Vibratory Dynamics of Four Types of Excised Larynx Phonations

*Lin Li, †Yu Zhang, §William Calawerts, and §Jack J. Jiang,*†‡Xiamen, Fujian, People’s Republic of China, and §Madison, Wisconsin

Summary: Objectives. There are four types of signals that are typical representations of vocal fold vibratory patterns. Type 1 signals are nearly periodic, type 2 signals contain subharmonic properties, type 3 signals are chaotic, and type 4 signals are characterized as white noise. High-speed imaging allows detailed observation of these vocal fold vibratory patterns. Therefore, high-speed imaging can explore the vibratory mechanism behind each of the four types of signals.

Methods. The glottal area time series of the four types of vocal fold vibrations were calculated from high-speed images of 10 excised canine larynges. Nonlinear dynamic parameters of correlation dimension ($D_2$) and Kolmogorov entropy ($K_2$) were used to quantify the characteristics of the glottal areas and acoustical signals for each voice signal type.

Results. The correlation dimension and Kolmogorov entropy of the glottal areas and acoustical signals for type 1, 2, and 3 voice signals were consistent with the results of previous studies. Interestingly, there was a difference between the glottal area and acoustical signals of type 4 voice signals ($P < 0.001$). Both the correlation dimension and Kolmogorov entropy of the type 4 glottal area were close to 0. In contrast, the type 4 acoustical signals had an infinite correlation dimension and a Kolmogorov entropy that was close to 1.

Conclusions. Turbulence in the vocal tract creates high-frequency breathiness, causing noise in the acoustical signal of type 4 voice, proving that the acoustical signal does not represent the motion mechanism behind type 4 voice. The results of this study demonstrate that high-speed imaging can provide a more accurate representation of the type 4 vocal fold vibratory pattern, and a more effective method to explore the mechanism of type 4 signals.

Key Words: High-speed imaging–Type 4 signal–Correlation dimension–Kolmogorov entropy.

INTRODUCTION

In 1995, Titze\textsuperscript{1} introduced a three-tiered classification scheme to quantify voice signals. According to his scheme, type 1 signals were defined as nearly periodic, type 2 signals exhibited strong modulations or subharmonic properties that approached the fundamental frequency in energy, and type 3 signals were irregular or aperiodic. Recently, Sprecher et al\textsuperscript{2} added a fourth voice type to Titze’s voice classification scheme. Sprecher updated type 3 voice to be defined as a signal that is chaotic with a finite dimension, and type 4 voices were defined as chaotic and dominated by stochastic noise features with infinite dimension. In contrast to the type 3 voice signals that are characterized by band-limited spectra with the energy centralized to lower frequencies, type 4 voice signals show a searing of energy across a broader range of frequencies, resembling that of broadband white noise.

Previous research has effectively analyzed the acoustic signals of the first three voice types using perturbation and nonlinear dynamic analysis methods.\textsuperscript{3–8} It has been proven that nonlinear dynamic methods are useful in quantifying complex voice systems, such as type 3 and type 4 voice signals.\textsuperscript{3–8} However, because of complicating factors such as turbulent noise and vocal tract filtering, the collected acoustic voice signals are not able to provide information about the vibratory dynamics of each signal type, leading to the development of other analysis methods. Because visualization of vocal fold vibrations is a powerful tool in the diagnosis of laryngeal pathology, digital kymography (DKG)\textsuperscript{9–11} and high-speed imaging\textsuperscript{12–15} have emerged as effective analysis methods in recent years. These imaging methods allow the objective evaluation of vocal fold vibratory parameters, leading to a better understanding of the mechanisms of disordered voice production and enhanced assessment of laryngeal pathology.

Sprecher et al\textsuperscript{2} systematically analyzed the four types of voice signals using spectrum and acoustical analysis in 2010. Additionally, Zhang et al\textsuperscript{16} provided a complement to this traditional signal classification techniques by using DKG for vibratory classification of voice signals.

The purpose of this study was to incorporate Sprecher’s acoustical classification scheme into the vocal fold vibratory pattern classification method presented by Zhang and investigate the differences between the vibratory patterns and acoustical signals of type 4 voice signals. Vibratory images and corresponding acoustical signals of the four types of voice signals from 10 excised canine larynges were recorded using high-speed imaging and a microphone. The glottal area time series of the four voice signal types were extracted from high-speed images, and nonlinear dynamic analysis was performed to quantify the glottal area and acoustical signal time series.
METHODS

Excised larynx tests

The excised larynx experimental setup is shown in Figure 1A. Ten canine larynges were harvested from healthy laboratory dogs and used in an experimental trial 12-36 hours after excision. Using a hose clamp, a segment of each trachea was secured to a pipe. A conventional air compressor, conditioned to 25-30°C, at 95% relative humidity and smaller than 45 dB noise level, was used to generate airflow. For each trial, subglottal pressure and flow were held constant until phonation was maintained for accurate high-speed measurement. The vocal fold vibratory images were recorded by a high-speed digital camera (Phantom Miro M110; AMETEK, Inc., Berwyn, PA) at a sampling rate of 4000 frames/s with a resolution of 512 × 256 pixels. The videos were recorded at 30 cm distance from each larynx. The microphone was used to record the acoustical signals during each trial.

To collect all four types of acoustical signals and high-speed images, the excised larynx experiments were conducted at subglottal pressures of 10 cm H2O, 30 cm H2O, 40 cm H2O, and 40 cm H2O with a 2-cm wide gap. During these experiments, all four types of acoustical signals and corresponding high-speed images were captured according to the classification scheme reported by Sprecher et al.

As demonstrated in Figure 1B, type 4 signals exhibited little vibration. To research the vibratory patterns of all four voice signal types, the glottal edges from the high-speed images were extracted using Lagrange interpolation analysis with Canny image edge detection as previously reported. The glottal areas were computed by counting the pixels within glottal edges of vocal fold.

Nonlinear dynamic analyses of glottal area and acoustical time series

As mentioned, the purpose of this article was to find a quantitative parameter capable of reflecting the difference between the vibratory pattern and the acoustical signal of all four types of voice signals. As previous research has reported, nonlinear dynamic analyses are usually used to describe the dynamic characteristics of a vocal fold system. In this article, the numerical algorithms of phase space reconstructions, correlation dimension (D2), and Kolmogorov entropy (K2) calculations based on Kantz’s algorithm were applied to analyze the glottal area and acoustical time series.

Phase space reconstructed. The reconstructed phase space shows that the dynamic behavior of a signal can be reconstructed by the time delay. A time series with length N is measured and recorded as \( x(t_1), x(t_2), \ldots, x(t_N) \), where \( x(t_i) \in \mathbb{R}, t_i = t_0 + i\tau (i = 1, 2, \ldots, N) \) at the discrete time interval \( \tau \). Then, the time delay vector \( \tau \) creates the reconstructed phase space as

\[
X(t) = \{ x(t), x(t - \tau), \ldots, x(t - (m - 1)\tau) \},
\]

where \( m \) is the embedding dimension.

Correlation dimension \( D_2 \). The correlation dimension (\( D_2 \)) is a quantitative measure that specifies the number of degrees of freedom needed to describe a dynamic system. The correlation dimension can be calculated as follows:

\[
D_2 = \lim_{r \to 0} \lim_{N \to \infty} \frac{\ln C(N, r)}{\ln r},
\]

where \( r \) is the radius around \( X_{i} \), and the correlation integral \( C(N, r) \) is

\[
C(N, r) = \frac{1}{N(N - 1)} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} \theta(r - \| X_i - X_j \|),
\]

where the Heaviside function \( \theta(x) \) satisfies

\[
\theta(x) = \begin{cases} 
1, & x > 0 \\
0, & x \leq 0
\end{cases}
\]

Using the correlation dimension method, chaos has been found to distinguish different dynamic characteristics from white noise. The estimated \( D_2 \) of white noise does not converge with the increase of embedding dimension \( m \), whereas the estimated \( D_2 \) of a chaotic system converges to a finite value with the increase of \( m \). Therefore, a more complex system has a higher dimension, meaning that more degrees of freedom may be needed to describe its dynamic state.

Kolmogorov entropy \( K_2 \). Kolmogorov entropy (\( K_2 \)) is a description of the rate of information loss in a dynamic

FIGURE 1. (A) An illustration of the excised larynx setup. (B) Nineteen frames of high-speed images of types 1, 2, 3, and 4 vibration of the vocal fold.
system.\textsuperscript{4,5,18} $K_2$ is a useful measure in distinguishing chaos from stochastic noise. For a regular system with little movement, $K_2 = 0$; for a stochastic system, $K_2$ approaches infinity; and for a chaotic system, $K_2$ is a positive constant. That is, the larger the Kolmogorov entropy is, the more complex the system is.

Kolmogorov entropy is usually estimated by second Renyi entropy $K_2$. The relationship between $K_2$ and $C(N,r)$ is as follows:

$$K_2 = \lim_{m \to \infty} \lim_{r \to 0} \frac{1}{m} \ln \frac{C(m,r)}{C(m+1,r)} \quad (5)$$

when discrete time interval $\tau$ and embedding dimension $m$ are fixed values.

**Statistical analysis**

Correlation dimension and Kolmogorov entropy of the glottal areas and acoustical signals from each voice signal type were analyzed using SigmaStat 3.5 (Systat Software, San Jose, CA). The voice signal groups were compared using individual Mann-Whitney rank-sum tests. A significance level of $P < 0.05$ was used throughout.

**RESULTS**

The extracted glottal area time series and recorded acoustic time series of four signal types have been illustrated in Figure 2. The vibratory patterns have a strong correlation with the corresponding acoustical waveforms for type 1, 2, and 3 signals, but not for that of type 4. The type 1 pattern displays regular and periodic waveforms, as shown in Figure 2A. In Figure 2B, the type 2 pattern appears periodic, alternating between small and large amplitudes that are defined as a subharmonic wave.\textsuperscript{1} The type 3 and type 4 waveforms have no apparent periodicity, as shown in Figure 2C and D, but the type 4 signals are characterized by a high-frequency signal.

Figure 3A–D displays the corresponding frequency spectrum of the vibratory patterns and acoustical signals for each voice signal type. The frequency spectrums of type 1 vibratory patterns and acoustical signals show a discrete fundamental frequency and integer multiples harmonic frequencies $nf_0$, $n = 1, 2, \ldots, N$. Type 2 signals exhibit subharmonic frequencies $f_0/2$. Type 3 signals display a broadband frequency spectrum and frequency peaks that mainly exist at low frequencies. In contrast to the type 3, type 4 signals exhibit a smearing of energy across a broad frequency range with no apparent energy peaks, as is characteristic of broadband white noise.

Both glottal area and acoustical time series of all four voice signal types were analyzed by nonlinear dynamic analysis methods. According to the Kantz algorithm, the relationship between correlation integral log $C(r)$ and ln$(r)$ of the four types of acoustical signals and glottal areas was calculated by increasing the embedding dimension $m = 1, 2, \ldots, 12$, as seen in Figure 4.
The results were illustrated in Figure 5. According to Figure 5A, the $D_2$ values of type 1, type 2, and type 3 acoustical signals converged to 1.05 ± 0.001, 2.18 ± 0.003, and 4.2 ± 0.001, respectively. The complexity of the voice signal was directly correlated with correlation dimension. However, the $D_2$ estimate of type 4 acoustical signal was infinite with the increasing of embedding dimension $m$, similar to white noise. As shown in Figure 5B, the $D_2$ values of type 1, type 2, and type 3 glottal areas converged to 1.08 ± 0.001, 2.27 ± 0.001, and 3.19 ± 0.002 similar to the acoustical signals. Interestingly, the $D_2$ estimate of the type 4 glottal areas was finite and converged to 0.11 ± 0.004.

To reflect the difference of the correlation dimensions between the vibratory pattern and acoustical signal of four types of signals, the statistical results were given in Figure 6. The correlation dimensions of type 1, type 2, type 3 glottal area, and acoustical signal series showed strong correlations ($P = 0.341$, $P = 0.974$, $P = 0.431$). However, the correlation dimension of type 4 acoustical signals showed no convergence with increasing embedding dimension. To quantify the difference of the type 4 signal from type 1, 2, 3, we chose the values at embedding dimension $m = 12$ as the correlation dimension of type 4 acoustical signals, as seen in Figure 6. This demonstrates that the statistical data of type 4 voice signals have obvious discrepancies between the vibratory and acoustical system ($P < 0.001$).

Additionally, the statistical graph of Kolmogorov entropy provided strong evidence of the inconsistency between the vibratory and acoustical systems of type 4 voice signals. As illustrated in Figure 7, the $K_2$ of the glottal area and acoustical signal series of type 1, 2, and 3 voice signals showed remarkable consistency ($P = 0.430$, $P = 0.743$, $P = 0.533$), whereas in type 4, the $K_2$ of the glottal area series was close to 0, but that of the acoustical signal series was infinite. Similar to the method for choosing $D_2$ of the type 4 signals, we chose the values at an embedding dimension $m = 12$ as the Kolmogorov entropy. This resulted in significant differences between the vibratory and acoustical systems ($P < 0.001$).

**DISCUSSION**

In this study, we collected the vibratory images and acoustical signals of the four voice signal types from excised larynges. This study aimed at investigating the differences between the vibratory patterns and modulated acoustical systems of these signal types. We developed Sprecher’s classification scheme of acoustic voice signals because it was not previously able to provide information about the vibratory system.
dynamics of the vocal folds because of complicating factors such as turbulent noise and vocal tract filtering. High-speed images reflect the original vibratory patterns of the vocal folds, especially for type 3 and type 4 voice signals. Current acoustical analysis methods such as perturbation analysis and nonlinear dynamic analysis are used to analyze voice signals. However, traditional perturbation techniques are not applicable to type 3 and type 4 signals because their outputs are neither valid nor reliable for aperiodic voice signals.\textsuperscript{1–7} Thus, nonlinear dynamic analysis techniques are needed to

FIGURE 4. The correlation integral $\log C(r)$ versus $\ln (r)$ of the four types (A) acoustical signals and (B) glottal areas.
provide valuable information about the nature of these complex systems.1–6 In this article, vibratory images were recorded using high-speed imaging techniques and analyzed using glottal area time series, frequency spectra, and nonlinear dynamics. The acoustical signals were analyzed using the same methods.

Figures 2 and 3 demonstrate that vibratory patterns have a strong correlation with the corresponding acoustical waveforms for type 1, 2, and 3 voice signals, but not for that of type 4. As shown in Figures 2A and 3A, type 1 displayed periodic time series and discrete frequency spectra, consistent with the findings of previous studies.2,16 This finding confirms that type 1 vibrations are dominated by first-order harmonic components. The estimated correlation dimension of the glottal areas and acoustical signals of type 1 signals were both close to 1, as seen in Figure 5. This suggests that type 1 voice signals are a one-dimensional system. In contrast to type 1, subharmonic components and half of fundamental frequency \((f_0/2)\) existed in the glottal areas, acoustical signal time series, and the corresponding frequency spectrum of type 2 signals, as seen in Figures 2B and 3B. These results suggest that type 2 signals consist of several modes that make the correlation dimension of type 2 higher than that of type 1, as shown in Figure 5.

Both type 3 and type 4 signals were aperiodic without any clear pattern in either amplitude or frequency, as seen in Figures 2C and 3C, and 2D, 3C, and 3D. The frequency spectra of type 3 voice signals were captured by the energy peaks. However, type 4 signals resembled that of white noise, of which the frequency spectra is completely flat over the entire range of frequencies. This is seen in Figures 3C and 3D. The estimated \(D_2\) of type 3 acoustical signals was finite, although that of type 4 was characterized by infinite dimensionality, as illustrated by Figure 5A. Surprisingly, the estimated dimension of the type 4 glottal area time series was close to 0, although the acoustic signals \(K_2\) was much higher, as shown in

**FIGURE 5.** (A) The estimated dimension versus embedding dimension \(m\) of the type 1, 2, 3, and 4 acoustical signals. (B) The estimated dimension versus embedding dimension \(m\) of the type 1, 2, 3, and 4 glottal areas.
Figures 6 and 7. This is because type 4 acoustical signals are disordered voice signals with a large breathy component produced by turbulence in the vocal tract. This breathiness destroys the self-similarity property of the chaotic system and dominates the signal as high-frequency noise. However, the glottal areas extracted from high-speed images of type 4 can eliminate the effects of turbulence and vocal tract filtering, which inverts the original movement patterns of the vocal fold without high-frequency noise. Thus, it is reasonable that the estimated dimensions \(D_2\) and Kolmogorov entropy \(K_2\) of type 4 glottal area time series are finite and converge to 0, as seen in this study.

Figure 7 demonstrates that the \(K_2\) values of the type 4 glottal area time series are similar to that of the type 1 voice signal. This is because the \(K_2\) is theoretically equal to 0 for regular movements. Therefore, it is suggested that correlation dimensions \(D_2\) better represents the differences among the four signal types than Kolmogorov entropy \(K_2\) and could be an effective tool in distinguishing the four signal types.

Current methods of voice signal analysis are time consuming and not ideal for quantifying the complexity of disordered voice characteristics. The results of this study indicate that high-speed vibratory imaging combined with nonlinear analysis such as correlation dimension \(D_2\) could provide an objective metric capable of evaluating all four voice signal types. This approach could allow for the expansion of evidence-based practices in dysphonia management as well as explore the mechanism underlying type 4 voice production. Future research can use the approach followed in this article to determine the critical turbulent energy, or the amount of turbulent energy required to produce type 4 voice signals. A more complete understanding of the mechanism responsible for producing type 4 voice would provide a better approach to the treatment of these disorders as well as examine which pathologies are usually responsible for producing that signal. Finally, this new analysis method can track how a voice signal changes during a treatment period, allowing the clinician to monitor the efficacy of a specific intervention.

CONCLUSIONS
Four types of vocal fold vibrations were directly observed via high-speed images of excised canine larynges. Next, the glottal area time series were extracted from these images. Nonlinear dynamic analysis methods were used to quantify the four types of vocal fold glottal area time series and acoustical signals. The results proved that the nonlinear dynamic parameters of correlation dimension \(D_2\) and Kolmogorov entropy \(K_2\) effectively distinguish all four voice signal types on the basis of their acoustical signals. The calculations of the vibration patterns and acoustical signals for type 1, 2, and 3 voice signals showed strong correlations. However, type 4 signals did not show this correlation. Although the nonlinear dynamic calculations of the vibration patterns of type 4 signals were approximately 0, their acoustic signal calculations were infinite. This phenomenon is due to the large breathy component that is seen in type 4 acoustical signals, which is produced by turbulence in the vocal tract. Therefore, we conclude that high-speed images provide a more original representation of the type 4 vocal fold vibratory pattern and a more effective method to explore the mechanism behind type 4 voice signals.

Acknowledgments
This work was financially supported in part by the National Science Foundation of China (grant nos. 41276040 and 11174240) and the Natural Science Foundation of Fujian Province of China (grant no. 2012J06010). The project was sponsored by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

REFERENCES