Detection of Chronic Laryngitis due to Laryngopharyngeal Reflux Using Color and Texture Analysis of Laryngoscopic Images

*Daniel R. Witt, †Huijun Chen, *Jason D. Mielens, *Kieran E. McAvoy, †Fan Zhang, *Matthew R. Hoffman, and *,†Jack J. Jiang, *Madison, Wisconsin and †Shanghai, People's Republic of China

Summary: Objective. To determine if pattern recognition of hue and textural parameters can be used to identify laryngopharyngeal reflux (LPR).

Methods. Laryngoscopic images from 20 subjects with LPR and 42 control subjects without LPR were obtained. LPR status was determined using the reflux finding score. Color and texture features were quantified using hue calculation and two-dimensional Gabor filtering. Five regions were analyzed: true vocal folds, false vocal folds, epiglottis, interarytenoid space, and arytenoid mucosae. A multilayer perceptron artificial neural network with varying numbers of hidden nodes was used to classify images according to pattern recognition. Receiver operating characteristic (ROC) analysis was used to evaluate diagnostic utility, and intraclass correlation coefficient analysis was performed to determine interrater reliability.

Results. Classification accuracy when including all parameters was 80.5% ± 1.2% with an area under the ROC curve of 0.887. Classification accuracy decreased when including only hue (73.1% ± 3.5%; area under the curve = 0.834) or texture (74.9% ± 3.6%; area under the curve = 0.852) parameters. Interrater reliability was 0.97 ± 0.03 for hue parameters and 0.85 ± 0.11 for texture parameters.

Conclusions. This preliminary study suggests that a combination of hue and texture features can be used to detect chronic laryngitis due to LPR. A simple, minimally invasive assessment would be a valuable addition to the currently invasive and somewhat unreliable methods currently used for diagnosis. Including more data will likely improve classification accuracy. Additional investigations will be performed to determine if results are in accordance with those provided by pH probe monitoring.

Key Words: Laryngopharyngeal reflux–Hue–Gabor filter–Artificial neural network–Laryngitis.

INTRODUCTION

Approximately 15% of all patients presenting to the otolaryngology office have chronic laryngopharyngeal reflux (LPR).[1–3] Twenty-four-hour pH probe measurements indicative of LPR have been observed in 50% of patients with voice complaints.[4] Despite the high incidence of this pathology in voice patients, current methods of diagnosing LPR can be unreliable.

LPR is the regurgitation of gastric contents onto the mucosal linings of the pharynx, larynx, and upper aerodigestive tract, resulting in a spectrum of nonspecific symptoms.[5] Both Goldberg[6] and Koufman[7] identified a causal relationship between the presence of acidic gastric juice and mucosal tissue damage in animals, suggesting that a similar damaging effect could occur on the laryngeal mucosa during LPR. The presence of acid and pepsin in this sensitive region causes a variety of physiological responses, such as laryngeal edema and erythema, mucosal hypertrophy,[4] granuloma, carcinoma, and subglottic stenosis.[8,9] These physical signs are often considered during LPR diagnosis along with common symptoms such as throat clearing, persistent cough, globus sensation, and changes in voice quality.[10] Thus, there is an array of nonspecific signs and symptoms that point to LPR as an underlying etiology, making diagnosis controversial.

Despite the widespread prevalence of this disorder, current diagnosis is fairly subjective and can be inaccurate. Branski et al[11] demonstrated low interrater reliability among otolaryngologists assessing the same physical laryngeal signs, such as erythema, edema, and granulation. Some clinicians accept 24-hour ambulatory pH probe testing to be the gold standard in LPR diagnosis.[9,12,13] Twenty-four-hour pH probes have shown 96% sensitivity and specificity for identifying the presence of acid reflux near the lower esophageal sphincter[14]; however, other studies have reported much lower sensitivity and reproducibility (55%) for identifying abnormal amounts of proximal esophageal acid reflux.[15] The accuracy of 24-hour pH probe testing can be limited due to incorrect probe placement, irregularity of reflux during testing, and low pH values used to identify periods of reflux.[13,16,17] Moreover, cost factors, inconvenient procedure length, and the invasive nature of this method can deter patients from undergoing pH probe impedance testing.[11–13] Given some of the technical limitations and low availability of pH probe equipment in some clinics, the reflux finding score (RFS) can serve as a reasonable, noninvasive surrogate for the pH probe. Park et al[18] reported high sensitivity (87.8%) and lower specificity (37.5%) for the RFS in diagnosing pH probe-validated LPR in patients with globus (n = 57).

Accepted for publication August 26, 2013.
Conflicts of interest: None.
From the *Department of Surgery, Division of Otolaryngology–Head and Neck Surgery, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin; and the †Department of Otolaryngology–Head and Neck Surgery, Shanghai EENT Hospital, Fudan University, Shanghai, People's Republic of China.
Address correspondence and reprint requests to Jack J. Jiang, 1300 University Ave, 2725 Medical Sciences Center, Madison, WI 53792. E-mail: jjjiang@wisc.edu
Journal of Voice, Vol. 28, No. 1, pp. 98-105
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2013.08.015

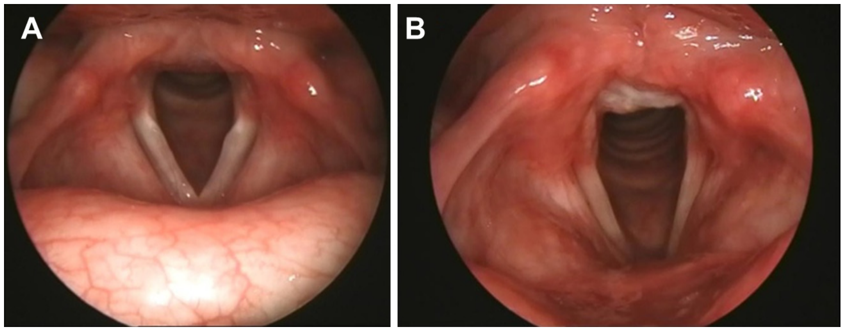

FIGURE 1. Images of larynges without (**A**) and with LPR (**B**).

New methods to diagnose LPR are warranted. Computer-based analysis of laryngeal image color and texture features offers an alternative which is objective, quantitative, and convenient. Few studies have been conducted on color and texture analyses of the larynx, with little investigation into the specific physical signs of LPR. Texture, as it pertains to digital images, constitutes descriptive metrics that describe human-perceived "textures" of an image. That is, how segmented, patterned, "bumpy," or "smooth" an image appears. Verikas et al[19] performed a study that examined the diagnostic capability of color and texture analysis in differentiating between healthy and abnormal vocal folds and between various subgroups of vocal fold mass lesions. Hanson et al[20] performed a study quantifying erythema in chronic posterior laryngitis resulting from LPR. Different areas of the larynx were analyzed for relative erythema, and treatment efficacy was monitored using Red-Green-Blue (RGB) color analysis. This method focused solely on obtaining an erythema index for a limited number of regions, with no consideration of texture patterns.

We propose using pattern recognition of color and texture features obtained from laryngoscopic images to identify LPR. We hypothesized that hue and a combination of two-dimensional (2D) Gabor textural parameters would distinguish images obtained from subjects with LPR from those obtained from controls. To test this hypothesis, we classified images based on hue and texture parameters using an artificial neural network (ANN).

MATERIALS AND METHODS

Selection criteria

This study was conducted under the approval of the ethics committee of the Shanghai EENT Hospital. Laryngeal images were obtained from 20 subjects with LPR (18 men; age range: 37–67 years; mean age: 54.3 years) and 42 control subjects (25 men; age range: 24–78 years; mean age: 48.7 years) who required laryngoscopy for other laryngeal abnormalities. Subjects for both groups had laryngoscopy performed at the Eye, Ear, Nose, & Throat Hospital of Fudan University in Shanghai, China. Subjects in the LPR group presented to the clinic with signs and symptoms of LPR and were diagnosed with LPR according to the RFS assessment.[5] A score exceeding 7 indicated the presence of LPR.[5] Subjects comprising the control group were individuals presenting with unknown laryngeal complications who had an RFS $\leq$7 (because LPR was diagnosed when RFS $\geq$8). Patients comprising the control group presented to the clinic with varying laryngeal complications, including nasopharyngeal discomfort, foreign body sensation in pharynx, cysts near tongue base, chronic neck discomfort, and chronic sore throat.

ANN analysis requires knowledge of the class to which each data set belongs before the pattern recognition algorithm can be evaluated. There is currently no perfect assessment of LPR. One currently available option, the RFS, demonstrates excellent intra- and interrater reliability[5] and the cutoff used in this study was demonstrated previously to be effective in distinguishing persons with and without LPR.[5] Although 24-hour ambulatory pH probe is considered a gold standard in LPR diagnosis, it has numerous well-noted limitations[13,16,17] and is not performed regularly at our clinic. Accordingly, the RFS was considered an adequate indicator of LPR presence or absence for this preliminary study.

Image collection

Sixty-two videos were obtained as a part of standard clinical assessment. A 70° laryngoscope (model 8706 CA; KARL STORZ; Tuttlingen, Germany) with a mounted GP-KS822 micro camera (Panasonic; Secaucus, NJ) was used. This model delivers images of 752 $\times$ 582 pixels at a sensitivity of 6 Lux/F1.4. Importantly, the white balance function was used to prevent any potential effects of variable lighting on image color.

Protocol

Videos were assessed on *VLC* media software, Version 1.1.11 (VideoLAN Organization; Paris, France), which enabled the extraction of individual still image frames (Figure 1). Images were saved as JPEG image files and then individually analyzed for hue and texture.

The following regions were evaluated: true vocal folds, false vocal folds, epiglottis, interarytenoid space, and arytenoids (Figure 2). Each region was manually selected using a sensor sketch pad (Wacom Intuos 4 PTK-640; Wacom Co., Ltd., Toyonodai Otonemachi, Kazo-shi, Saitama, Japan). These site selections were performed by two image analysts. Both analysts maintained similar technique to achieve a high level of standardization in region selection. The mean hue and texture parameters were derived from three independent boundary selection trials. Following site selection, mean hue and texture outputs were assessed for each selected region.