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[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 that approached the fundamental frequency in energy, and type 3 signals were irregular or aperiodic. Recently, Sprecher et al[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.[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.[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)[9–11] and high-speed imaging[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[2] systematically analyzed the four types of voice signals using spectrum and acoustical analysis in 2010. Additionally, Zhang et al[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.

Accepted for publication August 19, 2015.

From the *Department of Applied Marine Physics & Engineering, College of Oceanography and Environmental Science, Xiamen University, Xiamen, Fujian, People's Republic of China; †Department of Applied Marine Physics & Engineering, Key Laboratory of Underwater Acoustic Communication and Marine Information Technology, Ministry of Education, Xiamen University, Xiamen, Fujian, People's Republic of China; ‡Department of Applied Marine Physics & Engineering, State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, Fujian, People's Republic of China; and the §Division of Otolaryngology–Head and Neck Surgery, Department of Surgery, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin.

Address correspondence and reprint requests to Yu Zhang, Department of Applied Marine Physics & Engineering, State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen 361005, People's Republic of China. E-mail: yuzhang@xmu.edu.cn

Journal of Voice, Vol. ■, No. ■, pp. 1-7
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2015.08.012

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 H_2O, 30 cm H_2O, 40 cm H_2O, and 40 cm H_2O 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.[2] 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.[17] 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 (D_2), and Kolmogorov entropy (K_2) calculations based on Kantz's algorithm[18] 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.[19] A time series with length N is measured and recorded as $x(t_1)$, $x(t_2)$, $\ldots$, $x(t_N)$, where $x(t_i) \in R, t_i = t_0 + i\tau (i = 1, 2, \ldots, N)$ at the discrete time interval τ. Then, the time delay vector τ creates the reconstructed phase space as

$$X(t) = \{x(t), x(t-\tau), \ldots, x(t-(m-1)\tau)\}, \tag{1}$$

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:

$$D_2 = \lim_{r \to 0} \lim_{N \to \infty} \frac{\ln C(N, r)}{\ln r}, \tag{2}$$

where r is the radius around X_i, and the correlation integral $C(N, r)$ is

$$C(N, r) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} \theta\left(r - \left\|X_i - X_j\right\|\right), \tag{3}$$

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

$$\theta(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{4}$$

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.[4,5,18]

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

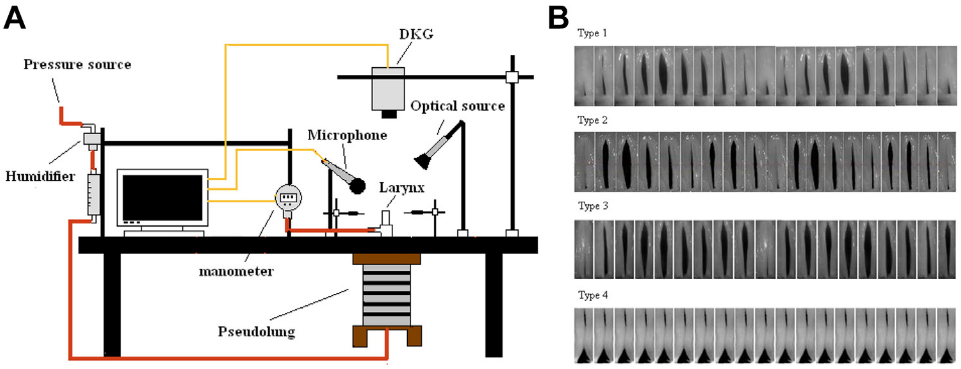

FIGURE 1. (**A**) An illustration of the excised larynx setup. (**B**) Nineteen frames of high-speed images of types 1, 2, 3, and 4 vibration of the vocal fold.