Spectral Analysis of Digital Kymography in Normal Adult Vocal Fold Vibration

*Wenli Chen, †Peak Woo, and ‡Thomas Murry, *†‡*New York, New York*

Summary: The purpose of this study was to examine the spectrum of normal adult vocal fold vibration obtained through direct visualization technique using digital kymography (DKG). DKG extracts high-speed kymographic images of each vocal fold sampled at a single or multiple points along the vocal folds. Thus, direct and objective quantification of the cycle-to-cycle movements of the left and right vocal folds were obtained. In this study, DKG spectrum configuration in males and females were examined. Samples were obtained from seven subjects (three males and four females) with no history of voice disorders. Subjects were asked to produce tokens obtained from a standard clinical laryngeal evaluation producing tokens at modal, low, and high frequency and at modal frequency with increased loudness. Results demonstrated that the spectrum of normal adult phonation consisted of a large and robust H_1 (fundamental frequency) and significant attenuation of power in the higher harmonics (H_2, H_3). The spectral peaks were quasiperiodic with no spectral smearing. Analysis of the spectral data revealed variations in the spectrum that were influenced by the frequency of phonation and gender. Subjects demonstrated a range of left-right asymmetry of the spectral peaks (2–26%). This study presents a preliminary database of direct spectral characteristics of normal adult vocal fold vibration over a range of frequencies and intensities. Furthermore, these measures provide data from which similar measures obtained from dysphonic pathologies can be compared.

Key Words: Digital kymography–Kymography–Spectrum–Vocal fold vibration–Normative data–High-speed videoendoscopy.

INTRODUCTION

Spectral analysis has been used to objectively quantify vibratory movements of the vocal folds. Because vibratory movement of the vocal folds critically determines the quality of voice production, spectral analysis can further elucidate the dynamic vibratory behavior of the vocal folds in both normal and pathologic voices. Recent applications using spectral analysis to study human adult phonation have focused on indirect approaches, such as using acoustic measures.[1–4] Although these indirect measures provide valuable evidence of the vibratory motion of normal and disordered vocal folds, there remains little data from direct spectral analysis of vocal fold vibration.

High-speed videoendoscopy (HSV) has been used to examine vibratory movements of the vocal folds in normal phonation.[5–8] HSV provides direct visualization of vocal fold movement in real-time and captures images at frame rates of 2000 to 10 000 frames per second. This improved capturing rate enables one to visualize the entire cycle-to-cycle vibratory motion of the vocal folds in normal and severely disordered phonation that cannot be tracked with stroboscopy.[9] HSV can also be used to generate kymographic data, also known as digital kymography (DKG). Kymography quantifies cycle-to-cycle movement of each vocal fold sampled from single or multiple lines across the vocal folds over time.[10,11] As a result, kymography is ideally suited for visual judgment of left-right asymmetry[12–14] and deriving the vocal fold vibratory spectrum in a wide variety of vibratory features in healthy and disordered voices.[15] Although DKG spectrum has been reported in excised canine larynges,[16] DKG studies in human have focused on their relationship to perceptual ratings[12–14] and to time series measurements.[17–19]

To our knowledge, spectral analysis of normal vocal fold vibration from DKG has not been examined across a wide range of frequencies or intensities in human phonation. Previous approaches to objectively quantify direct vocal fold movement in human phonation focused on measurements of glottal area[5–8,20,21] and laryngeal topogram.[22] Spectral analysis of kymographic data provided additional information (eg, frequency, periodicity, peak power, symmetry) on the vibratory behavior of each vocal fold movement. Such a technique may provide clinically useful information for diagnostic clarity in cases, in which other endoscopic techniques may have limited usefulness. Therefore, the purpose of this study was to examine the spectrum of normal male and female phonation during standard clinical laryngeal evaluation tasks using DKG obtained from HSV. Furthermore, the study aims to establish a preliminary normative database for the vibratory parameters derived from DKG spectrum of the vocal folds in normal adult human phonation.

METHODS

Subject selection criteria and description

Following a laryngeal examination by a laryngologist, seven adult subjects (four females and three males) with ages ranging between 24 and 65 years were entered into the study. All subjects had no reported history of smoking, neurologic disease, laryngeal surgery, voice disorders, hearing problems, and speech or language impairment. Examination of their vocal folds before obtaining the samples verified no lesions or significant asymmetry of vibration.

Accepted for publication October 21, 2013.
A version of the article was presented at the 42 Annual Symposium Care of the Professional Voice; June 1, 2013; Philadelphia, PA.
From the *Department of Biobehavioral Sciences, Teacher's College, Columbia University, New York, New York; †Department of Otolaryngology - Head and Neck Surgery, Icahn School of Medicine, New York, New York; and the ‡Department of Otolaryngology-Head and Neck Surgery, Weill Cornell Medical College, New York, New York.
Address correspondence and reprint requests to Wenli Chen, 145 East 48th Street, Apt 27G, New York, NY 10017. E-mail: wc2416@tc.columbia.edu
Journal of Voice, Vol. 28, No. 3, pp. 356-361
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2013.10.015

Phonatory tasks

All subjects performed four phonatory tasks typically used in standard laryngovideostroboscopy. In the modal task, subjects produced /i/ at a comfortable pitch and loudness. This was each subject's reference point for the other three tokens. In the low task, subjects produced /i/ at a low pitch and at a comfortable loudness. In the high task, subjects produced /i/ at a high pitch and at a comfortable loudness. Finally, in the modal loud task, subjects produced /i/ at their modal pitch with a loud voice. Using these four measures, the subjects established a broad frequency and intensity range. A second experimenter monitored the subjects' fundamental frequency (F_0) and intensity and provided verbal feedback to each subject to ensure that they met the target production established in the trials. Subjects were instructed to practice their individual targets in short, successive 2-second tokens. Throughout the tasks, the investigator provided modeling and feedback to assist the subject in maintaining target frequency and intensity. In short, the targets were chosen as common comfortable effort level phonation similar to a standard clinical laryngeal examination in which comfortable effort level, a high frequency, and low frequency relative to the comfortable effort level as well as a loud comfortable effort level were sampled.

High-speed video recording

HSV was acquired using KayPENTAX High-Speed Digital Imaging (HSDI) system (KayPENTAX Photronmotion; Pentax Medical, Montvale, NJ), which consisted of a 70-degree rigid endoscope (model 9100) coupled with a 300-W Xenon light source. The HSDI system acquired gray-scale images at a rate of 2000-frames per second with a spatial resolution of 256 × 120 pixels rotated to a vertical position for capture. Videolaryngoscopy was performed as in conventional videostroboscopy procedure. A condenser microphone was placed 6 in from the lips to monitor the intensity readings from an A-weighted sound level meter in the HSDI system. A contact microphone was held at the neck to monitor the fundamental frequency.

Following the instructions and practice trials, the samples were obtained when the examiner observed a clear and full view of the larynx and frequency and intensity of phonation were verified to be consistent by the second investigator. A template outlining the vocal folds was mounted on the monitor in order for the examiner to fit the visual image of the vocal folds within the template. This provided a consistent size of the vocal folds across tokens for later analysis. Six continuous 2-second tokens phonations were captured. The three best token with a clear and full view of the larynx were saved onto the hard drive for analysis. The F_0 and intensity of each phonation segment were manually recorded. All subjects tolerated the data collection procedure without any difficulties.

Data analysis

Kymography image processing. Kymograph analysis of the vibratory samples is critically dependent on accurate delineation of the edge of the vocal folds from HSV.[23] Therefore, HSV samples were preprocessed using video editing software (*VirtualDub*, V.1.9.11). Image rotation (bicubic resampling) was implemented such that the image was judged to be vertically aligned. An initial brightness and contrast adjustments were manually applied to images that were perceived to have a less distinct contrast between the vocal fold edges and the glottis. Images were not adjusted for size as they were obtained within the template described above. A 400- to 500-frame video segment that demonstrated a full view of the vocal fold with minimal movement of the subject was extracted from the recorded HSV samples. *Kay's Image Processing* Software (*KIPS*, model 9181; Pentax Medical, Montvale, NJ) was used to generate the kymogram by placing a transverse line across the glottis at the mid-membranous portion of the vocal folds, where vocal fold contact is greatest at this point.[13] Edge detection was subsequently applied to identify and trace the vocal fold edges (Figure 1). If automatic tracing was imprecise, a new kymographic sample was generated and manual correction functions internal to the *KIPS* (eg, brightening, darkening, and erase function) were applied onto the new kymograph. This process was repeated until the automatic tracing accurately delineated the vocal fold edges. When the vocal fold edges were delineated, Kymograph Edge Analysis function was applied. The resulting values were Kymograph Edge Data (KED), which described the coordinate values of the left and right edges of the vocal fold presented across time (Figure 2). Subsequently, Fourier transform function was applied to the KED, resulting in a spectrum ranging from 0 to 1000 Hz for the left and right edges of the vocal fold. The kymographic image processing was repeated for all phonatory tasks (Figure 3).

Spectral data analyses. The frequency and spectral peak power values of the F_0 (also known as H_1), second harmonic (H_2), and third harmonic (H_3) were obtained to examine the spectral shape of male and female phonation during the four laryngeal tasks. Twenty percent of the postprocessed 500-frame HSV tokens were randomly selected and reanalyzed by the same experimenter to evaluate error of measurement. Comparison of the H_1, H_2, and H_3 values between the original and reanalyzed sample yielded a reliability of 93%.

The following vocal fold vibratory parameters were examined: (1) spectral smearing; (2) spectrum shape; (3) spectral power configuration; (4) gender effects; and (5) spectral power symmetry. *Spectral smearing* was determined by examining the presence of the noise-band at a specific frequency along the spectrum. *Spectrum shape* was determined by summing the spectral peak values of H_1, H_2, and H_3 individually across all

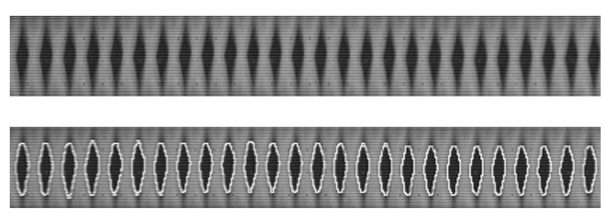

FIGURE 1. Kymogram of P3 during modal phonation task. The top image is a kymogram created with the sampling line selected at the mid-membranous portion of the vocal fold. The bottom image is the same kymogram following edge detection.