

INVESTIGATING THE POSSIBILITY OF IMPROVING RECOGNITION ROBUSTNESS WITH APPLIED CHAOS THEORY

E.K. Patterson and J.N. Gowdy

Department of Electrical and Computer Engineering
Clemson University
Clemson, SC 29634, USA
epatter@eng.clemson.edu, jgowdy@eng.clemson.edu

ABSTRACT

One of the primary difficulties involved with speech recognition is dealing with sounds other than speech that may occur while a person is speaking. There have been many techniques that attempt to deal with background “noise.” Many of these make use of some type of stochastic model to decrease the effect of the interfering signal. This paper, however, investigates the possibility of a new technique involving nonlinear chaos theory.

In the past few decades, the idea of chaos has been of growing interest in physics and has recently begun to enter the world of signal processing with some good results. In a simple sense, a chaotic signal is one that appears random but has some underlying, nearly deterministic structure. This paper considers using some of the knowledge from chaos theory in modeling background “noise” that may interfere with a recognizer or other speech system.

1. INTRODUCTION

Techniques that attempt to improve robustness in speech recognizers tend to take some stochastic approach. The “noise” that is interfering may not be random noise in the typical probabilistic, communications-system sense, though. It is usually some sound generated by the surrounding environment of the speaker, but is often treated as random and modeled as a signal that may be minimized in some probabilistic sense. Other techniques may treat the signal as a sum of periodic components and try to minimize the signal in some spectral sense. Both of these approaches achieve some good results, but perhaps more information can be obtained if the interfering signal is considered in a different manner.

Usually, data that is modeled as a random process has some underlying dynamic system generating it. It is convenient to model the system as random because the underlying dynamics are not known sufficiently well. Linear techniques or stochastic models are used because they are well known, even if they are not entirely accurate. The chaotic theory of nonlinear dynamical systems, though, assumes that even a relatively low-order system may create data that is irregular and seemingly random [1]. If the underlying system can be determined in some sense, then perhaps the data can be better modeled or predicted. This would aid in either removing the “noise” from a recognizer’s input or else adapting the recognizer.

In order to model a signal as chaotic, it is necessary to determine if the signal has chaotic properties. A chaotic signal is nonlinear, has a highly irregular waveform (often continuous and broadband in the frequency domain [2]), is extremely sensitive to

initial conditions, and can still be a relatively low-order system [3]. This paper looks at the possibility of modeling speech-interfering “noise” as a chaotic system. A very brief and cursory introduction to chaos and modeling a signal as such will be given in Section 2. Section 3 of this paper will concern investigating “noise” samples from the NOISEX database. The fourth section will discuss some possibilities of applying the results to improve speech recognition.

2. MODELING “NOISE” AS CHAOS

A chaotic system is a form of a nonlinear dynamical system that can be defined by the differential equation

$$\frac{dx}{dt} = F(x) \quad (1)$$

or by the discrete-time equation

$$x(t+1) = f[x(t)] \quad (2)$$

where x is a n -dimensional vector and F or f is a differentiable function [3, 4]. The *phase space* is the n -dimensional space within which the vector x moves during its time evolution based on Equation 1 or 2. Chaotic systems are extremely responsive to initial conditions; two systems started with only a minute difference will quickly separate by a large distance, creating the seemingly random time series and making the system difficult to predict. If the same system were started twice, however, with the *exact* same initial conditions then it would demonstrate deterministic behavior. This is what allows a chaotic system to be predicted for a short time. After the system has reached a steady-state boundedness, the portrait of the trajectories of x becomes what is known as the system *attractor*. A familiar example for purpose of demonstration is the Lorenz attractor in R^3 (with which Lorenz hoped to predict weather patterns) as shown in Figure 1 [3].

Based only on the incoming time series, establishing information about such a system would seem to be difficult, but Takens demonstrated that information about the underlying dynamical system may be found from embedding a single observable of the system, such as a time-series amplitude [5]. Embedding involves constructing a set of variables based on a delay time T , chosen from the data to help extract some of the “physical” sense of the data:

$$x_0(t), x_0(t+T), \dots, x_0(t+(n-1)T). \quad (3)$$

The delay T , known as the *embedding delay*, should be small enough so that $x(n)$ and $x(n-T)$ are still correlated to some

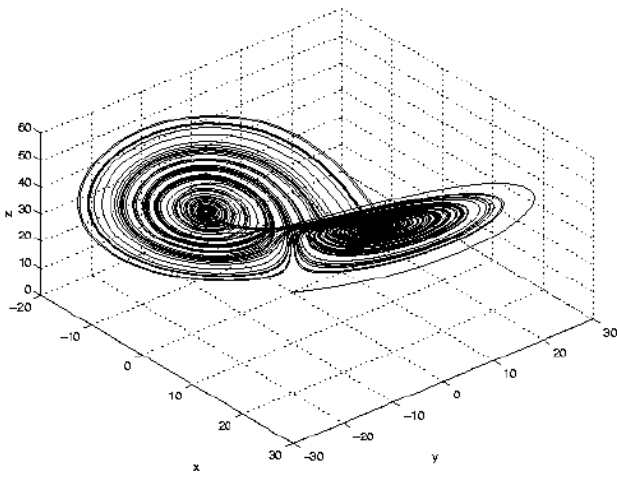


Figure 1: Lorenz attractor.

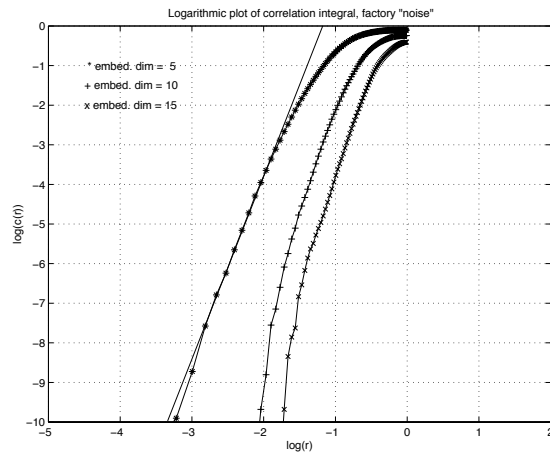


Figure 2: Calculating C_D for factory “noise.”

degree but large enough that they are sufficiently independent to serve as coordinates of a reconstruction space; it can be chosen based on some relation between $x(n)$ and $x(n - T)$ such as the first zero-crossing of the autocorrelation or the first minimum in the average mutual information [2]. The embedding theorem allows any choice of T if the data set is infinitely long. These techniques, however, help compensate for a finite time series [6].

The embedding dimension, D_E , must also be chosen when dealing with experimental data. This determines n in Equation 3. A sufficient but not necessary condition for choosing this is $D_E \geq 2d + 1$, where d is the dimension of the phase space of the system. D_E may be less than this, however, and still be sufficient for proper embedding. A technique for estimating the embedding dimension is the method of false nearest neighbors [6]. Dynamic reconstruction of the system may take place with any sufficiently large dimension D , but D_E is the minimum integer dimension that achieves reconstruction [7]. An attempt for reconstruction under smaller dimensions will not be accurate and under larger dimensions may be subject to corruption.

Once T and D_E are chosen, an attempt can be made to find the attractor dimension and determine if the data exhibits chaotic characteristics. The attractor dimension gives a lower bound on the order necessary to model the nonlinear dynamics of the system [8]. This is the amount of information necessary to specify points on the system attractor, which represents the equilibrium state of the dynamical system [7]. There are several forms of estimating this, but they all represent the *fractal dimension*. (If this is a non-integer, the attractor is called a *strange attractor*.) A popular method, although some prefer other methods [9], is the Grassberger-Procaccia sample correlation integral, which gives a method for finding D_C , the *correlation dimension*. A finite correlation function $C(r)$ based on the correlation integral can be used:

$$C(r) = \frac{1}{N_2} \sum_{\substack{i,j=1 \\ i \neq j}}^N \theta(r - |x_i - x_j|). \quad (4)$$

In Equation 4, θ is the Heaviside function for which $\theta(x) = 0$ if $x < 0$ or $\theta(x) = 1$ if $x > 0$. The correlation dimension is now determined by plotting $\log(C(r))$ versus $\log(r)$ and taking the slope

of the region which appears as a straight line, as shown in Figure 2. If the embedding dimension D_E is sufficiently chosen, the slope should be equivalent to D_C , the correlation dimension. If D_E is increased by increasing n in Equation 4, the slope should remain the same, despite the larger dimension. For normally distributed random data, however, the slope will continue to increase to match D_E . Also, for uncorrelated noise, the slope will be equal to one. A stable $D_C \geq 1$ is a possible indicator of a chaotic system, but some systems with colored noise have also exhibited fractional dimensions greater than one [2].

Finally, an important indicator of chaos is the Lyapunov spectrum. The Lyapunov spectrum is related to the fractional dimension of a strange attractor, and thus related to the correlation dimension. Lyapunov dimension may be calculated for a known system of equations, but typically other techniques such as the correlation dimension are easier for estimating attractor dimension from empirical data. Lyapunov exponents are characteristic exponents $\lambda_1, \lambda_2, \dots, \lambda_{D_L}$, ordered from largest to smallest. For a system to be chaotic, the largest exponent λ_1 has to be positive, and also $\lambda_1 + \lambda_2 + \dots + \lambda_{D_L} \leq 0$. As an example, the Lorenz attractor in Figure 1 has $\lambda_1 = 0.33$ and $D_L = 2.07$ [10]. It would need at least the next integer number of variables, 3, to reconstruct the dynamics of a series generated from the attractor. λ_1 is also an indicator of how predictable a system is [11].

3. CHARACTERIZING NOISE SAMPLES

Sounds that are considered “noise” in the speech-recognition sense are usually those that comprise the background of the environment where the speech is recorded. This is often modeled as white, Gaussian noise even though it is rarely close to that, except for possibly a small part introduced by the equipment; this makes the model and its evaluation simpler, though. (This discussion merely includes additive noise, disregarding effects from convolution in the channel.) In reality, the sounds are usually created by some physical system nearby such as an air conditioner, engine, or other person. The “noise” may be very simple and periodic in some cases but often tends to cover a broader, continuous spectrum, perhaps including a combination of interacting sounds. It is possible

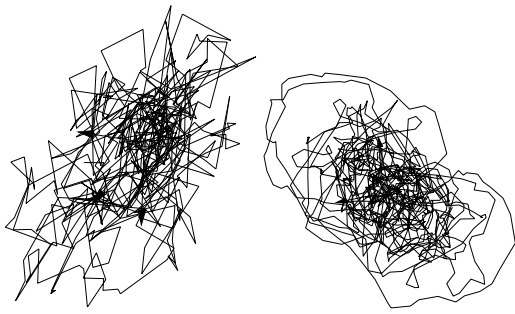


Figure 3: Attractors for factory and speech “noise.”

that the physical phenomena behind the sounds may lend itself to nonlinear dynamical system modeling. This could give insight into new techniques to model, predict, or adapt to the “noise.” We will now consider three sounds from the NOISEX database.

The NOISEX database includes samples of several background sounds that may occur in a recognition environment, such as vehicular noise. It is a common procedure to study a recognizer in terms of how well it deals with these intruders. Instead, though, we will consider the sounds in terms of dynamic modeling. The generating system will have a state space spanned by n system variables and some logical method of creating the sounds. This is likely true whether or not we have the mathematical ability to deal with the system without resorting to a stochastic model and whether or not the system is nonlinear or chaotic. We consider three noises from the NOISEX database as examples: car, factory, and speech (from a large crowd). All of these appear to be relatively broadband (in terms of the spectrum observed for speech recognition), although the NOISEX car sound does have augmented, lower-frequency components. (This is probably typical of quieter cars as the higher frequencies are easier to dampen. Also, this sample does not seem to include significant wind noise that may be incurred with the windows lowered). The factory recording is generated by all the systems interacting at the factory and may have a number of variables involved in creating the sound. The speech is the sum of a large number of voices resulting from conversations that are under psychological and social influences. Modeling these sounds or others as a nonlinear, dynamical system may achieve some desirable results in dealing with interfering sound. These techniques have obtained good results at least in radar signal processing where they have been used to predict and help negate the effect of sea-clutter on radar returns used for signal detection [6, 7, 11, 4].

The process of attempting to find the correlation dimension, C_D , for each of the three sounds was performed. A time series was taken from the amplitude of each of the NOISEX sound files used. The value of T was chosen based on the autocorrelation, as previously mentioned. Several values of T and also of D_E were considered. The method described in Section 2 for calculating D_C was used. Figure 2 shows the use of this method for the factory series. It includes a line fitted roughly to the slope that indicates D_C . Each of the three curves is from a different embedding dimension, D_E , and as also mentioned in Section 2, the slope tends to saturate even though D_E is increased. The approximate D_C for the factory series was determined to be around 4.6, indicating that a system would need at least 5 variables to describe the underlying

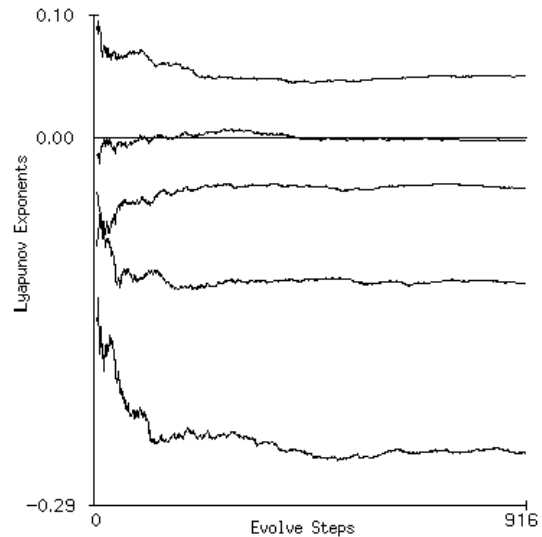


Figure 4: Factory “noise” Lyapunov exponents.

dynamics of generating the factory time series. This procedure was also repeated for the car and speech series. D_C for the car sound was found to be about 1.5 (We will see later, though, that this is likely the case where colored sound causes this value, as mentioned earlier.) Finally, the speech was found to have D_C of about 3.4. This seems consistent with some findings by Banbrook who has studied modeling phonemes as chaotic, finding them to be of dimensionality of 3 to 4 but not chaotic [12]. (It is possible that a whole collection of voices, though, may still be modeled as a chaotic system.)

Next, we looked at the attractors and Lyapunov exponents, using a package by Banbrook [13]. Figure 3 shows the factory and speech attractors, respectively. The Lyapunov exponents for these series were interesting. For the factory time series, λ_1 was positive (≈ 0.05), and $\sum_{i=1}^5 \lambda_i < 0$, indicating a chaotic series. See Figure 4. λ_1 for the car series was positive, but the sum of the exponents was positive, indicating meaningless results [2]. This is not entirely unexpected due to the lower-centralized frequencies of the car sound. For the speech series, though, λ_1 was positive (≈ 0.06), and the sum was less than zero, indicating a the presence of chaos. See Figure 5. We are still investigating some of the other noises included in the NOISEX database. The above results, though, suggest that some speech-interfering noise may be modeled from a chaotic, nonlinear dynamical standpoint.

4. IMPROVING RECOGNITION ROBUSTNESS

These results have indicated the possibility of modeling some of the sounds that interfere with speech recognition as chaotic. (There is also the possibility that a combination of periodic and chaotic signals can be used to represent the interfering sounds, especially with some of the other NOISEX samples.) If the system in question may be modeled as a chaotic dynamical system, what techniques might be available to aid in recognition? D_C and λ_1 provide some idea of the order and predictability of the dynam-

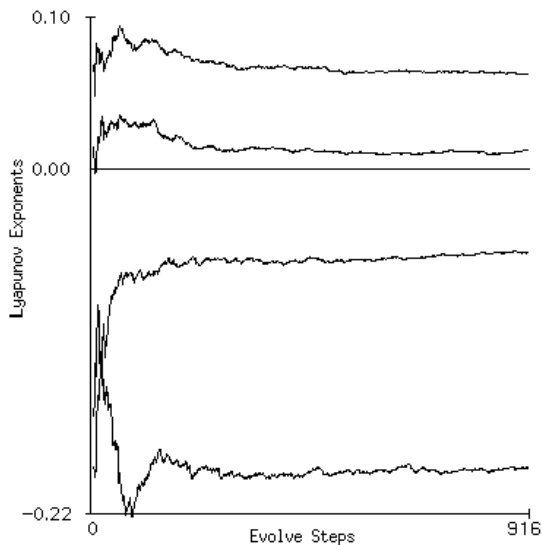


Figure 5: Speech “noise” Lyapunov exponents.

ical systems that create the “noise.” This information should be used to help minimize the effect of the “noise” by predicting and attempting to eliminate the background sound or by attempting to adapt to the background sound. Because of the nonlinear nature of a chaotic system, a neural network makes an excellent predictor because of its property as a nonlinear, universal approximator [14]. The information learned from the chaotic modeling also provides insight into what architecture to use. For instance, with a multilayer perceptron a lower bound is placed on N , the size of the input layer:

$$N \geq TD_E. \quad (5)$$

Other additive noises may corrupt the network, also, if the input layer is much larger than N [11]. D_E here should be chosen to be an integer larger than what D_C is ever expected to be, and T is chosen as mentioned in Section 2. Thus, if we were to construct a neural network to learn and predict the factory series, we could choose $N = 20 * 7 = 140$. It has been shown with chaotic radar-clutter return that a network constructed as such is successful in short-time prediction and partial removal of the chaotic system influences [6, 7, 11]. It is possible that this may also work well for some classes of speech-interfering signals such as the factory and speech samples studied from NOISEX. Some other nonlinear prediction methods have also been attempted with other data [15]. Other areas to investigate include determining which classes of background noise may be modeled as chaotic, what possible models may be constructed for combinations of periodic and chaotic signals, and techniques for model adaptation based on the embedding and dimensional information.

5. CONCLUSION

In this paper we have presented the possibility of modeling some speech-interfering sounds as chaotic, nonlinear dynamical systems. A brief overview of modeling time series by chaos theory has been given to attempt to provide some explanation and motiva-

tion. Background noises from the NOISEX database were used to study the feasibility of chaos-based modeling. Some of the time-series amplitudes of these sounds have evidenced chaotic properties. C_D was found to be between 3-6 for the factory and speech time series, and both produced positive Lyapunov exponents. Future work involves further determining which types of background noises that may be encountered in speech recognition are possibly chaotic. Also, evaluating the multilayer perceptron model mentioned here, as well as other possible prediction, elimination, and adaptation schemes are important steps. More research may reveal that chaotic, dynamical system modeling can be a valid and useful tool for improving noise robustness in speech systems.

6. REFERENCES

- [1] H. Schuster, *Deterministic Chaos: An Introduction*. Weinheim, Germany: VCH, 1989.
- [2] H. Abarbanel, T. Frison, and L. Tsimring, “Obtaining Order in a World of Chaos: Time-Domain Analysis of Nonlinear and Chaotic Signals,” *IEEE Signal Processing Magazine*, vol. 15, pp. 49–65, May 1998.
- [3] D. Ruelle, *Chaotic Evolution and Strange Attractors: The Statistical Analysis of Time Series for Deterministic Nonlinear Systems*. New York: Cambridge University Press, 1989.
- [4] H. Leung and S. Haykin, “Is There a Radar Clutter Attractor?,” *Applied Physics Letters*, vol. 56, pp. 592–595, 1990.
- [5] F. Takens, *Dynamical Systems and Turbulence*, pp. 366–381. Berlin: Springer, 1981. edited by D. Rank and L.S. Young.
- [6] S. Haykin and J. Principe, “Making Sense of a Complex World: Using Neural Networks to Dynamically Model Chaotic Events such as Sea Clutter,” *IEEE Signal Processing Magazine*, vol. 15, pp. 66–81, May 1998.
- [7] S. Haykin, “Neural Networks Expand Signal Processing’s Horizons,” *IEEE Signal Processing Magazine*, vol. 13, pp. 29–33, March 1996.
- [8] J. Farmer, E. Ott, and J. Yorke, “The Dimension of Chaotic Attractors,” *Physica D*, vol. 7, pp. 153–180, 1985.
- [9] H. Tong, “A Personal Overview of Nonlinear Time Series Analysis from a Chaos Perspective,” in *15th Nordic Conference on Mathematical Statistics*, (Lund, Sweden), August 1994.
- [10] A. Wolf, J. Swift, H. Swinney, and J. Vastano, “Determining Lyapunov Exponents from a Time Series,” *Physica D*, vol. 16, pp. 285–317, 1985.
- [11] S. Haykin and X. Li, “Detection of Signals in Chaos,” in *Proceedings of the IEEE*, vol. 83, pp. 95–121, January 1995.
- [12] M. Banbrook, “Nonlinear Dynamics from Time Series.” <http://www.ee.ed.ac.uk/~mb/research.html>.
- [13] M. Banbrook, “Chaos Analysis Software (for time series data).” http://www.ee.ed.ac.uk/~mb/analysis_progs.html.
- [14] K. Hornik, M. Stinchcombe, and H. White, “Multilayer Feedforward Networks Are Universal Approximators,” *Neural Networks*, vol. 2, pp. 359–366, 1989.
- [15] M. Casdagli, “Nonlinear Prediction of Chaotic Time Series,” *Physica D*, vol. 35, pp. 335–356, 1989.