NON-LINEARITY IN VOWEL WAVEFORMS

Leisa Condie,

Department of Computer Science,
University College, University of New South Wales.

ABSTRACT

Acoustic models of the vocal tract show that turbulence plays a significant role in speech production. Non-linear dynamics has recently provided tools which allow analysis in this very difficult area. In this note vowel waveforms are examined for evidence of irregular behaviours using such tools.

INTRODUCTION

Traditionally, fluid flows have been modelled linearly due to the extreme complexity of solving non-linear systems of equations. This linear approach has also been taken in modelling the physical system of speech production, which involves complex changes in parameters throughout the speech process. With the growing body of techniques available to study previously intractable non-linear processes comes the chance to examine speech waveforms from a new perspective.

DATA

The speech of an Australian female was recorded onto tape, amplified to ±1.25 volts by the main amplifier, then filtered through a 7.6kHz Butterworth low-pass filter to avoid aliasing. The signals were then sampled at 8kHz, with a gain of eight, with the DT2801 digital-to-analogue converter. Features were extracted with the ILS signal analysis package, and analysed with the S statistical package.

The sampled data was stored by ILS in 'coded ASCII form', which was not explained in the manual. By coded ASCII it was meant that each 16 bit sample was converted into three printable ASCII characters (excluding upper and lower case alphabetics) in the order low byte to high. It was clear that each character denoted five bits, with the third denoting six bits (five value bits and a sign bit). As the characters used were in the range octal 40 to octal 77, and octal 133 to octal 137, there was a subtraction of octal 40 from each character when processed to bring it down to the appropriate number of bits before being integrated into the final number. Examples are: '1?_' for -14, and '5- ' for 437.

Slices were taken from each of the vowels contained in the vocabulary 'one', 'two', 'three', 'four', 'five' and 'six'. A frame is of 5.375ms duration. To maintain the stationary nature of the signal, a slice of 129 data points - three frames - was used in the mappings described below. The frames to be used were selected arbitrarily, within the constraint that the frames were within the vowel section of the word.

ANALYSIS METHODS

System behaviour can be examined in the phase space: a random signal is seen as a cloud, whilst deterministic chaos lies on a form of curve. The latter occurs as such chaos is related to strange attractors with fractal dimension. Rather than study the attractor in phase space, it is usually easier to study Poincaré and next-amplitude maps. Poincaré maps show trajectory intersections with a particular plane: $(x(t), x(t+\tau))$, where τ is some time offset. This is termed τ-advance mapping (Parker and Chua, 1989). Next-amplitude maps consist of plotting successive maxima: $(x_{\max_i}, x_{\max_{i+1}})$. Again, deterministic chaos is indicated if there is one dimensional structure in these plots.

Given a time series $x(t_i)$ then, for almost all time delays τ and sufficiently large values of m, a new at-

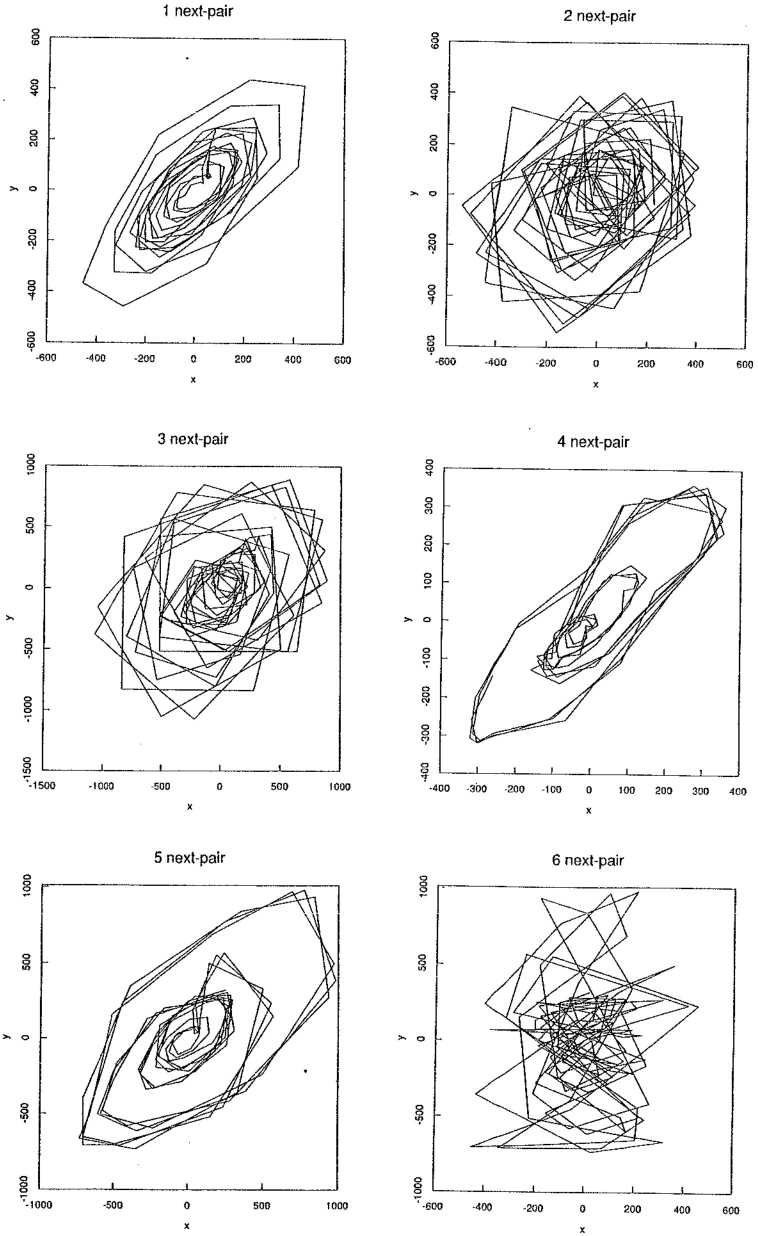

Figure 1. Next-pair maps

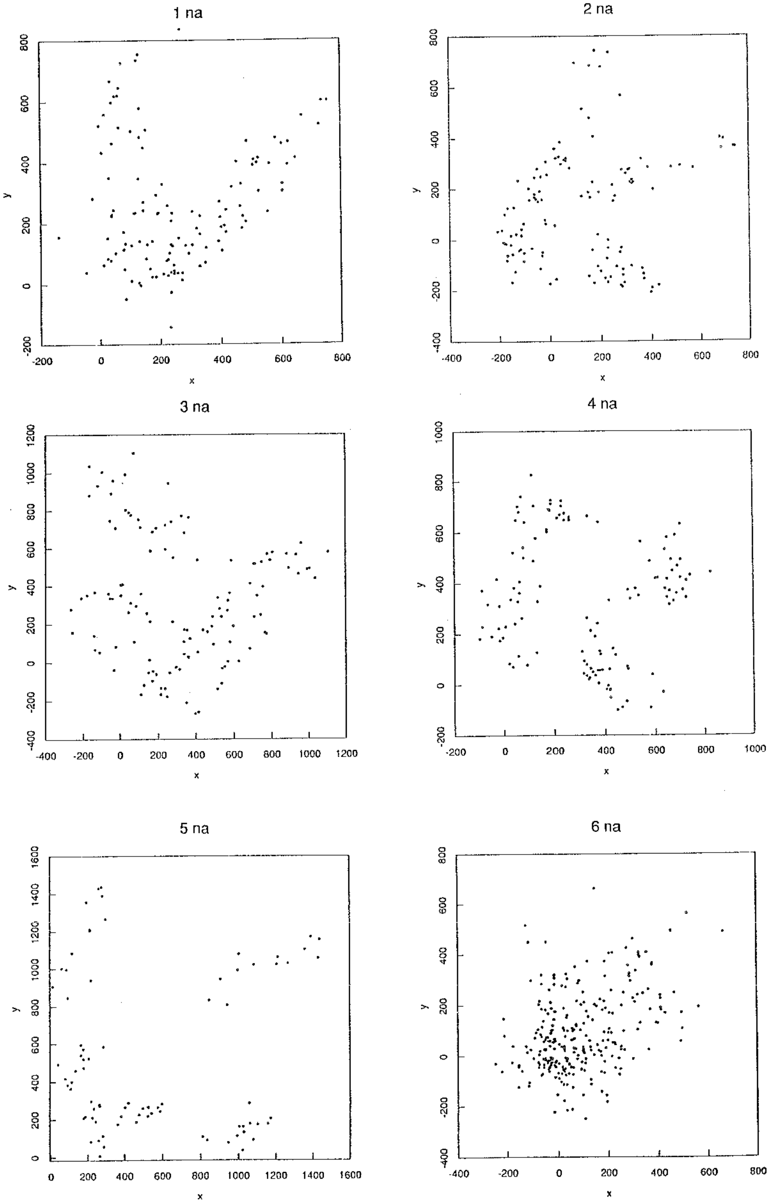

Figure 2. Next-amplitude maps

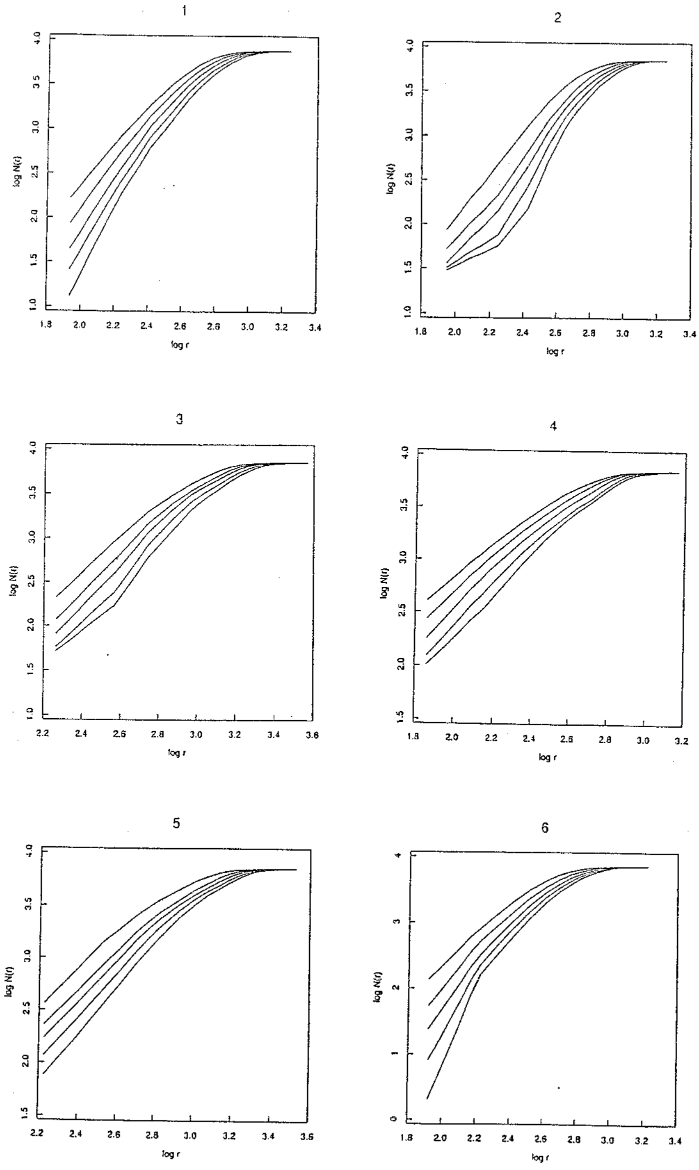

Figure 3. Log-log plots of interpoint distances for several embedding dimensions

tractor, defined by the trajectory $(x(t_i), x(t_i+\tau), x(t_i+2\tau),..., x(t_i+(m-1)\tau)$, will have the same dynamic properties (set of Lyapunov exponents, fractal dimension etc.) as the original. Takens (1981) showed that the embedding dimension m does not exceed $2n+1$, where n is the number of original variables, however this is only a sufficiency condition, and often a far smaller dimension will be enough. In particular, essentially two dimensional orbits can be embedded in three dimensions (Schaffer and Kot, 1986).

To estimate the attractor dimension an investigation is needed of the numbers $N(r, m)$ of pairs (x_i, x_j) where r is the maximum distance between the points. The correlation dimension D can be determined from log-log plots of $N(r, m)$ versus r as

$$D = \lim_{r \to 0} \frac{\log N(r, m)}{\log(r)}$$

If there is a range of r with a reasonably constant slope, then there is self-similarity. A permanent increase in slope with an increasing m indicates a random signal, whilst a saturation in slope indicates a low-dimensional attractor (Broomhead & Jones, 1989, Grassberger & Procaccia, 1983, Mende et al., 1990, Eckmann, 1985).

RESULTS

Initially next-pair plots were produced. In these, each data point x_i and its successor x_{i+1} form a pair to be plotted. If the sequence of time series points was purely random, then the resultant graph would be a unstructured tangle of lines. As the plots of Figure 1 show, this is certainly not the case for several of the vowels under examination, although others appear to have rather little structure. A search to determine what structure there is in these series then proceeded. It should be noted that only the vowel section of the word was examined, though each will be referred to by the full word to avoid any confusion.

Next-amplitude maps were then produced for each vowel. These are shown in Figure 2, and reveal clear structure, with the exception of 'six'. In conjunction with its next-pair plot it was concluded that randomness was displayed. When the frames were initially selected, there was no equipment available to double-check that the frames fell on the vowel section, and it was suspected that a sibilant section had been inadvertantly included. Later checks proved this to be the case. Hence sibilants display randomness, whilst vowels have structure in next-amplitude maps.

The next step was to measure the dimensionality. Firstly plots of $N(r, m)$ against r were made, then the log-log graphs of Figure 3 which show (from bottom to top) embedding dimensions $m = 3, 4, 5, 6, 7$. To check that the slopes saturated correctly, plots were additionally made of the slopes of the log-log plots. These revealed that the vowels demonstrated correlation dimensions between 3 and 4, or 4 and 5 in the case of 'two', which agrees with the observations of Tishby (1990). It is also to be noted that the slopes for the graph of 'six' do not saturate in the dimensions examined.

CONCLUSION

It has been shown that certain vowels display low-dimensional chaos, and further, that the correlation dimension lies between 3 and 4. A section of sibilant waveform, inadvertantly included, was found to be random rather than chaotic. To be complete, investigations are needed to find equations for the underlying non-linear behaviour that has been demonstrated here, however that is beyond the scope of this note.

REFERENCES

Broomhead, D. S. and Jones, R. (1989) *Time-series analysis*, Proc. of the Royal Society of London A 423.

Eckmann, J. -P. (1985) *Ergodic theory of chaos and strange attractors*, Reviews of Modern Physics 57, 3.

Grassberger, P. and Procaccia, I. (1983) *Measuring the strangeness of strange attractors*, Physica D 9.

Mende, W. and Herzel, H. and Wermke, K. (1990) *Bifurcations and chaos in newborn infant cries*, Physics Letters A 145, 8.

Parker, T.S. and Chua, L.O. (1989) *Practical Numerical Algorithms for Chaotic Systems*, Springer-Verlag.

Schaffer, W.M. and Kot, M. (1986) *Differential systems in ecology and epidermiology*, in Chaos, ed. A.V. Holden, Princeton University Press.

Tishby, N. (1990) *A dynamical systems approach to speech processing*, in Proceedings of IEEE ICASSP 90, S6b.5.