Proceedings of the Institute of Acoustics

# The Fractal Characterisation Of Isolated Human Speech

P.S. McDowell & S. Datta

Department of Electronic and Electrical Engineering
Loughborough University

## 1. Introduction

In 1975 Benoit Mandelbrot [1] introduced his concept of the word '*fractal*' to describe certain natural patterns and objects exhibiting the phenomena known as *self similarity*. Mandelbrot's own loose definition for the term is as follows:

> '*A fractal is a shape made of parts similar to the whole in some way*'

Objects such as mountains, clouds and coastlines are prime examples of natural objects that are self similar, ie., that at all scales of observation the object appears to exhibit the same properties. In order to characterise one fractal object from another, the so called *fractal dimension*, D, of the object can be calculated. The fractal dimension is a real number which falls between the limits of 1 and 3 and can be calculated in a number of ways.

The use of a fractal dimension has been derived from the fact that the conventional topological dimension, 1, 2 or 3, is too general to be useful in many areas of science and nature when trying to distinguish between similar objects that cannot be quantified using Euclidean mathematics. Essentially therefore, the fractal dimension gives us is a measure of the degree of irregularity or roughness for an object.

The fact that such complicated structures such as those described can be characterised by a single number has led to work being carried out in the area of acoustic and speech science. Speech waveforms themselves are highly irregular patterns which can be quantified to some degree using fractal mathematics.

Speech wave forms of the time scale 30 ms to 70 ms in increments of 10 ms have been extensively examined as this time scale covers the duration that phonemes are generally spoken for. The fractal techniques used to quantify the phonemes are the 'Box Counting' method [2], the Minkowski-Bouligand method [3] and the Richardson method [4].

This paper attempts to show that certain classes of phonemes of the English language can be characterised using the fractal dimension techniques indicated and that categories of elements in those classes can themselves be distinguished. It follows an earlier paper, A Fractal Approach to the Characterisation of Speech [5] which attempts to demonstrate principally how the 'Box Counting Method' [2] can be used to classify phonetic elements.

## 2. Phonetic Classification

The sound waves of speech are among the most complicated to be found in nature. Extreme changes in sound quality follow each other with great rapidity, indicating that the speech mechanism viewed as a generator of sound must work in a complicated manner and be capable of operating in a wide variety of ways.

**Vowels**
The articulation of vowels can be described in terms of tongue and lip positions. For the purposes of description, the tongue positions for making vowel sounds are compared with the positions used for making a number of reference or *cardinal vowels* [6]. The cardinal vowels are a set of standard reference vowels whose quality is defined independently of any language. They form a reference point against which the quality of any vowel can be measured. Of course, a strict definition of the term, 'cardinal vowel', is not possible since the quality of such a sound can only be perceived when it is correctly spoken.

**Consonants**
The English consonants can be described by their *place-of-articulation* their *manner-of-articulation* and further still by whether they are voiced or unvoiced as shown in Table 1 [7].

Table 1 Classification of English Consonants by Place and Manner of Articulation

| Place of articulation | Manner of articulation | | | | |
|---|---|---|---|---|---|
| | Plosive | Fricative | Semi-vowel | Liquids | Nasal |
| Labial | p b | | w | | m |
| Labio-Dental | | f v | | | |
| Dental | | 0 th | | | |
| Alveolar | t d | s z | y | l r | n |
| Palatal | | sh zh | | | |
| Velar | k g | | | | ng |
| Glottal | | h | | | |

### 3. Fractal Dimension Calculation

In his classical 1967 paper Mandelbrot [8] posed the question. 'How long is the coast of Britain ?'. One can of course attempt to reach an estimate to this problem in a variety of ways.

1. By using a set of dividers of length $\eta$, walking them along the coastline and counting the number of steps it takes to circumscribe the coast. The approximate length of the coast will be the length of the set dividers multiplied by the number of steps.

2. By attempting to cover the entire coastline with circular disks of radius $\eta$. In other words, considering all the points of both land and sea for which the distance to the coastline is no more than $\eta$. This, in effect, forms a kind of tape of width $2\eta$ which covers the coastline. The approximate length of the coastline can then be calculated by measuring the surface area of the tape and dividing it by $2\eta$.

3. By covering the entire coastline with a grid of cell sizes $\eta$ by $\eta$ and counting the number of cells that the coastline intersects. The approximate length can then be estimated by multiplying the number of cells by $\eta$.

The three methods described above, though distinct, have one very common similarity. Decreasing the size, $\eta$, of the measuring device, be it divider, disc or grid cell will result in a more accurate estimation of the coastline if one were to repeat the experiment. Further, no matter how small the measuring device is made, the estimation in length always increases. More and more detail of the coast could be measured as the device length decreased. This leads to the theoretical conclusion that the length of the coast of Britain or of any coastline for that matter is infinite. In practice the limiting factor is naturally the resolution of what the measuring device can be set to, not to mention the effects of the crashing sea and ebbing tide.

Having discussed the methods above attention can be focused again on Mandelbrot [1] and his derivation of the term *fractal dimension*. Following on from the work carried out by the mathematician, Lewes F. Richardson [4]. Mandelbrot suggested that the relationship between the measuring device length, $\eta$, and the number of steps, $N$, for the device to estimate the length of a coastline could be expressed by the parameter $D$, *the fractal dimension*.

$$N(\eta) = 1 / \eta^D \qquad (1)$$

Multiplying both sides of (1) by the device length, $\eta$, yields the estimation of coastline length, $L(\eta)$.

$$L(\eta) = \eta / \eta^D \qquad (2)$$

Rearranging (2) and taking logs of both sides gives the equation

$$D \log \eta = \log \eta - \log L(\eta) \qquad (3)$$

From (3) the *fractal dimension* is obtained from the slope of coastline length against device length.

$$D = \lim_{\eta \to 0} [1 - \{\log L(\eta) / \log \eta\}] \qquad (4)$$

## The Fractal Characterisation Of Isolated Human Speech

### 4. Experimental procedure

The phonemes tested were extracted from words recorded at 40 KHz, a sampling frequency chosen to preserve the fragmentation of the sampled signal as close as possible to that of the continuous time speech signal. The words used to extract the phonemes are listed in Table 2.

Table 2 - The Phonemes of General English

| | Vowels | | | Consonants | | | | |
|---|---|---|---|---|---|---|---|---|
| *ee* | - | heat | *t* | - | tee | *s* | - | see |
| *i* | - | hit | *p* | - | pea | *sh* | - | shell |
| *e* | - | head | *k* | - | key | *h* | - | he |
| *ae* | - | had | *b* | - | bee | *v* | - | view |
| *ah* | - | father | *d* | - | dawn | *th* | - | then |
| *aw* | - | call | *g* | - | go | *z* | - | zoo |
| *U* | - | put | *m* | - | me | *l* | - | law |
| *oo* | - | cool | *n* | - | no | *zh* | - | garage |
| Λ | - | ton | *ng* | - | sing | *r* | - | red |
| *uh* | - | the | *f* | - | fee | *y* | - | you |
| *er* | - | bird | θ | - | thin | *w* | - | we |
| *oi* | - | toil | | | | | | |
| *au* | - | shout | | | | | | |
| *ei* | - | take | | | | | | |
| *ou* | - | tone | | | | | | |
| *ai* | - | might | | | | | | |

Each of these words were recorded several times over a period of time to establish a database of usable phonemes. To overcome differences in amplitudes of recordings, each of the phonemes were amplitude normalised using standard deviation and scaling factor techniques.

In the testing process, each recording was divided into time slots ranging from the mid 30 ms in steps of 10 ms up to the mid 70 ms in order to give the experiments recording length independence and there by introducing a degree of time normalisation.

### 5. Computations and Results

Tables 3, 4 and 5 show the averaged results of calculations of D, using the three methods described, on numerous phonetic samples extracted from the words tabulated in section 4 and recorded at 40 KHz. The results shown are not cardinal values but the mean result of calculations made on several phonetic examples.

The figures given in Tables 3, 4 and 5 clearly show that, in statistical terms at least, there is good reason to believe that there exists a distinct difference in dimension at all time scales of interest between the phonetic groups shown for each fractal method used. The values themselves are not so significant as the actual difference that separates the phonetic categories at the time scales.

Table 3 The Box Counting method

| | 30 ms | 40 ms | 50 ms | 60 ms | 70 ms |
|---|---|---|---|---|---|
| fricatives | 1.60 | 1.67 | 1.73 | 1.76 | 1.77 |
| plosives | 1.47 | 1.57 | 1.62 | 1.63 | 1.64 |
| vowels | 1.36 | 1.48 | 1.55 | 1.57 | 1.57 |

Table 4 The Richardson method

| | 30 ms | 40 ms | 50 ms | 60 ms | 70 ms |
|---|---|---|---|---|---|
| fricatives | 1.72 | 1.73 | 1.74 | 1.73 | 1.72 |
| plosives | 1.63 | 1.66 | 1.65 | 1.66 | 1.66 |
| vowels | 1.50 | 1.52 | 1.54 | 1.53 | 1.52 |

Table 5 The Minkowski-Bouligand method

| | 30 ms | 40 ms | 50 ms | 60 ms | 70 ms |
|---|---|---|---|---|---|
| fricatives | 1.56 | 1.62 | 1.66 | 1.68 | 1.70 |
| plosives | 1.39 | 1.44 | 1.48 | 1.50 | 1.51 |
| vowels | 1.27 | 1.34 | 1.38 | 1.41 | 1.44 |

Table 6 and 7 details more explicitly the results of experiments carried out on fricative phonemes at 50 ms and plosive phonemes respectively using the three fractal methods.

Table 6 Fricative Phonetic Elements

| | Box Counting | Richardson | Minkowski-Bouligand |
|---|---|---|---|
| unvoiced | 50ms | 50ms | 50ms |
| f | 1.72 | 1.81 | 1.65 |
| θ | 1.78 | 1.80 | 1.71 |
| sh | 1.71 | 1.79 | 1.63 |
| s | 1.78 | 1.68 | 1.75 |
| voiced | | | |
| v | 1.71 | 1.67 | 1.65 |
| h | 1.67 | 1.70 | 1.56 |
| z | 1.76 | 1.69 | 1.71 |
| zh | 1.74 | 1.75 | 1.60 |
| th | 1.73 | 1.71 | 1.66 |

Table 7 Plosive Phonetic Elements

| | Box Counting | Richardson | Minkowski-Bouligand |
|---|---|---|---|
| unvoiced | 50ms | 50ms | 50ms |
| p | 1.62 | 1.70 | 1.53 |
| t | 1.73 | 1.72 | 1.68 |
| k | 1.67 | 1.80 | 1.52 |
| voiced | | | |
| b | 1.67 | 1.45 | 1.46 |
| d | 1.55 | 1.54 | 1.36 |
| g | 1.48 | 1.63 | 1.34 |

Theoretically, the unvoiced fricative phonemes should have a higher fractal dimension than their voiced counterparts because of the absence of a fundamental frequency and harmonics that are associated with voiced phonetic elements. This appears not to be the case because there is no distinct boundary between the two categories. However, the *highest* fractal dimension for all three methods does appear in the unvoiced category and similarly, the *lowest* fractal dimension appears in the voiced category for all three methods.

Between the unvoiced and voiced plosive elements the boundary between the two categories for all three fractal methods appears to be more significant. The *higher* fractal measurements consistently appear in the unvoiced category, /p/, /t/ and /k/ while the *lower* measurements appear in the voiced category, /b/, /d/ and /g/.

The remaining phonetic groups, nasal, liquids and semi vowels cannot be distinguished from the three main categories. However, because these groups are much smaller, the elements within them are easier to distinguish. For example in the nasal group, /n/ is consistently higher than /ng/ which itself is consistently higher than /m/. The semi- vowel /y/ is higher than /w/ at all time scales as is the liquid /l/ consistently higher than /r/.

## 6. Conclusions

The usefulness of fractal geometric mathematics as a model for characterising speech is inherently limited by the accuracy to which a fractal dimension can safely be calculated and further by the limited scope over which such dimensions (for the case of graphical waveforms) can exist. However, examining the achievements thus far with three simple algorithms there is some basis to suggest that fractal mathematics could become an important new linguistic tool.

Future developments in this area will depend very much on whether the fractal dimensions of all the phonetic elements can be proven to be consistent and more significantly, speaker independent. If this proves to be so then there is no question a new use for Mandelbrot's ingenious mathematics will have been found.

## 7. References

[1] Mandelbrot, B.B. 1977 'Fractals Form, Chance and Dimension'

[2] Feder J. 1988 'Fractals' Penum Press, New York and London

[3] Le Mehaute A. 1990 'Fractal Geometrics' Penton Press, London

[4] Richardson, L.F. 1961 'The Problem of contiguity: an appendix of statistics of deadly quarrels.' General Systems Yearbook 6, 139-187

[5] McDowell & Datta 1993 'A Fractal Approach to the Characterisation of Speech' Acoustics Letters Vol .17 No.1

[6] Denes & Pinson 1963 'The Speech Chain'

[7] Fletcher H. 1953 'Speech and Hearing in Communication' Van Nostrand. London New York Toronto

[8] Mandelbrot B.B 1967 'How Long is the coast of Britain ? Satatistical Similarity and Fractional Dimesion.' *Science* 155, 636-638