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

© 2017 The Voice Foundation. Published by Elsevier Inc. All rights reserved.

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).

Download English Version:

https://daneshyari.com/en/article/5124297

Download Persian Version:

https://daneshyari.com/article/5124297

Daneshyari.com