Published in final edited form as:

J Voice. 2012 September ; 26(5): 566–576. doi:10.1016/j.jvoice.2011.09.006.

Nonlinear dynamic-based analysis of severe dysphonia in patients with vocal fold scar and sulcus vocalis

Seong Hee Choi, Ph.D.^{*}, Yu Zhang, Ph.D.[†], Jack J. Jiang, M.D., Ph.D.[‡], Diane M. Bless, Ph.D.[‡], and Nathan V. Welham, Ph.D.[‡]

^{*}Department of Audiology and Speech-Language Pathology, Catholic University of Daegu, Gyeongsan, Korea

[†]Key Laboratory of Underwater Acoustic Communication and Marine Information, Xiamen University, Xiamen, China

[‡]Division of Otolaryngology, Department of Surgery, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, U.S.A.

Abstract

Objective—The primary goal of this study was to evaluate a nonlinear dynamic approach to the acoustic analysis of dysphonia associated with vocal fold scar and sulcus vocalis.

Study Design—Case-control study.

Methods—Acoustic voice samples from scar/sulcus patients and age/sex-matched controls were analyzed using correlation dimension (D_2) and phase plots, time-domain based perturbation indices (jitter, shimmer, signal-to-noise ratio [SNR]), and an auditory-perceptual rating scheme. Signal typing was performed to identify samples with bifurcations and aperiodicity.

Results—Type 2 and 3 acoustic signals were highly represented in the scar/sulcus patient group. When data were analyzed irrespective of signal type, all perceptual and acoustic indices successfully distinguished scar/sulcus patients from controls. Removal of type 2 and 3 signals eliminated the previously identified differences between experimental groups for all acoustic indices except D_2 . The strongest perceptual-acoustic correlation in our dataset was observed for SNR; the weakest correlation was observed for D_2 .

Conclusions—These findings suggest that D_2 is inferior to time-domain based perturbation measures for the analysis of dysphonia associated with scar/sulcus; however, time-domain based algorithms are inherently susceptible to inflation under highly aperiodic (i.e., type 2 and 3) signal conditions. Auditory-perceptual analysis, unhindered by signal aperiodicity, is therefore a robust strategy for distinguishing scar/sulcus patient voices from normal voices. Future acoustic analysis research in this area should consider alternative (e.g., frequency- and quefrency-domain based) measures alongside additional nonlinear approaches.

Keywords

auditory-perceptual analysis; chaos; correlation dimension; jitter; perturbation analysis; phase plot; signal-to-noise ratio; signal typing; shimmer; voice disorder

Correspondence: Nathan V. Welham Ph.D., K4/723 CSC, 600 Highland Avenue, Madison WI 53792, U.S.A. Tel: 608-263-0121. welham@surgery.wisc.edu.

The authors hold no financial or other conflicts of interest.

Introduction

Vocal fold scar and sulcus vocalis are fibroplastic disorders of the vocal fold mucosa.¹ The etiology of these related pathologies is often unknown (and in the case of sulcus vocalis, controversial); however, both conditions can arise from traumatic and/or inflammatory events.²⁻⁴ The pathogenesis of scar and sulcus vocalis can irrevocably alter vocal function by disrupting the biomechanical properties of the vocal fold lamina propria extracellular matrix (ECM), leading to reduced tissue pliability, mucosal wave disruption and glottic insufficiency.⁴⁻⁷ These conditions represent significant diagnostic and treatment challenges and there is currently no consensus on their management.

Vocal fold scar and sulcus vocalis correspond to a broad spectrum of dysphonia severity, presumably as a function of anatomic presentation. Ford et al.⁷ classified sulcus vocalis into three anatomic subtypes. Physiologic (type I) sulcus is confined to the superficial lamina propria; whereas pathologic (types II and III) sulcus extends into the intermediate and deep lamina propria (vocal ligament). The severity of the dysphonia resulting from each subtype is believed to be dependant on the extent of mucosal contour and pliability disruption arising from the sulcus deformity, which in turn corresponds to the likelihood of successful voice restoration via surgical and/or behavioral management.

Reliable and valid voice assessment is essential to accurately diagnosing vocal fold scar and sulcus vocalis, directing treatment, and evaluating outcomes. Although both conditions can be associated with profound dysphonia and voice handicap,^{5,8} many of the approaches used to assess these disorders hold significant deficiencies. Acoustic analysis, a noninvasive approach widely used to objectively quantify voice quality, and auditory-perceptual analysis, another noninvasive approach often cited as the gold standard of voice assessment, are routinely utilized in combination;⁹⁻¹¹ however, questions concerning the reliability and validity of each approach have been raised.¹²⁻¹⁶ Further, the relationship between acoustic and auditory-perceptual voice parameters is complex and controversial.¹⁷⁻²⁰ Auditory-perceptual measures have often been determined to be unreliable and acoustic measures have accounted for only a low-to-moderate percentage of variance in perceptual judgments of voice quality.²¹⁻²³

Acoustic analysis

Much of the inadequacy of traditional acoustic analysis algorithms stems from their dependence on near-periodic signals.²⁴⁻²⁷ Aperiodicity and nonlinear signal bifurcations invalidate time-domain based perturbation measures such as jitter, shimmer and signal-to-noise ratio (SNR) by impeding accurate F_0 extraction; however, these signal features are valuable descriptors of disordered voices and should therefore be represented in their analysis. Titze, summarizing a consensus workshop on acoustic analysis, suggested the adoption of signal typing to ensure the appropriate analysis of normal and disordered voices.²⁵ Type 1 signals, defined as near-periodic, are suitable for analysis using time-domain based perturbation analyses. Type 2 signals, characterized by at least one qualitative signal bifurcation (e.g., subharmonic modulation), are best analyzed using visual displays such as spectrograms or reconstructed phase plots. Type 3 signals, defined by complete aperiodicity/chaos, are best analyzed using auditory-perceptual ratings or nonlinear dynamic parameters.

Consistent with Titze's²⁵ recommendations, non F_0 -dependent measures have shown increasing value in the analysis of aperiodic (i.e., type 2 and 3) voice signals. Frequency-domain implementations of the harmonics-to-noise ratio (HNR) are favorable in that they do not require the identification of individual period boundaries,^{28,29} whereas time-domain based HNR³⁰ and SNR³¹ measures do. Frequency-domain measures such as cepstral peak

prominence (CPP) hold a similar advantage and have shown good correspondence to overall perceived voice quality.³² Nonlinear dynamic parameters such as reconstructed phase plots, Lyapunov exponents and the correlation dimension (D_2) have also been successfully employed to characterize and quantify highly aperiodic voice signals produced by speakers with vocal polyps, Parkinson's disease, unilateral vocal fold paralysis, and esophageal voice.^{33–38} Compared to time-domain based perturbation indices, nonlinear dynamic approaches can be applied to both sustained vowels and connected speech, and are more forgiving in terms of analysis window length, sampling rate and ambient noise levels.³⁹ Nevertheless, nonlinear dynamic algorithms (along with traditional perturbation algorithms) remain vulnerable to breakdown when faced with voice signals containing a high stochastic noise component.⁴⁰

Auditory-perceptual analysis

Auditory-perceptual analysis holds significant face validity, is considered a necessary component of any comprehensive voice evaluation, and is commonly used in the validation of instrumental (e.g., acoustic) measures.^{18,41} As noted above, perceptual ratings are viewed as particularly useful in the description of highly aperiodic voice signals.²⁵ Previous studies using rating scales to evaluate voice quality have varied widely in methodology.^{42–44}

The primary challenge associated with the auditory-perceptual evaluation of voice quality stems from intra- and inter-rater variability driven by rater disagreement, differences in perceptual strategy, and response errors.^{13,16,45} This variability, which manifests irrespective of rater experience, has driven the argument that raters are perceptually idiosyncratic and therefore averaging responses across multiple raters may be inappropriate.¹² Shrivastav et al.⁴⁶ demonstrated that measurement error is a meaningful contributor to rating variability and that inter-rater agreement and reliability are enhanced by averaging ratings from multiple presentations of the same stimulus to each rater (compared to a single rating from each rater). By averaging multiple ratings, this psychometric theory-based approach to auditory-perceptual analysis improves scale resolution and measurement accuracy and minimizes random errors; potentially yielding more reliable and valid voice quality ratings.

Hypotheses

The primary goal of this study was to evaluate the suitability of a nonlinear dynamic approach to the acoustic analysis of disordered voice signals associated with vocal fold scar and sulcus vocalis. We performed acoustic signal typing followed by time-domain based perturbation (jitter, shimmer, SNR), nonlinear dynamic (D_2 and reconstructed phase plots) and auditory-perceptual (mean rating of overall voice quality per Shrivastav et al.⁴⁶) analyses. We compared the performance of each measurement index in separating scar/sulcus patient voices from sex- and age-matched controls and delineating patient subgroups; we also evaluated the association between each acoustic index and auditory-perceptual ratings. Given the profound dysphonia that often accompanies these conditions, we hypothesized that patients with scar and sulcus would predominantly exhibit type 2 and 3 voice signals. We further hypothesized that the nonlinear dynamic index D_2 would outperform the time-domain based perturbation indices in both the experimental group comparisons and perceptual-acoustic associations.

Materials and Methods

Participants

Twenty-three patients (11 males, 12 females; mean age = 55.74 years [$SD = 10.69$ years]) with a clinical diagnosis of vocal fold scar and/or pathologic sulcus vocalis participated in

this study. All participants were recruited with Institutional Review Board approval. The initial diagnosis and classification of vocal fold scar and/or pathologic sulcus vocalis were made by a laryngologist using videostroboscopic data collected by a speech-language pathologist, and were subsequently confirmed by the laryngologist during direct microlaryngoscopy. Ten patients presented with isolated vocal fold scar in the absence of sulcus, seven patients presented with sulcus vocalis in the absence of scar, and six patients presents with sulcus vocalis with concomitant scar (defined as any degree of reduced tissue pliability and apparent fibrosis in a region distinct from the concomitant sulcus). Patients with sulcus vocalis were classified according to the definition of Ford et al.;⁷ three patients presented with bilateral type I sulcus, ten patients presented with type II sulcus on at least one vocal fold, and one patient presented with unilateral type III sulcus. Our relatively low recruitment rate (23 patients over ~7 years) was a direct consequence of excluding patients who presented with concomitant laryngeal disease (e.g., vocal fold scar in the setting of recalcitrant human papillomaviral infection) and/or had undergone intervention for their scar/sulcus disorder prior to presentation at our institution.

Twenty-three age- and sex-matched controls (11 males, 12 females, mean age = 42.95 years [$SD = 5.85$ years]) with no history of voice or speech impairment were selected from a commercially available voice disorders database.⁴⁷ Control voice samples were elicited by production of a sustained [a] vowel token and directly digitized using a 50 kHz sampling rate and 16-bit quantization.

Acoustic recordings

Acoustic voice recordings from patients with vocal fold scar and sulcus vocalis were performed according to American and European assessment guidelines.^{24,25} All samples were collected in a sound-treated room under low ambient noise conditions. Patients were instructed to produce a 3–4 s sustained [a] vowel token at a comfortable pitch and loudness level. Voice samples were recorded using a unidirectional cardioid microphone (SM58; Shure, Niles, IL) placed 10 cm from the lips at a 45° angle, connected to a preamplifier (Bluetube DP; PreSonus, Baton Rouge, LA) and digital audio tape (DAT) recorder (Fostex D-5; Foster Electric, Schaumburg, IL). Digitization was performed using a 44.1 kHz sampling rate and 16-bit quantization. A digital copy of each sample was transferred from DAT to desktop computer and a 1 s steady-state segment was extracted from the midpoint of each vowel token using Praat 5.1.04.⁴⁸ Control voice samples were edited in the same manner and sample identity was masked prior to further analysis.

Auditory-perceptual analysis

Auditory-perceptual ratings were performed in a quiet room and samples were presented at ~70 dB SPL using headphones (HD 238; Sennheiser, Old Lyme, CT). Three doctoral-level speech-language pathologists with clinical specialization in voice disorders (7–40 years mean post-graduate clinical experience; no reported history of voice, speech, language or hearing difficulties) rated overall voice quality for all vowel samples using a seven-point equal-appearing-interval (EAI) rating scale (1: normal; 7: largest deviation from normal).

The auditory-perceptual rating task was structured as follows. Five playlists were constructed, each containing all 46 vowel samples (23 scar/sulcus, 23 control) in random order. To avoid the possibility of within-task variability, playlists were presented to each rater in the same sequence. Each rater controlled the rate of sample presentation and was able to pause as needed; however, once a judgment was made, raters moved to the next sample and were not allowed revise a previous rating. All ratings were completed in a single setting. A 5–10 min break was allowed between each of the five playlists.

Each vowel sample was rated five times by each rater. Repeat ratings of each sample by a single rater were averaged to minimize intra-rater perceptual variability, as previously reported.⁴⁶ Next, ratings were averaged across raters to obtain a mean perceptual rating of overall voice quality representative of each sample.

Acoustic analyses

Acoustic samples were first subjected to spectrographic analysis to evaluate signal type.²⁵ Narrowband spectrograms (500-ms window) were generated using Praat 5.1.04 and visually inspected to identify bifurcations, subharmonic modulations and/or chaotic segments. Spectrography and signal type classification were initially performed independently by one speech-language pathologist and one voice scientist; samples with disagreement on initial analysis were then reviewed together and discussed until consensus was reached. Type 1 signals were near-periodic with no evidence of bifurcation in the analysis window; type 2 signals had at least one clear bifurcation; type 3 signals appeared completely aperiodic. All samples were subjected to subsequent perturbation and nonlinear dynamic analyses irrespective of signal type; however, signal type (particularly type 1 vs. non-type 1) was accounted for when interpreting data and performing statistical comparisons between experimental groups.

Perturbation values (jitter, shimmer, SNR) were calculated using CSpeech 4.0 with implementation of previously reported algorithms.³¹ Calculation of all three indices was dependent on period estimation using a time-domain based autocorrelation function.

Signals were downsampled to 25 kHz prior to nonlinear dynamic analysis using GoldWave 5.1 (GoldWave, St. John's, NL, Canada). D_2 and reconstructed phase space plots were generated for each acoustic sample as previously reported.⁴⁹ D_2 is a quantitative measure specifying the number of degrees of freedom required to describe a dynamic system. Generally, a higher D_2 value corresponds to greater signal aperiodicity, increased system complexity and a greater number of degrees of freedom required for characterization.³⁴ Algorithms for D_2 calculation have been described extensively in previous studies^{34,50–53} and are summarized briefly here. Voice signals were first represented as $x(t_1), x(t_2), x(t_3), \dots$, at time interval Δt , where $t_i = t_0 + i\Delta t$ ($i = 1, 2, \dots, N$) and $N = 2.5 \times 10^4$. Phase plot reconstruction was achieved by plotting a series of points representing the voice signal against itself at a given time delay.⁵⁴ From the voice signal $x(t)$, we therefore had the following m -dimensional time delay vector:

$$\mathbf{X}(t) = \{x(t), x(t+\tau), \dots, x(t + [m-1]\tau)\} \quad (1)$$

where τ is the time delay and m is the embedding dimension. τ was calculated using Fraser and Swinney's⁵⁵ mutual information method. It has been previously shown that when $m > 2D + 1$, where D is the dimension of the attractor, the reconstructed phase space containing lagged coordinates is topologically equivalent to the original phase space containing physical coordinates.⁵⁶ Thus, the dynamics of a voice signal of interest can be studied with respect to its reconstructed phase space; a periodic voice signal is closed in its reconstructed phase space, whereas a chaotic voice signal is irregular and noise-like in its reconstructed phase space. Based on this reconstructed phase space, D_2 was then calculated according to Grassberger and Procaccia⁵⁷ as

$$D_2 = \lim_{r \rightarrow 0} \frac{\log C(r)}{\log r} \quad (2)$$

where r is the radius around $\mathbf{X}_j$ and C is the correlation integral, determined using Theiler's⁵⁸ formula

$$C(W, N, r) = \frac{2}{(N+1-W)(N-W)} \sum_{n=W}^{N-1} \sum_{i=0}^{N-1-n} \theta(r - \|\mathbf{X}_i - \mathbf{X}_{i+n}\|) \quad (3)$$

where the constant W is set as the value of time delay τ and $\theta(x)$ satisfies

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

As m increased, the slopes of the $\log C(W, N, r)$ and $\log r$ curves initially increased and then converged in the scaling region. This convergence was used to obtain D_2 , which was estimated by fitting a linear curve to r in the scaling region. Scaling regions were identified manually using visual inspection of the data (Figure 1); scaling region determination was unambiguous for all samples. The reliability of each D_2 value was determined using a SD measure; all SD values were consistently below 5%.

Statistical analyses

Statistical computations were performed using SigmaStat 2.0 (Jandel Scientific, San Rafael, CA) and R for Windows.⁵⁹ The Wilcoxon rank-sum test was used to compare control and scar/sulcus groups; the Kruskal-Wallis one-way analysis of variance on ranks was used to compare subgroups of patients with sulcus alone, scar alone, and sulcus with concomitant scar. These non-parametric tests were implemented as our acoustic dataset did not meet the assumptions of a normal distribution with equal variance across groups. Inter-rater auditory-perceptual agreement was evaluated using Pearson product-moment correlation coefficients and Bland-Altman analyses.⁶⁰ Intra-rater auditory-perceptual agreement was evaluated using intraclass correlation coefficients. Probability of exact agreement was calculated for both inter- and intra-rater agreement data, following the approach for EAI scale data described by Shrivastav et al.⁴⁶ Pearson product-moment correlation coefficients were used to evaluate the association between each acoustic index and mean perceptual rating of voice quality. Although collected using an EAI (ordinal) scale, mean perceptual rating data were considered ratio level data due to within-rater averaging, and therefore judged suitable for parametric (i.e., Pearson product-moment) correlation analyses. An α -level of 0.01 was employed for all statistical testing; all p -values were two-sided.

Results

Signal typing

Narrowband spectrographic analysis and signal typing of acoustic samples from patients with vocal fold scar/sulcus vocalis and controls revealed a complete array of type 1, 2 and 3 signals in the dataset (Figure 2). The majority of control group samples were near-periodic type 1 signals; however, three type 2 signals, characterized by at least one bifurcation (e.g., arrow in Figure 2B), were also identified (Figure 2D). In contrast, the majority of samples in the scar/sulcus group were type 2 signals; three type 1 and three completely aperiodic type 3 signals were identified (Figure 2D). Additional signal characterization using reconstructed phase space plots revealed generally closed trajectories for type 1 signals (Figure 3A) and progressively irregular and chaotic patterns for type 2 (Figure 3B) and 3 (Figure 3C) signals.

Auditory-perceptual analysis of scar/sulcus patients and controls

Mean auditory-perceptual ratings of overall voice quality impairment were significantly higher in the scar/sulcus group compared to controls (Figure 4A: $p < 0.001$). Evaluation of the quality of mean ratings across raters was performed using a correlation matrix (Table 1) and Bland-Altman analyses (Figure 4B). Positive correlations were observed between all combinations of raters and ranged from $r = 0.84$ (rater 1 vs. rater 3) to $r = 0.92$ (rater 1 vs. rater 2). Positive correlations were also observed for each rater against the group mean and

ranged from $r = 0.94$ (rater 3) to $r = 0.97$ (rater 2). Mean bias between all possible pairings of the three raters ranged from -1.14 to 0.42 EAI scale intervals (95% CI: -1.46 to 0.71). The 95% limits of agreement ranged from 3.00 (rater 1 vs. 2) to 4.23 (rater 1 vs. rater 3) scale intervals. The probability of exact agreement across all three raters was 0.65 .

Evaluation of intra-rater agreement across the five repeated ratings performed by each rater revealed intraclass correlation coefficients of 0.94 , 0.93 and 0.93 for rater 1, 2 and 3 respectively. Probabilities of exact agreement were 0.79 , 0.80 and 0.80 for rater 1, 2 and 3 respectively.

Performance of perturbation and nonlinear dynamic indices in distinguishing scar/sulcus patients from controls

Acoustic perturbation and D_2 values were tightly clustered in the control group and demonstrated greater variance in the scar/sulcus group (Figure 5). Within the scar/sulcus group, type 1 signals were characterized by relatively low perturbation (low jitter and shimmer, high SNR) and low-to-midrange D_2 values, type 2 signals were characterized by midrange perturbation and D_2 values, and type 3 signals occupied relatively high perturbation (high jitter and shimmer, low SNR), and D_2 values (Figure 5). When analyzed by experimental group irrespective of signal type, all acoustic indices demonstrated significant deterioration in the scar/sulcus group compared with the control group (Figure 5: $p < 0.001$ for jitter; $p = 0.002$ for shimmer, $p < 0.001$ for SNR; $p < 0.001$ for D_2). Parallel analysis using only type 1 signals sharply reduced the scar/sulcus group n from 23 to 3 (the control group n was reduced from 23 to 20). As a result, no significant differences were observed between experimental groups for the three perturbation indices (Figure 5A–C: $p = 0.09$ for jitter; $p = 0.34$ for shimmer; $p = 0.09$ for SNR); whereas, a significant difference was preserved for D_2 (Figure 5D: $p = 0.009$).

Performance of auditory-perceptual and acoustic indices in distinguishing scar/sulcus patient subgroups

The auditory-perceptual and acoustic indices all demonstrated mean progressive deterioration in subgroups of patients with sulcus vocalis only (mean perceptual rating $M = 3.16$, $SD = 1.21$; jitter $M = 0.76$, $SD = 0.70$; shimmer $M = 2.25$, $SD = 1.61$; SNR $M = 22.39$, $SD = 4.85$; D_2 $M = 1.57$, $SD = 0.29$), followed by vocal fold scar only (mean perceptual rating $M = 4.11$, $SD = 1.37$; jitter $M = 1.41$, $SD = 1.60$; shimmer $M = 5.16$, $SD = 4.07$; SNR $M = 17.88$, $SD = 5.91$; D_2 $M = 2.16$, $SD = 0.51$), followed by sulcus vocalis with concomitant scar (mean perceptual rating $M = 4.66$, $SD = 2.57$; jitter $M = 1.45$, $SD = 1.48$; shimmer $M = 8.00$, $SD = 8.38$; SNR $M = 14.85$, $SD = 7.09$; D_2 $M = 2.87$, $SD = 1.48$); however, no subgroup comparison reached statistical significance (Figure 6: $p = 0.24$ for mean perceptual rating; $p = 0.52$ for jitter; $p = 0.10$ for shimmer; $p = 0.06$ for SNR; $p = 0.02$ for D_2). Type 1 and 2 signals were represented in all three patient subgroups; type 3 signals were only identified in patients with vocal fold scar (with or without concomitant sulcus vocalis). Analysis of perturbation measures using type 1 signals only was not performed as each patient subgroup contained only one type 1 signal.

Associations between auditory-perceptual ratings and acoustic indices

Associations between mean perceptual rating of overall voice quality and jitter, shimmer, SNR and D_2 were evaluated using scatter plots and correlation analyses. Data from scar/sulcus patients and controls were pooled for these analyses and associations were best represented using logarithmic scaling (Figure 7). Severity of mean perceptual rating and acoustic index corresponded generally to signal type. When all 46 voice samples were considered to predict the perceptual-acoustic relationship, the strongest correlation was observed for SNR (Figure 7C: $r = -0.88$) and the weakest (but still moderate) correlation

was observed for D_2 (Figure 7D: $r = 0.68$). Parallel analysis using only type 1 signals for the perturbation comparisons reduced the n for each comparison to 23 and sharply decreased the correlation coefficient for each acoustic index (Figure 7: $r = 0.52$ for jitter ; $r = 0.25$ for shimmer; $r = -0.44$ for SNR; $r = 0.37$ for D_2).

Discussion

The primary goal of this study was to evaluate the suitability of a nonlinear dynamic approach to the acoustic analysis of disordered voice signals associated with vocal fold scar and sulcus vocalis. We compared patient voice samples with those from age- and sex-matched controls using D_2 and reconstructed phase plots, time-domain based perturbation indices, and an auditory-perceptual rating scheme. We also performed acoustic signal typing to identify samples with bifurcations and apparent aperiodicity; qualitative signal features that have been demonstrated to invalidate time-domain based perturbation estimates.²⁵ Type 2 and 3 signals with irregular and chaotic phase plots were highly represented in the scar/sulcus patient group. When voice samples were analyzed irrespective of signal type, all perceptual and acoustic indices successfully distinguished scar/sulcus patients from controls, but did not separate patient subgroups. Removal of type 2 and 3 signals from the analysis led to near complete depletion of the scar/sulcus patient group n , eliminating the previously identified differences between experimental groups for all acoustic indices except D_2 . Evaluation of the relationship between auditory-perceptual ratings and acoustic indices revealed moderate-to-strong correlations for all measures; the highest correlation coefficient was observed for SNR. Removal of type 2 and 3 signals sharply reduced the correlation strength for all acoustic indices.

Overall, our findings supported our first hypothesis that patients with scar and pathologic sulcus vocalis predominantly exhibit type 2 and 3 voice signals, but failed to support our second hypothesis that D_2 is superior to time-domain based perturbation indices in separating scar/sulcus patients from controls and correspondence to auditory-perceptual ratings of voice quality.

Applicability of signal typing, time-domain based perturbation and nonlinear dynamic measures

Acoustic analysis has dominated the voice assessment literature for the past 40 years and has led to the development of a plethora of measures. Periodicity-based measures have long been considered requisite to any comprehensive voice signal analysis; however, many algorithms are unable to accurately characterize signals with high levels of perturbation. Generally, pathologic voice signals can be characterized as having *aperiodicity features* and *noise features*.⁶¹ Aperiodicity features derive from cycle-to-cycle variations in vocal fold vibratory behavior and have been used to define classic parameters such as jitter (cycle-to-cycle variation in period duration) and shimmer (cycle-to-cycle variation in peak amplitude or energy).⁶² Noise features are thought to arise from certain physiologic conditions that can present during the closed phase of the vibratory cycle, such as a glottal gap or incomplete glottal closure.⁶³ These features, which encompass collective perturbations in period, amplitude and waveform morphology, are captured by SNR and related measures (e.g., HNR).

Although jitter, shimmer and time-domain based SNR/HNR indices have been widely adopted as measures of voice signal perturbation, persistent concerns about their lack of reliability under certain conditions have led to the development of practical application guidelines. Titze and Liang concluded that perturbation values in excess of 5% are generally unreliable.⁶⁴ Likewise, Milenkovic and Read suggested that greater than 10 missed pitch period estimates per sample token (termed the *Err value* within CSpeech) is indicative of

poor F_0 extraction and therefore questionable waveform suitability for perturbation analysis.⁶⁵ Titze recommended reserving perturbation analyses for near-periodic type 1 signals identified by qualitative signal typing.²⁵ Signal typing is a potentially useful strategy for profiling phonatory dynamics in any sample of interest, has been successfully used in a number of studies,^{66,67} and was therefore implemented with our dataset.

Predetermination of analysis strategy using signal typing played a significant role in our findings, particularly in the scar/sulcus patient group. When all data were analyzed irrespective of signal type, jitter, shimmer and SNR performed equally as well as D_2 in distinguishing patients from controls, and outperformed D_2 in correlation analyses against mean perceptual rating of voice quality. The removal of type 2 and 3 signals from the analysis set dramatically reversed these findings. This observation is not surprising given the predominance of type 2 and 3 signals in the scar/sulcus patient dataset and the fact that the small number of remaining type 1 signals were near-periodic. Also, this finding was unquestionably influenced by the sharp decrease in n from 23 to 3 for the scar/sulcus and control group comparisons, and from 46 to 23 for the perceptual-acoustic comparisons. Our decision to deplete type 2 and 3 signals from both the perturbation and D_2 analyses (even though the D_2 algorithm is robust irrespective of signal type) was purposeful as it ensured parallel experimental treatment of each acoustic measure.

Although it may appear from our results that applying time-based perturbation measures to all voice signals is advantageous (particularly given their robust initial correlations with mean perceptual rating of voice quality), it is important to note that the values obtained from these algorithms under highly aperiodic (i.e., type 2 and 3) signal conditions are inherently susceptible to inflation due to inaccurate period boundary estimation, and are therefore quantitatively unreliable.^{25,64} Stated otherwise, while these measures may indicate a high level of perturbation in cases of severe dysphonia, they do not accurately represent the magnitude of perturbation and therefore should not be used to quantify treatment-induced change.

In contrast with time-domain based measures, a number of frequency-domain, quefrency-domain and nonlinear dynamic strategies are not invalidated by temporal instability and are therefore theoretically applicable to severely aperiodic and chaotic vibratory signals, such as those identified in our patient dataset. In this study, we elected to focus on the nonlinear dynamic measure D_2 , given its reported ability to differentiate a wide range of pathologic voices from normal controls;^{36,37,68} however, other measures such as the frequency-domain based HNR^{28,29} and the quefrency-domain based CPP³² are equally worthy of consideration. Given the superior auditory-perceptual correspondence of the time-domain based SNR measure in our dataset, the application of frequency-domain based HNR may be particularly relevant to the analysis of voice signals from patients with scar/sulcus, as it would capture the collective noise features of the signal (an inherent advantage of SNR) while eliminating algorithmic dependence on period boundary estimation (an inherent disadvantage of SNR).

Auditory-perceptual and acoustic relationships

We derived auditory-perceptual ratings of overall voice quality by averaging multiple ratings of the same stimulus presented to three independent raters. The decision to employ an overall voice quality metric came from research indicating that listeners' ability to detect individual elements or dimensions (e.g., roughness and breathiness) in a multidimensional voice signal is poor^{12,13} and that overall severity exhibits some correspondence to acoustic measures.^{23,69} The decision to employ averaging of multiple ratings was centered on Shrivastav et al.'s psychometric theory-based approach demonstrating reduced variability and improved reliability compared with single ratings.⁴⁶ Based on their findings, Shrivastav

et al. argued that the use of multiple average ratings can improve measurement accuracy and reduce random errors.

The strongest perceptual-acoustic correlation in our dataset was observed for SNR ($r = -0.88$); the strongest correlation following removal of type 2 and 3 signals was observed for jitter ($r = 0.52$). Many studies have attempted to examine the relationships between select acoustic parameters and perceptual dimensions; however, results have generally been inconsistent across studies and most parameters have demonstrated low sensitivity and specificity to particular disorders.^{21–23,70} Heman-Ackah et al. identified smoothed CPP as superior to time-domain perturbation indices when predicting overall voice quality, particularly during connected speech.^{71,72} This finding was confirmed in a recent meta-analysis focused on the relationship of acoustic measures to perceived overall voice quality,³² which concluded that acoustic parameters that are not dependent on F_0 extraction (such as smoothed CPP) hold stronger relationships with perceptual judgments than those that are F_0 dependent, in both vowels and connected speech. In light of this literature, it is surprising we did not observe greater correspondence between overall perceived voice quality and D_2 .

Biological sources of signal aperiodicity

Our data do not directly inform the etiology of the signal aperiodicity observed in the scar/sulcus patient group. It is likely, however, that the predominance of type 2 and 3 signals, elevated perturbation and D_2 values in these patients reflect impaired vocal fold anatomic and viscoelastic deficiencies. Studies of humans and animals with chronic vocal fold scar and pathologic sulcus vocalis suggest that the most common physiologic consequences of lamina propria ECM disruption in these disorders are blunted vibratory amplitude and mucosal wave, asymmetric and aperiodic cycle-to-cycle vibratory patterning, and impaired vibratory closure;^{5,73} all features which may correspond to significant cycle-to-cycle aperiodicity, bifurcations and chaos in the acoustic signal. Vocal fold vibratory asymmetry is a well-documented physiologic mechanism of voice signal irregularity and nonlinearity,^{52,53} and can manifest in the form of desynchronized left-right, anterior-posterior or even superior-inferior vibratory modes.^{74–79} Further, certain vibratory modes correspond to (and are enhanced by) natural vocal tract resonant frequencies.^{75,80} Given that the histopathologic features of vocal fold scar and sulcus vocalis in humans are generally isolated to discrete (and in certain cases unilateral) tissue regions, the increased dynamic complexity and nonlinear features observed in our scar/sulcus dataset may reflect vibratory asymmetries driven by vocal fold contour, viscoelastic, mass and tension asymmetries. In the future it may be possible to use new *in vivo* imaging tools coupled with nonlinear dynamic measurements of acoustic and/or movement behaviors to determine how increased dynamic complexity directly relates to biological tissue change.

Limitations and future research directions

This study holds several limitations. First, although vocal fold scar and sulcus vocalis patients are a valuable source of aperiodic voice signals and represent a worthy analytical challenge, their clinical presentation is often highly heterogeneous (i.e., a scar or sulcus deformity can present in a number of vocal fold tissue regions and might be unilateral or bilateral). Sulcus vocalis is also a relatively rare condition that does not feature in most diagnostic studies of the treatment-seeking population;^{81–83} Our primary analyses considered all patients as a single experimental group; however, subgroups of these patients have been shown to exhibit different analytic features.^{5,8} In this study, sample size may have contributed to our non-significant findings in subgroups of patients with sulcus only, scar only and sulcus with concomitant scar. Further, limited sample size did not allow us to

consider other subgroups of interest, such as sulcus vocalis subtype or unilateral versus bilateral pathology.

Second, given the variety of commercial and non-commercial algorithms used for F_0 extraction and perturbation measurement, it is possible that our findings may not directly correspond to all analysis packages. This limitation is not unique to this study and unfortunately represents a persistent challenge in all acoustic voice analyses.⁸⁴ Whereas we analyzed sustained vowel tokens in order to be comparable to the majority of the existing literature, the analysis of connected speech samples is an attractive future research direction given the dynamic complexity and inherent face validity of voice use in connected speech. Sustained vowel tokens are requisite input for most commercial voice analysis packages as standard perturbation algorithms cannot analyze the more complex signals seen in connected speech; in contrast, a number of measures such as smoothed CPP,^{71,72} D_2 ,³⁹ and certain implementations of SNR⁷⁰ are suitable for the analysis of connected speech and may help resolve some of the heretofore puzzles and paradoxes in the acoustic analysis of severely dysphonic voices.

Finally, future research focused on the physiologic source of the nonlinear dynamic features described here should pursue parallel analyses of aperiodic voice signals and vocal fold vibratory behaviors. Just as nonlinear dynamic approaches to the acoustic analysis of voice draw a significant analytic advantage from F_0 independence, high speed digital imaging (HSDI) can continue to track vocal fold vibratory motion under highly aperiodic conditions that are problematic for videostroboscopy.⁸⁵ As with acoustic signals, HSDI-derived movement trajectories can be analyzed using tools such as D_2 and reconstructed phase plots; simultaneous data collection and parallel analyses may provide direct evidence for the vibratory events that underpin acoustic events such as those reported here.

Conclusions

Patients with scar and pathologic sulcus vocalis predominantly exhibit type 2 and 3 voice signals. D_2 is inferior to time-domain based perturbation measures for the analysis of dysphonia associated with scar/sulcus; however, time-domain based algorithms are inherently susceptible to inflation under highly aperiodic (i.e., type 2 and 3) signal conditions. Auditory-perceptual analysis, unhindered by signal aperiodicity, is therefore a robust strategy for distinguishing scar/sulcus patient voices from normal voices. Future acoustic analysis research in this area should consider alternative (e.g., frequency- and quefrency-domain based) measures alongside additional nonlinear approaches.

Acknowledgments

This study was supported by grants R01 DC004428 and R01 DC006019 from the National Institute on Deafness and Other Communication Disorders. We gratefully acknowledge Seth H. Dailey, M.D. and Charles N. Ford, M.D. for performing laryngological diagnoses, Alejandro Muñoz del Río, Ph.D. for assistance with statistical analysis, and Rahul Shrivastav, Ph.D. for consultation regarding auditory-perceptual analysis.

References

1. Dailey SH, Ford CN. Surgical management of sulcus vocalis and vocal fold scarring. *Otolaryngol Clin North Am.* 2006; 39:23–42. [PubMed: 16469653]
2. Itoh T, Kawasaki H, Morikawa I, Hirano M. Vocal fold furrows. A 10-year review of 240 patients. *Auris Nasus Larynx.* 1983; 10(Suppl):S17–S26. [PubMed: 6651651]
3. Bouchayer M, Cornut G, Witzig E, Loire R, Roch JB, Bastian RW. Epidermoid cysts, sulci, and mucosal bridges of the true vocal cord: a report of 157 cases. *Laryngoscope.* 1985; 95:1087–1094. [PubMed: 4033333]

4. Benninger MS, Alessi D, Archer S, et al. Vocal fold scarring: current concepts and management. *Otolaryngol Head Neck Surg.* 1996; 115:474–482. [PubMed: 8903451]
5. Hirano M, Yoshida T, Tanaka S, Hibi S. Sulcus vocalis: functional aspects. *Ann Otol Rhinol Laryngol.* 1990; 99:679–683. [PubMed: 2396804]
6. Hansen JK, Thibeault SL. Current understanding and review of the literature: Vocal fold scarring. *J Voice.* 2006; 20:110–120. [PubMed: 15964741]
7. Ford CN, Inagi K, Khidr A, Bless DM, Gilchrist KW. Sulcus vocalis: a rational analytical approach to diagnosis and management. *Ann Otol Rhinol Laryngol.* 1996; 105:189–200. [PubMed: 8615582]
8. Welham NV, Dailey SH, Ford CN, Bless DM. Voice handicap evaluation of patients with pathologic sulcus vocalis. *Ann Otol Rhinol Laryngol.* 2007; 116:411–417. [PubMed: 17672242]
9. Welham NV, Rousseau B, Ford CN, Bless DM. Tracking outcomes after phonosurgery for sulcus vocalis: a case report. *J Voice.* 2003; 17:571–578. [PubMed: 14740937]
10. Pinto JA, da Silva Freitas ML, Carpes AF, Zimath P, Marquis V, Godoy L. Autologous grafts for treatment of vocal sulcus and atrophy. *Otolaryngol Head Neck Surg.* 2007; 137:785–791. [PubMed: 17967647]
11. Hsiung MW, Kang BH, Pai L, Su WF, Lin YH. Combination of fascia transplantation and fat injection into the vocal fold for sulcus vocalis: Long-term results. *Ann Otol Rhinol Laryngol.* 2004; 113:359–366. [PubMed: 15174762]
12. Kreiman J, Gerratt BR. Validity of rating scale measures of voice quality. *J Acoust Soc Am.* 1998; 104:1598–1608. [PubMed: 9745743]
13. Kreiman J, Gerratt BR. Sources of listener disagreement in voice quality assessment. *J Acoust Soc Am.* 2000; 108:1867–1876. [PubMed: 11051513]
14. Kreiman J, Gerratt BR, Ito M. When and why listeners disagree in voice quality assessment tasks. *J Acoust Soc Am.* 2007; 122:2354–2364. [PubMed: 17902870]
15. Oates J. Auditory-perceptual evaluation of disordered voice quality: pros, cons and future directions. *Folia Phoniatr Logop.* 2009; 61:49–56. [PubMed: 19204393]
16. Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *J Speech Hear Res.* 1993; 36:21–40. [PubMed: 8450660]
17. Shrivastav R, Sapienza CM. Objective measures of breathy voice quality obtained using an auditory model. *J Acoust Soc Am.* 2003; 114:2217–2224. [PubMed: 14587619]
18. Rabinov CR, Kreiman J, Gerratt BR, Bielamowicz S. Comparing reliability of perceptual ratings of roughness and acoustic measure of jitter. *J Speech Hear Res.* 1995; 38:26–32. [PubMed: 7731216]
19. Kreiman J, Gerratt BR, Precoda K. Listener experience and perception of voice quality. *J Speech Hear Res.* 1990; 33:103–115. [PubMed: 2314068]
20. Bhuta T, Patrick L, Garnett JD. Perceptual evaluation of voice quality and its correlation with acoustic measurements. *J Voice.* 2004; 18:299–304. [PubMed: 15331102]
21. Wolfe V, Cornell R, Palmer C. Acoustic correlates of pathologic voice types. *J Speech Hear Res.* 1991; 34:509–516. [PubMed: 2072674]
22. Kempster GB, Kistler DJ, Hillenbrand J. Multidimensional scaling analysis of dysphonia in two speaker groups. *J Speech Hear Res.* 1991; 34:534–543. [PubMed: 2072677]
23. Eskenazi L, Childers DG, Hicks DM. Acoustic correlates of vocal quality. *J Speech Hear Res.* 1990; 33:298–306. [PubMed: 2359270]
24. Dejonckere PH, Bradley P, Clemente P, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). *Eur Arch Otorhinolaryngol.* 2001; 258:77–82. [PubMed: 11307610]
25. Titze, IR. Workshop on acoustic voice analysis: Summary statement. National Center for Voice and Speech; Iowa City, IA: 1995.
26. Ma EP, Yiu EM. Suitability of acoustic perturbation measures in analysing periodic and nearly periodic voice signals. *Folia Phoniatr Logop.* 2005; 57:38–47. [PubMed: 15655340]

27. Karnell MP, Chang A, Smith A, Hoffman HT. Impact of signal type on validity of voice perturbation measures. *Natl Cent Voice Speech Status Prog Rep*. 1997; 11:91–94.
28. de Krom G. A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *J Speech Hear Res*. 1993; 36:254–266. [PubMed: 8487518]
29. Qi Y, Hillman RE. Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals. *J Acoust Soc Am*. 1997; 102:537–543. [PubMed: 9228816]
30. Yumoto E, Gould WJ, Baer T. Harmonics-to-noise ratio as an index of the degree of hoarseness. *J Acoust Soc Am*. 1982; 71:1544–1550. [PubMed: 7108029]
31. Milenkovic P. Least mean square measures of voice perturbation. *J Speech Hear Res*. 1987; 30:529–538. [PubMed: 2961933]
32. Maryn Y, Roy N, De Bodt M, Van Cauwenberge P, Corthals P. Acoustic measurement of overall voice quality: a meta-analysis. *J Acoust Soc Am*. 2009; 126:2619–2634. [PubMed: 19894840]
33. Giovanni A, Ouaknine M, Triglia JM. Determination of largest Lyapunov exponents of vocal signal: application to unilateral laryngeal paralysis. *J Voice*. 1999; 13:341–354. [PubMed: 10498051]
34. Jiang JJ, Zhang Y, McGilligan C. Chaos in voice, from modeling to measurement. *J Voice*. 2006; 20:2–17. [PubMed: 15964740]
35. Landini L, Manfredi C, Positano V, Santarelli MF, Vanello N. Non-linear prediction for oesophageal voice analysis. *Med Eng Phys*. 2002; 24:529–533. [PubMed: 12237049]
36. MacCallum JK, Cai L, Zhou L, Zhang Y, Jiang JJ. Acoustic analysis of aperiodic voice: perturbation and nonlinear dynamic properties in esophageal phonation. *J Voice*. 2009; 23:283–290. [PubMed: 18411036]
37. Rahn DA 3rd, Chou M, Jiang JJ, Zhang Y. Phonatory impairment in Parkinson's disease: evidence from nonlinear dynamic analysis and perturbation analysis. *J Voice*. 2007; 21:64–71. [PubMed: 16377130]
38. Zhang Y, Jiang J, Rahn DA 3rd. Studying vocal fold vibrations in Parkinson's disease with a nonlinear model. *Chaos*. 2005; 15:33903. [PubMed: 16252994]
39. Zhang Y, Jiang JJ. Acoustic analyses of sustained and running voices from patients with laryngeal pathologies. *J Voice*. 2008; 22:1–9. [PubMed: 16978835]
40. Sprecher A, Olszewski A, Jiang JJ, Zhang Y. Updating signal typing in voice: addition of type 4 signals. *J Acoust Soc Am*. 2010; 127:3710–3716. [PubMed: 20550269]
41. Eadie TL, Doyle PC. Classification of dysphonic voice: acoustic and auditory-perceptual measures. *J Voice*. 2005; 19:1–14. [PubMed: 15766846]
42. Wuyts FL, De Bodt MS, Van de Heyning PH. Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *J Voice*. 1999; 13:508–517. [PubMed: 10622517]
43. Wolfe V, Fitch J, Cornell R. Acoustic prediction of severity in commonly occurring voice problems. *J Speech Hear Res*. 1995; 38:273–279. [PubMed: 7596093]
44. Eadie TL, Doyle PC. Direct magnitude estimation and interval scaling of pleasantness and severity in dysphonic and normal speakers. *J Acoust Soc Am*. 2002; 112:3014–3021. [PubMed: 12509023]
45. Kreiman J, Gerratt BR, Berke GS. The multidimensional nature of pathologic vocal quality. *J Acoust Soc Am*. 1994; 96:1291–1302. [PubMed: 7962996]
46. Shrivastav R, Sapienza CM, Nandur V. Application of psychometric theory to the measurement of voice quality using rating scales. *J Speech Lang Hear Res*. 2005; 48:323–335. [PubMed: 15989395]
47. Kay Elemetrics. Voice disorders database [CD-ROM]. Version 1.03. Author; Lincoln Park, NJ: 1994.
48. Boersma, P.; Weenink, D. Praat: Doing phonetics by computer [computer program]. Version 5.1.04. <http://praat.org/>.
49. Jiang JJ, Zhang Y. Nonlinear dynamic analysis of speech from pathologic subjects. *Electron Lett*. 2002; 38:294–295.
50. Kumar A, Mullick SK. Nonlinear dynamical analysis of speech. *J Acoust Soc Am*. 1996; 100:615–629.

51. Zhang Y, McGilligan C, Zhou L, Vig M, Jiang JJ. Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps. *J Acoust Soc Am*. 2004; 115:2270–2277. [PubMed: 15139638]
52. Jiang JJ, Zhang Y, Ford CN. Nonlinear dynamics of phonations in excised larynx experiments. *J Acoust Soc Am*. 2003; 114:2198–2205. [PubMed: 14587617]
53. Herzel H, Berry D, Titze IR, Saleh M. Analysis of vocal disorders with methods from nonlinear dynamics. *J Speech Hear Res*. 1994; 37:1008–1019. [PubMed: 7823547]
54. Packard NH, Crutchfield JP, Farmer JD, Shaw RS. Geometry from a time series. *Phys Rev Lett*. 1980; 45:712.
55. Fraser AM, Swinney HL. Independent coordinates for strange attractors from mutual information. *Phys Rev A*. 1986; 33:1134–1140. [PubMed: 9896728]
56. Takens, F. Detecting strange attractors in turbulence. In: Rand, DA.; Young, LS., editors. *Lecture notes in mathematics*. Springer-Verlag; Berlin: 1981. p. 366-381.
57. Grassberger P, Procaccia I. Measuring the strangeness of strange attractors. *Physica D*. 1983; 9:189–208.
58. Theiler J. Spurious dimension from correlation algorithms applied to limited time series data. *Phys Rev A*. 1986; 34:2427–2432. [PubMed: 9897530]
59. R Development Core Team. [Accessed June 1 2009] R: A language and environment for statistical computing [computer program]. <http://www.R-project.org/>.
60. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 327:307–310. [PubMed: 2868172]
61. Michaelis D, Fröhlich M, Strube HW. Selection and combination of acoustic features for the description of pathologic voices. *J Acoust Soc Am*. 1998; 103:1628–1639. [PubMed: 9514027]
62. Pinto NB, Titze IR. Unification of perturbation measures in speech signals. *J Acoust Soc Am*. 1990; 87:1278–1289. [PubMed: 2324395]
63. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J Speech Hear Res*. 1994; 37:769–778. [PubMed: 7967562]
64. Titze IR, Liang H. Comparison of Fo extraction methods for high-precision voice perturbation measurements. *J Speech Hear Res*. 1993; 36:1120–1133. [PubMed: 8114479]
65. Milenkovic, P.; Read, C. *CSpeech version 4 user's manual*. University of Wisconsin-Madison; Madison, WI: 1992.
66. Behrman A, Agresti CJ, Blumstein E, Lee N. Microphone and electroglottographic data from dysphonic patients: type 1, 2 and 3 signals. *J Voice*. 1998; 12:249–260. [PubMed: 9649081]
67. van As-Brooks CJ, Koopmans-van Beinum FJ, Pols LC, Hilgers FJ. Acoustic signal typing for evaluation of voice quality in tracheoesophageal speech. *J Voice*. 2006; 20:355–368. [PubMed: 16185840]
68. Zhang Y, Jiang JJ, Biazzo L, Jorgensen M. Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis. *J Voice*. 2005; 19:519–528. [PubMed: 16301098]
69. Wolfe V, Martin D. Acoustic correlates of dysphonia: Type and severity. *J Commun Disord*. 1997; 30:403–415. [PubMed: 9309531]
70. Qi Y, Hillman RE, Milstein C. The estimation of signal-to-noise ratio in continuous speech for disordered voices. *J Acoust Soc Am*. 1999; 105:2532–2535. [PubMed: 10212434]
71. Heman-Ackah YD, Michael DD, Goding GS Jr. The relationship between cepstral peak prominence and selected parameters of dysphonia. *J Voice*. 2002; 16:20–27. [PubMed: 12008652]
72. Heman-Ackah YD, Heuer RJ, Michael DD, et al. Cepstral peak prominence: A more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol*. 2003; 112:324–333. [PubMed: 12731627]
73. Welham NV, Montequin DW, Tateya I, Tateya T, Choi SH, Bless DM. A rat excised larynx model of vocal fold scar. *J Speech Lang Hear Res*. 2009; 52:1008–1020. [PubMed: 19641079]
74. Berry DA, Herzel H, Titze IR, Krischer K. Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions. *J Acoust Soc Am*. 1994; 95:3595–3604. [PubMed: 8046149]
75. Berry DA. Mechanisms of modal and nonmodal phonation. *J Phon*. 2001; 29:431–450.

76. Regner MF, Robitaille MJ, Jiang JJ. Interspecies comparison of mucosal wave properties using high-speed digital imaging. *Laryngoscope*. 2010; 120:1188–1194. [PubMed: 20513038]
77. Mergell P, Herzel H, Titze IR. Irregular vocal-fold vibration--high-speed observation and modeling. *J Acoust Soc Am*. 2000; 108:2996–3002. [PubMed: 11144591]
78. Neubauer J, Mergell P, Eysholdt U, Herzel H. Spatio-temporal analysis of irregular vocal fold oscillations: biphonation due to desynchronization of spatial modes. *J Acoust Soc Am*. 2001; 110:3179–3192. [PubMed: 11785819]
79. Jiang JJ, Zhang Y, Kelly MP, Bieging ET, Hoffman MR. An automatic method to quantify mucosal waves via videokymography. *Laryngoscope*. 2008; 118:1504–1510. [PubMed: 18545215]
80. Švec JG, Horá ek J, Šram F, Veselý J. Resonance properties of the vocal folds: in vivo laryngoscopic investigation of the externally excited laryngeal vibrations. *J Acoust Soc Am*. 2000; 108:1397–1407. [PubMed: 11051466]
81. Van Houtte E, Van Lierde K, D'Haeseleer E, Claeys S. The prevalence of laryngeal pathology in a treatment-seeking population with dysphonia. *Laryngoscope*. 2010; 120:306–312. [PubMed: 19957345]
82. Herrington-Hall BL, Lee L, Stemple JC, Niemi KR, McHone MM. Description of laryngeal pathologies by age, sex, and occupation in a treatment-seeking sample. *J Speech Hear Disord*. 1988; 53:57–64. [PubMed: 3339868]
83. Coyle SM, Weinrich BD, Stemple JC. Shifts in relative prevalence of laryngeal pathology in a treatment-seeking population. *J Voice*. 2001; 15:424–440. [PubMed: 11575638]
84. Karnell MP, Hall KD, Landahl KL. Comparison of fundamental frequency and perturbation measurements among three analysis systems. *J Voice*. 1995; 9:383–393. [PubMed: 8574304]
85. Patel R, Dailey S, Bless D. Comparison of high-speed digital imaging with stroboscopy for laryngeal imaging of glottal disorders. *Ann Otol Rhinol Laryngol*. 2008; 117:413–424. [PubMed: 18646437]

Figure 1. Representative scaling region determination during nonlinear dynamic analysis. A: Representative plot showing $\log_2 C(r)$ against $\log_2(r)$. Each curve corresponds to an embedding dimension (m) integer value between 1 and 10. B: Plot showing the slope of each embedding dimension curve in panel A against r . Scaling region boundaries are indicated by dashed lines.

Figure 2.

Representative waveforms and narrowband spectrograms obtained during acoustic signal typing. A: Type 1 signal produced by a patient with bilateral vocal fold scar. B: Type 2 signal produced by a patient with bilateral type II sulcus vocalis. A bifurcation is indicated by an arrow at ~500 ms in the spectrogram. C: Type 3 signal produced by a patient with bilateral type II sulcus vocalis with concomitant bilateral scar. D: Distribution of signal types produced by controls and patients with vocal fold scar and pathologic sulcus vocalis ($n = 23$ in each group).

Figure 3.

Representative phase plots for the three patients presented in Figure 2. A: Type 1 signal produced by a patient with bilateral vocal fold scar. B: Type 2 signal produced by a patient with bilateral type II sulcus vocalis. C: Type 3 signal produced by a patient with bilateral type II sulcus vocalis with concomitant bilateral scar.

Figure 4.

Auditory-perceptual analysis of voice signals from patients with vocal fold scar and sulcus vocalis and controls ($n = 23$ in each group). Each signal token was rated five times each by three raters (i.e., 15 total ratings per token) using a seven point equal-appearing-interval scale representing overall voice quality (1: normal; 7: largest deviation from normal). Multiple ratings were then averaged within each rater followed by across raters. A: Comparison of mean perceptual rating of voice quality for patients and controls. B: Bland-Altman analysis of auditory-perceptual rating reliability across raters. Grey solid lines represent perfect agreement; black dashed lines represent mean bias; red dashed lines represent upper and lower limits of agreement (set at 95%).

Figure 5.

Comparison of linear (i.e., time-domain based perturbation) and nonlinear approaches to the acoustic analysis of voice signals from patients with vocal fold scar and sulcus vocalis and controls. A: Jitter. B: Shimmer. C: Signal-to-noise ratio (SNR). D: Correlation dimension (D_2). Type 1 signals are represented by purple open circles; type 2 signals are represented by green closed circles; type 3 signals are represented by red closed circles. Statistical p -values are presented for type 1 signal data only ($n = 20$ in the control group; $n = 3$ in the scar/sulcus group) and for all data ($n = 23$ in each group).

Figure 6.

Auditory-perceptual and acoustic analyses of patient subgroups: Sulcus vocalis only ($n = 6$), vocal fold scar only ($n = 7$) and sulcus vocalis with concomitant scar. ($n = 10$). A: Mean perceptual rating of voice quality (1: normal; 7: largest deviation from normal). B: Jitter. C: Shimmer. D: Signal-to-noise ratio (SNR). E: Correlation dimension (D_2). For acoustic data (panels B–E), type 1 signals are represented by purple open circles; type 2 signals are represented by green closed circles; type 3 signals are represented by red closed circles. Statistical P -values represent all data (statistical analysis of perturbation data using type 1 signals only was not possible as only one type 1 signal was identified in each patient subgroup).

Figure 7.

Comparison of the association between acoustic indices and mean perceptual rating of voice quality in patients with vocal fold scar and sulcus vocalis and controls. A: Jitter. B: Shimmer. C: Signal-to-noise ratio (SNR). D: Correlation dimension (D_2). Type 1 signals are represented by purple open circles; type 2 signals are represented by green closed circles; type 3 signals are represented by red closed circles. Correlation coefficients (r) are presented for type 1 signal data only ($n = 23$) and for all data ($n = 46$).

Table 1

Correlation matrix for mean perceptual rating of overall voice quality performed by three independent raters. Correlation coefficients (Pearson's r) are shown for each rater pair and each individual rater against the group mean.

Rater	1	2	3	Group mean
1	1.00	-	-	-
2	.92	1.00	-	-
3	.84	.85	1.00	-
Group mean	.97	.97	.94	1.00