Chapter 3-CURRENT TECHNIQUES IN SPEECH ANALYSIS

This chapter gives a detail account of various linear and nonlinear techniques used for speech analysis. Special emphasis is given to methods used for characterizing pathological conditions like vocal disorders and stuttering in speech process. The vocal and voice diseases should to be diagnosed in the early stages. It is well known that most of these diseases cause changes in the acoustic voice signal. Therefore, the voice signal can be a useful tool to diagnose them, In the bibliography, there are many algorithms to calculate acoustic parameters and it was shown that there is a great correlation between these parameters and the pathologies. For example, it was demonstrated that some acoustic parameters are deviated from the mean in the case of pathological voices.

Furthermore, the diagnosis of pathologies from the voice signal presents several advantages with regard to other methods:

1. It is a non-invasive tool and easy to use, the patient only has to speak.
2. It makes possible to build an automatic computer based diagnosis system.
3. The diagnostic is objective because it is based on the value of acoustic parameters.
4. It can also be used for the evaluation of surgical and pharmacological treatments and rehabilitation processes
3.1 LINEAR TECHNIQUES

The analysis of voice signal is usually performed by the extraction of acoustic parameters using digital signal processing techniques. After that, these parameters are analyzed to determine the characteristic of the voice: nonpathologic, pathologic and the type of pathology.

In the bibliography, there are a great number of acoustic parameters that can be extracted and analyzed, however it is not absolutely clear its usefulness for solving the problem. The selection of the most appropriate parameters is still an open problem.

NEEDS FOR DATABASE

There is a need for at least three types of databases, which should be recorded under standard acoustic conditions, on standard linguistic material within a language and comparable linguistic material across languages, on standard equipment, under standard procedures, with standard biographical documentation of the speakers involved. The first is a database of voice samples of patients known to be suffering from specific voice disorders, both pathological and functional. This database should include graded samples of a wide range of disorders at different degrees of severity, from men, women, and children. The second is a database of speakers suspected to be suffering from laryngeal pathology or dysfunction, recorded before diagnosis, with subsequent information available about the diagnoses. The third is a database of a control group of speakers, matched in sex, age, socioeconomic status, general physique, and general state of health to the speakers of the first two databases. This control database should be divided into two halves: smokers and nonsmokers. Neither group of control speakers should have a reported history of voice disorders.

A major part of the study of perturbation has examined the speech waveforms of young, healthy adults (usually university students), together with the phonation of speakers with abnormal laryngeal conditions. In the latter group, the speakers are often older than the healthy speakers, and predominantly male in sex. For a control group, data should be collected from a wider age and sex range of speakers to provide more representative
normative information. For the pathological group, data should be collected from as wide a range of disorders as possible. The voice samples in the databases should include not only the production of sustained vowels, but also a sample of continuous speech.

MAJOR ISSUES
The major issue in perturbation is determining the correlation between the acoustic characteristics of perturbed phonation and their anatomical, physiological, and aerodynamic counterparts. Once a better understanding is achieved of the mechanical consequences for phonation of different types of disruption of laminar tissue relationships, the search for an automatic system that is capable of providing diagnostic support to clinicians will be able to proceed on sounder footing. An important part of this development will be the optimization of the choice of perturbation parameters for different disorder types. Practically, a crucial objective is the development of a microprocessor-based version of perturbation analysis for low-cost use in otolaryngology clinics and speech pathology clinics, to provide quantified support for remedial and monitoring functions. In the meantime, progress in this area is likely to be focused on correlational studies of the connection between types and degrees of perturbation and the typology of laryngeal disorder that a combination of acoustic analysis, fiberoptic laryngoscopic inspection and histological examination can show. The potential usefulness of perturbation measures for assessing laryngeal pathology seems beyond doubt. The nature of the approaches to perturbation that will be the most profitable for patients suffering from laryngeal pathology remains to be established.

Scientist from various fields has developed a variety of methods and tools for the detection of normal as well as characteristic features in various vocal pathologies. This has provided the physician and speech therapist with several tradition tools for acoustic analysis. Various cases of acoustic analysis of normal and pathological voices are reported [1-4]. Acoustic methods have the potential to provide quantitative techniques for clinical assessment of laryngeal and vocal tract function. Though several methods like laryngoscopy, glottography, electromyography, stroboscopy and acoustic analysis [5] currently exists for laryngeal and vocal tract research, acoustic analysis have added
advantage over other methods because of its nonintrusive nature. Diagnosis of voice pathologies is mainly done using either subjective technique like evaluation of voice quality by the clinician or invasive methods like laryngeoscopical techniques. Several quantitative measures of voice quality assessment [6 - 9] are proposed in the recent years which help in the documentation of evolution of the pathological condition. Such measures can prove to be useful for application in fields like preventive medicine and telemedicine.

Acoustic tools are commonly used by speech clinicians, such as surgeons and speech therapists for recording changes in acoustic pressure at the lips or inside the vocal tract. These tools [10], amongst others, can provide potentially objective measures of voice function. Although acoustic examination is only one tool in the complete assessment of voice function, such objective measurement has many practical uses in clinical settings, augmenting the subjective judgement of voice function by clinicians. These measures find uses, for example, in the evaluation of surgical procedures, therapy, differential diagnosis and screening [10, 11], and often augment subjective voice quality measurements, for example the GRB (Grade, Roughness and Breathiness) scale [12]. These objective measures can be used to portray a "hoarseness" diagram for clinical applications [13], and there also exists a variety of techniques for automatically screening for voice disorders using these measures [14 - 16].

Traditionally voice signals has been modeled as a linear process and therefore, most of the tools are based on linear system theory. Most conventional linear time series analysis methods [17, 18] implicitly assume that the data come from a linear dynamical system, perhaps with many degrees of freedom and some added noise. Thus the variation is assumed to be a superposition of sine waves or exponentials that grow or decay in time. Most commonly used linear methods to characterise the system dynamics are autocorrelations, Fourier analysis and power spectrum representation. For stationary data with inherent periodicities, Fourier analysis [19] turned out to be extremely useful and this lead to the development of signal processing era in all experimental data. Signal processing continued to gain importance with the growth of electronic industry and
became extremely useful with the invention of Fast Fourier Transform computer program [20]. Spectral analysis saw another fantastic leap with the introduction of wavelets in the mid 1980s [21]. With the invention of information theory by Shannon and Weaver [21] time series could be understood in terms of symbolic dynamics.

In short linear methods interpret all regular structure in a data set as a dominant frequency, as linear correlations. This means that the intrinsic dynamics of the system are governed by the linear paradigm that small causes lead to small effects. Since linear equations can only lead to exponentially growing or periodically oscillating solutions, all irregular behaviour of the system has to be attributed to some random external input to the system. [22].

The most important vocal acoustic parameters for clinical use are measurements of acoustic spectrography, fundamental frequency (F0), Harmonic-to-noise ratio (HNR), vocal extension profile, and perturbation index - jitter and shimmer [23], zero crossing rate, wavelets.

1. Spectrogram

According to [24] fundamental frequency is determined physiologically by the number of cycles that the vocal folds make in a second, and they are the natural result of the length of these structures. During sound production the characteristic features of sound are produced by the interaction and complex interplay of frequencies coming from the vocal tract and the nasal tract.

The common acoustic method for speech analysis is the sound spectrogram which gives a graphical display of the time varying spectral characteristics of the speech signal [25, 26]. The spectrogram is fundamental to the analysis of speech sounds [27, 28]. The particular arrangement of the frequency components in a phoneme is a strong indicator of the associated phonetic category [27]. Under certain restrictions, similar and related analysis of speech sounds produced by patients can be a valuable aid to the diagnosis and progress monitoring in the course of medical treatment for voice disorders [28].
produces a two-dimensional pattern called a **spectrogram** in which the vertical dimension corresponds to frequency and the horizontal dimension to time. The darkness of the pattern is proportional to signal energy. Therefore, the resonance frequencies of the vocal tract show up as dark bands in the spectrogram.

Voiced regions are characterized by a striated appearance due to the periodicity of the time waveform, while unvoiced intervals are more solidly filled in. The spectrogram is basically a series of power spectra for small, consecutive time sections of the speech signal. The width of the time window for the Fourier transform has a great effect on the characteristics of the spectrogram produced. A long time window produces a narrow band-pass filter which allows the harmonic structure to be seen in spite of blurring the time definition. The spectrogram can be used to identify individual segments of speech, such as phonemes, each of which has its own spectral structure that can become clear to the trained eye. Furthermore the spectrogram can also be used to give information on the frequency of the glottal closure of the signal seen on the wide-band spectrogram as the reciprocal of the time period seen between the vertical lines. The spectrogram can also be used to identify the formant frequencies which are the dominant frequencies in the frequency spectrum.

### 3.2 PERTURBATION ANALYSIS:

The attempt to discover objective acoustic and physiological ways of characterizing laryngeal waveforms in terms of perturbation parameters has a history of over 30 years, beginning with the pioneering work of researchers such as Moore and von Leden [29], Moore [29], Lieberman [30], and Michel [31]. From this base, the topic of perturbation has attracted many contributions from speech science, signal processing, laryngology, and speech pathology [32 - 63]

A somewhat comparable approach is taken by Ludlow et al. [64], where perturbation is represented as a deviation from a locally calculated trend in terms of the difference between a given period and the average of the periods two cycles to the left and right. Alternative characterizations of perturbation have tended to focus on the relationship...
between adjacent phonatory cycles; for instance, in terms of how often the difference in value of adjacent cycles tends to change its algebraic sign (e.g., Hecker and Kreul's directional perturbation factor) [65]. Another approach is Horii's [47] jitter ratio, which calculates the average magnitude of the differences between adjacent periods. A more computationally intensive approach is to examine the serial correlation between periods as done by Baken [66]; Iwata and von Leden [67] and Iwata [68] explore the interesting possibility that different types of serial correlation may discriminate between different types of pathology.

The analysis of period and amplitude values from continuous speech is more difficult than from sustained monotone vowels because the detecting algorithm must examine a variety of signal structures produced by interactive segmental effects of the dynamic movements of the articulators, together with multiple voicing onset and offsets [69]. It is probably also thereby more subject to the risk of artifactual distortion. If continuous speech data are used for perturbation analysis, then a minimum duration of speech material is required to stabilize long-term measurements of perturbation (e.g., the mean and standard deviation of each perturbatory parameter). In an experiment reported by Hiller [70], it was found that a 40-second sample of read speech provided relatively stable long-term speaker-characterizing parameters of perturbation for healthy male and female speakers. This finding is in general agreement with the results of previous studies of the long-term features of the voice (in particular for long-term characteristics of fundamental frequency) studies described by Hiller [70] and Baken [66] provide the technical basis for the discussion that follows on the applications of automatic systems for quantifying the perturbatory characteristics of phonation. Emphasis in the discussion will be placed on noninvasive acoustic methods, but some comment will also be made on selected physiological methods.

**Jitter** refers to a short term(cycle-to-cycle) perturbation in fundamental frequency of voice. Acoustically, perturbation can characterize the laryngeal waveform in both the time domain and the frequency domain. In the time domain, dysperiodic perturbation of the cycle-to-cycle variation in fundamental frequency is called
jitter. Some of the early investigators (e.g., Lieberman [30], 1961) [69] displayed speech waveforms oscillographically and saw that no two periods were exactly alike. Perturbation can acoustically characterize the laryngeal waveform in both the time domain and the frequency domain. [69]. The fundamental frequency appeared jittery; hence, the term jitter. Jitter is affected mainly because of lack of control of vocal fold vibration, which are related with presence of noise at emission and breathiness.

Small and apparently random perturbations of the vocal waveform's period and amplitude (referred to as jitter and shimmer, respectively) have received a great deal of attention [69]. John Laver et. al. [69] suggests that automatic extraction methods make assessment of vocal perturbation a potential part of routine clinical assessments [69]. Careful attention will have to be paid to several technological issues, particularly to the quality of recorded voice samples used for this kind of analysis. Better databases must be developed, and the underlying voice physiology requires clarification.

Haydee et. al [71] carried out analyses of vocal characteristics related to intensity and fundamental frequency and their perturbations indices - jitter and shimmer, in children with phonological disorder [71]. The Computer Speech Lab was used to record and perform acoustic analyses of the vowels /a/, /e/, /i/, through the vocal parameters: fundamental frequency, intensity, jitter and shimmer. The results found were that frequency F0 for vowel /e/ was smaller, on an average, in the phonological disorder group and it was higher Hz in the Control group. No differences between the groups were found regarding the averages of jitter and shimmer. The results found in the present study point to the fact that these children with phonological disorders compared to children without disorders do not present any abnormality that affects the vocal folds, either muscle or neural activity involved with phonation, either lesions that may cause increase in aperiodicity of vocal fold vibration, which reflect the increased values of jitter [71].
The study also indicated that the characteristics such as reduction of glottic resistance, vocal fold mass lesions and greater noise at production, factors that could lead to affections of shimmer values, were not necessarily found.

**Fundamental frequency FO and Harmonic to noise ratio**

The spectral shape and level of the noise in the voice signals was estimated with a cepstrum-based technique described by De Krom for laryngeal voices [72]. Various features of a voice spectrum are disentangled in the corresponding cepstrum (the inverse Fourier transform of the log spectrum) and can then be manipulated separately. Overall spectral shape is represented at the low end of cepstrum. The harmonic structure gives rise to a few equidistant peaks in the cepstrum. The noise contributes to various cepstral parts: the envelope to the low end and the fine structure mainly to the higher regions. By application of a comb filter to the cepstrum, energy related to the harmonic structure in the spectrum can be removed. After a reverse transformation and a level correction, an estimated noise spectrum can be obtained. An example is an acoustical tracheoesophageal (TE)-voice spectrum and its estimated noise spectrum are shown Fig. 3.1. From these spectra the harmonics-to-noise ratio is calculated in the regions of first (HNR $F_1$) and second formant (HNR $F_2$), by subtraction of noise and signal intensity below 700 Hz and between 700 and 2300 Hz, respectively.

![Fig. 3.1 Spectrum of a TE voice (thin, dotted line) with its estimated noise spectrum (heavy line). The lower harmonics are clearly visible and show the periodicity of the](image-url)
voice but for increasing frequency the level of the noise increases and above 1 kHz the
noise spectral density is larger than the level of the harmonics.

When measuring fundamental frequency in signals with a low harmonics-to-noise ratio
the result may become very unstable due to random fluctuations of the noise. To reduce
the influence of noise the calculations were conducted in the spectral domain: not only
the fundamental but also the positions of the higher harmonics were used by calculating
the largest common divider of their frequencies.

This method was first described by Schroeder (1968) [72], applied by Festen et al. (1996)
[72], and is illustrated in Fig. 3.2. The spectrum of the signal was added to versions of the
same spectrum compressed along the frequency axis by small integer numbers. Compression by a factor of two brings the second harmonic at the $F_0$-position, compression by a factor of three the third harmonic, etc. The sum of these compressed spectra is called harmonic product spectrum and it generally shows a clear peak at the fundamental frequency, even for noisy signals. The vocal intensity is determined as the sound pressure level in dB.

![Fig. 3.2 The lower trace shows an arbitrary harmonic voice spectrum. The middle and upper trace show the same spectrum but compressed by factors two and three, respectively. This compression of the frequency scale brings successive higher harmonics](image)

32
at the position of F0. An addition of these spectra gives a strong peak at the fundamental frequency because here peaks in the various spectra coincide.

**Pitch Amplitude and coefficient of excess**

Two parameters used to determine the voice signal noise quantity are the deterministic harmonic to noise ratio (HNR) and the coefficient of excess (EX) that evaluate the noise from a statistical point of view [73]. The latter measures differences in shape of the distribution of the residue signals.

*Pitch Amplitude* gives a normalized measure of the amplitude of the pitch period peak of the residue signal autocorrelation function. This parameter is high for signals with clearly defined pitch period such as voiced sounds (vowels). For breathy vowels of pathological speakers, the PA is low because the signals have weak periodicity [73].

Parameters like jitter and shimmer evaluate perturbation or noise content in the voice signal.

**Shimmer**

Shimmer refers to the variation of the amplitude of sound wave, or intensity of vocal emission and it is the Perturbation of the cycle-to-cycle amplitude of successive laryngeal pulses [69]. Shimmer is affected mainly with reduction of glottic resistance and mass lesions in the vocal folds, which are related with presence of noise at emission and breathiness [69]. In the frequency domain, a frictional element can add spectral noise to the laryngeal waveform. Jitter and shimmer contribute to the rough perceptual effect usually called harshness in acoustory perspective. Spectral noise alone is associated with the auditory effect called whisperiness, and voices showing this effect can be called whispery voices. When spectral noise is added to jitter and/or shimmer, this has the auditory effect of adding whisperiness to harshness, and the composite quality is usually called hoarseness, or hoarse voice. Physiologically, the frictional element in whispery and hoarse voices is caused by incomplete glottal closure [69]. The incomplete closure can be the result either of individual, habitual adjustments of the phonatory muscle system, or of mechanical intrusion into the glottis of obstructions such as vocal nodules, polyps, and
other types of growths, or of paralysis of one or both vocal folds. Severe perturbations are almost always signs of either pathological or functional disorder, but slight perturbations are evident in all speaking voices, especially at the borders of voicing episodes. An important consideration in the ability of acoustic techniques to register perturbation data is therefore the sensitivity of the acoustic technique itself.

[69] showed that a simple bivariate plot for a directional perturbation factor in a shimmer parameter against mean fundamental frequency for a control group of 63 male speakers versus a group of 55 male speakers known to suffer from a variety of laryngeal pathologies. The results for the control group are expressed partly in terms of an ellipse surrounding two standard deviations of the data on the two parameters. The boundary of the ellipse can be used as a practical screening threshold for the detection of vocal pathology. This can be seen from the fact that only 9.5% of the control speakers fall outside the ellipse, and therefore register as false positives, whereas 90.1% of the known pathological speakers fall outside and are therefore successfully detected as probable pathological [69]. For a similar treatment of a group of 54 pathological female speakers versus a control group of 58 female speakers presumed healthy, 79.6% of the pathological speakers were successfully detected by this method, at a cost of 10.3% false positives.

Studies [74] used to analyze sustained vowels generated by patients before and after surgical excision of vocal polyps by Yu Zhanga and Clancy McGilligan showed that jitter shows a significant post surgical decrease. Uloza [75] found that jitter and shimmer are useful for studying the effects of surgery on vocal function; however, Zeitels et al. found that there was not a significant difference in jitter measurements after surgical excision of vocal fold lesions, although there was a significant postsurgical decrease in shimmer measurements. Methodological issues have been raised regarding jitter and shimmer because of the sensitivity of these two parameters to variations in recording systems, analysis systems, and extraction algorithms [75]. In addition, systems employed for these perturbation parameters cannot reliably analyze strongly aperiodic signals [76, 77]. It has been suggested that jitter and shimmer measurements are reliable for
voice analysis only when their values are less than 5% [78, 76] and may be unreliable when large instabilities in voice waveforms are observed [77]. Since jitter and shimmer only represent reliable parameters for nearly periodic voice signals under small perturbation conditions, seeking complementary objective measures capable of analyzing aperiodic voices and assessing the effects of surgery is important.

On the other hand, Zeitels et al. [79] found that there was not a significant difference in jitter measurements after excision of benign vocal-fold lesions, although there was a significant decrease in shimmer measurements. The study of Zeitels et al. and other studies suggest that these perturbation measures should be applied with caution, particularly when a periodic or chaotic voice signals are being analyzed. A similar conclusion was arrived at via their recent experiments with excised larynges [80]. Jitter and shimmer are recognized as inadequate for analysis of far from-periodic voices. The voices from patients showed a statistically significant postsurgical decrease in jitter, but not in shimmer.

[81] studied vocal jitter and shimmer in stuttering. The purpose of this study was to test for the presence of significantly different jitter and shimmer during vowel production in the fluent phonations of stutterers compared to nonstuttering matched controls.

Vocal jitter and shimmer measures of the fluent phonations of 14 stutterers, 12 male and two female, were compared with jitter and shimmer measures of a group of nonstutterers matched for age and sex. Each subject phonated four vowels nine times in random order. Each phonation was sustained for at least 5 sec and was tape-recorded. The mid-3-set portion of each recorded vowel phonation was subjected to jitter and shimmer analyses. Measures for stutterers were larger in both instances. Significant differences between stutterers and nonstutterers were obtained for shimmer measures. Differences on jitter measures were not significant. High variability in the stuttering group accounted for the nonsignificant finding in jitter measures and, in general, indicated heterogeneity among the stutterers. Findings led to the tentative conclusion that in fluent sustained phonation,
stutterers demonstrate less stable control of respiratory-laryngeal dynamics than nonstutterers.

Vocal jitter and shimmer are cycle-to-cycle variations in frequency and amplitude, respectively, of the quasiperiodic glottal tone. Jitter and shimmer are acoustic measures of these vocal perturbations obtained from sustained vowel phonations, which have proved useful in describing voice characteristics of normal and pathologic speakers (Deal and Emanuel, [82] Horii, [82 - 85] Smith, et al., [86]. Lower indices of magnitude on either jitter or shimmer indicate less vocal perturbation and greater stability in the larynx motor control of phonatory behavior. If the magnitude of vocal perturbation, either jitter or shimmer, in the fluent phonatory behaviors of stutterers was shown to be significantly greater than that of nonstutterers, this would provide additional support to the hypothesis that stutterers may demonstrate generally less competent neurophysiologic regulatory control over their peripheral mechanisms of phonation and respiration [87].

The jitter and shimmer analysis of the recordings was performed using a computer program SEARP developed by Horii [88, 83, 84]. The mid-3-see segments of accelerometric signals of vowel phonations recorded on magnetic tape were subjected to the SEARP analysis to obtain the jitter and shimmer data. The program used a peak-picking method to identify individual periods. Jitter is reported in percent. Jitter in percent was obtained by averaging absolute differences between periods in milliseconds from one cycle to the next, dividing the result by the average period and multiplying this value by 100. Shimmer in dB was defined as the average decibel difference between peak amplitudes of consecutive cycles.

A significant difference was found between nonstutterers and stutterers on measures of shimmer. However, differences in jitter between the two groups were not found to be statistically significant. This finding was unexpected in that Horii (1980) [84] had reported a significant, though modest, correlation between jitter and shimmer (r = 0.47, p < 0.001). Because of the supposed relation between the two phenomena, it was expected that either significant differences between stutterers and nonstutterers would be found for
both jitter and shimmer measures, or that no significant difference would be found for either measure. Means of both jitter and shimmer measures of the stutterers were larger than those of the nonstutterers, indicating that the sustained phonations of the stutterers were less stable than those of the nonstutterers in terms of both vocal frequency and intensity. The significant difference between stutterers and nonstutterers with respect to shimmer in dB and the differences observed with regard to jitter, coupled with the relation between the two phenomena, led to the probable conclusion that steady-state phonations of stutterers are different from those of nonstutterers. The direction of this difference suggests that stutterers have less stable neuromuscular control over the events regulating the aerodynamics of the laryngeal and respiratory systems during sustained fluent vowel articulations than nonstutterers [81]. Steady-state, sustained phonation involves an even maintenance of such forces as vocal fold tension, mass, length, and subglottic pressures, while it also maintains the supralaryngeal articulatory adjustments required for production of the vowel. Differences observed between the two groups suggest that nonstutterers are better able to control these forces than stutterers. The large standard deviation of jitter measures of stutterers accounts for the lack of significance between the two groups. In general, the higher variability revealed by the standard deviations of both jitter and shimmer measures of the stutterers suggests a greater heterogeneity in this population.

In order to arrive at a conclusive result this type of study on larger homogeneous samples of stuttering speakers are to be carried out.

Mean jitter measures of each of the four vowels were very similar within groups. This was also true for shimmer measures. Given that main effects for vowels and group interaction effects with vowels were not significant, it would seem appropriate to suggest that future studies need not include production of a variety of vowels for analysis. This is desirable, because the amount of effort and time required to obtain jitter and shimmer measures is considerable. This finding also indicates that the significant shimmer differences were due primarily to generalized factors underlying the
mechanisms of phonation and are relatively independent of the accompanying articulatory adjustments necessary to produce the English vowels /i a u/.

**Zero Crossing Rate**

A zero crossing is a point where an audio waveform crosses the zero-line from negative to positive. Zero crossing graph displays the count of zero crossing over a brief time interval. The oscillation around the zero line is a measure of high frequency sound context. The higher the zero crossing value, the greater the higher frequency component in the audio-date. Fricatives generally have a high zero-crossing value [89]. Zero Crossing Rate (ZCR) is defined as the number of time-domain zero-crossings within a defined region of signal, divided by the number of samples of that region.

The zero-crossing rate (ZCR) is obtained by counting the number of times the signal changes sign during one frame interval as given in the equation below:

$$ZCR = \frac{1}{N} \sum_{n=1}^{N-1} |Sgn(x(n)) - Sgn(x(n-1))| w(n-t)$$

ZCR gives a rough estimate of the frequency content of the speech signals, i.e. a sinusoidal signal of frequency F gives an average ZCR of 2F s\(^{-1}\). ZCR has been used by Rabiner [90] to improve the end-point detection algorithm. It has also been used to distinguish between voiced and unvoiced sounds, i.e. mean ZCR for voiced sounds is less than 1500 s\(^{-1}\) while for unvoiced sound it is more than 5000 s\(^{-1}\) [90].

Speech signals are broadband signals and the interpretation of average zero-crossing rate is therefore much less precise. However, rough estimates of spectral properties can be obtained using a representation based on the short-time average.

The ZCR of the time domain waveform is one of the most indicative and robust measures to discern voiced speech. It has widely used in practice as a strong measure to discern fricatives from voiced speech [89]. The ZCR is simply the count of crossing the zero throw fixed window size. It is said to occur if successive samples have different algebraic signs.
Zero crossing graph displays the count of zero crossing over a brief time interval. The oscillation around the zero line is a measure of high frequency sound context. The higher the zero crossing value, the greater the higher frequency component in the audio-dates. Fricatives generally have a high zero-crossing value. The zero crossing rate of speech signals is a parameter for many speech recognition purposes. [90] studied the speech signal as an analytic signal taking its zero crossing rate into consideration [90]. The zero crossing rate is one of the characteristic parameters in speech signal analysis. Varada. S and Sankar. R proposed a method for finding the end point of the speech signal using zero crossing rate and energy [92]. Digit recognition using energy envelope and zero crossing rate is proposed by [93]. [94] proposed a new feature extraction method based on zero crossings with peak amplitudes for robust speech recognition in noisy environments [94]. [95] present the result of experimental investigations that provides an interesting perspective on the relative importance of zero crossing interval for speech perception [95].

The zero crossing parameters of the speech signal are important for speech modeling and recognition purposes. Zero crossing rate is a reliable parameter for many speech recognition purposes. But the variation of zero crossing intervals of the speech signal and its uses in parametric estimation for recognition purpose is not yet carried out. Also short time energy of speech signal is utilized for boundary value detection and other analysis. But a steady average energy in the zero crossing interval for speech recognition has not yet been carried out.

Zero crossing information of the speech signal is a perceptually meaningful parameter because parameters like formant frequency can be extracted using zero crossing interval information of noise corrupted speech. T.V. Sreenivas and Niederjohn proposed a new method of finding the formant frequency of the noise corrupted speech signal using statistical properties of zero crossing intervals [95]. The method is compared with currently popular spectral analysis technique based on singular value decomposition and
found to provide generally better resolution and lower variance at low signal to noise ratio (SNR).

In the context of discrete-time signals, a zero crossing is said to occur if successive samples have different algebraic signs. The rate at which zero crossings occur is a simple measure of the frequency content of a signal. Zero-crossing rate is a measure of number of times in a given time interval/frame that the amplitude of the speech signals passes through a value of zero, Fig.3.3 and Fig.3.4. Speech signals are broadband signals and interpretation of average zero-crossing rate is therefore much less precise.

However, rough estimates of spectral properties can be obtained using a representation based on the shorttime average zero-crossing rate [89].

![Fig. 3.3. Definition of zero-crossings rate](image1)

![Fig.3.4. Distribution of zero-crossings for unvoiced and voiced speech [89](image2)

The model for speech production suggests that the energy of voiced speech is concentrated below about 3 kHz because of the spectrum fall of introduced by the glottal wave, whereas for unvoiced speech, most of the energy is found at higher frequencies.
Since high frequencies imply high zero crossing rates, and low frequencies imply low zero-crossing rates, there is a strong correlation between zero-crossing rate and energy distribution with frequency. A reasonable generalization is that if the zero-crossing rate is high, the speech signal is unvoiced, while if the zero-crossing rate is low, the speech signal is voiced [89].

**Phonation Number:** It is a qualitative parameter. The phonation number is defined as the number of single speech fragments separated by silent pauses. The phonation number was equal to the number of pauses. Kuniszyk-Jo’z’kowiak [96] have found that the total phonation time (i.e., the sum of all the phonation times in an utterance) is longer in fluent utterances than in nonfluent ones. The total phonation time depends on the sound level at which it is measured.

The aim of the work [96] focuses to find acoustic (objectively measurable) dimensions that characterize utterances of fluent speakers and of stutterers. Their earlier work [96] had stated that there were significant differences in courses of dependence of phonation number on sound level and in the statistical distributions of the times of phonations and pauses between them. The work focuses on acoustic characteristics of utterances that distinguishes stuttered and fluent speech. The interdependence of phonation number (the number of times a given sound level was crossed) and sound intensity level was determined in this study. Also the average distributions of phonation times and pause durations as a function of sound intensity level were examined. The data suggest that differences in speech envelopes of stutterers and non-stutterers may be used to evaluate the degree of speech non-fluency. The result support the hypothesis that the cause of stuttering is a lack of synchronization between laryngeal functioning and vocalizing activities. Their result states that transitional states between vowels and consonants are either prolonged or more abrupt than normal.

In this study on the basis of the course of speech envelopes the number of times the signal amplitude exceeded the preset criterion was calculated called the number of phonation.
Wavelet Analysis

Wavelet Analysis is the analysis based on time frequency localization of the signal which can be advantageous in nonstationary problems. Wavelet expansions often provide very concise signal representation and thereby can simplify subsequent nonlinear analysis and processing. Wavelet analysis is a specialist tool for deriving an indepth understanding of underlying components of complex curvilinear data. Wavelet analysis is based on a windowing technique employing regions of varying size and can provide the time and frequency information inspite of overcoming the resolution limitations of the Fourier transform (STFT). Long-time intervals are used to provide more precise information at low frequencies; shorter intervals are used to extract characteristics of high-frequency components of a signal. The wavelet transform is a powerful signal-processing tool that has been successfully used in the analysis of non-stationary data from a wide range of physical process, for example, from engineering systems and in meteorology, that exhibit multiscale features [97 - 99]. It has also been used extensively for speech compression and recognition [100, 101] and for a number of studies on speech waveform analysis [102 - 104].

One of the most well-known tools for signal processing is the Fourier transform. This breaks down a signal into its constituent sinusoids of different frequencies. One of the drawbacks of Fourier analysis is that in transforming to the frequency domain, time information is lost. For stationary signals, i.e., those that do not change much over time, this is not a serious drawback. However, for nonstationary data, spectral techniques that retain information from the time domain are more appropriate. Nonstationary waveforms are commonly encountered in experimental acoustics. Nobel prize winner Dennis Gabor adapted the Fourier transform in order to analyse only a small section of the signal, known as the spectral window, at a time. This so-called short-time Fourier transform (STFT) maps a signal into a two-dimensional function of time and frequency, so one can determine approximately when an event of a particular frequency occurs. However, although STFT has proved useful in numerous application, it is ill suited for the study of signals where the frequency content ranges over several orders of magnitude as the size of the window is the same for all frequencies.
This paper [ref above] study was to adapt wavelet analysis as a tool for discriminating speech sample taken from healthy subjects across two biological states. Speech pressure waveforms were drawn from a study on effects of hormone fluctuations across the menstrual cycle on language functions. Speech samples from the vowel portion of the syllable ‘pa’, taken at the low- and high-hormone phase of the menstrual cycle, were extracted for analysis. Firstly they applied fourier transform to examine the fundamental and format frequencies. Wavelet analysis was used to investigate spectral differences at a more microbehavioural level. The results found are that wavelet coefficients for the fundamental frequency of speech samples taken from the high-hormone phase had larger amplitude than those from the low-hormone phase. This study provided evidence for differences in speech across the menstrual cycle that affected the vowel portion of syllables [105]. This study provided a new tool for examination of behavioural differences in speech linked to hormonal variation.

The continuous wavelet transform (CWT) of a signal is defined as the sum over all time of the signal multiplied by scaled, shifted versions of the mother wavelet. Thus mathematically, the wavelet transform $C_{a,b}(t)$ of a signal $f(t)$, where $t$ represents any independent variable, is defined as

$$C_{a,b}(t) = \frac{1}{|a|^{1/2}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt$$

where the function $\psi(t)$ is a mother wavelet and the real number $a$ ($a \neq 0$) and $b$ denote the scaling and translation respectively. $C_{a,b}(t)$ are known as the wavelet coefficients. The constituent wavelets of the original signal can be found by taking the product of each coefficient with appropriately scaled and shifted wavelet. This facilitates the detailed analysis of the signal over a range of frequencies. The scale $a$ is related in a broad sense to frequency by:

$$F_a = \frac{\Lambda t x F c}{a}$$
Where \( F_a \) is the pseudo-frequency in Hz corresponding to the scale \( a \) in Hz, \( F_c \) is the center frequency in Hz, \( f_c \) is the center frequency in Hz of the mother wavelet and \( \Delta t \) is the sampling period. For the case of the db 10 mother wavelet, the center frequency is 0.68421 Hz.

3.3 NONLINEAR TECHNIQUES:

Despite the limited technological success of the linear model in several applications, such as speech coding, synthesis and recognition, there is strong theoretical and experimental evidence [106, 107 - 112] for the existence of important nonlinear aerodynamic phenomena during the speech production that cannot be accounted for by the linear model. Various factors contributing to nonlinearities in speech production are (i) turbulent air flow through vocal tract, (ii) coupling produced between different parts of vocal tract (iii) neuro muscular response to stimuli [106].

The investigation of speech nonlinearities can proceed in at least two directions: (i) numerical simulation of the nonlinear differential equations governing the 3-D dynamics of the speech air flow in the vocal tract [108, 110]. (ii) development of nonlinear signal processing systems suitable to detect such phenomenon and extract related information[112 - 115]. The phonetory system is time varying and consequently the speech signal is nonstationary. This can be clearly understood if one closely observes the amplitude of the speech samples.

Nonlinear time series analysis aims at understanding the dynamics of a system using the time series of a single available variable. Chaos theory states that random inputs are not the only source for irregular output of a system: nonlinear chaotic systems with purely deterministic equations of motion can produce very irregular data. Extensive research conducted on nonlinear dynamical systems over recent years have proved that conventional time domain and frequency domain approaches to real world systems are far from optimal.
Many quantities in nature fluctuate in time. Examples are the stock market, the weather, seismic waves, sunspots, heartbeats, and plant and animal populations. Until recently it was assumed that such fluctuations are a consequence of random and unpredictable events. With the discovery of chaos, it has come to be understood that some of these cases may be a result of deterministic chaos and hence predictable in the short term and amenable to simple modeling. Many tests have been developed to determine whether a time series is random or chaotic, and if the latter, to quantify the chaos. A positive maximal Lyapunov exponent derived from the time series, expresses irregular deterministic behavior, which is termed chaotic [116 - 119], whereas dynamical systems with solely non-positive exponents are usually referred to as regular. If chaos is found, it may be possible to improve the short-term predictability and enhance understanding of the governing process of the dynamical system. We mean by talk of a 'dynamical system': a real-world system which changes over time.

While there is a long history of linear time series analysis, nonlinear methods have only just begun to reach maturity. When analyzing time series data with linear methods, there are certain standard procedures one can follow, moreover, the behavior may be completely described by a relatively small set of parameters. Linear methods interpret all regular structure in the data set, such as dominant frequency, as linear correlations. This means, in brief, that the intrinsic dynamics of the system are governed by the linear paradigm that small causes lead to small effects. Since linear equations can only lead to exponentially growing or periodically oscillating solutions, all irregular behavior of the system has to be attributed to some random external input to the system [121]. A brief comparison between linear and nonlinear methods [120] can be found in Table 3.1. Chaos theory has taught us that random input is not the only possible source of irregularity in a system's output: nonlinear, chaotic systems can produce very irregular data with purely deterministic equations of motion plus Table 3.1.

Now, Nonlinear Time Series Analysis (NTSA) is the study of the time series data with computational techniques sensitive to nonlinearity in the data. This was introduced by the theory of chaos to characterize the source complexity [121]. The NTSA theory offers
tools that bridge the gap between experimentally observed irregular behavior and deterministic chaos theory. It enables us to extract characteristic quantities of a particular dynamical system solely by analyzing the time course of one of its variables [121-123]. In theory, it would then be possible to collect temperature measurements in a particular city for a given period of time and employ nonlinear time series analysis to actually confirm the chaotic nature of the weather. Despite the fact that this idea is truly charming, its realization is not feasible quite so easily. In order to justify the calculation of characteristic quantities of the chaotic system, the time series must originate from a (i) stationary, (ii) deterministic system.

A deterministic dynamical system is one for which there is a rule, and , given sufficient knowledge of the current state of the system one can use the rule to predict future states; i.e. the future state $x_{n+r}$ can be determined precisely from the current state $x_n$ at any instance $n$ for some value of $t>0$, by applying the deterministic rule for the dynamical system. The other important requirement before attempting to do quantitative analysis is identification of stationarity of the dynamical system. Dynamical systems that are not stationary are exceedingly difficult to model from time series. Unless one has a priori knowledge of the structure of the underlying system, the number of parameters will greatly exceed the number of available data [124]. It may be noted here that the definition for stationarity is not the same as for linear systems: a linear system is said to be stationary if all its moments remain unchanged with time. A non-stationary system is defined as one which is subject to temporal dependence based on some outside influence. If we extend our definition of the system to include all outside influences, the system is stationary.
Table 3.1. Comparison of linear and nonlinear signal processing techniques.

<table>
<thead>
<tr>
<th>Linear signal processing</th>
<th>Nonlinear signal processing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Finding the signal</strong></td>
<td><strong>Finding the signal</strong></td>
</tr>
<tr>
<td><strong>Finding the space</strong></td>
<td><strong>Finding the space</strong></td>
</tr>
<tr>
<td>Use Fourier space methods to turn difference equations in to algebraic forms $x(t)$ is observed $X(f) = \sum x(t) e^{\frac{j2\pi f}{T}}$ is used</td>
<td>Time lagged variables from coordinates for a reconstructed state space in m dimensions. $X(t) = [x(t), x(t + \tau), x(t + 2\tau), \ldots, x(t + (m - 1)\tau)]$ where $\tau$ and $m$ are determined by false nearest neighbours and average mutual information.</td>
</tr>
<tr>
<td><strong>Classify the signal</strong></td>
<td><strong>Classify the signal</strong></td>
</tr>
<tr>
<td>- Sharp spectral peaks</td>
<td>- Lyapunov Exponents</td>
</tr>
<tr>
<td>- Resonant frequencies of the system</td>
<td>- Fractal Dimension measures</td>
</tr>
<tr>
<td>Making models, predict:</td>
<td>- Unstable fixed points</td>
</tr>
<tr>
<td>$x(t+1) = \sum \alpha_k x(t - k)$</td>
<td>- Recurrence quantification</td>
</tr>
<tr>
<td>Find parameters $\alpha_k$ consistent with invariant classifiers- location of spectral peaks</td>
<td>- Statistical distribution of the attractor</td>
</tr>
<tr>
<td>Making models, predict:</td>
<td></td>
</tr>
<tr>
<td>$X(t) \rightarrow X(t + 1)$</td>
<td></td>
</tr>
<tr>
<td>$X(t + 1) = F[X(t), a_1, a_2, \ldots, a_p]$</td>
<td></td>
</tr>
<tr>
<td>Find parameters $a_j$ consistent with invariant classifier - Lyapunov Exponents,fractal Dimension</td>
<td></td>
</tr>
</tbody>
</table>

Methods and Implementation

Let us suppose that we have a dynamical system which is both stationary and deterministic. To apply the nonlinear time series methods, the dynamics of the system are
to be presented in a phase space. When the equations that govern process dynamics are not known, the phase space is reconstructed from a measured time series by using only one observation. The most basic step in this procedure is to rely on a time delayed embedding of the data, i.e. attractor reconstruction. For this purpose, we have to determine the proper embedding delay and embedding dimension. There exist two methods, developed in the framework of nonlinear time series analysis, that enable us to successfully perform these desired tasks. The average mutual information method [125] yields an estimate for the proper embedding delay, whereas the false nearest neighbor method [126] enables us to determine a proper embedding dimension.

In the following sub sections the methods of phase space reconstruction, average mutual information method and false nearest neighbors method are discussed.

**Phase Space Reconstruction – Taken’s Embedding Theorem**

The basic idea of the phase space reconstruction is that a signal contains information about unobserved state variables which can be used to predict the present state. Therefore, a scalar time series \( s(t) \) may be used to construct a vector time series that is equivalent to the original dynamics from a topological point of view.

The problem of how to connect the phase space or state space vector of dynamical variables of the physical system to the time series measured in experiments was first addressed in 1980 by Packard et. al [127] who showed that it was possible to reconstruct a multidimensional state-space vector \( X(t) \) by using time delays (or advances which we write as positive delays) with the measured, scalar time series, \( s(t) \). Takens [128] and later Sauer et. al [129] put this idea on a mathematically sound footing by proving a theorem which forms the basis of the methodology of delays. They showed that the equivalent phase space trajectory preserves the topological structures of the original phase space trajectory. Due to this dynamical and topological equivalence, the characterization and prediction based on the reconstructed state space is as valid as if it was made in the true state space. The attractor so reconstructed can be characterized by a
set of static and dynamic characteristics. The static characteristics describe the geometrical properties of the attractor whereas the dynamical characteristics describe the dynamical properties of nearby trajectories in phase space.

Thus, given a time series \( \{x(j)\} = x(1), x(2), x(3), \ldots, x(N) \) we define points \( X(i) \) in an \( m \)-dimensional state space as

\[
X(i) = [x(i), x(i + \tau), x(i + 2\tau), \ldots, x(i + (m-1)\tau)]
\]

for \( i = 1, 2, 3, \ldots, N - (m-1)\tau \) where \( i \) are time indices, \( \tau \), a time delay or sometime referred to as embedding delay, and \( m \) is called the embedding dimension. Time evolution of \( X(i) \) is called a trajectory of the system, and the space, which this trajectory evolves in, is called the reconstructed phase space or simply, embedding space.

While the implementation of Eq. (1), the mathematical statement of Takens' Embedding theorem, should not pose a problem, the correct choice of proper embedding parameters \( \tau \) and \( m \) is a somewhat different matter. The most direct approach would be to visually inspect phase portraits for various \( \tau \) and \( m \) trying to identify the one that looks best. The word "best", however, might in this case be very subjective. In practice this approach for finding embedding parameters are seldom advised since we usually want to analyze a time series that originates from a rather unknown system. Then we would not know if the underlying dynamics that produced the time series had two or twenty degrees of freedom. It is easy to verify that the time required to check all possibilities that might yield a proper embedding with respect to various \( \tau \) and \( m \) is very long. This being said, let it be a good motivation to discuss the average mutual information method and the false nearest neighbor method, which enable us to efficiently determine proper values of the embedding delay \( \tau \) and embedding dimension \( m \). Let us start with the mutual information method.

**Selecting the Time Delay-Average Mutual Information Method**

A suitable embedding delay \( \tau \) has to fulfill two criteria. First, \( \tau \) has to be large enough so that the information we get from measuring the value of \( x \) variable at time \( (i+\tau) \) is
relevant and significantly different from the information we already have by knowing the value of the measured variable at time $i$. Only then it will be possible to gather enough information about all other variables that influence the value of the measured variable to successfully reconstruct the whole phase space with a reasonable choice of $m$. Note here that generally a shorter embedding delay can always be compensated with a larger embedding dimension. This is also the reason why the original embedding theorem is formulated with respect to $m$, and says basically nothing about $\tau$. Second, $\tau$ should not be larger than the typical time in which the system loses memory of its initial state. If $\tau$ would be chosen larger, the reconstructed phase space would look more or less random since it would consist of uncorrelated points. The latter condition is particularly important for chaotic systems which are intrinsically unpredictable and, hence, loose memory of the initial state as time progresses. This second demand has led to suggestions that a proper embedding delay could be estimated from the autocorrelation function where the optimal $\tau$ would be determined by the time the autocorrelation function first decreases below zero or decays to $1/\exp$. For nearly regular time series, this is a good thumb rule, whereas for chaotic time series, it might lead to spurious results since it based solely on linear statistic and doesn’t take into account nonlinear correlations.

The cure for this deficiency was introduced by Fraser and Swinney [130]. They established that delay corresponds to the first local minimum of the average mutual information function $I(\tau)$ which is defined as follows:

$$I(\tau) = \sum P(x(i), x(i+\tau)) \log_2 \left[ \frac{P(x(i), x(i+\tau))}{P(x(i))P(x(i+\tau))} \right]$$

(2)

where $P(x(i))$, is the probability of the measure $x(i)$, $P(x(i+\tau))$ is the probability of the measure $x(i+\tau)$ and $P(x(i), x(i+\tau))$ is the joint probability of the measure of $x(i)$ and $x(i+\tau)$ [130]. The average mutual information is really a kind of generalization to the nonlinear phenomena from the correlation function in the linear phenomena. When the measures $x(i)$ and $x(i+\tau)$ are completely independent, $I(\tau) = 0$. On the other hand when $x(i)$ and $x(i+\tau)$ are equal, $I(\tau)$ is maximum. Therefore plotting $I(\tau)$ versus $\tau$ makes it
possible to identify the best value for the time delay, this is related to the first local minimum.

While it has often been shown that the first minimum of $I(r)$ really yields the optimal embedding delay, the proof of this has a more intuitive, or shall we rather say empiric, background. It is often said that at the embedding delay where $I(r)$ has the first local minimum, $x(i+r)$ adds the largest amount of information to the information we already have from knowing $x(i)$, without completely losing the correlation between them. Perhaps a more convincing evidence of this being true can be found in the very nice article by Shaw [131], who is, according to Fraser and Swinney, the idea holder of the above reasoning. However, a formal mathematical proof is lacking. Kantz and Schreiber [121] also report that in fact there is no theoretical reason why there should even exist a minimum of the mutual information. Nevertheless, this should not undermine one's trustworthiness in this particular method, since it has often proved reliable and well suited for the appointed task.

Once the time delay has been agreed upon, the embedding dimension is the next order of business. Let us now turn to establishing a proper embedding dimension $m$ for the examined time series.

Selecting Embedding Dimension- False Nearest Neighbors Method

In general, the aim of selecting an embedding dimension is to make sufficiently many observations of the system state so that the deterministic state of the system can be resolved unambiguously. It is best to remember that in the presence of observational noise and finite quantization this is not possible. Moreover, it has been shown that even with perfect observations over an arbitrary finite time interval, a correct embedding will still yield a set of states indistinguishable from the true state [132]. Most methods to estimate the embedding dimension aim to achieve unambiguity of the system state. The archetype of many of these methods is the so-called False Nearest Neighbors (FNN) technique [133, 134].
The false nearest neighbor method was introduced by Kennel et al. [126] as an efficient tool for determining the minimal required embedding dimension $m$ in order to fully resolve the complex structure of the attractor, i.e. the minimum dimension at which the reconstructed attractor can be considered completely unfolded. Again note that the embedding theorem by Takens [128] guarantees a proper embedding for all large enough $m$, i.e. that is also for those that are larger than the minimal required embedding dimension. In this sense, the method can be seen as an optimization procedure yielding just the minimal required $m$. The method relies on the assumption that an attractor of a deterministic system folds and unfolds smoothly with no sudden irregularities in its structure. By exploiting this assumption, we must come to the conclusion that two points that are close in the reconstructed embedding space have to stay sufficiently close also during forward iteration. If this criterion is met, then under some sufficiently short forward iteration, originally proposed to equal the embedding delay, the distance between two points $X(n)$ and $X(p)$ of the reconstructed attractor, which are initially only a small distance apart, cannot grow further as we fix a threshold value for these distances in computation. However, if an $n$-th point has a close neighbor that doesn't fulfill this criterion, then this $n$-th point is marked as having a false nearest neighbor. We have to minimize the fraction of points having a false nearest neighbor by choosing a sufficiently large $m$. As already elucidated above, if $m$ is chosen too small, it will be impossible to gather enough information about all other variables that influence the value of the measured variable to successfully reconstruct the whole phase space. From the geometrical point of view, this means that two points of the attractor might solely appear to be close, whereas under forward iteration, they are mapped randomly due to projection effects. The random mapping occurs because the whole attractor is projected onto a hyper plane that has a smaller dimensionality than the actual phase space and so the distances between points become distorted [135].

In order to calculate the fraction of false nearest neighbors, the following original algorithm is used.

52
Consider each vector $X(n) = [x(n), x(n + \tau), x(n + 2\tau), \ldots, x(n + (m - 1)\tau)]$ in a delay coordinate embedding of the time series with delay $\tau$ and embedding dimension $m$. Look for its nearest neighbor $X(p)$ and $X(p) = [x(p), x(p + \tau), x(p + 2\tau), \ldots, x(p + (m - 1)\tau)]$. The nearest neighbor is determined by finding the vector $X(p)$ in the embedding which minimizes the Euclidean distance $R_n = \|X(n) - X(p)\|$. Now consider each of these vectors under an $m + 1$ dimensional embedding,

$X'(n) = [x(n), x(n + \tau), x(n + 2\tau), \ldots, x(n + (m - 1)\tau), x(n + m\tau)]$

$X'(p) = [x(p), x(p + \tau), x(p + 2\tau), \ldots, x(p + (m - 1)\tau), x(p + m\tau)]$

In an $m + 1$ dimensional embedding, these vectors are separated by the Euclidean distance $R'_{n} = \|X'(n) - X'(p)\|$. The first criterion by which Kennel, et. al., identify a false nearest neighbor is if

$$\left[ \frac{R_{n}^2 - R_{n}'}{R_{n}^2} \right]^{1/2} \frac{|x(n + m\tau) - x(p + m\tau)|}{R_{n}} > R_{tol}$$

(3)

$R_{tol}$ is a unit less tolerance level for which Kennel, et. al. [124], suggest a value of approximately 15. This criterion is meant to measure if the relative increase in the distance between two points when going from $m$ to $m + 1$ dimensions is large. 15 was suggested based upon empirical studies of several systems, although values between 10 and 40 were consistently acceptable. The other criterion Kennel, et. al., suggest is

$$\frac{R_{n}}{R_{A}} > A_{tol}$$

(4)

Where $A_{tol}$ is called absolute tolerance. This was introduced to compensate for the fact that portions of the attractor may be quite sparse. In those regions, near neighbors are not actually close to each other. Here, $R_{A}$ is a measure of the size of the attractor, for which Kennel, et. al., use the standard deviation of the data. If either (3) or (4) hold, then $X(p)$ is considered a false nearest neighbor of $X(n)$. The total number of false nearest neighbors is found, and the percentage of nearest neighbors, out of all nearest neighbors, is measured. An appropriate embedding dimension is one where the percentage of false nearest neighbors identified by either method falls to zero.
The above combined criteria correctly identify a suitable embedding dimension in many cases. By now we have equipped ourselves with the knowledge that is required to successfully reconstruct the embedding space from an observed variable. This is a very important task since basically all methods of nonlinear time series analysis require this step to be accomplished successfully in order to yield meaningful results.

3.4 INVARIANT PARAMETERS:

Lyapunov Exponents

The exponential divergence of nearby trajectories is calculated by the Lyapunov exponent. It is a measure of the rate of attraction or repulsion. If two nearby trajectories on a chaotic attractor start off with a separation $d_0$ at time $t = 0$, then the trajectories diverge so that their separation at time $t$, denoted by $d(t)$ satisfies the expression

$$d(t) = d_0 e^{\lambda t} \quad (1)$$

where $\lambda$ is called the Lyapunov exponent for the trajectories.

The Lyapunov exponents provide a coordinate-independent measure of the asymptotic local stability of properties of a trajectory. In a geometry representation, we can imagine a small infinitesimal ball of radius $\varepsilon(0)$ centered on a point $\Phi(0)$ in state space. Under the action of the dynamics the center of the ball may move, and the ball becomes distorted. Since the ball is infinitesimal, this distortion is governed by the linear part of the flow. The ball thus remains an ellipsoid. Suppose the principal axis of the ellipsoid at time $t$ are of length $\varepsilon(t)$. The spectrum of Lyapunov exponents for trajectory $\Phi(t)$ is defined as

$$\lambda_i = \lim_{t \to +\infty} \lim_{\varepsilon(t) \to +\infty} \frac{1}{t} \log \frac{\varepsilon_i(t)}{\varepsilon(0)} \quad (2)$$

The Lyapunov exponents depend on the trajectory $\Phi(t)$. Their values are the same for any state on the same trajectory, but may be different for states on different trajectories.
Lyapunov exponents are convenient for categorizing steady state behaviour. The trajectory of an n-dimensional state space have n Lyapunov exponents. This is often called the Lyapunov spectrum and it is conventional to order them according to size. The qualitative features of the asymptotic local stability properties can be summarized by the sign of each Lyapunov exponent; a positive Lyapunov exponent indicating an unstable direction, and a negative exponent indicating a stable direction. If the exponent is positive then the trajectories diverge and the system is chaotic. However, for an attractor, contraction must outweigh expansion and so,

$$\sum_{j=1}^{n} \lambda_j < 0 \quad (3)$$

The geometrical meaning of positive Lyapunov exponents is that there exist directions in which the motion on average is unstable such that nearby trajectories in these directions will diverge from the original orbit. Although the orbits is unstable, its stable directions provide sufficient volume contraction so that the orbit is confined to some bounded region in state space. At least one Lyapunov exponent must be zero for any limit set other than an equilibrium point. To produce a strange attractor the system must be dissipative and hence must have at least one negative Lyapunov exponent. Furthermore, at least one Lyapunov exponent must be zero for any limit set other than an equilibrium point. Also for a chaotic system, at least one Lyapunov exponent must be positive. It follows that a strange attractor must have at least lyapunov exponents. Hence, chaos can only occur in third-order autonomous, second-order non-autonomous or higher order continuous time systems.

**Dimensions**

Dimension of an attractor corresponds to the number of equations (variables) necessary to describe a system. A system attractor could be defined to be n-dimensional if in a neighborhood of every point it looks like an open subset of $\mathbb{R}^n$. This is how in differential topology the dimension of a manifold is defined. For instance, fixed point is of 0 dimension, while a limit cycle while locally looks like an interval is one dimensional and a torus is two dimensional. The neighbourhood of any point of a strange attractor however has a fine structure and does not manifolds and do not have integer dimension.
There are several ways to generalize dimension to the fractional case, some of which are listed below.

**Capacity dimension:** The simplest type of dimension measure, which is also referred as a Fractal dimension. To compute this measure, the attractor in phase space is covered by a regular grid of volume elements (cubes, spheres etc.) of diameter $\varepsilon$. If the attractor is a D-dimensional manifold, the number of volume elements needed to cover it for small $\varepsilon$ is given by, $N(\varepsilon)=k\varepsilon^{-D}$ for some constant $k$. The definition of capacity dimension is obtained from this by taking the $\varepsilon$-limit.

$$D_0 = \lim_{\varepsilon \to 0} \frac{\ln(-\sum_{i=1}^{n} p_i \ln p_i)}{\ln(1/\varepsilon)}$$

For manifolds, $D_0$ is equal to the dimension of the manifold and is an integer while for objects that are not manifolds it gives a nonintegeral value.

**Information dimension:** While $D_0$ is a metric concept and does not utilize the information on the time behaviour of the system, information dimension is a probabilistic measure defined in terms of the relative frequency of visitation of a trajectory. This dimension is defined as,

$$D_1 = \lim_{\varepsilon \to 0} \frac{\ln(-\sum_{i=1}^{n} p_i \ln p_i)}{\ln(1/\varepsilon)}$$

Here, $p_i$ is the relative frequency at which a typical trajectory enters the with volume element of the covering is the amount of information needed to specify the state of the system to accuracy $\varepsilon$ if the state is known to be on the attractor. Hence the name information dimension.
**Correlation dimension**: Yet another probabilistic measure is the correlation dimension defined as

\[
D_2 = \lim_{\varepsilon \to 0} \frac{\ln \sum_{i=1}^{N(\varepsilon)} p_i^2}{\ln \varepsilon}
\]

An easy way to estimate this dimension is by determination of the correlation function for N points given by,

\[
C(\varepsilon) = \lim_{\varepsilon \to 0} \frac{1}{N^2} \sum \{ \text{the number of pairs } x_i \text{ such that } ||x_i - x_j|| < \varepsilon \}
\]

The Correlation Dimension or attractor dimension of a time series can be estimated from the Correlation Integral \(C(r)\) of the phase space vectors.

\[
C(r) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} \theta(r - ||x_i - x_j||)
\]

where \(\theta\) is Heaviside unit step function defined as 1 if the distance between vectors is less than \(r\) and 0 if distance is greater than \(r\), \(N\) is the number of data points in the phase space, \(x_i, x_j\) are points of trajectory in the phase space, the distance \(r\) is a radius around each reference point \(x_i\).

Grassberger and Procassia showed that \(C(r)\) obeys the scaling law

\[
C(r) = r^{D_2}
\]

Therefore \(D_2 = \lim_{r \to 0} \frac{\log C(r)}{\log (r)}\)

Where \(D\) is the dimension. \(C(r)\) is a measure of the probability that two arbitrary points \(x_i, x_j\) of phase space will be separated by distance \(r\). Plotting \(\log C(r)\) versus \(\log (r)\) allows us to calculate \(D_2\) from the slope of the curves. Main advantage of dimensional
analysis is the investigation of time series with noise like spectra. If the phase space representation of such signal converges to an attractor, its dimensionality is a finite number. Otherwise the time series has stochastic properties and cannot be considered as a signal from deterministic system.

$D_2$ is calculated for the standard Lorenz system using Grassberger and Procassia algorithm as shown in figure 3.5. We get the logarithm of the correlations sum, the correlation dimension.

$$\dot{X} = -\sigma (X - Y)$$
$$\dot{Y} = rX - Y - XZ$$
$$\dot{Z} = b (XY - Z)$$

were $\sigma$, $r$ and $b$ are parameters

were $\sigma = 10$, $b = 8/3$ and $r = 99$

Figure 3.5. Calculation of correlation dimension for Lorenz equation.
With these different measures defined, the natural question is to find out how these are related to each other. Capacity dimension considers the attractor as a static object completely ignoring that it is subject to a dynamical flow. In practical experiments and simulations the attractor is not directly seen, only the trajectories over a finite period are observed. Hence the probabilistic measures are of greater use in practical settings than the capacity dimension. In general, $D_2 \leq D_1 \leq D_0 \leq D_1$ since $D_2$ and $D_1$ being probabilistic in nature depend on the relative frequency at which each volume element is visited and hence will be equal to $D_0$ only when these frequencies are all equal.

**Entropies**

The dynamical characteristics of the system are studied using the measures of entropy. These measures, which help in understanding how the trajectories evolve in time, are specific of dynamical systems while dimensions can be defined for any fractal measure or point set. Entropy estimates the average information gained by observing a system’s state to a precision $\epsilon$. In the case of a fixed point or periodic orbit each orbit remains the same in time and thus no new information is gained by observing additional orbits. Or in other words, the uncertainty in the system is almost nil. Thus it becomes easy to predict the outcome of the system under such an evolution. However for a chaotic system, each new orbit contributes to information gain. Hence prediction of the next orbit without observing it becomes very difficult. Such a system is ‘creating’ information and a suitable measure of the amount of information produced or gained is required.

**Kolmogrov Sinai entropy**: The KS entropy is such a measure that quantifies the amount of information gained on an average by observing a portion of the system’s evolution. It measures the degree of ‘chaoticness’ of the system and is a long time average rate of information gain of the system. It is inversely proportional to the time interval over which the state of the system can be predicted, given the initial conditions to a precision $\epsilon$ as well as the evolution equations. In order to define this measure, let us consider partitioning the attractor as before into $N(\epsilon)$ boxes $S_1, S_2, \ldots, S_N$ with size $\epsilon$. If $m$ measurements at regular time intervals are made, these will yield a sequence of boxes.
visited by the tractectory. Let $P(s_1, s_2, \ldots, s_N)$ be the joint probability of finding the trajectory at time $\Gamma$ in box $s_1$, at time $2\Gamma$ in box $s_2$, and so on. The KS entropy is defined as,

$$K = \lim_{\tau \to 0} \lim_{\varepsilon \to 0} \lim_{\lambda \to 0} \lim_{m \to \infty} \frac{1}{m \tau} \sum_{s_1, \ldots, s_m} P(s_1, \ldots, s_m) \times \ln P(s_1, \ldots, s_m)$$

(1)

If $K$ approaches 0, or there is no change in information this means the system is fully predictable. In contrast for a stochastic process, $K$ approaches infinity and it attains in between values depending on the irregular nature for chaotic systems.

The entropies can be generalized to a set of order $q$ Renyl entropies, which are dynamical counterparts of Renyi dimensions. These are defined as

$$K_q = \lim_{\tau \to 0} \lim_{\varepsilon \to 0} \lim_{\lambda \to 0} \lim_{m \to \infty} \frac{1}{m \tau} \frac{1}{q - 1} \ln \sum_{s_1, \ldots, s_m} p^q(s_1, \ldots, s_m)$$

(2)

The $K$ entropy in eq. (1) for which $q=2$ in eq. (2) is however the easiest to compute among all the entropies.

Kolmogorov Entropy provides a measure of the rate of information flow in the system. It is thus the dynamical evolution of the system. $K_2$ is computed from time series data using the equation

$$K_2 = \frac{1}{T(d'-d)} \ln \frac{Cd(r)}{Cd'(r)}$$

(3)

Where $d$ and $d'$ are two embedding dimensions.

$K_2$ is calculated for the standard Lorenz system

$$\dot{X} = -\sigma (X - Y)$$
$$\dot{Y} = rX - Y - XZ$$
$$\dot{Z} = b (XY - Z)$$

using Grassberger and Procassia algorithm as shown in figure 3.6.
Figure 3.6. Calculation of Kolmogorov entropy ($K_d$)

**Entropy: Various forms and Interpretations:**

**Information and Entropy: Interpretations**

Information is one of the many interpretations of entropy. Chronologically it was relatively a very late interpretation. The concept of entropy was conceived and its application to dynamical interpretations was introduced in Clausius's theory of thermodynamics. In this theory, entropy is defined as the measure of unavailable energy which arises due to heat loss as well as other actions such as chemical reactions, mixing, change from solid to liquid to gas. These processes involve not only an increase in entropy but also an accompanying decrease in the orderly arrangement of constituent atoms implying an increase in disorder. For example, atoms and molecules are ordered in a solid thus having low entropy whereas in the case of liquids the entropy will be higher with less ordering of atoms or molecules. Therefore the measurement of entropy became regarded as the measurement of degree of disorder or disorganization of a system.
Still another interpretation is in terms of probability. Just as entropy is maximum for disordered conditions, it is also maximum for equally probable events i.e. when all possible outcomes are equally likely, the probability of any one outcome is low and entropy is high. In contrast, when a biased dice is thrown, the outcome will not be equally probable for all possible outcomes. In such cases, the entropy is lowest. So there is an inverse relation between entropy and probability. Another interpretation based on the probability notion is that of uniformity of the distribution of data. If the data is uniformly distributed among a certain number of compartments, the probability of getting each compartment in a single trial is equal. In that sense, a uniform distribution is a high entropy condition. Conversely, a very non-uniform distribution means low entropy, because one bin has a probability of one and other bins have a probability of zero.

Yet another notion of entropy is uncertainty. The uncertainty can be pertained to the outcome of an experiment about to be run, or it can pertain to the state of a dynamical system. When the outcome of an event is absolutely certain then uncertainty is zero indicating zero entropy. For eg: absolute certainty means probability P=1, for which case the entropy will be zero.

Another idea is related to randomly distributed observations versus reliable predictability. When there is disorder, and great uncertainty, predictions cannot be based on any known structure or pattern and can only be done probabilistically. In such cases where predictability is low, entropy will be high. In contrast, something well organized or nearly certain is usually very much predictable resulting in low entropy. The idea of many possible outcomes suggests diversity. Another idea of interpreting entropy is related to the information content of an event. In a given probability distribution there is an information value of so many bits. Furthermore, a relatively large number of bits means a relatively large number of information and vice versa. Hence entropy is the average amount of new information gained from a measurement. Table. 3.2 shows the different cases where low and high entropies occur with respect to the above discussed interpretations.
Table 3.2 Different cases where low and high entropies occur

<table>
<thead>
<tr>
<th>High Entropy</th>
<th>Low Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Large proportion of energy unavailable for doing work</td>
<td>Large proportion of energy available for doing work</td>
</tr>
<tr>
<td>2. Disorder, disorganization, thorough mix</td>
<td>Order, high degree of organization</td>
</tr>
<tr>
<td>3. Equally probable events, low probability of a selected event</td>
<td>Preordained outcomes, high probability of a selected event</td>
</tr>
<tr>
<td>4. Uniform distribution</td>
<td>Highly uneven distribution</td>
</tr>
<tr>
<td>5. Great uncertainty</td>
<td>Near certainty, high reliability</td>
</tr>
<tr>
<td>6. Randomness, unpredictability</td>
<td>Non randomness, accurate forecasts</td>
</tr>
<tr>
<td>7. Much information</td>
<td>Little information</td>
</tr>
</tbody>
</table>

Entropy Measures for Complexity Analysis

Approximate entropy characterizes the regularity of a signal by measuring the presence of similar patterns in a time series. Consider a time series of length N, from this time series short sequences or patterns $x_m(i)$ of length m are constructed and the quantity $C''_r$ with tolerance r defined as

$$C''_r(r) = N^{-1} \{ \text{number of } j \leq N - m + 1 \ | d \left[ x_m(i), x_m(j) \right] \leq r \}$$

(1)

is computed for each $x_m(i)$

This quantity measures the regularity of the patterns by comparing them to a given pattern. Here m is the detail level at which the signal is analysed and r is the threshold which filters out irregularities. The regularity parameter ApEn is defined as

$$ApEn(m, r) = \lim_{N \to \infty} \left| \phi''(r) - \phi''(r) \right|$$

(2)

63
where

\[ \phi^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \]  

(3)

This gives the relative frequency of finding a vector \( x_m(j) \) similar to vector \( x_m(i) \) within a tolerance of \( r \) and provides a quantitative measure of entropy of the time series. ApEn statistic gives good results and provides general information about the regularity and persistence of a signal. However, in the above evaluation method, vectors or pattern \( x_m(i) \) are allowed to self match and therefore results in biased statistics [136, 139, 138]. To overcome this drawback, a modification of ApEn algorithm named Sample entropy (SampEn) is developed [139] which avoid this self matching. SampEn shows a relative consistency compared to ApEn [139]. However, this measure strongly depends on length of the series and also gives biased results in the case of irregular signals.

As in the case of Komogorov-Sinai entropy, both sample and approximate entropy, provide a measure for the information increase over one step of dimension from \( m \) to \( m+1 \). To be able to resolve complexity on scales larger than this smallest scale, multiscale entropy is introduced [140, 141]. The efficiency of ApEn and SampEn are enhanced by estimating these measures at different time scales. These measures are widely used for characterising biological signals in clinical application [142, 143].

Though the above methods give reliable results, their applicability to real world signal analysis is limited due to the sensitivity to noise and computation cost. Therefore, a fast and efficient algorithm which is also robust to noise contamination is very essential for online applications. PE is one such measure suitable for analysis of real world data.

**Relative entropy:**

Relative entropy measures applied to healthy and pathological voice characterization. According to Cover and Thomas [145], entropy is a quantity defined for any probability distribution with properties that agree with the intuitive notion of
information measures. One of the first concept was presented in [146] as the definition of a measure of uncertainty of random variable. Considering a random variable $X$ that assumes values $x \in \mathcal{X}$ where $\mathcal{X}$ is a finite set, the entropy $H(X)$ can be defined by $H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)$.

The probability of $x$, $\Pr\{X = x\}$, is denoted by $p(x) = 0$, $p(x)\log_2 p(x) = 0$ by convention. This quantity is dependent on the distribution of $X$ instead of the actual values of the random variable. The entropy measures the average amount of bits necessary to store outcomes of the random variable.

It is possible to estimate the relative entropy of two distributions[144], $p(x)$ and $q(x)$, in discrete-time form, in the following equation

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{p(x)}{q(x)}$$

The relative entropy is also known as the kullback-Leibler(KL) entropy, which may be understood as a measure of the difficulty of discriminating between two distributions.

Relative entropy is well suited due to its sensibility to uncertainties. Recently Paulo Roge’rio Scalassara et. al., [73] applied relative entropy to characterize healthy patients from pathological voice of patients suffering from vocal disorders like nodule in vocal fold and Reinke’s edema. The study showed that nonlinear dynamical methods seem to be suitable technique for voice signal analysis. Due to the chaotic component in the signals the signals are characterized by an increase in the signal complexity and unpredictability. The results showed that pathological groups had higher entropy values in accordance with other vocal acoustic parameters presented viz., jitter, shimmer, Harmonic to noise ratio. Measures of entropy are intimately related to the predictability of signals. These measures can be used to evaluate forecast skill of a system.

Permutation Entropy

Permutation Entropy [147] is a complexity measure which has aspects of both dynamical systems and entropy measures. PE calculation relies on the order relations between neighboring values of a time series. It estimates complexity as the entropy of the
distribution of permutations of groups of time samples. PE can efficiently detect the regular and complex nature of any signal and extract useful information about the dynamics. Thus the variation of PE as a function of time can effectively indicate dynamical change in any real world data. As this method does not require direct calculations of embedding dimension and time delay, this gives faster output and makes it suitable for online application of real time processes. It is robust against dynamical as well as observational noise [147].

Calculation of Permutation Entropy

Computation of PE is based on comparison of neighbouring values in the time series of any dynamical variable of a system. It has been shown that any continuous time series representing a dynamical system can be mapped on to a symbolic sequence [147, 148, 149]. According to the embedding theorem, any arbitrary time series \( X = \{x_1, x_2, \ldots, x_T\} \) can be mapped on to an 'n' dimensional space with vectors \( X_i = \{x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(n-1)\tau}\} \) where \( n \) is the embedding dimension and \( \tau \) is the delay time for embedding calculated using appropriate methods like false nearest neighbour calculation and first minimum of autocorrelation function [150]. For any arbitrary vector \( X \), the components are \( n \) number of real values of the time series \( \{x_1, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(n-1)\tau}\} \) from time instant ' \( t \)' to ' \( t+(n-1)\tau \) '. Assuming \( \tau = 1 \), each point in the \( n \) dimensional space represented by its corresponding vector will therefore be equivalent to a short sequence of the time series consisting of \( n \) number of real values as \( \{x_1, x_{i+1}, x_{i+2}, \ldots, x_{i+(n-1)}\} \). If the components of each vector are arranged in ascending order, it will represent a pattern of evolution. Thus each of the vectors can be considered as a symbolic sequence which will be one of the \( n! \) possible permutations of \( 'n' \) distinct symbols. The probability distribution of each pattern \( \pi \) can be represented as

\[
p(\pi) = \frac{\#\{t | t \leq T-n, (x_{i+1}, \ldots, x_{i+n}) \text{ has type } \pi\}}{T-n+1}
\]

(6)
where $\pi$ represents a pattern and $\#$ represents the number of occurrences. Permutation entropy of order $n \geq 2$ is defined as the Shannon entropy of the $n!$ patterns or symbolic sequences and can be written as

$$H(n) = -\sum p(\pi) \log p(\pi) \quad (7)$$

where the sum runs over all $n!$ permutations or sequences. $H(n)$ lies between 0 and $\log(n!)$. For increasing or decreasing sequence of values, $H(n) = 0$, whereas for random series where all $n!$ possible permutations appear with same probability, $H(n) = \log(n!)$. For a time series representing some dynamics, $H(n) < \log(n!)$. Therefore, normalised PE per symbol of order $'n'$ is given by $H(n)/\log(n!)$. Thus PE characterizes the system dynamics, with low values indicating regular behaviour. Any increase in PE value will thus represent an increase in irregularity in the dynamics. For detection of dynamical changes from time series it is first partitioned into non-overlapping windows of suitable length $T$. PE for each window is calculated using Eq.(6) and Eq.(7). Any change in the dynamics of the system will be reflected in the variation of PE with respect to moving window. For a reliable estimation of PE, the window length $T$ should be greater than $n!$ [147]. The order of PE should not be too small as this will not give enough number of distinct states. Too large values of order $'n'$ will demand large values of window size which will not effectively detect dynamical changes and also will create memory restrictions. Optimum values of order of PE are reported to be around 3 to 8[147,148]. In our analysis PE of order 5 is used for a window size of 512 samples for the time series of the audio signal.

As the patterns can be calculated in a very fast and easy way, calculation time of PE is negligibly less compared to other classical nonlinear methods. In this, only two pairs of values are compared at a time. PE based method is 100 times faster than Lyapunov exponent based method [147] due to the fact that neighbourhood searching is not needed. Also we deal with order relations between values instead of values themselves, the permutation entropy is robust with respect to noise corrupting the data.
Permutation entropy has a practically invariance property. If \( y_i = f(x_i) \) where \( f \) is an arbitrary strictly increasing (or decreasing) real function, then \( H(n) \) is same for \( x_i \) and \( y_i \). Such nonlinear function \( f \) occurs, for example, when measuring physiological data with different equipments. Addition of observational noise causes only a small increase in the value of entropy [151] where as there is hardly any effect on the entropy due to dynamical noise. However in the presence of high noise the ability of PE to distinguish the change in dynamics decreases.

**Standard Data Test on PE**

Effectiveness of PE is verified on regular chaotic and random data sets. For this, normalised PE for a regular sine wave of amplitude (peak to peak) 0.2 and a random signal of amplitude 0.2 with 5000 data points each are calculated. Fig 3.7 (a) and (b) shows a sine wave and random signal respectively and their corresponding PE are shown in Fig 3.7 (c) and (d). Permutation entropy for regular sine wave is 0.114 and it is 0.9387 for random signal. Hence it is confirmed that PE values corresponding to regular signal is low whereas for random variation it shows high values.
Fig. 3.7 Variation of PE for regular and random signal. (a) sine wave (b) random signal (c) PE for sine wave (d) PE for random signal
When regular sine wave is connected with random signal as given in Fig 3.8 (a) the PE value suddenly jumps from 0.114 to 0.9387 as indicated in Fig 3.8(b). This clearly shows that the PE is sensitive to change in regularity. The sudden variation from regular to random state is clearly indicated by the abrupt change in PE values.

To get the feeling of the variations of entropy, results are also verified on chaotic signals with change in dynamics for different parameter values. Bifurcation diagram of logistic map \( x_{r+1} = r x_r (1 - x_r) \) is used to study the variation in PE in chaotic signal. Fig. 3.9(a) shows the bifurcation diagram of logistic map for 5000 parameter values corresponding to variation of ‘r’ from 3.5 to 4. For control parameter r less than 3.57 the logistic map exhibits period doubling phenomenon, and a chaotic dynamics is observed at 3.57. For
At $3.57 \leq r \leq 4$ the dynamics is more complicated and intricate. This interval of $r$ is not fully occupied by chaotic orbits alone, but many changes take place at different critical values of $r$. We can clearly see many changes of dynamics as a function of control parameter at 3.5748, 3.5925, 3.6785 and 3.828. At these points the chaotic nature is lost and the behavior is regular or quasi periodic.

PE of order 6 is calculated for sliding windows of 1024 samples. Fig 3.9 (b) shows the variation of PE for the same values of $r$ varying from 3.5 to 4. The variation in PE clearly indicates the change in dynamics corresponding to different $r$ values. Corresponding to $r$ values where there is a change in chaotic state PE drops indicating more regularity. This confirms the sensitiveness of PE to change in dynamics of any type of signal.

![Fig. 3.9 Logistic equation for varying control parameter 'r'](image)

(a) Bifurcation diagram (b) Variation of PE

71
3.5 REFERENCES


81. PARLEY W. NEWMAN, RICHARD W. HARRIS, and LAURENCE M. HILTON, VOCAL JITTER AND SHIMMER IN STUTTERING, J. Flency disorder 14, 87-95, 1989.


144. Timothy DelSole, Predictability and information theory. Part I: Measures of predictability, J. Atmos. Sci. 61 (20) 2425-2440, 2004


