Using Rate of Divergence as an Objective Measure to Differentiate between Voice Signal Types Based on the Amount of Disorder in the Signal

*William M. Calawerts, *Liyu Lin, †J.C. Sprott, and *Jack J. Jiang, *†*Madison, Wisconsin*

Summary: Objective/Hypothesis. The purpose of this paper is to introduce the rate of divergence as an objective measure to differentiate between the four voice types based on the amount of disorder present in a signal. We hypothesized that rate of divergence would provide an objective measure that can quantify all four voice types.

Study Design. A total of 150 acoustic voice recordings were randomly selected and analyzed using traditional perturbation, nonlinear, and rate of divergence analysis methods.

Methods. We developed a new parameter, rate of divergence, which uses a modified version of Wolf's algorithm for calculating Lyapunov exponents of a system. The outcome of this calculation is not a Lyapunov exponent, but rather a description of the divergence of two nearby data points for the next three points in the time series, followed in three time-delayed embedding dimensions. This measure was compared to currently existing perturbation and nonlinear dynamic methods of distinguishing between voice signals.

Results. There was a direct relationship between voice type and rate of divergence. This calculation is especially effective at differentiating between type 3 and type 4 voices ($P < 0.001$) and is equally effective at differentiating type 1, type 2, and type 3 signals as currently existing methods.

Conclusion. The rate of divergence calculation introduced is an objective measure that can be used to distinguish between all four voice types based on the amount of disorder present, leading to quicker and more accurate voice typing as well as an improved understanding of the nonlinear dynamics involved in phonation.

Key Words: Voice type–Nonlinear–Chaos–Parameter–Disorder.

INTRODUCTION

The existence of chaotic dynamics in phonation has been widely recognized since the early 1990s.[1,2] Subsequently, Titze, Baken, and Herzel separated voice signals into three types. Type 1 signals are periodic in nature, type 2 signals contain subharmonic or modulating frequencies, and type 3 signals have no apparent periodic structure.[3] Traditional perturbation measures such as jitter and shimmer have proven effective in analyzing type 1 and type 2 voice signals, but not type 3.[4–8] These measurements are determined by estimating the fundamental frequency and peak amplitude of each phonatory cycle, respectively. As voice type increases, estimates for jitter and shimmer have been proven to contain significantly larger trk and err values.[9] The limited robustness of these methods leads to poor reliability and large variance during irregular phonation.[9,10] Nonlinear dynamic measurements such as correlation dimension (D2) or largest Lyapunov exponent are successful in differentiating between normal and irregular phonations, whereas thes traditional perturbation methods fail.[9–12] Recently, Sprecher et al introduced the addition of a fourth voice type. This scheme reclassified type 3 voice as chaotic with a finite dimension, and defined type 4 voice as chaotic with infinite dimension.[9] These type 4 signals are heavily obscured by stochastic signal components, making it impossible to accurately calculate the D2 and the Lyapunov exponent.[9,11,13] Therefore, there are currently only subjective measures available for evaluating type 4 voice, such as spectrograms and perceptual analysis. These methods of analyses remain valid, but a method of objective evaluation is needed.

As mentioned above, previous studies have demonstrated that Lyapunov exponent calculations are capable of differentiating between regular and irregular phonation.[14–16] This is unsurprising since Lyapunov exponents, which are the average exponential rates of divergence or convergence of nearby orbits in phase space, are effective descriptors of chaos.[17,18] Exponential orbital divergence indicates that points with minuscule initial differences will soon diverge to drastically different values. The magnitude of the exponent reflects the time scale at which the system becomes unpredictable.[18] Thus, a higher maximum Lyapunov exponent indicates that the system is more chaotic. However to calculate a true Lyapunov exponent, a sufficient embedding dimension is required. This cannot be calculated for type 4 voice because the dimension of the signal is immeasurably high and potentially infinite. Using the correct embedding dimension allows the Lyapunov exponent to be measured in as many dimensions as there are present in the system.

Although we cannot calculate the Lyapunov exponent for type 4 voice, we reasoned that the rate of divergence in a certain dimension of each data sample could still be calculated. This value of divergence should increase as the amount of disorder in a sample increases. We hypothesized that the rate of divergence of two nearby points in a data series followed in three dimensions would have a direct relationship with voice type. That is, the rate of divergence would increase from type 1 to type 2, type

Accepted for publication January 7, 2016.

From the *School of Medicine and Public Health, Department of Surgery, Division of Otolaryngology—Head and Neck, University of Wisconsin-Madison, Madison, Wisconsin; and the †Department of Physics, University of Wisconsin-Madison, Madison, Wisconsin.

Address correspondence and reprint requests to Jack Jiang, 1300 University Avenue, 2745 Medical Sciences Center, Madison, WI 53706. E-mail: jiang@surgery.wisc.edu; jjjiang@wisc.edu

Journal of Voice, Vol. 31, No. 1, pp. 16–23
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.01.005

2 to type 3, and type 3 to type 4. The calculated result would provide an objective parameter to specifically distinguish between type 3 and type 4 voices because there is no measure currently available to do so based on nonlinear principles. Second, we hypothesized that the rate of divergence measure proposed in this paper would be comparable with the existing measures used to quantify voice signals, such as correlation dimension, Lyapunov exponent, and percent jitter and percent shimmer.

METHODS

Voice selection

We randomly selected 150 voice samples from the Disordered Voice Database Model 4337 (KayPEN-TAX, Lincoln Park, NJ). The sampling frequency for each of the samples used was 25 kHz. Each sample consisted of 0.75 seconds of sustained "ah" phonation. Table 1 displays the summary characteristics of the subjects selected for analysis.

Spectrogram analysis

A spectrogram was generated for each sample for voice classification. Based on the spectrogram classification system proposed by Sprecher et al,[9] each of the samples was subjectively sorted into one of the four voice types by three researchers. Samples that were not traditional representations of voice types or were sorted differently by any researchers were not used in the analysis. The final sample consisted of 22 type 1 samples, 49 type 2 samples, 50 type 3 samples, and 26 type 4 samples.

Sampling rate determination

The original sampling rate for each of the voice samples was 25 kHz. However, choosing the optimum time delay when conducting nonlinear dynamics calculations is critical to ensure accuracy. When the time difference between two points is too small, each data point is too close to its predecessors. This causes the attractor to stretch out along the diagonal in reconstructed space, and leads to a divergence calculation that is spuriously low. However, when time delay is too long, the system loses its determinism.

To determine the optimum time delay, we constructed the attractor of each sample with differing time delays. Figure 1 shows the relationship between a type 3 voice sample's attractor and the rate of down sampling used. Under the assumption that the system is chaotic, which is true for type 3 voice, this plot is of an attractor. We reviewed several voice samples and found that as the time delay increased, the attractor continued to expand until a downsampling rate of 8 and then began to fold over itself. This indicated that a sample frequency of 3.125 kHz was the optimum time delay between each point for our analysis. This method of calculating the time delay is consistent with the minimum mutual information calculation we performed. The minimum mutual information method of calculating the time delay provides a systematic method for choosing time delays and quantitatively describes the spatial patterns of chaotic signals by choosing the first minimum of the mutual information for the signal's attractor.[19–21]

Perturbation analysis

Perturbation analysis was conducted using the *TF32* software.[22] The measures of percent jitter and percent shimmer were calculated for each type 1 and type 2 signals. Jitter represents the cycle-to-cycle variation in signal frequency, whereas shimmer measures the cycle-to-cycle variation in signal amplitude.[22] Perturbation analysis was not conducted for either type 3 or type 4 voices because of the previous research that found that for those types of phonation, perturbation analysis is neither valid nor reliable.[4,5,9,10]

Correlation dimension and Lyapunov exponent analysis

Nonlinear dynamic analysis was applied to type 1, type 2, and type 3 voices using the same method that has been used in numerous publications.[9,10,14,15] Correlation dimension and Lyapunov exponent calculations cannot be accurately calculated for type 4 voice signals because of the extremely high dimensionality of those samples. The correlation dimension (D2) measures the number of degrees of freedom necessary to describe a system. Thus, a system with a higher degree of complexity needs more degrees of freedom to characterize its dynamic state.[17,19]

TABLE 1.
Subject Characteristics and Average and Standard Deviation for the Rate of Divergence Values of Each Voice Type for Each Signal Type Group

Voice Type	Number of Samples	Age in Years	Gender	Average Rate of Divergence	Standard Deviation
1	22	37.3 (22–63)	3 Males 19 Females	0.238761	0.06912
2	49	44.5 (23–75)	24 Males 25 Females	0.347352	0.102767
3	50	49.9 (7–80)	24 Males 26 Females	0.541578	0.13259
4	26	62.8 (30–85)	9 Males 17 Females	0.740452	0.047607

Age is displayed as mean age (age range).

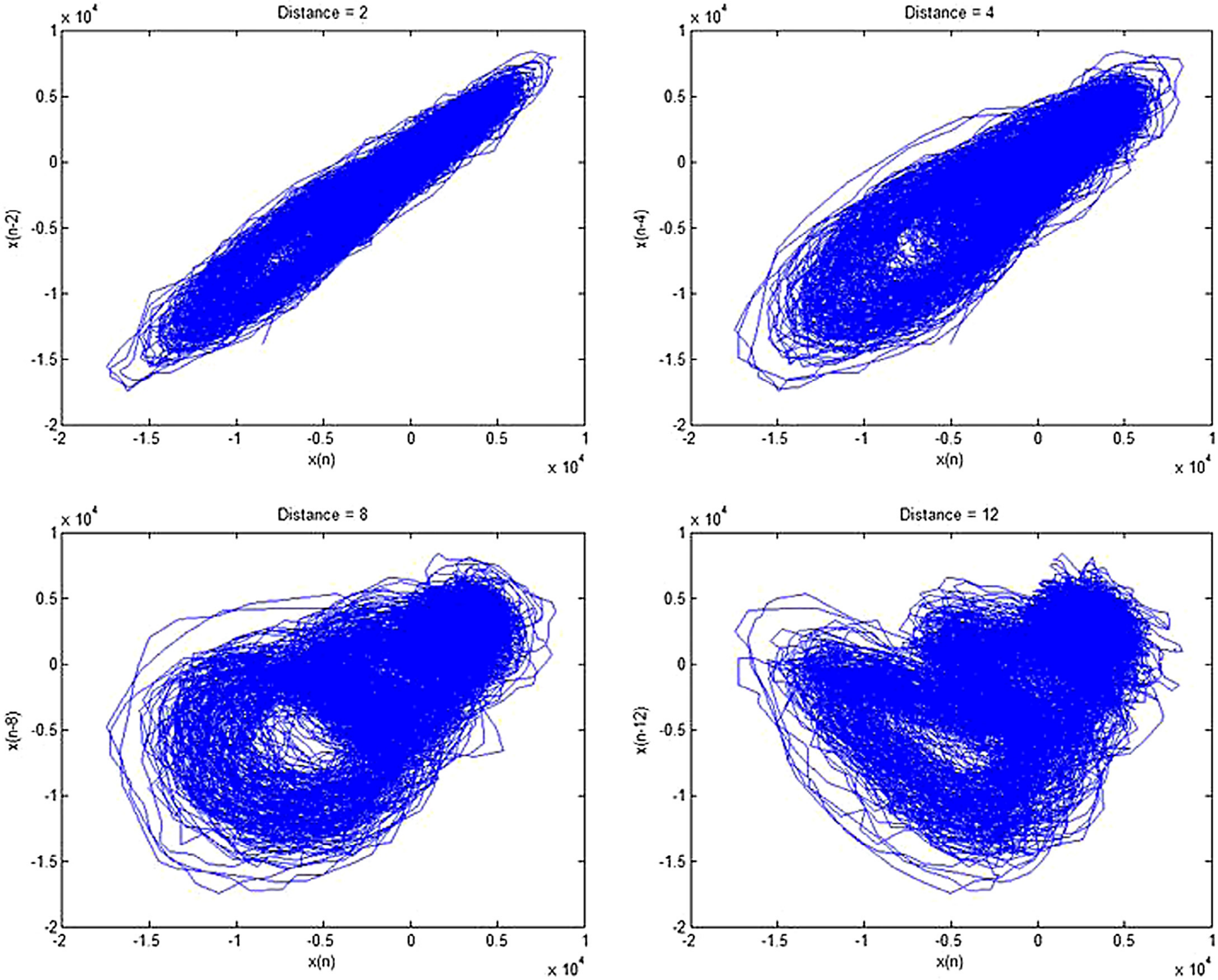

FIGURE 1. The shape of the attractor at each downsampling rate (distance). As we increase the time delay from a downsampling rate of 1–8, we see that the attractor begins to expand. Then, as we reach a downsampling rate of 12, the attractor begins to fold over itself.

Lyapunov exponents were calculated using Wolf's algorithm.[18] The embedding dimension to compute the Lyapunov exponent was calculated via the method presented in Cao.[23] Then, D2 was calculated using a time delay technique to create a phase space:

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

where m is the embedding dimension, and τ is the time delay.[21] A more detailed description of this calculation can be seen in Figure 2.

Chaos data analyzer program and rate of divergence analysis

After sorting the samples by voice type and determining the optimum time delay, the samples were analyzed using *Chaos Data Analyzer* (*CDA*) *Pro version 2.2* (Physics Academic Software, Raleigh, NC). The CDA performs various tests on a time series with the goal of detecting and quantifying hidden determinism. The files that are analyzed are standard ASCII text supplied by the user.[25]

To examine the rate of divergence of each of our samples, we used the Lyapunov exponent function in the CDA. The program utilizes the algorithm described in Wolf et al[18] to calculate the Lyapunov exponent. As mentioned above, because we cannot calculate the dimension of type 4 voice, it is impossible to set the embedding dimension necessary to calculate the Lyapunov exponent for type 4 voice samples. Thus, we chose an embedding dimension (D) of 3, which was kept consistent across the four different voice types for our calculation purposes. This value indicates that we start with two points close in time-lag space and follow their degree of separation in three dimensions. Although the system may consist of more than three dimensions, the calculation consistently chooses the same three dimensions for each sample. The next parameter set by the user is the number of sample intervals (n) over which each pair of points is followed before a new pair is chosen. We chose the default n value of 3, because following a pair of points too long causes the exponential divergence of the orbits to be lost and following too

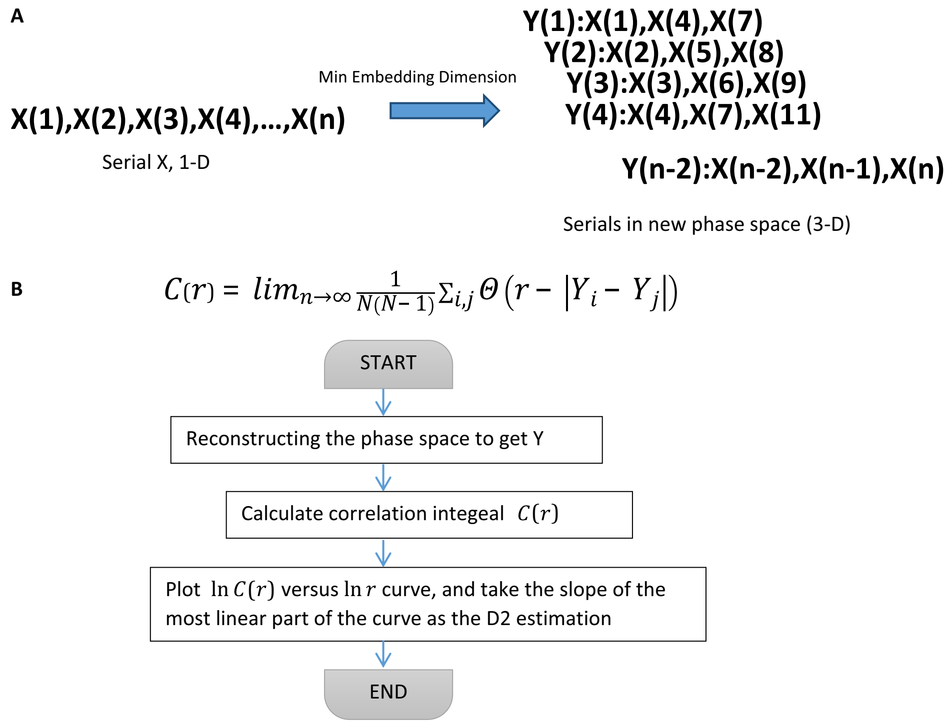

FIGURE 2. A flow chart to demonstrate the calculation of correlation dimension. (**A**) Reconstruction of phase space using a minimum embedding dimension of 3. (**B**) A correlation integral C(r) was calculated using the algorithm presented by Theiler.[24] 14. (Takens F. Detecting strange attractors in turbulence. In: Rand D. A. and Young L.-S, ed. *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, Vol. 898. Springer-Verlag; 1986:366–381). The radius around X_i is denoted as r. Using r to define the scaling region, curves of $\log_2 C(r)$ versus $\log_2 r$ were generated, and the value of D2 was calculated by taking the slope of the most linear part of the curve.

short makes the calculation less representative of the system. Lastly, the accuracy of the data was set to 10^{-4} to exclude outliers. Therefore, the output of this calculation is a measure of the rate of divergence in three dimensions of two nearby data points for the next three sample intervals. An output of 0 signifies that the sample is perfectly periodic, and an output of 1 signifies that the separation of nearby orbits doubles at each time step. A higher average rate of divergence corresponds to a system with a higher degree of chaos.

Each voice sample was downsampled by a factor of 8; thus, there were 8 fragments for each voice type. For example one of the fragments used data points 1 and 9, whereas the next fragment used data points 2 and 10, as seen in Figure 3. We opted to run each of the 8 fragments of every sample to completion to provide a more accurate representation of each sample. Once the rate of divergence of all the samples had been calculated, the average and standard deviation were calculated for each voice type.

Statistical analysis

A normality test determined that there was not a normal distribution across the four voice signal types. Thus, the rate of divergence, Lyapunov exponent, jitter %, shimmer %, and Correlation dimensions of different voice signal types were compared using a Kruskal-Wallis one-way ANOVA on ranks. Then, Dunn tests were used to compare each of the individual groups. All calculations and graphs were plotted using *SigmaPlot version 11.0* (Systat Software, San Jose, CA), and a significance level of $\alpha = 0.05$ was used throughout.

RESULTS

The results of the Kruskal-Wallis one-way ANOVA on ranks demonstrated that the differences between the voice type rate of divergence values were significant ($P < 0.001$). Table 1 lists the average, standard deviation, and number of samples for each voice type. The type 2 rate of divergence was significantly different than type 1 divergence ($P < 0.001$). Type 3 voice was significantly different from type 1 and type 2 voices ($P < 0.001$). Lastly,

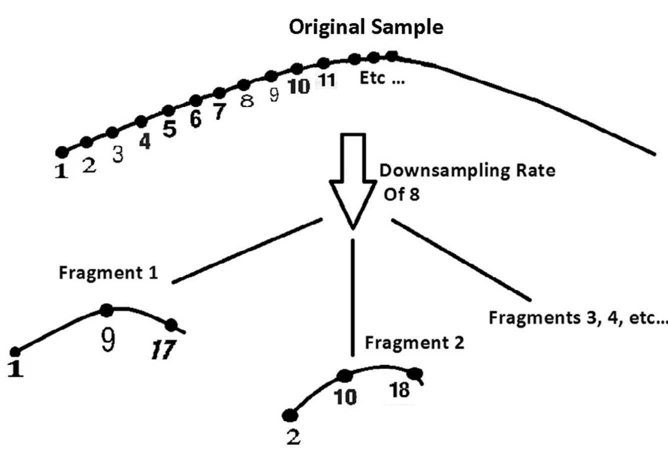

FIGURE 3. A visual representation of the fragment analysis used in this study. Because a downsampling rate of 8 was used, 8 fragments of each original signal sample were produced. Each fragment was analyzed to improve power and accuracy.

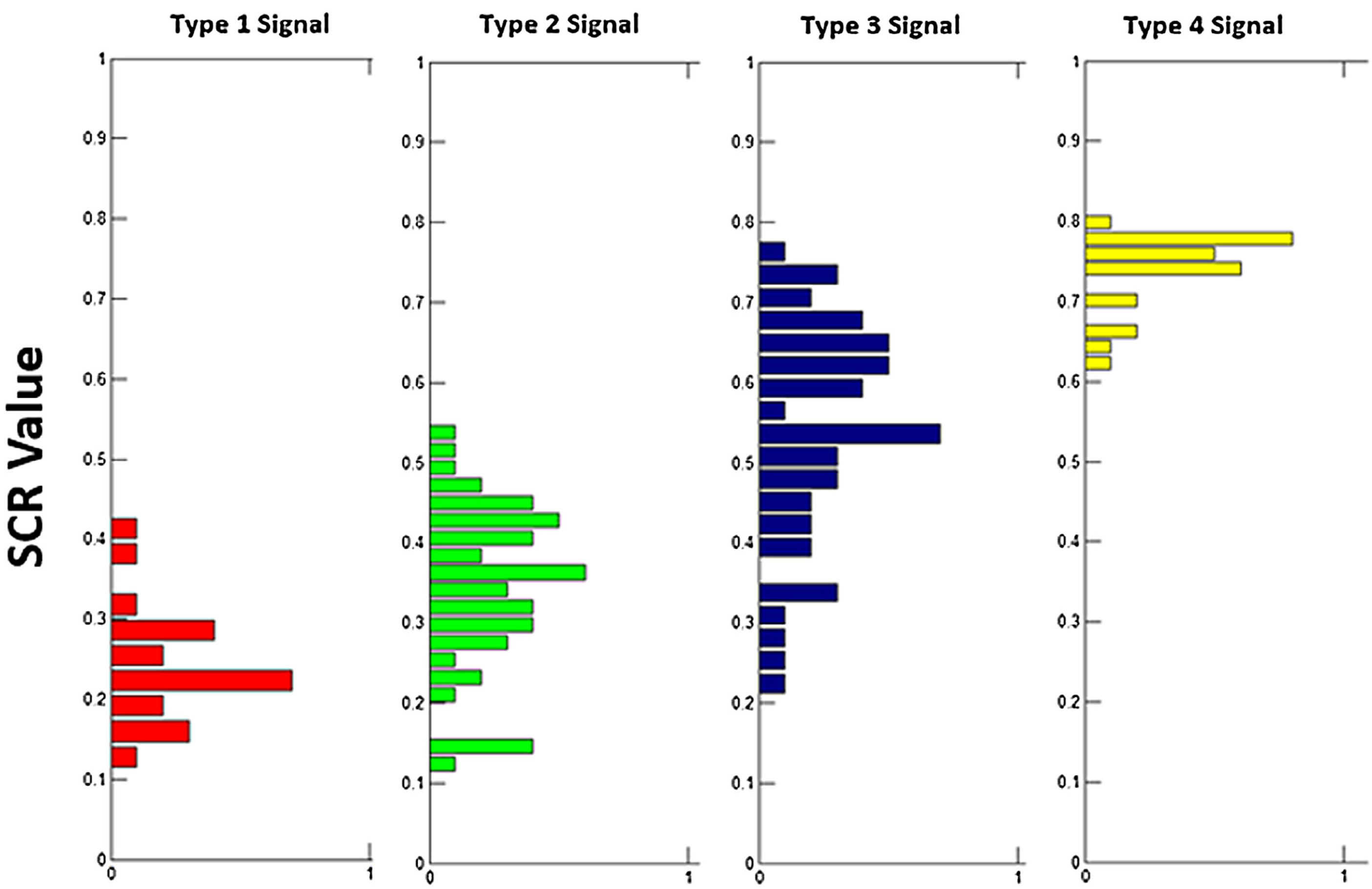

FIGURE 4. Histogram of the rate of divergence values for each voice type. The y-axis displays the actual rate of divergence value, and the x-axis shows the frequency of occurrence of each value range.

type 4 voice was significantly different from type 1, type 2, and type 3 voices ($P < 0.001$). Visual representations of these results can be seen in Figures 4 and 5A and in Table 2.

Perturbation analysis showed an increasing degree of complexity as voice type increased. Both jitter and shimmer % were significantly different between type 1 and type 2 voices ($P < 0.001$). Visual representations of these results can been seen in Figure 5B,C and in Table 2.

Lastly, the nonlinear analysis of the samples indicated increasing system complexity with increasing voice type. The Kruskal-Wallis ANOVA on ranks showed that there was a significant difference among the group's Lyapunov exponent values. Type 2 signals had significantly higher Lyapunov exponents than type 1 ($P < 0.001$), and type 3 signals had significantly higher Lyapunov exponents than type 1 and type 2 voices ($P < 0.001$). A box plot of this data can be seen in Figure 5D. The Kruskal-Wallis ANOVA on ranks of correlation dimension demonstrated that there was a significant difference among the groups ($P < 0.001$). D2 increased with voice type, with type 3 voice having significantly higher D2 values than type 2 and type 1 voices ($P = 0.01$ and $P < 0.001$, respectively). However type 1 and type 2 did not have significantly different D2 values ($P = 0.423$), which is consistent with results from Sprecher et al.[9] A box plot of these results can be seen in Figure 5E.

DISCUSSION

In this study, we define a new nonlinear, objective measure for quantifying voice signals. The calculation for this measure utilizes Wolf's algorithm for calculating Lyapunov exponents, but chooses to follow two close data points in only three dimensions as they separate during the next three time points. This method allows for the inclusion of voice samples with high dimensionality (type 4) and allows comparison across all voice types based on the amount of disorder present in each signal.

Currently, the leading methods of classifying high-dimensional voice types are spectrograms and perceptual analysis. These subjective measures leave room for error in interpretation of voice disorders that may affect treatment. Objective measures provide insight into the nonlinear dynamics of type 3 and 4 voice disorders as well as establish quantitative criteria for classifying voice types. Differentiating between type 3 and 4 voice signals is clinically relevant because it can provide insight into the complex biomechanical interaction involved in these types of phonations. The nature of the distinction between type 3 and type 4 signals is based on varying amounts of dimensionality, which suggests that there are functional differences between the two. Thus, if we can distinguish between these signals, it is possible to investigate the morphological or functional differences in larynges that produce these different voice signals.

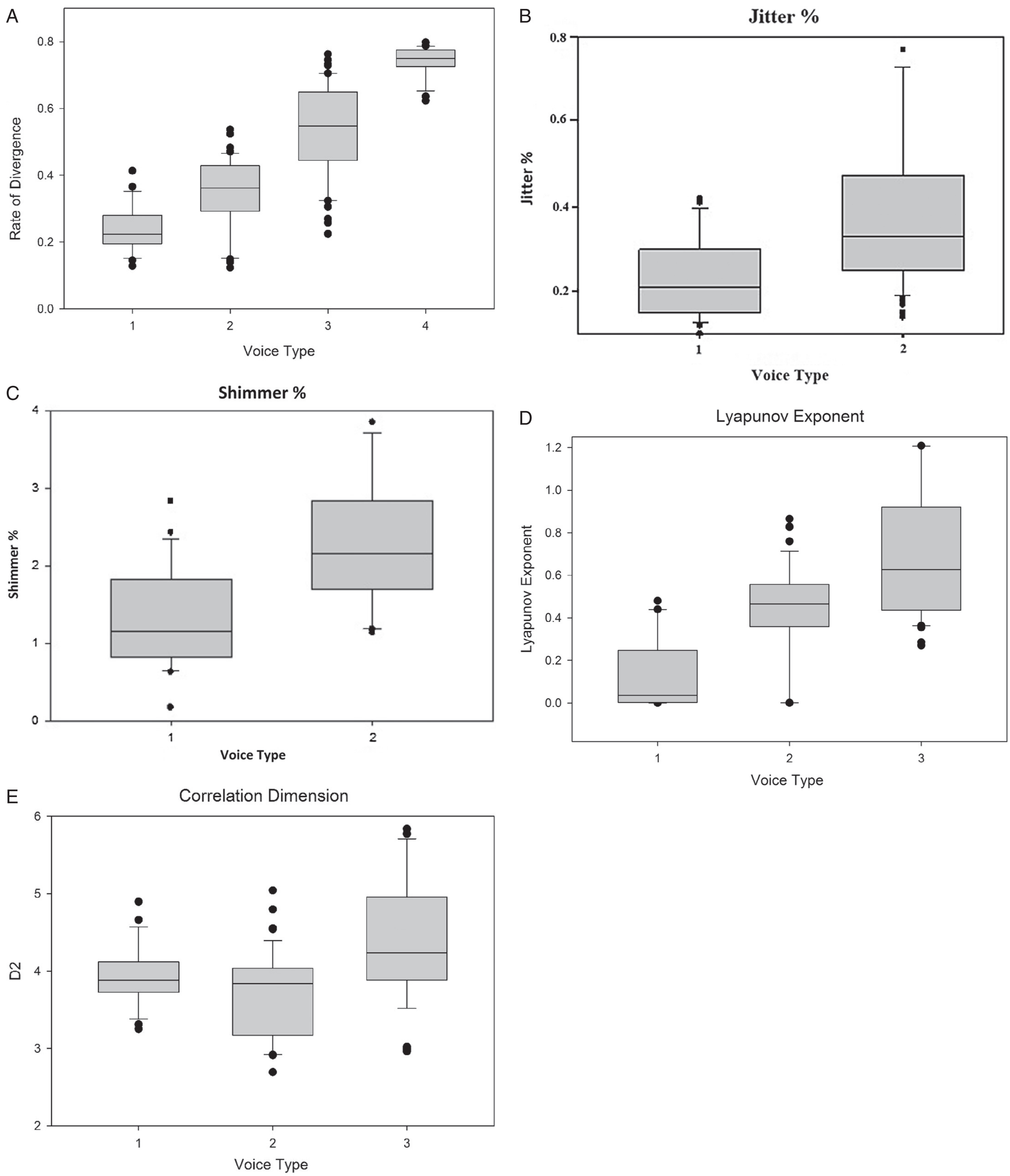

FIGURE 5. Box plots of (**A**) the rate of divergence of all four voice types, (**B**) percent jitter of type 1 and type 2 voices, (**C**) percent shimmer of type 1 and type 2 voices, (**D**) Lyapunov exponent values for type 1, type 2, and type 2 voices, **(E)** and correlation dimension values for type 1, type 2, and type 3 voices.

TABLE 2.
Description of Variables, Comparisons, and Statistical Tests Used in Analysis

Measure	Comparison	Statistical Test	*P* Value
Rate of divergence	All groups	Kruskal-Wallis one-way ANOVA on ranks	$P < 0.001$
Rate of divergence	Type 1 vs. type 2	Dunn *t* test	$P < 0.001$
Rate of divergence	Type 1 vs. type 3	Dunn *t* test	$P < 0.001$
Rate of divergence	Type 1 vs. type 4	Dunn *t* test	$P < 0.001$
Rate of divergence	Type 2 vs. type 3	Dunn *t* test	$P < 0.001$
Rate of divergence	Type 2 vs. type 4	Dunn *t* test	$P < 0.001$
Rate of divergence	Type 3 vs. type 4	Dunn *t* test	$P < 0.001$
Jitter %	Type 1 vs. type 2	Kruskal-Wallis one-way ANOVA on ranks	$P < 0.001$
Shimmer %	Type 1 vs. type 2	Kruskal-Wallis one-way ANOVA on ranks	$P < 0.001$
Lyapunov exponent	All groups	Kruskal-Wallis one-way ANOVA on ranks	$P < 0.001$
Lyapunov exponent	Type 1 vs. type 2	Dunn *t* test	$P < 0.001$
Lyapunov exponent	Type 1 vs. type 3	Dunn *t* test	$P < 0.001$
Lyapunov exponent	Type 2 vs. type 3	Dunn *t* test	$P < 0.001$
Correlation dimension	All groups	Kruskal-Wallis one-way ANOVA on ranks	$P < 0.001$
Correlation dimension	Type 1 vs. type 2	Dunn *t* test	$P = 0.423$
Correlation dimension	Type 1 vs. type 3	Dunn *t* test	$P < 0.001$
Correlation dimension	Type 2 vs. type 3	Dunn *t* test	$P = 0.01$

A significance level of $\alpha = 0.05$ was used throughout.

Our calculations suggest that type 4 voice samples have significantly higher divergence rates than any of the other voice types. Additionally, type 3 voice samples exhibit significantly higher divergence rates than both type 1 and type 2. The importance of this calculation is that the rate of divergence provides information about the amount of nonlinearity in a signal. A higher divergence rate indicates that the sample has a higher amount of disorder than a sample with a low divergence rate. Thus, our hypothesis that the rate of divergence (along with amount of disorder) would increase with increasing voice type was confirmed. This measure was effective in quantifying type 3 and type 4 voice signals. Additionally, the rate of divergence was proven effective as currently available measures in differentiating between voice types. This suggests that the rate of divergence should replace perturbation or other nonlinear measures in analyzing high level voice signals.

CONCLUSION

In this study, the rate of divergence was defined as an objective, nonlinear measure utilized to distinguish each of the four voice types. One limitation of this study is that there appeared to be an overlap at the transitions between each voice type. This can be seen in Figure 4. However, this outcome was expected for a several reasons. Because the voice samples were typed via subjective spectrogram analysis, there is a risk that some types were misclassified. There appear to be some outliers for each of the voice types, which may be due to subjective rating error. Furthermore, the only current analysis methods for voice typing are subjective, so it is unsurprising that there will be a "gray area" in which the labeling overlaps. The benefit of the rate of divergence as an analysis tool is that it offers a continuous variable capable of describing the amount of disorder present in a system.

Supplementing subjective analysis with this objective measure could reduce the chance of incorrect voice typing. This would help clinicians quickly and accurately evaluate voice type, which could lead to superior treatment for individuals suffering from voice disorders. Furthermore, using the rate of divergence calculation will allow the future investigation of mechanisms that underlie type 4 voice signals and help determine the critical turbulent energy, or amount of turbulent energy required to produce type 4 voice. A better understanding of the signal characteristics of a voice will allow clinicians to better measure progress in treatment and indicate when a treatment is not working. Lastly, this measure may be used to investigate which disorders are most likely to produce type 4 signals and which treatment intervention is warranted.

Acknowledgments

This study was supported by NIH Grant No. RO1 DC006019 from the National Institute of Deafness and other Communication Disorders. Dr. Liyu Lin is noted for equal contributions as the first author.

REFERENCES

1. Awrejcewicz J. Bifurcation portrait of the human vocal cord oscillations. *J Sound Vibrat*. 1990;136:151–156.
2. Titze IR, Baken RJ, Herzel H. Evidence of chaos in vocal fold vibration. *Vocal Fold Physiol*. 1993;143–188.
3. Titze IR. Workshop on acoustic voice analysis: summary statement. In: *Notional Center for Voice and Speech*. Denver, CO; 1995:26–27.
4. Karnell MP, Chang A, Smith A, et al. Impact of signal type of validity of voice perturbation measures. *NCVS Stat Prog Rep*. 1997;11:91–94.
5. Behrman A, Agresti CJ, Blumstein E, et al. Microphone and electroglottographic data from dysphonic patients: type 1, 2, and 3 signals. *J Voice*. 1998;12:247.
6. Brockmann-Bauser M, Beyer D, Bohlender JE. Reliable acoustic measurements in children between 5;0 and 9;11 years: gender, age, height, and weight effects on fundamental frequency, jitter and shimmer in phonations without and with controlled voice SPL. *Int J Pediatr Otohinolaryngol*. 2015. doi:10.1016/j.ijporl.2015.09.005.

7. Inwald EC, Dollinger M, Schuster M, et al. Multiparametric analysis of vocal fold vibrations in healthy and disordered voices in high-speed imaging. *J Voice*. 2011;25:576–590.
8. Brockman M, Drinnan MJ, Storck C, et al. Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task. *J Voice*. 2011;25:44–53.
9. Sprecher A, Olszewski A, Zhang Y, et al. Updating signal typing in voice: addition of type 4 signals. *Acoust Soc Am*. 2010;127:3710–3716.
10. Jiang JJ, Zhang Y, Ford CN. Nonlinear dynamic of phonations in excised larynx experiments. *J Acoust Soc Am*. 2003;114(4 pt 1):2198–2205. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14587617. Accessed December 3, 2000.
11. Choi SH, Zhang Y, Jiang JJ, et al. Nonlinear dynamic-based analysis of severe dysphonia in patients with vocal fold scar and sulcus vocalis. *J Voice*. 2012. doi: 10.1016/j.jvoice.2011.09.006.
12. Awan SN, Novaleski CK, Rousseau B. Nonlinear analysis of elicited modal, raised, and pressed rabbit phonation. *J Voice*. 2014; 28:538–547.
13. Awan S, Nelson R, Jiang JJ. Nonlinear dynamic analysis of disordered voice: the relationship between the correlation dimension (D2) and pre-/post-treatment change in perceived dysphonia severity. *J Voice*. 2010;24:285–293.
14. Zhang Y, Jiang JJ. Acoustic analyses of sustained and running voices from patients with laryngeal pathologies. *J Voice*. 2006;22:1–9.
15. Zhang Y, Jiang JJ, Biazzo L, et al. Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis. *J Voice*. 2004;19:519–528.
16. Giovanni A, Ouaknine M, Triglia JM. Determination of largest Lyapunov exponent of vocal signal: application to unilateral laryngeal paralysis. *J Voice*. 1999;13:1–14.
17. Jiang JJ, Zhang Y, McGilligan C. Chaos in voice, from modeling to measurement. *J Voice*. 2005;20:2–17.
18. Wolf A, Swift J, Swinney H, et al. "Determining Lyapunov exponents from a time series." 1984: 285–317.
19. Sprott JC. *Chaos and Time-series Analysis*. Oxford: 2006.
20. Fraser AM, Swinney HL. Independent coordinates for strange attractors from mutual information. *Am Phys Soc*. 1986;33:1134–1140.
21. Packard NH, Crutchfield JP, Farmer JD, et al. Geometry from a time-series. *Phys Rev Lett*. 1980;45:712–716.
22. Milenkovic P. *TF32 User's Manual*. Madison, WI: 2001.
23. Cao L. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D*. 1997;110:43–50.
24. Theiler J. Spurious dimension from correlation algorithms applied to limited time-series data. *Phys Rev A*. 1986;34:2427–2432.
25. Sprott JC, Rowlands G. *Chaos Data Analyzer: The Professional Version*. Raleigh, NC: Physics Academic Software; 1995.