Information Dimensions of Speech Phonemes

Chang-Young Lee*

ABSTRACT

As an application of dimensional analysis in the theory of chaos and fractals, we studied and estimated the information dimension for various phonemes. By constructing phase-space vectors from the time-series speech signals, we calculated the natural measure and the Shannon's information from the trajectories. The information dimension was finally obtained as the slope of the plot of the information versus space division order. The information dimension showed that it is so sensitive to the waveform and time delay. By averaging over frames for various phonemes, we found the information dimension ranges from 1.2 to 1.4.

Keywords: chaos, strange attractor, natural measure, information dimension, phoneme classification

I. Introduction

For several decades, the science of chaos [1] has been of much interest in numerous fields of academic concern and, for this reason, people gave the subject the title "one of the greatest discoveries in this century". Even though such enormous developments and discoveries have been achieved and found, the subject of chaos is said to be unsettled yet. One of the reasons is that the subject matters are too vast for systematic integration.

As geometrical objects which are related to chaos, the study of fractals reveals the hidden structures inherent in nature. An important subject in conjunction with fractals is the dimensional analysis, since it measures the complexity of the geometrical objects or the chaotic dynamical system concisely.

As an application of the theory of chaos and fractals developed so far, we try to calculate dimensions of various speech phonemes. Though the algorithm and procedure for calculation of the dimension is explained in the next section of this paper, the meaning of the dimension for the time series data of speech signals can not be expounded clearly for now. We might just say that the extracted parameters are somehow related to the mechanical structures and dynamical movements of human vocal tracts considered as a dynamical system.

We know that the random generator enters in the speech production of consonants and reflects somehow the dynamics of speech production [2]. When we consider the vocal tract as a dynamic system, an important parameter characterizing the system is the Ljapunov

* Department of Computer Engineering, College of Information System, Dongseo University

exponent [3]. Its concept has been generalized so that it applies to many interesting dynamic systems in mathematics and sciences. It has become one of the keys to measuring, evaluating and detecting chaotic behavior. However, the calculation procedure is not directly applicable to time series data such as speech signals. Therefore, we turn our calculation to the so-called information dimension which serves as a lower bound for the fractal dimension that is usually computed by the box-counting procedure.

II. Theory

Given a time-series data

$$x(0),\ x(1),\ x(2),\ \cdots\cdots \tag{1}$$

we construct a phase-space vector

$$(x(0),\ x(T)),\ (x(\tau),\ x(\tau+T)),\ (x(2\tau),\ x(2\tau+T))\ \cdots\cdots \tag{2}$$

Though it is generally formed in a larger-dimensional space, we work with the simple two-dimensional space for ease of graphics and calculation. In other terms, the above vector may be interpreted as a projection of a larger-dimensional one into two-dimensional subspace spanned by $x-$ and $y-$axes.

The parameter τ enters in conjunction with the frequency of A/D conversion and is thus limited to the values less than 5. On the other hand, the parameter T is expected to have an important effect on the results. If T is too small, then the $x-$ and $y-$components of the vectors have nearly equal values and thus their trajectories will lie around the line $y = x$. If T is too large, on the other hand, then there's no correlation between the components and we should fail in the search of any structures hidden in the given time series.

We divide the space into arrays of squares of size $s = 2^k$ and label each B_{ij}. Then we define the natural measure by

$$\mu_{ij} = \frac{1}{n}\sum_{k=0}^{n-1}\sigma_{ij}(\vec{v_k}) \tag{3}$$

which is to be calculated from the trajectories of the vectors $\vec{v_k}$ as given by Eq. (2). In (3), the following notation has been used:

$$\sigma_{ij}(\vec{v_k}) \equiv \begin{cases} 1 & \text{if the vector } \vec{v_k} \text{ is included in } B_{ij} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

In definition (3), the number of points n should in principle be infinite in order that the measured quantities be ergodic. However it is limited for practical reasons since we are

dealing with speech of finite time duration. In our study, we will use the fixed number $n = 256$ in order to be in parallel with the procedure of feature extraction of speech signals frame by frame in speech recognition technology. We will find out that the measured quantities are very sensitive to the speech waveforms and thus n should be limited roughly to the values of a frame length in order for the analysis to be meaningful on frame-basis. If n is so large, then the dimensional analysis is blurred out by averaging over frames.

The next step is to calculate the Shannon's Information

$$I(s) = -\sum_{ij} \mu_{ij} \log_2(\mu_{ij}) \tag{5}$$

for various values of the square size s and plot the results according to the prescription

$$I(s) = I_0 + D_I \log_2\left(\frac{1}{s}\right) \tag{6}$$

and get the slope by linear regression.

The slope D_I characterizes the amount of additional information in increasing the resolution of the phase-space by decreasing the mesh size s. In order to scrutinize and get more information, we should look into the space more closely. For this reason, the quantity D_I has been called "Information Dimension". It provides the lower bound for the fractal dimension of the given geometry as obtained commonly by the box-counting procedure [4].

We might hope that D_I could serve as a tool for phoneme classification like formant frequencies.

III. Experimental Results and Discussion

In order to avoid difficulties caused by speaker-dependence, we used speeches of only a single male speaker. For each phoneme, three words were pronounced and sampled at 10 kHz with 12 bit quantization after low-pass filtering. We performed our calculation frame by frame, each on consisting of 256 data points corresponding to 25.6 ms time duration. Figure 1 shows one of the speech phoneme /a/. Phase-space vectors were constructed according to Eq. (2) with $\tau = 1$. Figure 2 shows the trajectory of the phase-space vectors constructed from the time-series data of Fig. 1 with time delay $T = 50$.

Figure 1. The waveform of a phoneme /a/ which was pronounced by a male speaker, sampled at 10 kHz, and quantized by a 12 bit A/D converter. This is a small part of a speech for a word.

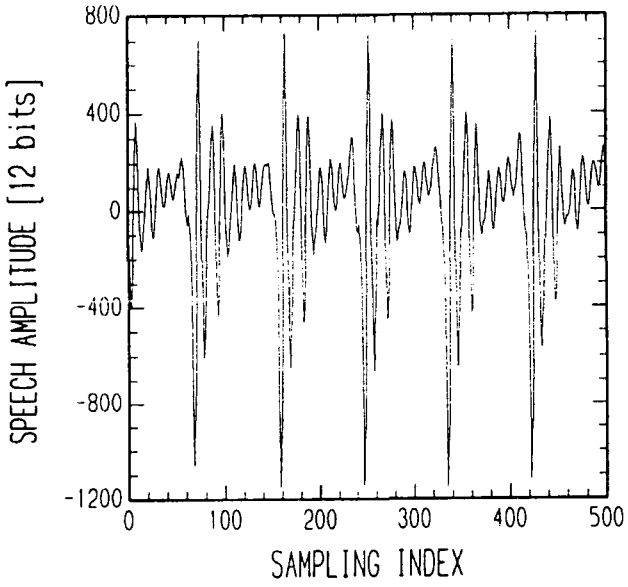
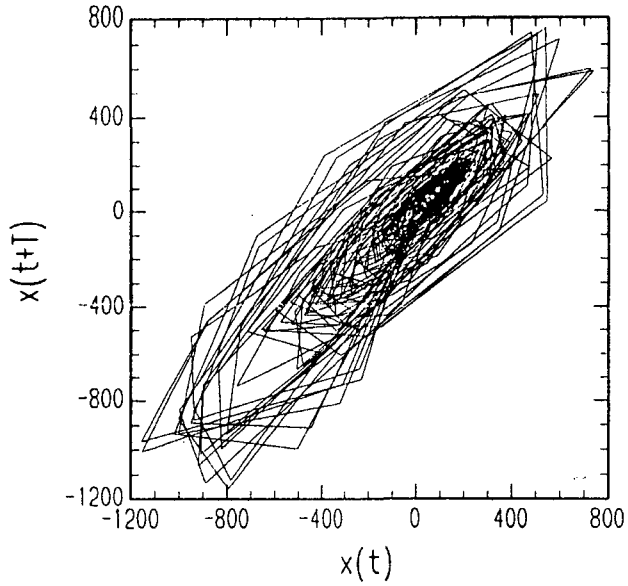


Figure 2. Phase-space construction of vectors and their trajectory in two-dimensional space with time delay 50. By dividing this box into square arrays, we obtain natural measure and the Shannon's information.

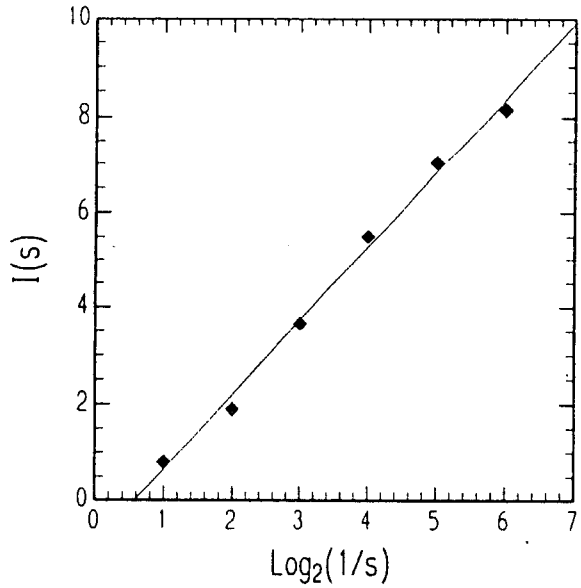


We consider the smallest rectangle which contains all of 256 vectors and divide it by a $(2^k \times 2^k)$ array. After calculation of the natural measure and the Shannon's information by

the procedures of Eqs. (3) and (5), we increase k by 1 and repeat the calculation. Figure 3 shows an example data and plot of the result, i.e., the Shannon's information versus the space division order $k = \log_2(1/s)$. The line represents the linear regression result and its slope gives the information dimension D_I.

Since we have only a rather small number (256) of vectors in a trajectory, the information contained in the space will be exhausted soon as we increase k and the information $I(s)$ is changed little. Actually, the data plot such as Fig. 3 always flattens off for large values of s even if the number of vectors goes to infinity. This effect can also be seen in Fig. 3, where there is a tint of sigmoid-like curvature. Though more elaborate methods may be devised and applied [5], we content ourselves with the result obtained from six divisions, i.e., $k = 1 \sim 6$. The results ranges from 1.2 to 1.4 for most phonemes and frames and the values are close to some famous two-dimensional strange attractors [6].

Figure 3. An example of the Shannon's information versus the space division order. The line is the linear regression result and its slope gives the information dimension.



The dependence of D_I on the time-delay T is drawn in Figure 4. By a comparison with Fig. 1, we can see that its dependence reveals the periodic nature of the speech signals as is expected. It is to be noted that D_I responds to the speech waveform very sensitively. By a shift of 1, i.e., $\Delta T = 1$, the result can differ as much as $\Delta D_I = 0.05$, which is the difference between the maximum and minimum in a frame. This sensitivity pervades all through the analysis in our study.

Figure 4. Information dimension versus time delay performed on the waveform as given in Figure 1.

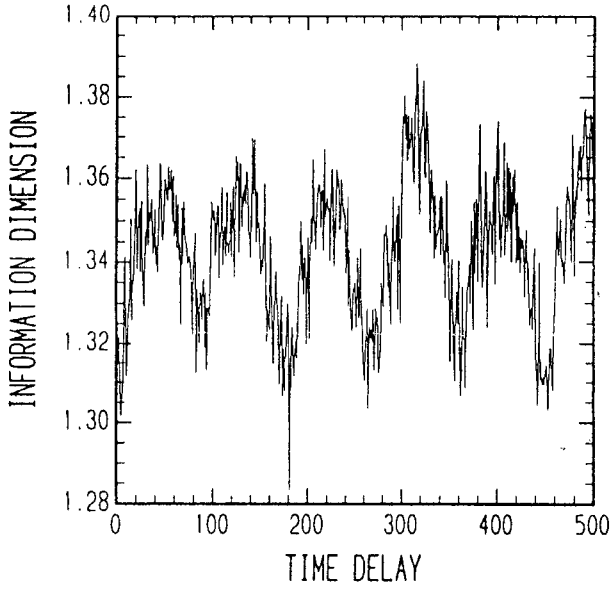


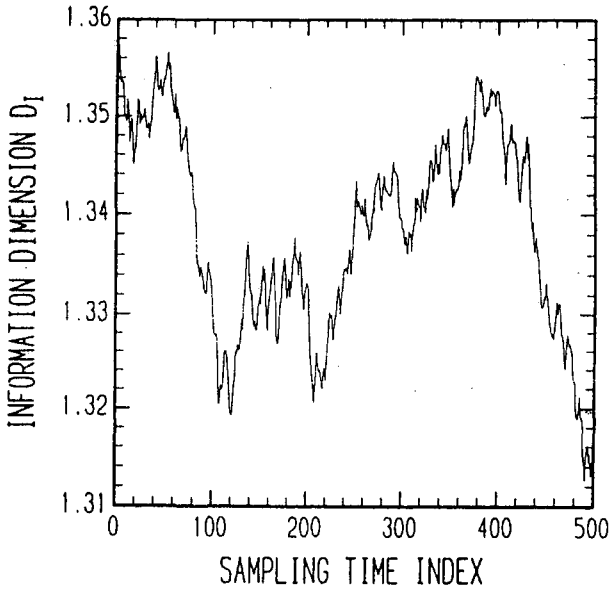
Figure 5 shows the information dimension calculated frame by frame. After 256 vectors were constructed and analyzed by Eq. (2) with $T = 50$, next frame is shifted by one sampling index, $\Delta t = 1$. Therefore, this figure shows the dimension as a function of time $D_I(t)$ which permits us to compare directly with the waveform as given in Fig. 1. We see that the result is so sensitive to the waveform that even the periodic nature could hardly be detected. This may be explained by the combined role of the time delay, waveform, and sensitivity of D_I.

As the final work of our study, we estimated D_I for various phonemes. The calculation was performed through the phoneme frame by frame, each frame shifted by 5 data points, and averages were taken from them. As the result in Table 1 shows, the values range from 1.2 to 1.35 for most phonemes.

Table 1. Information dimension for various phonemes

/g/	/n/	/d/	/m/	/b/	/s/	/J/	/C/	/k/	/t/	/p/	/h/	/a/	/x/	/o/	/u/	/I/
1.30	1.29	1.26	1.26	1.28	1.26	1.24	1.26	1.26	1.26	1.24	1.24	1.33	1.26	1.23	1.23	1.21

Figure 5. The information dimension as a function of time for the waveform as given in Figure 1. It shows how the result is sensitive to the combined role of the time delay and waveform.



IV. Summary and Conclusion

As an application of dimensional analysis in the theory of chaos and fractals, we studied and estimated the information dimension for various phonemes. By constructing phase-space vectors from the time-series speech signals, we calculated the natural measure and the Shannon's information from the trajectories. The information dimension was finally obtained as the slope of the plot of the information versus space division order.

The result showed that time delay is very effective and intricately related to speech signals. The analysis of the information dimension showed that it is very sensitive to waveform and time delay. By averaging over frames for various phonemes, we found the information dimension ranges from 1.2 to 1.4, which is reminiscent of some strange attractors.

REFERENCES

[1] For an extensive material and bibliography. 1992. see H. Peitgen, H. Jürgens, and D. Saupe, Chaos and Fractals, Springer-Verlag, New York.

[2] J Deller, J.G. Proakis, and J. Hansen. 1993. Discrete-Time Processing of Speech Signals, Macmillan, New York, 192-197.

[3] Ref. [1], pp. 509-519.

[4] Ref. [1], pp. 734-735.

[5] P. Grassberger, On the Fractal Dimension of the Henon Attractor. 1983. Phys. Lett. 97A 224-226.

[6] Ref. [1], pp. 659-708.

접수일자 : '98. 2. 21.
게재결정 : '98. 3. 22.

▲ Department of Computer Engineering, College of Information System
Dongseo University, Pusan 617-716 (617-716)
Tel: (051) 320-1572 (O), (051) 324-1325 (H)
Fax: (051) 312-2389
e-mail: seewhy@kowon.dongseo.ac.kr