

## Research Article

# Applied Chaos Level Test for Validation of Signal Conditions Underlying Optimal Performance of Voice Classification Methods

Boquan Liu,<sup>a</sup> Evan Polce,<sup>a</sup> Julien C. Sprott,<sup>b</sup> and Jack J. Jiang<sup>a</sup>

**Purpose:** The purpose of this study is to introduce a chaos level test to evaluate linear and nonlinear voice type classification method performances under varying signal chaos conditions without subjective impression.

**Study Design:** Voice signals were constructed with differing degrees of noise to model signal chaos. Within each noise power, 100 Monte Carlo experiments were applied to analyze the output of jitter, shimmer, correlation dimension, and spectrum convergence ratio. The computational output of the 4 classifiers was then plotted against signal chaos level to investigate the performance of these acoustic analysis methods under varying degrees of signal chaos.

**Method:** A diffusive behavior detection–based chaos level test was used to investigate the performances of different voice classification methods. Voice signals were constructed

by varying the signal-to-noise ratio to establish differing signal chaos conditions.

**Results:** Chaos level increased sigmoidally with increasing noise power. Jitter and shimmer performed optimally when the chaos level was less than or equal to 0.01, whereas correlation dimension was capable of analyzing signals with chaos levels of less than or equal to 0.0179. Spectrum convergence ratio demonstrated proficiency in analyzing voice signals with all chaos levels investigated in this study.

**Conclusion:** The results of this study corroborate the performance relationships observed in previous studies and, therefore, demonstrate the validity of the validation test method. The presented chaos level validation test could be broadly utilized to evaluate acoustic analysis methods and establish the most appropriate methodology for objective voice analysis in clinical practice.

Voice disorders can impart significant functional and psychological limitations on the lives of patients (Little, McSharry, Roberts, Costello, & Moroz, 2007). A comprehensive understanding of the acoustic basis of voice disorders becomes important to effectively diagnose and treat patients. To better characterize voice signals, Titze (1995) developed a classification scheme that designated voice signals into three signal types according to the nonlinearity present in the signal. Type 1 signals are predominantly periodic, Type 2 signals contain modulations and subharmonics, whereas Type 3 signals are aperiodic in nature. This classification scheme was modified by Sprecher,

Olszewski, Zhang, and Jiang (2010) to incorporate a fourth voice type, which primarily exhibits stochastic noise characteristics.

The ability to use acoustical analysis to differentiate between voice types represents both a noninvasive and objective means through which clinicians can obtain crucial information regarding normal and disordered phonation. In a traditional model of voice production, vocal fold collisions during phonation generate vibratory motions that produce quasiperiodic excitation signals in the vocal tract (Rabiner & Schafer, 1978). In this voice production model, voice type signals 1, 2, and 3 represent the low-dimensional vibratory system and are generated by periodic and quasiperiodic vibration patterns stemming from the vocal folds. On the contrary, Type 4 voice signals are produced by chaotic vocal fold collisions, nonlinear stress–strain tissue interactions, and infinite-dimensional turbulent airflow through the vocal tract (Jiang, Zhang, & McGilligan, 2006; Zhang, Jiang, Biazzo, & Jorgensen, 2005; Zhang, McGilligan, Zhou, Vig, & Jiang, 2004). The functional and morphological

<sup>a</sup>Department of Surgery-Division of Otolaryngology, University of Wisconsin School of Medicine and Public Health, Madison

<sup>b</sup>Department of Physics, University of Wisconsin-Madison

Correspondence to Jack Jiang: [jjjiang@wisc.edu](mailto:jjjiang@wisc.edu)

Editor-in-Chief: Krista Wilkinson

Editor: Preeti Sivasankar

Received June 27, 2017

Revision received November 20, 2017

Accepted January 30, 2018

[https://doi.org/10.1044/2018\\_JSLHR-S-17-0250](https://doi.org/10.1044/2018_JSLHR-S-17-0250)

**Disclosure:** The authors have declared that no competing interests existed at the time of publication.

complications present in vocal pathologies are manifested as Types 3 and 4 voice signals. Differentiation between Types 3 and 4 voice signals may provide further insight into the underlying functional or morphological changes simultaneously occurring during the transition from low-dimensional vibratory dynamics to infinite-dimensional biomechanical and turbulent flow dynamics in the vocal tract. In addition, periodically monitoring progression of the voice type profile during the course of treatment would allow clinicians to quantitatively assess patient voice improvement and compare the efficacies of different treatment interventions.

Currently, various linear and nonlinear methods are utilized in acoustical analysis and for classification of voices into their corresponding categories. Linear parameter-based perturbation analyses, such as jitter and shimmer, are calculated based on the fundamental frequency and peak amplitude of each phonatory cycle (Sprecher et al., 2010). However, low degrees of chaos inherent in irregular phonation diminish the ability of jitter and shimmer to produce stable estimates. Thus, previous studies have suggested jitter and shimmer should only be applied for analysis of nearly periodic Type 1 voice signals (Jiang et al., 2006).

The establishment of chaotic behavior in human phonation required application of nonlinear dynamics to analyze complex, nonlinear voice signals. Correlation dimension (D2) represents the number of degrees of freedom required to describe the complexity of a system and is useful for differentiating between periodic and irregular phonations (Jiang, Zhang, & Ford, 2003; Jiang et al., 2006). Although the D2 converges to a finite value with increasing embedding dimension when analyzing periodic and slightly chaotic signals, the D2 does not converge to a finite value when the chaos component of the signal becomes more extensive. Under these high chaos conditions, the signal exhibits infinite dimensionality, and therefore, accurate estimates of D2 for Type 4 voice signals become impossible. Spectrum convergence ratio (SCR) uses short-time Fourier transform (STFT) to evaluate the convergence of 250 generated segments for each voice signal (Lin, Calawerts, Dodd, & Jiang, 2016). SCR is sensitive to small variations in the periodicity of voice signals and is theorized to be capable of objectively classifying all four voice signal types.

The efficacy of these classifiers can vary tremendously when analyzing voice signals containing differing degrees of nonlinearity and chaos, causing them to produce invalid results under certain conditions (Calawerts, Lin, Sprott, & Jiang, 2017; Jiang et al., 2003, 2006; Lin et al., 2016; Sprecher et al., 2010). However, the performance and reason of invalidation for different classification methods are not well documented, limiting their practical application and reliability. In previous studies, the accuracy of classification methods was analyzed by comparing the voice type obtained from the linear or nonlinear parameter to the voice type assigned subjectively by a researcher in spectrogram analysis. The subjective nature of spectrogram analysis introduces classification errors and, therefore, decreases

confidence in the validity of this comparison method. Subjective perception is currently widely utilized in voice research and clinical practice and compared to the results of acoustical analysis methods; however, confounding factors, such as medical training, experience, and varying definitions of pathological phonation, generate variability among and between subjective raters (Gupta et al., 2016). To eliminate the variability caused by human subjective impression, we propose a validation test that directly relates the degree of noise present in the voice signal to the chaos level (CL), the degree of aperiodic or random behavior in a signal, to objectively determine the voice signal conditions under which classification methods perform optimally. We applied the validation test to determine how the degree of chaos in a signal, due to noise, affects the performance of various linear and nonlinear parameters.

Previous studies have demonstrated that the addition of external noise can induce spontaneous chaos. In the field of signal processing, noise-induced chaos has been observed in various models, such as the semiconductor superlattice and the Rössler model system (Kawata, Horita, Terachi, & Ogata, 1995; Yin et al., 2017). In addition, previous research in nonlinear dynamic analysis of voice production has demonstrated that the addition of a turbulent noise component to a two-mass model resulted in chaotic vocal fold vibrations (Jiang & Zhang, 2002). However, no research to our knowledge has investigated if the addition of varying levels of external noise to artificially constructed voice signals can model signal chaos.

In practice, differing noise powers in a voice signal lead to various degrees of chaos, resulting in different voice type signals. Under different noise powers, this article employs a diffusive behavior detection-based CL test to investigate the signal conditions that voice classification methods perform optimally under (Gottwald & Melbourne, 2005, 2009). We defined high chaos as a CL greater than or equal to 0.9, intermediate chaos as a CL between 0.1 and 0.9, low chaos as a CL between 0.01 and 0.1, and very low chaos as a CL less than or equal to 0.01. The applicability of the proposed method was investigated through performance analysis of the SCR, D2, and perturbation classification methods. We chose to only analyze SCR, D2, jitter, and shimmer because previous studies have thoroughly tested the voice types and signal conditions that these four parameters perform most optimally under. Performance analysis involved analyzing the computational output of SCR, D2, jitter, and shimmer at varying CLs from 0 to 1. A linear relationship between the acoustic analysis outputs obtained and CL would indicate that the specific acoustic analysis method under investigation can effectively quantify voice signals that widely vary from periodic to predominantly chaotic in nature.

Similar to the classification performance profiles observed in previous studies, we hypothesized that jitter and shimmer would only produce reliable estimates when the signal was predominantly periodic, which is when the CL is very low. Second, we hypothesized that D2 would perform optimally when analyzing signals with low CLs. Third, we

hypothesized that SCR would produce stable acoustic estimates under all signal CLs investigated in this study.

## Method

### Perturbation Analysis

Jitter is defined as the average absolute difference between consecutive periods, divided by the average period, expressed as

$$\text{Jitter} = \frac{\frac{1}{N-1} \sum_{i=1}^N |T_i - T_{i+1}|}{\frac{1}{N} \sum_{i=1}^N T_i} \quad (1)$$

where  $T_i$  are the extracted fundamental frequency period lengths and  $N$  is the number of extracted fundamental frequency periods (Brockman, Drinnan, Storck, & Carding, 2011; Sprecher et al., 2010).

Shimmer is defined as the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude, expressed as

$$\text{Shimmer} = \frac{\frac{1}{N-1} \sum_{i=1}^N |A_i - A_{i+1}|}{\frac{1}{N} \sum_{i=1}^N A_i} \quad (2)$$

where  $A_i$  are the extracted peak-to-peak amplitude data (Brockman et al., 2011; Sprecher et al., 2010).

### D2

The D2 describes the complex dynamic behaviors of a system. More complex systems might necessitate more state variables, whereas simpler systems may need fewer degrees of freedom to describe the system dynamics (Awan, Novaleski, & Rousseau, 2014; Lin et al., 2016; Sprecher et al., 2010).

A time series with length  $L$  is measured and recorded as  $x(n_1), x(n_2), x(n_3), \dots$ ; the D2 is calculated using the dimensional phase space (Packard, Crutchfield, Farmer, & Shaw, 1980):

$$X_i = \{x(n_i), x(n_i - \tau), \dots, x(n_i - (e-1)\tau)\} \quad (3)$$

where  $x(n_i)$  is the  $n_i$ th signal data,  $\tau$  is the delay, and  $e$  is the embedding dimension. The mutual information method is used to determine the appropriate time delay (Fraser & Swinney, 1986).

The correlation integral  $C(L, r)$  is represented by

$$C(L, r) = \frac{1}{L(L-1)} \sum_{i=1}^L \sum_{\substack{j=1 \\ i \neq j}}^L H(r - \|X_i - X_j\|) \quad (4)$$

where  $r$  is the radius around  $X_i$ ,  $H$  stands for Heaviside function, and the correlation integral  $C(L, r)$  is the probability

that the distance between two vectors on the attractor is smaller than a radius ( $r$ ). If the value of  $r$  is too small, random noise becomes a dominant factor and causes estimates of D2 to increase continuously with increasing embedding dimension ( $e$ ) (Awan, Roy, & Jiang, 2010). If the value of  $r$  is too large (i.e., approximately the size of the reconstructed phase space), estimations of D2 approach zero because all data pairs of interest in reconstructed phase space are smaller than  $r$ . The appropriate range of  $r$ , designated as the scaling region, lies between these two extremes and can be determined manually.

$$H(u) = \begin{cases} 1, & u > 0 \\ 0, & u \leq 0 \end{cases} \quad (5)$$

The D2 is defined as

$$D_2 = \lim_{r \rightarrow 0} \lim_{L \rightarrow \infty} \frac{\log C(L, r)}{\log r}. \quad (6)$$

The D2 was found by calculating the slope of the most linear part of the  $\log C(L, r)$  versus  $\log r$  plot (Fraser & Swinney, 1986; Grassberger & Procaccia, 1983).

### SCR

SCR was calculated based on applying the STFT to each voice signal sample. The discrete STFT is defined by

$$V(\omega, m) = \sum_{n=-\infty}^{\infty} y(n)g(n-m)e^{-j\omega n} \quad (7)$$

where  $y(n)$  is the time series;  $m$  is the number of segments,  $m = 1, 2, \dots, M$ ;  $g(n-m)$  is the window function; and  $\omega$  corresponds to the frequency. Briefly, discrete STFT is used to analyze a series of discrete segments that comprise a signal and compare these discrete segments to determine changes in frequency in the signal over time. To partition the time sequence into segments, a windowing function moves along the time axis and obtains local time segments. The window size, which dictates the number of sampled points, was set to 0.012 s, generating 250 segments for each voice sample. Fourier transformations were obtained for the 250 segments produced for each voice signal. A variable called the dynamic range of segments' spectrogram (DRSS) was then defined to quantify the variation in frequency between the different segments.

The DRSS was calculated by the following (Lin et al., 2016):

$$\text{DRSS} = \sum_{m=1}^M [C_{\max}(m) - C_{\min}(m)] \quad (8)$$

where  $C_{\max}(m)$  is the maximum energy curve and  $C_{\min}(m)$  is the minimum energy curve expression in the

segment  $m$ ,  $m = 1, 2, \dots, M$ , providing the difference between the maximum and minimum coefficient values of all segments. Next, a variable named the maximum energy (MAE) is defined as

$$\text{MAE} = \sum_{m=1}^M C_{\max}(m). \quad (9)$$

Finally, the SCR value was found using the formula

$$\text{SCR} = -\ln\left(\frac{\text{DRSS}}{\text{MAE}}\right). \quad (10)$$

Aperiodic or chaotic voice will have an SCR value that approaches 0, whereas the SCR value increases gradually as voice signals become more periodic. This trend in SCR values is due to the spectrum of a periodic signal being composed of extremely similar segments. Therefore, an aperiodic signal would be composed of segments of considerable dissimilarity. Accordingly, SCR decreases with increasing voice type.

### Validation Test

The voice signal  $x(z)$ ,  $z = 1, 2, \dots, Z$ , is modeled as

$$x(z) = s(z) + n(z) \quad (11)$$

where  $s(z)$  is purely periodic and  $n(z)$  is white, Gaussian-distributed noise (Jiang et al., 2003, 2006; Titze, 1995). A signal ( $s(z)$ ) amplitude of 0.6, sampling frequency of 25 kHz, and signal frequency of 180 Hz are employed in this study. The length of each constructed voice signal was 20,000 data points.

The  $p_c$  and  $q_c$  are two variable parameters that are constructed to represent the chaotic behavior of a signal:

$$\begin{aligned} p_c(T) &= \sum_{t=1}^T x(t) \cos ct \\ q_c(T) &= \sum_{t=1}^T x(t) \sin ct \end{aligned} \quad (12)$$

where  $c$  is randomly selected from  $\pi/5$  to  $4\pi/5$ , that is,  $c \in [\pi/5, 4\pi/5]$ ,  $T = 1, 2, \dots, Z$ . The choice to restrict the range of  $c$  to  $c \in [\pi/5, 4\pi/5]$  is implemented to avoid resonance distortions of CL around  $\pi$  (Gottwald & Melbourne 2005, 2009). Gottwald & Melbourne (2009) demonstrated that resonance and Brownian motion of  $p_c$  and  $q_c$  occurs when  $c = \pi$ , which would cause CL to increase regardless of if the signal is periodic or chaotic. As shown in Figure 1, the  $p_c$  and  $q_c$  exhibit bound property when the  $x(z)$  is a non-chaotic signal. However, the  $p_c$  and  $q_c$  demonstrate no bound property and behave like Brownian motion if  $x(z)$  is a chaotic signal.

The diffusive behavior of  $p_c$  and  $q_c$  can be investigated by analyzing the  $M_c(v)$ ,

$$M_c(v) = \frac{1}{Z-v} \sum_{k=1}^{Z-v} \left( [p_c(k+v) - p_c(k)]^2 + [q_c(k+v) - q_c(k)]^2 \right) \quad (13)$$

where  $v \ll Z$ . In practice, the  $v$  is chosen to be 1, 2, ...,  $Z/10$ , to optimize accuracy and computational efficiency. Using least square regression, the CL is defined by the asymptotic growth rate,

$$\text{CL} = \lim_{v \rightarrow \infty} \frac{\log M_c(v)}{\log v} \quad (14)$$

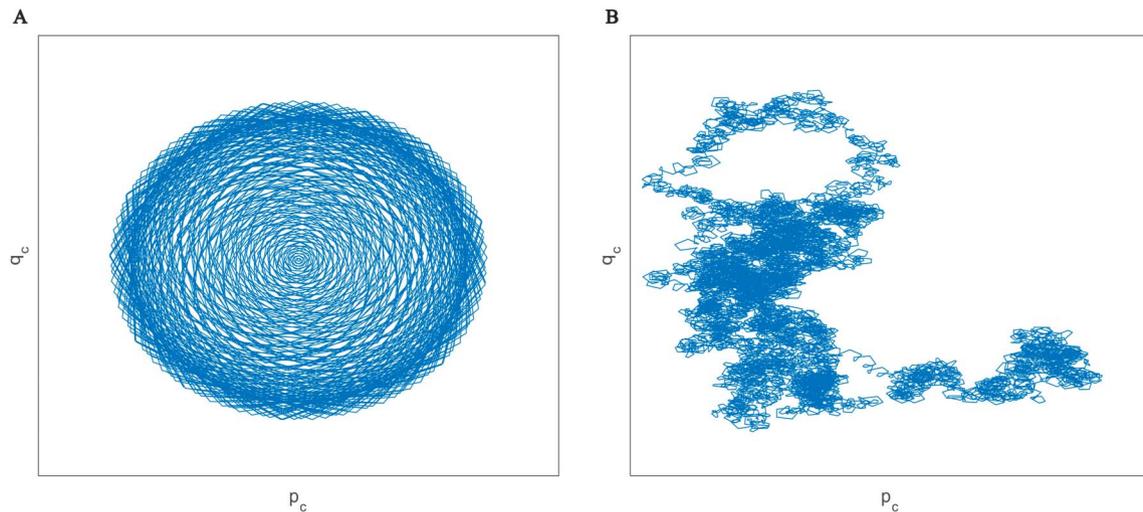
and determined by fitting a straight line to the graph of  $\log M_c(v)$  versus  $\log v$  through minimizing the absolute deviation (Gottwald & Melbourne, 2005, 2009). The periodicity or aperiodicity of a signal can then be ascertained from the slope of the fitted line of the  $\log M_c(v)$  versus  $\log v$  plot (Figure 2). Based on these mathematical equations, a custom MATLAB R2017a (MathWorks, Natick, MA) program was created for CL validation test analysis. A nonlinear signal processing software package named Open TSTool was used for D2 computations.

Constructed voice signals were used in this study instead of observational voice data because the degree of noise in the simulated signals could be controlled artificially. By changing the noise power, we obtained different  $x(z)$ ,  $z = 1, 2, 3, \dots, Z$ . In this article, we used the relationship between noise and CL to obtain differing degrees of signal chaos because it is simpler to directly control signal noise relative to signal chaos. The CL was then applied to represent the chaotic behavior of the constructed signals. A flowchart depicting the computational progression of the validation test is displayed in Figure 3. First, voice signals ( $x(z)$ ) were constructed with differing magnitudes of signal noise ( $n(z)$ ). Then the CL test computation was performed to calculate the CL for every generated  $x(z)$ . Next, the acoustical analysis methods of jitter, shimmer, D2, and SCR were used to evaluate and generate output values for each  $x(z)$ . In addition, 100 Monte Carlo experiments were performed within each noise power, and a completion of the 100 Monte Carlo iterations represents the stop criterion. Thus, once all 100 iterations were completed for a particular  $x(z)$ , the CL and acoustic analysis output values were averaged across the 100 Monte Carlo iterations and then graphically analyzed for linearity.

### Results

Figure 4 displays time domain representations for the constructed voice signals at differing CLs, where A, B, C, and D correspond to the waveforms at a CL of 0.05, 0.2, 0.5, and 0.8, respectively. The waveform fundamental frequency is represented by the inverse of the period,  $f_0 = 1/T$ . The waveforms exhibit increasing amplitudes as the signal CL increases. A signal amplitude of 0.6 was used for

**Figure 1.** Plot of  $p$  versus  $q$  for a (A) periodic signal and (B) chaotic signal.



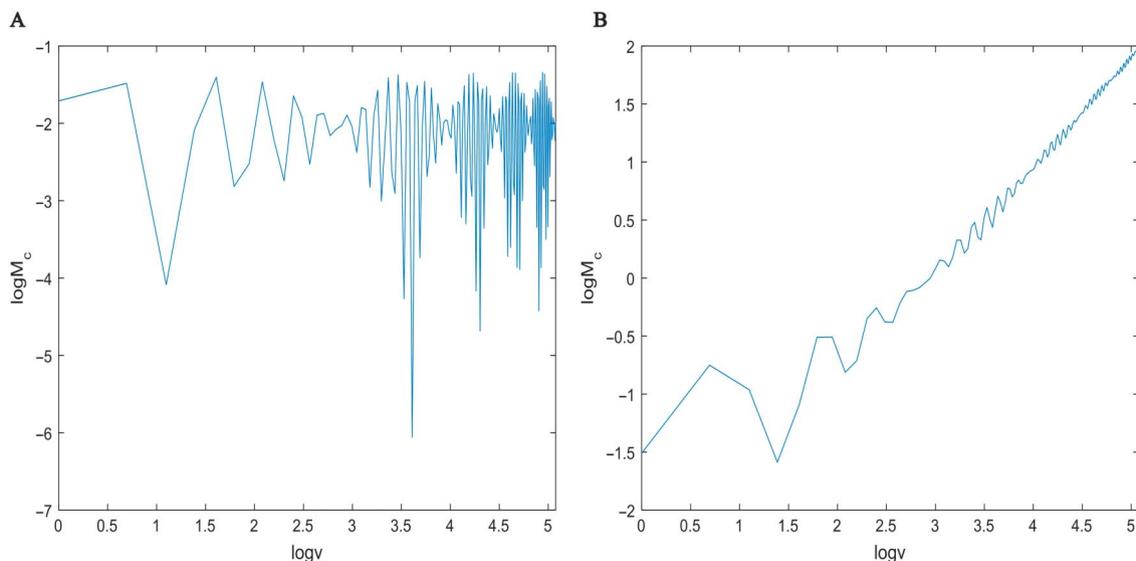
the construction of every signal; however, as the noise component ( $n(z)$ ) of the signal increases in the higher CLs, the waveform amplitudes are characterized by larger magnitudes. In addition, as the CL increases from 0.05 to 0.8, the phonatory cycles of the simulated waveforms transition from periodic to aperiodic.

As shown in Figure 5, the CL increases with increasing noise power. In other words, CL and signal-to-noise ratio are inversely related. The CL varying from approximately 0 to 1 corresponds to the signal dynamics ranging from periodic to predominantly chaotic, which coincides with the voice type definition. Consequently, an increasing CL would correlate with increasing voice type. Figure 5 was divided into four different magnitudes of chaos:

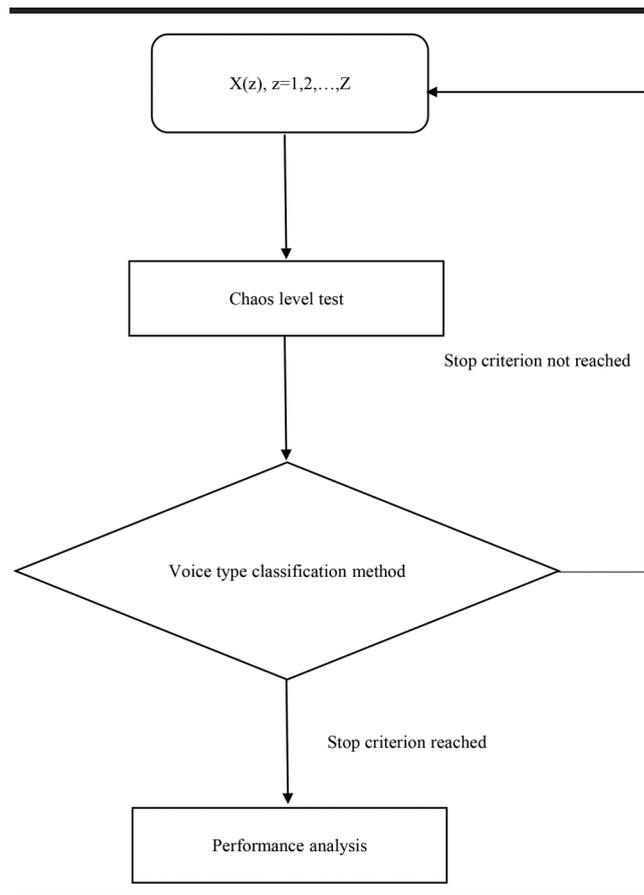
high chaos is defined as a CL greater than or equal to 0.9, intermediate chaos as the CL interval of 0.1–0.9, low chaos as the CL interval of 0.01–0.1, and very low chaos as a CL less than or equal to 0.01. As the CL approaches 1 the signal noise becomes increasingly larger, indicating that the chaos component dominates the signal.

Figure 6 shows the mean value of 100 Monte Carlo experiments performed within each CL for jitter and shimmer. Jitter and shimmer both exhibited nonmonotonic behavior with increasing CL. Overlapping of jitter and shimmer values are observed between CL values of 0.35 and 0.9. This indicates that different signal CLs correspond to identical values of jitter and shimmer. In addition, shimmer exhibited a more linear relationship with CL at very

**Figure 2.** Plot of  $\log M_c(v)$  as a function of  $\log v$  for a (A) periodic signal and (B) chaotic signal.



**Figure 3.** Flowchart demonstrating the validation test process for analysis of the performance of voice classification methods.



low CLs relative to jitter. Specifically, when the CL was smaller than 0.01, shimmer decreased linearly with decreasing CL. As CL increased beyond 0.01, shimmer no longer changed purely linearly with CL.

The D2 increased linearly with CL when the CL was smaller than 0.0179 (Figure 7). As the CL value further increased, D2 frequently resulted in an infinite value. The probability of D2 analysis producing an infinite result was positively correlated with increasing CL value. Therefore, because D2 did not converge at higher CLs, the graph only displays D2 values up to a CL of 0.0179.

The SCR monotonically decreased with increasing CL (Figure 8A). As displayed in Figure 8B, the standard deviations of SCR in the high CLs are smaller relative to the low CLs. The standard deviations of SCR were smaller than 0.002 when CL was larger than 0.1 and smaller than 0.0015 as CL rose up to 0.9. Table 1 displays the signal CL conditions under which the different classification methods analyzed performed optimally.

## Discussion

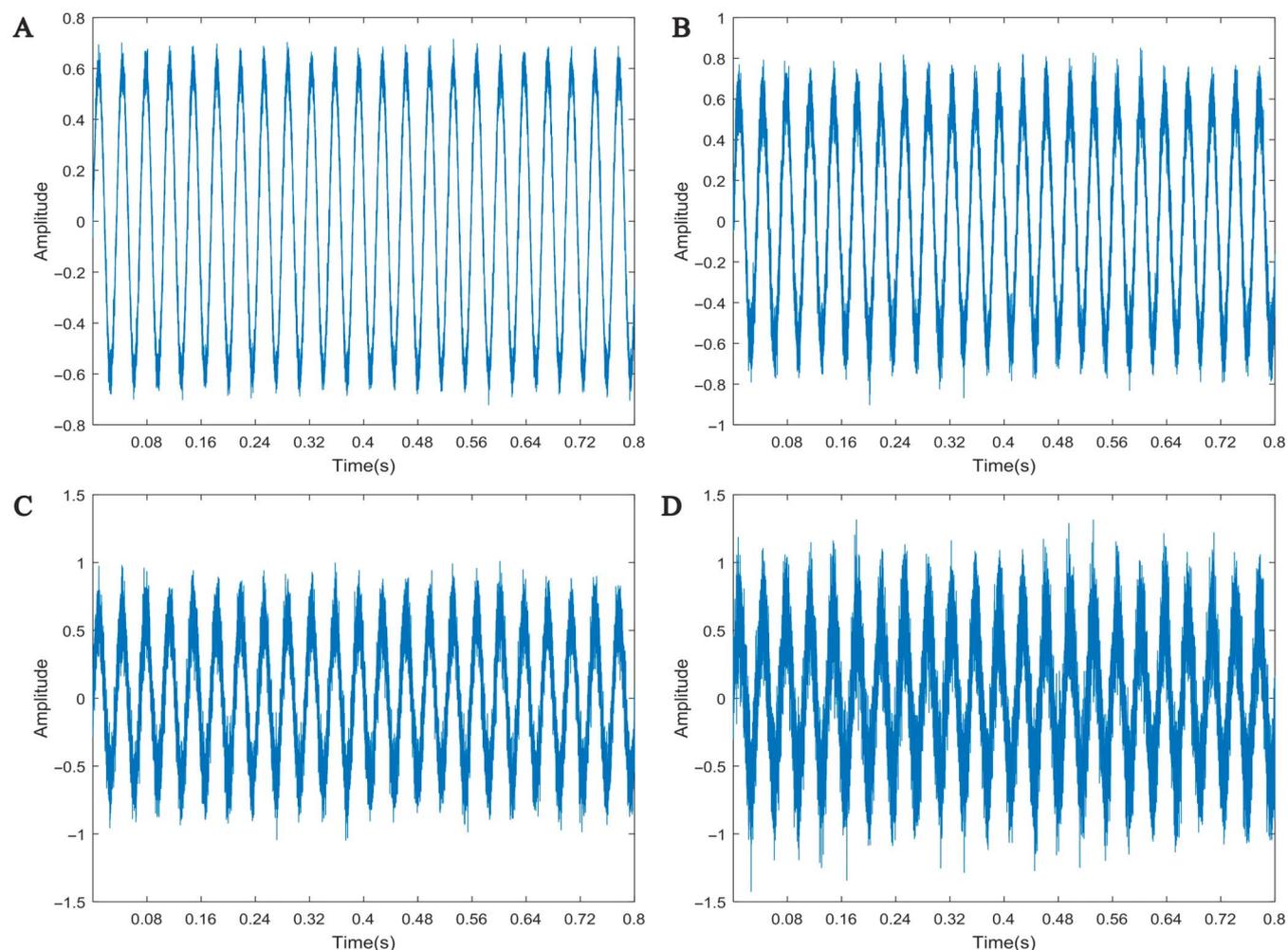
In this study, we proposed a method capable of objectively evaluating the performance and optimum signal

conditions for linear and nonlinear voice type classification methods. The validation test calculation utilizes a diffusive behavior detection-based CL test to determine the signal CL under which these classification methods become unreliable. One hundred Monte Carlo experiments were performed, and subsequently, mean values at each CL were calculated. We applied the validation test to analyze the performances of jitter, shimmer, D2, and SCR, across varying degrees of chaos in constructed voice signals.

The results of our study corroborated the observations and findings of previous studies. Analysis revealed that jitter and shimmer were only suitable for quantifying voice signals with very low degrees of nonlinearity. Shimmer performed slightly better than jitter; however, in practice, this difference is negligible. Sharp increases in the values of jitter and shimmer were observed from CLs of 0.3 to 0.5. These sharp increases in output values are due to the transition from a primarily periodic signal at a CL of 0.3 to a signal with considerable noise characteristics at a CL of 0.5. As the CL is increased further to 0.6, jitter and shimmer begin to decrease sharply because estimations of fundamental frequency and cycle periods become erroneous when noise becomes predominant in voice signals (Jiang et al., 2003; Little et al., 2007). This is consistent with previous studies that suggested jitter and shimmer should only be applied for analysis of Type 1 and some Type 2 voices (Jiang et al., 2003, 2006; Sprecher et al., 2010), which exhibit very low, if any, signal chaos. In this study, standard deviations were not reported for jitter, shimmer, or D2 because decreased classification performances and increased fluctuations from mean values as signal dynamics become increasingly aperiodic for these methods have been thoroughly investigated and observed previously in the literature. Consequently, as displayed in Figure 3, observed overlapping of jitter and shimmer values at differing CL values suggest that jitter and shimmer are incapable of producing stable estimations when the CL is greater than 0.01. Thus, our hypothesis that jitter and shimmer would only produce reliable estimates when the signal was nearly periodic was confirmed.

When the signal CL was greater than 0.0179, calculations of D2 frequently yielded infinite values and the probability of obtaining infinite values increased with the CL. Under periodic or low-dimensional chaos signal conditions, D2 converges to a finite value with increasing embedding dimension; however, D2 estimates of high-dimensional chaos signals do not converge with increasing embedding dimension (Jiang et al., 2003, 2006). Previous studies have suggested that D2 analysis becomes ambiguous under high-dimensional signal chaos conditions because stochastic noise characteristics dominate the voice signal (Jiang et al., 2003, 2006; Sprecher et al., 2010). Similarly, in this study, D2 was applicable for analysis of signals with very low CLs, as well as some signals with low CLs. Therefore, our second hypothesis that D2 would perform optimally when analyzing low-dimensional signals was confirmed.

**Figure 4.** Example waveforms of constructed voice signals at differing chaos levels (CLs). All signals were constructed with a signal ( $s(z)$ ) amplitude of 0.6, sampling frequency of 25 kHz, signal frequency of 180 Hz, and length of 20,000 data points. (A) CL = 0.05, (B) CL = 0.2, (C) CL = 0.5, (D) CL = 0.8.

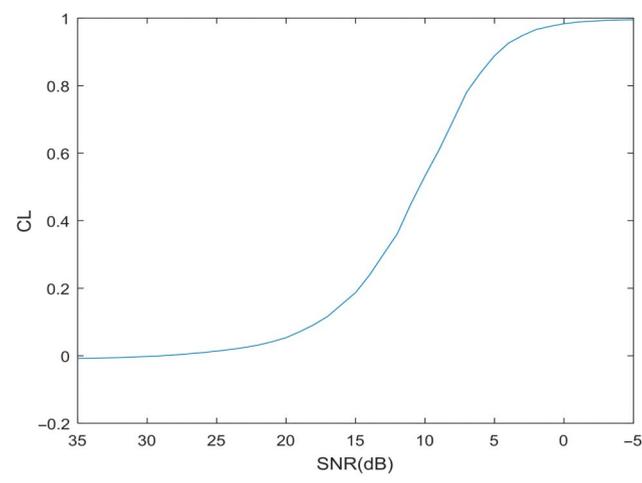


Previous studies have proposed that SCR might be capable of distinguishing between all four types of voices (Lin et al., 2016), which was confirmed by the findings in this study. Our results suggested that SCR analysis is effective in quantifying very low, low-, intermediate-, and high-dimensional chaos (Table 1). SCR values decreased linearly with increasing CL. The linear correlation demonstrates that SCR is applicable for analysis of all CLs, ranging from periodic to entirely chaotic, due to nonoverlapping regions in the SCR results distribution (Figure 8A). In addition, Figure 8B shows that the results of SCR analysis were characterized by higher variability and fluctuations for signals with lower CL values. This is not intuitive, as the classification performances of many objective linear and nonlinear techniques are characterized by greater accuracy for periodic signals, whereas decreased classification accuracy is observed as voice signals transition to chaotic dynamics. The observed difference in the standard

deviations of SCR is related to the differential performance of SCR when analyzing voices with low compared to high CLs. As shown in Lin et al. (2016), the SCR values encompassing Type 1 voices occupy a markedly larger range of SCR values relative to Type 4 voices. Consequently, SCR exhibited better classification performance during analysis of voice signals exhibiting high-dimensional chaos, relative to periodic and low-dimensional voice signals. Because of its proficiency at classifying chaotic voice signals, SCR could be applied to analyze voice disorders in clinical practice, which primarily exhibit Type 3 and Type 4 voice. Thus, our third hypothesis that SCR would generate stable acoustic estimates under all magnitudes of signal chaos analyzed in this study was confirmed.

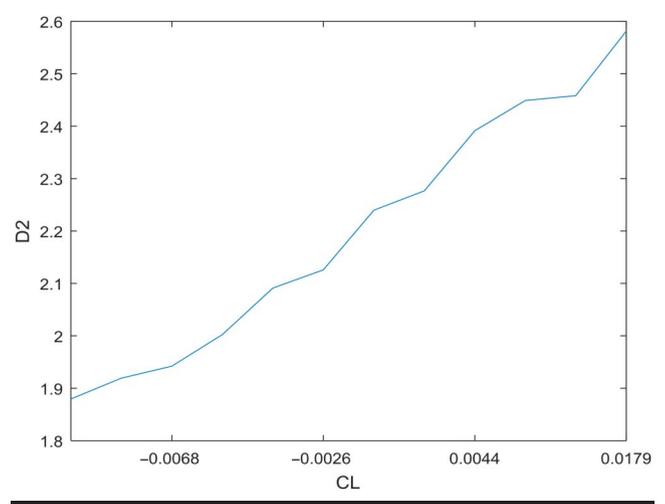
In this study, the addition of white, Gaussian distributed noise to artificially constructed periodic voice signals was used to model observational voice signals, which can exhibit varying degrees of chaos. The relationship between

**Figure 5.** Graph depicting chaos level (CL) varying with noise power. CL is designated on the y axis, and the signal-to-noise ratio (SNR) in decibels (dB) is on the x axis.



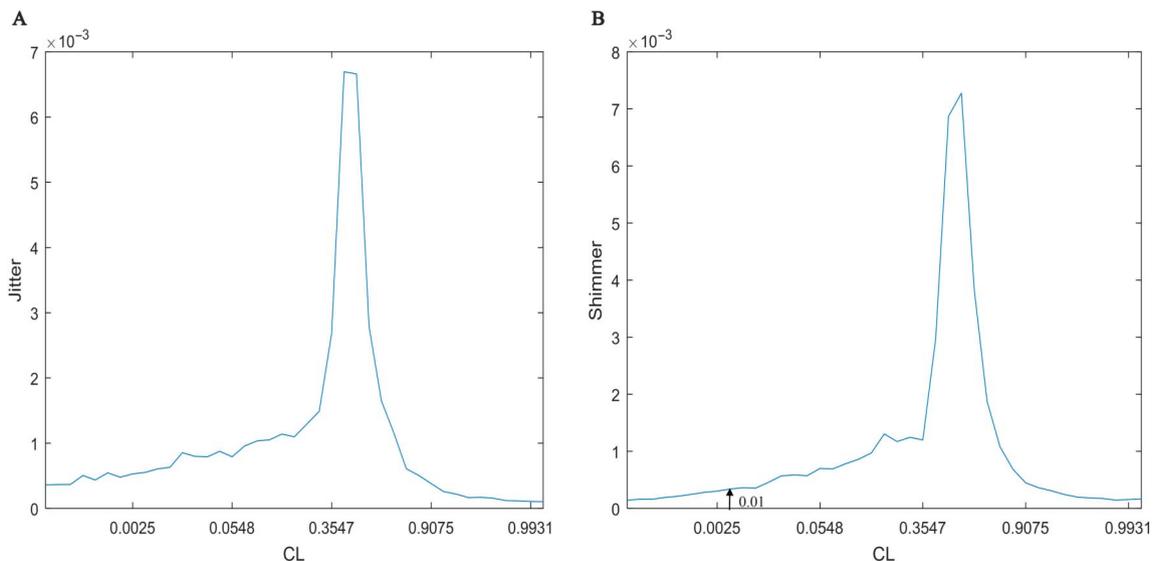
noise and chaos has been widely explored in the field of signal processing, where external noise or interference is commonly applied to induce chaotic dynamics in model systems. Previous research has investigated the effect of noise intensity on the chaotic dynamics of model attractors (Kawata et al., 1995). Results demonstrated that external observation noise added to the system output caused destruction of the trajectories of attractors in phase space, larger maximum Lyapunov exponents relative to the absence of noise, and increasing D2s with decreasing signal-to-noise ratios, indicating that the addition of random noise strongly influenced the chaotic dynamics of the system. Similarly, in

**Figure 7.** Distribution of correlation dimension (D2) values obtained from 100 Monte Carlo experiments performed at differing chaos levels (CLs). D2 did not consistently converge with increasing embedding dimension when the CL was greater than 0.0179.

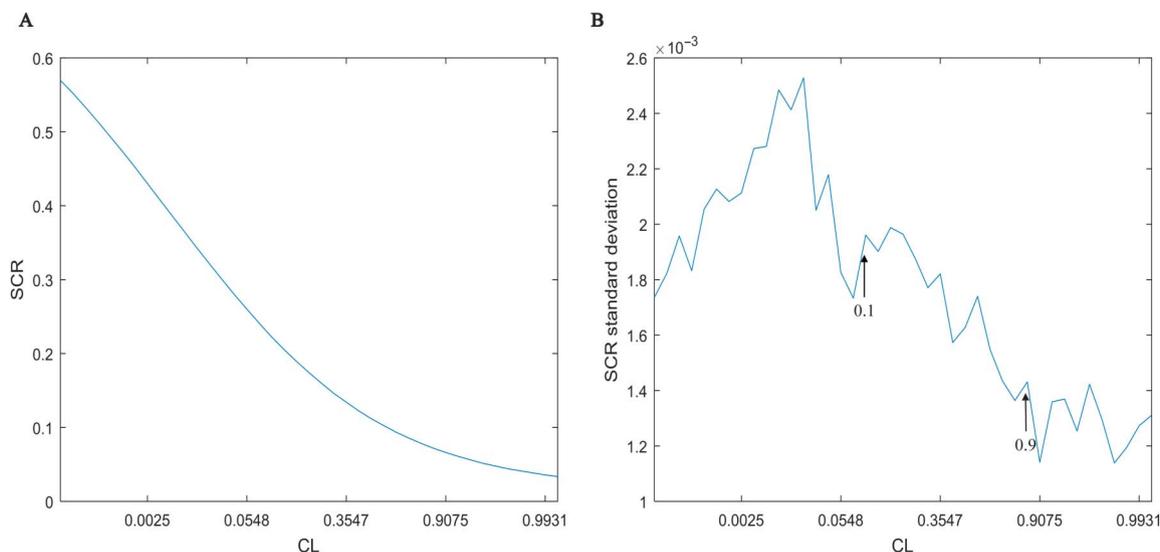


this study, we hypothesized that the addition of external noise was capable of controlling chaos in constructed voice signals. Lin et al. (2016) previously reported that the classification performance of SCR increased with increasing chaos in voice signal recordings from patients. In this study, we demonstrated that the enhanced classification performance of SCR could be explained by observed decreases in SCR standard deviations as the voice signals transitioned to high-dimensional chaos. Replicating the SCR performance relationship described in Lin et al. (2016) provides evidence supporting the validity of the artificial signal construction

**Figure 6.** Distribution of jitter and shimmer values obtained from 100 Monte Carlo experiments performed at chaos levels (CL) varying from periodic (CL = 0) to stochastic (CL = 1): (A) jitter and (B) shimmer.



**Figure 8.** Distribution and standard deviations of spectrum convergence ratio (SCR) values obtained from 100 Monte Carlo experiments performed at chaos levels (CLs) varying from periodic (CL = 0) to stochastic (CL = 1): (A) distribution and (B) standard deviation.



paradigm employed in this study. Although deterministic chaos is not equivalent to stochastic noise, a clear relationship between the addition of external noise and control of system chaos has been observed in various system models, including the signal model used in this study.

The voice signal construction model utilized in this study was not an attempt to model physiological voice production, which would require a more extensive description of turbulent flow sources and dynamics in the vocal tract. Rather, the proposed signal model is intended to simulate the different voice signal outputs that arise from complex biomechanical vocal fold collisions and airflow interactions throughout the vocal tract. That is, the current model is not focused on the source or creation of chaos in the vocal tract during voice production but instead is concerned with the corresponding signal output, which ranges from periodic to chaotic, irrespective of how precisely the signal is generated upstream in the vocal tract.

Currently, many studies utilize perceptual-, linear-, and nonlinear-based methods of acoustic analysis to investigate normal and pathological voices. The performance of existing classification methods can vary significantly

when analyzing voice signals with differing noise components, which might lead to erroneous results. Previous studies have suggested that certain linear and nonlinear analysis methods are unreliable for analyzing Types 3 and 4 voice signals; however, direct quantitative measurements of the degree of chaos in these voice signals that causes a transition to methodological unreliability have not been performed. The uncertainty surrounding the conditions that acoustic analysis methods might produce inaccurate results has limited their clinical application. The current gold standard for voice classification analysis is perceptual evaluation; however, the inherent error of this subjective method necessitates objective analysis as a supplement for clinical practice. The proposed validation test offers an objective means to analyze the performance of voice classification methods, effectively eliminating the subjectivity previously present when classifier efficacy was determined by comparison to the results of spectrogram analysis. Determination of the precise chaos conditions where objective classification methods perform optimally is clinically relevant because effective patient treatment is predicated on the ability of clinicians to obtain accurate diagnostic information. Thus, a more robust understanding of the nature of these validation conditions will enhance the accuracy and utility of these acoustic analysis tools.

**Table 1.** Chaos level application conditions for jitter, shimmer, correlation dimension (D2), and spectrum convergence ratio (SCR).

Method	Jitter	Shimmer	D2	SCR
Chaos level	≤ 0	≤ 0.01	≤ 0.0179	All

*Note.* A chaos level of 0 indicates a completely periodic voice signal, whereas a chaos level of 1 indicates a stochastic voice signal.

## Conclusion

In this study, we applied a CL test to objectively determine the efficacy of linear and nonlinear classification methods under varying signal CLs. Jitter and shimmer were applicable for acoustic analysis of signals with very low degrees of nonlinearity. D2 performed optimally when

analyzing voice signals with low-dimensional chaos. SCR was capable of distinguishing between very low, low-, intermediate-, and high-dimensional chaos. These performance results matched previously reported performance profiles for the classification methods analyzed. Thus, the proposed validation test provided accurate, objective information about the signal chaos conditions underlying optimal classification performance for each acoustical analysis method. Future studies could apply this validation test more broadly to observational voice data to identify the most accurate and efficient voice classification methods for clinical use.

Understanding the conditions under which voice classification methods perform optimally is not only critical for clinical use but also for generating a more robust understanding of the intrinsic signal characteristics of the four voice types. Future work with the validation test could identify additional nonlinear methods proficient in distinguishing Types 3 and 4 voice signals, which might assist in elucidating the mechanisms underlying disordered voices. Performance and validation condition analysis could assist clinicians in obtaining additional diagnostic tools, which may lead to superior treatment for individuals suffering from voice disorders.

## Acknowledgments

This study was supported by the National Institute on Deafness and Other Communication Disorders under award number DC006019 awarded to Dr. Jack Jiang.

## References

- Awan, S. N., Novaleski, C. K., & Rousseau, B. (2014). Nonlinear analyses of elicited modal, raised, and pressed rabbit phonation. *Journal of Voice*, 28, 538–547.
- Awan, S. N., Roy, N., & Jiang, J. J. (2010). Nonlinear dynamic analysis of disordered voice: The relationship between the correlation dimension (D2) and pre-/post-treatment change in perceived dysphonia severity. *Journal of Voice*, 24, 285–293.
- Brockman, M., Drinnan, M. J., Storck, C., & Carding, P. N. (2011). Reliable jitter and shimmer measurements in voice clinics: The relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task. *Journal of Voice*, 25, 44–53.
- Calawerts, W. M., Lin, L., Sprott, J. C., & Jiang, J. J. (2017). Using rate of divergence as an objective measure to differentiate between voice signal types based on the amount of disorder in the signal. *Journal of Voice*, 31, 16–23.
- Fraser, A. M., & Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33, 1134–1140.
- Gottwald, G. A., & Melbourne, I. (2005). Testing for chaos in deterministic systems with noise. *Physica D: Nonlinear Phenomena*, 212, 100–110.
- Gottwald, G. A., & Melbourne, I. (2009). On the implementation of the 0–1 test for chaos. *SIAM Journal on Applied Dynamical Systems*, 8, 129–145.
- Grassberger, P., & Procaccia, I. (1983). Characterization of strange attractors. *Physical Review Letters*, 50, 346–349.
- Gupta, R., Chaspari, T., Kim, J., Kumar, N., Bone, D., & Narayanan, S. (2016). Pathological signal processing: State-of-the-art, current challenges, and future directions. *The Institute of Electrical and Electronics Engineers, Inc. (IEEE) Conference Proceedings, 2016*, 6470–6474.
- Jiang, J. J., & Zhang, Y. (2002). Chaotic vibration induced by turbulent noise in a two-mass model of vocal folds. *The Journal of the Acoustical Society of America*, 112, 2127–2133.
- Jiang, J. J., Zhang, Y., & Ford, C. N. (2003). Nonlinear dynamics of phonations in excised larynx experiments. *The Journal of the Acoustical Society of America*, 114, 2198–2205.
- Jiang, J. J., Zhang, Y., & McGilligan, C. (2006). Chaos in voice, from modeling to measurement. *Journal of Voice*, 20, 2–17.
- Kawata, T., Horita, T., Terachi, S., & Ogata, S. (1995). Control of chaos by system noise and observation noise. *Proceedings of the SICE Annual Conference, 1995*, 1527–1530.
- Lin, L., Calawerts, W., Dodd, K., & Jiang, J. J. (2016). An objective parameter for quantifying the turbulent noise portion of voice signals. *Journal of Voice*, 30, 664–669.
- Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A., & Moroz, I. M. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Biomedical Engineering Online*, 6, 23.
- Packard, N. H., Crutchfield, J. P., Farmer, J. D., & Shaw, R. S. (1980). Geometry from a time series. *Physical Review Letter*, 45, 712–716.
- Rabiner, L. R., & Schafer, R. W. (1978). *Digital Processing of Speech Signals*. Upper Saddle River, NJ: Prentice Hall.
- Sprecher, A., Olszewski, A., Zhang, Y., & Jiang, J. J. (2010). Updating signal typing in voice: Addition of type 4 signals. *The Journal of the Acoustical Society of America*, 127, 3710–3716.
- Titze, I. R. (1995). *Workshop on acoustic voice analysis: Summary statement*. Salt Lake City, UT: National Center for Voice and Speech. Retrieved from <http://www.ncvs.org/freebooks/summary-statement.pdf>
- Yin, Z., Song, H., Zhang, Y., Ruiz-García, M., Carretero, M., Bonilla, L. L., . . . Grahn, H. T. (2017). Noise-enhanced chaos in a weakly coupled GaAs/(Al,Ga)As superlattice. *Physical Review E*, 95(1), 012218.
- Zhang, Y., Jiang, J. J., Biazzo, L., & Jorgensen, M. (2005). Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis. *Journal of Voice*, 19, 519–528.
- Zhang, Y., McGilligan, C., Zhou, L., Vig, M., & Jiang, J. J. (2004). Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps. *The Journal of the Acoustical Society of America*, 115, 2270–2277.