DETECTION OF CHAOTIC BEHAVIOUR IN SPEECH SIGNALS USING FRASER'S MUTUAL INFORMATION ALGORITHM

Hans–Peter Bernhard and Gernot Kubin

Institut für Nachrichtentechnik und Hochfrequenztechnik,
Technische Universität Wien,
Gusshausstrasse 25/389, A–1040 Vienna, Austria

Abstract

The apparent irregularity of speech signals is generally considered the effect of pure randomness or of time-varying control. This paper presents a new explanation in terms of deterministic chaos as the underlying model of the temporal fine structure of speech. Within this hypothesis, attractors of sustained vowel articulations are reconstructed using Takens' theorem. Next, Fraser's mutual information algorithm is exploited to estimate the marginal redundancy $R_n^{\prime T}$ of a signal sample given n past samples with delay time T. This results in the optimal choice of the delay time T_{opt} (in the vicinity of 1 msec), a saturation value of 3 for the embedding dimension and an estimate of the information production rate of roughly one bit per pitch period. This last result together with the independently measured correlation dimension (between 1 and 2) corroborate the chaos hypothesis for speech. The general use of the implemented algorithms of chaos-targetted analyses of natural signals is possible.

1 Introduction and motivation

When dealing with speech it is often very interesting to know which kind of signal it is. If signal processing algorithms are used e.g. to reduce the bit rate for speech transmission, they are targetted at the structure of speech signals (formants, pitch period, etc.). The observed irregularity of speech signals is at present considered the effect of pure randomness, or time-varying control by the speaker. Therefore, it is of great interest if there is a new explanation of the speech signal's nature. *Is it possible that speech signals are produced by a chaotic system ?* This question has been addressed only by a small number of investigations so far [KM90, MY90, QH90, Tis90]. However, the following points related to the speech production mechanism suggest such an interpretation:

1. If speech sounds are produced *unvoiced* the primary source consists in turbulent air flow originating at constrictions of

the vocal tract [Fla72, pp. 53–58]. Turbulent flow is one of the most extensively studied patterns of chaotic behaviour in fluid mechanics [Tak81].

2. If speech sounds are produced *voiced* there exists a nonlinear coupling between the sound source and the vocal tract [Fla72, pp. 41–53, 246–259]. Schoentgen [Sch90] shows that only a nonlinear model of the glottal waveform renders its correct amplitude and frequency dependence.

3. There is some experimental evidence that speech production is not characterized by linear propagation of acoustic waves but that the actual air stream mechanisms are much more complicated than generally believed [TT90].

To find out whether a signal can be regarded as chaotic or not we need a decision rule. The decision will be based on two measurements, the information production rate and the dimension of attractors in state space. The dimension of the attractor indicates chaos, if it is not an integer value. If the information production rate is positive it is related to a chaotic system.

2 Problem

Conventional analysis of a chaotic system is based on a system that is fully known. It is possible to calculate the Ljapunov exponents and to estimate the dimension. The attractor can also be computed, because we know the differential topology and the state variables.

In our problem we do not know the underlying differential equations of the system. We have only time series of recorded speech signals available. It is a classical situation for Takens' theorem [Tak81], which allowes to reconstruct an attractor from a single observation series. Let $b(t)$ be the observation and $b(t + kT)$ the same but time-delayed measurement where k is an integer and T is the delay time. If delayed observations are collected in a vector $[b(t), b(t+T), b(t+2T), \ldots, b(t+(n-1)T)]$ an n-dimensional embedding of the state space of the underlying system is reconstructed. Note: All further calculations and considerations refer to such reconstructed state spaces. It is also important to mention that at this point we have no idea about the required number of state variables (reconstruction dimension) and the delay time T. We use a method given by Fraser, to find these two important parameters.

3 Determining reconstruction dimension and delay time

3.1 Theory

The method originates from information theory. Fraser shows in [FS86] that this method is a reliable tool in determining the two parameters.

First we introduce the nomenclature which follows closely the one given by Fraser:

- $b(t)$ stands for a sample value at time t.

- $\vec{b}_n^T(t)$ denotes a vector with the index n for the dimension and the delay time T between the vector components.

$$\begin{aligned}\vec{b}_n^T(t) &= [b(t), b(t+T), b(t+2T), \ldots \\ &\quad \ldots, b(t+(n-1)T)].\end{aligned}$$

- As a concise notation we define $\vec{b}_n^T := \vec{b}_n^T(t)$.

- B: Capital letters represent the ensemble of $b(t)$.

- $B(\Delta T)$ specifies ensembles of $b(t+\Delta T)$.

- $\vec{B}_n^T$: Capital letters superscripted with an arrow denote the ensemble of vectors $\vec{b}_n^T(t)$.

3.1.1 The reconstruction dimension

Consider an n-dimensional vector ensemble generated by observing one output of a system. The vectors are constructed in the way given by Takens. Accordingly, it is no problem to construct state spaces of any given dimension. However, if a system has e g. a 3-dimensional attractor, a 5-dimensional state space does not contain more *information* about the system than a 3-dimensional state space. Based on this fact, it is possible to find out how many dimensions are necessary to describe an attractor.

How do we proceed in practice? If the vector ensemble $\vec{B}_n^T$ and the ensemble $B(nT)$ are known the redundancy $R_n'^T$ contained in $B(nT)$ can be calculated.

$$R_n'^T = I(B(nT), \vec{B}_n^T). \qquad (1)$$

$I(.,.)$ denotes the mutual information between the ensemble $B(nT)$ and the vector ensemble $\vec{B}_n^T$. Starting with $n = 1$, the quantity $R_n'^T$ is calculated for increasing n until saturation occurs, see figure 1. Let n_0 be the smallest index for which saturation is reached, $R_{n\geq n_0}'^T$ indicates the redundancy in the $(n+1)$st vector component. This component is not required for reconstruction anymore, it is totally irrelevant. In this curve the maximum dimension of the attractor is n_0.

The saturation value $R_{n\geq n_0}'^T$ has another important property; it is as high as the signal-to-noise ratio (or, the ratio of the deterministic signal component to the noise component) expressed in bit. In general, this is valid under the condition that the delay time $T \rightarrow 0^+$. It is clear that for $T = 0$ a singularity occurs with $R_n'^T\big|_{T=0} = +\infty$.

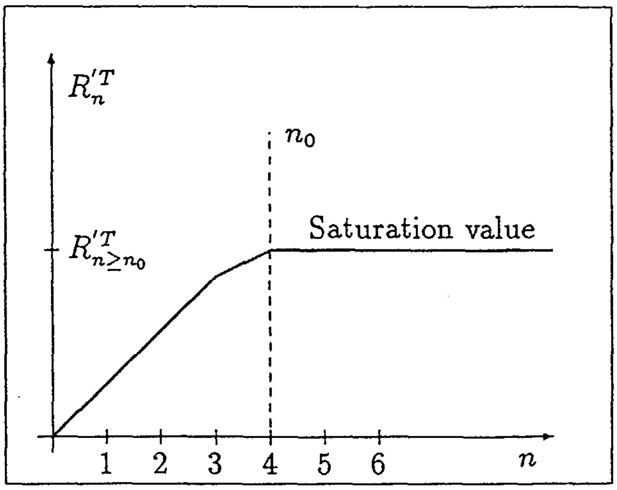

Figure 1: Schematic illustration of $R_n'^T$ as a function of n for chaotic signals

3.1.2 The delay time

To estimate the optimum delay time $T \doteq T_{opt}$, we let T vary. For $n \geq n_0$ we find a curve of $R_{n\geq n_0}'^T$ as a function of T which is called the saturation line (see figure 2). In this context, we use also the *total* redundancy R_n^T of the n-dimensional vector ensemble (see below equ. (3). The average redundancy per vector component is given by R_n^T divided by $(n-1)$. For the reconstruction of attractors we want to find a state space which contains as much information as possible.

It is impossible to find T_{opt} only by minimizing $\frac{R_n^T}{(n-1)}$ because in chaotic systems a steady information production takes place, i.e. $R_n^T \rightarrow 0$ as $T \rightarrow \infty$. The saturation line $R_{n\geq n_0}'^T$ as a function of T is, therefore, decreasing with increasing T, too. This negative slope indicates chaotic behaviour and is used for our decision chaotic/nonchaotic.

Hence we use a value of T that leads to the greatest distance between the curve of $\frac{R_n^T}{(n-1)}$ and the saturation line. In other words: The average redundancy per vector component has to be the smallest with respect to the maximal possible value (saturation line), cf. figure 2.

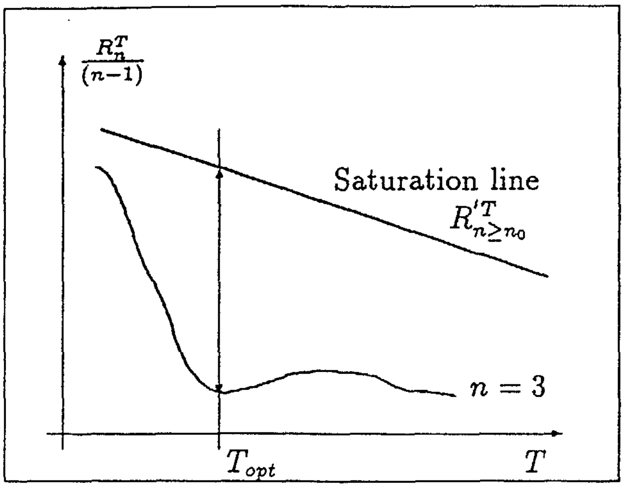

Figure 2: $\frac{R_n^T}{(n-1)}$ and $R'^T_{n\geq n_0}$ as a function of T

3.2 The Algorithm

The values of marginal redundancy are computed by the mutual information,

$$R'^T_n = I(B(nT), \vec{B}^T_n). \tag{2}$$

and the total redundancy as

$$R^T_n = \sum_{i=1}^{n-1} R'^T_i. \tag{3}$$

We can also write

$$R'^T_n = R^T_{n+1} - R^T_n \tag{4}$$

The problems in determining R^T_n are due to the problem of estimating probability density functions. We cannot determine continuous functions because we would have to consider an infinite number of continuous observations. The probability density function has to be approximated based on histogram classes. We have to find out suitable sizes for these classes in state space. It appears that no single size is best to determine the classes. If the size is fitted to the local conditions it is optimal. To this end, we define a class in the following way:

A range of signal space, which is considered as one class, possesses the property that the observed vectors are uniformly distributed over the class. It means that in such a class no further information is contained.

To find a partition of the state space into classes we begin to divide it up into 2^n hypercubes where n is the number of dimensions. Now, we proceed in that way for every single hypercube and divide it into 2^n further hypercubes. We continue this procedure recursively until a uniform distribution is found in a hypercube which is then accepted as a class. We stop dividing up this hypercube, but proceed in the other cubes where a uniform distribution was not been found yet. The distribution test has to be very carefully implemented because, on the one hand,no deterministic structure should remain hidden (otherwise information would be lost) and, on the other hand, random fluctuations should not be mistaken as a significant structure (otherwise the information content would be overestimated).

3.2.1 The distribution test

1. Consider a single hypercube at dividing depth m in the signal space. This hypercube is divided up into 2^n further cubes. Each of these hypercubes at depth $m+1$ is considered as one class and is used in a chi-square test [Spi75, p. 218] :

$$\chi^2 = \sum_{i=1}^{2^n} \frac{(N_i - Np_i)^2}{Np_i}, \tag{5}$$

where

- N is the total number of vectors in the hypercube at dividing depth m.
- p_i denotes the probability of the test distribution of the class i (i.e. the hypercube i at depth $m+1$) and
- N_i stands for the number of vectors observed in the class i.

We are testing for a uniform distribution. For a single class i, the probability is

$$p_i = \frac{1}{2^n}. \qquad (6)$$

We replace p_i in (5), hence

$$\chi^2 = \sum_{i=1}^{2^n} \frac{(N_i - \frac{N}{2^n})^2}{\frac{N}{2^n}} \qquad (7)$$

follows.

2. The test is only good for large amounts of data ($N > 30$). This property reduces the convergence of the algorithm. The chi-square test should be matched to our problem. Fraser uses a correction of the test based on an assumed multinomial distribution. For correction, the calculated distribution is related to σ_b^2 [Spi75, p. 216], the variance of the mutinomial distribution. This allows a more accurate determination of the distribution in areas with few observation vectors.

3. Another problematic point is the choice of the confidence level of the null hypothesis. It must be chosen right in between the two extreme cases, no structure will be detected or every single observation vector is accepted as a substructure. In both cases wrong results occur. Fraser found in his experiments an optimal confidence level of 20%. We have also experimented with several confidence levels and found, like Fraser, that 20% yields the best results.

4. Including the variance σ_b^2 for correction, the following test is performed:

   $$\sigma_b^2 \chi^2 < \chi^2_{20\%}. \qquad (8)$$

   If this inequality fails the distribution is not considered uniform. We conclude that in the investigated cube a substructure exists.

   The variance of the multinominal distribution is related to the number of classes by

   $$\sigma_b^2 = \frac{(2^n - 1)^2}{(2^n)^2} \qquad (9)$$

   The threshold value $\chi^2_{20\%}$ is usually taken from tables [BS83]. The value changes very much if the number of classes increases. If the number of classes is > 30, $\chi^2_{20\%}$ is calculated from an approximation formula [BS83, p. 690] .

5. The inequality (8) indicates whether a substructure exists or not. To verify this test another is done on the next finer dividing depth of 4^n hypercubes. If both inequalities are true the decision for a uniform distribution is accepted. If one test fails the division depth is increased and the distribution test has to be done in every of the 2^n smaller hypercubes and so on.

3.2.2 The equations of the algorithm

We want to compute the total redundancies R_n^T based on the marginal redundancies determined by (3). With (2) equation (3) can be written as

$$R_n^T = \sum_{i=1}^{n-1} I(B(iT), \vec{B}_i^T) \qquad (10)$$

and so

$$R_n^T = \sum_{i=1}^{n-1} H(B(iT)) - H(B(iT) \mid \vec{B}_i^T) \qquad (11)$$

$$= \sum_{i=1}^{n-1} H(B(iT)) + H(\vec{B}_i^T) - H(\vec{B}_{i+1}^T). \qquad (12)$$

In equation (12) we see: the entropy $-H(\vec{B}_{s+1}^T)$ in summation term ($i = s$) and the entropy $H(\vec{B}_s^T)$) in the summation term

$(i = s+1)$ eliminate each other and therefore we get

$$R_n^T = \sum_{i=1}^{n-1} H(B(iT)) - H(\vec{B}_n^T). \qquad (13)$$

The entropy of an ensemble can be written as

$$H(B(iT)) = -\sum_{K_m} P_{b_i}(R_m(K_m)) \\ \mathrm{ld} P_{b_i}(R_m(K_m)). \qquad (14)$$

b_i are the vector components in signal space and R_m defines a class in signal space indexed by the m-tupel K_m. The joint entropy is defined similarly. This results in R_n^T as

$$R_n^T = -\sum_{i=1}^{n-1}\sum_{K_m} P_{b_i}(R_m(K_m))\mathrm{ld}P_{b_i}(R_m(K_m)) \\ +\sum_{K_m} P_{b_1\ldots b_n}(R_m(K_m)) \\ \mathrm{ld}P_{b_1\ldots b_n}(R_m(K_m)). \qquad (15)$$

From this, we get a compact equation for the redundancy

$$R_n^T = \sum_{K_m} P_{b_1\cdots b_n}(R_m(K_m)) \\ \mathrm{ld}\frac{P_{b_1\cdots b_n}(R_m(K_m))}{P_{b_1}(R_m(K_m))\ldots P_{b_n}(R_m(K_m))}. \qquad (16)$$

The marginal distributions P_{b_i} have to be transformed into uniform distributions so that it is easier to calculate the several joint probabilities. After doing so we see that the marginal probability of one class is given at the dividing depth m by $(\frac{1}{2})^m$. Now we are allowed to separate the marginal distribution from the sum, therefore we can write

$$R_n^T = -m\,\mathrm{lg}\,(\frac{1}{2})^n + \\ \sum_{K_m} P_{b_1\cdots b_n}(R_m(K_m))\mathrm{ld}P_{b_1\cdots b_n}(R_m(K_m)). (17)$$

To find an efficient algorithm we pick out one hypercube in dividing depth m. This hypercube makes a contribution to the redundancy R_n^T which we denote as $R_n^T(R_m(K_m))$. In the following equations we use the abbreviation $P(R_m(K_m))$ instead of $P_{b_1\cdots b_n}(R_m(K_m))$.

$$R_n^T(R_m(K_m)) = -mP(R_m(K_m))\mathrm{ld}(\frac{1}{2})^n \\ +P(R_m(K_m))\mathrm{ld}P(R_m(K_m)) \qquad (18)$$

At depth $(m+1)$ the hypercube $R_m(K_m)$ is divided up into 2^n cubes denoted by (m+1)-tupels (K_m, j). Each of these elements contributes to $R_n^T(R_m(K_m))$:

$$R_n^T(R_{m+1}(K_m, j)) = \\ -(m+1)P(R_{m+1}(K_m, j))\mathrm{ld}(\frac{1}{2})^n \\ +P(R_{m+1}(K_m, j))\mathrm{ld}P(R_{m+1}(K_m, j)). \quad (19)$$

So we can calculate the redundancy $R_n^T(R_m(K_m))$ with the redundancies one dividing level deeper

$$R_n^T(R_m(K_m)) = \sum_{j=1}^{2^n} R_n^T(R_{m+1}(K_m, j)). \quad (20)$$

With (19) we can write

$$R_n^T(R_m(K_m)) = \\ \sum_{j=1}^{2^n}\Big[-(m+1)P(R_{m+1}(K_m, j))\mathrm{ld}(\frac{1}{2})^n \\ +P(R_{m+1}(K_m, j))\mathrm{ld}P(R_{m+1}(K_m, j))\Big]. \quad (21)$$

We simplify the terms containing the marginal distributions with

$$\sum_{j=1}^{2^n} P(R_{m+1}(K_m, j)) = P(R_m(K_m)) \qquad (22)$$

to get

$$R_n^T(R_m(K_m)) = -mP(R_m(K_m))\mathrm{ld}(\frac{1}{2})^n - \\ P(R_m(K_m))\mathrm{ld}(\frac{1}{2})^n + \sum_{j=1}^{2^n}\Big[P(R_{m+1}(K_m, j)) \\ \mathrm{ld}P(R_{m+1}(K_m, j))\Big]. \quad (23)$$

Now we see, stepping one dividing level deeper the term $P(R_m(K_m))\mathrm{ld}P(R_m(K_m))$ in equation (18) is replaced by

$$-P(R_m(K_m))\mathrm{ld}(\frac{1}{2})^n$$

$$+\sum_{j=1}^{2^n} P(R_{m+1}(K_m,j))\mathrm{ld}P(R_{m+1}(K_m,j)).$$

Hence we define a recursive function

$$F^*(R_m(K_m)) = -P(R_m(K_m))\mathrm{ld}(\frac{1}{2})^n$$
$$+\sum_{j=1}^{2^n} F^*(R_{m+1}(K_m,j)). \tag{24}$$

The recursion is performed until a uniform distribution is reached. The final hypercube is one class as defined before. The terminating equation for the recursion is similar to (18) without the first expression of the sum because the dividing level is contained in the recursion. The terminating equation is

$$F^*(R_m(K_m)) =$$
$$P(R_m(K_m))\mathrm{ld}P(R_m(K_m)). \tag{25}$$

The probability $P(R_m(K_m))$ can be expressed as a fraction of the observed state vectors in the class $N(R_m(K_m))$ over the total number of observations N_0.

$$P(R_m(K_m)) = \frac{N(R_m(K_m))}{N_0}.$$

From (24)

$$F^*(R_m(K_m)) = -\frac{N(R_m(K_m))}{N_0}\mathrm{ld}(\frac{1}{2})^n$$
$$+\sum_{j=1}^{2^n} F(R_{m+1}(K_m,j)). \tag{26}$$

and

$$F^*(R_m(K_m)) =$$
$$\frac{N(R_m(K_m))}{N_0}\mathrm{ld}\frac{N(R_m(K_m))}{N_0} \tag{27}$$

the factor $\frac{1}{N_0}$ can be extracted because $\frac{1}{N_0}$ is contained in all terms of the complete recursion. After that (27) may be written as

$$F^{**}(R_m(K_m)) = N(R_m(K_m))\mathrm{ld}N(R_m(K_m))$$
$$-N(R_m(K_m))\mathrm{ld}N_0. \tag{28}$$

We see that the second term does not depend on the structure, it is a constant contribution to the value of R_n^T

$$\sum_{K_m} N(R_m(K_m))\mathrm{ld}N_0 = N_0\mathrm{ld}N_0. \tag{29}$$

Hence results a start equation considering the factor $\frac{1}{N_0}$ and (29).

$$R_n^T =$$
$$\frac{1}{N_0}F(R_0(K_0)) - \mathrm{ld}N_0, \tag{30}$$

we proceed with the recursion

$$F(R_m(K_m)) =$$
$$nN(R_m(K_m)) + \sum_{j=1}^{2^n} F(R_{m+1}(K_m,j)) \tag{31}$$

and terminate if there exists no substructure with the terminating equation in every single class.

$$F(R_m(K_m)) =$$
$$N(R_m(K_m))\mathrm{ld}N(R_m(K_m)). \tag{32}$$

This recursive algorithm is the basic form to calculate the mutual information.

3.3 Implementation

The mutual information calculation is implemented on a IBM PC-AT 286 and on a MicroVax II minicomputer, respectively. The program is written in the programming language C. We implemented the computation of the recursions (30) to (32). The algorithm is very time and memory consuming. The third but most important point is that it is very sensitive to calculation accuracy. A disadvantage of Fraser's implementation of the algorithm is the huge amount of data needed to analyse a system. Even with 1 million sample values from a measurement the asymptotic saturation value is underestimated about 16%. We found out that this problem is mostly due to numerical effects. Next we optimized the implementation using a precedence list of arithmetic operations (summation with both signums, multiplication, division). Even if a division is within a summation it should be drawn out of summation by multiplication. We also have to take care of multiplication overflow. If we organize the computations such that no overflow occurs a sizeable amount of precision is lost even when using the IEEE floating point format. From our implementation, we could see that—when we process higher amounts of data—it is impossible to achieve all of the suggested numerical improvements simultaneously.

Last but not least the arithmetic unit is a very important factor in our system. We used the same software and the same number representation for our data on a PC with arithmetic coprocessor and without. Still, there appears a difference in the results of up to 10%.

Comparison to Fraser's implementation. The following values in percent are referred to the respective asymptotic saturation value of $R'^{T}_{n \geq n_0}$.

As mentioned before, Fraser reached with $1048576 = 1M$ samples only 16% accuracy, see [Fra89, figure 6]. We analysed the same quasiperiodic system and reached the saturation value within ±2%, moreover we used only $32k = 32768$ sample values. These results are obtained consistently over several runs of the analysis program. This provides for data saving by a factor of 16. If in Fraser's implementation the number of sample values is reduced to $32k$ the error even rises up to 26%.

The computation time needed in Fraser's and our implementations is approximately equal for the same amount of data. We expect improvements by optimizing the algorithm with respect to small sets of data.

4 Experiments with speech signals

Our experiments focus on long sustained vowels [a:, e:, i:, o:, u:] embedded in carrier sentences. They are spoken by three male native speakers of German, of age 24 to 30. We use sustained vowels to have sufficiently long stationary data for our analysis. The speech signals are recorded in an anechoic chamber and digitized with 16 bit resolution and 32 kHz sampling rate. All analyses are done on manually selected stationary segments of 32768 samples (i.e. roughly one second). There are several analysis tools available: the calculation of redundancy plots, computation of correlation dimension and visualization of three-dimensional plots of attractors. Time consuming analyses were done on a MicroVax II minicomputer.

4.1 Results

Figure 3 shows a three dimensional attractor projected on a two dimensional plane. The attractor correspondsto the vowel [i:] spoken by speaker 'F'and plotted with the optimum delay time $T_{opt} = 0.94$ msec. The graph displays three periodicities:

1. The trajectory almost replicates itself after every pitch period.

2. The first formant frequency is roughly twice the pitch frequency, so the overall structure of the graph is similar to a folded 8 (two cycles per 'pitch orbit')

3. The spectral peak related to the second and third formant frequencies is about 23 times the pitch frequency, so there are 23 small loops superimposed on the pitch orbit.

We estimated also the correlation dimension D of the attractor. The fractional value of $D = 1.7$ was found.

The overall structure of the attractor is maintained for other speakers. Figure 6 contains the attractor of vowel [i:] by speaker 'B'. The same looping structure as found for the [i:] spoken by speaker 'F' characterizes the attractor although the amplitude of the second-formant loops is significantly reduced due to the power differences in the respective formants of the two speakers.

It is very important to use the optimum delay time T_{opt}. We see in figure 4 again the vowel [i:] by speaker 'F', but with another delay time $T = 1.88$ msec. The optimal T-value produces results which are easier to interprete and most interestingly, this value is rather independent of the individual speaker or vowel under consideration. *The optimal delay time is on the order of 1 msec.*

The redundancy analysis of vowel [i:] by speaker 'F' is shown in figure 7. On display we have the marginal redundancies $R_n'^T$ for $n = 1, 2, 3$ and T between 1 and 500 sampling intervals (i.e. 0 to 15.6 msec). We see, the curves saturate at $n_0 = 2$. So we found the reconstruction dimension as $2 \times (n_0 - 1) + 1 = 3$. The slope of the saturation line is $h_\mu = 0.94$ bit over one pitch period. This is the information production rate of the underlying system.

Figures 5 and 8 show as a further example the reconstructed attractor and the redundancy plots for the vowel [a:] by speaker 'G'. While differing slightly in the numerical values, the general picture found in the previous figure is corroborated. For comparison, the pertaining measurement values are listed here: correlation dimension $D = 1.5$, optimal delay-time $T_{opt} = 0.94$ msec, information production rate $h_\mu = 0.9$ bit over one pitch period.

We give an overview of several of our measurements and calculations in tables 1, 2, and 3 where the SNR-value is obtained from $\lim_{T \to 0^+} R'^T_{n \geq n_0}$.

Table 1: VOWEL [i:] BY THREE SPEAKERS

Speaker	F	B	G
Pitch [Hz]	126	144	128
F 1 [Hz]	254	287	252
F 2 [Hz]	2330	2570	2450
SNR [dB]	28.88	36.37	24.2
T_{opt} [ms]	0.94	0.94	0.85
D	1.7	1.5	1.7
$h_\mu\ [\frac{Bit}{Pitchperiod}]$	0.94	1	0.6

Table 2: VOWEL [u:] BY THREE SPEAKERS

Speaker	F	B	G
Pitch [Hz]	126	146	128
SNR [dB]	30.9	40.8	26.41
T_{opt} [ms]	0.82	1	0.61
D	1.4	1.2	1.5
$h_\mu\ [\frac{Bit}{Pitchperiod}]$	0.84	0.65	0.76

Table 3: VOWEL [a:] BY THREE SPEAKERS

Speaker	F	B	G
Pitch [Hz]	115	117	126
SNR [dB]	45.7	35.75	45.5
T_{opt} [ms]	0.4	1.5	0.94
D	1.6	1.9	1.5
$h_\mu\ [\frac{Bit}{Pitchperiod}]$	0.9	0.9	0.9

5 Conclusion

Although the experimental basis is still rather limited, preliminary conclusions may be drawn.

- The gross shape of the reconstructed attractors can be interpreted in terms of standard phonetic theory. Their 'strangeness', however, goes beyond such theory which offers only explanations in terms of randomness or time-varying control.

- The investigated vowels can be interpreted in terms of deterministic chaos. This is supported through the positive information production rate of roughly one bit per pitch period and the fractional value of the correlation dimension between 1 and 2.

- As only sustained vowels with constant pitch have been investigated the observed chaotic behaviour can only be related to the temporal fine structure of speech that changes from pitch period to pitch period. No claim about the dynamic behaviour of spectral or articulatory parameters is made.

- In conclusion, if the results generalize to a statistically relevant group of speakers and to more speech sounds it should be possible to achieve more natural sounding speech synthesis by means of nonlinear signal models operating in the chaotic regime.

References

[BS83] I.N. Bronstein and K.A. Semendjajew. *Taschenbuch der Mathematik.* Teubner, Leipzig (Germany), 1983.

[Fla72] J.L. Flanagan. *Speech Analysis Synthesis and Perception.* Springer–Verlag, Berlin, Heidelberg , and New York, 2nd edition, 1972.

[Fra89] A.M. Fraser. Information and entropy in strange attractors. *IEEE Trans. on Information Theory*, IT-35(2):245–262, March 1989.

[FS86] A.M. Fraser and H.L. Swinney. Independent coordinates for strange attractors from mutual information.
Physical Review A, 33(2):1134–1140, Feb. 1986.

[KM90] A. Kumar and S.K. Mullick. Speech signal modeling *à la* chaos.
In *Proc. IEEE 1990 Digital Sig. Proc. Workshop*, paper 1.8, New Paltz (NY), Sept. 1990.

[MY90] P. Maragos and K.L. Young. Fractal excitation signals for CELP speech coders.
In *Proc. ICASSP'90*, pp. 669–672, Albuquerque (NM), April 1990.

[QH90] T.F. Quatieri and E.M. Hofstetter. Short-time signal representation by nonlinear difference equations. In *Proc. ICASSP'90*, pp. 1551–1554, Albuquerque (NM), April 1990.

[Sch90] J. Schoentgen. Non-linear signal representation and its application to the modelling of the glottal waveform. *Speech Communication*, 9(3):189–201, June 1990.

[Spi75] M.R. Spiegel. *Schaum's Outline of Theory and Problems of Probability and Statistics.* McGraw-Hill Book Company, New York etc., 1975.

[Tak81] F. Takens. Detecting strange attractors in turbulence. In D.A. Rand and L.-S. Young, eds., *Dynamical Systems and Turbulence, Warwick 1980*, pp. 366–381. Springer–Verlag, Berlin, 1981.

[Tis90] N. Tishby. A dynamical systems approach to speech processing. In *Proc. ICASSP'90*, pp. 365–368, Albuquerque (NM), April 1990.

[TT90] H.M. Teager and S.M. Teager. Evidence for nonlinear production mechanisms in the vocal tract. In W.J. Hardcastle and A. Marchal, eds., *Speech Production and Speech Modelling.* Kluwer Academic Publ., Dordrecht (The Netherlands), 1990.

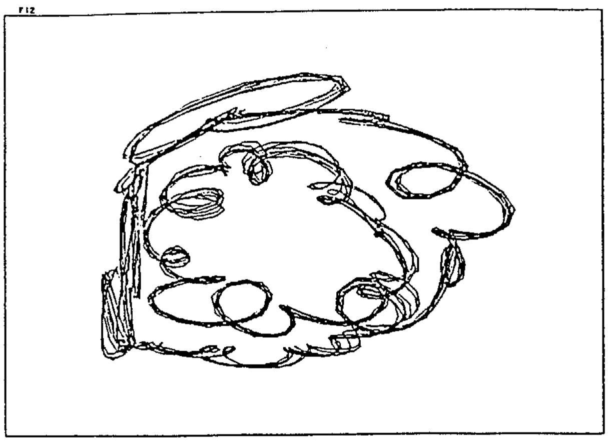

Figure 3: Attractor of vowel [i:] by speaker 'F' ($T_{opt} = 0.94$ msec).

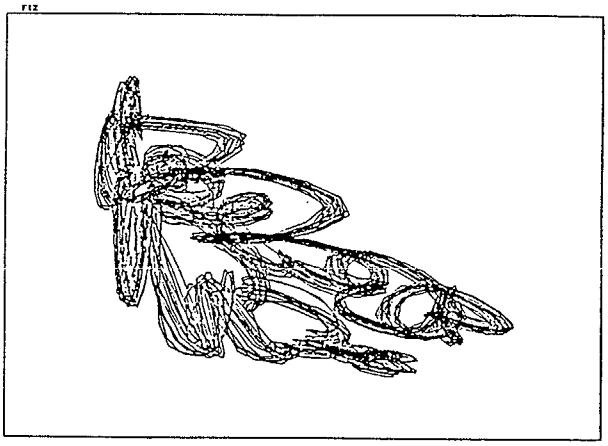

Figure 4: Attractor of vowel [i:] by speaker 'F' ($T = 1.88$ msec).

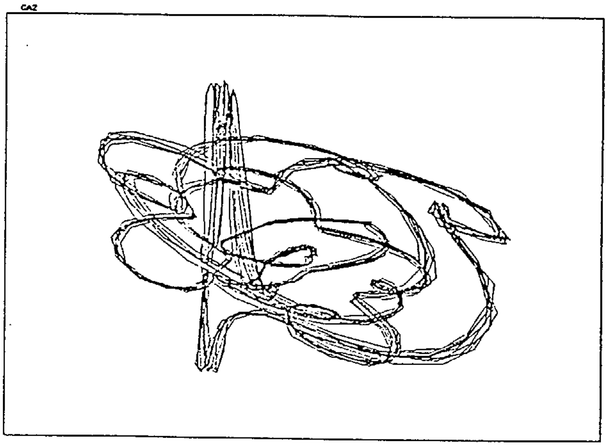

Figure 5: Attractor of vowel [a:] by speaker 'G' ($T_{opt} = 0.94$ msec).

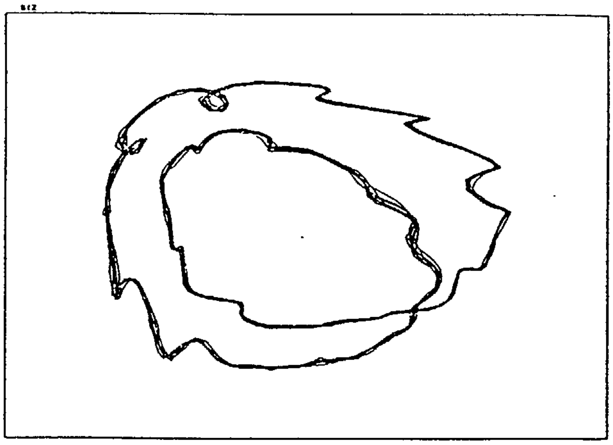

Figure 6: Attractor of vowel [i:] by speaker 'B' ($T_{opt} = 0.94$ msec).

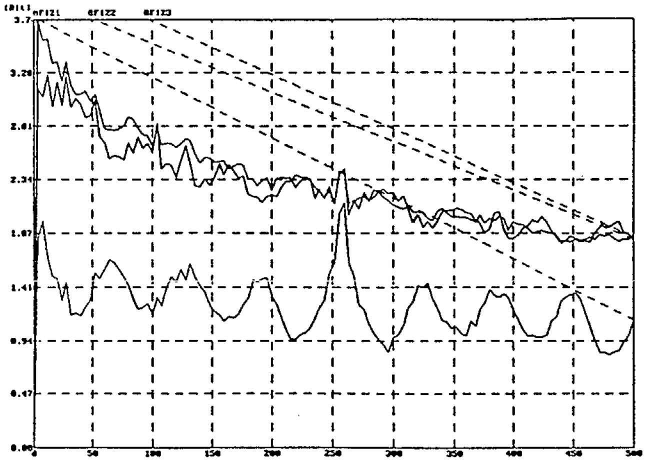

Figure 7: Marginal redundancy $R_n'^T$ of vowel [i:] by speaker 'F'.

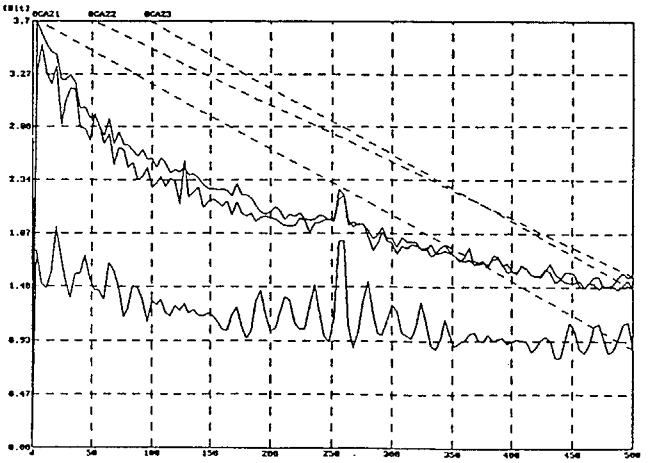

Figure 8: Marginal redundancy $R_n'^T$ of vowel [a:] by speaker 'G'.