ANALYSIS OF TIME DELAY AND EMBEDDING DIMENSION OF RECONSTRUCTED PHASE SPACE OF HUMAN VOCAL TRACT USING MALAYALAM VOWELS

[1]K. M. Muraleedharan,[2]R. K. Sunil Kumar,[3]Bibish Kumar K. T.,[4]Sunil John

[1]Assistant Professor,[2]Assistant professor,[3]Research Scholar,[4]Assitant Professor
[1]Computer Speech & Intelligence Research Centre, Department of Physics
[1]Govt. College, Madappally, Vadakara, Calicut, India

Abstract : Aim of this work is to optimise the time delay and embedding dimension for the process of phase space reconstruction from the time series of speech signals to analyze nonlinear behaviour of human vocal tract. According to Taken's theorem, the underlying dynamics of a non linear deterministic system can be identified from the Reconstructed Phase Space (RPS) of the system, if the phase space is reconstructed with optimum delay and dimensions. The delay and dimensions of RPS are optimised using five Malayalam vowel phonemes uttered by fifty female and fifty male speakers of the three different age groups. Each vowel sound is recorded ten times and time series corresponding to each sound is recorded in two sampling frequencies 16 kHz and 44.1 kHz. The optimum time delay for RPS of speech time series is estimated using Mutual Information (MI) method and the optimum dimension is determined using False Nearest Neighbour (FNN) method. The analysis shows that the optimum delay hasn't any noticeable dependence on gender and age. But the time delay increases almost linearly with sampling frequency. The embedding dimension of the reconstructed attractor is found to be same for all vowel utterances for all speakers at the studied sampling frequencies. The embedding dimension is directly related with the actual state space dimension of the vocal tract system and it will not vary with age, gender and sampling frequency. The feature extraction in the optimized delay and dimension from reconstructed phase space will help in a better understanding of the inherent non linear properties of vocal tract system. Further studies need to be carried out by reconstructing phase space with the optimised time delay and embedding dimension and to model the speech samples to analyse the dynamical behavior of vocal tract system.

Key words: Non linear dynamics-Reconstructed phase space-Mutual Information-False Nearest Neighbour.

I. INTRODUCTION

Speech signal pattern is determined by the configuration of human vocal tract which depends on the various speech articulators such as tongue, lips, jaw, velum and larynx. It is the behaviour of the articulators over time that produces continually varying acoustics. The listener extracts information about the speech production from the speech signal [1]. Since speech production system is essentially a deterministic chaotic system, methods based on nonlinear dynamics can be used to quantify chaos and irregularity in the vocal fold vibrations [2-6].

The phase space of a nonlinear system can be reconstructed from the time series generated from the nonlinear system. Various studies have been carried out in this direction in distinct biological and physical systems [7-9]. The feature extraction from the time series helps to determine the nature of nonlinear systems. Taken's theorem states that the phase space should be reconstructed with optimum delay and dimension so that the feature extracted from the time series will precisely quantify the system. Even though various feature extraction studies have been carried out to model the speech signal, the delay and dimension were taken on a trial and error basis [10-13].

If the phase space is reconstructed with optimum delay and dimension the features extracted from the phase space will help in improved speech identification, speaker identification, speech analysis and speech pathology. In this context, the Malayalam vowel samples are analyzed for optimizing delay and dimension of the reconstructed phase space(RPS) of speech signal. The purpose of present study is to analyze the change in the optimum delay and dimension of RPS with respect to different speech utterances and speakers

The enhancement in speech and speaker identification of Malayalam vowels using non linear parameters has been studied a number of times. But the dimension and delay is not yet optimised for this problem. This study gives a detailed analysis on time delay and embedding dimension in various samples including gender, age and sampling frequency. The detailed knowledge about the optimum time delay and embedding dimension will help to improve the identification of non linear behavior of human vocal tract system.

II. Reconstructed Phase space (RPS) using Time delay embedding

The accurate analytical characterization of a nonlinear system is impossible if the dimensionality and non-linearity of the system are high. However time-delay embedding can be used to reconstruct the phase space of the system and to draw useful information about the system. A reconstructed phase space can be exploited as a powerful signal processing tool especially when the dynamical system of interest is non-linear or even chaotic. In the case of a speech signal, phase space of the system can be reconstructed with the use of time delayed versions of the original scalar measurements [18].

2.1. Phase space reconstruction from time series

A reconstructed phase space (RPS) can be produced for a measured state variable Sn, n=1, 2, 3, 4 ….N, via the method of delays by creating vectors given by

$$Sn= [Xn, Xn+\tau ,Xn+2\tau\tau, \ldots ,Xn+ (m-1)\tau] \quad (1)$$

Where m is the embedding dimension and τ is the time delay value. The row vector Sn defines the position of a single point in the RPS. A speech signal can be treated as a one dimensional time series data. The row vectors then can be compiled in to a matrix called trajectory matrix to completely define the dynamics of the system and can create a reconstructed phase space

2.2. Importance of optimum delay and dimension

The quality of reconstruction of phase space mainly depends on the value chosen for time delay and there is no standard criteria found reported in the literature for choosing an optimum time delay. If the chosen time delay is too short, the coordinates used in reconstructed feature vector will not be independent enough and the trajectories of the RPS are squeezed along the identity line. Very high delay values leads to loss of information from the signal.

The features extracted from RPS will give reliable information only if the phase space is reconstructed with optimum dimension. If the dimension is less than the actual dimension lots of information regarding the system will be lost. The various feature vectors including Lyapunov exponent, Box counting dimension and Kolmogorov entropy of the two dimensional phase space of Malayalam vowels are analyzed and found that these features are not sufficient to describe the system. The higher dimensionality than the optimum value will cause increase in computational complexity and may lead to wrong interpretations.

III. Methods of estimating optimum delay and dimension

Mutual information supplies a measure of general dependence and is expected to provide a better measure of the shift from redundancy to irrelevance with non-linear systems. Mutual information is associated with the least redundancy and, therefore is the best for attractor reconstruction [14]. False nearest neighbour method is suitable for obtaining the embedding dimension of a short time series [17].

3.1. Mutual information

The mutual information, $I(Z,Y)$,which is a measure of the minimum uncertainity in time series ‘ z’ (when measurement of series ‘y ‘ is given)is

$$I(Z,Y) = \sum_{i,j} p_{yz}(y_i, z_i)\log[\frac{p_{yz}(y_i,z_i)}{p_y(y_i,z_i)p_z(z_j)}] \quad (2)$$

For a time series x(t), obtained from speech signal, the dependency of the values of $x(t+\tau)$ on the values of $x(t)$ can be determined by making the assignment $[y,z]= [x(t), x(t+\tau)]$. For speech time series x(t), the average mutual information between $x(t)$ and$x(t+\tau)$ can be stated as

$$I(\tau) = \sum_{x(t),x(t+\tau)} p(x(t), x(t+\tau)\log[p(x(t), x(t+\tau)/p(x(t)p(x(t+\tau)] \quad (3)$$

If the measurements x(t) and x(t+τ) are independent, then I(τ) will tend to zero. The delay at which first minimum of the average mutual information occurred is then taken as the delay for reconstruction of phase space. [14-16]

3.2. False Nearest Neighbours-FNN

FNN gives the fraction of false neighbours which has a strong relationship with the distances in between samples reconstructed in m-dimensional spaces. If the fraction for a given m is low a better reconstruction in m dimensions will be obtained. [17] This will helps to analyze how the dataset is reorganized and its information reflected according to the number of dimensions. When the fraction of false neighbour reaches zero, no significant modification in the distance of the points can be identified. Thus, there is enough information to unfold and understand the behaviour of studied system. The fraction of false nearest neighbours (Cd) for a given dimension d with time delay τ is defined in equation 4

$$C_d = \frac{\sum_1^{t-1} \sum_{j>1}^{t} sign[V_{i,j}^d - R_{tol}]}{\frac{\tau(\tau-1)}{2}} \quad (4)$$

IV. Experimental results and discussion

The speech database used in this study include five Malayalam vowels /a/, /i/, /e/, /o/, /u/ uttered by 50 male and 50 female speakers of the three age groups.(1)Between 5 and 10.(2).Between 20 and 25.(3).Between 60 and 65.The study is carried out with different sampling frequencies(16kHz and 44.1kHz). Each vowel sample is uttered 10 times by each speaker.

4.1 Embedding time delay

The variation of mutual information function with delay for each vowel is studied for optimising the time delay required for reconstructing phase space. Figure 1,Figure 2 and Figure 3 respectively show the variation of mutual information with delay for the five Malayalam vowels */a/, /i/, /e/, /o/, /u/* for male speakers of the age range(1) 5-10 sampled at frequency 16kHz (2) 20-25 sampled at frequency 16kHz and (3) 60-65 sampled at frequency 16 kHz.Figure 4 and Figure 5 show the variation of mutual information with time delay for female speaker vowel utterance sampled at frequencies 16 kHz and 44.1 kHz respectively. The average time delays for all speech samples with its standard deviation are tabulated in Table 1 and Table 2.

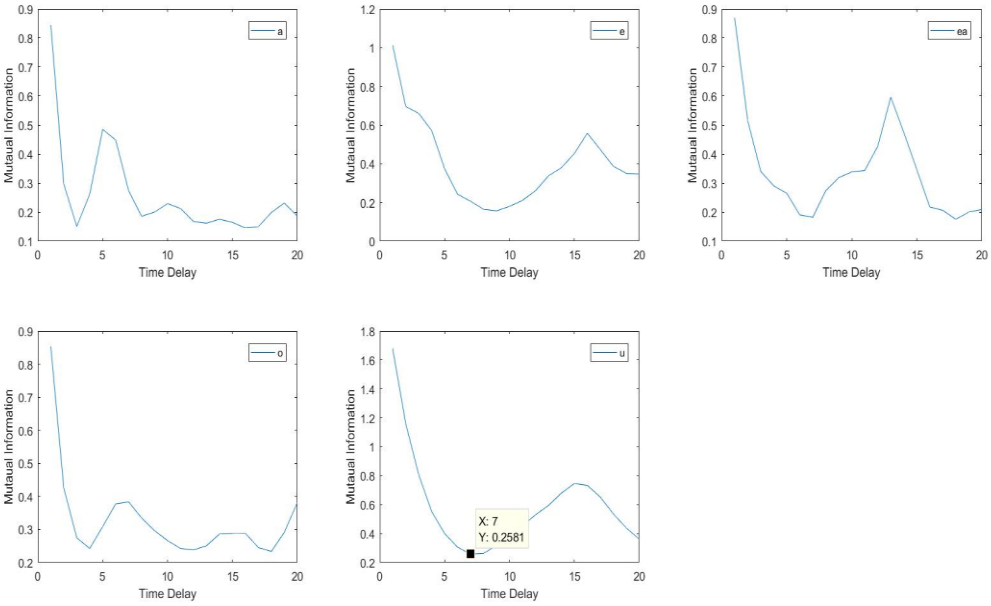

Fig. 1: Variation of mutual information with time delay for five vowels (Male speaker, age 5-10, sampling frequency16 kHz)

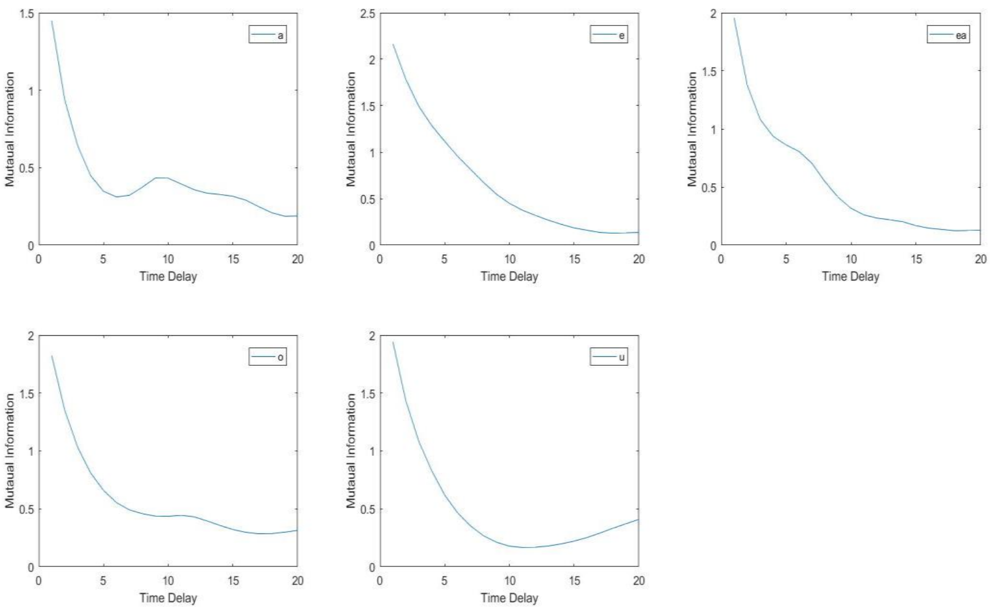

Fig. 2: Variation of mutual information with time delay for five vowels (Male speaker, age 20-25, sampling frequency16 kHz)

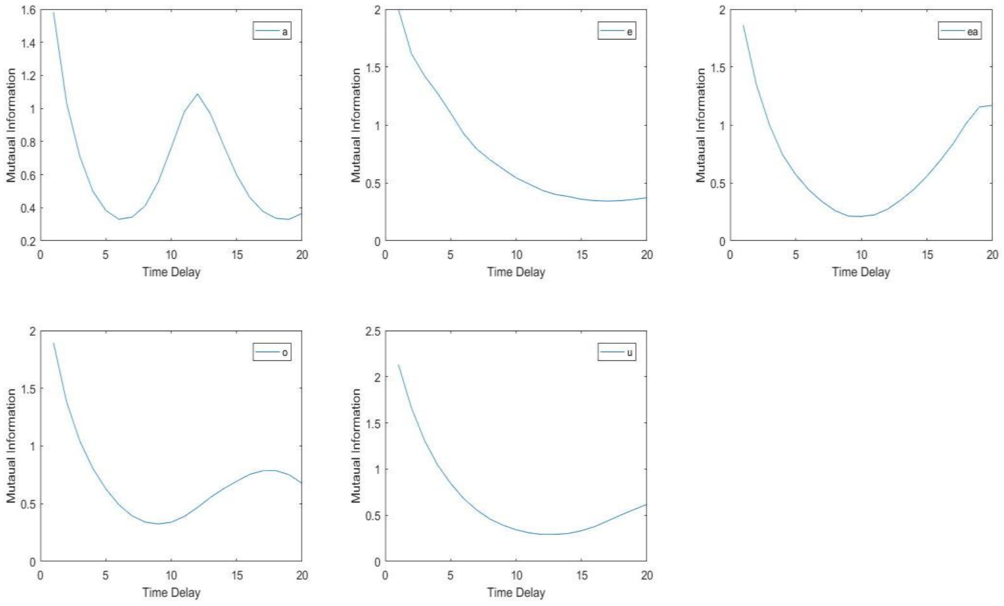

Fig. 3: Variation of mutual information with time delay for five vowels (Male speaker, age 60-65, sampling frequency16kHz)

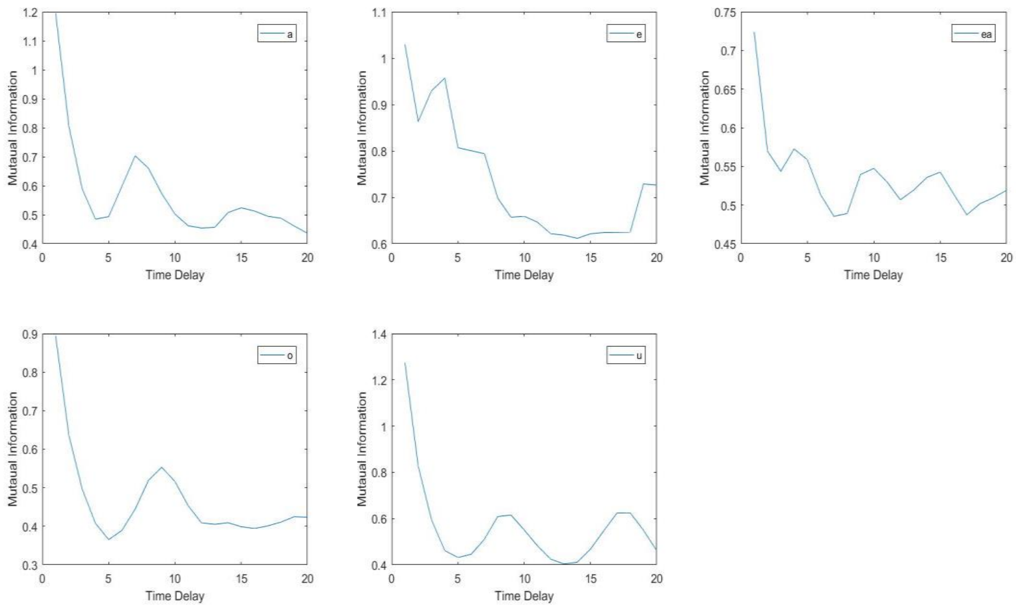

Fig. 4: Variation of mutual information with time delay for five vowels (Female speaker, age 20-25, sampling frequency16kHz)

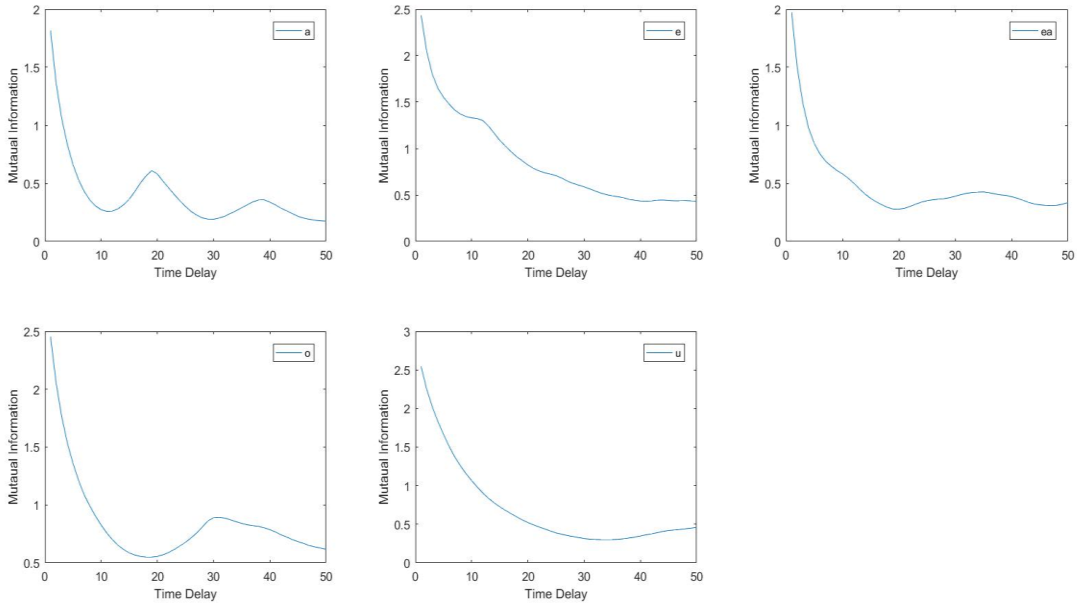

Fig. 5: Variation of mutual information with time delay for five vowels(Female speaker, age 20-25,sampling frequency16kHz)

Table 1: Average time delays for male vowel sounds

Sample	50 Male sounds(Number of samples for each vowel for each age group and frequency is 500)											
	Sampling Frequency 16 kHz						Sampling Frequency 44.1 kHz					
Age	05-10		20-25		60-65		05-10		20-25		60-65	
Vowel	Average Time Delay	Standard Deviation (SD)	Average Time Delay	SD	Average Time Delay	SD	Average Time Delay	SD	Average Time Delay	SD	Average Time Delay	SD
/a/	4	0.68	4	0.56	5	0.54	10	0.45	11	0.41	10	0.61
/e/	2	0.96	3	1.21	3	1.2	6	0.95	5	0.87	5	1.02
/ea/	5	0.54	5	0.35	6	0.85	12	0.52	13	0.54	13	0.59
/o/	5	0.32	6	0.29	5	0.74	13	0.41	15	0.61	14	0.24
/u/	6	0.33	7	0.35	7	0.69	15	0.32	12	0.34	13	0.38

Table 2: Average time delays for Female vowel sounds

Sample	50 Female sounds(Number of samples for each vowel for each age group and frequency is 500)											
	Sampling Frequency 16 kHz						Sampling Frequency 44.1 kHz					
Age	05-10		20-25		60-65		05-10		20-25		60-65	
Vowel	Average Time Delay	Standard Deviation (SD)	Average Time Delay	SD	Average Time Delay	SD	Average Time Delay	SD	Average Time Delay	SD	Average Time Delay	SD
/a/	5	0.64	4	0.66	5	0.34	11	0.45	12	0.41	12	0.61
/e/	3	0.86	3	0.98	4	0.87	7	0.95	6	0.89	8	1.05
/ea/	6	0.34	6	0.35	5	0.55	14	0.42	13	0.34	13	0.69
/o/	5	0.42	5	0.29	6	0.45	14	0.31	15	0.51	14	0.34
/u/	5	0.36	6	0.35	7	0.38	14	0.22	16	0.36	15	0.35

From Table 1 and Table 2, it is evident that the time delay values for a particular utterance and particular speaker cannot generalized. The delay values have considerable standard deviation especially for some utterance /e/.The specific dependence on gender and age is not visible because there are random fluctuations in the delay values even for same utterance of different speakers and different utterance of same speaker. But it is seen that time delay values depends highly on sampling frequency. As the sampling frequency increases time delay shows an almost linear increase.

4.2 Embedding dimension

The embedding dimension of the RPS is determined using the method of false nearest neighbors. The time delays optimized using mutual information function is used in calculating embedding dimension. The embedding dimension at which FNN falls below 1% is taken as the actual embedding dimension. Figure 6 and Figure 7 respectively show the variation of FNN for Malayalam vowel utterance of female speaker at sampling frequencies 16kHz and 44.1kHz.Table 3 shows the average embedding dimension and standard deviation of all samples at the different sampling frequencies used for the study.

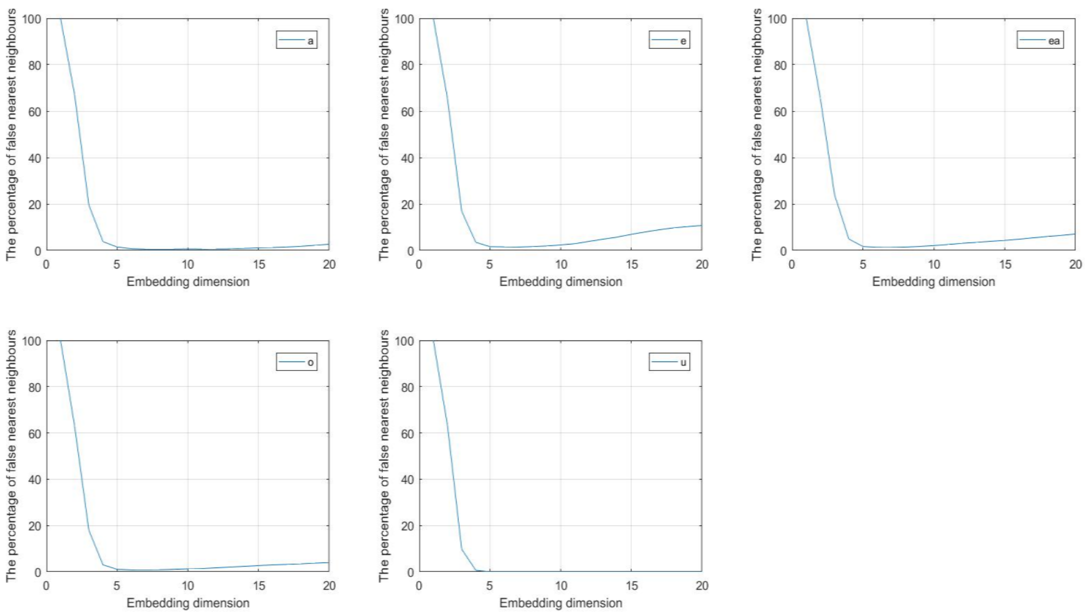

Fig. 6: Variation of FNN with Embedding dimension (Female speaker, sampling frequency 16 kHz)

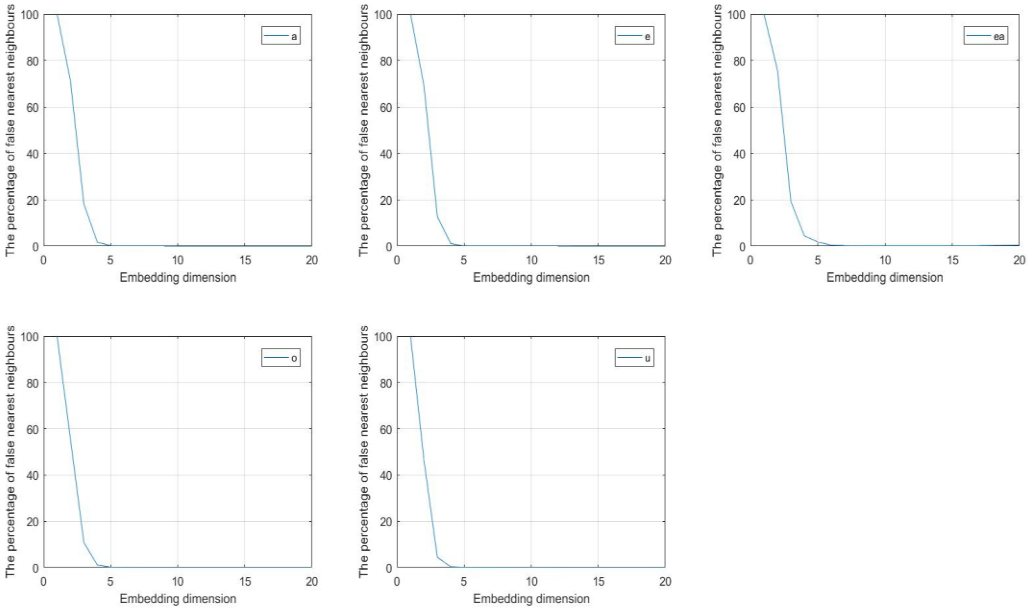

Fig. 7: Variation of FNN with Embedding dimension (Female speaker, sampling frequency 44.1 kHz)

Table 3: Average embedding dimension of speech vowel samples

Sampling Frequency	Average Embedding dimension of all vowel samples for all speakers under study(7500 samples)	Standard deviation
16kHz	5	0.31
44.1kHz	5	0.26

The average embedding dimension estimated is same for all Malayalam vowel utterances as shown in Table 2.It is independent of age,gender and sampling frequency. The frequency distribution of the estimated dimension values for all samples is shown in Figure 8.It is evident from the figure that embedding dimension almost show a Gaussian distribution about the mean value and hence the mean value is a better estimate of the optimum embedding dimension.

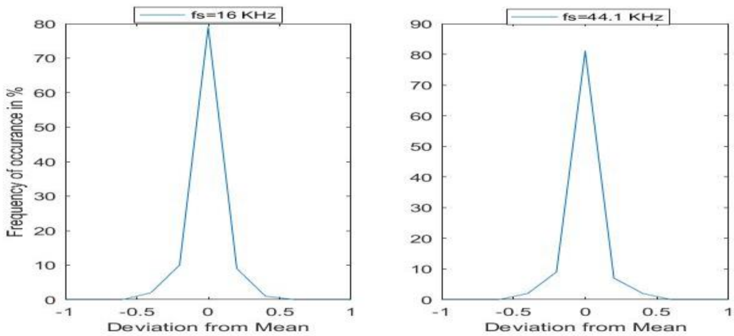

Fig. 8: Frequency Distribution of Embedding Dimension

V. CONCLUTION

The time series obtained from the speech signals are analyzed for estimating the optimum delay and dimension to reconstruct the phase space. The analysis shows that the optimum delay hasn't any noticeable dependence on gender and age. Delay shows some random variation with these parameters. But the time delay increases almost linearly with sampling frequency. Hence to reconstruct the underlying attractor delay should estimated for each sample separately and the generalization of delay values based on these parameters is very difficult. The embedding dimension of the reconstructed attractor is found to be same for all vowel utterances for all speakers at the studied sampling frequencies. The embedding dimension is directly related with the actual state space dimension of the vocal tract system and it will not vary with age, gender and sampling frequency. The feature extraction in the optimized delay and dimension from reconstructed phase space will help in a better understanding of the inherent non linear properties of vocal tract system. Further studies need to be carried out by reconstructing phase space with the optimized time delay and embedding dimension and to model the speech samples to analyze the non linear dynamical behavior of vocal tract system.

REFERENCES

[01] Rubin, Philip, and Eric Vatikiotis-Bateson. "*Measuring and modeling speech production." In Animal acoustic communication*, pp. **251-290**. Springer, Berlin, Heidelberg, 1998.

[02] Jiang, Jack J., Yu Zhang, and Clancy McGilligan. "*Chaos in voice, from modeling to measurement*." Journal of Voice **20**, no. 1 (2006): 2-17.

[03] Banbrook, Michael, Steve McLaughlin, and Iain Mann. "*Speech characterization and synthesis by nonlinear methods.*" IEEE Transactions on Speech and Audio Processing **7**, no. 1 (1999): 1-17.

[04] Kleijn, W. Bastiaan, and Kuldip K. Paliwal." *Speech coding and synthesis"*. Elsevier Science Inc., 1995.

[05] Kumar, Arun, and S. K. Mullick. "*Nonlinear dynamical analysis of speech*." The Journal of the Acoustical Society of America **100**, no. 1 (1996): 615-629.

[06] Pitsikalis, Vassilis, and Petros Maragos. "*Speech analysis and feature extraction using chaotic models.*" In Acoustics, Speech, 2002 IEEE International Conference on, vol**. 1**, pp. I-533. IEEE, 2002.

[07] Muda, Lindasalwa, MumtajBegam, and IrraivanElamvazuthi. "*Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques*." arXiv preprint arXiv:1003.4083 (2010).

[08] Harikrishnan, K. P., R. Misra, and G. Ambika. "*Efficient use of correlation entropy for analysing time series data.*" Pramana **72**, no. 2 (2009): 325-333.

[09] Harikrishnan, K. P., R. Misra, and G. Ambika. "*Combined use of correlation dimension and entropy as discriminating measures for time series analysis*." Communications in Nonlinear Science and Numerical Simulation **14**, no. 9-10 (2009): 3608-3614.

[10] Harikrishnan, K. P., R. Misra, and G. Ambika. "*Revisiting the box counting algorithm for the correlation dimension analysis of hyperchaotic time series*." Communications in Nonlinear Science and Numerical Simulation **17**, no. 1 (2012): 263-276.

[11] Vivek P Lagish VL Sunil kumar RK" *Improved vowel phoneme classification over the eigen values of reconstructed phase space using ANN*" National conference for Indian language computing 2015 -Cochin

[12] Shekofteh, Yasser, and FarshadAlmasganj. "*Feature extraction based on speech attractors in the reconstructed phase space for automatic speech recognition systems*." ETRI Journal **35**, no. 1 (2013): 100-108.

[13] Huang, Nanmu, Yu Zhang, William Calawerts, and Jack J. Jiang. "*Optimized nonlinear dynamic analysis of pathologic voices with laryngeal paralysis based on the minimum embedding dimension*." Journal of Voice **31**, no. 2 (2017): 249-e1.

[14] Fraser, Andrew M., and Harry L. Swinney. "*Independent coordinates for strange attractors from mutual information*." Physical review A **33**, no. 2 (1986): 1134.

[15] Pompe, Bernd. "*Measuring statistical dependences in a time series*." Journal of Statistical Physics **73**, no. 3 (1993): 587-610.

[16] M Palus, "*Testing for nonlinearly using redundancies qualitative and quantitative aspects* "physica D **80**, 186 (1993)

[17] Kennel, Matthew B., Reggie Brown, and Henry DI Abarbanel. "*Determining embedding dimension for phase-space reconstruction using a geometrical construction*." Physical review A **45**, no. 6 (1992): 3403

[18] Packard, N.H., Crutchfield, J.P., Farmer, J.D. and Shaw, R.S., 1980. *Geometry from a time series. Physical review letters*, ***45***(9), p.712.