Chapter 7
Optimum Embedding Parameters for Phase Space Reconstruction

7.1 Introduction

Phase space reconstruction is usually the first step in the analysis of dynamical systems. Typically, an experimenter obtains a scalar time series from one observable of a multidimensional system. State-space reconstruction is then needed for the indirect measurement of the system’s invariant parameters like, dimension, Lyapunov exponent etc. Several techniques for attractor reconstruction are currently employed, such as derivative coordinates [Packard.N.H, Shaw.R.S, et.al., 1980], [Taken.F, 1981] and principal components (or singular value decomposition) [Broomhead.D.S. and King. G.P, 1986], the method of delays [Packard.N.H., Shaw.R.S, et al. 1980], [Taken.F, 1981] etc.

Takens' theorem [Taken.F, 1980] demonstrates that in the absence of noise a multidimensional state space can be reconstructed from a scalar time series, using the method of delays. It is the most widespread approach because it is the most straightforward and the noise level is constant for each delay component [Casdagli.M, Eubank.S, et. al., 1991]. The method of delays reconstructs the attractor dynamics by using delay coordinates to form multiple state-space vectors, $X_n$. The reconstructed state of the system at each discrete time $n$ is
\[ X_n = [x_n \ x_{n+\tau} \ ... \ x_{n+(d-1)\tau}] \]

Where \( \tau \) is the reconstruction time delay and \( d \) is the embedding dimension. Taken’s theorem gives little guidance, about practical considerations for reconstructing a good state space. It is silent on the choice of time delay (\( \tau \)) to use in constructing \( d \)-dimensional data vectors. Indeed, it allows any time delay as long as one has an infinite amount of infinitely accurate data. However, for reconstructing state spaces from real-world, finite, noisy data, it gives no direction [Casdagli.M, Eubank.S, et. al., 1991].

Experiments show that the quality of reconstruction depends on the value chosen for ‘\( \tau \)’ and experimenters and theorists note that there are no criteria for choosing ‘\( \tau \)’ in the literature. The following facts are to be remembered while choosing proper time delay [Abarbanel.H.D.I, 1996.]

i) It must be some multiple of the sampling time \( T_s \), since we only have data at those times.

ii) If time delay is too short, the coordinates, which we wish to use in our reconstructed data vector, will not be independent enough. That is, not enough time will have evolved for the system, for its phase space to produce new information about that phase space. If ‘\( \tau \)’ is too small each coordinate is almost the same, and the trajectories of the reconstructed space are squeezed along the identity line; this phenomenon is known as redundancy.
iii) If $\tau$ is too large, in the presence of chaos and noise, the dynamics at one time become effectively causally disconnected from the dynamics at a later time, so that even simple geometric objects look extremely complicated; this phenomenon is known as irrelevance [Casdagli.M, Eubank.S, et. al., 1991].

For the embedding dimension ‘d’, the theorem states the sufficient (but not necessary) condition is $d \geq 2D$, where D is the fractal dimensions of the underlying attractor. Normally, one has no a priori knowledge regarding the topological dimension and it is unclear what values of d will satisfy the condition.

Most of the research on the state space reconstruction problem has centered on the problems of choosing ‘$\tau$’ and dimension ‘d’ for delay coordinates. In practice, the determination of the time lag can be difficult, because there is no theoretically well-founded method to ascertain it, where an obvious criterion function can be formulated. Despite this limitation, two heuristics have been developed in the literature for establishing a time lag [Kantz.H and Schreiber.T, 2003]. 1) The first zero of the autocorrelation function and 2) the first minimum of the auto-mutual information curve [Fraser.A.M and Swinney.H.L, 1986]. These heuristics are premised on the principle that it is desirable to have as little information redundancy between the lagged versions of the time series as possible.
There have been many discussions on how to determine the optimal embedding dimension from a scalar time series based on Taken’s theorem or its extensions [Sauer.T, Yorke.J.A., and Casdagli.M, 1991]. The standard way to find the minimum embedding dimension ‘d’ is to use some criterion, which the geometry of the attractor must meet and check for which embedding dimension this is fulfilled. Among different geometrical criteria, the most popular seems to be the method of False Nearest Neighbors [Kennel.M.B, Brown.R, and Abarbanel.H.D.I, 1992]. This criterion concerns the fundamental condition of no self-intersections of the reconstructed attractor.

In this chapter we are discussing the selection of proper time delay in terms of first zero of autocorrelation function and first minimum of auto mutual information function along with a new method in terms of Maximum Euclidean Distance Measure. For the embedding dimension we have used Cao’s algorithm in which the method overcomes the shortcomings of the conventional techniques for finding False Nearest Neighbors. With these optimum embedding parameters, Reconstructed Phase Space Distribution Parameter (RPSDP) is modified.

In the first session of this chapter, selection of optimum time delay for Phase Space Reconstruction based on first zero of Autocorrelation function is described. Following the limitations of this method, selection of ‘τ’ based on first minimum of average mutual information function is presented. Then a new technique for the selection of proper time delay is introduced. In the next session, minimum embedding dimension for Phase Space Reconstruction is
determined using Cao’s method. Finally modification of RPSDP with the optimum embedding parameters is presented.

7.2 Proper Time Delay based on Autocorrelation Function

Autocorrelation is a mathematical tool used frequently in signal processing for analysing functions or series of values, such as time domain signals. Informally, it is a measure of how well a signal matches a time-shifted version of itself, as a function of the amount of time shift. Autocorrelation is useful for finding repeating patterns in a signal, such as determining the presence of a periodic signal which has been buried under noise, or identifying the missing fundamental frequency in a signal implied by its harmonic frequencies.

The Autocorrelation function (ACF) shows the value of the autocorrelation coefficient for different time lags $\tau$ [Abarbanel H.D.I, 1996]:

$$C(\tau) = \frac{E[(X_n - \mu)(X_{n+\tau} - \mu)]}{\sigma^2}$$ ....(7.1)

Where $E$ is the expectation value, $\mu$ is the mean and $\sigma$ the standard deviation. The estimation of the autocorrelations from a time series is straightforward as long as the lag $\tau$ is small compared to the total length of the time series and $C(\tau)$ quantifies how these points are distributed. If they spread out evenly over the plane, or if they are uncorrelated, then $C(\tau) = 0$. If they tend to crowd along the diagonal $X_n = X_{n+\tau}$, then $C(\tau) > 0$, and if they are closer to the line $X_n = -X_{n+\tau}$ we have $C(\tau) < 0$. The latter two cases reflect
some tendency of $X_n$ and $X_{n+\tau}$ to be proportional to each other. $C(\tau)$ is maximized when the delay ‘$\tau$’ is zero.

The autocorrelation function is expected to provide a reasonable measure of the transition from redundancy to irrelevance as a function of delay. A common choice for delay ‘$\tau$’ is the time at which the autocorrelation function has its first zero, which makes the coordinates linearly uncorrelated.

Figure 7.1(a) shows the autocorrelation function for vowel $\text{ao/Λ}$. Here $C(\tau)$ crosses first zero when the delay is 4. Figure 7.2(a) gives the corresponding reconstructed phase space for vowel $\text{ao/Λ}$ with dimension 3.

Figures 7.1(b-e) show the autocorrelation functions for vowels $\text{ao/I}$, $\text{ae/α}$, $\text{o/o}$, $\text{u/u}$ and figures 7.2(b-e) show the corresponding Phase portraits.
Fig. 7.1(a): Delay Vs. Autocorrelation function for Malayalam vowel \( \text{a}/\Lambda/ \)

Fig. 7.2(a): Reconstructed Phase space for vowel \( \text{a}/\Lambda/ \) with delay 4 and dimension 3
Fig. 7.1(b): Delay Vs. Autocorrelation function for Malayalam vowel ओ/ौ/

Fig. 7.2(b) Reconstructed Phase space for vowel ओ/ौ/ with delay 9 and dimension 3
Fig. 7.1(c): Delay Vs. Autocorrelation function for Malayalam vowel \( \text{ae/} \)

Fig. 7.2(c): Reconstructed Phase space for vowel \( \text{ae/} \) with delay 6 and dimension 3
Fig. 7.1(d): Delay Vs. Autocorrelation function for Malayalam vowel ə/o/

Fig. 7.2(d) Reconstructed Phase space for vowel ə/o/ with delay 5 and dimension 3
Fig. 7.1(e): Delay Vs. Autocorrelation function for Malayalam vowel ɡ/ʊ/

Fig. 7.2(e) Reconstructed Phase space for vowel ɡ/ʊ/ with delay 6 and dimension 3
The autocorrelation based methods have the advantage of short computation times when calculated via the fast Fourier transform (FFT) algorithm. A quite reasonable objection of this procedure is that it is based on linear statistics and only measures linear dependence. It is not taking into account nonlinear dynamical correlations. Since the relationship between the spatial distribution of a reconstructed attractor and the temporal autocorrelation of a single time series is an ill-defined one, the method tends to be inconsistent [Fraser.A.M. and Swinney.H.L, 1986], [Albano.A.M, Rapp.P.E et. al., 1988], [Martinerie.J.M, Rapp.P.E et. al., 1992]. For this reason, Fraser and Swinney [Fraser.A.M. and Swinney.H.L, 1986] suggested a spatial measure based on mutual information.

7.3 Proper Time Delay based on Mutual Information Function

In contrast to the linear dependence measured by autocorrelation, mutual information, I(τ), supplies a measure of general dependence [Fraser.A.M. and Swinney.H.L, 1986]. Therefore, I(τ) is expected to provide a better measure of the shift from redundancy to irrelevance with nonlinear systems. Mutual information answers the following question: Given the observation of x(t), how accurately can one predict x(t + τ)? Thus, successive delay coordinates are interpreted as relatively independent when the mutual information is small. According to Fraser and Swinney greatest independence, or the lowest I(τ), can be associated with the least redundancy and, therefore is the best for attractor reconstruction. Hence the proper time delay ‘τ’ can be
selected as the lag that produces a local minimum of $I(\tau)$. In a system with positive metric entropy any measurement stripe will eventually spread back into the invariant measure. This is accomplished by the well known stretching and folding effect. To avoid this type of spreading, the first local minimum of $I(\tau)$ is preferred to later minima [Fraser.A.M. and Swinney.H.L, 1986], [Liebert.W and Schuster.H.G, 1989].

7.3.1 Average Mutual Information

In probability theory and information theory, the mutual information of two random variables is a quantity that measures the mutual dependence of the two variables, or it is the amount of information that is shared between two data sets. Shannon’s information theory provides a formalism for quantifying the concept of information among measurements.

Consider a set of possible measurements $s_1, s_2, \ldots, s_n$ of a system ‘S’. Let $P_s(s_1), P_s(s_2), \ldots, P_s(s_n)$ be the associated probabilities of the measurements. Then $P_s$ maps the measurements to probabilities. The amount of information gained from a measurement that specifies $s$ is the entropy ‘H’ of the system,

$$H(S) = -\sum_i P_s(s_i) \log P_s(s_i) \quad \ldots \quad (7.2)$$

If the log is taken to the base two, $H$ is in units of bits.

Now we consider a general coupled system ‘S’ and ‘Q’. We have given that ‘s’ has been measured and found to be $s_i$, then the uncertainty in a
measurement of ‘q’ is given by,

\[ H(Q|s_i) = - \sum_j P_{q|s}(q_j|s_i) \log[P_{q|s}(q_j|s_i)] \] .......(7.3)

where \( P_{q|s}(q_j|s_i) \) is the probability that a measurement of q will yield \( q_j \), given that the measured value of s is \( s_i \). But the probability associated with the combined system S and Q, that a measurement of q will yield \( q_j \), and the measurement of s yield \( s_i \), \( P_{sq}(s_i,q_j) \) is given by:

\[ P_{sq}(s_i,q_j) = P_{q|s}(q_j|s_i) \cdot P_s(s_i) \]

then,

\[ H(Q|s_i) = - \sum_j [P_{sq}(s_i,q_j)/P_s(s_i)] \log[P_{sq}(s_i,q_j)/P_s(s_i)] \] .......(7.4)

Then for a time series \( x(t) \), we can measure how dependent the values of \( x(t+\tau) \) on the values of \( x(t) \), by making the assignment \([s,q] = [x(t), x(t+\tau)]\). Given that \( x \) has been measured at time \( t \), then the average uncertainty in a measurement of \( x \) at time \( t+\tau \) is given by averaging \( H(Q|s_i) \) over \( s_i \), which yields

\[ H(Q|S) = \sum_i P_s(s_i)H(Q|s_i) \] ............(7.5)

\[ = - \sum_{i,j} P_{sq}(s_i,q_j) \log[P_{sq}(s_i,q_j)/P_s(s_i)] \]

\[ = - \sum_{i,j} P_{sq}(s_i,q_j) \{ \log[P_{sq}(s_i,q_j)] - \log[P_s(s_i)] \} \]

\[ = - \sum_{i,j} P_{sq}(s_i,q_j) \log[P_{sq}(s_i,q_j)] + \sum_{i,j} P_{sq}(s_i,q_j) \log[P_s(s_i)] \]
But \[- \sum_{i,j} P_{sq}(s_i, q_j) \log[P_{sq}(s_i, q_j)] = H(S, Q) \quad \ldots (7.6)\]

and \[\sum_{i,j} P_{sq}(s_i, q_j) \log[P_s(s_i)] = \sum_{i,j} P_{q|s}(q_j|s_i) P_s(s_i) \log[P_s(s_i)]\]

\[= \sum_i P_s(s_i) \log[P_s(s_i)] \quad \text{since,} \quad \sum_j P_{q|s}(q_j|s_i) = 1\]

so \[\sum_{i,j} P_{sq}(s_i, q_j) \log[P_s(s_i)] = - H(S) \quad \ldots (7.7)\]

Hence \[H(Q|S) = H(S, Q) - H(S)\]

\(H(Q)\) is the uncertainty of \(q\) in isolation, and \(H(Q|S)\) is the uncertainty of \(q\) given a measurement of \(s\). So the amount that a measurement of \(s\) reduces the uncertainty of \(q\) is

\[I(Q, S) = H(Q) - H(Q|S)\]

\[= H(Q) + H(S) - H(S, Q)\]

This is the mutual information. It is the answer to the question, ‘Given a measurement of \(s\), how many bits on the average can be predicted about \(q\)?’

From equation 7.6, we have \[H(S, Q) = - \sum_{i,j} P_{sq}(s_i, q_j) \log[P_{sq}(s_i, q_j)]\]

and from equation 7.7, \[H(S) = - \sum_{i,j} P_{sq}(s_i, q_j) \log[P_s(s_i)]\]

similarly, \[H(Q) = - \sum_{i,j} P_{sq}(s_i, q_j) \log[P_q(q_j)]\]

Hence,

\[I(Q, S) = \sum_{i,j} P_{sq}(s_i, q_j) \log[P_{sq}(s_i, q_j)] - \sum_{i,j} P_{sq}(s_i, q_j) \log[P_{s}(s_i)] - \sum_{i,j} P_{sq}(s_i, q_j) \log[P_{q}(q_j)]\]

\[I(Q, S) = \sum_{i,j} P_{sq}(s_i, q_j) \log[P_{sq}(s_i, q_j) / P_{s}(s_i)P_{q}(q_j)]\]
If the measurement of a value from Q resulting in $q_j$ is completely independent of the measurement of a value from S resulting in $s_i$, then $P_{sq}(s_i,q_j)$ factorizes:

$$P_{sq}(s_i,q_j) = P_s(s_i) P_q(q_j)$$

and the amount of information between the measurements, average mutual information is zero.

For time series $x(t)$, the average mutual information between $x(t)$ and $x(t+\tau)$ is given by

$$I(\tau) = \sum_{x(t), x(t+\tau)} P(x(t), x(t+\tau)) \log[P(x(t), x(t+\tau))/P(x(t))P(x(t+\tau))]$$

If the measurements $x(t)$ and $x(t+\tau)$ are independent, then $I(\tau)$ will tend to zero. It was the suggestion of Fraser that one can use the function $I(\tau)$ as a kind of nonlinear autocorrelation function to determine, when the values $x(t)$ and $x(t+\tau)$ are independent enough of each other to be useful as coordinates in a time delay vector but not so independent as to have no connection with each other at all. The actual prescription suggested is to take the ‘$\tau$’ where the first minimum of the average mutual information $I(\tau)$ occurs as that value to use in time delay reconstruction of phase space.

The algorithm suggested by Fraser and Swinney is implemented using Mathlab and the proper delay for Malayalam vowels are calculated.

Figures 7.3(a-e) show the average mutual information function for Malayalam vowels and figures 7.4(a-e) give the corresponding reconstructed phase spaces with dimension 3. In figure 7.3(a) first minimum of $I(\tau)$ is when delay is 3
Fig.7.3(a): Delay Vs. Average Mutual Information function (vowel /ɔ/Λ/)

Fig7.4(a): Reconstructed Phase space for vowel /ɔ/Λ/ with delay 3 and dimension 3
Fig. 7.3(b): Delay Vs. Average Mutual Information function (vowel $\mathbf{\alpha}/\mathbf{l}/$)

Fig. 7.4(b): Reconstructed Phase space for vowel $\mathbf{\alpha}/\mathbf{l}/$ with delay 8 and dimension 3
Fig. 7.3(c): Delay Vs. Average Mutual Information function (vowel $\alpha\theta$/ae/)

Fig. 7.4(c): Reconstructed Phasespace for vowel $\alpha\theta$/ae/ with delay 4 and dimension 3
Fig. 7.3(d): Delay Vs. Average Mutual Information function (vowel a/o/)

Fig. 7.4(d): Reconstructed Phase space for vowel a/o/ with delay 4 and dimension 3
**Fig. 7.3(e):** Delay Vs. Average Mutual Information function (vowel ə/u/)

**Fig. 7.4(e):** Reconstructed Phase space for vowel ə/u/ with delay 4 and dimension 3
Average mutual information $I(\tau)$ reveals quite a different insight about the nonlinear characteristics of an observed time series than the more familiar autocorrelation function, where the latter is tied to linear properties of the source. The primary drawback of this approach is the enormous computational costs. Martinerie et. al. showed that mutual information is also inconsistent in identifying the optimal value of ‘$\tau$’ [Martinerie J.M et. al., 1992].

The literature contains many more suggestions on methods of how to determine an optimal time lag. Some of them have a nice heuristic justification. For a nonlinear time series analysis, it is our impression that it might be more useful to optimize the time lag with respect to a particular source and application. Hence for speech signal analysis, we introduced a novel method for determining the proper time lag and is explained in the next section.

7.4 Geometry based method for Proper Time Delay

Geometry-based methods for determining $\tau$ may be interpreted as various attempts to answer the following question: What value for the delay results in the most space filling reconstruction? Or for what value of delay the attractor in the reconstructed phase space is most dilated? In the previous chapter we have utilized the reconstructed phase space for the determination of pitch period of the speech signal. The idea behind this approach is that, if the time delay corresponds to pitch period of the signal (T), the Euclidean
distance measure of phase space points from the phase space diagonal (the identity line) is the minimum. If the Euclidean Distance Measure from the Phase Space Diagonal is maximum, the Phase Space points are widely spread in the reconstructed space. The delay corresponding to the Maximum Euclidean Distance Measure can be interpreted as the delay, which results in the most space filling reconstruction.

The expansion from the main diagonal can be best quantified by measuring the Euclidean distance of the phase space points, as explained in the previous chapter. The delay corresponds to Maximum Euclidian Distance Measure (MEDM) can be used for the reconstruction of phase space, as it results in most space filling reconstruction and provides most dilated attractor in the Phase Space.

Figures 7.5(a-e) show the variation of Euclidean Distance Measure of Phase Space points from the Phase Space diagonal with delay for Malayalam vowels and figures 7.6(a-e) give the corresponding reconstructed phase spaces with dimension 3. In figure 7.5(a), Maximum Distance Measure from the phase space diagonal corresponds to time delay 6.
**Fig. 7.5(a):** Delay Vs. Distance Measure from the phase space diagonal (vowel ι/Λ/)

**Fig 7.6(a):** Reconstructed Phase space for vowel ι/Λ/ with delay 6 and dimension 3
Fig. 7.5(b): Delay Vs. Distance Measure from the phase space diagonal (vowel \( \mathcal{A}/l/ \))

Fig. 7.6(b): Reconstructed Phase space for vowel \( \mathcal{A}/l/ \) with delay 13 and dimension 3
Fig. 7.5(c): Delay Vs. Distance Measure from the phase space diagonal (vowel \( \alpha_g /ae/ \))

Fig 7.6(c): Reconstructed Phase space for \( \alpha_g /ae/ \) with delay 10 and dimension 3
Fig. 7.5(d): Delay Vs. Distance Measure from the phase space diagonal (vowel α/o/)

Fig. 7.6(d): Reconstructed Phase space for vowel α/o/ with delay 9 and dimension 3
Fig. 7.5(e): Delay Vs. Distance Measure from the phase space diagonal (vowel $\sigma/u/$)

Fig. 7.6(e): Reconstructed Phase space for vowel $\sigma/u/$ with delay 11 and dimension 3
Fig. 7.7: Phase portrait of Malayalam vowel ɡ/u/ with first zero of Autocorrelation function as time delay

Fig. 7.8: Phase portrait of Malayalam vowel ɡ/u/ with first minimum of Mutual Information function as time delay

Fig. 7.9: Phase portrait of Malayalam vowel ɡ/u/ with maximum distance measure from the phase space diagonal as time delay
The criteria chosen for the selection time delay in the reconstruction of phase space are compared in figures 7.7, 7.8 and 7.9. The phase portrait shown in Figure 7.7 is constructed for time delay corresponding to the first zero in the autocorrelation function and it occurs for delay 6. In Figure 7.8 the phase portrait corresponds to first minimum of mutual information function and here the delay is 4. In these figures it is clear that the trajectories are indistinguishable and the attractor is not dilated much. Therefore it is difficult to deduce quantitative information about the dynamics from these phase portraits. On the other hand in the phase portrait shown in figure 7.9 the criteria used for the selection of time delay is the delay corresponds to maximum distance measure from the main diagonal, and the delay selected is 11. Since it is the most space filling reconstruction compared to the other two methods, in applications like speech recognition, parameters can be quantitatively deduced from this portrait. Phase space distribution parameter extracted from these portraits can be effectively utilized for the improvement of recognition accuracy as explained in chapter 8

7.5 Minimum Embedding dimensions for Phase Space Reconstruction

The embedding theorem tells us that if the dimension of the attractor defined by the orbits is D, then we will certainly unfold the attractor in an integer dimensional space of dimension d, where \( d \geq 2D \). This is not the necessary dimension for unfolding, but is sufficient and certainly tells us when to stop adding components to the time delay vector. The box counting dimension of the strange attractor for the Lorenz model is \( D \approx 2.06 \), which
would lead us to anticipate \( d = 5 \) to unfold the Lorenz attractor. But it is shown that \( d = 3 \) will do well for this system [Abarbanel.H.D.I, 1996]. Therefore we are certain from the theorem that some dimension \( \leq 2D \) will do for \( d \), and in the following session we will discuss which dimension is to be selected.

There have been many discussions in the literature on selecting optimum embedding dimension from a time series data [Martinerie.J.M., AlbanoA.M..et al, 1992], [Kennel. M. B., Brown. R., and Abarbanel. H. D. I,1992]. Following are the three basic methods which are usually used to choose the minimum embedding dimension:

(1) computing some invariant on the attractor[ Grassberger.P and Procaccia.I, 1983]. By increasing the embedding dimension used for the computation one notes when the value of the invariant stop changing. The typical problem with this approach is that it is often very data intensive, certainly subjective, and time-consuming for computation. (2) singular value decomposition and (3) the method of false neighbors [Kennel.M.B, Brown.R, et.al., 1992]. It was developed based on the fact that choosing too low an embedding dimension results in points that are far apart in the original phase space being moved closer together in the reconstruction space.

7.5.1 The method of False Nearest Neighbors

The concept called false nearest neighbors, was introduced by Kennel, Brown & Abarbanel (1992). The basic idea is to search for points in the data set which are neighbors in embedding space. Imagine that the correct
embedding dimension for some data set is d. Now study the same data in a lower embedding dimension, \( d_0 \) (i.e. \( d_0 < d \)). The transition from \( d \) to \( d_0 \) is a projection, eliminating certain axes from the coordinate system. Hence points whose coordinates are eliminated by the projection can become ‘false neighbors’ in the \( d_0 \) dimensional space.

If \( d_0 \) is qualified as an embedding dimension by the embedding theorems, then any two points which stay close in the \( d_0 \) dimensional reconstructed space will be still close in the \( d_0 + 1 \) dimensional reconstructed space. Such pair of points are called true neighbors, otherwise they are called false neighbors. Perfect embedding means that no false neighbors exist. This is the idea of the false neighbor method by the authors Kennel, Brown & Abarbanel [Kennel.M.B, Brown.R, et.al., 1992].

For each point of the time series, take its closest neighbor in \( d_0 \) dimensions. Then compute the ratio of distances between these two points in \( d_0 + 1 \) dimensions and \( d_0 \) dimensions. If this ratio is larger than a threshold ‘\( r \)’, the neighbor was false. The process is repeated by increasing the dimension and the percentage of false nearest neighbors will drop from 100% to zero when the proper dimension ‘\( d \)’ is reached. Further it will remain zero from then onwards, since once the attractor is unfolded, it is unfolded.

But the criterion in this approach is subjective in some sense that, different values of parameters may lead to different results [Cao.L, 1997)]. For realistic time series data, different optimal embedding dimensions are
obtained if we use different values of the threshold value. Also with noisy
data this method gives spurious results. [Kantz.H and Schreiber.T, 1997].

7.5.2 Minimum Embedding Dimension using Cao’s Method

Consider a time series $x_1, x_2, \ldots \ldots, x_N$. The time-delay vectors can be
reconstructed as follows:

$$y_i(d) = (x_i, x_{i+\tau}, \ldots, x_{i+(d-1)\tau}), \quad i = 1, 2, \ldots, N-(d-1)\tau$$

where $d$ is the embedding dimension and $\tau$ is the time-delay. Note that $y_i(d)$
means the $i^{th}$ reconstructed vector with embedding dimension $d$. Similar to the

$$a(i, d) = \frac{||y_i(d+1) - y_{n(i,d)}(d+1)||}{||y_i(d) - y_{n(i,d)}(d)||}, \quad i = 1, 2, \ldots, N-d\tau$$

where $|| \|$ is measurement of Euclidean distance, $y_i(d+1)$ is the $i^{th}$
reconstructed vector with embedding dimension $d+1$, i.e., $y_i(d+1) = (x_i, x_{i+\tau}, \ldots, x_{i+d\tau})$ and $n(i,d)$, $(1 \leq n(i,d) \leq N-d\tau)$ is an integer such that $y_{n(i,d)}(d)$ is the
nearest neighbor of $y_i(d)$ in the m dimensional reconstructed phase space in
the sense of Euclidean distance $|| \|$.

In false nearest neighbors method, the authors diagnosed a false
neighbor by seeing whether the $a(i, d)$ is larger than some given threshold
value. The problem is how to choose this threshold value. It is very difficult
and even impossible to give an appropriate and reasonable threshold value,
which is independent of the dimension $d$ and each trajectory's point, as well as the considered time series data.

To avoid the above problem, in Cao’s method, the mean of all $a(i,d)$s is defined by,

$$E(d) = \frac{1}{N-d\tau} \sum_{i=1}^{N-d\tau} a(i, d)$$

$E(d)$ is dependent on only the dimension $d$ and the lag $\tau$. To investigate its variation from $d$ to $d+1$, a quantity is defined by

$$E_1(d) = \frac{E(d+1)}{E(d)}$$

It is found that $E_1(d)$ stops changing when $d$ is greater than some value $d_0$. Then $d = d_0 + 1$ is the minimum embedding dimension we are looking for.

The parameter $\tau$ is a necessary parameter, which must be given before the minimum embedding dimension is determined numerically. Although in principle the embedding dimension is independent of the time delay $\tau$, the minimum embedding dimension is dependent on $\tau$ in practice. Different values of $\tau$ may lead to different minimum embedding dimensions. We will use the method of Maximum Euclidian Distance Measure (MEDM) from the phase space diagonal to choose the parameter $\tau$ in this work, on account of reasons already explained. The algorithm is implemented using Matlab and the experimental results are shown in the figures 7.10(a-e).
Fig. 7.10(a): variation of $E_1(d)$ with dimension $d$ for Malayalam vowel $\mathfrak{m}/\Lambda/$.  

Fig. 7.10(b): variation of $E_1(d)$ with dimension $d$ for Malayalam vowel $\mathfrak{m}/l/$. 
Fig. 7.10(c): variation of $E_1(d)$ with dimension $d$ for Malayalam vowel $\text{ae}$.

Fig. 7.10(d): variation of $E_1(d)$ with dimension $d$ for Malayalam vowel $\text{o}$. 

180
From these figures it may be noted that $E_1(d)$ almost remains a constant when the embedding dimension is greater than 2, when time lag is selected by the method of MEDM. That is the minimum embedding dimension according to this result is 3. Therefore in the further analysis we have used the minimum embedding dimension for the reconstruction of phase space as 3.

7.6 RPS Features with Optimum Embedding Parameters

The concepts of minimum embedding dimension ‘$d$’ and the optimum time lag ‘$\tau$’ play a significant role in both the theoretical and practical aspects of Reconstructed Phase Spaces. In the previous session, we found that as far
as speech signal is concerned, the optimum time lag corresponding to the Maximum Euclidian Distance Measure (MEDM) from the phase space diagonal is a good choice. The heuristic procedures of false nearest neighbors for determining minimum embedding dimension contain subjective parameters. Cao’s algorithm is a solution to avoid such factors. Therefore we have optimized the parameters for reconstructing the Phase Space based on the above mentioned algorithms. In chapter 5 we have described an algorithm for extracting useful parameters from the Reconstructed Phase Space (RPSDP). In the next session modified Reconstructed Phase Space features are extracted with the optimum embedding parameters.

For each vowel, optimum time delay is determined using MEDM from the phase space diagonal. From Cao’s algorithm, we have got the minimum embedding dimension as 3. Hence phase space is reconstructed for each vowel with dimension 3 and optimum time delay. RPS distribution plot (scatter graph) for five Malayalam vowels are shown in figures 7.11(a-e).
Fig. 7.11(a): Reconstructed Phase Space Distribution for vowel əə/Λ/

Fig. 7.11(b): Reconstructed Phase Space Distribution for vowel ə/Ī/
Fig. 7.11(c): Reconstructed Phase Space Distribution for vowel /æ/.

Fig. 7.11(d): Reconstructed Phase Space Distribution for vowel /o/.
7.6.1 Modified RPS Distribution Parameter

To extract the distribution parameter from the three dimensional Reconstructed Phase Space, the entire three dimensional space is divided into 1000 locations. The number of Phase Space Points distributed in each location is calculated as follows.

RPS is divided into grids with 10 x 10 x 10 boxes. The box defined by co-ordinates (-1, .8, -.8), (-.8, 1, -1) is taken as location 1. Box just right side to it is taken as location 2 and it is extended towards X direction, with the last box in the row (.8, .8, -.8), (1, 1, -1) as location 10. This is repeated for the next row, taking the starting box as location 11 and repeated for all other rows. Then the entire steps are repeated in positive Z direction. The Reconstructed Phase Space Distribution Parameter (RPSDP) is calculated by

Fig.7.11(e): Reconstructed Phase Space Distribution for vowel ñ/u/
estimating the number of Phase Space points distributed in each location.

Figures 7.12(a-e) show the Modified Reconstructed Phase Space Distribution Parameter versus locations for the vowels /A/, /I/, /ae/, /o/, and /u/.

**Fig. 7.12(a)**: Modified RPS Distribution Parameter (Vowel /A/)

**Fig. 7.12(b)**: Modified RPS Distribution Parameter (Vowel /I/)
Fig. 7.12(c) : Modified RPS Distribution Parameter (Vowel $\alpha$/ae/)

Fig. 7.12(d) : Modified RPS Distribution Parameter (Vowel $\theta$/o/)
This operation is repeated for the same vowel uttered at different occasions. Figures 7.13(a-e) show the Modified Reconstructed phase space distribution parameters for each vowel uttered at different occasions. The graph thus plotted for different vowels shows the identity for each vowel as regard to pattern. Therefore this technique can be effectively utilized for speech recognition applications.
Fig. 7.13(a): Modified RPS Distribution Parameter for 15 repeated utterances (Vowel ι/Λ/ι)
Fig.7.13(b) : Modified RPS Distribution Parameter for 15 repeated utterances
(Vowel əː / I/)
Fig. 7.13(c): Modified RPS Distribution Parameter for 15 repeated utterances
(Vowel $\alpha$ /ae/)
Fig. 7.13(d): Modified RPS Distribution Parameter for 15 repeated utterances
(Vowel η/o/)

<table>
<thead>
<tr>
<th>Graph 1</th>
<th>Graph 2</th>
<th>Graph 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graph 4</td>
<td>Graph 5</td>
<td>Graph 6</td>
</tr>
<tr>
<td>Graph 7</td>
<td>Graph 8</td>
<td>Graph 9</td>
</tr>
<tr>
<td>Graph 10</td>
<td>Graph 11</td>
<td>Graph 12</td>
</tr>
<tr>
<td>Graph 13</td>
<td>Graph 14</td>
<td>Graph 15</td>
</tr>
</tbody>
</table>

192
Fig. 7.13(e): Modified RPS Distribution Parameter for 15 repeated utterances
(Vowel อ/อ/)
7.7 Conclusion

The problem of choosing the optimal time delay and the minimum embedding dimension for the reconstruction of phase space using the method of delays are addressed in this chapter. From the discussions of the methods of determining proper time delay, it can be concluded that the optimal delay depends upon the details of the time series as well as the dynamics of the underlying system. We developed a simple procedure that quantifies expansion from the identity line of embedding space. Such a procedure may be more useful in the estimation of the proper time delay. For determining the minimum embedding dimension we have used Cao’s method, which does not contain any subjective parameters except time delay for the embedding and is computationally efficient. This method gives consistent results with Malayalam vowels. With these optimum-embedding parameters, Reconstructed Phase Space Distribution Parameter (RPSDP) is modified, whose discriminatory power is illustrated in the next chapter.