Enhancing the Feature Extraction Process with Nonlinear Metrics for Malayalam Vowels Recognition

Fathima Kunhi Mohamed
Research Scholar
Farook College

***Abstract*—This paper presents a method for the speech recognition of 5 Malayalam vowels using nonlinear features, maximal Lyapunov exponent and Kolmogorov entropy. Reconstructed phase space concept is used in this method. The results show that phase space features contain significant discriminatory power. The results also indicates that nonlinear feature and MFCC used in combination within a classifier will yield increased accuracy for vowels recognition task.**

***Index Terms*—embedding dimension, Lyapunov exponent, Kolmogorov entropy, Nonlinear.**

I. INTRODUCTION

The human-machine interaction has made possible with the speech recognition system technology. Speech recognition is a special case of pattern recognition and so the accuracy of speech recognition system is affected by several factors. Researchers are continuously working in this area over the last few decades to increase the accuracy of the speech recognition system [1].

The foundation of most of the speech processing applications is the source-filter model and this model has headed speech recognition to great advances in the last 30 years. But it is found that this model disregards some structure known to be present in the speech signal thereby reducing the ability to discriminate speech sounds. The replacement of the linear model with nonlinear models must enable us to obtain an accurate description of the speech which in turn may result in better performance of speech recognition systems [2][3]. In literature, there is strong theoretical and experimental evidence for the existence of important nonlinear 3D fluid dynamics phenomena during the speech production that cannot be accounted for by the linear model such as modulations of the speech airflow and turbulence [4][5][6].

Human-machine interaction in one's native language is always a research area. It is especially important in a multi-lingual country such as India, where a large majority of the people will not be comfortable with communicating in English. Extensive research works for developing systems that enable human machine interaction in Indian languages for Hindi, Bangla, Telugu, Tamil, Kannada have been reported. The recognition of Malayalam (one of the South Indian Languages ranked as the eighth in the list of fifteen most popular languages in India) speech [7][8][9][10] has been studied by many re-searchers. But the works with the emphasis on recognition systems with nonlinear speech parameters is very few in number.

Several directions exist for the investigation of speech nonlinearities. In the proposed work, nonlinear signal processing approach is used. This approach consists of developing efficient computational models for detecting nonlinear phenomena in speech and extracting related acoustic signal features such as maximal Lyapunov exponent and Kolmogorov entropy from the 5 vowels given below.

അ ഇ ഉ എ ഒ

II. PHASE SPACE RECONSTRUCTION

Chaotic behavior of dynamical system is described either by non-linear mathematical equation or by experimental data. Since we don't have such an equation, we normally captures the important properties of the system by using the experimental data. Phase space reconstruction techniques forms the fundamental principles of dynamical system theory. It have been applied to a variety of time series analysis and nonlinear signal processing applications [11][12][13].

A reconstructed phase space is created by establishing vectors in R^m, the elements are time-lagged versions of the original time series as given in the equation,

$$\bar{x} = \{x[n], x[n-\tau], \ldots, x[n-(m-1)\tau]\}$$

Embedding dimension, τ is the time lag, and n is the time index.

The first step in the analysis of scalar time-series data from a nonlinear dynamical system is to embed it. Tisean is used for this purpose. It is a software project for the analysis of time series with methods based on the theory of nonlinear deterministic dynamical systems, or chaos theory [14]. The time delayed mutual information was proposed by Fraser and Swinney [15] as a tool to determine a reasonable delay. One way to do this is to use Tisean's mutual command, plot the results, and look for the first minimum in the curve. It is found to be 4 from the plot.

.

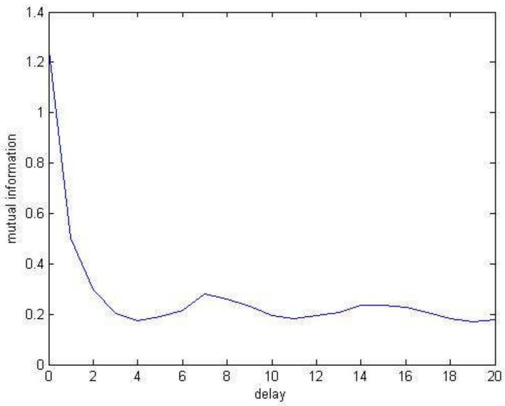

Figure 1. The Figure 1. The mutual information as a function of the time delay τ.

The second step is to estimate the embedding dimension m. A method to determine the minimal sufficient embedding dimension *m* was proposed by Kennel et al. [16]. It is called the *false nearest neighbor* method, which can be done with Tisean's false nearest command. According to it, the embedding dimension m where the number of false near neighbors falls to 10% is taken. It is found to be 5 from the plot below.

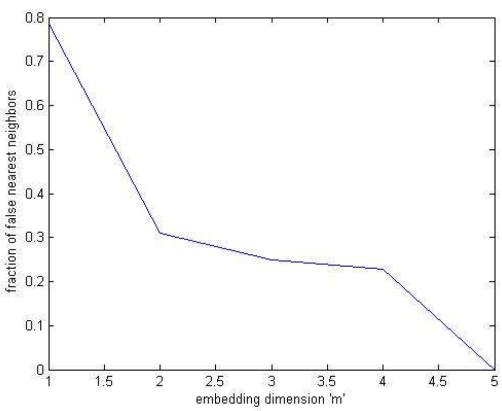

Figure 2. The fraction of false nearest neighbors as a function of the embedding dimension

The Reconstructed Phase Space Distribution plots in two dimensions are constructed for five vowels / അ/, / ഇ /, / ഉ /, / എ /, and / ഒ / as follows.

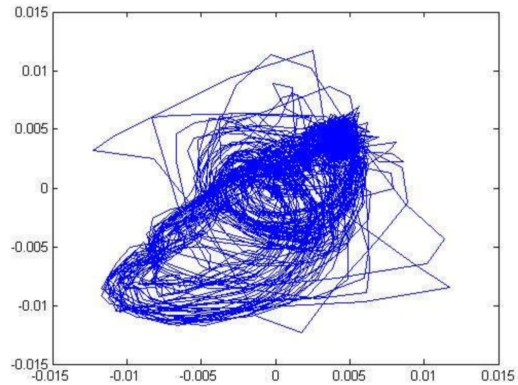

Figure 3. Reconstructed Phase Space (RPS) for Malayalam vowel അ with τ=4 and m=2.

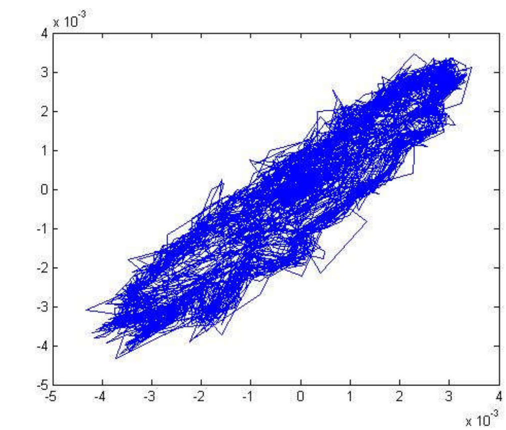

Figure 4. Reconstructed Phase Space (RPS) for Malayalam vowel ഇ with τ=4 and m=2.

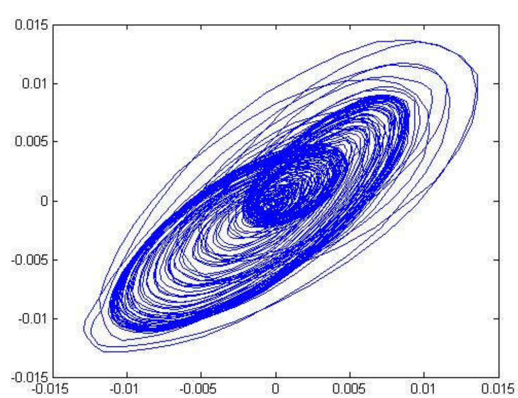

Figure 5. Reconstructed Phase Space (RPS) for Malayalam vowel ഉ with Ʈ=4 and m=2.

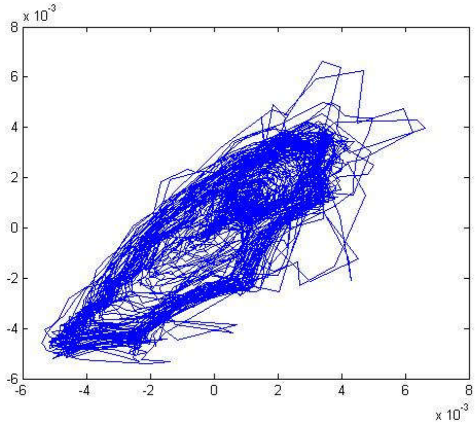

Figure 6. Reconstructed Phase Space (RPS) for Malayalam vowel എ with Ʈ=4 and m=2.

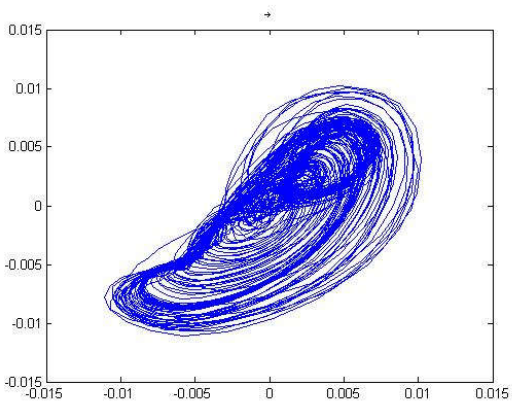

Figure 7. Reconstructed Phase Space (RPS) for Malayalam vowel ഒ with Ʈ=4 and m=2.

III. LYAPUNOV EXPONENT

Lyapunov exponents provide a quantitative and qualitative characterization of dynamical behavior. It is a measure for describing a dynamical system that remains intact by the embedding procedure and a measure of degree of chaos. They measure the exponential rate of divergence or convergence of orbits [17][18].

If we consider two neighboring points in the phase space at time 0 and at time t, the Lyapunov exponent (LE) is then defined by the average growth rate:

$$\lambda_i = \lim_{t\to\infty} \frac{1}{t} \log_2 \frac{||\partial x_i\,(t)||}{||\partial x_i\,(0)||}$$

A negative value indicates convergence of nearby orbits. A positive LE means divergence of nearby orbits. For a conservative system, the sum of Lyapunov exponents is negative, so that the orbits are bounded. Lyapunov exponents can be obtained from the time series. In this paper, maximal Lyapunov exponent is calculated using Kantz method [19][20][21] with time delay 4 and embedding dimension 5.

IV. *Kolmogorov Entropy*

The concept of the entropy is fundamental for the study of statistical mechanics and thermodynamics. Entropy is a thermodynamic quantity describing the amount of disorder in the system.

$$K \leq \sum_{\lambda^{(i)}>0} \lambda^{(i)}$$

Entropy is a measure of extent one is able to predict the next step in the sequence. The Kolmogorov-Sinai-entropy is an entropy which is a rich generalization of Shannon entropy. It measures the unpredictability of a dynamical system. For a system with higher unpredictability, entropy is higher [22]. There is a relationship between the Kolmogorov entropy K and the Lyapunov exponents λ(i). The sum of positive exponents gives the Kolmogorov entropy [23].

V. EXPERIMENTAL RESULTS

Present study investigates the vowels recognition capabilities of the method using Lyapunov exponent and feed forward multilayer perceptron (FFMLP).

The maximal Lyapunov exponent and Kolmogorov entropy are extracted from five vowels for the vowels recognition. The number of input layer is fixed according to the feature vector size and the five output nodes are present representing the five vowels. The experiment is repeated by changing the number of hidden layers and the number of iteration in order to obtain the successful architecture.

The network is trained with the maximal Lyapunov exponent and MFCC features extracted from the vowels samples. The recognition accuracy obtained for vowels classification based on above said features using FFMLP classifier are tabulated in table 1.

Table 1. Vowels Recognition Results.

Parameters used	**Size of the parameter**	**Vowels recognition accuracy (%)**
Maximal Lyapunov exponent	1	22.67
Kolmogorov Entropy	1	21.56

.

MFCC	12	74.39
Maximal Lyapunov exponent+MFCC	13	78.30
Kolmogorov Entropy + MFCC	13	75.50

Even though the experimental results indicate that the accuracy of this approach by itself below that of MFCC features, the combined feature gives considerable increase in the vowels classification accuracy.

VI. CONCLUSION

In this work, we modelled the vowel classification based on the non-linear properties of the speech samples. The experimental results indicate that the nonlinear feature accuracy of nonlinear features MLE and KE are 22.67% and 21.56 respectively. Even if the accuracy is below that of MFCC (74.39%), the combined approach offers improvement in vowels classification accuracy. From these results, it is inferred that nonlinear features have considerable discriminating power in speech recognition.

REFERENCES

[1] Anusuya, M. A., & Katti, S. K. (2010). Speech recognition by machine, a review. *arXiv preprint arXiv:1001.2267*.

[2] Faundez-Zanuy, M., S. McLaughlin, Arianna Esposito, A. Hussain, Jean Schoentgen, G. Kubin, W. B. Kleijn, and Petros Maragos. "Nonlinear speech processing: overview and applications." *Control and intelligent systems* 30, no. 1 (2002): 1-10.

[3] Pitsikalis, V., & Maragos, P. (2002, May). Speech analysis and feature extraction using chaotic models. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on* (Vol. 1, pp. I-533). IEEE.

[4] Chetouani, M., Faundez-Zanuy, M., Gas, B., & Zarader, J. L. (2005). Non-linear speech feature extraction for phoneme classification and speaker recognition. In *Nonlinear Speech Modeling and Applications* (pp. 344-350). Springer Berlin Heidelberg.

[5] McLaughlin, S., & Maragos, P. (2006). Nonlinear methods for speech analysis and synthesis. *Advances in nonlinear signal and image processing*,*6*, 103.

[6] Esposito, A., & Marinaro, M. (2005). Nonlinear speech modeling and applications.

[7] Sunny, S., David Peter, S., & Jacob, K. P. (2012). Development of a Speech Recognition System for Speaker Independent Isolated Malayalam Words. *International Journal of Computer Science & Engineering Technology*, *3*(4), 69-75.

[8] Kurian, C., & Balakrishnan, K. (2009, December). Speech recognition of Malayalam numbers. In *Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on* (pp. 1475-1479). IEEE.

[9] Krishnan, V. V., Jayakumar, A., & Anto, P. B. (2008, January). Speech Recognition of Isolated Malayalam Words Using Wavelet Features and Artificial Neural Network. In *Electronic Design, Test and Applications, 2008. DELTA 2008. 4th IEEE International Symposium on* (pp. 240-243). IEEE.

[10] Thasleema, T. M., Kabeer, V., & Narayanan, N. K. (2007, December). Malayalam vowel recognition based on linear predictive coding parameters and k-nn algorithm. In *Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on* (Vol. 2, pp. 361-365). IEEE.

[11] Lindgren, A. C., Johnson, M. T., & Povinelli, R. J. (2003, April). Speech recognition using reconstructed phase space features. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on* (Vol. 1, pp. I-60). IEEE.

[12] Dimitriadis, D., Maragos, P., Pitsikalis, V., & Potamianos, A. (2002). Modulation and chaotic acoustic features for speech recognition. *Control and intelligent systems*, *30*(1), 19-26.

[13] Lajish, V. L., Kumar, S. R., & Vivek, P. (2012). Speaker identification using a nonlinear speech model and ANN. *International Journal of Advanced Information Technology*, *2*(5), 15.

[14] Hegger, R., Kantz, H., & Schreiber, T. (1999). Practical implementation of nonlinear time series methods: The TISEAN package. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *9*(2), 413-435.

[15] A. M. Fraser and H. L. Swinney, *Independent coordinates for strange attractors from mutual information*, Phys. Rev. A **33**, 1134 (1986).

[16] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, *Determining embedding dimension*

.

for phase-space reconstruction using a geometrical construction, Phys. Rev. A **45**, 3403 (1992).

[17] Friha, S., & Mansouri, N. (2007). Speech Recognition via Lyapunov Metrics Using Neuro-Fuzzy Models.

[18] Fathima Kunhi Mohamed, Lajish V L. Nonlinear Speech Analysis and Modeling for Malayalam Vowel Recognition. *Procedia Computer Science-Journal- Elsevier*. ISSN: 1877-0509. Vol. 93, 2016.

[19] Kantz, H. (1994). A robust method to estimate the maximal Lyapunov exponent of a time series. *Physics letters A*, *185*(1), 77-87.

[20] Rosenstein, M. T., Collins, J. J., & De Luca, C. J. (1993). A practical method for calculating largest Lyapunov exponents from small data sets.*Physica D: Nonlinear Phenomena*, *65*(1), 117-134.

[21] Kantz, H., & Schreiber, T. (2004). *Nonlinear time series analysis* (Vol. 7). Cambridge university press.

[22] Paul Boersma and David Weenink. Praat: doing phonetics by computer, http://www.praat.org,(Accessed on 25-December 2018).

[23] Ogunmolu, Olalekan and Gu, Xuejun and Jiang, Steve and Gans, Nicholas. Nonlinear systems identification using deep dynamic neural networks arXiv preprint arXiv:1610.01439, 2016.