3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

STRANGE ATTRACTORS AND CHAOTIC DYNAMICS IN THE PRODUCTION OF VOICED AND VOICELESS FRICATIVES

Shrikanth S. Narayanan and Abeer A. Alwan

Department of Electrical Engineering, UCLA
405 Hilgard Avenue
Los Angeles, CA 90024, USA

ABSTRACT

In this study, analysis techniques based on chaos theory were used to analyze far-field acoustic pressure waveforms of the fricatives /s/, /ʃ/, /z/, and /ʒ/. Results indicate the presence of low-dimensional chaos in the fricative time-series data. In addition, it was found that phase trajectories of the steady-state parts of the voiced fricatives were distinctly different from those of the voiceless ones; this observation suggests the potential use of phase trajectories as an effective voiced/voiceless classification technique for fricatives.

Keywords: Chaos theory, fricatives, fractal dimensions.

1. INTRODUCTION

The use of deterministic approaches in analyzing and modeling turbulent flows has received wide attention in fluid dynamics and physics. In the heart of non-linear deterministic approaches lies the notion of chaos and bifurcation theory. These approaches have been used in analyzing classic turbulence experiments on Rayleigh-Bénard convection, Taylor–Couette flow [1] and Acoustic Cavitation [2]. The first attempt at fractal characterization of speech waveforms using results from chaos theory was reported by Pickover and Khorasani [3]. Techniques based on chaos theory have also been used to characterize irregularities in the periodicity and amplitude of vocal fold vibration [4], and to analyze newborn infant cries [5]. Geometrical nonlinear techniques, such as phase portraits, have been used to analyze articulatory data describing lip movement [6]. In this paper, acoustic turbulence in fricatives is investigated using results of chaos theory.

The production of fricatives involves the formation of a narrow supraglottal constriction and the generation of turbulence in the vicinity of that constriction when air flows through the vocal tract. In addition to turbulence, the vocal folds may vibrate as in the case of voiced fricatives. Theoretical concepts and analysis techniques will be presented in the first part of the paper, followed by a presentation and discussion of experimental results.

2. THEORY: ANALYSIS TECHNIQUES

Linear signal analysis techniques, such as fourier transforms and autocorrelation functions, may not be sufficient to characterize a chaotic time series. Geometrical techniques, such as phase-portrait constructions, followed by a careful evaluation of invariant measures, such as the attractor dimensions and lyapunov spectra, are required to analyze a chaotic time series.

2.1. RECONSTRUCTION OF THE PHASE SPACE

The experimental data we consider are typically measurements of a single scalar observable $\{p(t_k)\}$, not necessarily corresponding to a state variable, at a fixed spatial point. Fourier analysis of signals generated from nonlinear sources, such as turbulence data, does not result in useful information since the processes that give rise to chaotic behavior are fundamentally multivariate and may possess fractal dimensions. The reconstruction of the system phase space (state space) from such scalar measurements, for an effective time-asymptotic behavior description of the observed dynamics, is possible by using results of mathematical topology theory [7, 8, 9]. This approach views turbulence as exhibiting deterministic behavior described in terms of attracting sets in a phase space (resulting in *strange* or *chaotic* attractors). A multidimensional phase portrait can be constructed from the measurement of a single observable by using time delays. A point $\mathbf{X}(t_k)$ in such a d-dimensional phase space is given by

$$\mathbf{X}(t_k) = \{p(t_k), p(t_k + \tau), \cdots, p(t_k + (d-1)\tau\}. \quad (1)$$

The arguments in [7,9] suggest that the choice for the time delay τ is essentially arbitrary and that a sufficient condition for the choice of d, an integer, depends on the attractor dimension d_A, which can be fractional. These arguments suggest that if $d > 2d_A$ then the attractor as seen in the space with the lagged coordinates will be smoothly related to the attractor as viewed in the original physical coordinates. The procedure of choosing sufficiently large d is formally known as embedding and any dimension that reveals the attractor structure is called an embedding dimension (d_e). Once a large enough $d = d_e$ has been achieved, any $d \geq d_e$ will also provide an embedding. The time of the first zero crossing is typically calculated from the average mutual information of the time delay τ. Time-delay embedding is the most commonly used technique for mapping scalar data into the multidimensional phase space. The broad-band 'noise' generated during the production of fricatives may arise from stochastic or deterministic forces. To decide whether the dynamics of the system is generated by 'classical noise' or by deterministic chaos we construct and analyze phase–space portraits and evaluate lyapunov spectra. If the system were truly stochastic, assuming the time-domain trajectory of m variables, the corresponding m-dimensional state space will be filled in a uniform and dense manner however high m is chosen. On the other hand, the presence of an attractor structure implies that the trajectories will get attracted to a lower dimensional subset, in the sense of Mandelbrot, meaning that the attractor has an intricate structure when observed on any space scale. A stochastic signal will not exhibit such a structure.

2.2. ATTRACTOR DIMENSIONS

Attractor dimensions are the most widely used invariant measures for chaotic non-linear dynamical systems. The dimensions associated with chaotic systems are fractional in contrast to the integer dimensionality of regular, integrable systems [10]. Beyond the simplest concept of 'dimension', as the number of coordinates needed to specify a state, is the geometrically related concept of how (hyper)volumes (V) scale as a function of a characteristic length parameter (L): $V \propto L^D$. Suppose a strange attractor set S is covered with boxes of size ϵ, and there are $N(\epsilon)$ such boxes. If the i^{th} box is being visited with probability p_i, then the generalized q^{th} order information dimension D_q is defined as

$$D_q = \lim_{\epsilon \to 0} \frac{1}{q-1} \frac{\log \sum_{i=1}^{N(\epsilon)} {p_i}^q}{\log \epsilon}. \quad (2)$$

In practice, numerical implementation of such a box counting procedure is inefficient. Instead, algorithms have been developed to calculate the the *correlation* dimension D_2; D_2 has been found to characterize well the attractor dimension in experimental situations. The Grassberger-Procaccia [11] algorithm is the best known algorithm for computing the the *correlation integral* $C(\epsilon)$ and, hence, D_2. The correlation integral is expressed as:

$$C(\epsilon) = \frac{2}{N(\epsilon)(N(\epsilon)-1)} \sum_{i \neq j}^{N(\epsilon)} \theta(\epsilon - ||y(j) - y(i)||) \quad (3)$$

The separation ($||\mathbf{y}(j) - \mathbf{y}(i)||$) between the embedded vectors $\mathbf{y}(j)$ and $\mathbf{y}(i)$ is evaluated in terms of an L_2 or an L_∞ norm. The function $\theta(.)$ refers to the Heaviside's unit-step function of its argument. It has been proven [11] that this correlation integral shows an exponential behavior ($C(\epsilon) \propto \epsilon^\nu$) revealing the geometrical scaling property of the strange attractor. In practice a plot of $\log C(\epsilon)$ versus $\log(\epsilon)$ for a set of increasing embedding dimensions is made and the slope of this plot, which will reach a saturation value after a minimum embedding dimension, yields the attractor dimension D_2.

2.3. LYAPUNOV SPECTRA

While attractor dimensions characterize the (spatial) distribution of points in the state space, lyapunov exponents (LE) describe the dynamics of the (temporal) evolution of the trajectories. Lyapunov exponents describe the exponential divergence and/or convergence of trajectories towards an attractor in a multidimensional flow and reflect the properties of the underlying attractor by their sign and magnitude. LEs are defined by the time dependent behavior of small deviations from the flow. The set of d lyapunov exponents $\{\lambda_i\}$ for a d-dimensional embedding constituting the lyapunov spectrum is usually ordered as $\lambda_1 > \lambda_2 > \lambda_3 \ldots$. Positive lyapunov exponents, hallmarks of chaotic behavior, indicate a strong instability within the attractor. Presence of one positive LE indicates a *simple* chaos and, the presence of more than one positive LE indicates *hyper* chaos. Chaotic time series have many common features with stochastic processes, such as decaying autocorrelations and limited predictability, but the detection of attractors and positive LEs help to distinguish between chaos and noise. Two methods are used to obtain the LEs: the WSSV algorithm [12], which determines the largest LE, and the Sano-Sawada algorithm [13] which estimates the local Jacobian matrix from the time evolution of a number of adjacent trajectories and yields all LEs. In this paper, we used a modified Sano-Sawada algorithm to evaluate the lyapunov spectrum; the modified algorithm is described in [13]. The algorithm has been found to systematically underestimate the negative exponents. This limitation does not pose a problem here since we are interested only in positive exponents. A second limitation is the occurrence of spurious exponents when a reconstructed attractor with an embedding dimension greater than the actual attractor dimension is used. These spurious exponents, however, tend to wander with varying embedding dimensions while the true exponents show a plateau-behavior with changing embedding dimensions.

3. EXPERIMENTAL DATA ANALYSIS

The time-series data used in this study were far-field acoustic pressure waveforms of the voiceless fricatives /s/ and /ʃ/ and their voiced counterparts /z/ and /ʒ/, respectively, in V_1CV_2 utterances where the vowel was /i/, /a/ or /u/. Both symmetric and non-symmetric contexts were considered. There were five repetitions of each utterance as spoken by two female native speakers of American English. The data also included continuous repetitions of the symmetrical utterances till the speakers ran out of breath (e.g. *asa asa asa...*). The data were digitized at 48 kHz (16 bits/sample).

3-D phase portraits were constructed from the fricative time-series data. The portraits were used to identify the presence or absence of attractor structures. The constructions used 1000 samples (20.8 ms) of the steady-state fricative segment in each token. In addition, time-series data of pure tones (500 Hz, 2000 Hz), uniform and Gaussian noises, and synthetic and natural vowels (/i/, /a/ and /u/ by the same speakers) were analyzed to help us in interpreting the results. Clear strange attractor structures were found for the fricatives (Fig 1). The attractor structure is patently distinct from the stable orbit of a tone, a torus structure of a vowel or the absence of structure for noise. The presence of strange attractors suggest the possibility of chaotic dynamics (as opposed to stochastic dynamics) in the fricative time-series and led us to further chaotic analyses such as evaluating fractal dimensions and lyapunov spectra. It should be noted here that the strange attractor structure could not be revealed if less than 200-300 samples (4.2-6.4 ms) were used in constructing the phase portraits. In addition, we examined the effects of using different sampling rates (8, 12, 24, 48 and 96 kHz) on the structure of the phase plots. Our results indicate that attractor structures are revealed at a minimum sampling rate of 12kHz; higher sampling rates are obviously preferable. On the other hand, inadvertent oversampling can lead to artifacts in the dimension calculations [14]. For our study, a sampling rate of 48kHz was used.

The correlation integral $C(\epsilon)$ was evaluated from the scalar time-series data using the Grassberger-Procaccia algorithm under varying time delay embedding dimensions. The plot of $\log C(\epsilon)$ versus $\log(\epsilon)$ (Fig. 3a) or alternatively, the slopes obtained by a linear regression on $\log C(\epsilon)$ and its two neighbors versus $\log(\epsilon)$ (Fig. 3b) can be used to evaluate D_2. The fractal correlation dimension D_2 for the fricatives (/s/, /z/, /ʃ/, /ʒ/) was found to be /4.2, 5.9, 8.0, 6.3/, respectively, for speaker AK and /4.4, 4.8, 7.9, 6.3/, respectively, for speaker BB. The D_2 values were obtained from analyzing /aCa/ utterances. Similar results were obtained for the other vocalic contexts and in non-symmetrical contexts.

The lyapunov spectrum for each fricative was calculated using the algorithm described in [13]. 4000-5000 samples (83-104 ms) per token were used in the analysis. The analysis was repeated for varying embedding dimensions. The mutual information in the data was evaluated using a separate computer program and those results were used to specify the correlation lengths. The evolution time for the lyapunov spectrum evaluation program was chosen to be 0.25 milliseconds and the Gram-Schmidt Reorthonormalization was repeated every 0.3 milliseconds.

Results indicate a consistent presence of a relatively small positive exponent suggesting the presence of low-dimensional simple chaos. The presence of spurious exponents did not pose a problem since we were mainly interested in the largest (positive) exponent. Results were found to be consistent across repetitions and across tokens. Fig. 4a shows a sample lyapunov spectrum for the /s/ data in *asa* context for speaker AK. Exponent values calculated for varying embedding dimensions for the same data is plotted in Fig 4b.

We also found the phase trajectories for the voiced fricatives to be distinctly different from those of the voiceless ones (Fig. 1). In fact, phase trajectories seem to be more powerful than both the time waveform and the DFT spectrum (Fig. 2) in distinguishing between voiced and voiceless fricatives.

4. DISCUSSION

Results using modern chaotic time-series analysis techniques suggest the presence of chaotic dynamics in the acoustic pressure time-series data of fricatives. Our results also showed that phase trajectories can be used as a voiced/voiceless classification technique for fricatives. Analysis of a larger data set from several speakers is needed to determine the generality of our conclusions. Furthermore, it is not clear how closely far-field acoustic pressure data represent the production dynamics of fricatives. Analysis of turbulent flow and pressure measured *inside* the vocal tract near the constriction is needed to determine the chaotic or otherwise stochastic nature of fricative production.

5. REFERENCES

[1] G.Pfister, T. Buzug, and N. Enge, "Charaterization of experimental time series from taylor-couette flow," *Physica D*, vol. 58, pp. 441–454, 1992.

[2] W. Lauterborn and J. Holzfuss, "Acoustic chaos," *Int'l Journal of Bifurcation and Chaos*, vol. 1, no. 1, pp. 13–26, 1991.

[3] C. Pickover and A. Khorasani, "Fractal characterization of speech waveform graphs," *Comput. & Graphics*, vol. 10, no. 1, pp. 51–61, 1986.

[4] I. R. Titze, R. J. Baken, and H. Herzel, "Evidence of chaos in vocal fold vibration," NCVS Status and Progress report pp 39–63, National Center for Voice and Speech, Dec. 1992.

[5] W. Mende, H. Herzel, and K. Wermke, "Bifurcations and chaos in newborn infant cries," *Physics Letters A*, vol. 145, pp. 418–424, Apr. 1990.

[6] J. A. S. Kelso, E. Vatikiotis-Bateson, E. L. Saltzman, and B. Kay, "A qualitative dynamic analysis of reitrant speech production: Phase portraits, kinematics and, dynamic modeling," *JASA*, vol. 77, pp. 266–280, Jan. 1985.

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

[8] D. Ruelle, "Strange attractors as a mathematical explanation of turbulence," in *Statistical Models and Turbulence* (J. Ehlers, K. Hepp, and H. A. Weidenmuller, eds.), vol. 12 of *Lecture Notes in Physics*, pp. 292–299, Springer-Verlag, Berlin, 1971.

[9] R. Mañé in *Dynamical Systems and Turbulence* (D. A. Rand and L. S. Young, eds.), p. 230, Springer-Verlag, Berlin, 1981.

[10] J. D. Farmer, E. Ott, and J. A. Yorke, "The dimension of chaotic attractors," *Physica D*, vol. 7, pp. 153–180, 1983.

[11] P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Phy. Rev. Lett*, vol. 50, pp. 346–349, Jan. 1983.

[12] A. Wolf, J. Swift, H. Swinney, and J. Vastano, "Determining lyapunov exponents from a time series," *Physica D*, vol. 16, pp. 285–317, 1985.

[13] T.-M. Kruel, M. Eiswirth, and F. W. Schneider, "Computation of lyapunov spectra: Effect of interactive noise and application to a chemical oscillator," *Physica D*, vol. 63, pp. 117–137, 1993.

[14] G. Mayer-Kress, "Application of dimension algorithms to experimental chaos," in *Directions in Chaos* (H. Bai-lin, ed.), vol. 1, pp. 122–135, World Scientific, 1987.

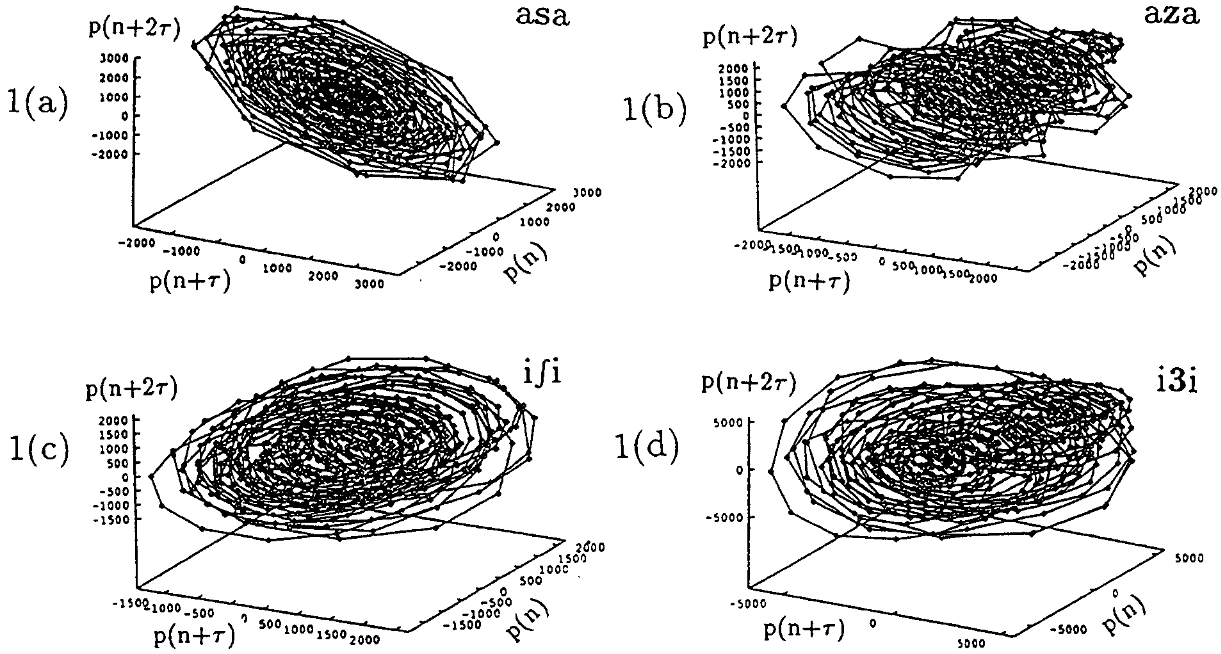

Fig. 1 a–d: Phase trajectories of the consonant portion in symmetric VCV syllable for one speaker(BB). data length 20.833 msec; sampled at 48kHz; Delay $\tau = 2$.

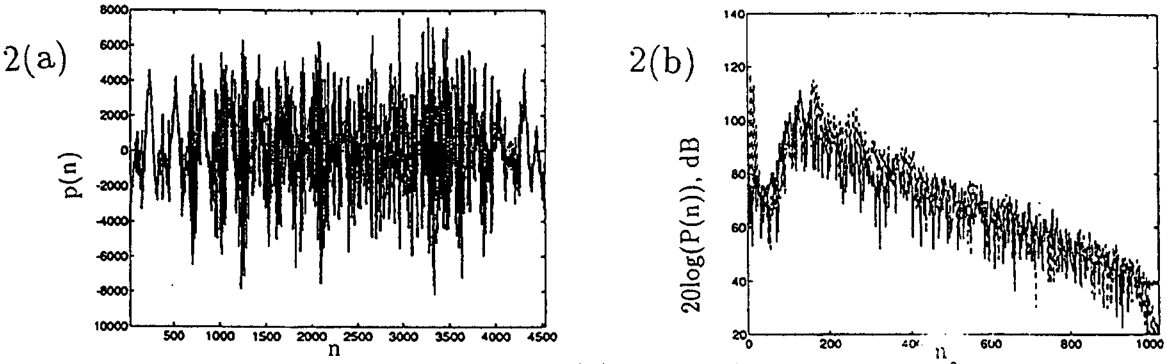

Fig. 2 (a): Time waveform of /3/ in *i3i* (b) DFTs of /3/ (dotted) and /∫/ (solid) using 2048 Hamming windowed samples.

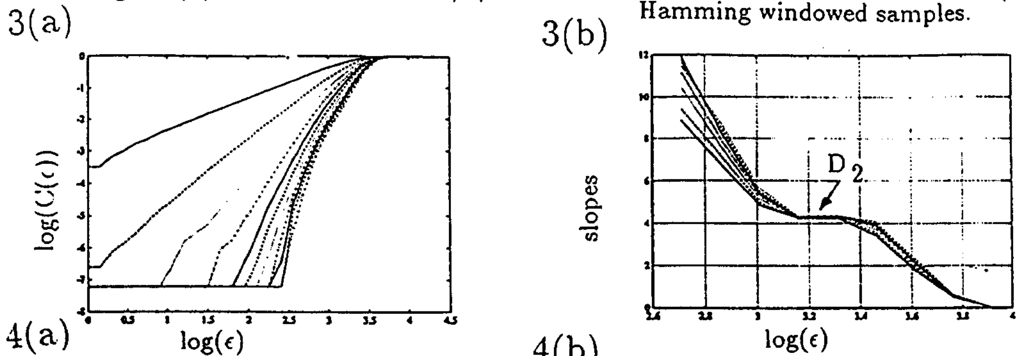

Fig. 3 (a–b): Correlation dimension calculation for /s/ in *asa*. (d = 3 to 15)

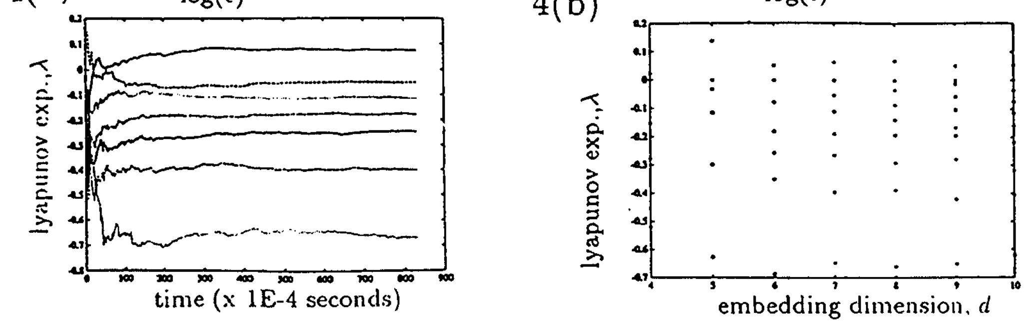

Fig. 4 (a): Lyapunov Spectrum for /s/ in *asa* with $d = 7$.

(b) Average Lyapunov Exponents for varying d for /s/ in *asa*.