Nonlinear Dynamic-Based Analysis of Severe Dysphonia in Patients With Vocal Fold Scar and Sulcus Vocalis

*Seong Hee Choi, †Yu Zhang, ‡Jack J. Jiang, ‡Diane M. Bless, and ‡Nathan V. Welham, *Gyeongsan, Korea, †Xiamen, China, ‡Madison, Wisconsin

Summary: 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 data set was observed for SNR and 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 (ie, 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.

Key Words: Auditory-perceptual analysis–Chaos–Correlation dimension–Jitter–Perturbation analysis–Phase plot–Signal-to-noise ratio–Signal typing–Shimmer–Voice disorder.

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 the 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 used in combination;[9–11] however, questions concerning the reliability and validity of each approach have been raised.[12–16] Furthermore, 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

Accepted for publication September 15, 2011.
The authors hold no financial or other conflicts of interest.
From the *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; and the ‡Division of Otolaryngology, Department of Surgery, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin.
Address correspondence and reprint requests to Nathan V. Welham, Division of Otolaryngology, Department of Surgery, University of Wisconsin School of Medicine and Public Health, K4/723 CSC, 600 Highland Avenue, Madison, WI 53792. E-mail: welham@surgery.wisc.edu
Journal of Voice, Vol. 26, No. 5, pp. 566-576
0892-1997/$36.00

doi:10.1016/j.jvoice.2011.09.006

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,[25] summarizing a consensus workshop on acoustic analysis, suggested the adoption of signal typing to ensure the appropriate analysis of normal and disordered voices. Type 1 signals, defined as near-periodic signals, are suitable for analysis using time-domain based perturbation analyses. Type 2 signals, characterized by at least one qualitative signal bifurcation (eg, 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's[25] recommendations, non F_0-dependent measures have shown increasing value in the analysis of aperiodic (ie, 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 HNR[30] and SNR[31] measures do. Quefrency-domain measures such as cepstral peak prominence (CPP) hold a similar advantage and have shown good correspondence to the 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 used 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 with 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 (eg, 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 the measurement error is a meaningful contributor to rating variability and that the inter-rater agreement and reliability are enhanced by averaging ratings from multiple presentations of the same stimulus to each rater (compared with 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, standard deviation [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 presented 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, 10 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 approximately 7 years) was a direct consequence of excluding patients who presented with concomitant laryngeal disease (eg, vocal fold scar in the setting of recalcitrant human papilloma viral