
Copyright © 1991 IEEE

Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.
Dimension of the speech space

M. D. Alder, BSc, PHD, MEngSc
R. Togneri, BE, PhD
Y. Attikiouzel, BSc, PhD, MIEE

Abstract: Tattersal et al. have described an attempt to use the Kohonen algorithm for locating a hypothetical two-dimensional speech space in a space of filter-bank values. The existence of this space is moot and its dimension is more moot, and the Kohonen algorithm itself does not yield any information on the intrinsic dimension of the set to which it converges. There is therefore some interest in trying to decide by other means whether or not the speech space does have an intrinsic dimension of two. This paper falls into two main sections. In the first, we define a statistic for estimating the intrinsic dimension of a finite set of points on the assumption that they lie on a smoothly embedded manifold, when, of course, the dimension is an integer. We test the method on finite sets drawn from known manifolds and show that it is robust. We also apply it to the Lorenz attractor, which is a well known set of nonintegral dimension. Finally we apply it to speech data of the same type as that used by Tattersal et al. We conclude that the speech space is not discernibly a low-dimensional manifold at all, and that a more plausible hypothesis is that the space is an open subset of the enclosing space. In the second section, we construct a measure of the extent to which the surface that the Kohonen algorithm fits to the speech space is buckled or wrinkled related to the mean absolute curvature. The intention is to test (a) the hypothesis that the points of the speech space constitute a muralium, a two-manifold with noise, and (b) the hypothesis that the Kohonen process will find the muralium. We conclude that it is indeed possible to approximate the speech space with a low-dimensional manifold, but that it has dimension greater than two.

1 Introduction

Tattersal et al. [1] have described an attempt to use the Kohonen process [2] for locating a hypothetical two-dimensional speech space in a space of filter-bank values. It is conjectured that such a space exists, and moreover exists independently of the enclosing space, so that a space of, for example, LPC coefficients would also contain the same space differently embedded. If it exists, the Kohonen process will produce what we have called in Reference 3 a numerical parameterisation of the space, allowing a very considerable data reduction. The reasons for thinking the space exists are of two sorts. One is the familiar vowel chart known to phoneticians which places the common vowel sounds in a two-dimensional array. Two objections can be raised: first, that there is no room in this space for nasals or for other nonvowels; and secondly, that in any case there is a third dimension of lip roundedness (as exemplified by the French /eu/ sound contrasted with the vowel sound at the end of /kangaroo/). To the latter it can be replied that the dimension in this case might be three rather than two, but at least it is still very much lower than the enclosing space. The other argument in favour of a two-dimensional speech space is neurological, and depends upon the observation that the cortex is laminar in structure and therefore the speech processing part of the brain maps the signal into a planar array. Objections to this are: first, the inherent dimension of the processing elements is not necessarily related to the dimension of the data represented, an example being the linear address space of a computer which can still represent data such as trees and multidimensional arrays; and secondly, that speech is a small part of the acoustic signal, and the argument for believing the speech part is two dimensional would presumably apply to the whole acoustic signal, and it is hard to believe that an intrinsically two-dimensional representation is adequate for this.

The existence of this space as a low-dimensional manifold then is moot, and the Kohonen algorithm itself does not yield any information on the intrinsic dimension of the set to which it converges. There is therefore some interest in trying to decide by other means whether or not the speech space does have an intrinsic dimension of two, or if not two then some number small compared with the dimension of the enclosing space. More generally, the space might be merely approximable by a manifold. An object such as a pancake is not a two-manifold but is what is called in De Hoff [4] a muralium, and it is likely enough that a finite set of data from speech would be affected by noise and would at best be only approximately low dimensional. More thought about the nature of the vocalisation process leads to the conjecture that there might well be a low-dimensional vowel space, but that the regions corresponding to nasals would occupy some largely disjoint region of perhaps different dimension; similarly for other phonemic elements. Natural pessimism, on the other hand, leads to the null hypothesis that there is no particular structure to the speech space and that it is merely an irregular blob in the enclosing space.

Between the optimism that the speech space is a two-dimensional manifold, and the pessimism that it is a structureless blob, lies the middle ground that it may be approximated tolerably well by some low-dimensional
manifold. It is of some interest to determine where the
truth lies. If an approximation by a low-dimensional
manifold is feasible then the problem of quantising the
space efficiently becomes easier. Also, the belief of pho-
eticians that speech may be characterised in a speaker-
independent way using a relatively small number of
parameters becomes much more plausible, and the task
of translating between the categories of phoneticians and
features of the signal becomes accessible; while if it
appears that the speech space is essentially featureless,
then this gives support to current engineering practice of
trying to store trajectories in the space, and suggests that
phoneticians' categories are incapable of being made
precise.

This paper falls into two main sections. In Section 2
we investigate the strong hypothesis that the speech
points actually lie very close to a low-dimensional mani-
fold. In Section 3 we investigate the weaker hypothesis
that the set many be approximated globally by a low-
dimensional manifold. In both cases we take it that we
have to dispose of a null hypothesis that the space is a
random collection of points in some region of the space,
and that we have to devise statistical measures of good-
ess of fit to an unknown manifold. We are obliged there-
fore in both cases to test the measures we devise on
known data sets.

Other workers have written on estimating the dimen-
sion of a set on the basis of a finite sample. Farmer et al.
[5] is an example, and the main application has been to
the case where the dimension is not expected to be an
integer, the case of sets of fractal dimension. It has been
studied in the case where the set is a strange attractor,
when it is not infrequently problematic that the set has a
dimension that is not an integer. Here we take it that
Mandelbrot [6] is, of course, the genesis of much current
work on fractal dimension. There are a small number of
standard methods for measuring the dimension or the
Hausdorff measure of finite sets; the primary one is to
look at the number of points in a ball of radius \(x\) as a
function of \(x\). If the set is of dimension \(k\) and is flatly
embedded in \(\mathbb{R}^n\), then we expect the count to go up as the
\(k\)th power of the radius; we may plot the logarithm of the
count against the logarithm of the radius and hope to
find a constant slope line; the gradient is then the dimen-
sion. Other methods are essentially derived from this
one; we may look at the distribution of distances between
points, for example. The methods suffer from the draw-
back that they do not behave well when the embedding is
not flat, or when the distribution of points is nonuniform.
See Farmer et al. [5] for a discussion on these matters.

In Section 2, then, we define a local statistic for esti-
mating the intrinsic dimension of a finite set of points on
the assumption that they lie on or at least close to a
smoothly embedded manifold, in which case, of course,
the dimension is an integer. Then locally the manifold is
flat, and if we have a sufficiently large number of points
we can hope to approximately triangulate the manifold
using the points. The deviation of the manifold from the
points in each simplex will be small and will decrease as
the density of the points increases. We propose to esti-
mate these deviations. Specifically, we test the plausibility
that the points were drawn from a manifold of dimension
\(k\) embedded in \(\mathbb{R}^n\) by evaluating a statistic which has an
expectation that tends to zero as the point density
increases if the hypothesis is true, and a strictly positive
limiting expectation otherwise.

We test the method on finite sets drawn from known
manifolds and show that it is robust. We also apply it to
the Lorenz attractor, which is a well known set of non-
integral dimension. Finally we apply it to speech data of
the same type as that used by Tattersal et al. [1]. We
conclude that the hypothesis that points of the speech
space are drawn from a low-dimensional manifold is not
tenable.

In Section 3 we investigate the hypothesis that the
points are drawn from a generalised muralium, a set
which in a neighbourhood of each point may have the
dimension of the enclosing space but which may be glob-
ally approximable by a low-dimensional manifold. An
example would be a pancake or a sheet of paper, a
which although at some scales appears to be three dimen-
sional in the neighbourhood of one of its points, is clearly
globally approximable by a surface. The statistic we use here
employs the Kohonen process used in Reference 1 to
immers a grid of dimension \(k\) in the speech space, and
then measures the mean amount of crinkling or buckling
of the grid once it has converged. It is intuitively plain
that if a set of points lies close to a surface, then a linear
one-dimensional grid approximating it will require a lot
of folding; similarly a three-dimensional cube approxi-
mating the surface would require a large amount of
buckling to squash it into two dimensions, while a rec-
tangular grid of dimension two could be fitted to the
surface with a relatively small amount of crinking. By
measuring what we have described informally as the
amount of crinking, actually the mean absolute curva-
ture of the grid, we produce a statistic for testing the
hypothesis that the speech space is approximable by a
\(k\)-dimensional manifold globally. It is of course necessary
to establish that the statistic behaves in the manner that
intuition would lead one to expect, and so it was tested
on the same set of data points used in Section 2.

We also ran the speech data through a number of
dimension estimators developed by others [7, 8] for use
on strange attractors. The results of all methods indicate
that the speech space is not a featureless blob. Our
methods suggest that it is approximable by a relatively
low-dimensional manifold, but that the dimension is
greater than two.

The speech data were obtained by taking frames of
25 ms duration advanced by 10 ms sampled at 10 kHz
with 16 bit precision, and performing an FFT yielding
128 points from 0 to 5 kHz. We reduced the dimension
to 12 by two different methods: one was a mel spaced set
of intervals overlapping by one quarter from 100 to
5000 Hz; the other was a uniformly spaced non-
overlapping set of intervals from 0 to 5 kHz. We used
5392 frames for one speaker, and then augmented the set
by adding another speaker. The speech comprised com-
plete utterances of the digits in English. Extended speech
consisting of more speakers and a comprehensive
phoneme range to over 16000 points was also used. It
was found that neither the different choices of representa-
tion of the data nor the amount of data used made any
substantive difference to the results.

2 The local statistic

We shall assume that the set \(X\) is given as a set of points
in \(\mathbb{R}^n\) for fixed \(n\), and consider the hypothesis that \(X\) is a
finite sample of points from a \(k\)-manifold \(M\) smoothly
embedded in \(\mathbb{R}^n\). Take a point \(p\) of \(X\) and hence of the
putative \(k\)-manifold. Choose its \(k + 1\) nearest neighbours
in \(X\). Then we expect the distance of the selected point
\(p\) from the \(k\)-simplex \(S\) spanned by the \(k + 1\) nearest neigh-
bours to be small compared with the distance we should
expect if the dimension exceeded $k$. If the embedding is flat it will indeed be zero.

In the case where we have a surface of dimension two, for example, we expect that the point $p$ will be 'close to' the plane determined by the two-simplex, a triangle consisting of the three nearest neighbours of $p$, but not that it will be particularly close to the line segment (one-simplex) determined by its two nearest neighbours. Again, if the surface really is flatly embedded, i.e. if it is a plane, then the distance from the plane of the triangle will be zero.

Since we want a deviation which is scale invariant, we require that the unit of measurement of the distance be given by the size of the $k$-simplex; we take the average of the lengths of the sides of the simplex, and divide the distance of the point from the simplex by this value. Clearly the result is scale invariant. We also observe that the deviation of an equilateral $k$-simplex, defined as the shortest distance between any vertex and the opposite face of the $k$-simplex, is the square root of $(k-1)/2k$. We divide by this number also in order to compensate for an obvious effect on the distance of increasing the dimension.

Now we consider the distribution of these deviations. If the manifold really is $k$-dimensional and we measure the deviation from neighbouring $k$-simplexes, we expect the (relative) deviation to be close to zero and to get closer and closer to zero as the number of points taken from the manifold increases and their mean separation decreases; this must happen if the manifold is smoothly embedded. If on the other hand the manifold has dimension greater than $k$, then we expect the relative deviation to be larger, and not to depend on the density of points. It is not immediately clear what the expected deviation will be; specifically, if we take points in the unit hypercube in $R^n$, say, with uniform density, and ask what is the expected value of our deviation as a function of $k$, about all that can be easily said is that if $k = n$ it is zero, otherwise it is positive, not greater than one, and will be a decreasing function of $n$. This is because we can see that the distance of $p$ from $S$ can never be greater than the distance of $p$ from the most remote vertex of $S$, and can be less than the distance of $p$ from the closest vertex of $S$. Its expectation therefore is not greater than the diameter of the equilateral $k+1$ simplex, which is one, by virtue of our normalisation. Moreover, as $k$ increases the likelihood that the $k+1$ nearest neighbours of $p$ will contain a point that is very close increases, and the deviation is sensitive to the nearer points much more than the remote points. Although it is possible in principle to calculate the expectations and distributions for the cases of different $k$ of a uniformly distributed randomly generated set of points in a cube, it is very much simpler in practice to estimate them by Monte Carlo methods, which is the course we have followed here.

### 2.1 Data

A C program was run under UNIX on a Silicon Graphics IRIS computer to compute deviations for various data. The first tests consisted of choosing randomly points which lie on flat manifolds of dimension $k$, for the cases where we have a surface of dimension one and two, where the points were taken to be in $R^n$ for $n$ taking various values. We also performed the Monte Carlo estimate of the expected results of a 12-dimensional set of points in a cube in $R^{12}$.

There was no difficulty in distinguishing the dimension of flat submanifolds of low dimension (hyperplanes) even with small numbers of points:

$P[1][2]$ was defined as the set $\{x_1, x_2, x_3 \in R^3: 0 < x_1, x_2 < 100, x_3 = 0\}$, where $x_1, x_2$ were generated randomly.

$P[2][2]$ was defined as the set $\{x_1, x_2, x_3 \in R^3: 0 < x_1, x_2 < 100, x_3 = 50x_1 + 50x_2\}$, where $x_1, x_2$ were generated randomly.

In both cases the deviation dropped to zero for $k \geq 2$, precisely what one would expect.

We next tried the method on $k$-spheres in $R^n$ as before, again for different numbers of points. We did this using two probability distributions: the first, giving a uniform density on the sphere, was obtained by choosing random points throughout the $(k+1)$-cube and normalising; the second was obtained by taking random points uniformly on the $k$-cube and projecting via a Mercator projection, which gives a density which increases towards the poles of the $k$-sphere. Again, there is no difficulty in distinguishing the dimension:

$S[1][2]$ is defined as the set $\{(x_1, x_2, x_3) \in R^3: -100 \leq x_1 \leq 100, -\sqrt{(100^2 - x_1^2)} \leq x_2 \leq \sqrt{(100^2 - x_1^2)}, x_3 = \sqrt{(100^2 - x_1^2 - x_2^2)}\}$

$S[2][2]$ is defined as the set $\{(x_1, x_2, x_3) \in R^3: 0 \leq \theta \leq \pi/2, 0 \leq \phi \leq 2\pi; x_1 = 100 \cos \theta \cos \phi, x_2 = 100 \cos \theta \sin \phi, x_3 = 100 \sin \phi\}$.

$x_1, x_2, \theta$ and $\phi$ were generated uniformly at random, subject to the given constraints.

We also experimented with tori of various dimensions. As the dimension $k$ increases, the number of points required to establish that our statistic is indicating a significant difference goes up, with a consequent increase in compute time. For this reason we never went beyond dimension four in $R^8$. We looked at a torus embedded in a conventional manner in three-space, and also the same torus (or a quadrant of it) embedded in four-space with a nonuniform distribution of points. We then took the three-torus, topologically $S^1 \times S^1 \times S^1$, and the four-torus, $S^1 \times S^1 \times S^1 \times S^1$, embedded in $R^6$ and $R^8$ respectively in the obvious manner. The sets were defined as follows:

$T[1][2] = \{(x_1, x_2, x_3) \in R^3: 0 \leq \theta \leq 2\pi, 0 \leq \phi \leq 2\pi; x_1 = \cos \theta (150 + 50 \cos \phi), x_2 = \sin \theta (150 + 50 \cos \phi), x_3 = 50 \sin \phi\}$

$T[2][2] = \{(x_1, x_2, x_3) \in R^3: -100 \leq x_1 \leq 100, x_2 = \sqrt{(100^2 - x_1^2)}, -100 \leq x_3 \leq 100, x_4 = \sqrt{(100^2 - x_3^2)}\}$

$T[3][3]$ and $T[4][4]$ were defined in the same way as $T[2][2]$, as a three-torus and four-torus in $R^6$ and $R^8$ respectively.

We considered the case of a nonmanifold, the well known Lorenz attractor which has (fractal) dimension just over two. It is not to be expected that a method which depends upon the points having come from a smooth manifold will yield unambiguous results, but it might be concluded that the result is consistent with naive expectations. It is noteworthy that, as with speech data, there is a sense in which the intrinsic dimension is one, as the data come from continuous movement in a higher-dimensional space which has been sampled. This is apparent in the results below; nevertheless, we see that the gaps are filled in satisfactorily by different orbits, and a two-dimensional model accords better with the results than a one-dimensional one.

We also considered a rather badly embedded surface of dimension two in 12-space, as we wanted to test the statistic in a case which might be close to the actual
results if the speech space should indeed constitute a two-

Finally, we took speech data obtained as described in the

2.2 Results

The histograms of Fig. 1 show the distribution of devi-

Table 1: 12-dimensional set in $\mathbb{R}^{12}$

<table>
<thead>
<tr>
<th>$k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>0.71</td>
<td>0.62</td>
<td>0.57</td>
<td>0.53</td>
<td>0.50</td>
<td>0.46</td>
<td>0.42</td>
<td>0.39</td>
<td>0.34</td>
<td>0.27</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Table 2: Two-sphere $S[2][2]$

<table>
<thead>
<tr>
<th>$k$</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>0.67</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 3: Two-toruses

<table>
<thead>
<tr>
<th>$T[1][2]$</th>
<th>$T[2][2]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>1</td>
</tr>
<tr>
<td>$m$</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Table 4: Three- and four-toruses

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>1</td>
</tr>
<tr>
<td>$m$</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Consistently, the standard deviation was about the same

This begs the question of what the expected means ought to be. Since we are interested not in dimen-

Table 6: Two-dimensional manifold $M[1][12]$
2.3 Conclusions

In Section 2 we have described a local statistic for testing the hypothesis that a finite set of points in $\mathbb{R}^n$ was drawn from a smoothly embedded $k$-manifold for some $k < n$. We have tested the relevant statistic on simulated data of a variety of types, and shown that it behaves as one might reasonably expect. We have applied this statistic to speech data with the aim of testing a hypothesis implicit in some work of Tattersal et al. [1], that there is an intrinsically defined speech space of dimension two.

The results make it plain that with about 6000 points the local flatness of the manifold is not apparent. We cannot, of course, rule out the possibility that the space in fact is a low-dimensional manifold; however, if it is, this structure is not discernible with this number of points. By comparison, only 128 points suffice to detect the two-dimensionality of a known two-manifold also in $\mathbb{R}^2$, and moreover one which is far from a flat embedding.

We infer that either the speech space is so embedded that it is extremely high local curvature and hence needs a prohibitively large number of points to detect this structure, or it is not in fact a two-manifold at all. In the former case, this imposes a bound on the size of the grid used in the Kohonen algorithm, and also on the number of data points used to attract this grid; the crudest possible scaling argument suggests that the minimum number of points in the space before any such structure might be found will be of the order of 30000.

This leaves the possibility that the local statistic is affected by noise and that there is an approximate manifold structure. The possibility we wish to investigate next is that the speech space, although not, in the light of Section 2, definitively a surface, might be a muralium in the sense of DeHoff [4] or some higher-dimensional generalisation. If so, it is possible that the Kohonen algorithm might still yield a useful approximation to the speech space. It is far from clear that the Kohonen process will find anything like a least squares best fit surface to a finite set of points obtained from a muralium; in effect we test this at the same time.

3 The global statistic

Here we measure directly the amount of crinkling or buckling of the approximating Kohonen grid. We compare the amount of crinkling for the speech space with that obtained for a randomly generated set of points in a similar sized region of $\mathbb{R}^2$, and also for sets of points known to lie on the manifolds of various dimensions discussed in Section 2. We also produce points drawn from manifolds to which 12-dimensional noise has been added.

There are many possible measures of what we have referred to informally as the amount of crinkling of the grid. In Baddeley and Averback [9] and DeHoff [4] the merits of the total mean curvature are discussed in a similar context. Our crinkle index is related to the mean absolute curvature but is simpler to compute for the case of a rectangular grid embedded in $\mathbb{R}^2$, and is insensitive to the nonuniformity of the distribution of the points.

If it should turn out that the amount of crinkles for the Kohonen grid is significantly less for the speech space than for a random collection of points drawn uniformly from a cube in $\mathbb{R}^2$, we may consider the muralium hypothesis to be tenable.

Again, we suppose that there is a set of points $X$ in $\mathbb{R}^n$ for some fixed $n$, and we consider two hypotheses: first, the null hypothesis that the set is obtained by taking points uniformly at random inside a cube in $\mathbb{R}^n$; and secondly, the hypothesis that there is a surface in the enclosing space, say $M$, such that the points are all close to $M$. Clearly the alternative hypothesis can never be shown to be wrong; it can however be shown to be unnecessary, while the null hypothesis can only be shown to be implausible.

We adopt as an auxiliary hypothesis the assumption that the Kohonen algorithm will give a reasonably good fit of the grid to any muralium of points. We will give some experimental evidence to support this hypothesis, but we point out that the attempt to find the speech space by means of the Kohonen algorithm also makes this assumption. If it is false, then again doubt is cast on the methods of Tattersal et al. [1], although from a rather different direction.

The crinkle index is computed as follows. For a rectangular grid we can take, for each point $C$ not on the boundary, four neighbours $N$, $S$, $E$, $W$. We may take it that in a flat embedding of the grid in a plane, the angle subtended by the line segments $NS$ and $EW$ is $180^\circ$; likewise the angle subtended by the line segment $NW$; the deformation of the grid is maximal in the NS direction if the angle is zero; likewise in the EW direction. We therefore measure the cosine of half the angle from $N$ to $C$ to $S$, and to simplify the arithmetic take its square. This number is of course zero if the three points lie in a straight line with $C$ at the centre, and one if $N$ and $S$ are in the same direction from $C$. No account is taken of the length of the lines. The same number is computed for the points $E$, $W$, $C$ and the results added. Thus we obtain, for the case where $C$ is at the origin, the following simple expression for the crinkle at 0 of a grid point:

$$c = 1 + \frac{\langle n, s \rangle}{n |x|} + \frac{\langle e, w \rangle}{2 |e| |w|}$$

where $\langle x, y \rangle$ denotes the usual inner product and $|x|$ denotes the derived norm.

This certainly gives some sort of measure of the extent to which the grid is not a regular rectangular array. By summing over all centre points $C$ in the interior of the grid and dividing by the number of such points, we obtain a mean crinkle per grid point. We observe that the resulting index is invariant under rotations, translations and similarities. The generalisation to higher dimensions is obvious.
3.1 Data

Whether or not the index is a good measure depends on the extent to which it gives results which are compatible with our expectations on known data sets. We therefore ran a program implementing the Kohonen algorithm on the data sets described in Section 2. These comprised two planes in $R^2$; two two-spheres in $R^3$; two two-tori, one in $R^2$ and one in $R^3$; a badly embedded surface in $R^{12}$; a three-torus in $R^4$; a four-torus in $R^8$; a randomly generated set of points with a uniform distribution in a cube in $R^{12}$; the Lorenz attractor in $R^3$; a plane in $R^{12}$; and finally the space of speech data obtained from a number of utterances from one speaker obtained as described in the introduction and embedded in $R^{12}$. We also investigated the results for a speech space obtained from two speakers, and the effect of additive noise on points taken from a manifold. The program was written in C and run under UNIX on a Silicon Graphics IRIS microcomputer. If the speech space formed a muralium, if it was 'thin' in all but two dimensions, and if the Kohonen process fitted a surface to it adequately, then it might be hoped that the data would reveal this. More generally, if the speech data could be approximated adequately by some $k$-cube of dimension higher than two but less than that of the enclosing space, we could hope to determine a suitable $k$.

3.2 Results

The crinkle per grid point, defined as above, was calculated for a $10 \times 10$ grid, a $32 \times 32$ grid, a $50 \times 50$ grid, a $100 \times 100$ grid and also for a three-dimensional grid $10 \times 10 \times 10$, all using the Kohonen algorithm of Reference 1. Table 8 summarises the mean crinkle per grid point when the $(32 \times 32)$ grid is supposed to converge to a data set of about $16000$ points drawn from a (two-dimensional) plane in $R^{12}$. The repetitions involved both a new set of points randomly drawn from the plane and a new initial location of the grid.

When the data points were chosen at random from the hypercube in the whole space under the same conditions, the values in Table 9 were obtained.

It is clear that the crude methods we employ are adequate to distinguish extreme cases. It is difficult to compute an expected crinkle index for even the simplest cases. It is clear that when the attracting points are on a plane the index may be expected to be low for a two-dimensional grid, higher for a three-dimensional grid. When the points are randomly scattered in $R^{12}$ it can be expected that much more folding of the grid will occur and that this will yield a higher index, as is indeed the case. One would expect that a three-dimensional set would yield a lower crinkle index for a three-dimensional grid than for a two-dimensional grid. We computed the index for a number of sets of known dimension, embedded in $R^d$ for various $n$. The variance of the crinkle index is comparable in all other cases, and from now on we give means to a more realistic precision.

We found that with smaller grids the mean crinkle per point was always lower than with larger grids; we investigated $10 \times 10$, $32 \times 32$, $50 \times 50$ and $100 \times 100$ grids, with the larger grids usually exhibiting a higher mean index. The exception was the 'badly embedded surface', a two-manifold which we had embedded in $R^{12}$ in a highly nonlinear way; the $10 \times 10$ grid showed a high crinkle index, which we ascribed to the structure being too crudely approximated. We concluded that the usually higher index average per grid point for larger grids was measuring the number of data points in the set. We believe that with a smaller alpha (step size) parameter and more points we would have obtained more similar numbers, but this would have increased the time taken to obtain the results considerably. We therefore compromised on a $32 \times 32$ grid for the two-dimensional fitting, and $10 \times 10 \times 10$ for the three-dimensional case.

We first tested the index on a number of sets of known dimension; these sets have been, in the main, described in Section 2. We defined planes, two-spheres and two-toruses embedded in $R^3$, $R^4$ and $R^{12}$. Points were obtained from these by taking random numbers and parameterising the manifolds. In addition, we used data obtained from generating the Lorenz set in $R^3$, which is a strange attractor with a fractal dimension of about 4.1; we took a three-torus in $R^4$ and a four-torus in $R^8$; and finally we took the 'twisted surface' $M12$, the two-dimensional surface embedded in $R^{12}$ by a highly non-linear embedding alluded to above. The results of repeating five times with randomised starting locations for the grid points and a new set of randomly generated points of the set is shown in Table 10, both for the $32 \times 32$ two-dimensional grid and also for the $10 \times 10 \times 10$ three-dimensional grid. The precision is approximately indicated by rounding the last digit given, and the standard deviation was comparable with those in Tables 8 and 9.

Recall that the two planes in $R^d$ differed in their orientation only. The two spheres differed in the density of points on them (and one was a hemisphere only). The three-torus and four-torus were symmetrically embedded in $R^d$ and $R^d$ in the obvious way. The two-torus embedded in $R^d$ was of course asymmetric.

It can be seen that the index is behaving in a manner which accords well with expectations. It is lower for the three-dimensional grid on a three-dimensional set than on a two-dimensional set; it is higher for the two-dimensional grid on a curved surface such as a torus or the badly embedded surface; but it seems to be largely independent of the enclosing dimension.

The results of applying the two-dimensional grid to random points in a cube in $R^{12}$ have been given in Table 9. It has been suggested that random number generators

<table>
<thead>
<tr>
<th>Table 8: Two-plane in $R^{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$32 \times 32$ 2D rectangular grid, Euclidean norm, sequential access; alpha = 0.1; neighbourhood = 16; ordering steps = 50000; convergence steps = 200000</td>
</tr>
<tr>
<td>P32 rr res</td>
</tr>
<tr>
<td>Crinkle per neuron = 0.003620, 0.003866, 0.003475, 0.003596, 0.003690</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 9: Random noise in $R^{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>R12 rr res</td>
</tr>
<tr>
<td>Crinkle per neuron = 0.275382, 0.265168, 0.272776, 0.270593, 0.285478</td>
</tr>
</tbody>
</table>
are at best chaotic rather than random, and that we should ensure that anomalies were not being obtained from this source. We therefore conducted a separate experiment with two random number generators, and some 'genuine' random numbers obtained from cobalt-60 decay counts. It was found that there was no distinguishable difference in the resulting crinkle indices of the grids, and we therefore concluded that our random number generator was adequate for these purposes.

We explored the extent to which the Kohonen process would find a good fit to noisy data by adding random noise to the points taken from a plane to see how this affected the crinkle index. Our hope was that the process would still yield a reasonable fit to a muralium, and that the speech space is not one but does not have the featureless characteristics of a uniform distribution in the enclosing space. There is some reason for conjecturing that the dimension may be about four, in so far as the set can be said to have a dimension.

Two conclusions are apparent. First, the hypothesis that the speech space is a structureless uniform set of points randomly distributed through a cube is not tenable, although this information comes mainly from the three-dimensional grid. Secondly, the hypothesis that the space is a two-manifold may also be rejected. This latter is in accord with our earlier results. We note that the results most similar to those for the speech space obtained from manifolds are those from the four-torus.

### 3.3 Noise

We may conclude from Tables 10–12 that the Kohonen process can produce a reasonable fit to a muralium, and that the speech space is not one but does not have the featureless characteristics of a uniform distribution in the enclosing space. There is some reason for conjecturing that the dimension may be about four, in so far as the set can be said to have a dimension.

<table>
<thead>
<tr>
<th>Space</th>
<th>Dimension of space</th>
<th>Dimension of enclosing $\mathbb{R}^3$</th>
<th>32 $\times$ 32 grid, mean crinkle index</th>
<th>10 $\times$ 10 $\times$ 10 grid, mean crinkle index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plane 1</td>
<td>2</td>
<td>3</td>
<td>0.064</td>
<td>0.016</td>
</tr>
<tr>
<td>Plane 2</td>
<td>2</td>
<td>3</td>
<td>0.004</td>
<td>0.016</td>
</tr>
<tr>
<td>Plane 3</td>
<td>2</td>
<td>12</td>
<td>0.004</td>
<td>0.017</td>
</tr>
<tr>
<td>Two-sphere 1</td>
<td>2</td>
<td>3</td>
<td>0.014</td>
<td>0.01</td>
</tr>
<tr>
<td>Two-hemisphere 2</td>
<td>2</td>
<td>3</td>
<td>0.014</td>
<td>0.008</td>
</tr>
<tr>
<td>Two-torus 1</td>
<td>2</td>
<td>3</td>
<td>0.04</td>
<td>0.08</td>
</tr>
<tr>
<td>Two-torus 2</td>
<td>2</td>
<td>4</td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td>Three-torus</td>
<td>3</td>
<td>6</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>Four-torus</td>
<td>4</td>
<td>8</td>
<td>0.19</td>
<td>0.1</td>
</tr>
<tr>
<td>Nonlinear surface M12</td>
<td>2</td>
<td>12</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>Lorenz attractor</td>
<td>2.1</td>
<td>3</td>
<td>0.08</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Table 10: Crinkle indices for sets of known dimension

<table>
<thead>
<tr>
<th>Space</th>
<th>Noise level $q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.20</td>
</tr>
<tr>
<td>2</td>
<td>0.20</td>
</tr>
<tr>
<td>4</td>
<td>0.20</td>
</tr>
<tr>
<td>6</td>
<td>0.20</td>
</tr>
<tr>
<td>8</td>
<td>0.20</td>
</tr>
<tr>
<td>10</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Table 12: Effects of noise on two-dimensional data: crinkle per neuron

We explored the extent to which the Kohonen process would find a good fit to noisy data by adding random noise to the points taken from a plane to see how this affected the crinkle index. Our hope was that the process would still yield a reasonable fit to a muralium, and that the speech space is not one but does not have the featureless characteristics of a uniform distribution in the enclosing space. There is some reason for conjecturing that the dimension may be about four, in so far as the set can be said to have a dimension.

<table>
<thead>
<tr>
<th>Space</th>
<th>Extra speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>0.20</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Table 13: Normalised versus unnormalised speech data

<table>
<thead>
<tr>
<th>Crinkle factor without normalisation</th>
<th>Crinkle factor with normalisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Noise level $q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Noise level $q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Space</th>
<th>Noise level $q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.20</td>
</tr>
<tr>
<td>2</td>
<td>0.20</td>
</tr>
<tr>
<td>4</td>
<td>0.20</td>
</tr>
<tr>
<td>6</td>
<td>0.20</td>
</tr>
<tr>
<td>8</td>
<td>0.20</td>
</tr>
<tr>
<td>10</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Tattersall et al. have suggested normalising the speech data to factor out the effect of two speech trajectories in the filter bank space which differ only in loudness; it is clear that two utterances identical except for the total energy will lie at different distances from the origin, and dividing by the total energy will reduce them to more comparable trajectories. We therefore tried a very crude frame normalisation which effectively projected each point of the trajectories on the 11-sphere. This has the consequence of factoring out not only the total energy but also the variation in energy throughout the utterance of each word. The results of doing this are shown in Table 13 using the 32 $\times$ 32 grid. This suggests that the radial (energy) component of the trajectories contributes significant information. A normalisation which scales by...
the total energy of an utterance may well improve matters, but using an inner product measure is likely to make matters worse. This accords with our findings using the Kohonen process to recognise phonemes.

3.4 Alternative dimension estimators
The Department of Mathematics at the University of Western Australia has developed a number of algorithms for computing the dimension of strange attractors. These are not entirely appropriate to our application, where the set may not have a dimension. It may be taken that strange attractors can be expected in general to have a fractal dimension, and in cases where the dimension is known the methods work at least as well as the conventional methods and are generally considerably faster. We ran the speech data through three of these algorithms, which may be found elsewhere [7, 8]. They yielded results with a mean of 3.4 for the dimension of the speech space, with an uncertain degree of confidence of approximately 1.5.

3.5 Conclusions
The hypothesis that the speech space is approximately a surface of dimension two is not sustainable. There is some evidence to suppose that it is approximately a noisy four-manifold, but more speakers and a wider range of phonemes might make the dimension higher. On the other hand we may reject the hypothesis that the speech space is nothing more than an unstructured blob in the enclosing space. Since even this represents a very considerable data reduction, there is some interest in trying to find a best fit manifold which approximates the speech space. The Kohonen process is not an entirely satisfactory way of finding a best fit four-manifold; projection pursuit methods have allowed us to establish that the fitting of two- and three-dimensional grids occurs with the grid converging well in the interior of the cluster of points, a matter which will be described elsewhere. The general problem of fitting a smooth manifold to a set of points in a high-dimensional space is difficult when the set is not even approximately affine, as is the case with the speech space. Modifications of the Kohonen process would seem a useful line of investigation.

4 Acknowledgments
We are grateful to James Glover of the Mathematics Department, University of Western Australia, for supplying us with the data for the Lorenz attractor. We also acknowledge gratefully a grant by the Australian Research Council which supported some of this work.

5 References
4 DeHOFF, R.T.: 'Stereological uses of the area tangent count', in 'Geometrical probability and biological structures: Buffon's 200th anniversary' (Springer, 1977)
7 GLOVER, J.N.: 'Detecting folds in chaotic processes by mapping the convex hull'. Proc. Conf. on Dynamics of Complex Interconnected Biological Systems, Albany, Western Australia, 1989
8 JUDD, K.: 'Chaos in complex systems'. Proc. Conf. on Dynamics of complex interconnected biological systems, Albany, Western Australia, 1989