Objective Acoustic Analysis of Pathological Voices from Patients with Vocal Nodules and Polyps

Jack J. Jiang\textsuperscript{a, b}, Yu Zhang\textsuperscript{a}, Julia MacCallum\textsuperscript{a}, Alicia Sprecher\textsuperscript{a}, Liang Zhou\textsuperscript{b}

\textsuperscript{a}Department of Surgery, Division of Otolaryngology, Head and Neck Surgery, University of Wisconsin School of Medicine and Public Health, Madison, Wisc., USA; \textsuperscript{b}Eye, Ear, Nose, and Throat Hospital, Fudan University Medical School, Shanghai, China

\textbf{Key Words}
Acoustic analysis \cdot Acoustic measurements \cdot Voice disorders \cdot Voice research

\textbf{Abstract}
\textbf{Objective:} We aim to examine the abilities of objective acoustic analysis methods (nonlinear dynamic and traditional perturbation measures) to describe voices from individuals with vocal nodules and polyps.\textbf{Subjects and Methods:} Sustained vowel recordings from normal subjects, patients with vocal nodules, and patients with vocal polyps were analyzed. Perturbation measures of jitter and shimmer were obtained with the Multi-Dimensional Voice Program (MDVP) and CSpeech. Signal-to-noise ratio was calculated using CSpeech. Nonlinear dynamic measures of phase space reconstruction and correlation dimension were also applied to analyze the voices.\textbf{Results:} A significant difference between normal and polyp groups was found in jitter and shimmer obtained from MDVP, as well as in jitter and signal-to-noise ratio from CSpeech. However, no parameters significantly differentiated between normal and nodule groups. Shimmer from CSpeech did not reveal any significant differences among any of the groups. Correlation dimension values for the nodule and polyp groups were significantly higher than the normal group.\textbf{Conclusion:} Nonlinear dynamic analysis has great potential value for the characterization of voice from patients with vocal nodules and polyps. The combination of traditional perturbation and nonlinear dynamic measures may improve our ability to provide objective clinical analysis of voices with vocal mass lesions.

\textbf{Introduction}
Vocal nodules and polyps are benign mass lesions of the vocal folds that can interfere with vocal fold closure, introduce asymmetry to the vocal folds, and produce severely rough voices and aperiodic acoustic waveforms [1]. Vocal nodules typically occur as bilateral lesions, while vocal polyps typically occur as unilateral lesions. These lesions of the lamina propria often result in a perceptual quality of hoarseness during phonation. Traditionally, perturbation measures such as jitter, shimmer, and signal-to-noise ratio (SNR) have been applied as noninvasive, objective methods to assess these mass lesions [2–9]. Perturbation analysis is commonly performed using the Multi-Dimensional Voice Program (MDVP) [10] and CSpeech [11], popular commercial programs used in both clinical practice and in research laboratories. However, it has recently been suggested that perturbation analysis is only reliable for nearly periodic signals and might not be applicable to aperiodic signals [12, 13]. Thus, complementary objective measures capable of analyzing severely rough and aperiodic voices are both important and necessary.
In recent years, nonlinear dynamic analysis, including phase space reconstruction and correlation dimension calculation, has been applied to investigate chaotic behaviors of biomedical systems in many fields, including laryngology [14–21]. There has been substantial evidence to demonstrate that nonlinearity is inherently involved in laryngeal physiology and voice production, and nonlinear dynamic analysis has recently shown potential application to clinical voice analysis [8, 14, 15, 17, 19–21]. Baken [14] originally applied fractal analysis to quantify the irregularity in period and amplitude of normal voices. Titze [12] suggested methods of improving understanding of pathological voices through nonlinear dynamic analysis. Jiang et al. [22] found that correlation dimension ($D_2$) values of disordered pathological voices are statistically different from those of normal voices. Ping et al. [8] demonstrated the efficacy and possible clinical application of the Lyapunov exponent in analyzing disordered voice. Box-counting dimension, another method of estimating fractal dimension, has also been used as a quantitative measure of the phonatory irregularities in laryngeal pathologies [14]. Thus, applying nonlinear dynamic analysis to voices with vocal mass lesions may lead to valuable clinical diagnostic methods based upon objective, reliable acoustic analysis, enhancing the understanding and assessment of these laryngeal pathologies.

The purpose of this study was to examine the capabilities of nonlinear dynamic analysis and perturbation analysis to differentiate between normal voices from healthy subjects and pathological voices from patients with vocal nodules and polyps. Perturbation analysis was performed with MDVP and CSpeech software. Traditional parameters of SNR, percent jitter, and percent shimmer were calculated. Nonlinear dynamic analysis methods of phase space reconstruction and correlation dimension parameters were applied to quantify irregular phenomena. The study examined the potential application of traditional perturbation and nonlinear dynamic analyses to the objective assessment of laryngeal mass lesions.

### Materials and Methods

#### Experimental Design

The Institutional Review Board of Fudan University EENT Hospital approved the protocol and consent procedure used in this study. Three subject cohorts participated in this study: normal subjects (17 females and 4 males, ages 27–67), patients with vocal nodules (21 females, ages 19–59), and patients with vocal polyps (26 females and 13 males, ages 27–65). Diagnoses were made by an attending otolaryngologist based upon the subjects’ medical histories and laryngoscopic examinations of the subjects’ vocal folds. The normal subjects were healthy volunteers with no current or past evidence of voice disorders and with normal larynges, as determined by clinical examination performed by an otolaryngologist.

Subjects were asked to sustain the vowel /a/ at a comfortable pitch and intensity as steadily and as long as possible while audio recordings were made. This task was performed in a double-walled, sound-attenuated room. Recordings were collected with a condenser microphone, model 4144 (Brüel & Kjær, Nærum, Denmark) and were digitized using an acquisition card with a 12-bit analog-to-digital converter, AT-MIO-E-2 (National Instruments, Austin, Tex., USA) at a sampling rate of $f_s = 20$ kHz via a custom-made LabVIEW 4.0 (National Instruments) data acquisition program. Voice onset and offset were excluded, and a middle stationary segment with a length of 500 ms was chosen for analyses from each subject’s recording.

#### Acoustic Analysis

Traditional voice parameters of percent jitter, percent shimmer, and SNR were computed for normal and pathological voices using MDVP and CSpeech. Percent jitter is a cycle-to-cycle frequency perturbation measure, and percent shimmer is a cycle-to-cycle amplitude perturbation measure. MDVP and CSpeech algorithms requiring accurate extraction of fundamental frequency and amplitude have been designed to calculate these measures [10, 11].

SNR is the ratio between the total energy of a signal and the energy of the noise components of that signal. A lower SNR value suggests lower harmonic components and higher noise components of the signal. SNR of the voices was obtained with CSpeech.

Previous studies have suggested that perturbation analysis of aperiodic voices is unreliable, because peak-to-peak pitch periods may be inaccurately extracted for irregular signals. These studies suggest that only nearly periodic voices can be reliably analyzed by perturbation analysis [11, 12]. Titze [12] qualitatively classified voice signals into three types: type 1 signals are defined as nearly periodic, type 2 signals contain strong modulations or subharmonics, and type 3 signals are aperiodic. It is suggested that perturbation analysis is appropriate only for analysis of type 1 signals. Therefore, this study only applied perturbation measurements to nearly periodic voices. Three strongly aperiodic type 3 voices from the polyp group and one strongly aperiodic type 3 voice from the nodule group were excluded from perturbation analysis.

In contrast, nonlinear dynamic analysis does not rely upon accurate pitch extraction and thus can be applied to both periodic and aperiodic voices. A reconstructed phase space shows the dynamic behavior of a signal; a periodic signal produces a closed trajectory while an aperiodic signal produces an indiscriminate trajectory [23]. A qualitative distinction between phase space reconstructions of normal versus pathological voices can easily be made. The phase space reconstruction of a normal voice (fig. 1a) shows a closed loop of trajectory lines with a regular pattern. This reconstruction illustrates the regular, periodic waveforms presented in the normal voice. In contrast, the phase space reconstruction of a pathological voice (fig. 1b) shows scattered, diverg-
ing trajectory lines with an irregular pattern. This reconstruction illustrates the irregular, aperiodic waveforms exhibited by the pathological voice. Therefore, irregularity in a reconstructed phase space gives a qualitative indication of the pathological condition of a voice, while a regular pattern in the reconstructed phase space corresponds to a normal voice.

In order to quantify the irregularity of the reconstructed phase space, the algorithm proposed by Grassberger and Procaccia [24] was used to calculate the correlation dimension. Correlation dimension is a measure used to quantify the aperiodic behavior of a voice signal and is a geometric measure that describes how strongly two points are correlated in a phase space. A more complex and irregular system with a higher correlation dimension (e.g., a pathological voice) requires more variables to describe its behavior, while a simpler, regular system with a lower correlation dimension (e.g., a normal voice) requires fewer variables to describe its dynamic state. To ensure reliable dimension estimation, the standard deviation of the estimated value should be less than 5%. In this study, phase space reconstruction and correlation dimension calculation were performed using nonlinear dynamic analysis software developed by the Laryngeal Physiology Laboratory at the University of Wisconsin. Calculations made by the software were based on the numerical algorithms detailed in previous studies analyzing excised larynx phonations [22, 25], laryngeal modeling [26, 27], and pathological human voices [21, 28].

**Statistical Analysis**

Percent jitter, percent shimmer, SNR, and correlation dimension were compared for the normal, vocal nodule, and vocal polyp groups. Because it could not be predefined whether the tested groups were from normally distributed populations, we applied the Kruskal-Wallis one-way analysis of variance (ANOVA) on ranks. Percent jitter and percent shimmer from MDVP, percent jitter, percent shimmer, and SNR from CSpeech, and correlation dimension from the University of Wisconsin software were the dependent variables, and the subject groups (i.e., normal, nodule, and polyp) were the independent variables. Statistical significance level was set at $p = 0.05$. In order to determine how any two groups were statistically different, a post hoc Dunn test, used for multiple comparisons of groups with unequal sample size, was performed. The Dunn test has been shown to maintain familywise type I error at a minimum [29]. SigmaStat 3.0 (Jandel Scientific) software was used for statistical analysis.

**Results**

**Preliminary Analysis**

The voice samples used for preliminary analysis are representative of the groups from which they are drawn, providing estimation of typical acoustic analysis results and illustration of typical signal waveforms and frequency spectra for their respective cohorts. Figure 2 shows the typical waveforms and frequency spectra of both a normal voice and a pathological voice from a patient with vocal nodules. The normal voice showed a nearly periodic waveform and a discrete harmonic spectrum (fig. 2a). Using MDVP, percent jitter and percent shimmer for this normal voice were calculated as 0.49 and 0.55%, respectively. Using CSpeech, percent jitter, percent shimmer, and SNR were estimated as 0.19%, 0.62%, and 29.7 dB, respectively. Compared to the normal voice, the pathological voice showed an aperiodic waveform and a broad-band spectrum (fig. 2b). Because of the large error in determining voice pitch periods of this pathological voice signal, perturbation parameters cannot be reliably estimated using MDVP and CSpeech [11], and the use of nonlinear dynamic methods is necessary.

![Fig. 1.](image-url)

**Fig. 1.** a The reconstructed phase space of the normal voice. b The reconstructed phase space of the pathological voice from a patient with vocal nodules.
Figure 1 shows the results of nonlinear dynamic analysis for the normal voice and the pathological voice from a patient with vocal nodules. The reconstructed phase space in figure 1a shows the regular, closed trajectory structure of the normal voice. The estimated correlation dimension of this normal voice was calculated as $1.25 \pm 0.01$. In contrast with the normal voice, the irregular phase space of the pathological voice is shown in figure 1b. Preliminary analysis shows that with an estimated correlation dimension of $3.10 \pm 0.03$, the pathological voice has a higher dimension and greater complexity than the normal voice. To confirm this, comparisons among the three cohorts were made with all 11 healthy subjects with normal voices, 21 patients with vocal nodules, and 39 patients with vocal polyps.
Statistical Analysis

The results of the statistical analysis of percent jitter and percent shimmer from MDVP, percent jitter, percent shimmer, and SNR from CSpeech, and $D_2$ from the normal and pathological groups are shown in Table 1. Figures 3 and 4 show the distributions of percent jitter and percent shimmer, respectively, as derived from both MDVP and CSpeech. The distribution of SNR given by CSpeech is shown in Figure 5. The Kruskal-Wallis one-way ANOVA test revealed a statistically significant difference between groups for percent jitter from MDVP, percent jitter from CSpeech, and SNR ($p < 0.05$). A significant difference was also found between groups for percent shimmer from MDVP, but not for percent shimmer from CSpeech. Figure 6 illustrates the distribution of $D_2$ values. The Kruskal-Wallis one-way ANOVA performed on the $D_2$ values revealed a statistically significant difference between groups at the $p = 0.05$ significance level. Thus, the pathological voices demonstrated significantly higher dimensionality than the normal voices.

Table 1. Mean, standard deviation, and Kruskal-Wallis H values of percent jitter, percent shimmer, SNR and correlation dimension ($D_2$) from normal and pathological groups

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Vocal nodules</th>
<th>Vocal polyps</th>
<th>Kruskal-Wallis H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jitter, % (MDVP)</td>
<td>0.40±0.19</td>
<td>0.59±0.23</td>
<td>1.02±1.17</td>
<td>9.71*</td>
</tr>
<tr>
<td>Jitter, % (CSpeech)</td>
<td>0.18±0.13</td>
<td>0.23±0.13</td>
<td>0.65±0.60</td>
<td>19.00*</td>
</tr>
<tr>
<td>Shimmer, % (MDVP)</td>
<td>0.91±0.72</td>
<td>1.28±0.69</td>
<td>1.59±1.15</td>
<td>6.40*</td>
</tr>
<tr>
<td>Shimmer, % (CSpeech)</td>
<td>0.82±0.51</td>
<td>1.19±0.83</td>
<td>1.34±0.99</td>
<td>3.98</td>
</tr>
<tr>
<td>SNR (CSpeech)</td>
<td>31.2±3.08</td>
<td>29.3±3.67</td>
<td>25.5±8.2</td>
<td>11.95*</td>
</tr>
<tr>
<td>$D_2$</td>
<td>1.39±0.18</td>
<td>1.92±0.59</td>
<td>1.63±0.24</td>
<td>18.40*</td>
</tr>
</tbody>
</table>

The Kruskal-Wallis test statistics demonstrated statistically significant differences between groups for all parameters except percent shimmer from CSpeech. * $p < 0.05$. 

Fig. 3. Comparisons of distribution of percent jitter derived from MDVP and CSpeech among the normal, nodule and polyp groups. The line inside the box marks the median, whiskers show 10th and 90th percentiles, and the dots are the outlying points.

Fig. 4. Comparisons of distribution of percent shimmer derived from MDVP and CSpeech among the normal, nodule and polyp groups. The line inside the box marks the median, whiskers show 10th and 90th percentiles, and the dots are the outlying points.

Fig. 5. Comparisons of distribution of SNR among the normal, nodule and polyp groups. The line inside the box marks the median, whiskers show 10th and 90th percentiles, and the dots are the outlying points.
Results of the post hoc multiple comparison procedures using the Dunn test are shown in table 2. Percent jitter from MDVP and CSpeech, percent shimmer from MDVP, and SNR for the polyp group were significantly different from the normal group \((p < 0.05)\). However, all traditional parameters of the nodule group were not significantly different from either the normal or the polyp group \((p > 0.05)\). No significant differences between the normal, nodule, and polyp groups were found using percent shimmer from CSpeech \((p > 0.05)\). In contrast, correlation dimension values for both pathological groups were significantly higher than those for the normal group \((p < 0.05)\). This indicates the greater complexity of voices resulting from laryngeal mass lesions. No significant difference between vocal nodules and polyps was found for either perturbation or nonlinear dynamic parameters \((p > 0.05)\).

**Discussion**

Jitter, shimmer, and SNR have traditionally been used as objective parameters to provide noninvasive, quantitative assessment of vocal fold lesions, but these parameters have produced inconsistent results. Peppard et al. [2] found that patients with nodules have significantly different jitter values than the normal group, while the shimmer of the normal and pathological groups showed no significant difference. Similarly, Lin et al. [3] reported that percent jitter, but not percent shimmer, was successful in differentiating between a normal group and a vocal mass lesion group. In contrast, Rosen et al. [5] found a significant difference between normal and nodule groups in shimmer but not in jitter. Perturbation analysis applied to assess treatment effects for vocal fold lesions has also produced mixed results. Uloza [4] found significant decreases in both jitter and shimmer after surgical excision of vocal nodules and polyps, while Zeitels et al. [7] found that there was a significant decrease in shimmer measurements, but not in jitter measurements, after lesion excision. The contradictions in these results of mass lesion perturbation analysis may be attributable to differing population selection; however, the most conspicuous methodological dissimilarity among the studies is the application of different analysis systems and software for signal processing. Percent jitter and percent shimmer are dependent on pitch extraction algorithms and thus are sensitive to variations in analysis systems [11, 12, 30].

Furthermore, recent studies have suggested that jitter and shimmer might not be applicable to aperiodic voices, such as voices from patients with vocal mass lesions [12, 13, 31].

The present study corroborates previous studies; MDVP and CSpeech perturbation parameters show inconsistent results in differentiating normal voices from pathological voices. For example, a significant difference between normal and polyp groups can be found for percent shimmer from MDVP but cannot be found for percent shimmer from CSpeech (table 1). The discrepancy between MDVP and CSpeech may be a product of the different algorithms used by the two programs for ex-

---

**Table 2.** Post hoc Dunn test multiple comparisons between pairs of subject groups (normal, vocal nodules, and vocal polyps) on percent jitter, percent shimmer, SNR, and correlation dimension \((D_2)\) values

<table>
<thead>
<tr>
<th></th>
<th>Normal vs. vocal nodules</th>
<th>Vocal nodules vs. vocal polyps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jitter, % (MDVP)</td>
<td>18.67</td>
<td>0.68</td>
</tr>
<tr>
<td>Jitter, % (CSpeech)</td>
<td>8.99</td>
<td>24.80*</td>
</tr>
<tr>
<td>Shimmer, % (MDVP)</td>
<td>13.81</td>
<td>15.38*</td>
</tr>
<tr>
<td>Shimmer, % (CSpeech)</td>
<td>9.67</td>
<td>0.49</td>
</tr>
<tr>
<td>SNR (CSpeech)</td>
<td>12.77</td>
<td>22.51*</td>
</tr>
<tr>
<td>(D_2)</td>
<td>31.47*</td>
<td>19.73*</td>
</tr>
</tbody>
</table>

\* \(p < 0.05\).
traction of the pitch period. Moreover, perturbation analysis may be simply unreliable for characterization of the analyzed aperiodic voices from patients with vocal mass lesions. Thus, traditional perturbation methods demonstrate great incongruity in differentiating pathological voice from normal voice. Accordingly, these parameters should be applied with caution to aperiodic voices from laryngeal diseases.

This study demonstrates the objective, reliable results of using nonlinear dynamics to analyze pathological voices from patients with vocal nodules and polyps, further emphasizing the value of nonlinear dynamic methods in voice analysis. Differing from jitter and shimmer, correlation dimension describes the geometric properties of a voice and does not require determination of cycle periods. Correlation dimension values show a significant difference between normal and pathological voices. Perturbation and nonlinear dynamic analyses provide different but complementary information, and thus a combination of the two methods might provide more precise information than traditional perturbation analysis alone for clinical and research-based voice analysis.

Correlation dimension values of normal voices and voices from patients with vocal masses were finite (D2 < 4; fig. 6). This implies that voices from patients with vocal nodules and polyps have low-dimensional characteristics, indicating that finite degrees of freedom may be sufficient to describe the vibratory characteristics of such vocal folds. This is in agreement with the finite element simulation of Jiang et al. [26], where the vibrations of vocal folds with vocal nodules were dominated by the first few vibratory modes. The correlation dimension values of the pathological groups were significantly higher than those of the normal group, suggesting that the system variables may be needed to describe the dynamic state of voices from patients with vocal mass lesions than to describe the dynamic state of normal voices. This supports the recent computer model study of Zhang and Jiang [27], where more system variables were needed to describe the vibratory dynamics of vocal fold systems with vocal nodules or polyps. Nonlinear dynamic analysis of voices from patients with vocal nodules and polyps may help us examine modeling accuracy and increase our understanding of the predictions of computer models.

Nodules and polyps, which typically present as bilateral and unilateral lesions, respectively, represent different types of mass asymmetry within the vocal folds. Though no significant difference between vocal nodules and polyps was found for either perturbation or nonlinear dynamic parameters, with refinement, noninvasive, objective acoustic analysis might have a limited ability to predict types of anatomical asymmetries caused by laryngeal mass lesions, that is, to differentiate between vocal nodules and polyps. Other measurement methods for clinical voice assessment include perceptual evaluation, electroglottography, aerodynamic measurement, and high-speed imaging techniques. The abilities of these measurements to differentiate between the mass lesions of vocal nodules and polyps have not been completely determined, though different analysis methods provide unique information for the description of disordered voice production. Objective acoustic parameters, such as nonlinear dynamic analysis, should not replace existing techniques, but should serve as a complement to the array of existing voice analysis methods available to the clinician. The combination of traditional acoustic and nonlinear dynamic analyses could potentially improve our ability to objectively describe the acoustic properties of pathological voices from laryngeal diseases.

Conclusion

In this study, we applied nonlinear dynamic and traditional perturbation methods to analyze sustained vowels from normal subjects and patients with vocal mass lesions of vocal nodules and polyps. Jitter and shimmer from MDVP, as well as jitter and SNR from CSpeech, showed a significant difference between normal and polyp groups. Shimmer from CSpeech did not reveal any significant differences between any of the three cohorts. The correlation dimension values of the nodule and polyp groups were both significantly higher than those of the normal group. These results show that acoustic nonlinear dynamic analysis may improve objective characterization of laryngeal disorders and could be developed into a valuable approach for the clinical evaluation and diagnosis of laryngeal mass lesions.

Acknowledgment

This study was supported by NIH grant No. R01 DC006019 from the National Institute of Deafness and Other Communication Disorders.
References