

Automatic Tracing of Vocal Fold Edges in High Speed Laryngeal Imaging

Erik Bieging
ECE 533
Final Project

1. Introduction

Vibratory patterns of the vocal folds of interest in the field of laryngeal physiology. In the case of vocal fold pathologies, irregular motion of the vocal folds is known to occur. In order to better understand how vocal pathologies affect vocal fold vibration, video data is needed to quantify the spatial and temporal dynamics of the vocal folds. As the main functional tissue in the larynx, the vocal folds open and close rapidly while under tension when air is forced passed them. Opening between the vocal folds is known as the glottis. (Figure 1) The vocal folds oscillate at 100 to 400 Hz during normal phonation. [3]

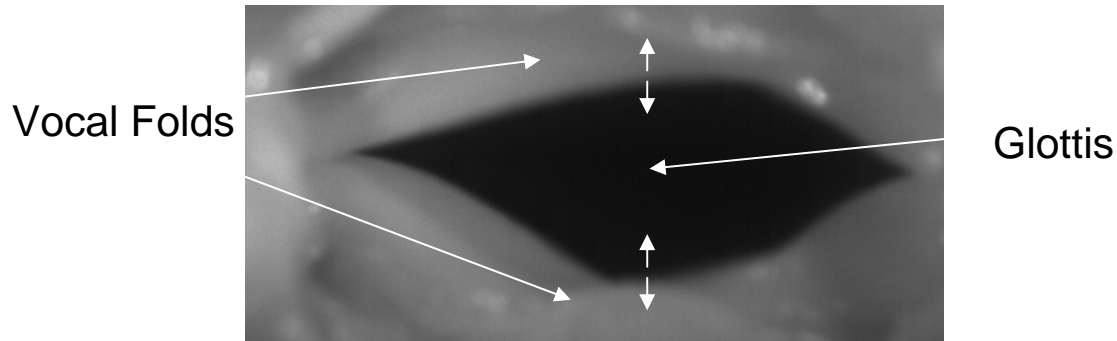


Figure 1: A typical laryngeal image obtained from an HSDI system, with the glottis and vocal folds designated.

Former methods for capturing this motion include videostroboscopy, in which a strobe light was used to illuminate the vocal folds, and images were acquired at approximately 25 Hz. This method gave an approximate picture of vocal fold motion, but the sampling rate was much too low to reconstruct the actual motion of the vocal folds.

Advances in technology have led to the use of high speed digital imaging (HSDI) systems when imaging the vocal folds. Images can be acquired at rates up to 4000 frames/sec. This sampling rate allows for complete capture of the vocal fold motion in one individual oscillation. However due to the high rate of image acquisition, large amounts of image data must be analyzed. In a typical 1 to 4 second sampling period, several thousand images are acquired and need to be analyzed. The vocal fold edges must be traced systematically in each frame of the video so that the glottal area can be extracted. Determined from an image series, a glottal area waveform can be extracted and used to determine if the vocal fold vibratory motion is normal or abnormal. [3]

2. Approach

The goal of this project is to implement and test a new method for the extraction of the vocal fold edges and the glottal area in high speed laryngeal images. This method is then compared to existing methods of edge extraction that are currently used in laryngeal physiology. This method was conceived by myself and other members of the UW-Laryngeal Physiology Lab.

2.1. Our New Method

The method developed has two main steps, differentiation and Canny edge detection. The high speed laryngeal images under investigation are all grey scale images, so each pixel has a single value denoting its intensity. The images are converted to 8-bit MatLab images from Audio Video Interleave (.avi) files, and thus have the intensity range [0, 255]. The first step is to ensure that the glottis is oriented horizontally in the image frame, as it is in Figure 1. Each image in the video is then cropped to minimize the amount of unnecessary surrounding tissue in the image frame. In most laryngeal images the significant regions are of relatively low intensity. High intensity pixels are normally caused by reflective areas on the vocal fold tissue due to moisture. To remove high intensities from the images, a user defined threshold is applied to the image such that every pixel above the threshold is set to the threshold value, eliminating errors caused by high intensities. The default value of this threshold is 160.

After these preprocessing steps take place, each column of the image is processed individually. A second user defined threshold is applied to the minimum value of each column. If the column's minimum value does not drop below this threshold, then it is assumed that there is no glottal opening in the column, and no edge needs to be detected. This threshold is normally between 40 and 80. However if the threshold is exceeded, a five-point differentiating filter is applied to the intensity levels of each column, which smoothes the function as it differentiates reducing the effect of noise. The filter's transfer function is as follows.

$$H(z) = \frac{1}{6} (2z^2 + z + z^{-1} + 2z^{-2})$$

The maximum and minimum of the differentiated column correspond to the two most rapid points in image intensity in the column. These are assumed to be the vocal fold edges. Points lying between the max and min are assumed to be part of the glottis and are given the value 1 and points lying outside are assumed to be vocal fold tissue and are given the value zero. After this process is applied to each column, a binary image is created (Figure 2).

$$L(y) = \min \left(\frac{\partial}{\partial x} I(x, y) \right)$$

$$R(y) = \max \left(\frac{\partial}{\partial x} I(x, y) \right)$$

In the theoretical equations for determining the vocal fold edges, $L(y)$ and $R(y)$ are the positions of the two vocal fold edges in column y , and $I(x, y)$ is the intensity image. The binary image is based on the following equation

$$I_{binary}(x, y) = \begin{cases} 1 & L(y) < x < R(y) \\ 0 & x < L(y) \text{ or } x > R(y) \end{cases}$$

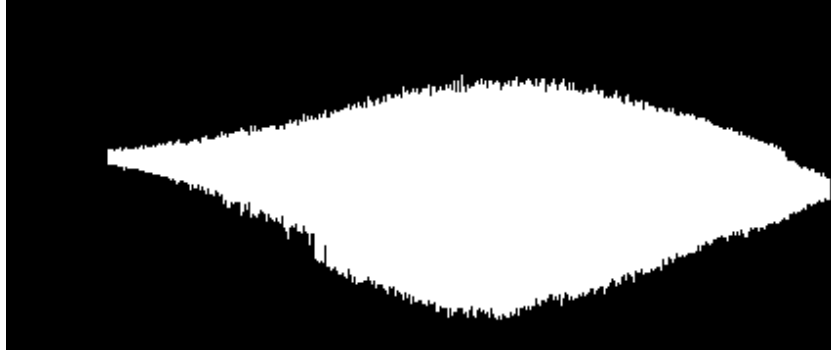


Figure 2: The binary image found after the differentiation step has taken place.

This process alone gives a reasonable approximation of the vocal fold edge location. However due to inconsistencies in the location of the maximum and minimum slope in intensity, the edge is very rough. The Canny edge detection method is used to smooth the vocal fold edges, and is implemented automatically in MatLab. In Canny edge detection, an edge detection mask is applied to the image and two thresholds are used to determine the edge points. The first threshold is used to start an edge, and the second, lower threshold is used to continue an existing edge. This method both smoothes the edge and eliminates some unwanted edge points when errors occur in the first step of the algorithm. In this way, if an error occurs in the differentiation step in a single column, it will not affect the final edge. The output after the Canny edge detection step is shown with the original image in Figure 3.

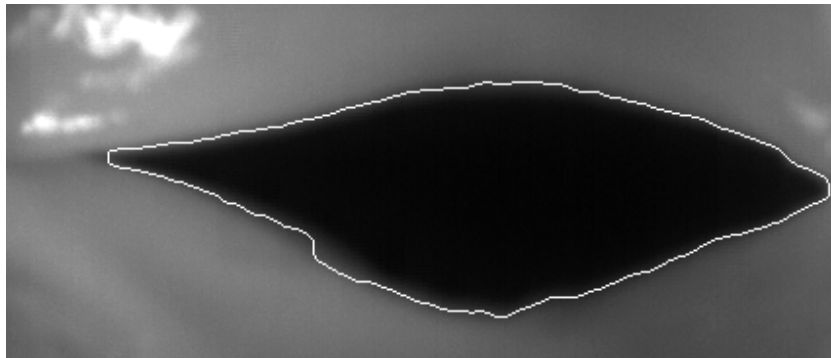


Figure 3: The result of the edge detection algorithm after Canny edge detection has taken place.

From the Canny edge detection algorithm new functions denoting the glottal edge $L'(y)$ and $R'(y)$ are obtained. To extract the glottal area from the frame, the glottal width function is first calculated to be $Wid(y) = |L'(y) - R'(y)|$. Although infrequent, errors can occur in the glottal edge detection that causes the width function to change drastically from one point to the next. To account for these drastic jumps in width, a filter is applied to the width function. The function is differentiated to obtain points a function of drastic increase using the following equation.

$$J(y) = \frac{d}{dy} |L'(y) - R'(y)|$$

A user defined threshold is then applied to $J(y)$, to obtain a set of points where $J(y)$ crosses the threshold. Every two crossings are interpreted as a range of error in the width function. Each area of error is replaced using linear interpolation of points on either side of the area of error. This gives a new width function $Wid'(y)$. The glottal area is then determined by integrating this function over the length of $Wid'(y)$.

$$GA = \sum_{y=0}^N Wid'(y)$$

This complete process is applied to each image in the video, and the glottal area curve is extracted from the image series.

2.2. Other Methods

To verify the effectiveness of the above method, other methods that are currently used for vocal fold edge detection were investigated for comparison. The methods investigated are histogram, region growing, and active contour.

2.2.1. Histogram

Histogram image segmentation methods are a simple thresholding method in which the threshold is determined based on the histogram of each image. The glottis is in the range of darker intensity levels than the vocal fold tissue, so the correct threshold can be used to differentiate between the glottis and the vocal folds. The threshold is determined based on the probability distribution of the object pixels and the background pixels. A typical histogram for a high quality laryngeal image is shown in Figure 4.

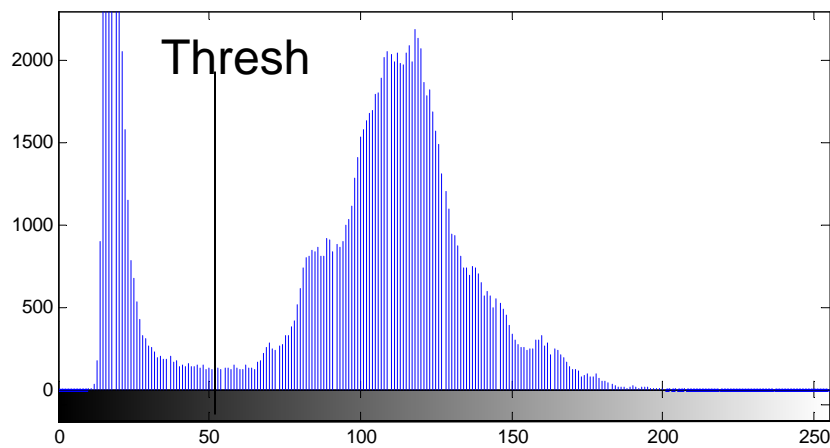


Figure 4: A typical histogram for a laryngeal image. The threshold is marked approximately where it would be found using the histogram method.

2.2.2. Region Growing

Region growing methods are based on starting with a single pixel and then growing a region based on similarity to the surrounding pixels. Thus the image is divided into homogeneous regions. In laryngeal images, starting pixel, or seed, is ideally placed with the glottal region. The region grows outward from this point until large increases in intensity are reached at the glottal edges. [4]

2.2.3. Active Contour

Active contour methods involve a defined initial region, $v(s)$, (snake) that is deformed iteratively based on the gradient of the images. The iteration aims to minimize the energy of the snake based on the following equation.

$$E = \int_s \frac{1}{2} (\alpha |v_s|^2 + \beta |v_{ss}|^2) + E_{ext}(v(s)) ds$$

v_s and v_{ss} represent the first and second derivatives of v with respect to s , and α is a measure of the snake's tension and β is a measure of the snake's rigidity. E_{ext} is the external energy acting on the snake determined from the image gradient. α and β tend to shrink the region while E_{ext} causes the region to expand. [2] In laryngeal images, the initial region must be set automatically, and the histogram method is commonly used. As a set number of iterations are performed, the snake tends toward the vocal fold edges. [1]

3. Work Performed

A MatLab program and GUI were used to implement our new algorithm, as well as each of the other listed image detection methods. The GUI was designed, and each of the associated functions was written. Code from software previously designed in the UW Laryngeal Physiology Lab was incorporated into the new program. The histogram method was implemented originally, and the active contour and region growing algorithms were implemented using freely distributed code designed by others. Also a program was designed such that the user could manually determine the glottal edge based on direct observation of each laryngeal image.

In order to demonstrate the effectiveness of the new edge detection algorithm, a representative sample of 10 laryngeal high speed videos were chosen. The method, along with the three other popular methods, was applied to the first 100 frames of each video. The glottal area in each frame was calculated from each of the methods, and these areas were compared to those determined manually by looking at each individual frame and verifying the calculated area. In addition, a timing function was built into each program so that the processing time for each algorithm could be recorded and compared.

4. Results

In the case of high quality image data and regular vocal folds, all of the included methods adequately extracted the glottal area from the image data. Figure 5 shows regular vocal

folds in desirable recording conditions. Each of the four algorithms adequately detects the vocal fold edge, and computes the glottal area. Figure 6 shows vocal folds which are oscillating in an irregular pattern and under less than ideal recording conditions. In this case, our method correctly detects the glottal edge, while other methods fail. The histogram method is fairly accurate, but the threshold has been set slightly below the optimal value, making the glottal area slightly below its actual value. The region growing method detects dark regions outside of the glottis, and tabulates them as part of the glottal area. In the active contour method, the initial region is set incorrectly due to error in the histogram thresholding, making the final area misshaped and inaccurate. Our method provides the most accurate detection of the glottal edge.

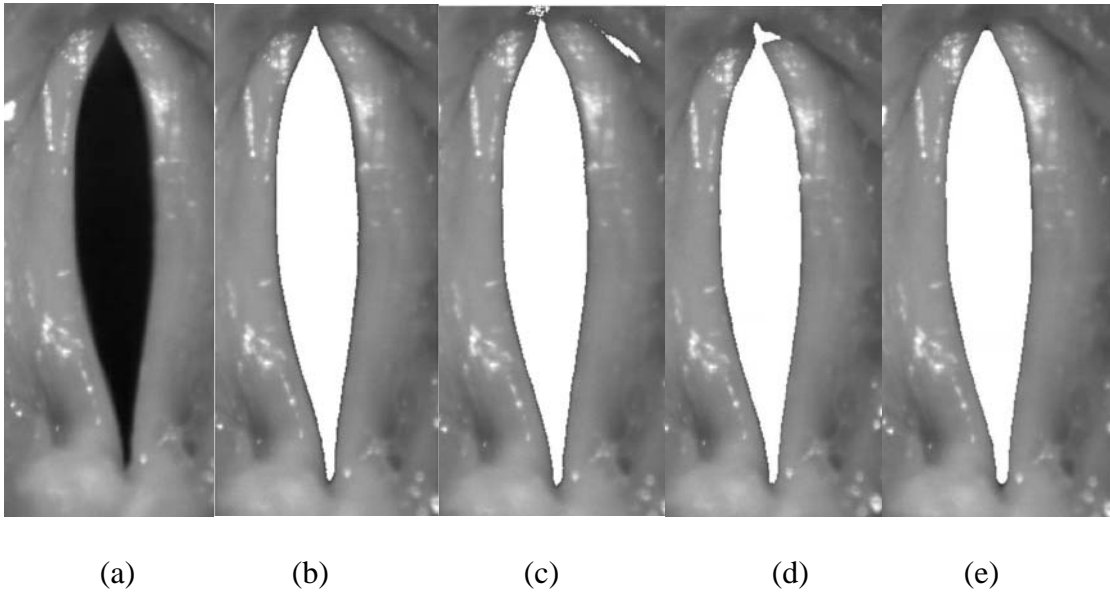


Figure 5: Edges detected using four different algorithms in high quality regular vocal folds. (a) Original image (b) Histogram (c) Region Growing (d)Active Contour with Histogram (e) Our Method

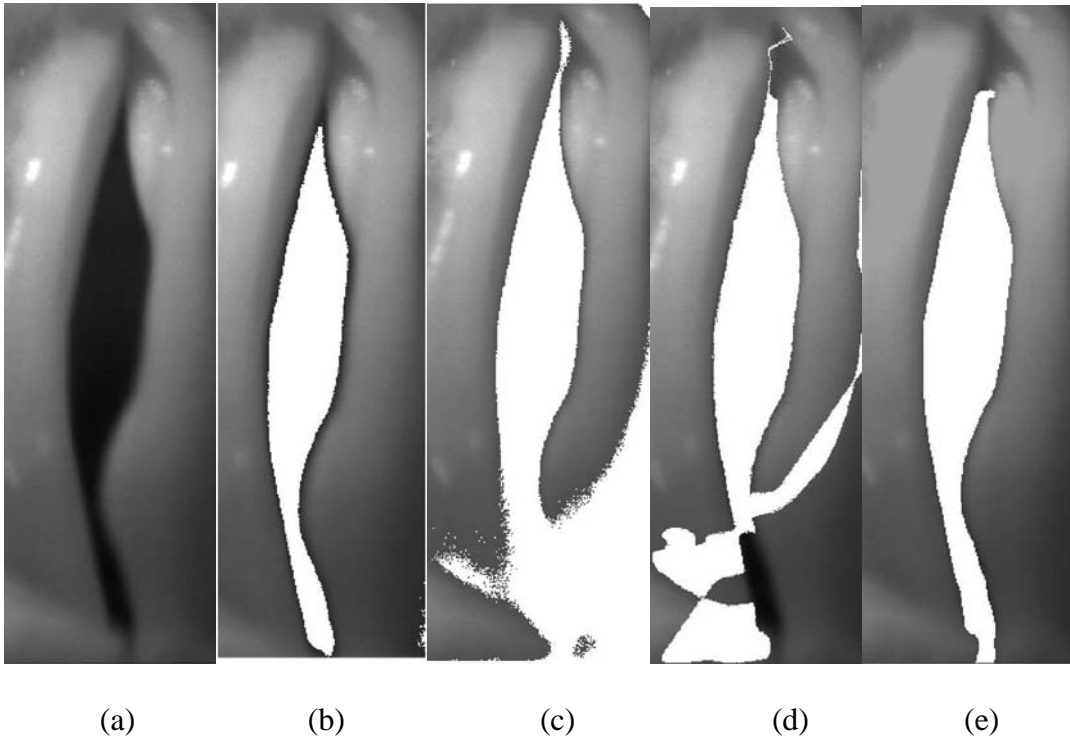


Figure 6: Edges detected using four different algorithms in low quality irregular vocal folds. (a) Original image (b) Histogram (c) Region Growing (d)Active Contour with Histogram (e) Our Method

The extracted glottal area curves for one of the selected videos are shown in Figure 7, and are compared to the manually determined glottal area. The histogram and region growing methods both become inaccurate when the glottal area is small. The active contour method becomes inaccurate when the glottal area is large. Our method accurately detects the glottal area throughout the sample period.

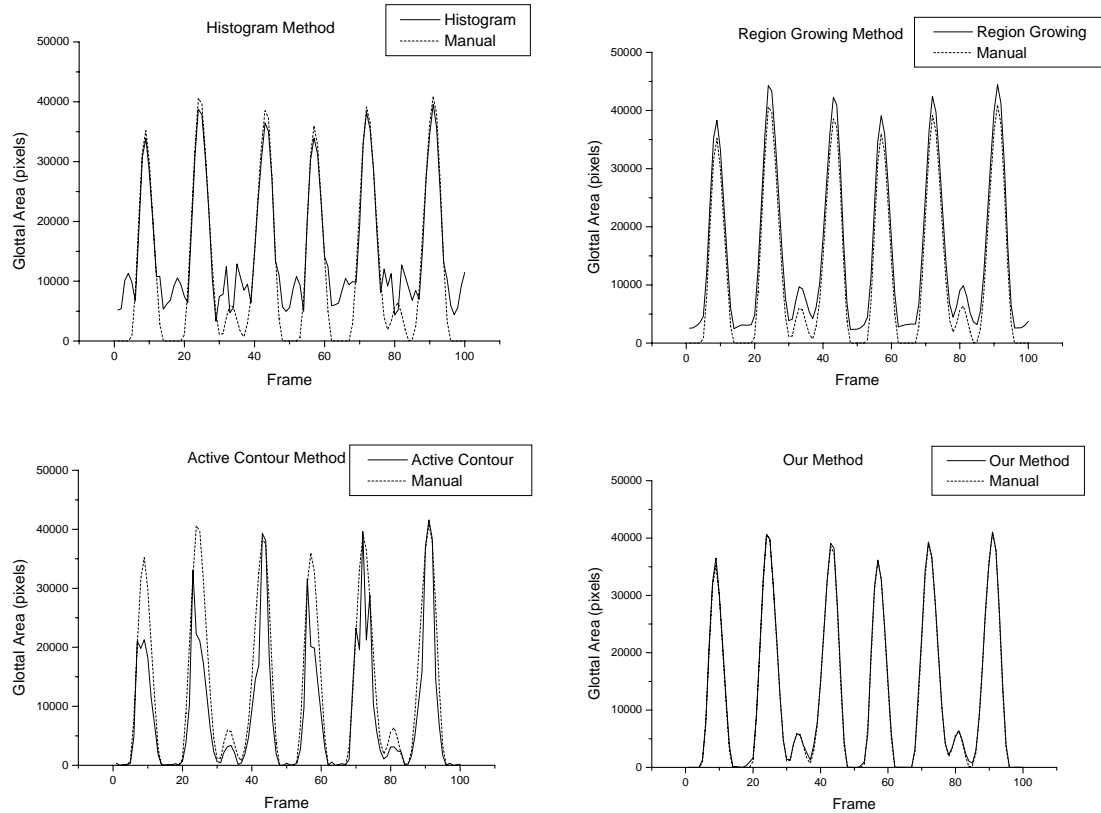


Figure 7: Glottal area curves extracted using each method and plotted against the manually determined glottal area.

The results from 100 frames of ten different high-speed vocal fold videos are shown in Table 1. In all sample videos but one, our algorithm shows the lowest deviation from the manually determined glottal area. When the deviations are averaged, our method deviates from the actual glottal area three times less than any other method.

Video	Histogram	Region Growing	Active Contour	Our Method
1	110.66%	197.55%	62.98%	15.53%
2	18.89%	26.58%	38.73%	5.71%
3	3.93%	5.50%	40.38%	1.85%
4	45.64%	16.31%	44.95%	4.02%
5	30.80%	21.58%	43.85%	18.72%
6	10.22%	7.31%	57.54%	14.02%
7	30.69%	18.21%	34.11%	3.05%
8	6.23%	4.42%	41.13%	4.75%
9	47.13%	47.29%	38.92%	12.97%
10	10.75%	10.73%	39.77%	10.60%
Average	31.49%	35.55%	44.24%	9.12%

Table 1: Deviation comparison of four algorithms

In order to show the efficiency with which our method processes image data, each of the four algorithms were timed while processing each set of 100 frames of data. The results are shown in Table 2. Our algorithm processes the data faster than any of the others. The region growing and active contour methods require numerous iterations of detailed computation, making them much more computationally expensive compared to the histogram method or our method.

Video	Histogram	Region Growing	Active Contour	Our Method
1	2.2	40.6	88.7	1.5
2	2.2	41	80.4	1.5
3	2.3	40.8	55.5	1.1
4	2	40.9	65.8	1.3
5	2.3	40.9	79.8	1.3
6	1.8	41	74.7	1.1
7	2.5	41.8	52.3	1
8	2.4	40.9	59.2	1.2
9	2.1	41	56.2	1.1
10	1.1	40.9	64.3	1
Average	2.09	40.98	67.69	1.21

Table 2: Speed comparison of four algorithms (in minutes)

5. Discussion

Our method for vocal fold edge detection takes advantage of the specific properties of high-speed vocal fold video data. The glottis is the single object in the images that should be detected. The glottis can be divided into two or more separate objects in a frame, but we know that there can be either two or zero boundary points at each position along the glottal axis. Since the glottal axis is oriented horizontally in the image frame, two (or zero) edges will be detected in each column of the image. By searching for a maximum of two edge points, the effect of noise and image irregularities is minimized. The algorithm does not use fixed thresholds to find edge points. It simply finds the two most significant boundaries and eliminates all others.

The computation time table shows the simplicity of our algorithm. Our algorithm has a set sequence of computations in extracting the glottal edge. Region growing and active contour algorithms require numerous iterations to find the boundaries. The calculations are carried out numerous times resulting in long computation times. Our method, as well as the histogram method, uses one direct process to determine the vocal fold edges, making them much more efficient.

Histogram methods are limited by the quality of the histogram of each image. In order for histogram methods to be effective there must be a distinct gap between the object and

background intensities. This would be shown by two distinct peaks in the histogram, which is frequently not the case in vocal fold images. In addition, histogram methods apply a single threshold to the entire image, which does not account for changes in intensity boundary within the image.

Region growing methods require one or more initial regions or seeds. Thus their effectiveness is limited by our ability to automatically initialize the region correctly. If a seed is placed outside of the desired object, i.e. the glottis, the algorithm will incorrectly detect an area outside of the actual glottal area. In lower quality recording conditions, there is a high probability that the algorithm will be initialized incorrectly, making it ineffective.

Setting a correct initial region becomes a problem in active contour methods as well. Active contour methods are highly dependant on initial region and other parameters. Histogram methods can be used to set the initial region, but they will only be effective if the histogram method alone is reasonably accurate, which is not always possible with lower quality image data. Our method does not require an initial region or many significant parameters. The method is not very sensitive to the few parameters involved; their default values can be used successfully in almost all cases.

Our new method overcomes many of the limitations that other popular methods, and has been shown to be an effective method for vocal fold edge detection. Its implementation into the field of vocal fold physiology may lead to a better understanding of vocal fold motion.

6. References

- [1] Marendic, B.; Galatsanos, N. and Bless, D. "A new active contour algorithm for tracking vibrating vocal folds." *Proceedings 2001 International Conference on Image Processing.*, pt. 1, vol.1, pp. 397-400. 2001.
- [2] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models.", *International Journal of Computer Vision.* vol. 1, n. 4, pp. 321-331, 1987.
- [3] Wittenberg, T., Moser, M., Tigges, M., and Eysenholdt, U. "Recording, processing, and analysis of digital high-speed sequences in glottography." Vol. 8 pp. 399-404. 1995.
- [4] Yuling, Y., Chen, X., and Bless, D. "Automatic tracing of vocal fold motion from high speed digital images." *IEEE Trans. Biomed. Engr.* Vol. 53-7 pp. 1394-1400. 2006.

7. Work Breakdown

Activity	Me	Others in Lab
Implementation of our method	90%	10%
Implementation of other methods	50%	50%
Data processing	100%	0%
Presentation	100%	0%
Report	100%	0%

Erik Bieging