PHASE SPACE BEHAVIOUR OF SPEECH

Dr Leisa Condie

Dept of Mathematics, Statistics and Computing Science,
University of New England - Armidale

ABSTRACT - Characterisation of phase space behaviour of speech gives an alternate view of speech waveforms. Correlation dimension and false nearest neighbour measures are just two techniques for investigating the dynamical behaviour of speech. The recent false nearest neighbours technique for determining the embedding dimension for phase space reconstruction of a time series is used on both voiced and unvoiced speech. Results from false nearest neighbours are compared with the correlation dimension results.

THEORETICAL CONCEPTS

An n-dimensional manifold M is a space in which it is possible to set up a coordinate system near each point such that locally the space looks like a subset of the Euclidean space R^n. A dynamical system D is a smooth manifold M, together with a vector field v defined on M (Casti, 1989). An attractor is a subset of n-dimensional phase space which almost all sufficiently close trajectories approach asymptotically.

There are three major areas of interest in the characterisation of attractors: Lyapunov (or characteristic) exponents, entropy and dimensions. Lyapunov exponents give the average rate of expansion (if positive) or contraction (if negative) near a limit set. They generalise the linear stability criteria for fixed points and limit cycles. A non-chaotic system is asymptotically stable if all its Lyapunov exponents are negative. For any limit sets on continuous, time-dependent systems, except an equilibrium point, the exponent is zero. Although it is required that the sum of exponents is negative for dissipative systems, the presence of one or more positive exponents indicates a strange attractor (Condie, 1991).

Dimensions

The dimension of an attractor can be expressed in many ways, but the underlying idea is that there is a lower bound on the number of variables needed to describe the steady state behaviour of the system. A non-chaotic system has integer dimension, whilst a chaotic system almost always has fractal (non-integer) dimension.

Given an attractor in $\mathcal{R}^n$ covered by $N(r)$ n-dimensional hyperspheres of radius r then

- The fractal, or Hausdorff, dimension is calculated as the limit is taken. As $r \to 0$, $N(r) \propto r^{-D_H}$.
- The capacity dimension is defined as

$$D_{cap} = \lim_{r \to 0} \frac{\ln N(r)}{\ln(1/r)}$$

  or alternately, $N(r) \propto (1/r)^{D_{cap}}$ for some $k > 0$.

- Let P_i be the relative frequency of visitation of a typical trajectory to the ith hypersphere. The information dimension is then defined to be

$$D_I = \lim_{r \to 0} \frac{H(r)}{\ln(1/r)}$$

  where $H(r) = -\sum_{i=1}^{N(r)} P_i \ln P_i$, or alternately $H(r) = kr^{-D_I}$ for some $k > 0$.

- The correlation dimension can be expressed as

$$D_C = \lim_{r \to 0} \frac{\ln C(r)}{\ln r}$$

where

$$C(r) = \lim_{N \to \infty} \frac{1}{N^2} \sum_{i \neq j} \Theta(r - ||x_i - x_j||)$$

$\Theta(x)$ is the Heaviside function and $||\cdots||$ represents the Euclidean norm. Alternately $C(r) = kr^{D_{cap}}$ for some $k > 0$.

- Given Lyapunov exponents $\lambda_1, \lambda_2, \cdots, \lambda_j$ where $\sum_{i=1}^{j} \lambda_i \geq 0$ then the Lyapunov dimension

$$D_L = j + \frac{\lambda_1 + \cdots + \lambda_j}{|\lambda_{j+1}|}$$

If no j exists to satisfy the condition, D_L is defined to be 0. The sum of positive exponents is essentially the Kolmogorov entropy.

False Nearest Neighbours

From the set of observations $x(n)$, multivariate vectors in d dimensional space $y(n) = [x(n), x(n + T), \cdots, x(n + (d-1)T)]$ are used to trace out the orbit of the system, where T is the time delay (usually chosen to be the first minimum of the average mutual information). The aim of the method proposed by (Kennel et al., 1992) is to find the embedding dimension d by examining the topology of the embedding. The purpose of time delay is to unfold the projection of multivariate state space onto the one dimensional time series back to a multivariate system representative of the original. Takens (1981) states that $d > 2d_A$ where d_A is the dimension of the attractor, but this is only a sufficient condition, and in experiments it is preferable to use d_E, the minimum embedding dimension.

The idea of false nearest neighbours is that in examining d and then $d + 1$ one can distinguish true neighbours of a given point from false neighbours. False neighbours are points on the data set that are neighbours solely because we are viewing the attractor in too small an embedding space ($d < d_E$) and in a higher dimension those points will have moved apart and no longer be neighbours. The value for d_E is then obtained when there are no false nearest neighbours in successive embeddings. Other methods involve computation of some invariant on the attractor until it becomes independent of d, but such methods tend to be data intensive, and somewhat subjective in determining d_E.

Neighbours are determined by simply using the square of the Euclidean distance between points on the attractor. The rth nearest neighbour $y^{(r)}(n)$ of $y(n)$ has distance

$$R_d^2(n, r) = \sum_{k=0}^{d-1} (x(n + kT) - x^{(r)}(n + kT))^2$$

In $d+1$ dimensions $R_{d+1}^2(n,r) = R_d^2(n,r) + (x(n+dT) - x^{(r)}(n+dT))^2$. Thus a false nearest neighbour can be defined as any neighbour such that

$$\sqrt{\frac{R_{d+1}^2(n,r) - R_d^2(n,r)}{R_d^2(n,r)}} = \frac{|x(n+dT) - x^{(r)}(n+dT)|}{R_d(n,r)} > R_{tol}, \quad R_{tol} \geq 10$$

It is sufficient to use $r = 1$, for the nearest neighbour only, and interrogate on all points $n = 1, \cdots, N$ of the attractor. Unfortunately this condition is not sufficient for determining d_E: the nearest neighbour is not necessarily a *close* neighbour - indeed $R_d(n, 1)$ can be comparable to the size of the attractor for a noise waveform, leading to erroneous results. Thus a second measure is used in conjunction with the first, and both must be satisfied for true neighbours.

$$\frac{R_{d+1}(n)}{R_A} > A_{tol}$$

where

$$R_A^2 = \frac{1}{N} \sum_{n=1}^{N} (x(n) - \frac{1}{N} \sum_{n-1}^{N} x(n))^2$$

It should be noted that any estimate of attractor dimension can be used here. Recommended tolerances were $A_{tol} < 2.0$ and $R_{tol} > 15.0$.

For small data sets and noisy data sets the percentage of false nearest neighbours will not fall to zero. For small data sets a threshold of 1% was recommended. Experiments with noisy data sets revealed that the increase in embedding dimension was quite slow, until at a SNR of zero where it was indistinguishable from noise. With an increase in d it was predicted that the number of false nearest neighbours (expressed as a percentage of all neighbours) should plateau, effectively giving a 'calibration' figure.

ANALYSIS METHODS

Ted Bullen, of the University of Technology, South Australia, provided 19 sets of input data consisting of three seconds of each of /a/, /i/, /u/ and /n/ for one female and four male speakers. The data was sampled at 8kHz, with 3500Hz bandpass filtering, at 14 bits A/D conversion. Not all the data was valid, as phoneme segmenting errors at recording resulted in background noise taking up half the points at the start of one file. Another file was even worse – only 100 points of phoneme onset were recorded. This had the beneficial effect, however, of allowing tests to be run on that background noise, to determine whether it was the cause of any structure that was found. The speakers are identified only by their initials: tb, jo, gdb, gb and pt. For a second set of tests an Australian female speaker was recorded uttering vowels (in sung form), plosives, semi-vowels, fricatives and nasals, using 10kHz and 14kHz sampling, 5kHz bandpass filtering, and 8 bits A/D conversion. With the same equipment several notes from a set of pitch pipes were also recorded for analysis.

The correlation dimension exponent gives a useful measure of the local structure of the attractor, and is easily calculated from time series. If $C(r) \approx r^n$ then the system is random. If there is a strange attractor then $C(r) \approx r^\nu$ where ν is independent of the embedding dimension n. The system is embedded in progressively higher dimensional space to see which scaling occurs. For comparison both the Euclidean norm and max norm were tried in the correlation dimension calculations. The Euclidean norm is calculated by taking the sum of the squares of the coordinate differences, then taking the square root of that total. The max norm (Haucke et al., 1986) is defined as the largest of all the coordinate differences. The mod norm is calculated as the sum of the absolute values of each coordinate difference. For completeness the correlation dimension was calculated using the mod norm. A slightly different set of correlation dimension calculations were next performed: a single embedding dimension was chosen and the correlation dimension was calculated for a series of different delays.

A program implementing the false nearest neighbours technique described earlier was provided by Matthew Kennel (Kennel et al., 1992). It required all samples to be unique, rather than quantised, so small amounts of noise had to be added to make each data point unique. Differing levels of this noise were tried to see if they had a significant effect on the results. Another test undertaken was to use only 3000 data points to see what, if any, effect this had on the embedding dimension returned by the program. A number of time delays were tested, the R_{tol} parameter varied, and the results examined.

RESULTS

Phase Space Analysis

The phase space portraits show there is three dimensional (at least) structure in the voiced samples, and that structure is directional. Each portrait is different for each speaker and each phoneme: no speaker has identical portraits for different phonemes, however the portraits for /m/ and /n/ are very similar. Noise, sibilants, plosives and africtatives are unstructured in phase space, whilst nasals, semi-vowels and vowels are structured. This result is not surprising: the phase space portrait of periodic, quasi-periodic, and chaotic (close to, but not exactly, either of these) sources are expected to show structure, whilst noise is not.

Correlation Dimension

The correlation dimension for embedding dimensions two through to twelve were calculated using the Euclidean, max and mod norms. Table 1 shows the average dimension for embedding dimensions three

File	Euclid	Max	Mod
gbnn	1.01	1.16	0.95
gdbnn	1.53	1.59	1.55
jonn	1.01	1.00	0.98
tbnn	1.03	0.99	1.07
noise	1 → 6	1 → 4	1 → 5
gbah	1.20	1.19	1.05
gdbah	1.91	1.98	1.77
joah	1.74	1.66	1.90
tbah	2.19	2.12	2.24
ptah	1.88	2.00	1.76
gbee	1.62	1.67	1.60
gdbee	1 → 7	1 → 7	1 → 5
joee	2.09	2.05	2.18
tbee	2.20	2.27	2.21
ptee	1.68	1.87	1.66
gboo	1.46	1.67	1.38
gdboo	1.93	2.11	1.82
jooo	1.61	1.67	1.55
tboo	1.24	1.39	1.16
ptoo	1.36	1.61	1.26

Table 1: Correlation dimensions using Euclidean, max and mod norm

File	Set 1	Set 2
jonn	1.01	1.04
gbah	1.20	0.99
joah	1.74	1.80
gbee	1.62	1.59
joee	2.09	2.02
jooo	1.61	1.56

Table 2: Correlation dimensions using Euclidean norm for 2000 and 2500 points

to nine (inclusive). The noise file consisted of the first 1400 points of the gdbnn file, leaving 1200 points for gdbnn, otherwise 2000 data points were used. Both the noise file and gdbee – which consisted only of noise – could not be averaged as their dimensions rose rapidly. For those two files the lower and upper dimension are given. Comparing these results to earlier ones (Condie, 1990) shows these dimensions are lower. This is attributed to having used more data points for a more reliable result in this experiment. The question then arises: would even more points change the results significantly? To answer this the dimensions were recalculated (using the Euclidean norm) for 2500 data points, and the results presented (as Set 2), along with those for 2000 data points (Set 1), in Table 2. This table shows that there is little difference in correlation dimension for most of the data sets tested.

It is interesting to compare the correlation dimensions with the phase space portraits. Gb, tb and jo to all have very tight patterns, whilst gdb is quite loosely structured. Table 1 shows gb, tb and jo to have a (Euclidean norm) correlation dimension very close to one, whilst gdb is 1.5. At the other end of the dimension scale are joee, tbee and tbah, all with dimensions over two. Examination of their phase space portraits reveals them to be strange block patterns. Of the other data sets, it appears that there is a link between the correlation dimension and the sharpness of the portrait, with the average pattern giving a correlation dimension around 1.5. Turning to the supplemental data, whose correlation dimensions averaged for dimensions 2 to 8 inclusive is shown in Table 4, the portraits for /e/, /n/ and /m/ are similar, with loosening of the loop reflected in increasing dimensions. The pitch pipe notes all show intricate patterning - 'a' more so than the others. The sibilants showed a wildly unstructured pattern (as did a file of noise created with ILS), but the africitive and plosives showed random patterns -

File	Dim	File	Dim	File	Dim	File	Dim
gbnn	1.02	gbah	0.94	gbee	1.59	gboo	1.37
gdbnn	1.49	gdbah	1.64	gdbee	2.14	gdboo	1.82
jonn	1.05	joah	1.91	joee	1.99	jooo	1.54
tbnn	1.00	tbah	1.85	tbee	2.11	tboo	1.22
noise	2.28	ptah	1.59	ptee	1.60	ptoo	1.42

Table 3: Correlation dimensions using Euclidean norm for delays 1 to 10

File	Dim	File	Dim	File	Dim	File	Dim
a(pitch)	1.63	b(pitch)	1.14	d(pitch)	1.16	ah	1.26
ee	0.92	mm	1.15	nn	1.04	oo	1.18
dd	0.22	jj	0.51	tt	0.49	ss	$1 \rightarrow 4$
rr	1.51	vv	1.45	ww	1.24	ll	1.23
zh	$1 \rightarrow 5$	sh	$1 \rightarrow 5$	th	1.3	noise	$1 \rightarrow 7$

Table 4: Correlation dimensions (Euclidean norm) with 3000 points

unstructured, but loose. Their correlation dimension is extremely low. Liquids and glides showed a neat pattern, as did the labiodental fricatives - however the lingudental fricative was loosely structured.
Table 3 shows the correlation dimension calculated using 2000 data points with the Euclidean norm averaged over the delays one to ten (inclusive). The first point to note is that noise (and the noise file gdbee) give a single, stable dimension over two: the delay approach does not show the explosive rise across embedding dimension because here the embedding dimension is fixed: it has value four in this experiment. Comparing the results with Table 1 shows comparable results for some sounds (for example /n/) but very different results for others (for example /a/). With one notable exception, the correlation dimension across the delays was stable: there was less than a 0.05 spread from highest to lowest. The exception was jo for /i/: there the spread was 0.5.
Tests with the false nearest neighbours program were next undertaken. Table 5 shows the embedding dimension at which the percentage of false nearest neighbours first dropped below 0.1%. The time delay was 5 and $R_{tol} = 15.0$. In all experiments $A_{tol} = 2.0$. A range of time delays were tested, with some slight variations recorded. Where they occurred, they were usually a rise of 1 in the embedding dimension. The value of R_{tol} was then varied from 13.0 to 17.0 in increments of 1.0, with the time delay held steady, and the embedding dimensions rechecked. It was found that little variation occurred even with a variety of noise levels. Again holding the delay steady, the embedding dimensions reported when only 3000 data points (instead of up to 28000) were used were examined. The dimensions were found to be slightly higher in this case, but not significantly so.

An interesting point raised during these experiments was whether pitch played any role in the results.

File	Dim	File	Dim	File	Dim	File	Dim
joah	3	joee	4	jonn	3	jooo	3
tbah	4	tbee	4	tbnn	3	tboo	3
gbah	5	gbee	3	gbnn	4	gboo	3
gdbah	4	gdbee	> 6	gdbnn	3	gdboo	3
ptah	3	ptee	3	ptoo	4	noise	> 8
a(pitch)	5	b(pitch)	5	d(pitch)	5	ss	> 8
rr	5	zh	4	sh	> 8	th	7
ll	6	vv	5	ww	5	pp	> 8
mm	7	nn	7	jj	7	dd	> 8

Table 5: Embedding dimension where false nearest neighbours < 0.1

A brief experiment was undertaken where an Australian female speaker recorded utterences of /a/ and /e/ each at three different pitches. The /e/ was also recorded twice at the same pitch (different from the previous three). It was found that pitch did have an influence on the correlation dimension: from lowest to highest pitch /a/ returned dimensions 1.31, 1.49 and 1.71 (dimensions three to eight inclusive), whilst /e/ returned 0.83, 1.04, 1.3 (as for /a/). The two repeated utterances of the same pitch returned 1.16 and 1.19. This phenomena deserves further investigation. Interestingly, when these data samples were tested with the false nearest neighbours program, there was little variation in minimum embedding dimensions for /e/ and none for /a/.

CONCLUSION

Many of the data sets examined revealed three dimensional structure which could not be called random: there were clear orbitals in most. Background noise could not account for this structure as tests revealed it to have a distinctive two dimensional appearance. Correlation dimension measures showed noise to have a rapidly rising dimension as the embedding dimension was increased, but non-noise data showed stable correlation dimensions over embedding dimensions three to nine. In comparing the correlation dimension to the phase space portrait the more tightly structured the phase space portrait, the lower the correlation dimension. Comparing the use of Euclidean, max and mod norms reveals there are sometimes rather large differences in results, but it is unclear which is the more accurate. Other dimensional measures would have to be made to determine this. It was also shown that pitch can influence correlation dimension. False nearest neighbours showed noise and noise-like waveforms do not embed in a low dimension, whereas other waveforms have a minimum embedding dimension around 4. This will influence methodology in other dimensional measures.

References

Ben-Mizrachi, A., Procaccia, I., and Grassberger, P. (1984). Characterization of experimental (noisy) strange attractors. *Physical Review A*, 29(2).

Casti, J. L. (1989). *Alternate Realities: Mathematical models of nature and man.* John Wiley and Sons.

Condie, L. (1990). Non-linearity in vowel waveforms. In *Third International Australian Speech Science and Technology Conference.*

Condie, L. (1991). *Speech Signal Analysis and Investigations in Cryptography.* PhD thesis, Dept. of Computer Science, University College, University of New South Wales, Australian Defence Force Academy.

Devaney, R. L. (1987). *An Introduction to Chaotic Dynamical Systems.* Addison-Wesley.

Haucke, H., Ecke, R. E., and Wheatley, J. C. (1986). Dimension and entropy for quasiperiodic and chaotic convection. In Mayer-Kress, G., editor, *Dimensions and Entropies in Chaotic Systems.* Springer-Verlag.

Kennel, M. B., Brown, R., and Abarbanel, H. D. I. (1992). Determining embedding dimension for phase space reconstruction using the method of false nearest neighbors. Preprint.

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

Swinney, H. L. (1986). Experimental observations of order and chaos. In Thompson, J. M. and Stewart, H. B., editors, *Nonlinear Dynamics and Chaos.* John Wiley and Sons.

Takens, F. (1981). Detecting strange attractors in turbulence. In Rand, D. A. and Young, L.-S., editors, *Lecture Notes in Mathematics 898.* Springer-Verlag.

Thompson, J. M. and Stewart, H. B. (1986). *Nonlinear Dynamics and Chaos.* John Wiley and Sons.

Wolf, A., Swift, J., Swinney, H., and Vastano, J. (1985). Determining Lyapunov exponents from a time series. *Physica D*, 16.