Perturbation and Nonlinear Dynamic Analysis of Adult Male Smokers

*Lingying Chai, †Alicia J. Sprecher, *Yi Zhang, *Yufang Liang, *Huijun Chen, and *,†Jack J. Jiang, *Shanghai, China and †Madison, Wisconsin

Summary: Objective. Smoking results in a voice change, and the perception by smokers of an abnormal voice may encourage quitting behavior. Moreover, a disordered voice is often the first sign of vocal pathology. Efforts to evaluate voice have focused on classical acoustic analysis; however, nonlinear dynamic analysis has been shown to be a reliable objective method for the evaluation of voice. We compare the discriminatory ability of these two methods when applied to normal and smokers' voices.

Study Design. Prospective study.

Methods. The study included 73 subjects, 36 nonsmokers and 37 smokers. A segment of sustained vowel production was obtained from each subject. Acoustic dimension and correlation dimension (D2) analyses were applied to the data. Results were compared with a Mann-Whitney rank sum test, logistic regression, and receiver operating characteristics (ROC) analysis.

Results. D2 values for smokers were significantly higher than D2 values for nonsmokers ($P < 0.001$). Jitter and shimmer analysis showed higher values for these parameters among smokers. Logistic regression indicated a higher predictive power with D2, and ROC analysis found no significant difference between the analysis methods.

Discussion. This study indicated that D2 is highly sensitive to changes associated with smoking and has the potential to be implemented clinically as an indicator of abnormal voice. Further research could focus on using nonlinear dynamic analysis to create a normative database, producing standards for monitoring voice changes caused by cigarette smoking.

Key Words: Acoustic analysis–Nonlinear dynamic analysis–Voice disorder–Smoking.

INTRODUCTION

Cigarette smoking is a major predisposing factor for an assortment of potentially fatal respiratory ailments.[1–4] Within the upper airway, smoking is a risk factor for diseases, including rhinitis, sinusitis, and laryngitis.[5] Chronic use of cigarettes often leads to laryngeal problems, such as chronic inflammation, erythema, dryness, and itching laryngeal mucosa.[6] Moreover, smoking increases rates of laryngeal reflux and Reinke's edema and is the most important factor associated with laryngeal carcinoma.[6,7] Pathological changes in the larynx result in voice disruptions, and disordered voice is often the first symptom and chief complaint indicating laryngeal disease.[8,9]

Previous research indicated that smoking affects perceptual, acoustic, and aerodynamic performance during phonation. Smoking results in a voice characterized as rough and breathy with a decreased pitch.[7] Sorensen and Horii[10] reported a lower fundamental frequency (F_0) for both running and spontaneous speech among male smokers compared with nonsmokers. Gonzalez and Carpi[11] observed elevated perturbation parameters (jitter and sPPQ) in early stage smokers of both genders and higher tremor parameters (amplitude tremor intensity index and frequency tremor intensity index) in male smokers.[11] In studies examining aerodynamic aspects of phonation, lower vital capacities and maximum phonation times were noted. In addition, smokers were observed to produce greater peak and mean pressures during the voiceless stop-plosive task.[12]

Much of the previous research on voice in smokers has focused on F_0 and perturbation analysis. However, such analyses may not be appropriate for the smoker's voice. Titze[13] recommended the classification of voices into three types before acoustic analysis. He defined a type I voice as nearly periodic and indicated that only this type of voice was suitable for acoustic analysis. Type II signals contained strong subharmonics or modulations, whereas type III signals are aperiodic. Titze[13] recommended against applying perturbation analysis to type II and III signals because of difficulties associated with accurate extraction of F_0. Normal voices are predominantly type I and are therefore easily analyzed with perturbation measures.[13] Disordered voices can be characterized by any of Titze's[13] three voice types. The decreased voice quality associated with smokers' voices may correspond with an increased likelihood of these voices being classified as type II or III and unsuitable for classic acoustic analysis.[7,10–12]

Unlike perturbation, nonlinear dynamic methods are capable of analyzing all three of Titze's voice types. Nonlinear analysis has been established as a viable means of measuring acoustic characteristics of phonation through computer modeling and excised larynx experiments.[14–18] The method has been applied to both pathological and nonpathological voice samples in numerous studies to quantify the degree of aperiodicity and irregularity.[16,19–21] Moreover, research has indicated that nonlinear measures require shorter samples, are more robust in response to noise, and require a lower sampling rate than classical acoustic measures.[16] These features increase the

Accepted for publication January 11, 2010.
Grant support: This work was supported by the National Institutes of Health grant number RO1DC006019 from the National Institute of Deafness and other Communication Disorders.
From the *Shanghai EENT Hospital of Fudan University, Shanghai, China; and the †Division of Otolaryngology–Head and Neck Surgery, Department of Surgery, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin.
Address correspondence and reprint requests to Jack J. Jiang, Department of Otolaryngology, Shanghai Eye and ENT Hospital of Fudan University, 83 Fenyang Road, Shanghai 200031, China. E-mail: jjjiang@wisc.edu
Journal of Voice, Vol. 25, No. 3, pp. 342-347
0892-1997/$36.00

doi:10.1016/j.jvoice.2010.01.006

potential for applicability of nonlinear parameters in situations, such as the clinic, where the recording environment may not be ideal and computer storage may be limited.

Nonlinear dynamic methods have not been previously applied to voices of smokers. In this study, we attempt to quantify the effect of smoking on voice with perturbation and nonlinear dynamic analysis. Next we compare the results of both methods in evaluating the voices of long-term Chinese male smokers and nonsmokers using logistic regression and receiver operating characteristics (ROC) analysis. We evaluate the potential to incorporate nonlinear measures into the typical acoustic analysis protocol.

METHODS

Subjects

The protocol used in the study was approved by the Ethics Committee of Shanghai Eye, Ear, Nose and Throat (EENT) Hospital of Fudan University and the Institutional Review Board (IRB) of the University of Wisconsin. Female smokers are not common in China; therefore, all subjects were male. Seventy-three subjects were recruited from the EENT Hospital; 36 nonsmokers and 37 smokers. All the subjects participated voluntarily and neither had a history of laryngeal disorder or trauma nor were voice complaints associated with their visit to the clinic. In accordance with routine clinical procedure, before acoustic analysis, all subjects received a rigid endoscopic examination. As we intended to use acoustic analysis as a screening measure apart from endoscopic examination, these results did not influence inclusion in the study; however, significant endoscopic findings were noted. Exclusion criteria included: excessive use of voice, overexposure to noise, alcohol addiction, symptoms consistent with gastroesophageal reflux disease (GERD) and laryngopharyngeal reflux, symptoms of a cold within the previous 2 weeks, and chronic rhinitis or postnasal drip. Smokers were defined as subjects who smoked at least five cigarettes everyday for at least 5 consecutive years before the study. Nonsmokers were operationally defined as subjects who had never smoked. Subject information is provided in Table 1, including age, height, weight, number of cigarettes consumed per day, and number of years as a smoker. Smokers and nonsmokers were similar in age, height, weight, and body mass index.

Measurement

Subjects were seated in a sound proof room during data collection. A Shure PG48 microphone (Shure, Niles, IL) was used to record voice samples and was positioned 10 cm away from the subjects' mouths. Subjects were asked to produce an /a/ at a comfortable pitch. Each phonation lasted for at least 3 seconds and was monitored by a sound level meter (Radio Shack, Fort Worth, TX) to maintain intensity between 75 and 80 dB. Subjects were allowed to practice several times before recording. Voice samples were obtained using the *Multi-Dimensional Voice Program model 5105*, Version 3.1.4 at a sampling rate of 44.1 kHz (KayPENTAX, Lincoln, NJ). Three trials were recorded for each subject.

A 1 second segment was spliced from a central steady portion of each voice sample and analyzed by both perturbation and nonlinear dynamic methods. Voice onset and offset were excluded to avoid effects of speech intonations and interactions between the larynx and vocal tract on acoustic analyses. All voices were classified as type I, type II, or type III according to Titze's definitions using a narrowband spectrogram generated by *PRAAT* (Boersma & Weenink, Amsterdam, The Netherlands http://www.praat.org/). Voices with clearly defined harmonics and without subharmonics, bifurcations, or strong modulations in the F_0 were classified as type I, whereas the presence of any of these features resulted in a classification of type II. Type III voices were defined by spectrograms where harmonics were not clearly defined (energy smeared across the frequencies). Figure 1 presents sample spectrograms of each voice type generated from voices included in the *Disordered Voice Database 4337* (KayPENTAX, Lincoln Park, NJ).

Perturbation analysis

Perturbation analysis was performed with *CSpeech software*, version 4.0 (Milenkovic & Read, Madison, WI). Percent jitter, percent shimmer, and signal-to-noise ratio (SNR) were measured. Jitter measures the cycle-to-cycle frequency variation of the signal, whereas shimmer measures the cycle-to-cycle amplitude variation, and SNR provides a measure of the amount of noise in the speech waveform.[16,22]

The reliability of perturbation analysis was determined by a parameter called error (ERR) calculated by the *CSpeech* program. ERR counts the number of times the calculated pitch period varies in excess of one sample position.[23] A high ERR indicates numerous pitch breaks, which negatively affect the quality of the F_0 estimate. Data with a ERR greater than 10 are usually considered unreliable, and the sample is deemed unsuitable for classical acoustic analysis.[23]

TABLE 1.
Subject Information

Character	Nonsmokers (n = 36)	Smokers (n = 37)	*P* value
Age (yr)	42.47 (8.48)	44.59 (6.80)	0.477
Height (m)	1.71 (0.05)	1.72 (0.06)	0.240
Weight (kg)	68.44 (8.00)	70.23 (9.28)	0.340
Body mass index	23.48 (2.04)	23.79 (2.70)	0.778
Cigarettes smoked per day		17.68 (7.33)	
Total years smoked		20.32 (6.82)	

Reported are the means with standard deviation given in parentheses.