.12: “Lacunarity” example. (a) Shows the log of the correlation integral plotted as a function of $\log \epsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension $D_2$ as a function of $d_E$. These estimates were made using the unvoiced speech data \"s\" for a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjc1) database. The number of speech samples $N$ is 2723 starting at sample number 4364. Using the plateau region between $\log \epsilon \in [6.9, 7.5]$, the estimated correlation dimension $D_2$ is 3.31.
Figure 5.13: "Lacunarity" example. (a) Shows the log of the correlation integral plotted as a function of $\log \epsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension $D_2$ as a function of $d_E$. These estimates were made using the unvoiced speech data \["z"\] for a female speaker, taken from the ISOLET (/ISOLET/isolet1/fdcf0) database. The number of speech samples $N$ is 7667 starting at sample number 785. Using the plateau region between $\log \epsilon \in [5.5, 7.0]$, the estimated correlation dimension $D_2$ is 3.3562.
Figure 5.14: "Mixed" example. (a) Shows the log of the correlation integral plotted as a function of log $\epsilon$ for different embedding dimensions $d_{E}$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension $D_{2}$ as a function $d_{E}$. These estimates were made using the unvoiced speech data \"z\" for a female speaker, taken from the ISOLET (/ISOLET/isolat1/fmbd0) database. The number of speech samples $N$ is 8841 starting at sample number 1441. Using the plateau region between log $\epsilon \in [7.0, 7.6]$, the estimated correlation dimension $D_{2}$ is 1.0867. This is a "mixed" example because of the lacunarity exhibited between log $\epsilon \in [4.5, 5.5]$ in addition to the normal plateau region log $\epsilon \in [7.0, 7.6]$. 

134
Figure 5.15: "Mixed" example. (a) Shows the log of the correlation integral plotted as a function of \( \log\epsilon \) for different embedding dimensions \( d_E \) from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension \( D_2 \) as a function \( d_E \). These estimates were made using the unvoiced speech data \"z\" for a male speaker, taken from the ISOLET (\texttt{/ISOLET/isolet1/mjfv0}) database. The number of speech samples \( N \) is 3210 starting at sample number 2665. Using the plateau region between \( \log\epsilon \in [7.0, 7.9] \), the estimated correlation dimension \( D_2 \) is 0.4749. This is a "mixed" example because of the lacunarity exhibited between \( \log\epsilon \in [4.5, 5.3] \) in addition to the normal plateau region \( \log\epsilon \in [7.0, 7.9] \).
a single scalar realization or data sequence in equation (2.47). \( M \) is the total number of embedded vector points generated on the reconstructed attractor in phase space using the Taken’s embedding theorem [56] as described in section 2.8. When the embedding dimension \( d_E \) is small, \( M \) is high, i.e., the number of inter-point distances used in computing \( C_{q,d_E}(\epsilon) \) is higher. This produces a flat plateau region. However, for a fixed number of data points but increasing embedding dimensions, the total number of inter-point distances available for the computation of \( C_{q,d_E}(\epsilon) \) is reduced dramatically. The plateau region where the dimension is estimated is no longer flat. This makes it increasingly difficult to estimate the correlation dimension \( D_2 \). Fig. 5.16 shows the plot of the correlation integral and the local slope as a function of \( \log \epsilon \) for different embedding dimensions \( d_E \) from 5 to 15. The unvoiced speech data \("s"\) for a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. Notice the crossing local slope curves (Fig. 5.16 (b)) in the plateau region around \( \log \epsilon = 7 \) which is due to the increase in the embedding dimension \( d_E \). However, the ambiguity is resolved when the correlation integrals and local slopes are plotted for the same speaker at only the low range of embedding dimensions \( d_E \), from 5 to 8 in Fig. 5.17. Fig. 5.18 also shows the plot of the correlation integrals and the local slopes as a function of \( \log \epsilon \) for different embedding dimensions \( d_E \) from 5 to 15 for a different speaker. Notice the crossing local slopes in the plateau region around \( \log \epsilon = 6.2 \). This ambiguity is also resolved when the correlation integrals and the corresponding local slopes are plotted using only lower embedding dimensions \( d_E \) from 5 to 9 in Fig. 5.19. The unvoiced speech data \("s"\) for a male speaker are taken from the ISOLET (/ISOLET/isolet1/mjfv0) database.

The effects of using overly high embedding dimensions on a fixed length data sequence for male speakers and female speaker were most severe in the unvoiced speech sound \("s"\). This is due to the relatively small segment length \( N \) caused by the removal of the long silence region and as well as the quasi-periodic part in \("s"\) as compared to the truncation of only the silence in the unvoiced speech \("z"\).

In the next section, we consider the turbulent nature of unvoiced speech signals. We describe a method by which the nature of the turbulence involved in unvoiced speech production can be characterized.
Figure 5.16: (a) Shows the log of the correlation integral plotted as a function of $\log \epsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension, $D_2$ as a function of $d_E$. The estimates are computed for the unvoiced speech data "s" for a female speaker, taken from the ISOLET (/ISOLET/isolet1/fmc0) database. The number of speech samples $N$ is 4158 starting at 3810. Observe the ambiguity (artifacts) in the value of $D_2$ created by the overlapping or crossing of the local slope curves at higher embedding dimensions near the range of $\log \epsilon \in [6.8, 7.8]$. 

137
Figure 5.17: (a) Shows the log of the correlation integral plotted as a function of $\log \epsilon$ for different embedding dimensions $d_E$ from 5 to 8 for the same speaker as in Fig. 5.16. (b) depicts the corresponding local slope curves also plotted as a function of $\log \epsilon$ and $d_E$. Notice the four distinct regions discussed in section 2.13 and a better resolution of the ambiguity in Fig. 5.16 in the plateau region located near $\log \epsilon \in [6.6, 7.4]$. For this sound, the number of speech samples $N$ is 2665 starting at sample number 3210. The estimated correlation dimension $D_2$ is 2.5542.
Figure 5.18: (a) Shows the log of the correlation integral plotted as a function of log $\epsilon$ for different embedding dimensions $d_E$ from 5 to 15. (b) depicts the corresponding local slope curves which are estimates of the correlation dimension, $D_2$ as a function of $d_E$. The estimates are computed for the unvoiced speech data “s” for a male speaker, taken from the ISOLET (/ISOLET/ISOLET1/mjfv0) database. The number of samples $N$ is 4820 starting at sample number 858. Observe the ambiguity (artifacts) in the interval $\log \epsilon \in [6.0, 6.8]$ created by overlapping local slope curves at higher embedding dimensions.
Figure 5.19: (a) Shows the log of correlation integral plotted as a function of log $\epsilon$ for different embedding dimensions $d_E$ from 5 to 8 for the same speaker as in Fig. 5.18. (b) depicts the corresponding local slope curves also plotted as a function of log $\epsilon$ and $d_E$. Notice the four distinct regions discussed in section 2.13 and a better resolution of the ambiguity in Fig. 5.18 (b) in the plateau region located near log $\epsilon \in [6.0, 6.5]$. For this sound, the number of speech samples $N$ is 4820 starting at sample number 858. The estimated correlation dimension $D_2$ is 2.2468.
5.8 Low dimensional turbulence in unvoiced speech signals

In the dynamical modeling of what is technically called turbulence or "hard" turbulence, the estimated correlation dimension $D_2$ is so high that such modeling is invalid and discouraged [14]. The multifractal analysis of unvoiced speech signal has revealed that what is generically referred to in the speech modeling literature as turbulence is actually "soft" turbulence [128].

The major difference between soft turbulence and "hard" turbulence is their relative value of the Reynolds number $Re$ defined in section 3.2.2. According to Heslot et al. [128] in their study of helium gas at low temperatures, three regimes are possible depending on the value of the Reynolds number $Re$. There is a transition between the three states. A chaotic state is possible up to $Re = 2.5 \times 10^5$, a soft turbulent state from $Re > 2.5 \times 10^5$ to $Re = 4 \times 10^7$, and hard turbulence at higher $Re$.

Although the range of Reynolds number $Re$ found in helium gas is higher than that of speech, the transitions from noise to soft turbulence to chaos reported in the helium gas can also be observed in the local slope plot of the "normal" structures of unvoiced speech signals defined in section 5.6. For example, in Figure 5.20, there is a region (marked by the vertical dashed lines) in the $\frac{\partial}{\partial \log \epsilon} \log C_{2,d_E}(\epsilon)$ versus log $\epsilon$ plots after the noisy region but just before the plateau region in which as the embedding dimension $d_E$ is increased, the correlation dimension $D_2$ of the speech data varies almost linearly with embedding dimension $d_E$. This may be defined as the "soft" turbulent region following the transition pattern in helium gas. By taking a slice of the log $C_{2,d_E}(\epsilon)$ curves along log $\epsilon$ in this regions, the strength and nature of the soft turbulence can be determined [128]. Fig. 5.20 (a) shows the local slopes versus log $\epsilon$ for unvoiced speech data "s" for a male speaker taken from the ISOLET (/ISOLET/isolet1/mtdw0) database. A slice in the soft turbulent region taken along log $\epsilon = 6$, is shown using the vertical dashed line. Other examples of the local slopes versus log $\epsilon$ plot for various embedding dimensions are also shown in Fig. 5.20 (b), (c) and (d). In Fig. 5.20 (b), the unvoiced speech data for the sound "s"
spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmg0) database. The slice in the soft turbulent region, taken along $\log \varepsilon = 6.5$, is shown using the vertical dashed line. Fig. 5.20 (c) shows the local slopes plot for the unvoiced speech data for the sound \"z\" spoken a male speaker taken from the ISOLET (/ISOLET/isolet1/mrmh1) database. A slice in the soft turbulent region is taken along $\log \varepsilon = 5.5$ as shown by the dashed line. Finally Fig. 5.20 (d), shows the local slopes for the unvoiced speech data for the sound \"z\" spoken by a female speaker taken from the ISOLET (/ISOLET/isolet1/fec0) database. A slice in the soft turbulent region is taken along $\log \varepsilon = 6$ as shown by the dashed line. The corresponding correlation dimensions $D_2$ as a function of the embedding dimension $d_E$ for each speaker in the soft turbulent region in Fig. 5.20 (a)-(d) are depicted in Fig 5.21 (a)-(d), respectively. All plots exhibits an approximately linear increase with $d_E^2$.

This is in contrast to curves aa and cc in Fig. 5.22 (b). They represent slices through the plateau region, soft turbulence region bb and the noisy region, respectively, in Fig. 5.22 (a) of typical local slope plots for unvoiced speech signals. In Fig. 5.22 (b), the three slices in Fig. 5.22 (a) are plotted as a function of increasing embedding dimension, $d_E$. Notice in Fig. 5.22 (b) that the convergence of curve aa from the plateau region, the steady linear increase in correlation dimension $D_2$ in curve bb from the soft turbulent region and the erratic behavior of curve cc illustrating the noise region as the embedding dimension $d_E$ increases.

Also notice that in all the speech examples used in Fig. 5.21, the estimates of the correlation dimension $D_2$ are less than 6 even with an embedding dimension $d_E$ as high as 15. This suggests that the dynamical modeling of the soft turbulence observed during unvoiced speech production is now possible since the estimated correlation dimension $D_2$ in the soft turbulent region is low as compared to the suggested high dimensional turbulence explanation offered by Narayanan et al. [14] which discouraged such dynamical models in the turbulent region.

In the next section, we present the results of applying the procedure for estimating

---

2In Appendix D, more examples of the correlation dimension $D_2$ as a function of the embedding dimension $d_E$ in the soft turbulent region will be given for rest of the unvoiced speech samples analyzed in this dissertation.
Figure 5.20: (a)-(d) Shows the local slope curves as a function of $\log \epsilon$ for embedding dimensions $d_E$ from 5 to 15. The speech data are from both male and female speakers taken from the ISOLET database (/ISOLET/isoletl/mtdw0.fcmg0.mrmhi.fec0, respectively). Notice that in all cases, there exists regions of soft turbulence between the noisy and plateau regions, marked by the vertical dash lines. In each subfigure, the vertical-axis corresponds to the correlation dimension $D_2$ whereas the horizontal-axis corresponds to $\log \epsilon$. 
Figure 5.21: (a)-(d) Show the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in Fig. 5.20 (a)-(d), respectively. In each figure, the vertical-axis corresponds to the correlation dimension $D_2$, whereas the horizontal-axis corresponds to embedding dimension $d_E$. 

144
Figure 5.22: (a) shows the local slope curves of the correlation integral as a function of $\log \epsilon$ and $d_E$ of the speech sound "z" spoken by a female speaker from the ISOLET (/ISOLET/isolet1/fec0) database. (b) Cross-sectional slices of the local slope curves through the plateau region, the soft-turbulent region and the noise region are displayed, aa corresponds to the plateau region in Fig. 5.22 (a) located at $\log \epsilon = 7.0$, bb corresponds to the soft-turbulent region at $\log \epsilon = 6.0$ and cc is the noisy region at $\log \epsilon = 4.5$. 

145
generalized dimensions $D_q$ in section 5.5 to unvoiced speech signals. We describe the different types of generalized dimension $D_q$ plots encountered and explain the significance of the plots to unvoiced speech signals.

5.9 Generalized dimension $D_q$ of unvoiced speech signals

Multifractal analysis has been used extensively in studying the nature of turbulence in fluid flow [102, 129, 130, 101]. Due to the existence of soft turbulence and its associated variability in the correlation integral and local slope structures as well as the fluid flow nature of unvoiced speech signals [76, 87], we analyze unvoiced speech signals using multifractal measures to determine the presence of various fractal regimes in speech attractors. These fractal regimes are revealed by the estimation of spectra of generalized dimensions $D_q$. Each non-constant region of $D_q$ corresponds to a different fractal structure present on the attractor of the unvoiced speech signal.

In this section, we present and discuss the result of multifractal analysis on unvoiced speech signals. The generalized dimensions $D_q$ were obtained from unvoiced speech data using the procedure discussed in section 5.5. We estimate the generalized dimensions $D_q$ for $q$ from -5 to 5 for all the unvoiced speech data.

5.9.1 Result of $D_q$ estimation for unvoiced speech signals

In the multifractal analysis of unvoiced speech signals, using the procedures itemized in section 5.5, three different types of generalized dimensions $D_q$ curves were identified. They include:

- **Normal**: For a chaotic dynamical system, which exhibits multifractal behavior, the estimated generalized dimensions give a set of $D_q$ values along a smooth $D_q$ curve that satisfies

$$D_q < D_{q'} \quad \text{for} \quad q > q'$$

(5.6)
as explained in section 4.7.2. An example of such a smooth $D_q$ curve is shown in Fig. 4.5 for the triadic Cantor set. In unvoiced speech multifractal analysis, whenever the expression for $D_q$ in equation (5.6) is satisfied for $|q| \leq 5$, the estimated generalized dimensions curve shall be referred to in this thesis as “normal”. Fig. 5.23 shows an example of a “normal” curve of generalized dimensions $D_q$ versus $q$, estimated for the unvoiced speech sound \“s\” spoken by a female speaker taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. The corresponding log-log plots of $C_{q,d}(\epsilon)$ versus $\epsilon$ from which the individual $D_q$ values in Fig. 5.23 were estimated are depicted in Appendix C. Figures 5.24 and 5.25 also show examples of “normal” generalized dimension $D_q$ curves for female and for male unvoiced speech data from the ISOLET (/ISOLET/isolet1/fdcf0, mtdw0) database, respectively. These plots can be compared with the $D_q$ curve for the triadic Cantor set in Fig. 4.5 (a) to reveal a monotonic decrease in the $D_q$ values as a function of $q$ in the range $|q| \leq 5$.

\begin{itemize}
  \item \textbf{Phase-Transitions}: In the generalized dimensions $D_q$ curve for some chaotic dynamical systems that exhibit multifractal behavior, the derivative of “normal” $D_q$ versus $q$ curve sometimes has discontinuities. These discontinuities in the derivative of the $D_q$ curve are analogous to the phenomena of phase-transitions in condensed matter physics [131, 110, 111], hence the term \textbf{phase-transitions}. These phase transitions are related to free energy in thermodynamics. Phase transitions in a chaotic attractor introduce a small amplitude, high-frequency content into the otherwise low-frequency $D_q$ curve. A notable chaotic attractor which exhibits phase-transitions is the Logistic Map\footnote{The closed form expression [3] for the generalized dimensions $D_q$ of the Logistic Map is 

$$D_q = \begin{cases} 
1, & \text{if } q < 2 \\
\frac{q}{2(q-1)}, & \text{if } q \geq 2 
\end{cases}.$$ 

The phase-transition occurs at $q = 2$.} [3]. The generalized dimensions $D_q$ curve as a function $q$ for the Logistic Map is shown in Fig. 5.26. Notice the discontinuity in the derivative of the \textit{Phase-Transitions:} In the generalized dimensions $D_q$ curve for some chaotic dynamical systems that exhibit multifractal behavior, the derivative of “normal” $D_q$ versus $q$ curve sometimes has discontinuities. These discontinuities in the derivative of the $D_q$ curve are analogous to the phenomena of phase-transitions in condensed matter physics [131, 110, 111], hence the term \textbf{phase-transitions}. These phase transitions are related to free energy in thermodynamics. Phase transitions in a chaotic attractor introduce a small amplitude, high-frequency content into the otherwise low-frequency $D_q$ curve. A notable chaotic attractor which exhibits phase-transitions is the Logistic Map\footnote{The closed form expression [3] for the generalized dimensions $D_q$ of the Logistic Map is 

$$D_q = \begin{cases} 
1, & \text{if } q < 2 \\
\frac{q}{2(q-1)}, & \text{if } q \geq 2 
\end{cases}.$$ 

The phase-transition occurs at $q = 2$.} [3]. The generalized dimensions $D_q$ curve as a function $q$ for the Logistic Map is shown in Fig. 5.26. Notice the discontinuity in the derivative of the

$$D_q = \begin{cases} 
1, & \text{if } q < 2 \\
\frac{q}{2(q-1)}, & \text{if } q \geq 2 
\end{cases}.$$ 

The phase-transition occurs at $q = 2$.}
Figure 5.23: Shows the "normal" generalized dimensions $D_q$ versus $q$ curve for an unvoiced speech sound \"s\" uttered by a female speaker. The speech data are from the ISOLET (/ISOLET/isolet1/fcmc0) database.
Figure 5.24: Shows the “normal” generalized dimensions $D_q$ versus $q$ curve for an unvoiced speech sound /“z”/ uttered by a female speaker. The speech data are from the ISOLET (/ISOLET/isolet1/fdcf0) database.
Figure 5.25: Shows the generalized dimensions $D_q$ versus $q$ curve for an unvoiced speech sound "z" uttered by a male speaker. The speech data are from the ISOLET (/ISOLET/isolet1/mtdw0) database.
$D_q$ curve at $q = 2$. Examples of unvoiced speech signals with suspected phase-transitions are shown in Fig. 5.27 and Fig. 5.28. Note the change in the shape of the $D_q$ curves at $q = 0$ and $q = 1$, respectively.

- **Abnormal**: In the multifractal analysis of experimental data, the inadequate length $N$ of the time series segment used in the analysis sometimes shows up when the estimated generalized dimensions $D_q$ violate the monotonic expression in equation (5.6). Therefore, in unvoiced speech multifractal analysis, any estimated $D_q$ curve that violates this expression within $-5 \leq q \leq 5$ shall be referred to in this thesis as “abnormal”. Fig 5.29 shows an example of an abnormal $D_q$ curve encountered in the multifractal analysis of unvoiced speech signals. A non-constant monotonic decreasing $D_q$ curve causes significant errors in the calculation of the singularity spectrum, $f(\alpha)$, and hence, will not be used in the multifractal measure calculations.

In the next section, we use the min-max Parks-McClellan filter design approach of section 4.9 to evaluate the Legendre transform of the “normal” generalized dimensions $D_q$ estimated in the last section in order to obtain an estimate of the $f(\alpha)$ singularity spectrum for unvoiced speech signals.

### 5.10 $f(\alpha)$ singularity spectrum for unvoiced speech

In chapter 4, we discussed the $f(\alpha)$ singularity spectrum formalism related by a Legendre transform to the generalized dimensions $D_q$. We also introduced the Park/McClellan min-max filter design method that can be used to improve the accuracy of the conventional interpolation and differentiation approximations needed to compute the Legendre transform between $D_q$ and $f(\alpha)$. We now apply the min-max filter method to the generalized dimensions $D_q$ of unvoiced speech signals.

In order to be consistent, the $f(\alpha)$ singularity spectrum of unvoiced speech signals will only be evaluated using monotonic decreasing $D_q$ values in the range of $q$ between $-3$ and $3$. This limit was chosen because most of the generalized dimensions $D_q$...
Figure 5.26: Schematic plot of the generalized dimensions $D_q$ versus $q$ showing the phase-transition for of Logistic Map [3]. The discontinuity in the derivative of the $D_q$ occurs at $q = 2$. 
Figure 5.27: Shows a "phase transition" in the estimated values of the generalized dimension $D_q$ versus $q$ for an unvoiced speech sound "z" uttered by a male speaker. The speech data are taken from the ISOLET (/ISOLET/isolet1/mrmh1) database. Notice the change in the $D_q$ curve around $q = 0$. 
Figure 5.28: Shows a "phase transition" in the estimated values of the generalized dimension $D_q$ versus $q$ for an unvoiced speech sound \"s\" uttered by a male speaker. The speech data are taken from the ISOLET (/ISOLET/isolet1/mrmb1) database. Notice the change in the $D_q$ curve around $q = 1$. 

154
Figure 5.29: Shows an “abnormal” curve of estimated generalized dimensions $D_q$ versus $q$ values for an unvoiced speech \"s\" uttered by a male speaker. The speech data are taken from the ISOLET (/ISOLET/isolet1/mrs0) database. This is an invalid $D_q$ curve. The expression in equation (5.6) that requires $D_q$ to be monotonic decreasing has been violated around $q < -3$. 

155
Figure 5.30: Shows an "abnormal" curve of estimated generalized dimensions $D_q$ versus $q$ values for an unvoiced speech "z" uttered by a female speaker. The speech data are taken from the ISOLET (/ISOLET/isoletl/fmc0) database. This is an invalid $D_q$ curve. The expression in equation (5.6) that requires monotonic decreasing $D_q$ has been violated around $q < -2$ and $q < -1$. 
Figure 5.31: Shows an "abnormal" curve of estimated generalized dimensions $D_q$ versus $q$ values for an unvoiced speech \"z\" uttered by a male speaker. The speech data are taken from the ISOLET (/ISOLET/isoleti/mrs0) database. This is an invalid $D_q$ curve. The expression in equation (5.6) that requires monotonic decreasing $D_q$ has been violated around $|q| < 1$. 
curves were “normal”, i.e. monotonic decreasing in the range of $|q| < 3$. Accurate $D_q$ estimates for $q < -3$ required large data lengths than were available.

For each speech signal, the generalized dimensions $D_q$ were obtained using equation (2.32) at integer $q$ values ($\Delta q = 1$). The resulting sparse generalized dimension $D_q$ samples were then optimally interpolated by a factor of 4 using a 33$^{rd}$ order min-max lowpass filter with passband between normalized frequencies 0 and $F_p = 0.065$. The transition region is between the normalized frequencies $F_p = 0.065$ and $F_s = 0.185$ and the stopband region is between normalized frequencies $F_s = 0.185$ and $F_N = 0.5$. The derivatives in the Legendre transform in equations (4.18) and (4.19) were evaluated with a 5$^{th}$ order min-max differentiator whose frequency response is approximately linear in the passband region with the normalized frequencies between 0 and $F_p = 0.15$. The transition region is between $F_p = 0.15$ and $F_s = 0.25$. The stopband extends from $F_s = 0.25$ to $F_N = 0.5$. Fig. 5.32 (a) shows the optimally interpolated generalized dimensions $D_q$ for an unvoiced speech signal \"s\" uttered by a female speaker. This corresponds to the optimally interpolated $D_q$ curve between $q$ from -3 to 3 for the speech signal whose $D_q$ curve is shown in Fig. 5.23. The original $D_q$ samples are shown in asterisks “*” and the interpolated $D_q$ values are depicted with plus signs “+”. The original estimates appear at multiples of $\Delta q = 1$, while the interpolated samples appear at multiples of $\Delta q' = 0.25$. Fig. 5.32 (b) shows the corresponding $f(\alpha)$ singularity spectrum. This is a valid spectrum with no spurious end or interior points. It is convex and has one maximum at $D_0 = 6.02$ which is exactly the same as the $D_0$ value from the interpolated $D_q$ curve of Fig. 5.32(a). Fig. 5.33 shows how little these plots change if we use the entire range of $D_q$ data points for $|q| < 5$ as compared to Fig. 5.34. The singularity spectrum in Fig. 5.34, was computed using only the $D_q$ data points from $|q| < 3$. In contrast, Fig. 5.35 shows the less accurate estimate of the singularity spectrum that results from conventional techniques which use sub-optimal linear interpolation and a first difference approximation to differentiation on sparsely sampled generalized dimensions $D_q$ using the same range $|q| < 3$.

Fig. 5.36 to Fig. 5.39 show the results of applying the min-max filter design method to the other generalized dimensions $D_q$ curves, $|q| = 3$, for both the
Figure 5.32: (a) Shows the optimally interpolated generalized dimensions (GD) $D_q$, $|q| = 3$ of an unvoiced speech signal "s". The speech data for a female speaker are taken from the ISOLET (/ISOLET/isolet1/fcm0) database. Fig. 5.32(b) shows the corresponding $f(\alpha)$ singularity spectrum using the min-max filter design method for both interpolation and differentiation filters. From the $f(\alpha)$ curve, $D_0 = 6.1$ and the estimated $D_{-\infty}$ and $D_{\infty}$ values are 9.25 and 1.4, respectively.
Figure 5.33: Shows the $f(\alpha)$ singularity spectrum using the min-max filter design method on the generalized dimensions (GD) $D_q, \ |q| = 5$ of an unvoiced speech signal "z". The speech data for a male speaker are taken from the ISOLET (/ISOLET/isolet1/mdtw0) database.
Figure 5.34: Shows the $f(\alpha)$ singularity spectrum using the min-max filter design method on the generalized dimensions (GD) $D_q$, $|q| = 3$ of the same speech data as the one used in estimating the $D_q$ in Fig. 5.33. The speech data for a male speaker are taken from the ISOLET (/ISOLET/isolet1/mdtw0) database.
Figure 5.35: (a) Shows the $f(\alpha)$ singularity spectrum using the conventional techniques which uses linear approximation and a first difference approximation to differentiation. (b) shows the $f(\alpha)$ singularity spectrum using the min-max filter design method on the generalized dimensions (GD) $D_q$, in the range $|q| = 3$. The speech data \"s\" for a female speaker are taken from the ISOLET (/ISOLET/isolet1/mdtw0) database.
“normal” and “phase-transitions” type of $D_q$ curves in section 5.9.1. In particular, Fig. 5.36 (a) shows the optimally interpolated generalized dimensions $D_q$ curve for an unvoiced speech signal \"z\"\ uttering by a female speaker from the ISOLET (/ISOLET/isolet1/fdcf0) database. From the $f(\alpha)$ curve in Fig. 5.36(b), we obtain $D_0 = 5.03$ and the estimated $D_{-\infty}$ and $D_{\infty}$ values are 9.6 and 1.8, respectively. Fig. 5.37 depicts the interpolated $D_q$ curve and the corresponding $f(\alpha)$ spectrum for an unvoiced speech \"z\"\ for a male speaker. The speech segment is taken from ISOLET (/ISOLET/isolet1/mtdw0) database. From the $f(\alpha)$ singularity spectrum in Fig. 5.37 (b), we obtain $D_0 = 4.25$ and the estimated $D_{-\infty}$ and $D_{\infty}$ values are 6.35 and 1.85, respectively.

For the unvoiced speech signals with \"phase-transition\" shown in Figures 5.27 and 5.28, we also show their interpolated $D_q$ curves as well as the $f(\alpha)$ singularity spectrum. Fig. 5.38 (a) corresponds to the generalized dimensions $D_q$ for unvoiced speech \"z\"\ uttered by a male. The speech segment is taken from the ISOLET (/ISOLET/isolet1/mrmh1) database. From the $f(\alpha)$ singularity curve of Fig. 5.38 (b), we obtain $D_0 = 3.73$ and the estimated $D_{-\infty}$ and $D_{\infty}$ values are 8.7 and 0.9, respectively. Finally, Fig. 5.39 (a) shows the optimally interpolated generalized dimensions $D_q$ curve for an unvoiced speech signal \"s\"\ uttering by a male speaker from the ISOLET (/ISOLET/isolet1/mrmh1) database. From the $f(\alpha)$ curve in Fig. 5.39(b), we obtain $D_0 = 4.45$ and the estimated $D_{-\infty}$ and $D_{\infty}$ values are 8.5 and 1.5, respectively.

In all the unvoiced speech cases discussed above, the sparse ($\Delta q = 1$) $D_q$ curves were well interpolated ($\Delta q' = \frac{1}{4}$ and there were no spurious endpoints in the $f(\alpha)$ singularity spectrum. Also, the speech examples which had $D_q$ with phase-transitions, are always difficult for other algorithms to handle [12]; however, our algorithm produces $D_q$ that were well interpolated by the min-max filter method producing valid $f(\alpha)$ singularity spectrum. One of this (and other) algorithms error is that there are some points on the singularity curve that are below the zero line, i.e., $f(\alpha) < 0$. These are physically meaningless results which may occur as a result of artifacts of relatively large variability in $D_q$ for high values of $|q|$ [102]. We also conjecture that the error at the ends of the $f(\alpha)$ singularity curve where $f(\alpha) < 0$ may be due to
artifacts resulting from the approximations in the filter design.
Figure 5.36: (a) Shows the optimally interpolated generalized dimensions $D_q$, $|q| = 3$ of an unvoiced speech signal “z”. The speech data for a female speaker are taken from the ISOLET (/ISOLET/isolet1/f00c0) database. (b) shows the corresponding $f(\alpha)$ singularity spectrum using the min-max filter design method. From the $f(\alpha)$ curve, we obtain $D_0 = 5.03$ and the estimated $D_{-\infty}$ and $D_{\infty}$ values are 9.6 and 1.8, respectively.
Figure 5.37: (a) Shows the optimally interpolated generalized dimensions $D_q$, $|q| = 3$ of an unvoiced speech signal \"z\". The speech data for a male speaker are taken from the ISOLET (/ISOLET/isolat1/mtdw0) database. (b) shows the corresponding $f(\alpha)$ singularity spectrum using the min-max filter design method. From the $f(\alpha)$ curve, we obtain $D_0 = 4.25$ and the estimated $D_{-\infty}$ and $D_{\infty}$ values are 6.35 and 1.85, respectively.
Figure 5.38: (a) Shows the optimally interpolated generalized dimensions $D_q, |q| = 3$ of an unvoiced speech signal \\"z\". The speech data for a male speaker are taken from the ISOLET (/ISOLET/isolet1/mrmh1) database. Fig. (b) shows the corresponding $f(\alpha)$ singularity spectrum using the min-max filter design method. From the $f(\alpha)$ curve, we obtain $D_0 = 3.73$ and the estimated $D_{-\infty}$ and $D_{\infty}$ values are 8.7 and 0.9, respectively.
Figure 5.39: (a) Shows the optimally interpolated generalized dimensions $D_q$, $|q| = 3$ of an unvoiced speech signal "s". The speech data for a male speaker are taken from the ISOLET (/ISOLET/isolet1/mrmh1) database. Fig. 5.39(b) shows the corresponding $f(\alpha)$ singularity spectrum using the min-max filter design method. From the $f(\alpha)$ curve, we obtain $D_0 = 4.45$ and the estimated $D_{-\infty}$ and $D_{\infty}$ values are 8.5 and 1.5, respectively.
5.11 Discussion of $D_q$ and $f(\alpha)$ results

Theoretically, the generalized dimensions $D_q$ of a multifractal chaotic sequence are known for all values of $q$, but in practice, they are only computed for a finite range of $q$. Due to data length constraints, it is almost impossible to extract the extrema generalized dimensions $D_{-\infty}$ and $D_{\infty}$ from real data because to do so would require infinite data records as alluded to in section 4.5. However, from a novel implementation of the Legendre transform relating $f(\alpha)$ to $D_q$ by using the min-max filter design method, the two extrema values of $D_q$, i.e., $D_{-\infty}$ and $D_{\infty}$, can now be estimated at $f(\alpha) = 0$, even when only a few (e.g. 7) generalized dimensions $D_q$ samples are known. These extrema points $(D_{-\infty}, D_{\infty})$ describe the densest and least populated areas, respectively, of the speech attractor and where the calculation of the $D_q$ is very difficult.

The analysis in this thesis has shown that unvoiced speech signals \"s\" and \"z\" may be multifractal in nature. The estimated generalized dimensions $D_q$ for the unvoiced fricatives \"s\" for selected male speakers from the ISOLET database range from 1.32 to 8.77. For the same speech sound, the estimated generalized dimensions $D_q$ for selected female speakers from the ISOLET database range between 1.20 and 6.97. In the voiced fricatives \"z\", the estimated generalized dimensions $D_q$ for selected male speakers from the same database range from 1.07 to 7.72 whereas the estimated generalized dimensions $D_q$ values for selected female speakers range from 0.73 to 7.19. These results (also shown in Table 5.1) reveal that multifractality is not peculiar to a certain gender or unvoiced speech type, but that it is found in both male speakers and female speakers as well as voiced and unvoiced fricative sounds.

The variability and inadequacy of the correlation dimension $D_2$, which may be due to variations in the Reynolds number $Re$ of the underlying soft turbulence in the speech production for different speakers, has now been resolved by the new result that unvoiced speech sounds may be multifractal. From the multifractal analysis, each unvoiced speech sound is characterized by its own set of non-constant generalized dimensions $D_q$ which clearly define the structure of its attractor. In all the unvoiced speech signals analyzed, the set of valid generalized dimensions $D_q$ values were different for the various speakers.
From each group of unvoiced speech signals analyzed, the $f(\alpha)$ singularity spectrum was evaluated from the valid generalized dimensions $D_q$ values with $q$ from -3 to 3. By valid, we mean that the generalized dimensions $D_q$ values estimated from unvoiced speech sound strictly satisfy

$$D_q < D_{q'} \quad \text{for} \quad q > q'.$$

(5.7)

in the range of $q$ defined above. This is to ensure that $f(\alpha)$ singularity spectrum evaluated from the generalized dimensions $D_q$ is valid. The $f(\alpha)$ singularity spectrum are not computed for generalized dimensions $D_q$ for which the equation is not satisfied.

Using the min-max filter design method of evaluating the Legendre transform of equation (4.19), discussed in section 4.9, the values of $\alpha$, i.e., $\alpha_{\text{min}}$ and $\alpha_{\text{max}}$ at $f(\alpha) = 0$ for the unvoiced fricative speech sound "s" for some of the selected male speakers from the ISOLET database ranges from 0.5 to 11.6. These $\alpha$ values are also the estimate of $D_{-\infty}$ and $D_{\infty}$, the dimensions of the most rarefied and densest parts of the attractor respectively. For some of the selected female speakers, the values of $\alpha_{\text{min}}$ and $\alpha_{\text{max}}$ for the unvoiced fricative speech sound "s" ranges from 1.4 to 9.25. In the case of the voiced fricative "z", $\alpha_{\text{min}}$ and $\alpha_{\text{max}}$ for some of the selected male speakers range from 0.9 to 8.7 whereas for the female counterparts, the values are from 1.01 to 9.6. The results for the values of $\alpha_{\text{min}}$ and $\alpha_{\text{max}}$ for the unvoiced fricative speech sounds are also summarized in Table 5.2. The scaling exponent $\alpha$ values obtained from the $f(\alpha)$ singularity spectrum suggest that the probability density functions (pdf) of unvoiced speech signals are predominantly low in intensity since the scaling exponent values $\alpha$ are mostly greater than 1.

The asymmetric spread of $\alpha$ values with the corresponding $f(\alpha)$ around the maximum of $f(\alpha)$ reveal the inhomogeneity in the attractors of unvoiced speech signals as do the variations in the $D_q$ values. In particular, if we define $\alpha_0$ to be the $\alpha$ value at which the $f(\alpha)$ attains its maximum, then the inhomogeneity in the attractors of unvoiced speech signals can be seen clearly by noticing that the interval $[\alpha_{\text{min}}, \alpha_0]$ is wider or smaller than the interval $[\alpha_0, \alpha_{\text{max}}]$. Based on the asymmetric spread of $\alpha$ values around the maximum of $f(\alpha)$, we regard the turbulent energy fields generated
during unvoiced speech production as non-homogeneous fractals. The scaling property of this turbulent energy is now completely described by the estimated generalized dimensions $D_q$.

In Appendix D, more examples of the interpolated generalized dimension $D_q$ curve and their corresponding $f(\alpha)$ singularity spectrum will be shown for unvoiced speech signals "s" and "z" uttered by American male and female speakers in the ISOLET database. A summary of important parameters obtained from all of the plots in the appendix D are displayed in Tables 5.3 and 5.4.
Table 5.1: Shows a summary of the generalized dimension $D_q$ results for unvoiced speech signals \"s\" and \"z\".

<table>
<thead>
<tr>
<th>Sound</th>
<th>Gender</th>
<th>min $D_q$</th>
<th>max $D_q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>Male</td>
<td>1.32</td>
<td>8.77</td>
</tr>
<tr>
<td>s</td>
<td>Female</td>
<td>1.20</td>
<td>6.97</td>
</tr>
<tr>
<td>z</td>
<td>Male</td>
<td>1.07</td>
<td>7.72</td>
</tr>
<tr>
<td>z</td>
<td>Female</td>
<td>0.73</td>
<td>7.19</td>
</tr>
</tbody>
</table>

Table 5.2: Shows a summary of the values of $\alpha_{\text{min}}$ and $\alpha_{\text{max}}$ for the unvoiced fricative speech sound \"s\" and \"z\". The values correspond to the estimates of $D_\infty$ and $D_{-\infty}$ obtained by the min-max method described in chapter 4.

<table>
<thead>
<tr>
<th>Sound</th>
<th>Gender</th>
<th>$\alpha_{\text{min}}$</th>
<th>$\alpha_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>Male</td>
<td>0.50</td>
<td>11.6</td>
</tr>
<tr>
<td>s</td>
<td>Female</td>
<td>1.40</td>
<td>9.25</td>
</tr>
<tr>
<td>z</td>
<td>Male</td>
<td>0.90</td>
<td>8.70</td>
</tr>
<tr>
<td>z</td>
<td>Female</td>
<td>1.01</td>
<td>9.60</td>
</tr>
</tbody>
</table>
Table 5.3: Shows a summary of the multifractal results for the unvoiced speech sound \"s\" from the ISOLET database, spoken by female (f) and male (m) speakers and located in the file whose name is given in the third column. The $D_0$ data points were obtained from the $D_q$ estimates at the index $q = 0$. The $D_{-\infty}$ and $D_{\infty}$ values were estimated from the singularity spectrum $f(\alpha)$ estimates to correspond to the points $\alpha_{\min}$ and $\alpha_{\max}$ respectively at which $f(\alpha) = 0$. $D_{-\infty}$ and $D_{\infty}$ were not estimated for sounds with invalid (non-monotonic) $D_q$ values in the range $|q| < 3$. 

<table>
<thead>
<tr>
<th>Sound</th>
<th>Gender</th>
<th>ISOLET file</th>
<th>$D_{-\infty} = \alpha_{\min}$</th>
<th>$D_0$</th>
<th>$D_{\infty} = \alpha_{\max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>f</td>
<td>fmbd0</td>
<td></td>
<td>5.5696</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>f</td>
<td>fme0</td>
<td></td>
<td>5.7238</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>f</td>
<td>fmb0</td>
<td></td>
<td>5.7992</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>f</td>
<td>fet0</td>
<td></td>
<td>6.0745</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>f</td>
<td>fews0</td>
<td></td>
<td>6.0917</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>f</td>
<td>fkh0</td>
<td></td>
<td>6.0917</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>f</td>
<td>fcmc0</td>
<td>9.25</td>
<td>6.10</td>
<td>1.40</td>
</tr>
<tr>
<td>s</td>
<td>f</td>
<td>fcmg0</td>
<td></td>
<td>6.1630</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>f</td>
<td>fec0</td>
<td></td>
<td>6.1677</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>f</td>
<td>fdcf0</td>
<td></td>
<td>6.4545</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>m</td>
<td>mrmh1</td>
<td>8.50</td>
<td>4.4512</td>
<td>1.50</td>
</tr>
<tr>
<td>s</td>
<td>m</td>
<td>mjp0</td>
<td>8.50</td>
<td>4.7854</td>
<td>1.10</td>
</tr>
<tr>
<td>s</td>
<td>m</td>
<td>mjcl</td>
<td>7.40</td>
<td>5.0601</td>
<td>1.60</td>
</tr>
<tr>
<td>s</td>
<td>m</td>
<td>mnjh</td>
<td>7.80</td>
<td>5.2375</td>
<td>1.40</td>
</tr>
<tr>
<td>s</td>
<td>m</td>
<td>mjrs0</td>
<td></td>
<td>5.3148</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>m</td>
<td>mtdw0</td>
<td></td>
<td>5.4643</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>m</td>
<td>mrs0</td>
<td>9.70</td>
<td>5.8887</td>
<td>1.40</td>
</tr>
<tr>
<td>s</td>
<td>m</td>
<td>mnre0</td>
<td>10.8</td>
<td>6.3136</td>
<td>1.20</td>
</tr>
<tr>
<td>s</td>
<td>m</td>
<td>mjfv0</td>
<td>11.6</td>
<td>7.0103</td>
<td>1.20</td>
</tr>
<tr>
<td>s</td>
<td>m</td>
<td>msa0</td>
<td></td>
<td>7.0352</td>
<td></td>
</tr>
</tbody>
</table>
### Table 5.4: Shows a summary of the multifractal results for the unvoiced speech sound “z” from the ISOLET database, spoken by female (f) and male (m) speakers and located in the file whose name is given in the third column. The $D_0$ data points were obtained from the $D_q$ estimates at the index $q = 0$. The $D_{-\infty}$ and $D_{\infty}$ values were estimated from the singularity spectrum $f(\alpha)$ estimates to correspond to the points $\alpha_{\min}$ and $\alpha_{\max}$ respectively at which $f(\alpha) = 0$. $D_{-\infty}$ and $D_{\infty}$ were not estimated for sounds with invalid (non-monotonic) $D_q$ values in the range $|q| < 3$.

<table>
<thead>
<tr>
<th>Sound</th>
<th>Gender</th>
<th>ISOLET file</th>
<th>$D_{-\infty} = \alpha_{\min}$</th>
<th>$D_0$</th>
<th>$D_{\infty} = \alpha_{\max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>z</td>
<td>f</td>
<td>fmbd0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>f</td>
<td>fme0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>f</td>
<td>fcmg0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>f</td>
<td>fec0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>f</td>
<td>fmb0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>f</td>
<td>fcmc0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>f</td>
<td>fet0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>f</td>
<td>fdcf0</td>
<td>9.6</td>
<td>5.0300</td>
<td>1.8</td>
</tr>
<tr>
<td>z</td>
<td>f</td>
<td>fews0</td>
<td>7.6</td>
<td>5.0525</td>
<td>1.01</td>
</tr>
<tr>
<td>z</td>
<td>f</td>
<td>fkh0</td>
<td></td>
<td>5.2829</td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>m</td>
<td>mrmhl</td>
<td>8.7</td>
<td>3.7249</td>
<td>0.9</td>
</tr>
<tr>
<td>z</td>
<td>m</td>
<td>msa0</td>
<td></td>
<td>3.7992</td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>m</td>
<td>mjfv0</td>
<td></td>
<td>3.9589</td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>m</td>
<td>mjcl</td>
<td></td>
<td>4.0221</td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>m</td>
<td>mrs0</td>
<td></td>
<td>4.1537</td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>m</td>
<td>mtdw0</td>
<td>6.35</td>
<td>4.2551</td>
<td>1.85</td>
</tr>
<tr>
<td>z</td>
<td>m</td>
<td>mjrs0</td>
<td></td>
<td>4.8070</td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>m</td>
<td>mjpo</td>
<td></td>
<td>5.3001</td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>m</td>
<td>mnre0</td>
<td></td>
<td>6.3615</td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>m</td>
<td>mnjh</td>
<td></td>
<td>6.4390</td>
<td></td>
</tr>
</tbody>
</table>


Chapter 6

Conclusions and Future work

The complexity of speech production and modeling cannot be overemphasized. A lot of progress has been made in the linear modeling of speech [88, 86, 9]. However, nonlinear dynamical analysis and modeling of speech signals is still at its infancy. Rapid progress can be made by abstracting the essential features of the nonlinear dynamical models underlying speech production. The extraction of these key features is the right direction towards finding the "ultimate" speech model. Therefore, the problem addressed in this dissertation is the analysis of unvoiced speech signals with the aid of nonlinear dynamical systems, i.e., chaos. We have used multifractal measures to determine the presence of various fractal regimes present in the attractors of unvoiced speech signals.

The primary objective of the dissertation is to capture the chaotic properties of unvoiced speech signals in the embedded vector space more succinctly by not only estimating the correlation dimension $D_2$ which we have shown to be incomplete and inadequate, but by also estimating the generalized dimension $D_q$. The generalized dimensions were estimated from phase space reconstructed vectors of single scalar variable realization of unvoiced speech signals. Results from the multifractal analysis have shown that unvoiced speech signals indeed may be multifractal in nature. There exist an infinite hierarchy of generalized dimensions $D_q$ in unvoiced speech signals. The generalized dimensions $D_q$ are unequal. The largest of these dimensions can now be used as an indication of the minimum dimension required in the phase space reconstruction of any realistic dynamical modeling of speech signals [63].
The multifractal analysis also shows that unvoiced speech signals exhibit low-dimensional chaos as well as "soft" turbulence. Contrary to the opinion that unvoiced speech signals are generated from what is technically know as "hard" turbulent flow, in which the dimension of a dynamical model is very high [14, 63], unvoiced speech signals may actually be generated from "soft" turbulent flow as discussed in section 5.8. Although, the estimated correlation dimensions $D_2$ in the soft turbulent region increased with increasing embedding dimensions $d_E$, the estimated $D_2$ is low and the increase is gradual. Hence, this presents us with the possibility that dynamical modeling of soft turbulence in unvoiced speech signals may be a reality.

In this dissertation, we also explore the relationship between the estimated generalized dimensions $D_q$ and the $f(\alpha)$ singularity spectrum. With the aid of signal processing methods of min-max filter design, the $f(\alpha)$ singularity spectrum has been accurately evaluated from the estimated generalized dimensions $D_q$. The importance of this new approach to singularity spectrum calculation is its improved accuracy and its computational reduction. This is due to the fact that the $D_q$ curve can be coarsely sampled and later optimally interpolated to obtain the smooth curve that is needed to compute the $f(\alpha)$ singularity spectrum. The proposed method also provides estimates of the generalized dimensions at $D_\infty$ and $D_{-\infty}$ which are almost impossible to obtain from real data with limited number of data samples. From the accurate min-max filter design method of evaluating the Legendre transform between the generalized dimensions $D_q$ and $f(\alpha)$ singularity spectrum, the estimates of these extrema $D_q$ values can now be extracted more easily. The method works well for attractors that exhibit multifractal behavior as well as phase transitions as long as the $D_q$ satisfy equation (5.7) in the range between $q_{\min}$ and $q_{\max}$. The proposed method exhibits improved accuracy over conventional linear interpolation [12].

We have evaluated the $f(\alpha)$ singularity spectrum for unvoiced speech signals from the generalized dimensions $D_q$. The $f(\alpha)$ curves, in addition to the generalized dimension $D_q$ curves, contain all the information about the scaling property of the attractor of an unvoiced speech signal. If we assume that unvoiced speech signals have bifurcations just as in infant cries [20], and from the work of Kakita et al., where vibrations in the vocal folds are related to the folding and twisting in the Henon Map
[16], then given an unvoiced speech signal, we can estimate the generalized dimension $D_q$ using the GCI method described in chapter 2. With the $f(\alpha)$ singularity spectrum evaluated from the $D_q$ curve, it is now possible to obtain a lot of information about the properties of a nonlinear map that gives rise to the evaluated $f(\alpha)$ curve using the proposed method of Feigenbaum et al. [132]. This could lead to simple maps that may be used to model how unvoiced speech signals are produced in the vocal cavity. The resulting maps may be useful in understanding the complexity involved in speech production thereby leading to better nonlinear dynamical modeling of speech. Also, symbolic dynamic representation [133, 134] of unvoiced speech signals from these nonlinear maps may be possible. With this plausible scenario, this dissertation has presented evidence for the potential of multifractal analysis to become an important tool in characterizing and describing unvoiced speech sounds.

The computation of the generalized dimensions $D_q$ is on the order of $\mathcal{O}(N \log N)$ for each $q$ [47]. Therefore, as the range between $q_{\min}$ and $q_{\max}$, defined in chapter 4, is increased and the number of speech sample $N$ gets longer, the estimation of $D_q$ becomes more expensive. But with the advent of faster computers and with better efficient algorithm for finding the nearest neighbors of any arbitrary point on the attractor [61], the computation time required to estimate generalized dimensions $D_q$ will be considerably reduced. With this computational reduction, the generalized dimensions $D_q$ which are different and unique for different speakers may be a useful speaker identification or speaker verification criterion. In the next section, we describe some of the unanswered questions which this dissertation has raised.

### 6.0.1 Future Research

In this dissertation, the multifractal analysis of unvoiced speech signals has been performed on a few sounds in the alphabet of the English language spoken in isolation. This is not the ideal scenario since human being converse in continuously spoken sentences. The multifractal analysis of unvoiced speech signals in this dissertation is only the tip of the iceberg in the rich nonlinear dynamical study of speech signals. Therefore it would be very instructive to analyze unvoiced speech signals that are part of a spoken sentence. Continuous spoken sentences have more intervocalic
structure than alphabets spoken in isolation. Also, in the evaluation of the \( f(\alpha) \) singularity spectrum of unvoiced speech signal from their generalized dimensions \( D_q \), many of the \( D_q \) values were declared invalid because the curves were not monotonic decreasing. It would also be interesting to extrapolate between this invalid \( D_q \) values under a monotonic decrease constraint. This may lead to further reduction in the computation of the generalized dimensions. Finally, we hope to perform multifractal analysis on existing biomechanical model of the vocal tract [21] in order to compare theoretical multifractal results from the biomechanical models to results obtained from multifractal analysis of unvoiced speech signals.

Future modeling of speech as a nonlinear dynamical system could profitably consider entropy. Schuster [2] stress the importance of entropy, especially the Kolmogorov-Sinai entropy \( (K_1) \), as a measure by which chaotic motion in phase space can be characterized. \( K_1 \) is the average rate of increase of disorder in a dissipative dynamical system, and contains information both about the time evolution (as the Lyapunov exponents) and the static structure of the asymptotic orbit (as do the generalized fractal dimensions \( D_q \)). When chaotic orbits are multifractals, there exists generalized entropies \( K_q \) for which \( K_q > K_{q'} \) if \( q < q' \). Implementing the calculation of \( K_q \) poses a number of technical problems, these include stabilization of the \( D_q \) at high embedding dimensions \( d_E \), choosing appropriate metric, and determining that the formulas for \( K_q \) (which are only correct asymptotically) are valid [51]. Unfortunately, there is a paucity of good bench-mark entropy calculations.
Appendix A

Numerical Computation of the Generalized Dimensions $D_q$

From equation (2.29),

$$D_q = \frac{1}{q-1} \lim_{\epsilon \to 0} \frac{1}{\log \epsilon} \log \sum_{i=1}^{N(\epsilon)} p_i^q, \quad -\infty < q < \infty$$  \hspace{1cm} (A.1)

where the sum in equation (A.1) is over all the $N(\epsilon)$ volume elements of size $\epsilon$ needed to cover the attractor as defined in section 2.5.1 [73]. For the special case when $q = 1$, we have from L'Hospital's rule

$$D_1 = \lim_{\epsilon \to 0} \frac{\sum_{i=1}^{N(\epsilon)} p_i \log p_i}{\log \epsilon}$$  \hspace{1cm} (A.2)

which is the information dimension.

A numerical approximation to the generalized correlation integral is evaluated as

$$C_{q,d_B}(\epsilon) \approx \left\{ \frac{1}{M_{seed}} \sum_{i=1}^{M_{seed}} \left\{ \frac{1}{M-1} \sum_{j=1 \atop j \neq i}^{M} \Theta(\epsilon - \| X_i - X_j \|) \right\} \right\}^{q-1} \left( \frac{1}{\epsilon} \right)^{1/(q-1)}$$  \hspace{1cm} (A.3)

where $\Theta$ is the Heaviside unit step function, $\| \cdot \|$ denotes maximum norm, the index $j$ runs over the $(M-1)$-element fractal subset and the index $i$ runs over $M_{seed}$ equally-spaced points which form a subset of the $M$ vector points [60]. For simple implementation, we can now split equation (A.3) into two cases.
Case 1: \( q = 1 \):

Let \( S_{ij} = \sum_{j=1}^{M} \Theta(\epsilon - \|X_i - X_j\|) \) be the total number of inter-point distances on the attractor that are less than the volume element of size \( \epsilon \). Then from equation (A.3),

\[
\log C_{1,dB}(\epsilon) = \frac{1}{M_{seed}} \sum_{i=1}^{M_{seed}} \log \frac{S_{ij}}{M}
\]

\[= \frac{1}{M_{seed}} \sum_{i=1}^{M_{seed}} \log S_{ij} - \frac{1}{M_{seed}} \sum_{i=1}^{M_{seed}} \log M \]

\[= \frac{1}{M_{seed}} \left( \sum_{i=1}^{M_{seed}} \log S_{ij} \right) - \log M \]

\[= < \log S_{ij} > - \log M \]

where \( <> \) defines an average value over the \( M_{seed} \) reference vector points on the attractor.

Case 2: \( q \neq 1 \):

\[
\log C_{q,dB}(\epsilon) = \frac{1}{q-1} \log \left( \frac{\sum_{i=1}^{M_{seed}} \{S_{ij}\}^{q-1}}{M_{seed} M^{q-1}} \right)
\]

\[= \frac{\log \left( \sum_{i=1}^{M_{seed}} \{S_{ij}\}^{q-1} \right) - \log M_{seed}}{q-1} - \log M
\]

These forms of the generalized correlation integrals (equations (A.7) and (A.9)) are necessary to avoid taking the logarithm of zero and division by zero, which would give an undefined values.
Appendix B

Frequency domain implementation of $f(\alpha)$ singularity spectrum

From equations (4.17)-(4.19), the relationship between the generalized dimension $D_q$ and the $f(\alpha)$ singularity spectrum is as follows:

\[ \tau(q) = (q - 1)D_q \quad (B.1) \]
\[ \alpha(q) = \frac{d}{dq} \tau(q) = D_q + (q - 1) \frac{d}{dq} D_q \quad (B.2) \]
\[ f(\alpha) = f(\alpha(q)) = q\alpha(q) - \tau(q). \quad (B.3) \]

Using the Fourier transform duality property that

\[ F\{\dot{x}(q)\} = j2\pi \hat{f} X(f) \quad (B.4) \]

and

\[ F\{-j2\pi q x(q)\} = \dot{X}(f) \quad (B.5) \]

where $x(q)$ and $X(f) = F^1\{x(t)\}$ is a Fourier transform pair, we can now derive the Fourier transforms of the Legendre transform of equation (B.3).

The Fourier transform of $\tau(q)$ in equation (B.1) is given as

\[ \mathcal{T}(\hat{f}) = F\{\tau q\} = -\frac{\hat{D}(\hat{f})}{j2\pi} - D(\hat{f}) \quad (B.6) \]

\[ ^1F \text{ represents the Fourier transform and } f \text{ is equal to } \omega \frac{2\pi}{2\pi}. \]
where $D(\hat{f})$ is the Fourier transform of the generalized dimension $D_q$ and $\mathcal{T}(f)$ is the Fourier transform of $\tau(q) = (q - 1)D_q$. $\dot{D}(\hat{f}) = \frac{d}{d\hat{f}}D(\hat{f})$ is the derivative of $D(f)$ with respect to frequency.

From equation (B.3), the Fourier transform of $f(\alpha(q))$ is now

$$\mathcal{F}(\hat{f}) = -\frac{d}{d\hat{f}}(f\mathcal{T}(\hat{f})) - \mathcal{T}(\hat{f})$$

$$= -\hat{f}\dot{\mathcal{T}}(\hat{f}) - 2\mathcal{T}(\hat{f}). \quad \text{(B.7)}$$

If we now substitute equation (B.6) into (B.8), we obtain

$$\mathcal{F}(\hat{f}) = \frac{\hat{f}}{2\pi} \ddot{D}(\hat{f}) + \dot{D}(\hat{f}) \left( \hat{f} + \frac{1}{j\pi} \right) + 2D(\hat{f}) \quad \text{(B.9)}$$

where $\ddot{D}(\hat{f})$ is the second derivative of $D(\hat{f})$ with respect to frequency.

Equation (B.9) is the Fourier transform equivalent of the Legendre transform relating the generalized dimensions $D_q$ and the $f(\alpha)$ singularity spectrum. In the implementation of the min-max filter design method of accurately estimating the $f(\alpha)$ singularity spectrum from the generalized dimension $D_q$, all the calculations were done in the “$q$” domain. The derivation in this appendix provides a frequency domain implementation which may be useful for evaluating the Legendre transform between the generalized dimensions $D_q$ and the $f(\alpha)$ singularity spectrum.
Appendix C

Plot of the Generalized Correlation Integrals

In this appendix, we plot the log of the generalized correlation integrals $C_{q,d_E}(\epsilon)$ versus $\log \epsilon$ for embedding dimension $d_E$ in the range of 5 to 15. The corresponding generalized dimensions $D_q$ are estimated from the linear regions of the log-log plots.
Figure C.1: (a) Shows the log-log plot of the generalized correlation integrals \( C_{q,d_\varepsilon}(\varepsilon) \) as a function of the volume element \( \varepsilon \) for embedding dimension \( d_\varepsilon \) ranging from 5 to 15. In this plot, \( q = -5 \). The unvoiced speech \"s\" spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcm0) database.
Figure C.2: (a) Shows the log-log plot of the generalized correlation integrals $C_{q,d_E}(\epsilon)$ as a function of the volume element $\epsilon$ for embedding dimension $d_E$ ranging from 5 to 15. In this plot, $q = -4$. The unvoiced speech 's' spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database.
Figure C.3: (a) Shows the log-log plot of the generalized correlation integrals $C_{q,d_E}(\epsilon)$ as a function of the volume element $\epsilon$ for embedding dimension $d_E$ ranging from 5 to 15. In this plot, $q = -3$. The unvoiced speech \textquotedblleft s\textquotedblright\ spoken by a female speaker is taken from the ISOLET (/ISOLET/isoleti/fcmc0) database.
Figure C.4: (a) Shows the log-log plot of the generalized correlation integrals $C_{q,d_E}(\epsilon)$ as a function of the volume element $\epsilon$ for embedding dimension $d_E$ ranging from 5 to 15. In this plot, $q = -2$. The unvoiced speech "s" spoken by a female speaker is taken from the ISOLET (/ISOLET/isole1/fcmc0) database.
Figure C.5: (a) Shows the log-log plot of the generalized correlation integrals $C_{q,d_E}(\epsilon)$ as a function of the volume element $\epsilon$ for embedding dimension $d_E$ ranging from 5 to 15. In this plot, $q = -1$. The unvoiced speech "s" spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database.
Figure C.6: (a) Shows the log-log plot of the generalized correlation integrals $C_{q,d_E}(\epsilon)$ as a function of the volume element $\epsilon$ for embedding dimension $d_E$ ranging from 5 to 15. In this plot, $q = 0$. The unvoiced speech "s" spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database.
Figure C.7: (a) Shows the log-log plot of the generalized correlation integrals $C_{q,d_E}(\epsilon)$ as a function of the volume element $\epsilon$ for embedding dimension $d_E$ ranging from 5 to 15. In this plot, $q = 1$. The unvoiced speech "s" spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmco) database.
Figure C.8: (a) Shows the log-log plot of the generalized correlation integrals $C_{q,d_E}(\epsilon)$ as a function of the volume element $\epsilon$ for embedding dimension $d_E$ ranging from 5 to 15. In this plot, $q = 2$. The unvoiced speech "s" spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmc0) database.
Figure C.9: (a) Shows the log-log plot of the generalized correlation integrals $C_{q, d_E}(\epsilon)$ as a function of the volume element $\epsilon$ for embedding dimension $d_E$ ranging from 5 to 15. In this plot, $q = 3$. The unvoiced speech "s" spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcm0) database.
Figure C.10: (a) Shows the log-log plot of the generalized correlation integrals $C_{q,d_E}(\varepsilon)$ as a function of the volume element $\varepsilon$ for embedding dimension $d_E$ ranging from 5 to 15. In this plot, $q = 4$. The unvoiced speech \"s\" spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcmco) database.
Figure C.11: (a) Shows the log-log plot of the generalized correlation integrals $C_{q,d_E}(\epsilon)$ as a function of the volume element $\epsilon$ for embedding dimension $d_E$ ranging from 5 to 15. In this plot, $q = 5$. The unvoiced speech "s" spoken by a female speaker is taken from the ISOLET (/ISOLET/isolet1/fcm0) database.
Appendix D

Plots of various measures of chaos in unvoiced speech signals

In this appendix, we give a number of examples of multifractal analysis of unvoiced speech signals. All the spoken letters of the English alphabet "s" and "z" were taken from the ISOLET database. The speech data are relatively noise free as discussed in section 5.2. The sampling rate for the speech data in the ISOLET database is 16 kHz. The noise-like portions in the unvoiced speech signals were selected by looking at the speech waveforms. For a particular speaker, labeled as /ISOLET/isolet1/fmbd0, there was no convergence in the correlation integral as the embedding dimension $d_E$ increased. i.e., there exist no plateau region where generalized dimensions $D_q$ can be estimated. For the rest of the speakers analyzed in the dissertation however, multifractal measures were estimated. A summary of the multifractal results for the unvoiced speech sound "s" and "z" that are shown in this appendix are also displayed in Tables 5.3 and 5.4.
Figure D.1: (a) Shows an unvoiced fricative "s" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fcm0) database. (b) shows an isolated, truncated length 4158 sample segment of the unvoiced fricative "s" starting at sample number 3810. Notice the noise-like appearance of the speech signal. \( T = 6.25 \times 10^{-5} \) sec.)
Figure D.1 (continued) Show measures of chaos that were extracted from an unvoiced fricative "s" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fmc0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$ and (f) shows the corresponding $f(\alpha)$ singularity spectrum computed using the min-max interpolation and differentiation filter method.
Figure D.2: (a) Shows an unvoiced fricative \"s\" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fcmg0) database. (b) shows an isolated, truncated length 4520 sample segment of the unvoiced fricative \"s\" starting at sample number 4380. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.2 (continued) Show measures of chaos that were extracted from an unvoiced fricative "s" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fcmg0) database. (c) shows the local slope curves as a function of log ε for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.3: (a) Shows an unvoiced fricative "s" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fDCF0) database. (b) shows an isolated, truncated length 4711 sample segment of the unvoiced fricative "s" starting at sample number 5210. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.3 (continued) Show measures of chaos that were extracted from an unvoiced fricative "s" spoken by a female, taken from the ISOLET (/ISOLET/isolet1/fdcf0) database. (c) shows the local slope curves as a function of \( \log \epsilon \) for embedding dimensions \( d_E \) from 5 to 15. (d) shows the plots of the estimates of the correlation dimension \( D_2 \) as a function of embedding dimension \( d_E \) in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions \( D_q \) versus \( q \). This is an invalid \( D_q \) curve, so \( f(\alpha) \) was not estimated.
Figure D.4: (a) Shows an unvoiced fricative "s" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fec0) database. (b) shows an isolated, truncated length 3352 sample segment of the unvoiced fricative "s" starting at sample number 4100. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.4 (continued) Show measures of chaos that were extracted from an unvoiced fricative "s" spoken by a female, taken from the ISOLET (/ISOLET/isolet1/fec0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.5: (a) Shows an unvoiced fricative “s” spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fet0) database. (b) shows an isolated, truncated length 4990 sample segment of the unvoiced fricative “s” starting at sample number 4323. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.5 (continued) Show measures of chaos that were extracted from an unvoiced fricative "s" spoken by a female, taken from the ISOLET (/ISOLET/isolet1/fet0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.6: (a) Shows an unvoiced fricative "s" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fkh0) database. (b) shows an isolated, truncated length 4300 sample segment of the unvoiced fricative "s" starting at sample number 4318. Notice the noise-like appearance of the speech signal. (T = 6.25 x 10^{-5} sec.)
Figure D.6 (continued) Show measures of chaos that were extracted from an unvoiced fricative \"s\" spoken by a female, taken from the ISOLET (/ISOLET/isolet1/fkh0) database. (c) shows the local slope curves as a function of log $\varepsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.7: (a) Shows an unvoiced fricative "s" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fmb0) database. (b) shows an isolated, truncated length 4934 sample segment of the unvoiced fricative "s" starting at sample number 3879. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.7 (continued) Show measures of chaos that were extracted from an unvoiced fricative "s" spoken by a female, taken from the ISOLET (/ISOLET/isolet1/fmb0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.8: (a) Shows an unvoiced fricative "s" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fmbd0) database. (b) shows an isolated, truncated length 3908 sample segment of the unvoiced fricative "s" starting at sample number 4849. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.8 (continued) (c) No convergence in the generalized correlation integrals, i.e., there exist no plateau region where generalized dimensions $D_q$ can be estimated. The speech sound is the same as in Fig. D.8.
Figure D.9: (a) Shows an unvoiced fricative "s" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fme0) database. (b) shows an isolated, truncated length 4324 sample segment of the unvoiced fricative "s" starting at sample number 4849. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.9 (continued) Show measures of chaos that were extracted from an unvoiced fricative "s" spoken by a female, taken from the ISOLET (/ISOLET/isolet1/fme0) database. (c) shows the local slope curves as a function of \( \log \varepsilon \) for embedding dimensions \( d_E \) from 5 to 15. (d) shows the plots of the estimates of the correlation dimension \( D_2 \) as a function of embedding dimension \( d_E \) in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions \( D_q \) versus \( q \). This is an invalid \( D_q \) curve, so \( f(\alpha) \) was not estimated.
Figure D.10: (a) Shows an unvoiced fricative "s" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjc1) database. (b) shows an isolated, truncated length 2723 sample segment of the unvoiced fricative "s" starting at sample number 4364. Notice the noise-like appearance of the speech signal. (T = 6.25 \times 10^{-5} \text{ sec.})
Figure D.10 (continued) Show measures of chaos that were extracted from unvoiced fricative "s" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjc1) database. (c) shows the local slope curves as a function of $\log \epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$ and (f) shows the corresponding $f(\alpha)$ singularity spectrum computed using the min-max interpolation and differentiation filter method.
Figure D.11: (a) Shows an unvoiced fricative "s" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjfv0) database. (b) shows an isolated, truncated length 4434 sample segment of the unvoiced fricative "s" starting at sample number 3658. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.11 (continued) Show measures of chaos that were extracted from unvoiced fricative \“s\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjfv0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$ and (f) shows the corresponding $f(\alpha)$ singularity spectrum computed using the min-max interpolation and differentiation filter method.
Figure D.12: (a) Shows an unvoiced fricative "s" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjp0) database. (b) shows an isolated, truncated length 4739 sample segment of the unvoiced fricative "s" starting at sample number 4323. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.12 (continued) Show measures of chaos that were extracted from unvoiced fricative "s" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjp0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$ and (f) shows the corresponding $f(\alpha)$ singularity spectrum computed using the min-max interpolation and differentiation filter method.
Figure D.13: (a) Shows an unvoiced fricative "s" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mnjh0) database. (b) shows an isolated, truncated length 7275 sample segment of the unvoiced fricative "s" starting at sample number 4429. Notice the noise-like appearance of the speech signal. \( T = 6.25 \times 10^{-5} \) sec.)
Figure D.13 (continued) Show measures of chaos that were extracted from unvoiced fricative \"s\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mnjh0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$ and (f) shows the corresponding $f(\alpha)$ singularity spectrum computed using the min-max interpolation and differentiation filter method.
Figure D.14: (a) Shows an unvoiced fricative "s" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mnre0) database. (b) shows an isolated, truncated length 4601 sample segment of the unvoiced fricative "s" starting at sample number 4655. Notice the noise-like appearance of the speech signal. \( T = 6.25 \times 10^{-5} \) sec.)
Figure D.14 (continued) Show measures of chaos that were extracted from unvoiced fricative "s" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mnre0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$ and (f) shows the corresponding $f(\alpha)$ singularity spectrum computed using the min-max interpolation and differentiation filter method.
Figure D.15: (a) Shows an unvoiced fricative "s" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mrmh1) database. (b) shows an isolated, truncated length 3713 sample segment of the unvoiced fricative "s" starting at sample number 4628. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.15 (continued) Show measures of chaos that were extracted from unvoiced fricative "s" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mrmhl) database. (c) shows the local slope curves as a function of $\log \epsilon$ for embedding dimensions $d_\epsilon$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_\epsilon$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$ and (f) shows the corresponding $f(\alpha)$ singularity spectrum computed using the min-max interpolation and differentiation filter method.
Figure D.16: (a) Shows an unvoiced fricative \textquotedbl{}s\textquotedbl{} spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mrs0) database. (b) shows an isolated, truncated length 3215 sample segment of the unvoiced fricative \textquotedbl{}s\textquotedbl{} starting at sample number 4460. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.)
Show measures of chaos that were extracted from unvoiced fricative "s" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mrs0) database. (c) shows the local slope curves as a function of log $\varepsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$ and (f) shows the corresponding $f(\alpha)$ singularity spectrum computed using the min-max interpolation and differentiation filter method.
Figure D.17: (a) Shows an unvoiced fricative "s" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/msa0) database. (b) shows an isolated, truncated length 3927 sample segment of the unvoiced fricative "s" starting at sample number 4133. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.17 (continued) Show measures of chaos that were extracted from unvoiced fricative "s" spoken by a male speaker, taken from the ISOLET/ISOLET/isolet1/msa0 database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.18: (a) Shows an unvoiced fricative \"s\" \space spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mtdw0) database. (b) shows an isolated, truncated length 7247 sample segment of the unvoiced fricative \"s\" \space starting at sample number 4572. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5}\) sec.)
Figure D.18 (continued) Show measures of chaos that were extracted from unvoiced fricative "s" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mtdwc0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.19: (a) Shows voiced fricative \"z\" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fcmc0) database. (b) shows an isolated, truncated length 9587 sample segment of the voiced fricative \"z\" starting at sample number 1552. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5} \text{ sec.}\)
Figure D.19 (continued) Show measures of chaos that were extracted from voiced fricative \"z\" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fcmco) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.20: (a) Shows voiced fricative \"z\" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fcmg0) database. (b) shows an isolated, truncated length 4050 sample segment of the voiced fricative \"z\" starting at sample number 3950. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5} \text{ sec.}\)
Figure D.20 (continued) Show measures of chaos that were extracted from voiced fricative "z" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fcng0) database. (c) shows the local slope curves as a function of \( \log \epsilon \) for embedding dimensions \( d_E \) from 5 to 15. (d) shows the plots of the estimates of the correlation dimension \( D_2 \) as a function of embedding dimension \( d_E \) in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions \( D_q \) versus \( q \). This is an invalid \( D_q \) curve, so \( f(\alpha) \) was not estimated.
Figure D.21: (a) Shows voiced fricative \"z\" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fdcf0) database. (b) shows an isolated, truncated length 7667 sample segment of the voiced fricative \"z\" starting at sample number 785. Notice the noise-like appearance of the speech signal. \(T = 6.25 \times 10^{-5}\) sec.)
Figure D.21 (continued) Show measures of chaos that were extracted from voiced fricative \"z\" spoken by a female, taken from the ISOLET (/ISOLET/isolet1/fdca0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$ and (f) shows the corresponding $f(\alpha)$ singularity spectrum computed using the min-max interpolation and differentiation filter method.
Figure D.22: (a) Shows voiced fricative "z" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fec0) database. (b) shows an isolated, truncated length 7184 sample segment of the voiced fricative "z" starting at sample number 1316. Notice the noise-like appearance of the speech signal. (\(T = 6.25 \times 10^{-5}\) sec.)
Figure D.22 (continued) Show measures of chaos that were extracted from voiced fricative "/z/" spoken by a female, taken from the ISOLET (/ISOLET/isolet1/fec0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.23: (a) Shows voiced fricative "z" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fet0) database. (b) shows an isolated, truncated length 7400 sample segment of the voiced fricative "z" starting at sample number 2882. Notice the noise-like appearance of the speech signal. \( T = 6.25 \times 10^{-5} \text{ sec.} \)
Figure D.23 (continued) Show measures of chaos that were extracted from voiced fricative \"z\" spoken by a female, taken from the ISOLET (/ISOLET/isolet1/fet0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.24: (a) Shows voiced fricative \textquoteleft \textquoteleft z\textquoteright \textquoteleft \textquoteleft spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fews0) database. (b) shows an isolated, truncated length 7677 sample segment of the voiced fricative \textquoteleft \textquoteleft z\textquoteright \textquoteleft \textquoteleft starting at sample number 1662. Notice the noise-like appearance of the speech signal. \((T = 6.25 \times 10^{-5} \text{ sec.})\)
Figure D.24 (continued) Show measures of chaos that were extracted from voiced fricative \"z\" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fews0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$ and (f) shows the corresponding $f(\alpha)$ singularity spectrum computed using the min-max interpolation and differentiation filter method.
Figure D.25: (a) Shows voiced fricative "z" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fkh0) database. (b) shows an isolated, truncated length 8592 sample segment of the voiced fricative "z" starting at sample number 1468. Notice the noise-like appearance of the speech signal. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.25 (continued) Show measures of chaos that were extracted from voiced fricative \"z\" spoken by a female, taken from the ISOLET (/ISOLET/isolet1/fkh0) database. (c) shows the local slope curves as a function of $\log \epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.26: (a) Shows voiced fricative \”z\” spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fcm0) database. (b) shows an isolated, truncated length 7843 sample segment of the voiced fricative \”z\” starting at sample number 1468. Notice the noise-like appearance of the speech signal. \( T = 6.25 \times 10^{-5} \) sec.)
Figure D.26 (continued) Show measures of chaos that were extracted from voiced fricative "z" spoken by a female speaker, taken from the ISOLET (/ISOLET/isolet1/fmc0) database. (c) shows the local slope curves as a function of log ε for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$ and (f) shows the corresponding $f(\alpha)$ singularity spectrum computed using the min-max interpolation and differentiation filter method.
Figure D.27: (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjc1) database. (b) shows an isolated, truncated 5404 length sample segment of the voiced fricative \"z\" starting at sample number 2182. \(T = 6.25 \times 10^{-5}\) sec.)
Figure D.27 (continued) Show measures of chaos that were extracted from voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET ([/ISOLET/isolet1/mjc1]) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. 

249
Figure D.28: (a) Shows voiced fricative \textquoteleft\textquoteleft z\textquoteright\textquoteright spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjfv0) database. (b) shows an isolated, truncated 3210 length sample segment of the voiced fricative \textquoteleft\textquoteleft z\textquoteright\textquoteright starting at sample number 2665. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.28 (continued) Show measures of chaos that were extracted from voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjfv0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.29: (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isoleti/mjp0) database. (b) shows an isolated, truncated 8092 length sample segment of the voiced fricative \"z\" starting at sample number 1358. \(T = 6.25 \times 10^{-5} \text{ sec.}\)
Figure D.29 (continued) Show measures of chaos that were extracted from voiced fricative "z" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjp0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.30: (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isole1/mjrs0) database. (b) shows an isolated, truncated 8591 length sample segment of the voiced fricative \"z\" starting at sample number 1108. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.30 (continued) Show measures of chaos that were extracted from voiced fricative "z" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mjrs0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.31: (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mnjh0) database. (b) shows an isolated, truncated 8601 sample length segment of the voiced fricative \"z\" starting at sample number 1544. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.31 (continued) Show measures of chaos that were extracted from voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mnjh0) database. (c) shows the local slope curves as a function of $\log \epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. 

257
Figure D.32: (a) Shows voiced fricative \textquotedbl\textquotedblright z\textquotedbl\textquotedblright\ spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mnre0) database. (b) shows an isolated, truncated 6758 length sample segment of the voiced fricative \textquotedbl\textquotedblright z\textquotedbl\textquotedblright\ starting at sample number 1690. \(T = 6.25 \times 10^{-5} \text{ sec.}\)
Figure D.32 (continued) Show measures of chaos that were extracted from voiced fricative "z" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mnre0) database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.33: (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mrmh1) database. (b) shows an isolated, truncated 5705 length sample segment of the voiced fricative \"z\" starting at sample number 1616. (\(T = 6.25 \times 10^{-5}\) sec.)
Figure D.33 (continued) Show measures of chaos that were extracted from voiced fricative "z" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mrmh1) database. (c) shows the local slope curves as a function of log $\varepsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$ and (f) shows the corresponding $f(\alpha)$ singularity spectrum computed using the min-max interpolation and differentiation filter method.
Figure D.34: (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mrs0) database. (b) shows an isolated, truncated 7538 length sample segment of the voiced fricative \"z\" starting at sample number 1524. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.34 (continued) Show measures of chaos that were extracted from voiced fricative "z" spoken by a male speaker, taken from the ISOLET /ISOLET/isolet1/mrs0 database. (c) shows the local slope curves as a function of log$\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.35: (a) Shows voiced fricative \textit{"z"} spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/msa0) database. (b) shows an isolated, truncated 7205 length sample segment of the voiced fricative \textit{"z"} starting at sample number 1718. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.35 (continued) Show measures of chaos that were extracted from voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET /ISOLET/isolet1/msa0 database. (c) shows the local slope curves as a function of log ε for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$. This is an invalid $D_q$ curve, so $f(\alpha)$ was not estimated.
Figure D.36: (a) Shows voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET (/ISOLET/isolet1/mtdw0) database. (b) shows an isolated, truncated 10359 length sample segment of the voiced fricative \"z\" starting at sample number 1385. ($T = 6.25 \times 10^{-5}$ sec.)
Figure D.36 (continued) Show measures of chaos that were extracted from voiced fricative \"z\" spoken by a male speaker, taken from the ISOLET/ISOLET/isolet1/mtdw0 database. (c) shows the local slope curves as a function of log $\epsilon$ for embedding dimensions $d_E$ from 5 to 15. (d) shows the plots of the estimates of the correlation dimension $D_2$ as a function of embedding dimension $d_E$ in the soft turbulence regions marked by the vertical dashed lines in (c). (e) shows the estimated generalized dimensions $D_q$ versus $q$ and (f) shows the corresponding $f(\alpha)$ singularity spectrum computed using the min-max interpolation and differentiation filter method.
References


271


Bibliography


