THE GLOTTALTOPOGRAPH: A METHOD OF ANALYZING HIGH-SPEED IMAGES OF THE VOCAL FOLDS

Gang Chen[1] *Jody Kreiman*[2] *Abeer Alwan*[1]

[1]Department of Electrical Engineering, University of California, Los Angeles
[2]Division of Head and Neck Surgery, School of Medicine, University of California, Los Angeles
gangchen@ee.ucla.edu, jkreiman@ucla.edu, alwan@ee.ucla.edu

ABSTRACT

High-speed video provides a way to record physiological vibrational patterns of the vocal folds. Due to the large amount of data it produces, many methods have been proposed to reduce raw images to the underlying vibrational patterns. Previous methods either focus on a certain location on the vocal folds, or a certain frequency of vibrational activity. In this paper, we propose the "glottaltopograph" which is based on principal component analysis of pixels' gray scale time courses. This method reveals the overall synchronization of the vibrational patterns of the vocal folds. Experimental results showed that this method is effective in visualizing pathological and normal vocal-fold vibrational patterns.

Index Terms— high-speed video, vocal fold, principal component analysis

1. INTRODUCTION

Clinicians and speech scientists have used several clinical instruments to observe vocal-fold vibrations [1, 2]. Recently, high-speed imaging of the larynx has become more common due to the decreasing cost of high-speed recording devices. With the increased recording frame rate and resolution, high-speed video recording has become a powerful tool to visualize vocal-fold vibrations.

Speech scientists have developed many techniques to analyze high-speed video data. The kymograph is extracted from high-speed images in [2, 3] to analyze the movement of the vocal folds, but the region of interest is restricted to a single line of the image. In [4], the temporal oscillation patterns in the entire laryngeal area are visualized by using the Fourier transform of the light intensity of consecutive high-speed images. The resulting signal contains amplitude and phase information as a function of frequency and is displayed in several ways to analyze vibrational characteristics of the entire laryngeal area. The laryngotopography was proposed in [5] to extract spatial characteristics of the larynx, and was applied to the high-speed images of normal subjects and patients. Results showed that the laryngotopography was effective in visualizing various vibrational modes of the vocal folds of patients with paralysis and cysts.

The spatial limitation of the kymograph is the restricted region: a single line of the image. The laryngotopography maintains the spatial characteristics of the entire image but only focuses on several frequency components of the spectrum of the vibrational pattern. In this paper, we propose the "glottaltopograph" based on principal component analysis (PCA) to visualize high-speed video data. The proposed method reveals the overall spatial and temporal synchronization pattern of the vocal-fold vibration, rather than focusing on a certain location or frequency. Comparison between analyses of pathological and normal data shows that the proposed method is effective in visualizing a wide variety of vocal-fold vibrational patterns.

2. METHOD

2.1. High-speed images

High-speed images were recorded at 4000 frames/second with a Color High-Speed Video System (CHSV), Model 9710 (KayPENTAX, Montvale, NJ). The image resolution was 512×256 pixels. The color mode was 8 bit RGB. Audio signals were recorded at a sampling rate of 40 kHz simultaneously with high-speed recordings. Nine subjects (8 males and 1 female) with voice disorders were recorded while saying the vowel /i/ using their habitual voice. Four normal subjects (3 males and 1 female) were also recorded saying the same vowel with breathy, modal, and pressed voice quality.

2.2. Image adjustment

High-speed images were first converted from RGB to gray scale. Due to illumination conditions, some areas in the high-speed images displayed ultra high brightness. The brightness of these glare spots needed to be adjusted before subsequent pixel gray scale analysis, because their brightness did not reflect actual vocal-fold movement. Histogram adjusting was implemented to map a certain brightness interval in the histogram to an 8 bit gray scale (0-255), enhancing edge contrast of the vocal folds. The brightness interval to be rescaled was

This work was supported in part by NSF grant IIS-1018863 and NIH/NIDCD grant DC01797.

chosen according to the actual brightness condition of each recording.

2.3. PCA implementation

A rectangular window was manually selected to isolate the vibrating part of vocal folds. For each pixel inside the rectangular window, the gray scale time course was extracted across the entire video, as shown in Figure 1. 300 consecutive frames were used to extract the gray scale functions to ensure the representativeness of each function. PCA was performed on the gray scale time courses of the pixels.

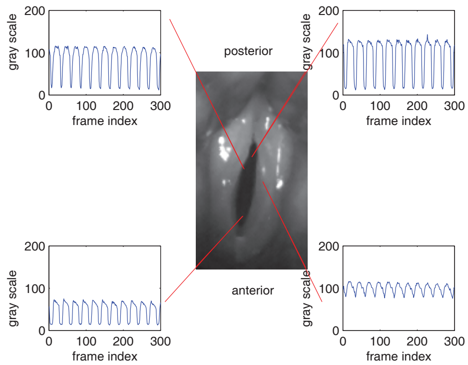

Fig. 1. Center: image obtained from high-speed camera. Panels around: Gray scale time functions of pixels at different locations of the vocal folds

2.4. Analysis and Visualization

For the gray scale time courses, the first 2 coefficients (projection on principal components) were calculated. The coefficients were normalized to an 8 bit scale (0-255) and visualized at the original pixel location in terms of color saturation (for better visualization). The gray scale curve was then reconstructed using the first 2 coefficients and principal components. Reconstruction errors (mean square errors) were calculated and visualized in the same way. In the final stage, the percentage of variance explained was calculated for the first 2 principal components.

3. RESULTS

3.1. Variations in voice quality among normal subjects

The glottaltopograph is applied to modal, breathy, and pressed voices from normal speakers.

3.1.1. Modal

Figure 2 (a)-(e) shows the glottaltopograph of a normal female subject (speaker 1). Compared to the original image in (a), contrast of vocal fold edges has been enhanced after the brightness adjustment in (b). As shown in (c), the first principal coefficient distribution displays a highly symmetric pattern, representing the vocal-fold structure. A similar symmetric pattern is also observed in (d), the second principal coefficient distribution. In (e), the reconstruction error distribution is visualized. The posterior and anterior end regions of the vocal folds display the highest reconstruction error. This can be attributed to the regions' small vibration amplitude and slight phase lag compared to the middle portion of the vocal folds, which characterizes the main vibrational activity. Over 80% of energy is captured in the first principal component, reflecting the high regularity of the vocal-fold vibration.

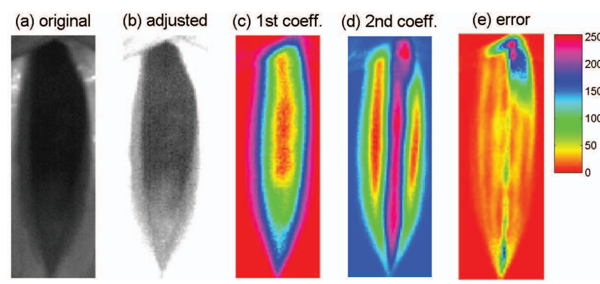

Fig. 2. The glottaltopograph of a modal voice from a normal subject (speaker 1). (a) and (b): the original and adjusted images of the vocal folds. (c) and (d): the first and second principal coefficients, displayed in terms of color saturation. (e): reconstruction error using the first two principal coefficients, displayed in terms of color saturation.

3.1.2. Breathy

Figure 3 shows the glottaltopograph of a breathy voice produced by speaker 1. Compared to the original image in (a), glare areas have been removed after the brightness adjustment in (b). Similar to Figure 2, highly symmetric patterns were observed in (c) and (d). The largest reconstruction error in (e) is from the epiglottis region. (d) reveals that the main vibration activity is at the anterior region of the vocal folds, which is characterized by large vibration amplitude. The shape of the posterior glottal gap is also visualized in (d).

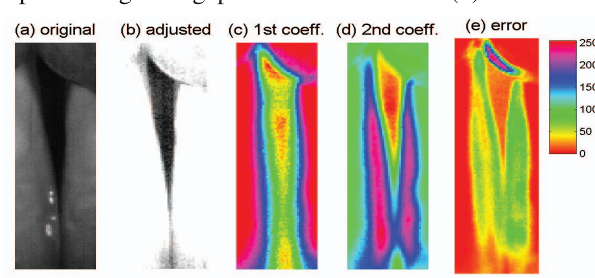

Fig. 3. The glottaltopograph of a breathy voice from a normal subject (speaker 1)

3.1.3. Pressed

The glottaltopograph of a pressed voice from a normal male subject (speaker 2) is shown in Figure 4. Again, (c) and (d) exhibit highly symmetric patterns. The first two principal components account for over 90% of the variance, indicating a compact energy distribution.

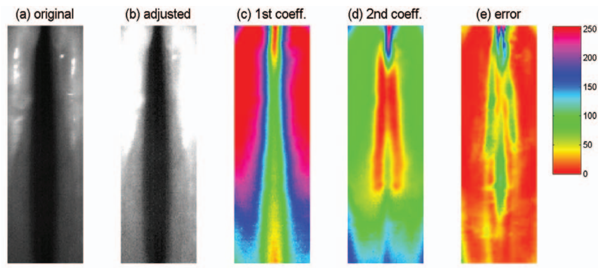

Fig. 4. The glottaltopograph of a pressed voice from a normal subject (speaker 2)

3.2. Patients with voice disorders

The glottaltopograph is applied to patients with voice disorders.

3.2.1. *Speaker 3*

Figure 5 (a)-(e) shows the glottaltopograph of a patient (speaker 3) exhibiting complex asymmetry in vibrational amplitude between the left and right vocal folds, with the left fold vibrating with larger amplitude than the right. Further, the vibrating amplitude of the right fold alternates cycle by cycle, although both left and right vocal folds have the same vibratory frequency. The corresponding acoustic signal exhibits creaky voice quality.

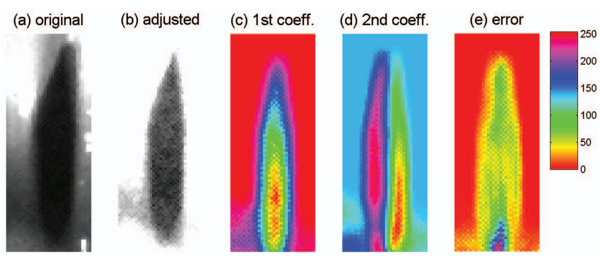

Fig. 5. The glottaltopograph of a patient (speaker 3)

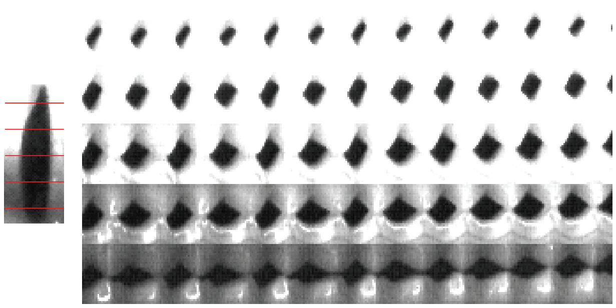

Fig. 6. Multi-line kymograph of a patient (speaker 3). Upper is left and lower is right.

Similar to Figure 2, the first principal coefficient draws the shape of the glottis in (c). However, the second principal coefficient shows different color patterns between left and right vocal folds, indicating the difference in vibrational patterns. Less then 60% of the variance is explained by the first principal component, indicating the irregularity of vibration.

The multi-line kymograph for this subject is shown in Figure 6, where alternating vibrational amplitudes are displayed. It is revealed in the glottaltopograph that the vibrational patterns are distinct between the left and right vocal folds, but similar within either side. Compared to the kymograph, the glottaltopograph provides a better spatial resolution in visualizing the different vocal-fold vibrational patterns.

3.2.2. *Speaker 4*

The glottaltopograph of a second patient (speaker 4) with a voice disorder is shown in Figure 7. The corresponding audio signal exhibits a breathy voice quality. The glottaltopograph shows an interesting pattern. The right anterior portion of the vocal folds has a phase lag of about π, which means that the right anterior portion is in opposite phase with regard to the rest of the vocal folds. This is manifested in (d) as two distinct portions in the second principal coefficient distribution: the right anterior portion versus the rest of the vocal folds. In (e), the right middle portion of the vocal folds has the largest reconstruction error. Note that a signal with a perfect opposite phase could be reconstructed perfectly. The right middle portion is the border where normal phase and opposite phase meet, which produces an irregular vibrational pattern. Only 62% of the variance is explained by the first principal component, suggesting the irregularity in vibrational pattern.

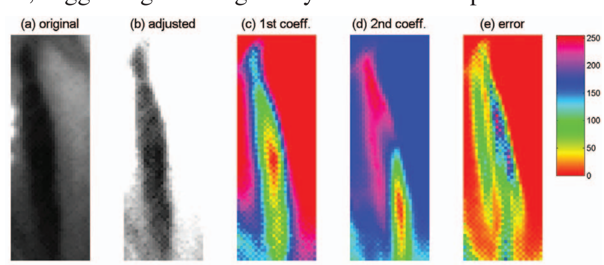

Fig. 7. The glottaltopograph of a patient (speaker 4)

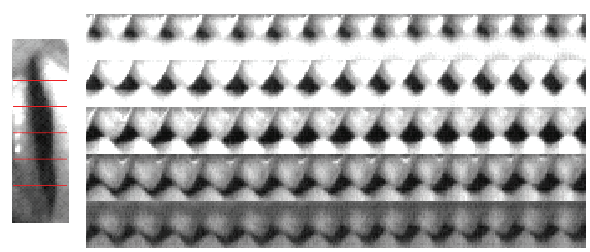

Fig. 8. Multi-line kymograph of a patient (speaker 4)

Figure 8 shows multi-line kymograph of speaker 4, where the **anterior** portion shows the phase difference between the left and right vocal folds. However, vocal-fold activity in the vertical direction is not well captured in the kymograph. The glottaltopograph clearly shows that the "out-of-phase" region is the **right anterior** portion. The size and position of this problematic region is also visualized. Compared to the kymograph, the glottaltopograph again provides better spatial information about the overall vocal-fold vibrational pattern.

3.3. A normal subject with asymmetric vibration

The glottaltopograph and the kymograph of a modal voice from a normal male subject (speaker 5) are shown in Figures 9 and 10. This subject has a phase difference between the left and right vocal folds. In this case, the asymmetric vibrational patterns are captured in both Figures 9 (d) and 10.

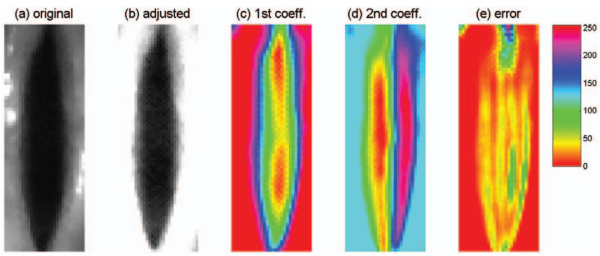

Fig. 9. The glottaltopograph of a modal voice from a normal subject (speaker 5)

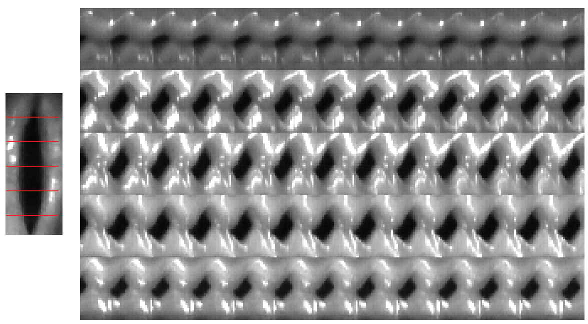

Fig. 10. Multi-line kymograph of a modal voice from a normal subject (speaker 5)

4. DISCUSSION

Glottaltopographs of pathological phonation reveal irregular patterns. The low energy concentration indicates the existence of multiple vibration modes, in either the spatial or temporal domain. These irregularities result in the asymmetric pattern of the glottaltopographs. Observing those special patterns could aid in identifying the portion of the vocal folds that might be the physiological cause of the vocal pathology. Note that certain unsynchronized vibrational patterns can also be observed among normal subjects, where incorporating the acoustic signals would aid in the analysis and evaluation.

When applying this method, one thing that must be kept in mind is that there is a non-linear transformation from physical position to light intensity. A brightness mapping technique was applied before PCA to remove the glare spots and also to increase the contrast between the glottal open area and vocal folds. After this preprocessing, the brightness of vocal folds approaches maximum value and the brightness of the glottal open area approaches 0, so that brightness curves better represent movements of the vocal folds. If the vocal-fold vibrational pattern is synchronized over the entire region, the brightness curves for different locations should exhibit synchronized behavior, which would be displayed as symmetric patterns in the glottaltopograph.

5. COMPARISON WITH OTHER TECHNIQUES

The kymograph visualizes the vocal-fold activity within the specified line in the horizontal direction, but lacks information along the vertical direction. The laryngotopography is based on Fourier analysis and focuses on several frequency components of the vibrational pattern. Compared to the kymograph and the laryngotopography, the glottaltopograph reveals the overall spatial and temporal vibrational pattern of the entire laryngeal area, rather than a single line or several frequency components. The glottaltopograph also provides an automatic way of finding the region of interest from the entire image, instead of restricting the analysis region to several pre-selected lines. The explained-variance percentages from PCA also provide quantitative measures of synchronization of the vibrational pattern. The glottaltopograph may be used in conjunction with the kymograph and the laryngotopography to analyze high-speed recordings of the vocal folds.

6. CONCLUSIONS

The glottaltopograph is proposed and is based on PCA of gray scale time courses to visualize high-speed images of the vocal folds. This method focuses on the synchronization of movements across the entire laryngeal region. Results show that this method is effective in visualizing spatially and temporally irregular patterns of the vocal-fold vibration, suggesting this method may provide insight into the phonatory mechanisms underlying perceived voice disorders.

7. REFERENCES

[1] P. Kitzing, "Stroboscopy–a pertinent laryngological examination," *The Journal of Otolaryngology*, vol. 14(3), pp. 151–157, 1985.

[2] M. Tigges, T. Wittenberg, P. Mergell, and U. Eysholdt, "Imaging of vocal fold vibration by digital multiplane kymography," *Comput. Med. Imaging Graph.*, vol. 23(6), pp. 323–330, 1999.

[3] H. Larsson, S. Hertegård, P.-Å. Lindestad, and B. Hammarberg, "Vocal fold vibrations: high-speed imaging, kymography and acoustic analysis: a preliminary report," *Laryngoscope*, vol. 110, pp. 2117–2122, 2000.

[4] S. Granqvist and P.-Å. Lindestad, "A method of applying Fourier analysis to high-speed laryngoscopy," *J. Acoust. Soc. Am.*, vol. 110(6), pp. 3193–3197, 2001.

[5] K. Sakakibara, H. Imagawa, M. Kimura, H. Yokonishi, and N. Tayama, "Modal analysis of vocal fold vibrations using laryngotopography," in *Interspeech*, 2010, pp. 917–920.