Proceedings of the Technical Sessions, 24 (2008) 36-43
Institute of Physics – Sri Lanka

Analysis of Human Voice Using Non-Linear Techniques

H. P. Mahabaduge and M. K. Jayananda
Department of Physics, University of Colombo, Colombo 3.

ABSTRACT

This paper presents an analysis of five long vowel sounds of Sinhala language, carried out using non-linear techniques. The vowel sounds recorded at 8 kHz sampling rate with 16-bit resolution from both male and female speakers were analyzed by reconstructing the phase portraits using the SVD embedding technique. The required embedding dimension was determined using the false nearest neighbour method. The reconstructed phase portraits were compared with the position of these vowels in the International Phonetic Association (IPA) formant charts. In addition, in order to check the presence of chaos, the Lyapunov exponents of the recorded sounds were calculated.

The results indicate that the speech production system can be modeled using just three dimensions and that the five vowels analyzed can be identified by their geometries in the embedded phase portraits. A correlation exists between their geometry and the position in the IPA formant charts. A comparison of phase portraits with frequency spectra showed the existence of bifurcations. Both positive and negative Lyapunov exponents were found, indicating the presence of chaos.

1. INTRODUCTION

The production of human speech is initiated in the lungs. Air pressure is generated by the lungs and this wave is modulated as it flows through the larynx. Larynx is made up of two almost symmetric masses known as the vocal folds which are capable of closing completely together or, as they move apart, creating a triangular opening called the glottis. During the normal respiration and the production of unvoiced sounds, air passes freely through the glottis. When vocal folds vibrate voiced sounds are produced. The resulting waveform excites the vocal tract, which is the region extending from the larynx to the lips. Different configurations of the vocal tract will result in different modulations of the glottal waveform and thus produce specific sound.

Due to the non-linearities present in the pressure-flow relation in the glottis, the stress-strain curves of vocal fold tissues, the vocal fold collisions etc., one can easily expect very complex, possibly chaotic, behaviours in the human speech production mechanism. Many studies suggest that at least some of the complexities observed in human voices are caused by the intrinsic nonlinear dynamics [1,2]. The work described in this paper was motivated by these observations which justify the attempts to analyze human voice based on nonlinear dynamical techniques.

2. DYNAMICAL THEORY

The behaviour of a dynamical system can be represented by a phase space diagram or a *phase portrait* in which the trajectory of a point corresponding to the current state of the system is plotted in an *n*-dimensional phase space. In many systems, one can observe that, after a transient period, the trajectory is attracted to a fixed region of space, and this is called an *attractor*. While non-chaotic systems have simple attractors such as fixed points or limit cycles, a chaotic systems display very complex attractors called strange attractors. The study of strange attractors gives very useful information about the underlying dynamical system.

A digitized voice sample is only a one dimensional time series and therefore, one cannot directly plot an *n*-dimensional phase portrait out of such data. However, a landmark paper in non-linear dynamics [3] describes a way of reconstructing a phase space diagram from a single observed parameter using the technique now known as *time delay embedding*.

Another problem that arises in working with experimental data is the presence of noise that can seriously affect the reconstruction of the phase portrait. For overcoming this problem, time delay embedding algorithm has been improved by incorporating a noise filtering technique and it is known as singular value decomposition (SVD) embedding [4]. In this work, the SVD embedding was used to reconstruct the phase portraits of Sinhala vowel sounds.

However, before employing SVD embedding, the embedding dimension must be known. Although a real dynamical system such as the human speech production system can have a large number of dimensions, most of them may not play a major role in the observed dynamics and hence the embedding dimension can be much lower than the true dimension of the system. In this work, the algorithm developed by Kennel, Brown, and Abarbanel [5] was used to determine the embedding dimension.

While visual inspection of a phase portrait can be used to determine whether a system is chaotic, often, in experimental data sets, visual inspection does not give conclusive evidence regarding chaos. Therefore, in order to determine whether the voice data that were analyzed in this work were from chaotic systems, Lyapunov exponents were calculated. Lyapunov exponents indicate the rate of contraction (if negative) of expansion (if positive) of the phase space trajectory from an equilibrium point. The method described by Sano and Savada [6] was used for calculating the Lyapunov exponents.

3. THE DATA SET

The required voice samples were collected from two males and two females in their early twenties. For the recording purposes a Creative voice recorder was used with an 8 kHz sample rate and 16-bit resolution. Necessary precautions were taken to minimize the noise while recording the voice samples.

The sounds were five long vowels in Sinhala, , , , , and (International phonetic association symbols /a:/, /i:/, /e:/, /o:/ and /u:/). The reason for choosing long vowels was to obtain higher time duration. The standard practice for obtaining this kind of data is to record a sentence or a word containing the required vowel [7] and then extract it or to use a data set library[8]. The reason for deviating from this standard practice was the adequacy of the data set obtained in the above mentioned procedure for the analysis purposes.

These data samples were analyzed using software packages called Chaos Analyzer [9], TISEAN [10] and Praat [11]. The tools used for analysis from these software packages were time delay embedding, singular value decomposition, formant calculation and Lyapunov spectra. Finally, the obtained results were compared with well known chaotic and non-chaotic systems.

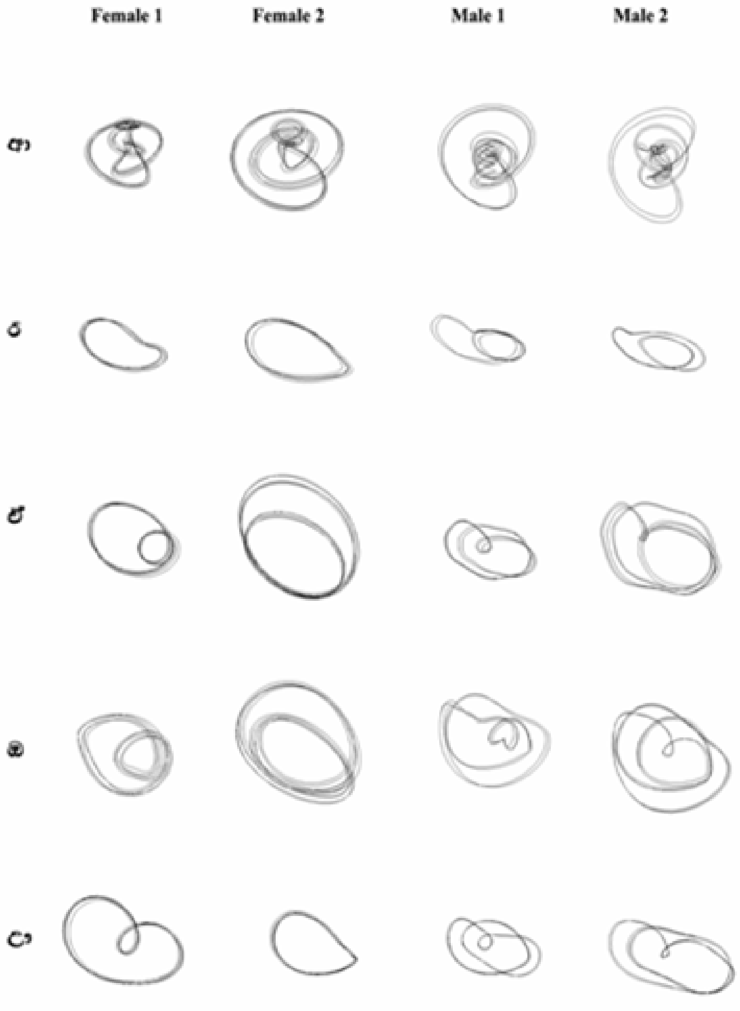


Figure 1: Summary of the attractor patterns

4. RESULTS

A reasonable choice of the delay values is important due to the fact that one always has to deal with a finite amount of noisy data. Both noise and limitation of number of data points prevent us from having access to infinitesimal length scales, so that the structure to be investigated should persists up to the largest possible length scales. A suitable time delay value has to be chosen depending on the type of structure to be explored.

The time delayed mutual information [12] as well as visual inspection of delay representations with various lags provide important information about reasonable delay times while the false neighbors statistic [5] can give guidance about the proper embedding dimension. Using these methods, the embedding dimension was chosen to be 3 while the delay window length was chosen to be 50.

4.1 Phase portraits

Chaos Analyzer [9] was used to observe the attractor patterns. The Signal Analysis Tool in Chaos Analyzer was used to visualize the phase space reconstruction of the sound wave. The reconstructed time delay embeddings for the voice samples from four speakers are shown in figure 1. The simple nature of these diagrams show that the major features of human voice production mechanism can be modeled by just three dimensions. In addition, one can observe that a given sound has some speaker independent features and these may be useful for speech recognition and synthesis.

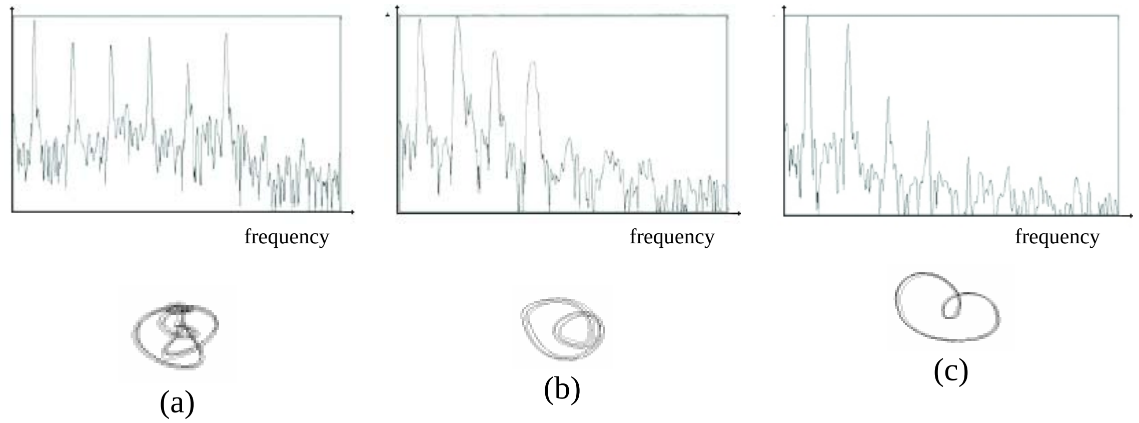
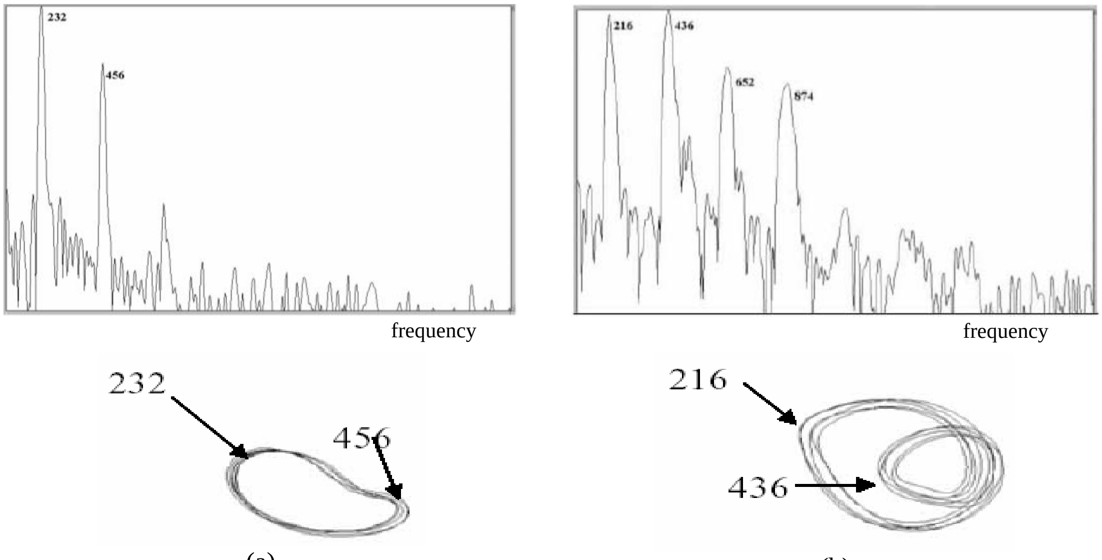


Figure 2: Comparison of the attractor patterns with the frequency spectrum for the sounds (a) (b) and (c) by female I.

Figure 2 shows the frequency spectra (plotted using Praat [11]) along with the respective attractor patterns. A closer observation of the above figure shows a strong correlation between the number of loops present in the attractor with the number of significant peaks in the frequency spectrum.

Figure 2(a) shows the most complex attractor pattern, along with the highest number of peaks in the frequency spectrum. Figure (b) has two loops with four peaks in the spectrum while figure (c) has a single loop with a small knot and four peaks (two large and two small) in the spectrum.

The comparisons of the frequency values of these peaks unveil another interesting feature. That is the existence of period doubling bifurcations. The phase portrait in figure 3(a) appears to be a single loop with a frequency of 456 Hz. However, at close inspection, one can observe that it is split into two, leading to a period doubling. This period doubling is clearly visible in the spectrum where a clear peak corresponding to 232 Hz (approximately half of 456 Hz). Another such example is shown in figure 3(b). Although period doubling bifurcations do not appear to play a major role in speech production period doubling bifurcations exist in speech.



(a) (b)

Figure 3: Period doubling

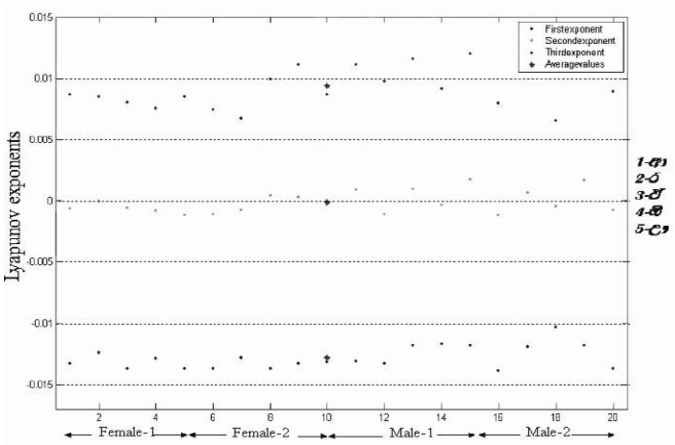


Figure 4: The values of the Lyapunov exponents.

4.2 Lyapunov Exponents

The standard measure for determining whether or not a system is chaotic is the Lyapunov exponent. Lyapunov exponents provide a quantitative characterization of stretching and folding of trajectories in state space. Essentially, across the whole attractor, a positive exponent indicates the divergence of trajectories, where as a negative exponent indicates convergence. If both positive and negative exponents are present then this indicates a strange attractor and chaos.

Figure 4 shows the three Lyapunov exponents calculated for each of the sounds analyzed. The important feature is that all of them have one nearly zero exponent and two exponents with opposite signs, indicating chaos. In order to verify the correctness of the algorithm used for this calculation, Lyapunov exponents were calculate for three data sets obtained from a computer simulation of the Chua's circuit corresponding to a fixed point, a limit cycle and a double scroll strange attractor (shown in figure 5).

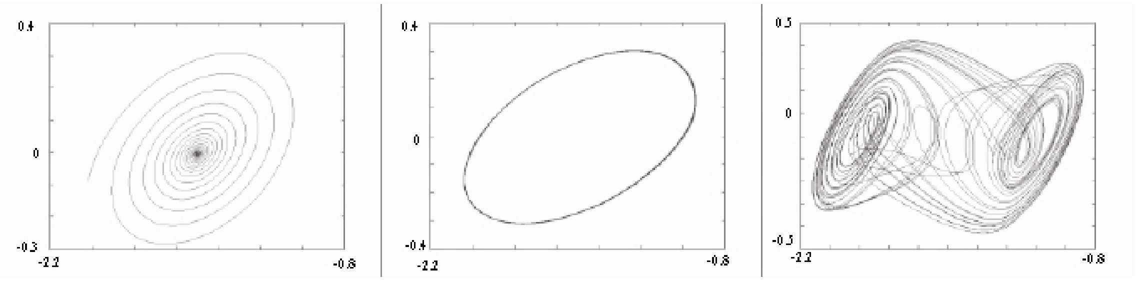


Figure 5: Attractor patterns for the Chua's circuit.

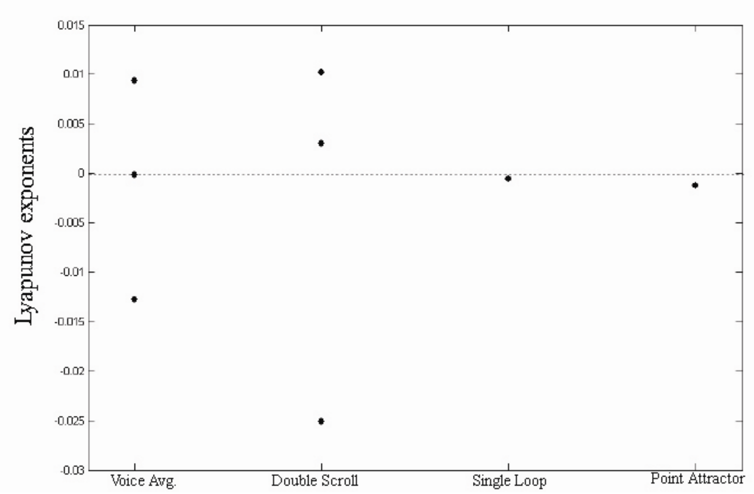
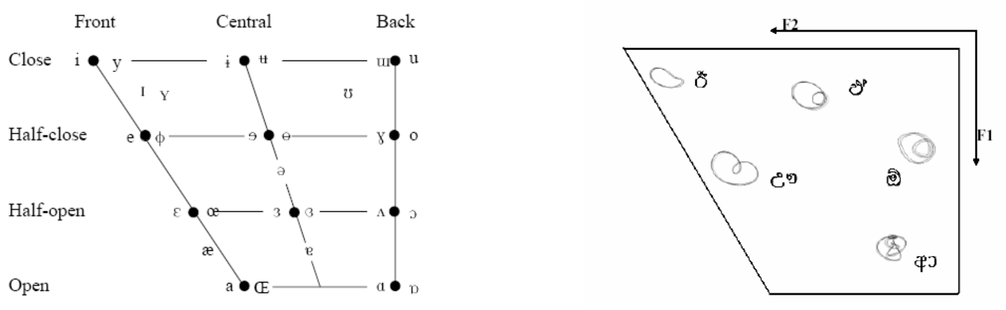


Figure 6: Comparison of Lyapunov exponents for Chua's circuit with those from human voice.

Figure 6 shows the values obtained for the above attractors along with the values obtained for the human voice. Double scroll is a chaotic attractor with two positive and one negative Lyapunov exponents, which confirms the presence chaos. The single loop (limit cycle) and the point attractor each had two Lyapunov exponents and both of them were negative which confirms the absence of chaos.

4.3 Formant charts vs phase portraits

The International Phonetic Association (IPA) has produced a set of phonetic symbols to define all of the individual speech sounds in terms of their place of articulation. The IPA vowel chart is shown in figure 6(a), and provides an alphabet of all vowel sounds by place of articulation.



(a) (b)

Figure 6: (a) The IPA vowel chart, showing vowel phonetic symbols according to tongue position. (b) Attractor patterns shown in the formant chart (Male-1).

In figure 6(b), the formant charts for the sounds analyzed and the respective attractor patterns have been drawn together for comparison. There seems to be a correlation between the position in the chart and the number of loops or folds in the attractor. Progressing anticlockwise around the chart the complexity of the attractor grows. The attractor belonging to the sound “ ” (/a:/) which has the highest complexity compared to the other attractors lies in the bottom right corner and then progressing further upward, the complexity starts to reduce.

5. CONCLUSIONS

One of the important conclusions that can be derived from this analysis is that, although the human speech production is quite complex, the major features of speech can be modeled using a low dimensional dynamical system. This analysis also indicates that different vowel sounds, irrespective of the speaker, have some common features which might lead to a way of speech recognition and speech synthesis based on phase portraits. The combination of formant charts with the respective attractor patterns confirms this possibility.

The other important observations are the existence of bifurcations, and the finding of both positive and negative Lyapunov exponents which indicate chaos. However, it must be emphasized that the magnitudes of the Lyapunov exponents are quiet small and it is possible that noise (both external noise and quantization noise introduced in digitization process) could reverse their signs. Therefore more investigations are required in this respect.

REFERENCES

1. L. Matassini, R.Hegger, H. Kantz and C. Manfredi, *Analysis of Vocal Disorders in a Feature Space*, arXiv:cond-mat/0009188 v1 (2000).
2. W. T. Fitch, J. Neubauer and H. Herzel, *Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production*, Animal Behaviour, , 63, 407–418 (2002).
3. N. Packard, J. Crutchfield, D. Farmer and R. Shaw, *Geometry from a time series*, Phys. Rev. Lett. 45, 712 (1980).
4. Iain Mann, *An investigation of nonlinear speech synthesis and pitch modification techniques*, Ph.D. thesis, University of Edinburgh, (1999).
5. 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).
6. M. Sano and Y. Sawada, *Measurement of the Lyapunov spectrum from a chaotic time series*, Phys. Rev. Lett. **55**, 1082 (1985).
7. S. Haykin and J. Principe, *Making sense of a complex world*, IEEE Signal Processing Magazine, 15, 66 – 81, (1998).
8. F. Plante, G. F. Meyer, and W. A. Ainsworth, *A pitch extraction reference database*, *EUROSPEECH'95*, 1, 837 – 840, (1995).
9. M. Banbrook, G. Ushaw and S. McLaughlin, *CHAOS ANALYSER version 1.0* (1996)
10. R. Hegger, H. Kantz, and T. Schreiber, *Practical implementation of nonlinear time series methods: The TISEAN package*, CHAOS 9, 413 (1999).
11. Boersma, Paul & Weenink, David, *Praat: Doing phonetics by computer (Version 4.3.17)* [Computer program]. Retrieved July 7, 2005, from http://www.praat.org/.
12. A. M. Fraser and H. L. Swinney, *Independent coordinates for strange attractors from mutual information*, Phys. Rev. A 33, 1134 (1986).