Published in final edited form as:

J Voice. 2010 January ; 24(1): 21–29. doi:10.1016/j.jvoice.2008.03.003.

Efficient and effective extraction of vocal fold vibratory patterns from high-speed digital imaging

Yu Zhang, Erik Bieging, Henry Tsui, and Jack J. Jiang

Department of Surgery, Division of Otolaryngology Head and Neck Surgery, University of Wisconsin Medical School, Madison, WI 53792-7375

Abstract

Objectives—High-speed digital imaging can provide valuable information on disordered voice production in voice science. However, the large amounts of high-speed image data with limited image resolutions produce significant challenges for computer analysis, and thus effective and efficient image edge extraction methods allowing for the batch analysis of high-speed images of vocal folds is clinically important. In this paper, a novel algorithm for automatic image edge detection is proposed to effectively and efficiently process high-speed images of the vocal folds.

Methods—The method integrates Lagrange interpolation, differentiation, and Canny edge detection, which allow objective extraction of aperiodic vocal fold vibratory patterns from large numbers of high-speed digital images. This method and two other popular algorithms, histogram and active contour, are performed on 10 sets of high-speed video data from excised larynx experiments in order to compare their performances in analyzing high-speed images. The accuracy in computing glottal area and the computation time of these methods are investigated.

Results and Discussion—The results show that our proposed method provides the most accurate and efficient detection, and is applicable when processing low resolution images. In this study, we focus on developing a method to effectively and efficiently process high-speed image data from excised larynges. However in addition we show the clinical potential of this method by use of example high-speed image data obtained from a patient with vocal nodules.

Conclusions—The proposed automatic image-processing algorithm may provide a valuable biomedical application for the clinical assessment of vocal disorders by use of high-speed digital imaging.

Keywords

High-speed digital imaging; vocal folds; Lagrange interpolation; Canny image edge detection; automatic image processing

I. INTRODUCTION

Biomedical imaging has become an important tool in the field of otolaryngology [1–3]. In voice science, commonly used acoustic measurement cannot directly record vocal fold vibrations or

© 2009 The Voice Foundation. Published by Mosby, Inc. All rights reserved

5745 MSC 1300 University Ave Madison WI, 53706 jiang@surgery.wisc.edu Tel: 608 265 9854 Fax: 608 265 2139.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

provide visual information about the vibratory patterns of laryngeal systems [4]. Laryngeal imaging has played an important role in visualizing vocal fold vibrations [5]. Laryngeal stroboscopy provides a visual image of the glottal movement at a low video rate (25 Hz). However, it is applicable only for periodic vocal fold vibrations and cannot provide an accurate depiction of aperiodic vibratory patterns. Vocal folds oscillate at a fundamental frequency of 100 Hz ~ 400 Hz during normal phonation and therefore faster imaging systems are desired to capture the entire oscillation cycle of the vocal folds. High-speed digital imaging is emerging as a valuable clinical tool for the direct measurement of vocal fold vibrations [6–15]. High-speed imaging systems record images of the larynx at a rate as high as 4000 frames/second, and are capable of resolving actual vibrations of the vocal folds. Yan *et al.* [2,6] applied automatic tracing of vocal fold movement to extract glottal area of normal subjects and patients. The analysis of the glottal area function has also been used for the assessment of vocal disorders from laryngeal pathologies. Hong *et al.* [8] used high speed imaging of the vocal folds to characterize the vocal fold movement during stop production using a simple threshold based edge detection algorithm. As the mechanism and properties of human voice production are ultimately dependent on the dynamics of vocal fold vibrations, high-speed imaging is expected to be applied as a clinical tool to provide new information on the vibratory mechanisms of vocal disorders [16–18]. Our previous study [18] showed that this tool could be used to extract system parameters of a vocal fold model where a manual, frame by frame, edge detection process was used, and thus progress toward clinical use of vocal fold parameter extraction has been limited by the difficulty in processing large amounts (thousands) of image data. In order to realize clinical application of vocal fold parameter extraction, the edges of the vocal folds need to be detected automatically and efficiently. Finite spatial and temporal resolution of high-speed imaging gives rise to noise in the image data which also challenges the automatic extraction of vocal fold vibratory patterns. Therefore, developing an effective and efficient image edge extraction method allowing for the batch analysis of high-speed images of vocal folds is clinically important.

Excised larynx experiments facilitate direct observation and measurement of vocal fold vibrations, and have proven to be advantageous in the study of laryngeal physiology [19,20]. High speed imaging of excised larynges offers controllable sample data that is difficult to achieve in clinical recording. The effects of image resolution and signal length on image processing can be systematically monitored and independently studied, which can be valuable for the initial testing of the vocal fold edge detection algorithm. More importantly, in order to study the effects of computation error and processing time, long image series and high-quality image data should be included as reference samples, which are difficult to obtain in a clinical setting. Excised larynx experiments offer unique benefits in examining the effectiveness and limitations of a glottal edge detection method before its clinical application. Recent study has applied high-speed imaging in excised larynx experiments to investigate the mechanism of disordered voice production. Manual image edge detection methods have been applied on a frame-by-frame basis to extract vocal fold edges [11,16,18]. Traveling mucosal waves have been measured by use of high-speed photography [11]. Irregular vocal fold vibrations have been observed at extremely high subglottal pressures [16]. Although a manual frame-by-frame procedure has been widely applied to extract vocal fold vibratory patterns, such a procedure is very time consuming for large numbers of high-speed images, and therefore is difficult to apply in clinical practice.

Several automatic methods for determination of the glottal edge from high-speed imaging have been employed in previous studies, including histogram and active contour based methods. Histogram methods assume that there is a significant difference between the intensity of the object pixels and the intensity of the background pixels. When the glottis is nearly or completely closed, such a difference becomes difficult to recognize and a proper segmentation threshold cannot be determined. Active contour algorithms are popular in medical image processing, as

well, [12–14] but are limited in their effectiveness in processing thousands of high speed images due to their high computation times. In addition, the active contour snake may be drawn toward false edges during the glottal closure phase. Therefore, development of a new method that overcomes the limitations of these methods while accurately and efficiently detecting vocal fold edges is clinically important for the application of high speed imaging.

The purpose of this paper is to propose a new automatic algorithm for extracting the vocal fold vibrations from high-speed images in excised larynx experiments. Our method integrates the features of the Lagrange interpolation, differentiation, and Canny edge detection. Some commonly used methods, histogram and active contour, have also been employed. A comparison among these three methods will be made on computation error and processing time. We will show that our proposed method is more effective and efficient in processing large amounts of high-speed image data. We will show that this method is capable of extracting vocal fold vibrations at low image resolutions. Finally, the potential biomedical application of our analytical approach to high-speed imaging for the analysis of vocal disorders will be examined by extracting the glottal area series from a patient with vocal nodules.

II. METHOD

A. Excised Larynx Experiment

An excised larynx was used as an organ model of the vocal folds. Figure 1 shows that the experimental system consists of an excised larynx setup [20] and a high-speed camera system, as described by our previous studies [16,18]. Ten canine larynges harvested from healthy laboratory dogs were used in an experimental trial 12 to 36 hours after excision. Each freshly excised canine larynx with structures of vocal glottal, surrounding tissue, and all laryngeal cartilages, was mounted with a section of trachea on top of a pipe, and the trachea was tightly clamped to the pipe with a hose clamp. A 3-pronged device was used to stabilize the arytenoid cartilages bilaterally, allowing for micrometer control of adduction and abduction. A second micrometer system was attached by stitching a rod to the anterior tip of the thyroid lamina. Turning this micrometer system controlled the elongation of the vocal folds. During the experiments, vocal fold elongation and adduction were symmetrically held constant since the micrometer systems were unchanged. An Ingersoll-Rand (Type 30) conventional air compressor was used to generate the airflow. The input air was conditioned to 35° C to 38° C and 95% to 100% relative humidity by two ConchaTherm III heater-humidifiers (Respiratory Care, Inc.) placed in series. The subglottal pressure in the artificial lung was measured with an open-ended water manometer (Dwyer No. 1211). When subglottal pressure was sufficiently increased, the larynx vibrated. In experiments, subglottal pressure and vocal fold configurations could be held constant, so that long duration vocal fold vibrations were ensured for analysis. The vibratory patterns of these 10 excised larynges were recorded with a high-speed digital camera (Fastcam-ultima APX). The high-speed camera was mounted on a track-system over vocal folds. The high-speed digital camera system acquired images at a sampling rate of 4000 frames/second with a resolution of 256 × 512 pixels, and the image data was transmitted into a computer for analysis.

B. Image Detection Algorithm

In each high-speed image, each pixel (x, y) has an intensity level $E(x, y)$. In comparison to other structures present in the image, such as false vocal folds and arytenoid cartilages, vocal fold edges represent the boundaries at which the change in intensity is most rapid, as shown in Fig. 2(a). In order to reduce noise and unwanted textures (or smooth the boundary), and extract glottal edges from high-speed images, we apply an automatic algorithm by integrating three basic steps: Lagrange interpolation [21], differentiation, and Canny image edge detection

[22]. The algorithm was implemented using MatLab® software on an Intel® Pentium® 4 CPU 2.80 GHz processor.

B.1 Lagrange interpolation—Lagrange interpolation is used as the first step to reduce the amount of detail present in each image and produce a continuous function of the intensity level $E(x, y)$. Filtering out the noise and other small structures allows for determination of the vocal fold edges. At each line y_i ($i = 1, 2, \dots, N$) of an individual frame in Fig. 2(b), the Lagrange interpolating polynomial method will be applied to intensity function $E(x, y_i)$ as [21]:

$$E(x, y_i) = \sum_{j=1}^M E(x_j, y_i) \prod_{\substack{k=1 \\ k \neq j}}^M \frac{x - x_k}{x_j - x_k} \quad (1)$$

denoting a differentiable polynomial, where $E(x_j, y_i)$ is the intensity at the position (x_j, y_i) , and M and N denote the horizontal and vertical image resolutions. This equation creates a polynomial that represents the intensity data in one column of the image, as it theoretically minimizes the error between the data and a polynomial. The created intensity function is continuous and therefore can be differentiated. The image intensity functions before and after Lagrange interpolation are shown in Fig. 2(b). In order to further smooth the image and reduce noise and unwanted details and textures, we apply a Gaussian filter

$h_\sigma(x, y) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$ to the intensity function $E(x, y)$. The width of the Gaussian filter is specified by σ . Increasing σ results in more smoothing. For this study σ was set to a value of 5 pixels. These two operations are applied to the image simultaneously.

B.2 Differentiation—The glottal edge is normally represented by the point of greatest change in the image intensity corresponding to the extreme values of the image intensity. In the second step, differentiation horizontally along the rows of the image is used to determine the extreme values of the image intensity as well as the glottal edges. $E(x, y)$ and $h_\sigma(x, y)$ are continuous functions, and therefore the points of maximal change in the filtered intensity function $E(x, y) * h_\sigma(x, y)$ can be obtained from the derivative using the derivative theorem of convolution, as [23]:

$$f(x, y) = \frac{\partial}{\partial x} [E(x, y) * h_\sigma(x, y)] = \left[\frac{\partial}{\partial x} E(x, y) \right] * h_\sigma(x, y), \quad (2)$$

where $*$ denotes the convolution operator. From Eqs. (1) and (2), the filtered intensity function can be described as,

$$f(x, y) = \left[\sum_{j=1}^M E(x_j, y) \sum_{l=1, l \neq j}^M \left(\frac{1}{x_j - x_l} \prod_{k=1, k \neq j}^M \frac{x - x_k}{x_j - x_k} \right) \right] * \left[\frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \right]. \quad (3)$$

This equation represents the derivative of the smoothed interpolating polynomial taken with respect to x . With increasing x , $f(x, y)$ shows the greatest magnitude at the two glottal edges, and thus its minimal and maximal values give the left and right boundary of the vocal folds in the images. Then the left $L(y)$ and right $R(y)$ vocal fold edges can be approximated as

$$L(y) = \min_{0 \leq x \leq x_0} f(x, y), \quad (4)$$

$$R(y) = \max_{0 \leq x \leq x_0} f(x, y). \quad (5)$$

where x_0 is the x -axis boundary of the image. The slow gradient of the flattened $f(x, y)$ function may cause difficulty in determining the specific location of $L(y)$ and $R(y)$. We then rewrite the intensity as $E'(x, y) = E(x, y) W(x, y)$ in order to increase the contrast between low intensity (glottal gap) and high intensity (surrounding tissues) and the gradients at $L(y)$ and $R(y)$, where $W(x, y)$ represents the window function described as,

$$W(x, y) = \begin{cases} 1, & L(y) \leq x \leq R(y), \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

$W(x, y)$ can be well approached using the continuous function

$$0.5 \tanh [c(x - L(y))] - 0.5 \tanh [c(x - R(y))] \quad (7)$$

with a sufficiently large constant c . At this point, the glottal edge has been detected, but further processing is necessary to create a smooth continuous edge that more accurately represents the exact glottal edge.

B.3 Canny edge detection—In the final step, we use Canny edge detection, a double thresholding algorithm [22], to obtain a continuous glottal edge. In our algorithm, this process effectively detects the edge by eliminating less prominent portions of a jagged vocal fold edge. The Canny edge thresholding algorithm ensures that the sharp protrusions of the glottal edge are not included in the in the final delineated edge.

In order to perform Canny edge detection, the gradient of the intensity function $E'(x, y)$ is computed, determining the rate of change in pixel intensity. Local maxima of this gradient that are above a specified threshold are initially identified as edges, and non-maximal pixels in the edges are removed to create a thin edge. After non-maxima suppression, a low threshold T_l and a high threshold T_h are applied to obtain double thresholded edge images. If the gradient at a pixel is above T_h , there is a clear edge at that point and it is declared as an edge pixel. If the gradient at a pixel is below T_l , the edge at that point is very faint and thus it is declared as a non-edge-pixel. If the gradient at a pixel is between T_l and T_h then may or may not be an actual desired edge. Therefore, it is declared as an edge pixel if and only if it is connected to a definite edge pixel directly or via other pixels between T_l and T_h .

Each of the above three steps have the same importance in the detection process. From Eqs. (1) to (7), the vocal fold edges can be extracted from the surrounding tissue by integrating Lagrange interpolation, differentiation, and the Canny image edge detector. An extracted glottal area from high-speed image data is shown in Fig. 2(c), where the glottal area is marked as white curve, $T_l = 0.25$, and $T_h = 0.65$. Since the edge is nearly determined prior to the Canny detection, the detection is not greatly dependent on the selected values of T_l and T_h . Figure 2 (d) shows the detected glottal area using $T_l = 0.1$, and $T_h = 0.8$, with no substantial change in the glottal edge. Our image detection algorithm still effectively detects the vibratory patterns of vocal folds, and the values $T_l = 0.25$ and $T_h = 0.65$ can be generally applied for high-speed

image data. We will compare this method with some commonly used automatic methods, histogram and active contour.

C. Other Popular Analytical Algorithms

Since glottal edge can be perceptually segmented from the surrounding vocal fold tissue, the simplest, but very time consuming method is to manually determine the glottal edges on a frame-by-frame basis [5,11,16]. In this manual or frame-by-frame procedure, for each individual image a threshold is chosen by the user to segment the glottal area. If any error is present after thresholding, the areas of error are manually removed by the user, so that the glottal area can be obtained. Different image frames may require different thresholds, and thus this manual method requires significant user-program interface. Such a manual frame-by-frame procedure is very time consuming for large numbers of high-speed images, and thus is difficult to use in clinical practice. In recent years, with the development of high speed digital imaging, various automatic image detection algorithms have been applied [6,7,9,12–14]. In this study, we consider two of the most popular methods, the histogram and active contour algorithms, to automatically detect vocal fold edges.

The histogram algorithm is widely used for content-based image processing [6,10,24]. A histogram is the graphical version of a table which shows what proportion of image intensity falls into each of several or many specified grayscale values. As the vocal fold image contains both high intensity (surrounding tissue) and low intensity (glottal gap) areas, then the value in between the two can be treated as the threshold level. The threshold is determined by finding the first local minimum in the histogram [24], similarly to the method used by Wittenberg *et al.* [10]. Based on the threshold derived from the histogram algorithm, we may separate the glottal gap from the surrounding tissues. Due to its simplicity, the histogram algorithm has high efficiency in batch processing large amounts of image data.

The active contour (snake) model introduced by Kass *et al.* [25] is an automatic method for image segmentation. The contour, or snake, is defined as a curve $v(s) = (x(s), y(s))$, $s \in [0,1]$ that moves through the spatial domain of an image $I(x,y)$ to minimize the energy function

[12,14,25,26], $E = \int_{\frac{1}{2}} \left(\alpha |v_s|^2 + \beta |v_{ss}|^2 \right) + E_{ext}(v(s)) ds$, where v_s and v_{ss} represent the first and second derivatives of v with respect to s . Then the snake satisfies the following Euler equation, $\alpha v_{ss} - \beta v_{ssss} - \nabla E_{ext} = 0$, where α and β are the weights to control the snake's tension and rigidity, respectively. v_{ssss} denotes the fourth derivative of v with respect to s . We apply E_{ext} from the image gray level $f(x, y)$ that will take on smaller values at the boundaries. The final position of the contour will have a minimum value of energy E , so finding the object boundary becomes an energy minimization problem. After initializing a curve close to the object (or glottal gap) boundary, the snake, which is seen as an energy minimizing spline, starts deforming to fit the local minima so as to move towards the desired object boundary and finally settles on the object boundary. Considering that the initial curve will significantly affect the results of the active contour method, in this study, we will apply the glottal gap obtained by the histogram algorithm as the initial curve of the active contour algorithm. Although it is time consuming and sensitive to noise, the active contour algorithm has become an important method in the medical imaging community [12–14,25–29].

In order to investigate the effectiveness and efficiency of our image detection algorithm, we will compare the results with those obtained from the above two popular analytical algorithms. While improvements to these algorithms have been made to address their deficiencies, we compare the results of our method to these in computation accuracy and computation time in order to show that our algorithm overcomes the obstacles commonly encountered in high speed vocal fold image processing.

III. Results and Discussions

In order to show the effectiveness of our method in detecting vocal fold edges, we apply it to high-speed image data and compare the results obtained using the histogram and active contour methods. Figure 3 illustrates eight successive frames of high-speed images during the open and closing phases of the vocal folds, where the curves in panels (a), (b), (c), and (d) correspond to high-speed image sequences, histogram, active contour method, and our automatic method, and the white area represents the extracted glottal area. During the open phase of the glottal cycle, two sides of vocal folds have no contact with each other. A distinct boundary between the glottis and the surrounding tissue leads to successful detection of the vocal fold edges using each of these four methods. In the closing stage of the glottal cycle, the vocal folds collide at the midpoint of the glottis. The proportional area of glottal gap, and therefore the proportion of low intensity level pixels, was sufficiently reduced. As shown in Fig. 3(b), the histogram method applied in this study can not give an accurate extraction, which has also been found in the study by Yan *et al.* [6]. In addition, vocal fold closure cannot be correctly extracted using the active contour method. The small area marked by the red lines in Fig. 3(c) shows that the final snake or shape sufficiently deviates from real glottis and moves toward the surrounding tissue. This may be improved by choosing the thresholds based on several consecutive frames, although the propagation of computation error to the following frames may be an issue. In comparison with these two methods, our method provides a more accurate detection of vocal fold edges during the entire vibratory cycle of the vocal folds, as shown in Fig. 3(d). Figure 4 displays the extracted glottal area of 100 high-speed frames using the manual frame-by-frame method and our automatic method. The glottal area time series have clear valley-cut sinusoid shapes [11], indicating the vocal fold openings and closures. High-speed imaging using our method reveals the aperiodic properties of vocal fold vibrations, which cannot be obtained using traditional laryngeal stroboscopy imaging [5]. The detected glottal edge using our automatic method sufficiently approaches the glottal area shown in the high-speed image sequence of Fig. 3(a), demonstrating the effectiveness of our automatic image detection method. In addition, in this study, we focus on the extraction of glottal area from high speed imaging which has been shown to be applicable to the assessment of laryngeal pathologies in previous studies [6,9]. Since the entire vocal fold vibratory patterns can be extracted using our automatic method, other valuable information, such as mucosal wave phase difference [5,11], spatiotemporal irregularity [7,16] can also be derived. Further investigation of the application of the proposed method should be performed in these important topics.

An effective automatic image detection method should have good computation accuracy in extracting vocal fold edges from high-speed image data. In order to investigate the computation accuracies of these image edge detection methods, we define the error function

$$\Delta_1 = \sqrt{\frac{1}{T} \int_0^T [a(t) - a_s(t)]^2 dt} / \sqrt{\frac{1}{T} \int_0^T a_s(t)^2 dt}$$

averaged over 100 high-speed image frames, where $T = 100$, $a_s(t)$ represents the glottal area time series extracted using the manual frame-by-frame detection method, and $a(t)$ represents the glottal area time series extracted using each automatic detection method. Figure 5 shows the error function distribution of ten larynges obtained using the histogram method, active contour method, and our method. With respect to the manual frame-by-frame detection method, the averaged error over 10 larynges of the histogram method, active contour algorithm, and our method are 34.2%, 11.7%, and 7.7%, respectively. The histogram method gives the worst results because the method is limited by the quality of the histogram of each image and assumes the presence of an object in the image. In order for the histogram method to be effective, there should be a distinct gap between the object and background intensities [6,10,24]. During the stage of glottal closure, there is no glottal gap to detect and thus no low intensity peak in the histogram, making it difficult to determine an accurate threshold value. These factors dramatically increase the computation error in finding the glottal area from the histogram. The active contour method is also

fundamentally limited in its effectiveness. This algorithm is designed to detect the boundaries of one and only one object in an image, which is frequently not the case in vocal fold imaging. There is no object present during the glottal closure phase, and in the presence of vocal nodules the glottis may be divided into two or more image objects. Setting a correct initial region is also a significant obstacle, and the active contour method is highly dependant on the initial region and other parameters. Previous studies [6,9,12–14] use the result from the previous frame as the initial region of the following frame in the image sequence. However, the propagation of computation error may be an important factor, and an unsuccessful detection may result in difficulty of detection in subsequent images. In this study, the histogram method was used to dynamically set the initial region of each individual image frame. However, the active contour method is only effective if the initial region determined by the histogram method is already reasonably close to the actual edge, making it not valuable. In recent years, improvements to the histogram and active contour methods have been made [13]. Whether or not these improved methods have value in laryngeal physiology requires further examination and is beyond the scope of this study. In this study, we compare our new method to some typically used methods to demonstrate its effectiveness. In comparison with the histogram and active contour methods, our method does not require predefining any initial region, nor does it require histogram quality. In addition, our method does not require the determination of many parameters, as is required by the active contour method. Thus, our method represents a more convenient and effective procedure to detect vocal fold vibratory patterns in comparison to these two automatic methods.

Clinically, real-time processing of thousands of high-speed image frames requires that an image detection method be automatic, and in particular, should be highly efficient in extracting vocal fold vibratory patterns [7,9]. The implementation of the parameter extraction methods in our previous study [18] depend on efficient extraction of the vocal fold edges. In order to investigate the computation efficiency, we calculate the computation times of the histogram method, active contour method, and our method to process 100 high-speed image frames from each larynx, as shown in Fig. 6. For 10 larynges, the average computation time of our method was 1.2 minutes. The average computation time of the histogram method was 2.1 minutes, much shorter than that of the active contour method (67.7 minutes). Here, the computation times are based on the implementation of each of the three algorithms in MatLab, an interpreted programming language that is easily used and but has low computation efficiency, and could be greatly improved by using compiled programs or hardware-based firmware programs. The computation error of the histogram method is much larger than of the active contour method, as shown in Fig. 5. The active contour method includes numerous iterations in computation [12–14,24–29]. Thus the active contour method is much more computationally expensive than the histogram method or our method. This makes the active contour method difficult to apply to the real-time batch processing of thousands of high-speed images. Our method uses a direct process to determine the vocal fold edges. It does not involve iterations in computation or operations in determining the histogram quality, and thus is more efficient than the other two automatic methods. The higher computation accuracy and computation efficiency of our method makes it a better candidate to automatically extract glottal area data from numerous high-speed images, and thus a better candidate for use in the extraction of vocal fold parameters [17,18].

Recent study has suggested that the limited image resolution of high-speed imaging systems may affect the computed results [14]. Image resolutions lower than those in our excised larynx experiments have been applied in clinical practice [6–14]. In addition, due to the presence of arytenoid cartilages and other surrounding tissues, it is often necessary to reduce the size of the field of view in order to capture the vocal folds, which results in a decreased image resolution for edge detection. Because of these two facts, it is necessary to examine the results of our method under reduced image resolution conditions. In addition to high computation

accuracy and efficiency, our algorithm is robust to low image resolution. In order to investigate the effects of low image resolution, we reduce the resolution of our original data by downsampling the image data by a power of 2 along the horizontal and vertical directions. We

define the normalized error function as $\Delta_2 = \sqrt{\frac{1}{T} \int_0^T [a_{i,j}(t) - a_{M,N}(t)]^2 dt} / \sqrt{\frac{1}{T} \int_0^T a_{M,N}(t)^2 dt}$, where $a_{i,j}(t)$ is the glottal area determined under the horizontal image resolution i ($i \leq M$) and the vertical image resolution j ($j \leq N$), and $a_{M,N}(t)$ corresponds to the maximal image resolution, where $M = 256$ and $N = 512$. Figure 7 shows the relationship between Δ_2 and the coefficient k averaged over 10 larynges, where $(M/k, N/k)$ represents the horizontal resolution $i = M/k$ and the vertical resolution $j = N/k$, k is increased from $2^0, 2^1, \dots, 2^4$. When the horizontal image resolution i is decreased from 256 below 64, the error function sufficiently increases, indicating its important effect on image edge detection. However, the error function does not significantly change even when the vertical image resolution j is decreased from 512 to 32. This shows that our automatic image detection procedure has robustness to low image resolution, and that high horizontal image resolution is more important than high vertical resolution (from anterior to posterior sides of the vocal folds) when automatically processing clinical high-speed images using our method. These results show that our method remains effective at resolutions as low as 64×32 .

Because of inconvenient operation, high-speed images recorded in clinical routine may have a lower image quality than that was used in this excised larynx study. To examine the possible clinical application, we examine the performance of our automatic image detection method to a human high speed image data. In Fig. 8(a), the high-speed image with the image resolution 128×256 was recorded from a patient (female, 48 years old) with vocal nodules using the standard procedure [30]. Despite of the lower image resolution in this clinical case, our automatic image detection method still can successfully extract glottal area curve of this clinical data, as shown in Fig. 8(b). The proposed method may be potentially applicable for extraction of vocal fold vibrations from high-speed digital images generated by patients with laryngeal diseases.

IV. Conclusion

In this study, using Lagrange interpolation, differentiation, and Canny image edge detection, we have proposed a novel automatic image processing algorithm to determine the vocal fold vibratory patterns from high-speed imaging. Other commonly used methods, such as histogram and active contour, have also been employed. Comparisons of our method with other methods have been performed on computation error and time. Our method has shown the smallest computation error and shortest computation time, which offers great advantage in effectively and efficiently processing large amounts of aperiodic high-speed image data. In particular, this method is capable of extracting vocal fold vibrations at low image resolutions. High-speed photography has recently become an important technique in the field of otolaryngology. Canine and human larynges possess physical similarities that allow the canine larynx to be used as an organic model for studying mechanisms of human phonations. Clinical applicability of our method has been further examined by extracting the glottal area series from a patient with vocal nodules. The proposed automatic image-processing algorithm may provides a valuable tools for analyzing thousand of high-speed images and offer a valuable biomedical application for the clinical diagnosis of laryngeal pathologies.

Acknowledgments

This study was supported by NIH Grant No. (1-RO1DC05522) and NIH Grant No. (1-RO1DC006019) from the National Institute of Deafness and Other Communication Disorders.

References

1. Svec JG, Schutte HK. Videokymography: high speed line scanning of vocal fold vibration. *J Voice* 1996;10:201–205. [PubMed: 8734395]
2. Yan Y, Damrose E, Bless D. Functional Analysis of Voice Using Simultaneous High-Speed Imaging and Acoustic Recordings. *J Voice*. 2006
3. Wittenberg L, Tigges M, Mergell P, Eysholdt U. Functional imaging of vocal fold vibration: Digital multislice high speed kymography. *J Voice* 2000;14:422–442. [PubMed: 11021509]
4. Baken, RJ. Clinical measurement of speech and voice. Singular; San Diego: 2000.
5. Hirano, M.; Bless, D. Videostroboscopic examination of the larynx. Singular; San Diego: 1993.
6. Yan YL, Chen X, Bless D. Automatic Tracing of vocal-fold motion from high-speed digital images. *IEEE Trans Biomed Eng* 2006;53:1394–1400. [PubMed: 16830943]
7. Westphal LC, Childers DG. Representation of glottal shape data for signal processing. *IEEE Trans Acoustics, Speech, and Signal processing* 1983;31:766–769.
8. Hong KH, Kim HK, Niimi S. Laryngeal Gestures During Stop Production Using High Speed Digital Images. *J Voice* 2002;16:207–214. [PubMed: 12150373]
9. Yan Y, Ahmad K, Kunduk M, Bless D. Analysis of vocal fold vibrations from high speed laryngeal images using a Hilbert transform based methodology. *J Voice* 2005;19/2:161–175. [PubMed: 15907431]
10. Wittenberg T, Moser M, Tigges M, Eysholdt U. Recording, processing, and analysis of digital high-speed sequences in glottography. *Mach Vis Applicat* 1995;8:399–404.
11. Jiang JJ, Yumoto E, Lin SJ, Kadota Y, Kurokawa H, Hanson DG. Quantitative measurement of mucosal wave by high-speed photography in excised larynges. *Ann Otol Rhinol Laryngol* 1998;107:98–103. [PubMed: 9486902]
12. Saadah AK, Galatsanos NP, Bless D, Ramos CA. Deformation analysis of the vocal folds from videostroboscopic image sequences of the larynx. *J Acoust Soc Am* 1998;103:3627–3641. [PubMed: 9637044]
13. Marendic B, Galatsanos N, Bless D. A new active contour algorithm for tracking vocal folds. *Proc. IEEE Int. Conf. On Image Processing*. 2001
14. Lohscheller J, Dollinger M, Schuster M, Schwarz R, Eysholdt U, Hoppe U. Quantitative investigation of the vibratory pattern of the substitute voice generator. *IEEE Trans Biomed Eng* 2004;51:1394–1400. [PubMed: 15311824]
15. Lohscheller J, Toy H, Rosanowski F, Eyesholdt U, Dollinger M. Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos. *Med Image Analysis* 2007;11:400–413.
16. Zhang Y, Jiang JJ. Spatiotemporal chaos in excised larynx vibrations. *Phys Rev E* 2005;72:35201–35204.
17. Schwarz R, Hoppe U, Schuster M, Wurzbacher T, Eysholdt U, Lohscheller J. Classification of unilateral vocal fold paralysis by endoscopic digital high-speed recordings and inversion of a biomechanical model. *IEEE Trans Biomed Eng* 2006;53:1099–1108. [PubMed: 16761837]
18. Tao C, Zhang Y, Jiang JJ. Extracting physiologically relevant parameters of vocal folds from high-speed video image series. *IEEE Trans Biomed Eng* 2007;54:794–801. [PubMed: 17518275]
19. Van den Berg J, Tan TS. Results of experiments with human larynxes. *Pract Oto-Rhino-Laryngologica* 1959;21:425–450.
20. Jiang JJ, Zhang Y, Ford CN. Nonlinear dynamics of phonations in excised larynx experiments. *J Acoust Soc Am* 2003;114:2198–2205. [PubMed: 14587617]
21. Jeffreys, H.; Jeffreys, BS. *Methods of Mathematical Physics*. 3rd ed.. Cambridge University Press; Cambridge: 1988. Lagrange's Interpolation Formula; p. 260
22. Canny J. A computational approach to edge detection. *IEEE Pattern Anal Machine Intell* 1986;PAMI-8:679–698.
23. Bracewell, RN. *The Fourier Transform and Its Applications*. 3rd ed.. McGraw-Hill; New York: 2000.

24. Bocker W, Muller W-U, Streffer C. Comparison of different automatic threshold algorithms for image segmentation in microscope images. *Proceedings of the SPIE - The International Society for Optical Engineering* 1995;2564:230–41.
25. Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. *Int J Compu Vision* 1988;1:321–331.
26. Xu C, Prince J. Snakes, shapes, and gradient vector flow. *IEEE Trans Image Processing* 1998;7:359–369.
27. Xu J, Chutatape O, Chew P. Automated Optic Disc Boundary Detection by Modified Active Contour Model. *IEEE Trans Biomed Eng* 2007;54:473–482. [PubMed: 17355059]
28. Shan ZY, Ji Q, Gajjar A, Reddick WE. A knowledge-guided active contour method of segmentation of cerebella on MR images of pediatric patients with medulloblastoma. *J Mag Res Imag* 2005;21:1–11.
29. Yezzi A, Kichenassamy S, Kumar A, Olver R, Tannenbaum A. A geometric snake model for segmentation of medical imagery. *IEEE Trans Med Imag* 2006;16:199–209.
30. Kay Elemetrics Corp. High-Speed Video (HSV) Model 9700: Instruction Manual. Kay Elemetrics Corporation; Lincoln Park, NJ: 2002.

Figure 1.
High speed digital imaging setup of excised larynx experiments.

Figure 2.

Image detection using our method. (a) A single high-speed image frame; (b) The intensity level distributions of the black line in (a) before and after Lagrange interpolation; (c) The detected glottal area with $T_l = 0.25$ and $T_h = 0.65$, where the detection of the vocal fold edge is marked by the white line; (d) The detected glottal area with $T_l = 0.1$, and $T_h = 0.8$.

Figure 3.

Detected edges from high-speed image sequences using four different algorithms. (a) High-speed image sequence; (b) Histogram; (c) Active contour method, where the circled area represents the vocal fold edge detected by this method; (d) Our automatic method.

Figure 4. Glottal area time series extracted from 100 high-speed frames of an excised larynx using manual frame-by-frame detection method and our automatic method.

Figure 5. The distribution of the error function for ten larynges of histogram, active contour algorithm, and our method.

Figure 6. The distribution of the computation time of histogram, active contour algorithm, and our method.

Figure 7.
The effects of the image resolutions on the error function.

Figure 8. Detected glottal edges from high-speed image sequences from a patient with vocal nodules using our automatic method. (a) A single high-speed image frame; (b) Glottal area time series extracted from 50 high-speed image frames.