A nonlinear dynamical systems analysis of fricative consonants

Shrikanth S. Narayanan and Abeer A. Alwan
Department of Electrical Engineering, UCLA, 405 Hilgard Avenue, Los Angeles, California 90024

(Received 17 May 1994; accepted for publication 16 December 1994)

Acoustic waveforms of the strident fricatives /s/, /z/, /ʃ/, and /ʒ/ spoken by two native American English speakers are analyzed using modern chaotic analysis techniques. Fricative data are extracted from both intervocalic and sustained utterances. For comparison, acoustic waveforms of the vowels /a/, /i/, and /u/ are also analyzed. For 44% of the unvoiced fricative tokens in VCV contexts and 59% of the sustained voiced fricatives, indications of low-dimensional dynamics could be found with the given limitations of stationarity. The low-dimensional chaotic behavior is exhibited by a correlation dimension (D_2) ranging between 3 and 7.2, and by positive maximum Lyapunov exponents (LEs). For the remaining fricatives, results suggest that the dimensional complexity therein is greater than the maximum D_2 value that could be reliably estimated from the available data (about 7.8 for the intervocalic cases and 9 for the sustained cases). Intervocalic voiced fricatives are excluded from the analysis due to stationarity requirements. Analysis of vowels, on the other hand, indicates nonchaotic behavior demonstrated by folded limit cycles and nonpositive maximum LEs; this is consistent with results of previous studies. Findings are interpreted in terms of posited articulatory and aerodynamic parameters of turbulence in the production of fricative consonants.

PACS numbers: 43.70.Aj, 43.70.Bk, 43.72.Ar, 43.25.Rq

INTRODUCTION

Turbulence phenomena in fluids may be studied using different approaches. Derivation of analytical models of turbulence from physical principles is, in general, difficult and requires several simplifying assumptions regarding the system and its geometry. Hence in many practical situations, experimental data form the basis for model formulation rather than just serving as a vehicle for the validation of a specific analytically derived model. The irregular behavior exhibited by the physical variables representing turbulence, typically pressure or velocity signals, may be analyzed from either a stochastic or a deterministic point of view. The irregularity may be manifested in either, or a combination of, the amplitude, phase or period of the signal. The stochastic view point assumes the signal to be a realization of a random process and uses concepts such as autocorrelation functions and power spectra for signal characterization. The application of deterministic, nonlinear-systems theoretic concepts in analyzing and modeling turbulent flows, however, has recently received wide attention in fluid dynamics and physics (Tatsumi, 1984; Helleman, 1986; Dwoyer *et al.*, 1985). In the heart of such nonlinear deterministic approaches lies the notion of chaos and bifurcation theory. In principle, phenomena such as turbulent fluid flows are modeled by an infinite-dimensional system. It is also known that the asymptotic system behavior in dissipative dynamical systems may relax on to a small invariant subset of a full state space. Application of nonlinear systems concepts to various experimental data has demonstrated that the turbulent behavior therein may be characterized by a low-dimensional attractor instead of an infinite-dimensional system. Examples of such experimental investigations cover a wide range of areas including Rayleigh–Bénard convection, Taylor–Couette flow (Brandstater and Swinney, 1987; G. Pfister *et al.*, 1992), lasers (Stoop and Meier, 1988), chemical oscillators (Kruel *et al.*, 1993), acoustic cavitation (Lauterborn and Holzfuss, 1991), solar activity (Kurths and Herzel, 1987), and radar (Haykin and Leung, 1992).

The results of these studies have led to a better understanding of the underlying physical phenomena, disregarded as "noise" till then, in terms of categorization, and provided grounds for the construction of low, finite-dimensional dynamical models from an observed time series.

The dimension underlying 3D turbulence can be arbitrarily large; in fact, it has been argued by Manley (1984) that although a fluid system described by Navier–Stokes equations has essentially finite degrees of freedom, the actual problem arises when one attempts to estimate the dimensionality of the system. Low dimensionality in turbulence occurs for a relatively small range of the Reynolds number. Limitations in existing numerical techniques and data length requirements prohibit reasonable estimates for dimensions greater than ten. Hence, the question that remains to be answered is how one might attempt to identify, and model, low-dimensional turbulent systems. Time-series techniques such as power-spectral analysis which characterize the irregular behavior as broadband noise do not distinguish between the low- and high-dimensional system dynamics that resulted in the signal. Nonlinear deterministic techniques, on the other hand, can provide information about the dimensionality and other dynamical properties of the underlying system.

In speech, certain sounds, such as fricatives, are produced by generating turbulence in the vocal tract. In this study, a nonlinear deterministic approach to the analysis of fricatives is undertaken. The objective is to find out whether fricative turbulence is low dimensional or high dimensional. Detection of low-dimensional deterministic behavior may help in the development of better source models for these sounds. No study has used this approach before in analyzing

 0001-4966/95/97(4)/2511/14/$6.00

fricatives although there has been a considerable interest in applying nonlinear dynamical systems principles in speech analysis and modeling. Concepts of fractal geometry have been used for speech waveform characterization (Pickover and Khorasani, 1986; Baken, 1990; Maragos, 1991). Tishby (Tishby, 1990) suggested the possibility of modeling speech as an output of a chaotic dynamical system. Townshend's study (Townshend, 1992), an elaboration of Tishby (1990), reports a low dimensionality of about 3.3 for speech signals. These calculations were made from a "global" point of view of speech without distinguishing between specific sound classes. A nonlinear predictor for speech was then developed based on local approximation techniques (Sidorowich, 1992). The 3-dB prediction gain demonstrated by the nonlinear predictor over a linear predictor was offered as a converging evidence for deterministic nonlinear attributes in the speech signal. The computational complexity and speaker-dependent training required by the nonlinear predictor limit the advantages of such a scheme. Other studies have used nonlinear deterministic techniques in the analysis of vocal fold vibration in normal and pathological voiced speech (Herzel *et al.*, 1994; Titze *et al.*, 1993; Herzel, 1993; McLaughlin and Lowry, 1993), and newborn infant cries (Mende *et al.*, 1990). Results for nonpathological vowels have revealed a nonchaotic low-dimensional behavior. Geometrical techniques, such as phase portraits, have also been used to analyze articulatory data describing lip movement (Kelso *et al.*, 1985).

In this paper, a brief review of the theory of fricative production mechanisms and a description of the algorithms used in our analysis is first presented. In the sections following, experimental results are described followed by a discussion and suggestions for future work.

I. THEORY AND ANALYSIS TECHNIQUES

A. Fricative mechanisms

Fricatives are produced by the formation of a narrow supraglottal constriction in the vocal tract and the generation of turbulence in the region downstream the constriction when air flows through the vocal tract (Fant, 1960; Stevens, 1971). The generation of turbulence occurs near the vocal-tract walls and/or the teeth which act as an obstacle to the airflow. In addition to turbulence, the vocal folds may vibrate, at least for part of the frication period, as in the case of voiced fricatives. The eight fricative consonants in English, specified in terms of their place of articulation in unvoiced-voiced pairs, are the labiodentals /f/ and /v/, dentals /θ/ and /ð/, alveolars /s/ and /z/, and palatals /ʃ/ and /ʒ/. It is believed that the production of fricatives is characterized by complex, nonlinear, fluid dynamical phenomena. The source mechanisms for fricatives are not completely understood. The vocal tract is relatively inaccessible for direct area function measurements and for *in vivo* pressure and flow studies making direct physical modeling of the production mechanisms difficult. The acoustic signal radiated from the lips typically forms the basis for the analysis. The spectra of fricatives are characterized by the presence of high-frequency broadband energy, typically in the range above 3 kHz. In the time domain, the acoustic signal of voiceless fricatives exhibits a highly irregular behavior whereas that of the voiced fricatives is "nearly periodic" due to voicing.

During fricative production, the air flow becomes turbulent at a critical value of the Reynolds number, $\mathrm{Re}_{\mathrm{crit}}$. The squared Reynolds number can be expressed as $\mathrm{Re}^2 = 4\rho^2 V^2/\pi\mu^2 A_c$ where V is volume velocity at the constriction, A_c is the area of constriction, ρ is the fluid density and μ is the viscosity (Schroeter and Sondhi, 1992). For air flow in tubes with rough surfaces, $\mathrm{Re}_{\mathrm{crit}}$ is about 2000 (Streeter, 1962). Typical flow rates during strident fricative production are in the range of 200–500 cc/s and the supraglottal constriction areas range between 0.075–0.4 cm^2 (Stevens, 1971; Narayanan *et al.*, 1994). The flow rate and area data suggest that Re values range between 2700 and 12625; this in turn indicates that varied degrees of turbulence may be expected during fricative production. Evidence of intra- and interspeaker variabilities in the production of fricative consonants has been illustrated in several studies (Subtelny *et al.*, 1972; Hardcastle and Clark, 1981; Warren *et al.*, 1981; Stone *et al.*, 1992; Narayanan *et al.*, 1994). Intraspeaker variabilities for sustained fricative utterances are relatively minimal, in contrast with those observed in vocalic contexts (VCV utterances, for example). The articulatory events in a sustained utterance correspond to a relatively static vocal tract shape in comparison with the dynamical events that occur in a vocalic context. A factor that introduces further variability in the degree and manner of turbulence is the transition between a vowel and a fricative in a vocalic context. During the vowel-fricative transition, the flow pattern changes from a presumably laminar pattern, during vowel production, to a turbulent one in the fricative. In VCV utterances, for example, the onset of turbulence for unvoiced fricatives may occur even before full constriction is achieved and continue even after the constriction area starts increasing (Stevens *et al.*, 1992). Hence varying degrees of turbulence is expected in the fricative segment. A third factor that might affect turbulence generation is devoicing of the voiced fricatives. Devoicing may affect the aerodynamic interaction between the voicing source at the glottis and the turbulence generated at the supraglottal constriction.

B. Analysis techniques

Although, linear signal analysis techniques such as Fourier transforms and autocorrelation functions, provide converging evidence for a signal's deterministic attributes, they are not sufficient to characterize a chaotic time series. Geometrical techniques, such as phase-portrait constructions, followed by a careful evaluation of invariant characteristics, such as the attractor dimensions and Lyapunov spectra, are required to analyze a chaotic time series. This approach views turbulence as exhibiting deterministic behavior described in terms of attracting sets in a phase space and resulting in *strange* or *chaotic* attractors. Additional evidence for deterministic behavior in the analyzed data is provided by an exponential decay in the power spectrum at high frequencies, or equivalently a linear decay in a semilogarithmic plot (Brandstater and Swinney, 1987).

1. Reconstruction of the phase space

The ideal scenario for dynamical state-space modeling would be one where all the system states are accessible for measurement. In most practical situations, however, the experimental data considered are typically measurements of a single scalar observable $\{p(t_k)\}$ at a fixed spatial point. Hence, the first step in our modeling is to reconstruct the system state space (phase space) from the observed measurements. Time-delay embedding (Takens, 1981; Ruelle, 1971) is the most commonly used technique for mapping scalar data into the multidimensional phase space especially when analyzing experimental data. A point $\mathbf{P}(t_k)$ in such a d-dimensional phase space is given by $\mathbf{P}(t_k)=\{p(t_k),p(t_k+T),\ldots,p(t_k+(d-1)T\}$; the choice for the time delay T is essentially arbitrary. A sufficient condition for the choice of d, an integer, depends on the attractor dimension d_A, which can be fractional (Takens, 1981; Ruelle, 1971). 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 attractor dimension d_A, however, is not known *a priori*. Since, d_A is not known, the selection of the embedding dimension is essentially by trial and error. The procedure of choosing a sufficiently large d is formally known as embedding and the minimum dimension that reveals the attractor structure is called the embedding dimension (d_e). Once a large enough $d=d_e$ has been achieved, any $d\geqslant d_e$ will also provide a valid embedding. Although the embedding theorem poses no constraints on the choice of T, the mutual information is typically used for calculating the time delay for embedding. Based on Fraser and Swinney's criterion (Fraser and Swinney, 1986), T for reconstruction is chosen from the first minimum time of the mutual information function evaluated over all t_k. In practice, in order to preserve the fragmentation of the sampled signal in the time domain as close as possible to the continuous-time speech signal, the signal needs to be sampled above a minimum sampling rate. Inadvertent oversampling, however, can lead to artifacts in the dimension calculations (Mayer-Kress, 1987; Theiler, 1990).

2. Phase portraits

Irregular signals may arise from stochastic or deterministic forces. 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 to be. 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. The attractor will then have an intricate structure when observed on any space scale. A truly stochastic signal will not exhibit such a structure. As a first step in chaotic analysis, it is customary to make the phase plots, even though human visual capabilities are limited to three dimensions $(d_e=3)$. In many cases, such plots provide an indication of the underlying structure even though neither the absence nor presence of structure is a conclusive evidence of chaos. For illustrative purposes, we show in Fig. 1

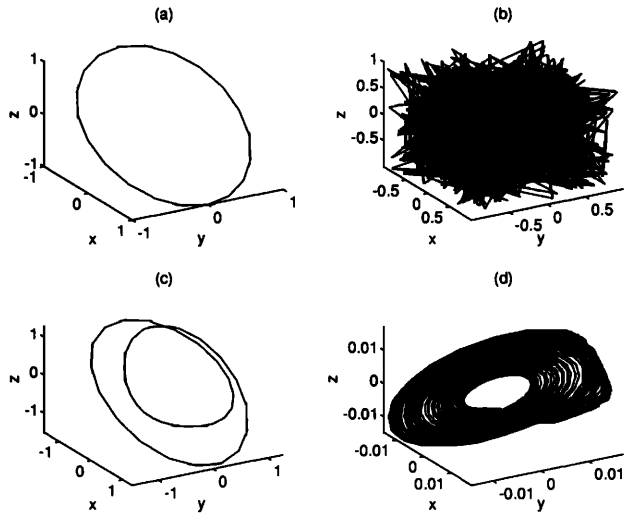

FIG. 1. Phase plots in embedding dimension $d_e=3$ with $T=2$ and $N=1000$. (a) Tone at 500 Hz. (b) Uniform noise. (c) Folded limit cycle. (d) Rössler attractor.

some phase plots of typical attractors. The phase plot corresponding to a tone at 500 Hz [Fig. 1(a)] reveals a *stable orbit* while that of uniform noise [Fig. 1(b)] is characterized by the absence of any structure. The phase plots corresponding to a folded limit cycle with two loops is shown in Fig. 1(c). Figure 1(d) shows a chaotic attractor generated from a mathematical model of the Rössler attractor (Wolf *et al.*, 1985).

3. Attractor dimensions

Attractor dimensions are the most widely used invariant characteristics for chaotic nonlinear dynamical systems (Farmer *et al.*, 1983; Theiler, 1990). Among the several definitions that exist for attractor dimensions, the *correlation* dimension (D_2) has been found to be the most useful due to the relative ease of its evaluation in practical situations. The correlation integral $C(r)$, denoting the number of pairs of points with Euclidean distance $\leqslant r$ in the attractor, shows a power law behavior, ν, $(C(r)\propto r^{\nu})$, for small values of r, revealing the geometrical scaling property of the strange attractor (Grassberger and Procaccia, 1983). The power-law dependence of the correlation integral, ν, is taken to be a measure of D_2. The correlation dimension estimate for a time series of length N may be expressed as $D_2=\lim_{r\to 0}\lim_{N\to\infty}[\log C(N,r)/\log r]$ (Theiler, 1990). The algorithm for evaluating $C(N,r)$ is computationally simple and has been found to converge rapidly. In practice, for a given N, a plot of $\log C(r)$ versus $\log(r)$ 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 (Grassberger and Procaccia, 1983). Since, in most cases, the plot is dominated by the effects of noise and digitization in the region of $[r\to 0]$, intermediate length scales, where constant slopes can be found reliably, are typically used for D_2 estimation. Moreover, D_2 can be directly estimated by finding the "local" slopes $\{d[\log C(r)]/d[\log(r)]\}$ (Theiler, 1990); one way of estimating this derivative is by performing a

linear regression on each point and its two neighbors on the $\log C(r)$ versus $\log(r)$ curve. In practice, the choice of scaling region length depends on various factors such as data length and signal-to-noise ratio. Some practical issues involved in estimating D_2 are discussed in Sec. II C.

4. 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 indicate the exponential divergence or convergence of trajectories toward an attractor in a multidimensional flow and reflect the properties of the underlying attractor by their sign and magnitude. The set of d Lyapunov exponents $\{\lambda_i\}$ for a d-dimensional embedding constituting the Lyapunov spectrum is usually ordered as a decreasing sequence: $\{\lambda_1 \geq \lambda_2 \geq \lambda_3 ...\}$. Positive Lyapunov exponents, hallmarks of chaotic behavior, indicate a strong instability within the attractor. Presence of one positive LE indicates *simple* chaos (Wolf *et al.*, 1985, for example) and the presence of more than one positive LE indicates *hyper* chaos (Kruel *et al.*, 1993, for example). 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 can help in distinguishing between chaos and random noise.

There are two popular methods in obtaining Lyapunov exponents: The WSSV algorithm (Wolf *et al.*, 1985), which determines the largest LE, and the Sano–Sawada algorithm (Sano and Sawada, 1985) which estimates the local Jacobian matrix from the time evolution of a number of adjacent trajectories and yields all LEs. In this paper, the Sano–Sawada algorithm that was modified for dealing with experimental data by Stoop and Meier (Stoop and Meier, 1988) was used to evaluate the Lyapunov spectrum. The modifications in the algorithm were essential to deal with problems such as finite precision and measurement noise found in experimental data. The reason for selecting this particular algorithm for our analysis was that it provided an opportunity to explore the entire Lyapunov spectra (more than one positive exponent is possible), a feature not provided by the WSSV algorithm. The computer program that implements the algorithm was obtained from T. Kruel and is described in Kruel *et al.* (1991, 1993).

This algorithm has certain limitations. It was found to systematically underestimate the absolute values of the negative exponents. This limitation does not pose a problem here since we are mainly interested 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.

C. Practical issues and limitations

Dimension and Lyapunov exponent estimation algorithms have several limitations, particularly when dealing with experimental data. The fundamental limitation of the analysis techniques is posed by the available number of data points, N. The maximum correlation dimension that can be estimated using N data points is $2 \log_{10} N$ (Eckmann and Ruelle, 1992). Other researchers have posed more stringent requirements on the estimate's reliability (Mayer-Kress, 1985). The accuracy of the estimates, in general, increases as the amount of data (points per orbit of the attractor) increases. Increasing data lengths, however, results in a significant increase in computational intensity. Other sources for systematic and statistical errors include measurement noise, edge effects and autocorrelation effects. One of the most important requirements that has to be met in order to minimize these errors is to use long data sets of high quality, a difficult requirement to attain in most practical situations. The autocorrelation effects in the time series data result in erroneous dimension estimates due to the breakdown of the scaling behavior, particularly at large values of embedding (Theiler, 1990; Theiler, 1987). One way to minimize this problem is to avoid taking points that are within some specified autocorrelation length, typically calculated by evaluating the mutual information contained in the signal (Kennel, 1993).

The concepts of dimension and Lyapunov spectra refer to asymptotic properties of the attractor and require stationarity of the time series. An operational definition for stationarity in chaos signal analysis, considering long data sets, is that there is no significant power at low frequencies (Theiler, 1991). For example, linearly correlated noise with a $1/f^{\alpha}$-power spectrum ($\alpha > 1$) has been found to possess fractal properties. The time series corresponding to such correlated noise, however, is characterized by long autocorrelation lengths and significant power at low frequencies, clearly violating stationarity conditions. Hence, the fractal properties of noise with a $1/f^{\alpha}$-power spectrum do not correspond to chaotic data. Care has to be taken to ensure stationarity of the signal prior to the analysis.

In distinguishing chaotic data from data resulting from a random process, in general, it is useful to compare analysis results with results obtained from a corresponding surrogate data set. A stringent test would involve creating surrogate time series with the same power spectra as the original data and subjecting the surrogate data to the same analysis as that done on the original data. If the results of the surrogate data are significantly different from those of the original data, then the null hypothesis that the original time series is random noise can be rejected. The use of surrogate data sets has been adopted in experimental situations (Grassberger, 1986; Gober *et al.*, 1992; Kurths and Herzel, 1987). One method of creating a surrogate data set is to take the Fourier transform of the original data, randomize the phase and then inverse the Fourier transform to obtain the surrogate time series. This approach was adopted in our study. Phase randomization using Gaussian random sequences was implemented in such a way that the resulting Fourier transform of the surrogate data had conjugate symmetry, thereby yielding a real signal when inverse Fourier transformed.

Given the limitations of existing numerical techniques, precise estimates of the attractor dimension and Lyapunov exponents (LE) are difficult to achieve. Hence, actual num-

bers of the dimension and LEs are not the final aim of this study. These values, however, indicate the "dimensional complexity" of the analyzed data (Mayer-Kress, 1985).

II. EXPERIMENTAL DATA ANALYSIS

As shown in the previous section, determining whether or not a system is chaotic involves analyzing experimental time series data in several stages. First, reconstruction of the phase space with an appropriate choice of the time delay T is carried out. The projection of the phase trajectories is drawn in 3D to provide initial insight into the problem. Presence of attractor structure in the phase plots can be an indication of chaos; this would prompt further analysis. Linear decay at high frequencies in the semilogarithmic power spectral plot can also provide preliminary evidence for deterministic attributes in the signal. Attractor dimensions and the Lyapunov spectra are then evaluated, whenever possible, to provide quantitative characterization of the attractor in the phase space.

In this paper, time-series data of fricatives, both in sustained and intervocalic contexts, were analyzed. In addition, sustained vowel utterances and surrogate data were analyzed to aid in the interpretation of the results.

A. Data acquisition

The time-series data used in this study were far-field acoustic pressure waveforms of the unvoiced fricative consonants /s/ and /ʃ/ and their voiced cognates /z/ and /ʒ/, respectively, in symmetric VCV utterances (V: vowel, C: consonant) where the vowel was /i/, /a/, or /u/. Four repetitions of each utterance spoken by two phonetically trained female native talkers of American English (AK, BB) were recorded. The VCV syllables, set in a phrase "*Say* VCV *again*," had the initial stress on the first vowel. The data also included, three tokens of steady sustained fricative utterances and sustained vowels made at normal speech levels. The recordings were made in a sound proof facility (IAC 3696 model 1202 A) with an AKG C460B microphone. A fixed reference was used to help the speaker maintain a steady head position during the recording; the microphone was retained in position relative to this fixed reference. The speech material was directly digitized into a digital audio tape (DAT) recorder at a sampling rate of 48 kHz (16 bits/sample). The acoustic recording was done in a manner similar to that described by Shadle (1985). The microphone was placed approximately at six to eight inches from the talker's mouth at an angle of 15 deg to the left from the line perpendicular to the plane of the lips to avoid picking up wind noise. The recording microphone had a low-frequency cutoff at 70 Hz and the DAT recorder (Sony DTC 75ES) had built-in antialiasing filters that ensure a flat frequency response from 2 to 22 000 Hz within ±0.5 dB. The signals were redigitized into a SUN SPARC workstation using an audio-DSP port interface (Ariel ProPort model 656) at a sampling rate of 48 kHz. The sigma-delta oversampling converters used in the A/D conversion employ near-perfect antialiasing filters with extremely flat passband (±0.003 dB) and high stop-band attenuation (>96 dB). The passband cutoff is automatically fixed at half the sampling rate used.

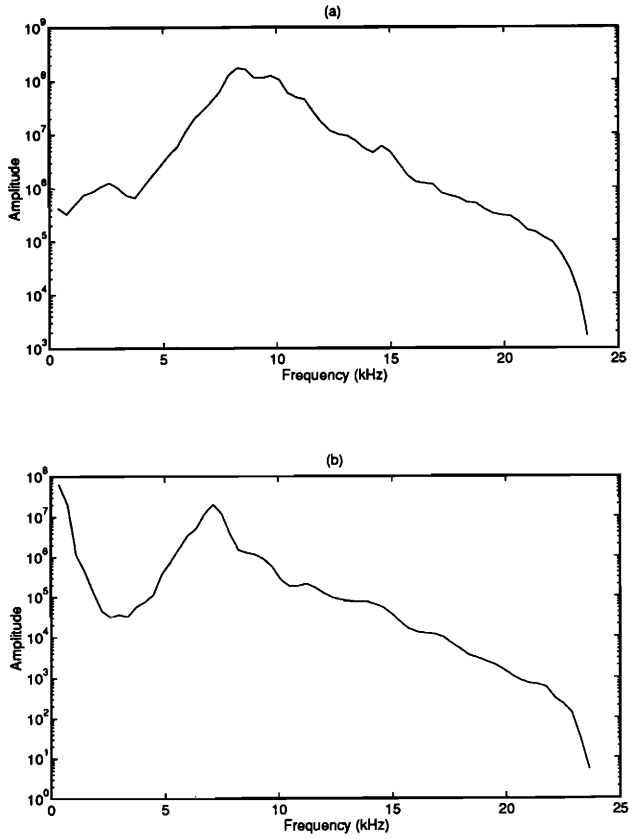

FIG. 2. Power spectra for the fricatives (a) /s/ and (b) /z/ calculated using $N=8192$ (speaker AK).

The fricative segments obtained from the VCV utterances correspond to the "steady-state" portion of frication which was identified with the aid of a time-aligned spectrographic display. In a manner similar to Stevens *et al.* (1992), the beginning of this steady-state portion was taken to be the time when the spectrum shape departed significantly from the canonical spectrum shape of the preceding vowel. A similar criterion was used to determine the end of the fricative segment. For the sustained utterances, the transient segments, both at the beginning and end of the utterances, were excluded from the analyses. In VCV contexts, the number of data points were in the range of 5800 to 8000 (120.8 to 166.6 ms) for the unvoiced fricatives and 2800 to 4500 (58.3 to 93.7 ms) for the voiced fricatives. For the sustained utterances, the available data lengths were in the range of 1.9×10^4 to 2.4×10^4 (4 to 5 s).

B. Power spectra

Power spectra for adjacent, nonoverlapping, 256-point segments were first computed. The resulting spectra were then averaged over the entire duration of the analyzed utterance. Typical power spectra for sustained /s/ and /z/ are shown in Fig. 2. The power spectra clearly show a linear decay at high frequencies. This high-frequency behavior is distinct from that exhibited by the power spectra of vowels and provides converging, but nonconclusive, evidence re-

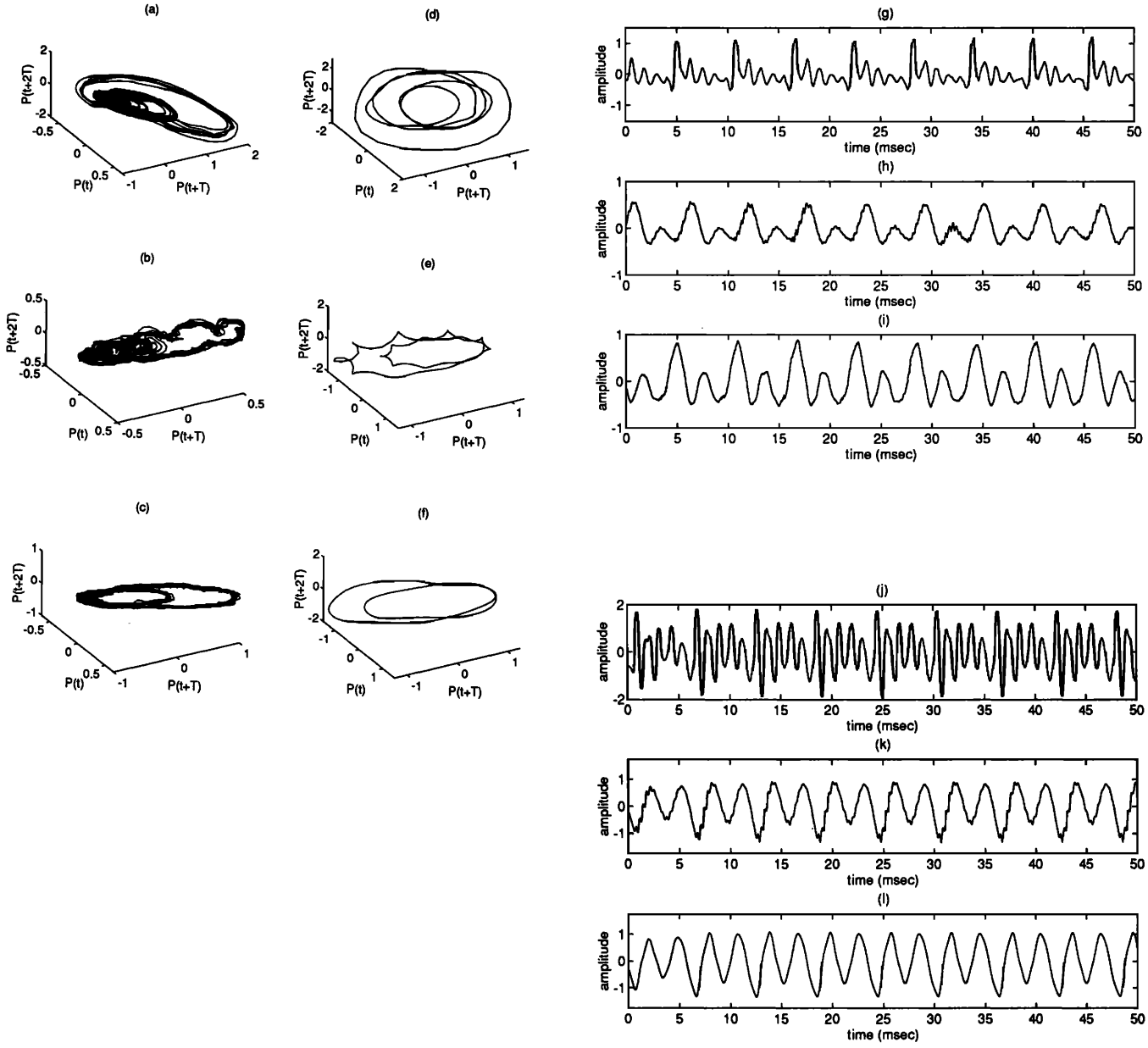

FIG. 3. (a)–(f). Phase plots for vowels in embedding dimension $d_e=3$ with $T=2$ and $N=2500$ (about 9 pitch cycles). (a)–(c) Natural utterances of /a/, /i/, and /u/, respectively, (speaker AK). (d)–(f) Synthetic vowels /a/, /i/, and /u/, respectively. (g)–(i) Time waveforms for natural utterances of the vowels /a/, /i/, and /u/, respectively (speaker AK). (j)–(l) Time waveforms for the synthetic vowels /a/, /i/, and /u/, respectively.

garding the nonlinear deterministic behavior of the data. Similar power spectra were obtained for fricatives in VCV utterances.

The fricative power spectra were evaluated at different sampling rates: 11.025, 16, 32, and 48 kHz. It was found that linear decay at high frequencies was evident only for sampling rates greater than 16 kHz. For our study, a sampling rate of 48 kHz was used. This high sampling rate did not introduce artifacts in the dimensional analysis as shown in a later section.

C. Stationarity

Absence of significant power at low frequencies for the unvoiced fricatives [for example, Fig. 2(a)] is in agreement with the operational definition of stationarity suggested by Theiler (1991). This definition of stationarity can not be used in the case of the voiced fricatives since voicing contributes to power in the low-frequency end of the spectrum. Instead, a durational constraint is imposed on voiced sounds so that a large number of pitch cycles is included in the analysis and these signals could then be assumed stationary. A duration of 100 ms or longer seems to be sufficient in this regard (Herzel, 1993). For our data, the available data lengths for the intervocalic voiced fricatives are, in general, shorter than 100 ms to warrant reasonable stationarity assumptions. Furthermore, we found that about 25% (3/12) and 17% (2/12) of the intervocalic /z/ tokens for speakers AK and BB, respectively, were devoiced. Similarly, about 33% (4/12) of the intervocalic /ʒ/ tokens, for both speakers, showed devoicing. As a result of devoicing, near-periodicity is not evident in the en-

tire fricative segment and stationarity of the data is questionable. Hence, due to data length and stationarity restrictions, the intervocalic voiced fricatives were excluded from further analysis. Sustained voiced fricatives analysis, on the other hand, were considerably longer than 100 ms, were not devoiced, and, hence, were subjected to analysis.

D. Phase plots

Three dimensional phase portraits were constructed from the fricative time-series data. Synthetic and naturally produced vowels (/a/, /i/, and /u/) were analyzed for comparison. For the phase plots $d_e=3$ and $T=2$ were used. The value of T corresponds to the smallest delay indicated by the first minimum time of the mutual information function calculated from the unvoiced data.

1. Vowels

The naturally spoken vowels (/a/, /i/, and /u/) showed normal variations in fundamental frequency (*jitter*) and period-to-period amplitude (*shimmer*). The pitch and first four formant frequencies were estimated using the Entropics signal processing software (ESPS, 1994). The range of pitch variation was 165–171 Hz for AK and 155–177 Hz for BB. Formant and pitch values were averaged over the duration of each utterance and the average values were then used to generate synthetic vowels using the formant synthesizer SENSYN (SenSyn, 1993). Formant frequencies higher than $F4$ were set at the default values of the synthesizer. The synthesis was performed at the maximum sampling rate of the synthesizer (20 kHz) and the signals were later up-sampled by a factor of 5:2 (to 50 kHz) to enable comparison with the natural utterances sampled at 48 kHz.

The phase plots corresponding to the vowels /a/, /i/, and /u/ for speaker AK are shown in Fig. 3(a)–(c) (natural) and Fig. 3(d)–(f) (synthetic). The corresponding time waveforms are shown in Fig. 3(g)–(i) (natural) and Fig. 3(j)–(l) (synthetic). The phase plots were constructed using 2500 samples (41.66 ms, approximately eight to nine pitch periods). The plots reveal a folded limit cycle structure for the vowels; this structure is not an indication of chaos. The limit cycle structures of the naturally produced vowels show a "dispersed" form, when compared to the synthetic cases, highlighting some jitter and shimmer present in the natural utterances. Phase plots for the synthetic vowels, however, do not reveal any dispersive behavior. Recall that the vowel synthesis employed time-invariant values of the pitch and formant frequencies. The correspondence between the time waveforms and the phase plots is clear. The number of loops in a phase plot equals the number of significant peaks per pitch period in the time waveform. The relative size of the loops corresponds to the relative amplitudes of the peaks in the time waveform. For example, in the synthetic /a/, the five significant peaks in the time waveform [Fig. 3(j)], manifest themselves as five loops in the phase plot [Fig. 3(d)]. The loops corresponding to the fourth and fifth peaks in the natural /a/ [Fig. 3(a)] are, however, indistinguishable due to their relatively small, and comparable, amplitudes in the time waveform [Fig. 3(g)]. High-frequency oscillations present in the time waveform of the vowel /i/ [Fig. 3(h) and (k)] result in nonsmooth contours in the phase plots [Fig. 3(b) and (e)]. The relationship between the vowel phase plots and time waveforms for the other tokens of speaker AK, and for speaker BB is similar.

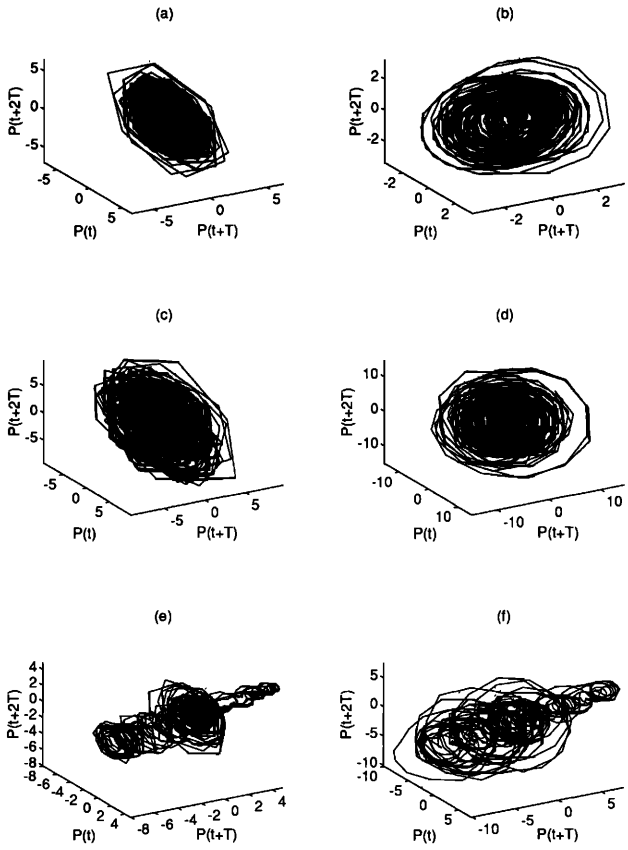

FIG. 4. Phase plots for (a) /s/ and (b) /ʃ/ in symmetric VCV context with the vowel /a/. Phase plots for sustained fricatives: /s/ (c), /ʃ/ (d), /z/ (e), /ʒ/ (f). Embedding dimension $d_e=3$ with $T=2$ and $N=4800$.

2. Fricatives

Sample phase plots of /s/ and /ʃ/ in symmetric VCV contexts with the vowel /a/ (speaker AK) are shown in Fig. 4(a) and (b). Figure 4(c)–(f) show phase plots of the fricatives /s/, /ʃ/, /z/, and /ʒ/ obtained from sustained utterances by the same speaker. The phase plots were constructed using 100-ms data segments (4800 samples); this data length corresponds to about 17 pitch cycles for the voiced fricatives. Phase plots constructed using speaker BB's data were similar.

The phase plots for fricatives reveal a structure that is different from the stable orbit of a tone, (folded) limit cycle of a vowel, or the absence of structure for random noise (Figs. 1, 3). As illustrated in Fig. 4, phase trajectories of the voiced fricatives differ from those of the unvoiced ones. The 3D plots show that the underlying periodicity in the voiced cases manifests itself in a fashion similar to an "extended spring." We also found that removing the low-frequency energy (by high-pass filtering with the cutoff frequency around the pitch value) of the voiced fricatives results in a phase plot similar to that of the unvoiced fricatives. A high-pass filtered data set of the voiced fricatives, however, was not used for

TABLE I. Results of dimensional analysis of the unvoiced fricatives in VCV context (four tokens each) indicating the presence or absence of converging power-law behavior in their correlation integrals. D_2 estimates are provided for the cases that showed convergence (speakers AK and BB). Tokens that did not reveal convergence are marked with "N."

Fricative in VCV	Speaker AK: Token No.				Speaker BB: Token No.			
	1	2	3	4	1	2	3	4
/asa/	5.8	N	5.9	5.2	N	5.2	N	N
/isi/	N	4.2	6.5	N	4.5	5.8	5.0	5.6
/usu/	N	N	4.5	N	4.6	N	4.8	N
/aʃa/	N	7.2	7.0	N	N	N	6.4	N
/iʃi/	6.7	N	N	N	6.8	N	6.4	N
/uʃu/	N	N	N	N	N	N	7.0	6.2

any further analysis. Although results of experimental and modeling studies offer evidence of aerodynamic interaction between the glottis and constriction in the vocal tract (Bickley and Stevens, 1986), the exact nature of the interaction between the voicing and fricative sources is not completely understood. Hence, at this point, it is not clear if separation of voicing effects from the turbulence "noise," through high-pass filtering, which assumes simple linear superposition, is warranted. A possible explanation of this particular appearance of the voiced fricatives' phase plots is that voicing is functioning as a carrier for the chaotic excursions resulting in the stretching of the phase plot along a certain plane.

The phaseplots for fricatives suggest two possibilities: The underlying time series is linearly correlated noise or is a chaotic time series. This necessitates further analysis that includes the evaluation of the dimension and Lyapunov spectra. It should be noted that the observed pattern of the phaseplots could not be revealed if less than 200–300 samples (4.2–6.3 ms) were used in their construction. For the voiced fricatives, this corresponds to about one pitch period.

The observations from the power spectra and phaseplots led us to evaluate attractor dimension and Lyapunov spectra to obtain a quantitative characterization of the signals analyzed.

E. Correlation dimension

The correlation integral $C(r)$ was evaluated from the scalar time-series data using a box-assisted-type implementation of the Grassberger–Procaccia algorithm (Goldberg, 1993) under varying time-delay embedding dimensions. Plots of $\log C(r)$ versus $\log(r)$ were made and the correlation dimension was estimated by calculating the slope of the curves in cases where a saturation of the slopes over a reasonable scaling region, for increasing embedding dimensions was revealed. The local slopes may be directly calculated through a linear regression on $\log C(r)$ and its two neighbors versus $\log(r)$; these slopes can then be used to estimate D_2. Effects of different data lengths and time delays were examined.

1. Intervocalic fricatives

Four tokens of each unvoiced fricatives, /s/ and /ʃ/, in symmetric VCV contexts (a total of 48 tokens) were considered for dimension analysis. As mentioned in Sec. II C, the

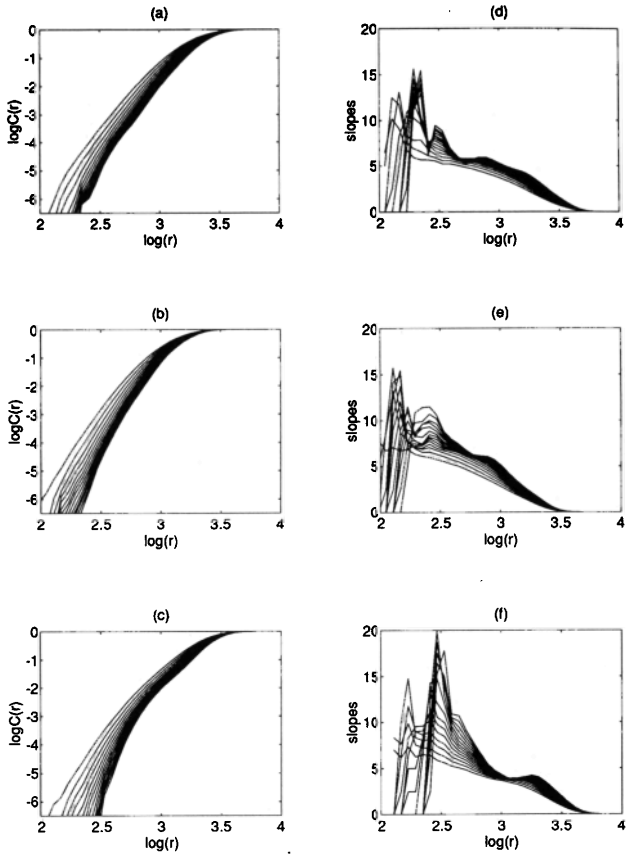

FIG. 5. Examples showing converging scaling regions in correlation integrals for /s/ in symmetric VCV contexts (speaker AK). (a)–(c) Correlation integrals $\log C(r)$ vs $\log(r)$ for /s/ with the vowels /a/ ($N=7195$), /i/ ($N=6416$), and /u/ ($N=5841$), respectively. (d)–(f) Local slopes versus $\log(r)$ corresponding to (a)–(c), respectively. Embedding dimension $d_e=7-20$; curves for increasing d_e's move from left to right in steps of one.

intervocalic voiced fricatives were excluded from the analysis due to stationarity requirements. For the available data lengths, (5800–8000 samples), the upper bound of a reasonable estimate of the dimension D_2 is 7.5–7.8 (Eckmann and Ruelle, 1992). Results of the dimensional analysis are summarized in Table I. The table shows the cases where the presence/absence of converging scaling regions for the fricatives was found [regions were typically within the range $2<\log(r)<3$]. For the cases with a converging scaling behavior, the value of the slope was taken as the estimate of the dimension D_2. It should be pointed out that the D_2 estimates may not be precise due to relatively small scaling regions revealed by the data; these values, however, provide an indication of the dimensional complexity of the underlying physical processes. The results shown in Table I imply that the behavior of the scaling regions in the correlation integral plots and their convergence with increasing embedding dimensions do not show a generalizable trend across different tokens, contexts or speakers. There were cases that offered converging slopes over reasonable scaling zones and cases that showed no convergence at all. About 54% (13/24) of the /s/ tokens and 33% (8/24) of the /ʃ/ tokens revealed a low-dimensional behavior. For illustration purposes, we show in

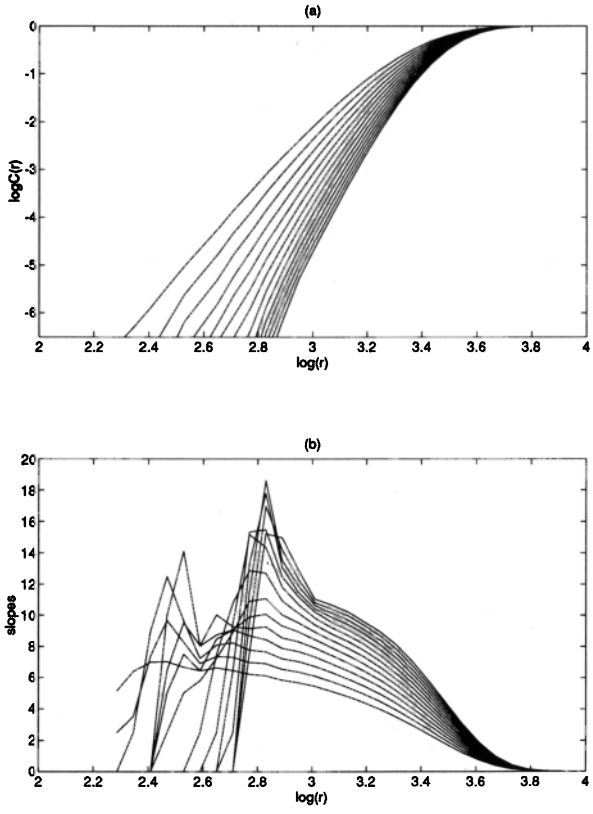

FIG. 6. Examples showing nonconvergence of the scaling regions in correlation integrals for /s/ in symmetric VCV context with the vowel /a/ ($N=6125$, speaker AK). (a) Correlation integrals $\log C(r)$ vs $\log(r)$. (b) Local slopes versus $\log(r)$ corresponding to (a). Embedding dimension $d_e=7-20$; curves for increasing d_e's move from left to right in steps of one.

Figs. 5 and 6 examples of both converging and nonconverging cases. Figure 5(a)–(c) show the correlation integrals for /s/ in symmetric intervocalic contexts with /a/, /i/, and /u/, respectively, for speaker AK and for $d_e=7$ to 20. The corresponding plots of the local slopes, shown in Fig. 5(d)–(f), reveal converging regions from which an estimate of the correlation dimension may be made. In contrast, a plot of the correlation integrals and local slopes for an /s/ token that does not reveal converging scaling behavior is shown in Fig. 6.

Variability in D_2 estimates for different tokens may be due to variations in the Reynolds number (Re) of the underlying turbulence (an increase/decrease in Re results in an increase/decrease in the dimension). The results where no convergence of the correlation integral was found suggest that the underlying turbulence may be high dimensional, i.e., well beyond the upper bound of the dimension value that could be reliably estimated with the available data (in our case $D_{2,\max}=7.8$). Other studies have shown that at high values of the Reynolds number, the underlying turbulence may possess a dimension much higher than the value that an experimentalist could estimate (Sieber, 1987; Brandstater and Swinney, 1987). This is because the data requirements increase (at least) exponentially with increasing dimensions.

Two important issues arise at this juncture. The first is to determine whether, or not, longer data sets can be used for analysis. The sampling rate could not be increased inadvertently because of artifacts introduced in the dimensional calculations. Alternatively, sustained fricative utterances may be used to generate considerably longer data sets (Shadle, 1985; Badin, 1989). It should be noted that steady sustained fricatives are produced by a relatively static vocal tract shape and do not involve any apparent laminar to turbulent flow transitions, such as those found in vocalic contexts, implying relatively small intraspeaker variabilities. The second issue is to verify the validity of the low-dimensional behavior exhibited by some fricative data. This can be addressed by analyzing surrogate data.

2. Sustained utterances

In the second part of the dimensional analysis, long data sets obtained from sustained fricative utterances (three tokens for each /z/ and /ʒ/) were analyzed. The duration of the sustained fricatives was four to five seconds as opposed to 60 to 170 ms for the fricative segments in VCV contexts. Effects of different data lengths, delays, and embedding dimensions were studied.

a. Unvoiced fricatives. Due to computational limitations, the maximum data length considered for analysis was $N=1.5\times10^5$ data points (3.125 s); this value of N results in an upper bound of a reliable D_2 estimate to be $D_{2,\max}=10.3$ (Eckmann and Ruelle, 1992). In addition, two segments of lengths $N=10^4$ and $N=4\times10^4$ (implying $D_{2,\max}$ of 8 and

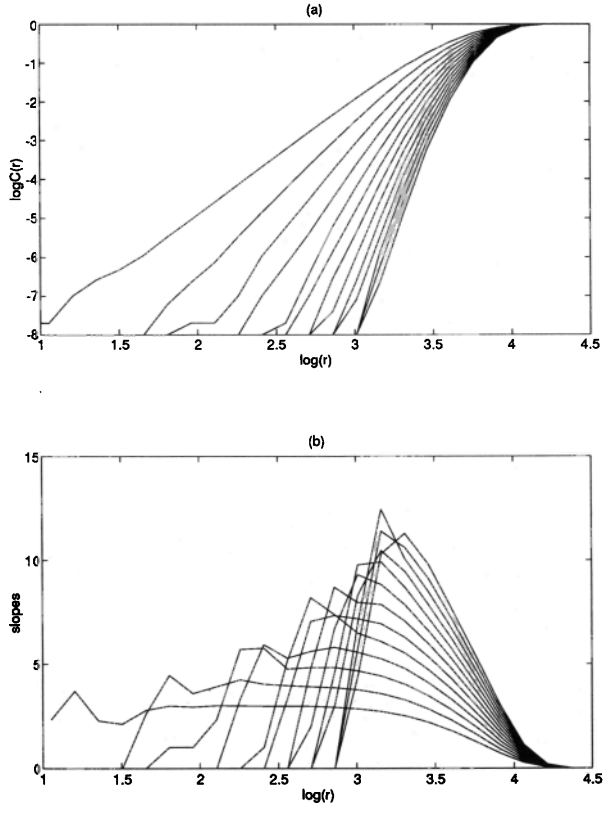

FIG. 7. Correlation integrals under different embedding dimensions for a sustained /s/ ($N=40000$, speaker AK). $d_e=3-15$, in steps of one. Curves for increasing d_e's move from left to right in the plots. Plots of local slopes corresponding to (a) appear in (b) and reveal no convergence in scaling regions.

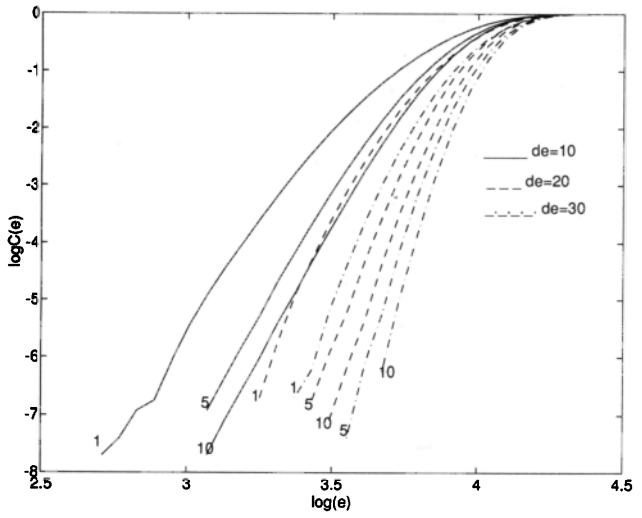

FIG. 8. Correlation integrals for different time delays ($T=1,5,10$) under different embedding dimensions $d_e=10$ (solid), $d_e=20$ (dashed), and $d_e=30$ (dot-dashed) for sustained /s/ ($N=40000$, speaker AK). No convergence in scaling regions is found.

9.2, respectively) were selected from each sustained fricative token to enable us to study the effect of increasing data lengths on the dimension estimates. Correlation integral plots for /s/ and /ʃ/ calculated using $N=10^4$ and $N=1.5\times10^5$ for embedding dimensions of $d_e=10$ and 20 revealed no convergence of slopes over reasonable scaling regions (i.e., power law behavior of the correlation integral). Since data requirements increase exponentially with maximum estimable dimensions $D_{2,\max}$, we could not pursue data lengths longer than those mentioned above. In addition, there are limitations on how long an utterance could be sustained comfortably by a speaker. Our phonetically trained subjects could sustain fricatives in a steady manner up to 10–12 s. After excluding transients, say we consider for analysis $N=5\times10^5$ points at a sampling rate of 48 kHz, this would yield $D_{2,\max}=11.4$ which, from a practical point of view, is not much higher than $D_{2,\max}=9.2$ corresponding to $N=4\times10^4$ points. On the other hand, the computational load increases tremendously with increasing data lengths while not offering much of an advantage in terms of the maximum estimable dimensions. Hence, for the rest of the analysis, we used $N=4\times10^4$ points (833.3 ms). Figure 7 shows the correlation plots for /s/ under varying embedding dimensions of $d_e=3-15$. Again, no convergence of scaling regions was found. Similar results were obtained for the /ʃ/ tokens. The effect of various delays on the correlation integrals was examined and the correlation integral plots (Fig. 8) did not reveal any "tail effects," (Caputo *et al.*, 1989), ascertaining that oversampling was not a possible source of error in our analysis.

b. Voiced fricatives. Three tokens of sustained /z/ and /ʒ/ utterances with $N=4\times10^4$ (about 140 pitch periods, $D_{2,\max}=9.2$) were analyzed to examine how the results compare with those for the unvoiced cases. The tokens analyzed were not devoiced. For /z/, two out of the three tokens, of both speakers, revealed a low dimensional behavior. For /ʒ/, two tokens of AK and one token of BB revealed low-

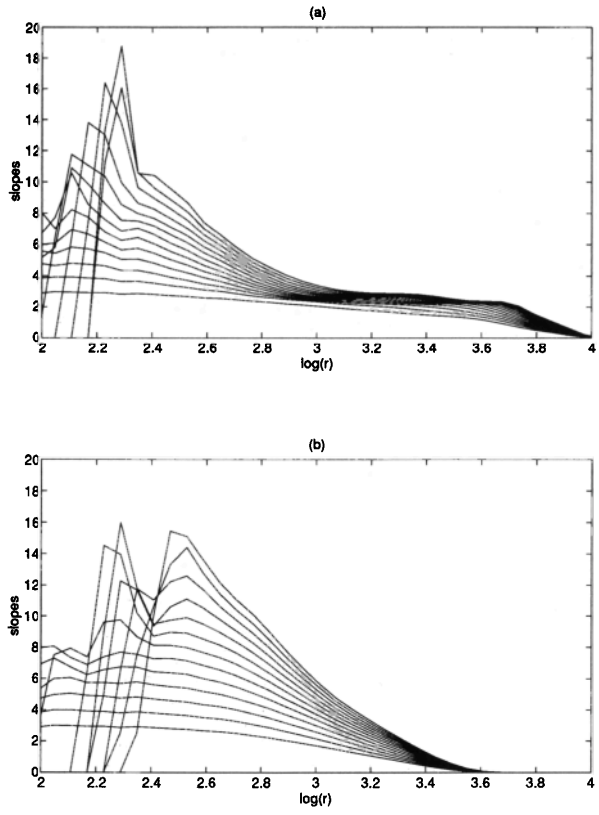

FIG. 9. Dimensional analysis for sustained voiced fricatives for speaker AK. (a) /z/ (b) surrogate data for /z/. $d_e=8-15$ and $N=40\ 000$. Convergence of scaling is demonstrated by the original data and not by the surrogate data.

dimensional behavior. The D_2 estimates for the low-dimensional voiced fricatives were in the range of 3.0–4.8. Figure 9(a) shows typical results of the local slopes versus $\log r$ of /z/ for speaker AK where a low-dimensional behavior was detected. Analysis results of the surrogate data are shown in Fig. 9(b). The results clearly indicate that a low-dimensional behavior is not revealed in the surrogate data case, when compared to the original data, supporting the possibility of a nonlinear deterministic behavior in the fricative data. In the next part of the analysis, we analyzed the tokens that exhibited low-dimensional behavior by evaluating the Lyapunov spectra.

F. Lyapunov spectra

As mentioned in Sec. I B, a modified Sano–Sawada algorithm was used to evaluate the Lyapunov spectra. Recall that the data length requirements for the Lyapunov exponent estimation are more stringent than those required for reliable D_2 estimates. Hence, only tokens with low-dimensional behavior ($D_2<9$) were analyzed since it was not possible to assume a reasonable value for the embedding otherwise. In addition, the Lyapunov spectra of a simple tone, natural and synthetic vowels were evaluated for comparison. The calculations were repeated for varying embedding dimensions and evolution lengths. Here again, the mutual information in the data was used to specify the time interval to be excluded from the calculations in order to avoid autocorrelation ef-

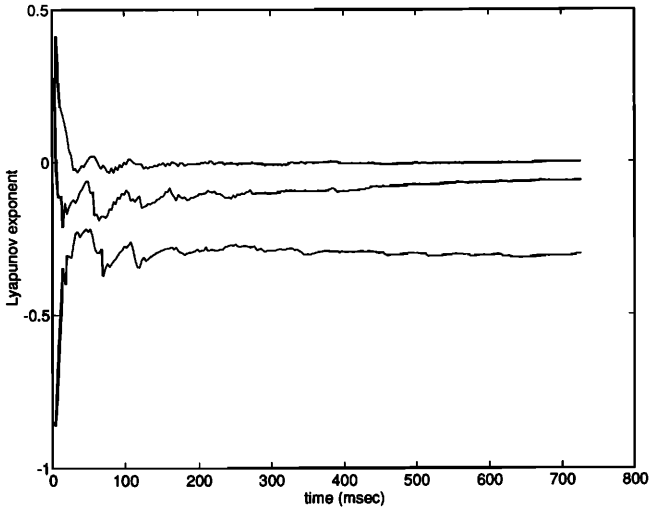

FIG. 10. Lyapunov spectrum for the vowel /a/ (speaker AK). $d_e=3$ and $N=35\ 000$.

fects. The evolution time for the Lyapunov spectrum evaluation program was 0.7 ms and the Gram–Schmidt reorthonormalization was repeated every 1.2 ms.

The Lyapunov spectra corresponding to a tone (at 500 Hz) revealed a $(0,-,-)$ pattern for $(\lambda_1,\lambda_2,\lambda_3)$, indicating one zero-valued exponent and two negative-valued exponents, which is typical for limit cycles. Figure 10 shows the Lyapunov spectra corresponding to the vowel /a/ for speaker AK. The figure reveals the maximum exponent λ_1 of zero which is consistent with previous observations on nonpathological vowels (Herzel, 1993). The value of λ_2 is close, but not equal, to zero for the vowels. The Lyapunov spectrum is shown for $d_e=3$. Higher embedding dimensions were found to present consistent results but with additional spurious exponents. The results for the vowels /i/ and /u/ revealed similar Lyapunov spectra.

The Lyapunov spectra of the low-dimensional fricatives indicate a consistent presence of a relatively small positive exponent suggesting the presence of simple chaos. Figure 11(a)–(b) show sample Lyapunov spectra for /s/ and /ʃ/, respectively, spoken in /aCa/ context by subject AK. The effect of varying embedding dimensions was examined and sample results for the same /s/ token are shown in Fig. 12. The figure illustrates the convergence of the positive exponent with increasing embedding dimensions. Average values calculated over the plateau region (typically, final 95% of the segment) are shown and the error bars, shown only on the positive exponent value for the sake of clarity, indicate deviations over the duration of averaging.

The behavior of the maximum Lyapunov exponent with evolution length may also be used to distinguish stochastic noise from chaos (Kurths and Herzel, 1987). Maximum Lyapunov exponents of stochastic signals do not exhibit plateaulike behavior with evolution time. Chaotic signals, on the other hand, exhibit plateau behavior in the variation of the maximum Lyapunov exponent over reasonable evolution lengths implying exponential growth over several scales. The effect of varying the evolution length on the maximum

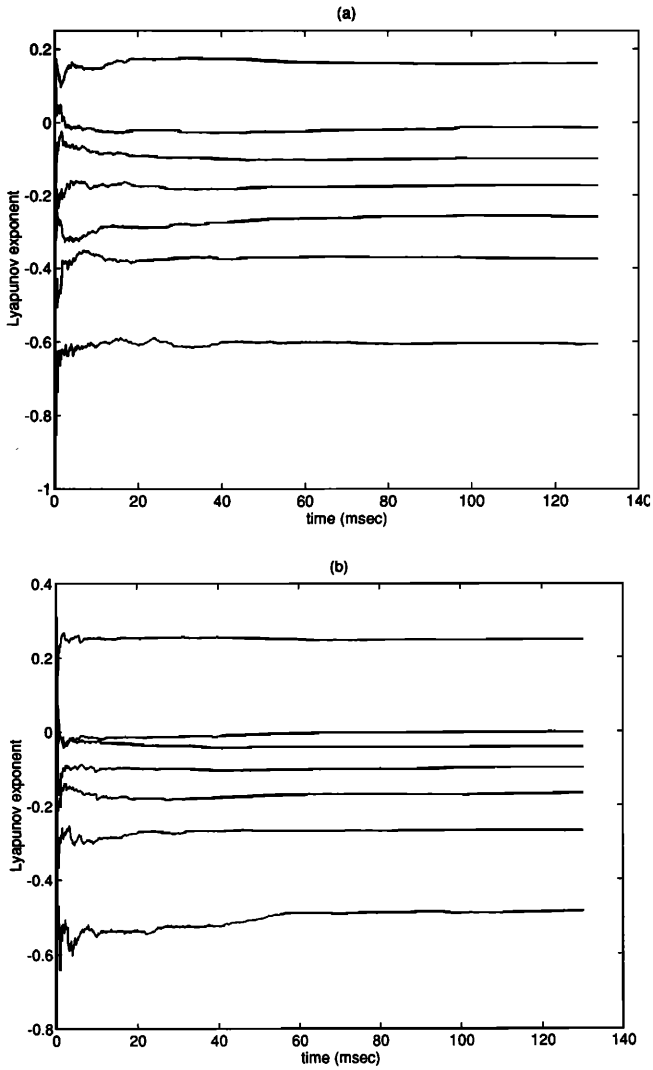

FIG. 11. Examples of Lyapunov spectra for low-dimensional fricatives ($d_e=7$). (a) /s/ in /asa/ context ($N=7195$, $D_2\approx4.2$). (b) /ʃ/ in /aʃa/ context ($N=6108$, $D_2\approx6.7$), speaker AK.

Lyapunov exponent, for the /s/ token considered in Fig. 11(a), is shown in Fig. 13. As shown in this figure, the relative plateau behavior of the Lyapunov exponents of the /s/ token, when compared to its surrogate data, suggests the possibility of an underlying chaos.

Finally, the sustained voiced fricative tokens that had previously revealed a low-dimensional behavior were analyzed. Figure 14(a)–(b) show the Lyapunov spectra for a /z/ and /ʒ/ token, respectively. Analysis of the variation of the maximum Lyapunov exponent with varying evolution lengths and comparison with the behavior of a corresponding surrogate data set suggests nonlinear deterministic attributes in the fricative data.

III. SUMMARY AND DISCUSSION

In this study, acoustic waveforms of the strident fricatives in English /s, ʃ, z, ʒ/, obtained from VCV and sustained utterances, were analyzed using modern chaotic analysis techniques. Time-series data of the sustained vowels /a/, /i/, and /u/ were also analyzed for comparison. Phase plots and

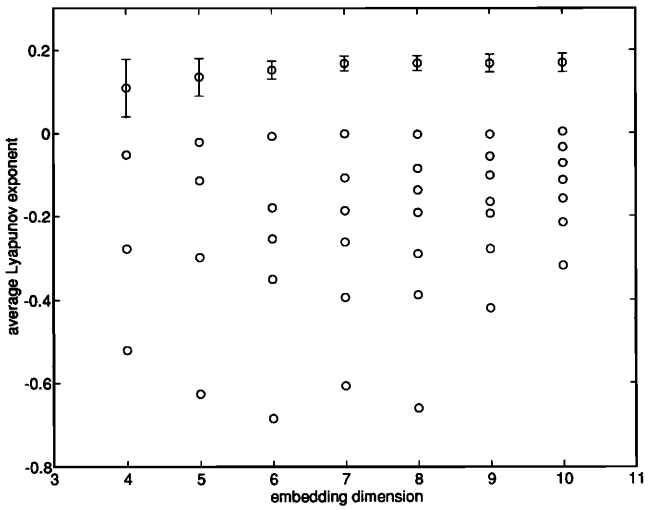

FIG. 12. Average Lyapunov exponent values under varying embedding dimensions $d_e=4-10$ for /s/ in /asa/ context [same token as in Fig. 11(a)]. Error bars marking standard deviation in the maximum (positive) exponent values show a decreasing trend with increasing d_e.

power spectra were suggestive of a nonlinear deterministic behavior in the fricative data at sampling rates greater than 16 kHz. To determine the dimensional complexity of the speech sounds analyzed, attractor dimensions were evaluated.

Results of the dimensional analysis for the unvoiced fricatives in VCV contexts were mixed: about 54% (13/24) of the /s/ tokens and about 33% (8/24) of the /ʃ/ tokens exhibited a power-law behavior in the correlation integral values. Dimension (D_2) estimates were in the range of 4.2–6.5 for /s/ and 6.2–7.2 for /ʃ/. The lower percentage of convergence in D_2 estimates for /ʃ/ together with the relatively higher D_2 values, compared to those for /s/, suggest that the dimensional complexity of /ʃ/ is likely to be higher than that of /s/.

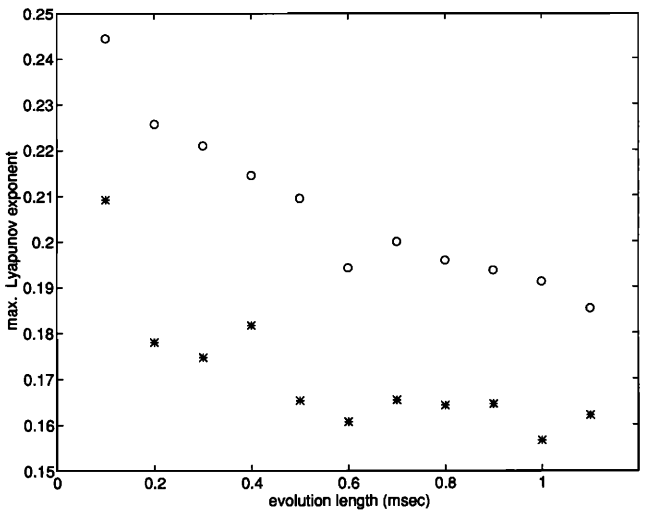

FIG. 13. Variations in maximum Lyapunov exponent for varying evolution lengths ($d_e=7$) for /s/ in /asa/ context [asterisks, same token as in Fig. 11(a)], show a plateau-behavior when compared to the corresponding surrogate data set (open circles).

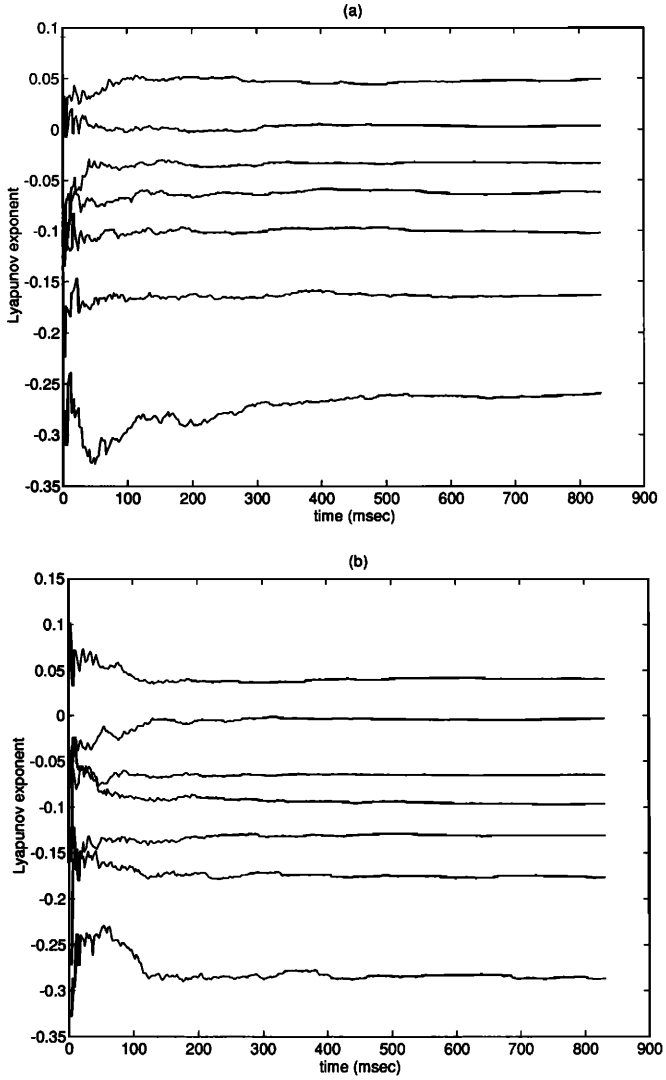

FIG. 14. Examples of Lyapunov spectra for sustained voiced low-dimensional fricatives with $d_e=7$ and $N=40\ 000$ for speaker AK. (a) /z/ ($D_2\approx3$). (b) /ʒ/ ($D_2\approx4$).

Relatively higher values for the flow rates for /ʃ/, compared to /s/ (Hixon, 1966), and hence higher Reynolds numbers, can explain our results. The voiced fricatives in VCV contexts were excluded from the dimensional analysis due to data length and stationarity restrictions.

In the second stage of the analysis, sustained fricative utterances, which are considerably longer than fricatives in VCV contexts, were analyzed. Dimensional analysis revealed no converging power-law behavior in the correlation integrals for the sustained unvoiced fricatives, indicating that the dimensional complexity therein is greater than nine. Any further analysis was limited by data requirements in terms of both data availability and computational complexity. Results for the sustained voiced fricatives revealed low-dimensional behavior for 59% of the tokens (7/12) with $D_2\approx3-4.8$. It should be noted that the phase plots did not provide an indication of the variability in the dimensional behavior.

Variability in D_2 estimates is attributed to posited variabilities in the underlying articulatory and aerodynamic parameters such as constriction areas and flow rates. As a re-

sult, the Reynolds number, and hence the dimensionality of the fricative turbulence, can assume a wide range of values. In addition, in a vocalic context, the flow pattern changes from laminar to turbulent when the speech produced changes from a vowel to a fricative. Assuming that dimensional estimates provide a measure of the "degree" of turbulence, we can expect varied degrees of turbulence in fricative production. Although it is theoretically possible for the turbulence in fricative data to posses finite dimensional complexity, limitations posed by finite data lengths prohibit verification of such a claim.

The differences in the dimensional behavior of the unvoiced fricatives in VCV contexts when compared to those in sustained utterances may be attributed to: (1) coarticulatory influences of the adjacent vowels, (2) the more deliberate production of the sustained utterances as opposed to the more natural and speechlike production in VCVs. As a result, it is possible that the aerodynamical/articulatory conditions (flow rates, constriction areas) for a fricative in a VCV context may not always attain values that lead to a high-dimensional turbulence such as that evident for the sustained fricatives. Differences in the dimensional behavior of the sustained unvoiced and voiced fricatives suggests that the dimensional complexity for the voiced fricatives is less than that of their unvoiced cognates. These differences may be due to the interaction between the voicing source at the glottis and the turbulence generated at the supraglottal constriction for the voiced fricatives.

To find out if the low-dimensional fricative tokens are truly chaotic, Lyapunov spectra were evaluated. A consistent presence of a positive exponent for these fricatives suggests a low-dimensional chaos. Analysis of surrogate data sets provided additional evidence of the nonlinear deterministic behavior of these fricatives. Furthermore, the D_2 estimates for the fricatives are found to be consistent with the Lyapunov dimension estimates (Farmer *et al.*, 1983). The maximum Lyapunov exponent values for the vowels, on the other hand, were close to zero indicating that the underlying behavior is likely to be nonchaotic.

Numerical limitations of chaos analysis techniques when dealing with experimental data were discussed in this paper. Care was taken to minimize effects of systematic errors such as experimental noise and autocorrelation effects. Nevertheless, fundamental limitations posed by finite data lengths could not be handled with the presently available numerical techniques. Due to these limitations, actual values of the attractor dimensions and Lyapunov exponents can not be relied upon and are not the final aim of this study. These values, however, can be used as an indication of the dimensional complexity of the underlying turbulence.

Another limitation of our study is that data from only two subjects were analyzed. Analysis of a larger data set from several subjects is needed to determine the generality of our conclusions. From a purely signal-analysis perspective, it is deemed reasonable to assume that the far-field acoustic speech waveforms contain necessary information, and hence can be analyzed to gain insight into the production process. Our study is based on this assumption. Nevertheless, it is not clear how closely far-field acoustic pressure data represent the actual production mechanisms of fricatives. From a physical point of view, it is important to analyze turbulent flow and pressure measured *inside* the vocal tract near the constriction. The option of using inverse-filtered data was not pursued in this study due to unknown exact source locations (and characteristics) for fricatives. In the future, we hope to measure and analyze turbulence parameters inside the vocal tract.

A natural extension of this study is to analyze the transition regions between the vowels and the fricatives to gain a better understanding of how the flow pattern changes, from a presumably laminar flow (vowel) to a turbulent flow (fricatives). Again, data-length limitations might severely limit the scope of such an attempt since the duration of these transitions is typically in the range of 10–30 ms. Alternatively one can use mechanical models of the vocal tract, such as those described in (Shadle, 1991; Shadle, 1990), vary the Reynolds number, and visualize the flow behavior. In addition, it would be interesting to use the nonlinear approach to analyze the nonstrident fricatives /f/, /v/, /ð/ and /θ/.

The ultimate objective of this study is to derive analytical chaotic models for fricative production. Results of our analysis have indicated that derivation of generalized models is not possible in a simple manner. Nevertheless, approximate nonlinear models may be derived, particularly for fricatives in vocalic contexts, and used for devising a nonlinear prediction scheme similar to that proposed by Tishby (1990) and Townshend (1992). The adequacy, performance, and practicality of such nonlinear models are yet to be investigated.

ACKNOWLEDGMENTS

The authors thank H. Herzel, A. Löfquist, K. Munhall, and two other anonymous reviewers for insightful comments and suggestions.

Badin, P. (**1989**). "Acoustics of voiceless fricatives: Production theory and data," STL-QPSR 2–3, Apr.–Sept. pp. 45–52.

Baken, R. J. (**1990**). "Irregularity of vocal period: A first approach to the fractal analysis of voice," J. Voice **4**(3), 185–197.

Bickley, C. A., and Stevens, K. N. (**1986**). "Effects of a vocal-tract constriction on the glottal source: experimental and modelling studies," J. Phon. **14**, 373–382.

Brandstater, A., and Swinney, H. L. (**1987**). "Strange attractors in weakly turbulent Couette–Taylor flow," Phys. Rev. A **35**(5), 2207–2220.

Caputo, J. G., Malraison, B., and Atten, P. (**1989**). "Determination of attractor dimension and entropy for various flows: An experimentalist's viewpoint," in *Dimensions and Entropies in Chaotic Systems*, edited by G. Mayer-Kress, Springer Series in Synergetics Vol. 32 (Springer-Verlag, Berlin), 2nd ed., pp. 180–190.

Dwoyer, D. L., Hussaini, M. Y., and Voigt, R. G. (**1985**). *Theoretical Approaches in Turbulence*, Vol. 58 of *Applied Mathematical Sciences* (Springer-Verlag, New York).

Eckmann, J.-P., and Ruelle, D. (**1992**). "Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems," Physica D **56**, 185–187.

ESPS (**1994**). *Entropic Signal Processing System* (Entropic Research Laboratory Inc., Washington, DC).

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Farmer, J. D., Ott, E., and Yorke, J. A. (**1983**). "The dimension of chaotic attractors," Physica D **7**, 153–180.

Fraser, A. M., and Swinney, H. L. (**1986**). "Independent coordinates for strange attractors from mutual information," Phys. Rev. A **33**(2), 1134–1146.

Gober, M., Herzel, H., and Graf, H.-F. (**1992**). "Dimension analysis of El Niño/Southern Oscillation time series," Ann. Geophys. **10**, 729–734.

Goldberg, J. (**1993**). "Using the correlation dimension code: Grass," UCSD, La Jolla, CA.

G. Pfister, Buzug, T., and Enge, N. (**1992**). "Characterization of experimental time series from Taylor-Couette flow," Physica D **58**, 441–454.

Grassberger, P. (**1986**). "Do climatic attractors exist?," Nature **323**, 609.

Grassberger, P., and Procaccia, I. (**1983**). "Characterization of strange attractors," Phy. Rev. Lett. **50**(5), 346–349.

Hardcastle, W. J., and Clark, J. E. (**1981**). "Articulatory, aerodynamic and acoustic properties of lingual fricatives," Tech. Rep. 5, Speech Research Laboratory work in progress, University of Reading, pp. 51–78.

Haykin, S., and Leung, H. (**1992**). "Model reconstruction of chaotic dynamics: first preliminary radar results," in *IEEE Proc. ICASSP* (San Francisco, CA), Vol. 4, pp. IV–125–128.

Helleman, R. H. G. (**1986**). "Chaotic and turbulent behavior of simple nonlinear systems," in *Frontiers in Physical Acoustics*, edited by D. Sette, Vol. XCIII of *Proceedings of the International School of Physics Enrico Fermi* (North-Holland, Amsterdam), pp. 102–114.

Herzel, H. (**1993**). "Bifurcations and chaos in voice signals," Appl. Mech. Rev. **46**(7), 399–413.

Herzel, H., Berry, D., Titze, I., and Saleh, M. (**1994**). "Analysis of vocal disorders with methods from nonlinear dynamics," J. Speech Hear. Res. **37**, 1008–1019.

Hixon, T. (**1966**). "Turbulent noise sources for speech," Folia Phoniatr. **18**(3), 168–182.

Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E. L., and Kay, B. (**1985**). "A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics and, dynamic modeling," J. Acoust. Soc. Am. **77**, 266–280.

Kennel, M. (**1993**). "The multiple-dimensions mutual information program," Inst. for Nonlinear Science, UCSD, La Jolla, CA.

Kruel, T., and Eiswirth, M. (**1991**). "LCE-EXP: A program for the calculation of the complete spectrum of Lyapunov exponents from a time series of experimental data," Wurzburg, Germany.

Kruel, T.-M., Eiswirth, M., and Schneider, F. W. (**1993**). "Computation of Lyapunov spectra: Effect of interactive noise and application to a chemical oscillator," Physica D **63**, 117–137.

Kurths, J., and Herzel, H. (**1987**). "An attractor in a solar time series," Physica D **25**, 165–172.

Lauterborn, W., and Holzfuss, J. (**1991**). "Acoustic chaos," Int. J. Bifurcation Chaos **1**(1), 13–26.

Manley, O. (**1984**). "Finite dimensional aspects of turbulent flows," in *Chaos in Nonlinear Dynamical Systems*, edited by J. Chandra (SIAM, City), pp. 165–175.

Maragos, P. (**1991**). "Fractal aspects of speech signals: Dimension and interpolation," in *IEEE Proc. ICASSP* (IEEE Signal Processing Society, Ontario, Canada), Vol. 1, pp. 417–420.

Mayer-Kress, G. (**1985**). *Dimensions and Entropies in Chaotic Systems*, Springer Series in Synergetics Vol. 32 (Springer-Verlag, Berlin), 1989 ed.

Mayer-Kress, G. (**1987**). "Application of dimension algorithms to experimental chaos," in *Directions in Chaos*, edited by H. Bai-lin (World Scientific, Teaneck, NJ), Vol. 1, pp. 122–135.

McLaughlin, S., and Lowry, A. (**1993**). "Nonlinear dynamical systems concepts in speech analysis," in *Proceedings of EUROSPEECH'93*, Vol. 1 of *3rd European Conf. on Speech Commn. and Tech.* (ESCA, Berlin, Germany), pp. 377–380.

Mende, W., Herzel, H., and Wermke, K. (**1990**). "Bifurcations and chaos in newborn infant cries," Phys. Lett. A **145**(8,9) 418–424.

Narayanan, S., Alwan, A., and Haker, K. (**1994**). "An MRI study of fricative consonants," in *Proceedings of ICSLP* (Yokohoma, Japan), Vol. 2, pp. 627–630.

Pickover, C., and Khorasani, A. (**1986**). "Fractal characterization of speech waveform graphs," Comput. Graphics **10**(1), 51–61.

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

Sano, M., and Sawada, Y. (**1985**). "Measurement of the Lyapunov spectrum from a chaotic time series," Phys. Rev. Lett. **55**, 1082–1085.

Schroeter, J., and Sondhi, M. M. (**1992**). "Speech coding based on physiological models of speech production," in *Advances in Speech Signal Processing*, edited by S. Furui and M. M. Sondhi (Dekker, New York), pp. 231–268.

SenSyn (**1993**). "SenSyn v1.1," Sensimetrics Corp., Cambridge, MA.

Shadle, C. H. (**1985**). "The Acoustics of Fricative Consonants," Tech. rep. 506, MIT, Cambridge, MA.

Shadle, C. H. (**1990**). "Articulatory-acoustic relationships in fricative consonants," in *Speech Production and Speech Modelling*, edited by W. Hardcastle and A. Marchal (Kluwer Academic, New York), pp. 187–209.

Shadle, C. H. (**1991**). "The effect of geometry on source mechanisms of fricative consonants," J. Phonetics **19**, 409–424.

Sidorowich, J. J. (**1992**). "Modeling of chaotic time series for predictions, interpolation and smoothing," in *IEEE Proc. ICASSP* (San Francisco, CA), Vol. 4, pp. IV–121–124.

Sieber, M. (**1987**). "Experiments on the attractor-dimension for turbulent pipe flow," Phys. Lett. A **122**(9), 467–470.

Stevens, K. N. (**1971**). "Airflow and turbulence noise for fricative and stop consonants: Static considerations," J. Acoust. Soc. Am. **50** (P.2), 1180–1192.

Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., and Kurowski, K. (**1992**). "Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters," J. Acoust. Soc. Am. **91**(5), 2979–3000.

Stone, M., Faber, A., Raphael, L. J., and Shawker, T. H. (**1992**). "Cross-sectional tongue shapes and linguopalatal contact patterns in [s], [S], and [l]," J. Phon. **20**(2), 253–270.

Stoop, R., and Meier, P. F. (**1988**). "Evaluation of Lyapunov exponents and scaling functions from time series," J. Opt. Soc. Am. B **5**(5), 1037–1045.

Streeter, V. L. (**1962**). *Fluid Mechanics* (McGraw-Hill, New York).

Subtelny, J. D., Oya, N., and Subtelny, J. D. (**1972**). "Cineradiographic study of sibilants," Folia Phoniatr. **24**, 30–50.

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

Tatsumi, T. (**1984**). *Turbulence and Chaotic Phenomena in Fluids*. The International Union of Theoretical and Applied Mechanics (Elsevier, North-Holland).

Theiler, J. (**1987**). "Efficient algorithm for estimating the correlation dimension from a set of discrete points," Phys. Rev. A **36**, 4456–4462.

Theiler, J. (**1990**). "Estimating fractal dimension," J. Opt. Soc. Am. A **7**(6), 1055–1073.

Theiler, J. (**1991**). "Some comments on the correlation dimension of $1/f^{\alpha}$ noise," Phys. Lett. **155**(8,9), 480–493.

Tishby, N. (**1990**). "A dynamical systems approach to speech processing," in IEEE Proc. ICASSP **4**, 365–368.

Titze, I. R., Baken, R. J., and Herzel, H. (**1993**). "Evidence of chaos in vocal fold vibration," in *Vocal Fold Physiology: New Frontiers in Basic Science*, edited by I. R. Titze (Singular Publishing Group, San Diego), Chap. 4, pp. 143–188.

Townshend, B. (**1992**). "Nonlinear prediction of speech signals," in *Nonlinear Modelling and Forecasting*, edited by M. Casdagli and S. Eubank, SFI Studies in the Sciences of Complexity, Vol. XII (Addison-Wesley, New York), pp. 433–449.

Warren, D. W., Hall, D. J., and Davis, J. (**1981**). "Oral port constriction and pressure-airflow relationships during sibilant productions," Folia Phoniatr. **33**, 380–394.

Wolf, A., Swift, J., Swinney, H., and Vastano, J. (**1985**). "Determining Lyapunov exponents from a time series," Physica D **16**, 285–317.