Nonlinear Analysis of Arabic Vowel Phonemes Using Lyapunov Algorithm

Mohammed Mohsen Olama
Department of Applied Science, University of Arkansas at Little Rock
Little Rock, AR, 72204, USA
Email: mmolama@ualr.edu

and

Moussa Abdallah
Department of Electronics Engineering, Princess Sumaya University
P.O.Box 1438 Al-Jubaiha 11941, Amman, Jordan
moussa@psut.edu.jo Email:

ABSTRACT

It is somewhat surprising to find that certain dynamical systems can exhibit extremely complex behavior, and the resulting time series may seem entirely random. This is defined as the presence of chaos and the system that exhibits such a behavior is called chaotic system. Later, Lyapunov exponents will be shown to give a measure of how predictable a system is. A recent study by Banbrook (Banbrook, 1996) attempted to resolve this issue, by developing a robust algorithm for the calculation of Lyapunov exponents, and then applying this algorithm to a carefully collected database of Arabic stationary vowel sounds. Previous work on measuring Lyapunov exponents in speech signals is examined. Results on measuring chaos in speech are still contradictory, and some new work is presented that may help to explain why this is so

Keywords: Nonlinear dynamic systems, Lyapunov Algorithm, Arabic vowels, chaotic system.

INTRODUCTION

It is suggested that speech is a nonlinear process in terms of speech production process. This led to a large amount of speculation about whether the underlying mechanism of speech has any chaotic properties (Mann, 1998). The requirement is to determine the dimensionality of the state-space in which the speech waveform is embedded into higher dimensional space by using Takens theorem and to ensure that the reconstructed space is a non-chaotic system.

SPEECH AS A NONLINEAR DYNAMICAL SYSTEM

In nonlinear processing the state-space of a dynamical system can be reconstructed from a time series of only one observed variable by using Takens theorem. This theorem states that there will be a one to one mapping between the reconstruction and the actual attractor of the underlying system (Takens, 1980). If the embedding is carried out in a space of dimension of at least *m*, where:

$$m \geq 2d + 1 \qquad (1)$$

where *d* is the phase-space dimension of the original attractor, then this mapping is guaranteed. In practice, forming the reconstructed trajectory matrix is relatively simple, and involves moving a window of length *m* through the data to form a series of *x* vectors. So for the discrete time series $x_n = (x_0, x_1, x_2 \quad x_i,$), the trajectory matrix *X* is:

$$X = \begin{bmatrix} x_0 & x_t & \cdots & x_{(m-1)t} \\ x_1 & x_{(1+t)} & \cdots & x_{(1+(m-1)t)} \\ x_2 & x_{(2+t)} & \cdots & x_{(2+(m-1)t)} \\ \vdots & \vdots & & \vdots \end{bmatrix} \qquad (2)$$

In real systems, where noise is an issue, the reconstruction produced by time delay embedding inherits the noise of the time series, which will adversely affect the results of any subsequent analysis. The Singular Value Decomposition (SVD) embedding technique was developed to resolve this problem (Broomhead, 1986). The principle behind it is to partition the state-space into two subspaces, one containing the signal and the other the noise. As a first step, the time delay embedding matrix X is formed with t = 1 sample and a window length of *w*. *w* will be referred to as the SVD window length and is chosen to be much greater than the supposed required embedding dimension. The singular value decomposition of X is defined by (Golub, 1989):

$$X = UWV^T \qquad (3)$$

where W is diagonal and contains the singular values $w_0 > w_1 > w_2 \quad 0$, and U and V are orthogonal and contain the singular vectors associated with W. However, if the last elements of W are much smaller than their predecessors, it should be legitimate to discard the final columns, thus reducing the dimension and removing noise effects (Mees *et al*, 1987). Therefore, the reduced trajectory matrix can be written as:

$$X = X V_d \qquad (4)$$

where V_d only contains the columns of V corresponding to the significant values of W.

Lyapunov exponents provide a quantitative characterization of chaos property and are often used as an indicator of the presence or otherwise of chaos. Essentially, across the whole attractor, a positive exponent indicates the divergence of trajectories, whereas a negative exponent indicates convergence. If both positive and negative exponents are present then this indicates a strange attractor and chaos.

Calculating the exponents from a time series, when the system equations are not known, is a difficult process, due mainly to noise and short data lengths. Figure 1 shows the vowel /u:/

embedded in a 3D embedding space, in this figure τ was set to 1.875 millisecond (τ = 15 samples). Examining the phase-space reconstruction from all angles in three dimensions shows that there are no crossings of the manifold, implying a sufficient embedding according to the theory of no intersections. Therefore, vowels can be reconstructed onto some form of attractor in a low dimensional space, and so it is possible to model them using a nonlinear dynamical approach by:

$$x_{i+1} = F\{X_i\} \tag{5}$$

where F is some nonlinear mapping between a vector of previous samples X_i and the next sample x_{i+1} . This low-dimensionality characteristic is shared by all voiced sounds, but may not be true for unvoiced sounds. By Tokens theorem, there will be a one-to-one mapping between F and the underlying nonlinear system. Therefore, correctly modeling F implies that the dynamics of speech production have been captured.

LYAPUNOV EXPONENTS ESTIMATION ALGORITHM

There is still no complete conclusion to the subject of chaos within vowel sounds. A recent study by Banbrook (Banbrook, 1996) attempted to resolve this issue, by developing a robust algorithm for the calculation of Lyapunov exponents, and then applying this to a carefully collected database of stationary Arabic vowel sounds.

As with all Lyapunov exponent estimation techniques, the algorithm is complex. To produce correct results, it relies on the appropriate choice of a number of parameters, which must be determined experimentally. Figure 2 shows an overview of the functionality of the algorithm.

As can be seen, the process can be broken down into a number of distinct steps, which are described in the following:

1. *Embedding:* The speech signal is embedded into d-dimensional state-space using SVD embedding.
2. *Choose neighborhood:* The practical method to calculate Lyapunov exponents from experimental data involves constructing a d-dimensional hyper-sphere containing a set number of points in state-space, and then observing the evolution of this hyper-sphere. The hyper-sphere is constructed by finding a sufficient number of points within the locally linear space around a chosen point, x_i, and this forms the neighborhood vector set Γ_i.

   A check is made that points within Γ_i are not false nearest neighbors, by ensuring that they evolve along suitable trajectories. Now, the neighborhood matrix B_i can be formed as:

$$B_i = \begin{bmatrix} \gamma_1 - x_i \\ \gamma_2 - x_i \\ \vdots \\ \gamma_M - x_i \end{bmatrix} \tag{6}$$

   where γ_n is the nth entry of Γ_i. It is necessary to reform this matrix after a number of steps, a, since the points will become so far separated as to break the assumption of local linearity. The choice of this parameter, called the number of evolve steps, as well as the number of points in the hyper-sphere, must be made experimentally.

3. *Calculate tangent map:* If the step between B_i and the evolved neighborhood matrix, B_{i+a}, is small enough to be considered locally linear, then the eigenvalues of the tangent map between the two matrices will give the Lyapunov exponents. The tangent map, T_i, defines the linear transformation between B_i and B_{i+a}:

$$B_{i+a}^T = T_i \, B_i^T \tag{7}$$

   The evolved neighborhood matrix B_{i+a} is constructed as per B_i, but the neighborhood vector Γ_{i+a} contains points a steps on from Γ_i. Equation 8 can be solved for T_i by writing:

$$B_i^+ \, B_{i+a} = T_i^T \tag{8}$$

   where B_i^+ is the pseudo-inverse of B_i calculated by SVD.

4. *Average exponents:* The tangent map T_i gives a description of the local system dynamics. In order to calculate the global dynamics, a series of these tangent maps across the whole phase-space structure needs to be combined together. This is achieved using the SVD-factorization technique, which allows any matrix A to be written as:

$$A = UWV^T \tag{9}$$

   where W is diagonal and contains the positive singular values and U and V are orthogonal and contain the singular vectors associated with W. This allows the global Lyapunov exponents to be calculated as:

$$\lambda_i = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \log(W_j)_{ii} \qquad 1 \le i \le d \tag{10}$$

   where W_j is the jth W matrix, corresponding to the jth tangent map, and n is the total number of tangent maps.

As mentioned above, there are a number of parameters that must be selected in order to obtain correct results from the algorithm. These are:

- Number of points in hyper-sphere, which should be as large as possible without breaking the assumption of local linearity.
- Embedding dimension. This is indicated by other (e.g., correlation dimension) method, but must be experimented with.
- Size of SVD window, in order to remove any noise present.
- Re-initialization delay. This is a trade-off between a large number of steps which will show greater change, against the fact that the neighborhood will become very dispersed and no longer be locally linear.
- Total number of tangent maps, which should be as large as possible, but shall involve at least one revolution of the phase-space reconstruction.
- Number of data points, which should also be large so as to adequately fill the areas of state-space covered by the reconstruction.

The algorithm has an additional noise-reduction feature; this involves a different formulation of the B_i matrix, which allows an averaging operation to be performed. Experiments on noisy Lorenz data show a considerable improvement in the estimation of the positive Lyapunov exponent using this technique (Banbrook, 1996).

DISCUSSION AND RESULTS

Initially the recording of database was performed in general room environment (laboratory room in the department). Six subjects were recorded, three males and three females, each subject read thirty utterances (words). The subjects were requested to extend the vowel portion of each utterance. The speech data were obtained by sampling at 8 kHz and digitized at 8-bit resolution (A/D conversion). The speech waveform is firstly segmented to remove other phonemes and silence, and obtain the required phoneme alone (vowel phoneme).

Lyapunov exponents are calculated for some of the subjects in the database over a wide range of values of the parameters described above in order to arrive at the best estimates for use in the full analysis. Figure 3 shows how the exponents vary for varying SVD window length and local embedding dimension for the female speaker. Here, the stability of the results for SVD window lengths above approximately 20 samples led to the choice of 30 samples for this parameter. The local embedding dimension was chosen as three, since the first two exponents show little change for dimensions above this. This analysis was applied to all parameters in order to arrive at the final set of values. The value in all cases was found to be one zero and two negative Lyapunov exponents, leading to the conclusion that the data is not chaotic (Banbrook, 1996). This result is unchanged by any increase in local embedding dimension, which was increased from three to seven without any other significant exponents appearing.

The results described above which find that vowels are not chaotic directly contradict the findings from a number of other papers, which maintain that these signals do exhibit chaotic behavior. In the above study, there is no positive exponent found for the parameter set used (SVD window length of 30; local embedding dimension of 3). However, for shorter SVD window lengths, implying less noise reduction, a positive exponent is seen. Hence, it is important to consider what effect the process of SVD may have on the results.

EFFECTS OF SINGULAR VALUE DECOMPOSITION

The Lyapunov exponents estimation algorithm as developed by Banbrook *et al.* relies on the use of SVD to help reduce noise present in the signal. While it has been demonstrated that this gives good results on real data, some concern has been expressed that valid parts of the signal may also be removed by this process when applied to speech. This would then alter the values of the Lyapunov exponents that are calculated subsequently, and so could mask the presence of chaos. In an effort to resolve this issue, some work was carried out to investigate the effects of SVD on voiced speech signal characteristics.

As explained previously, given the SVD of the embedding matrix X, the X is the reduced trajectory matrix corresponding to the *d* most significant values of *W*. It is this matrix X that is used in the subsequent analysis, so it is now necessary to question whether any important dynamical information has been lost when discarding the other less significant values. Figure 4 shows the SVD embedding for the vowel /u:/ shown in Figure 1. Some noise reduction can be seen in the SVD case, resulting in a smoother phase-space reconstruction, but it is very difficult to assess the quality of the signal.

For this reason, it is necessary to return back to the time domain where speech-specific characteristics, such as pitch and formant structure, can be examined. As can be seen from Equation 3, the elements of X can be expressed as:

$$X_{ij} = \sum_{k=0}^{w-1} U_{ik}\, W_k\, V_{jk} \quad (0 \le i < N \,,\, 0 \le j < w) \qquad (11)$$

where N is the data length. In forming X, all w_n, n d, were discarded, so the SVD processed trajectory matrix can be written as:

$$X_{ij} = \sum_{k=0}^{d-1} U_{ik}\, W_k\, V_{jk} \quad (0 \le i < N \,,\, 0 \le j < w) \qquad (12)$$

Note that the dimensions of X are the same as that of X and not X. Since the time delay τ between subsequent values in X is one sample, the SVD processed one-dimensional time series x can be formed from the column vector X_{i0} (0 i < N). It is now possible to carry out a direct comparison between the original time series *x* and the SVD processed time series x.

The above SVD process was then applied to the data to produce a set of SVD processed signals suitable for comparison with the originals. An SVD window length of w = 30 samples was used. Across all signals, it was found that the time structure was not significantly altered, and that the pitch was unchanged. However, examining the frequency spectra shows that the higher frequencies undergo some attenuation. Figure 5 shows an example of an original time domain signal and the resulting SVD processed signal for the vowel /u:/ for a female speaker. It is evident that the general structure is preserved, although some fine details have been removed.

CONCLUSIONS

Given that this attenuation of the higher frequencies takes place, it must now be considered how this could affect any subsequent dynamical analysis, such as the calculation of Lyapunov exponents. If SVD is not used then noise may remain on the signal, and this may produce the artifact where a positive Lyapunov exponent is found, which is in fact zero. The SVD process has been shown to alleviate this problem, due to its noise reduction properties. It would appear that it is extremely difficult to find a middle ground between these two in real world applications: with a Lyapunov analysis of speech not using SVD, the system can never be guaranteed to be noise-free. However, employing SVD does attenuate the higher frequencies, and this may imply the loss of some important higher frequency information. It is argued that the best currently available Lyapunov exponent calculation techniques have been used. Neither method can give a definite answer to whether the largest Lyapunov exponent is zero, or slightly positive, but taking into account their known limitations, it would seem clear that if voiced speech is at all chaotic, then it is extremely so weakly.

REFERENCES

- Banbrook, M. 1996. *Nonlinear Analysis of Speech from a Synthesis Perspective*. Ph.D. Thesis, University of Edinburgh, Edinburgh.
- Broomhead, D. and King, G. 1986. Qualitative Analysis of Experimental Dynamical Systems, In: *Nonlinear Phenomena and Chaos*, Bristol Adam Hilger.
- Golub, G. and Van Loan, C. 1989. *Matrix Computations*, Second Edition. The John Hopkins University Press, Baltimore and London.
- Mann, I. and McLaughlin, S. 1998. A Nonlinear Algorithm for Epoch Marking in Speech Signals Using . *Proceedings of the 9th European Signal Processing Conference*, vol. 2, pp. 701-704.
- Mann, I. 1999. *An Investigation of Nonlinear Speech Synthesis and Pitch Modification Teqniques*. Ph.D. Thesis, University of Edinburgh, Edinburgh.

- Mees, A. Rapp, P. and Jennings L. 1987. Singular Value Decomposition and Embedding Dimension. *Physical Review A*, vol. 36, pp. 340-346.
- Takens, F. 1980. Detecting Strange Attractors in Turbulence. *Proceedings of Symposium on Dynamical Systems and Turbulence*, pp. 366-381.

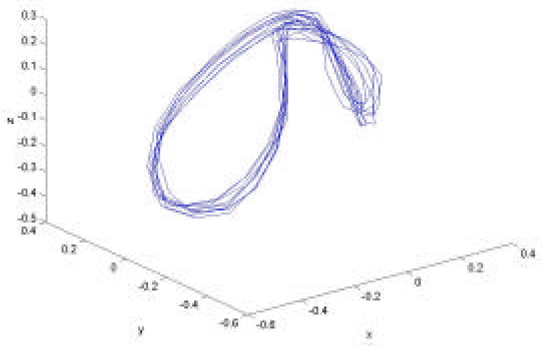

Figure 1: Time delay embedding of the vowel /u:/ spoken by a female speaker in 3D space, with τ = 1.875 msec.

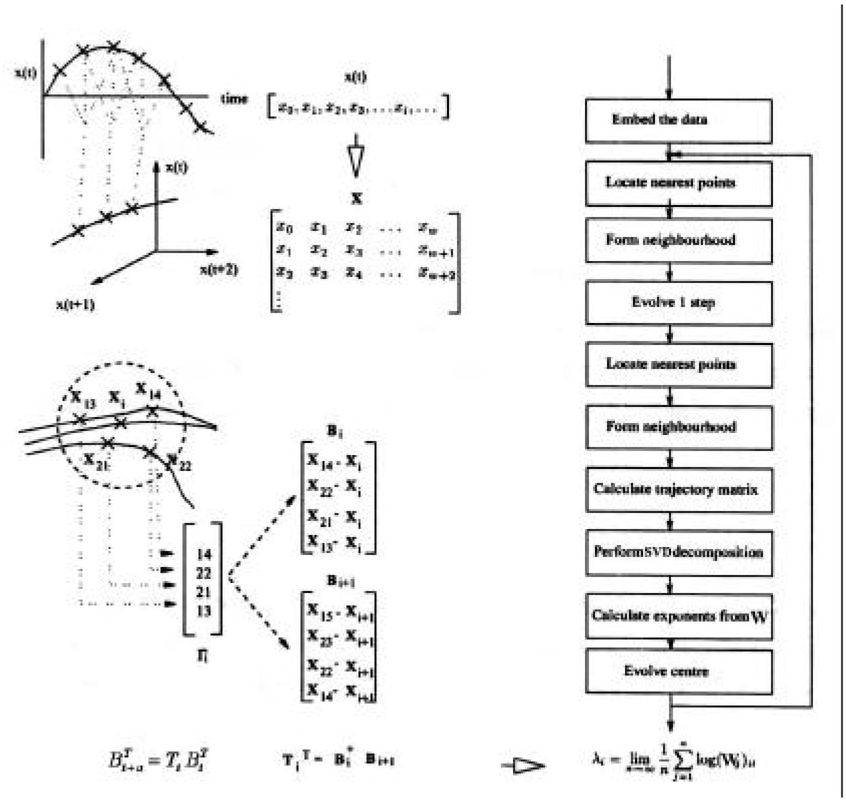

Figure 2: Overview of the Lyapunov exponent estimation algorithm by Banbrook (Banbrook *et al*, 1996).

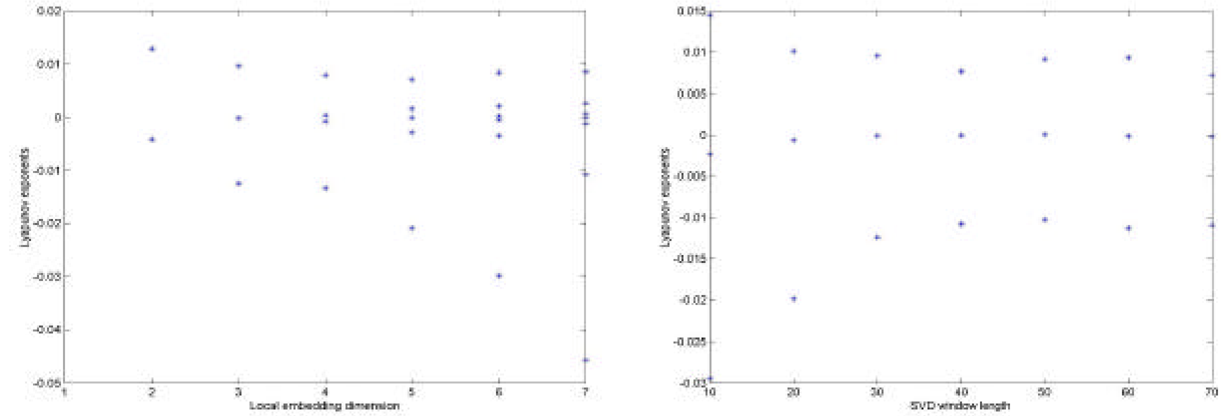

Figure 4: The SVD embedding (w = 30) for the vowel /u:/ spoken by a female speaker in 3D space.

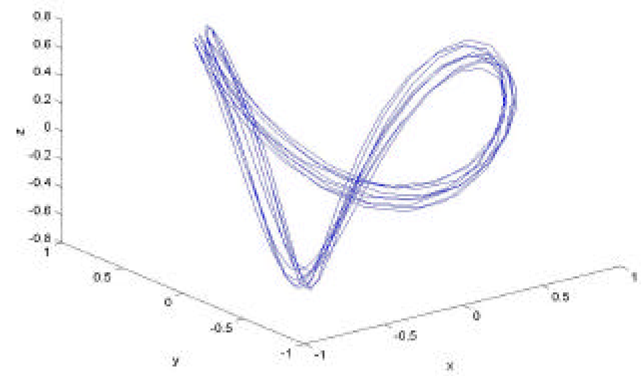

Figure 3: Choosing parameters for Lyapunov algorithm: (a) Exponents for variable SVD window length and (b) variable local embedding dimension (vowel /u:/). Other parameters are 50 neighbors; 20 vectors in each neighborhood set.

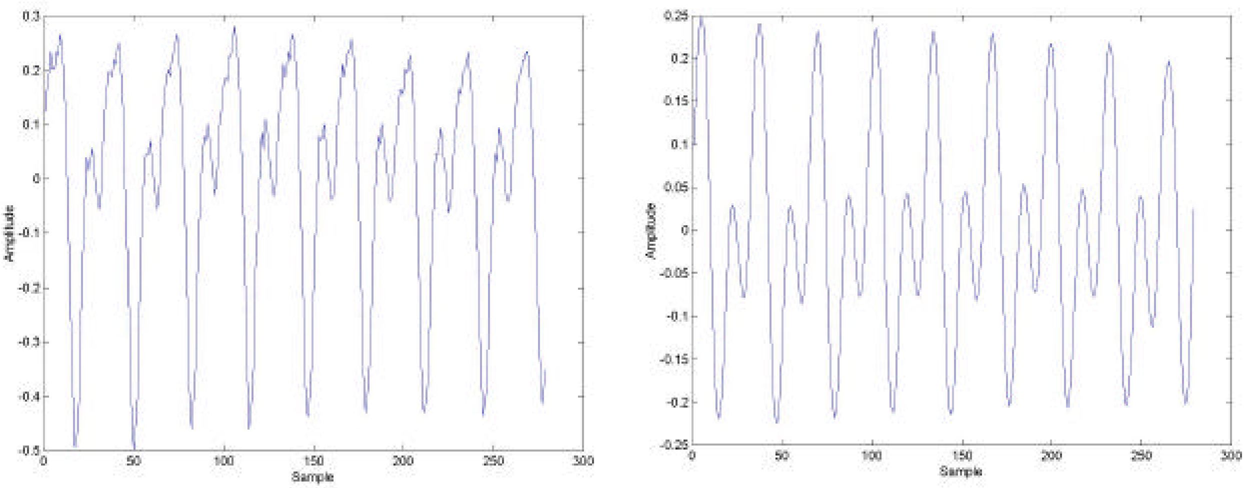

Figure 5: Comparison of (a) original and (b) SVD processed time domain signals, showing the preservation of time and pitch structure