

# GlottalImageExplorer 1.0 Manual

Peter Birkholz

*Institute of Acoustics and Speech Communication, TU Dresden,  
Germany*

December 2, 2015

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Graphical user interface</b>	<b>2</b>
2.1	Menu . . . . .	3
2.2	Control panel . . . . .	3
2.3	Images . . . . .	3
2.4	Area waveform . . . . .	4
<b>3</b>	<b>Segmentation guide</b>	<b>4</b>
	<b>References</b>	<b>5</b>

## 1 Introduction

GlottalImageExplorer (GIE) is a PC software to support the segmentation of the glottis in endoscopic high-speed videos of the larynx. The software processes videos with a resolution of 256x256 pixels, which is the resolution of the ENDOCAM high-speed camera systems of the company Richard Wolf GmbH, Knittlingen, Germany. You can export the results of the segmentation in terms of the glottal area waveform or the glottal contours in the images as text files.

GIE is free and open source. It is available for download at [www.vocaltractlab.de](http://www.vocaltractlab.de) and contains the source code along with the project files for Microsoft Visual Studio Professional 2013, and the executable for the Windows platform. The executable "GlottalImageExplorer.exe" has been tested under Windows 7 and Windows 10 (64 bit) and can be started without any special requirements.

GIE was written in C++ using the cross-platform GUI library wxWidgets 2.8.12 (<https://www.wxwidgets.org/>). Therefore, if you want to compile GIE, this library must be installed. Note that by now there is the new major version 3.x.x of wxWidgets, which is not fully compatible with wxWidgets 2.8.12 used for GIE. Apart from wxWidgets, no other libraries are required. If you want to use Visual Studio to compile the program for Windows, you will have to adapt the project settings, especially the path settings to the wxWidgets library. Since wxWidgets is a portable library, it should be possible to compile the sources on other platforms. However, there are a few parts of the GIE source code that use Windows-specific function calls (for example for loading AVI files) that would have to be adapted to the other platform.

Note that GIE comes with no warranty of any kind. The whole software is under continual development and may undergo substantial changes in future versions. Please feel free to report any bugs to [peter.birkholz@vocaltractlab.de](mailto:peter.birkholz@vocaltractlab.de).

The segmentation method implemented in this software is based on the paper by Lohscheller et al. (2007). In brief, the glottis in each frame (image) is segmented by seeded region growing, a well-established method in computer vision. Starting with an initial set of three seed points (pixels) in the area of the glottis, the method evaluates the intensity of the neighboring pixels and adds them to the segmented area if their intensity is below a certain threshold (the glottis is darker than the surrounding vocal folds). This process is iterated until no more neighboring pixels satisfy the threshold criterion. As proposed by Lohscheller et al. (2007), we do not use a single threshold for a whole image, but the threshold may vary for each pixel row. This is an important concept, because it allows to compensate for gradients in the illumination of the larynx. The threshold progression is defined by three threshold points, each of which specifies the threshold value at a certain y-coordinate. Between the threshold points, the threshold value is linearly interpolated. Both the threshold progression and the seed points can be set by the user for any frame of the film. In the current version, the number of seed points and threshold points that can be adjusted in a given frame if fixed to three. This proved to be sufficient for the videos that we analyzed so far. Between the frames with user-defined settings, seed positions and threshold points are linearly interpolated over time. In most films, it is sufficient to make adjustments in a handful of frames to achieve a satisfactory segmentation of the glottis in all frames, because the illumination conditions vary rather slowly from one frame to the next.

For a more elaborate description of the method, please refer to Birkholz (2016) and Lohscheller et al. (2007). When you use the software for your research, please cite Birkholz (2016).

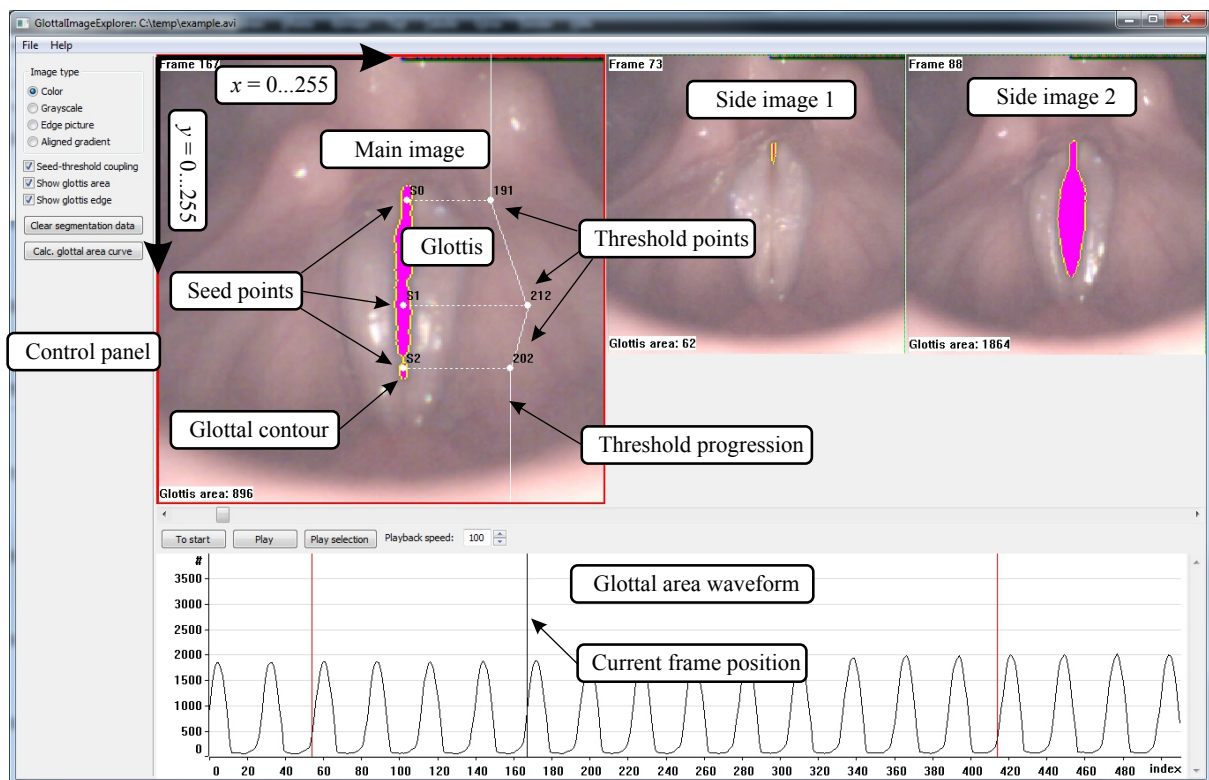


Figure 1: Graphical user interface of GlottalImageExplorer.

## 2 Graphical user interface

Figure 1 shows the graphical user interface of the program. Its main components are a menu, a control panel at the left side, three images in the top right area, and the glottal area waveform at the bottom.

## 2.1 Menu

The File menu allows to load a film from an AVI file (\*.avi), a raw data file of the HRES ENDOCAM system of the company Richard Wolf GmbH (\*.bld), or from a set of BMP images (\*.bmp) in a folder. The image size in the AVI and BMP files must be 256x256 pixels.

Furthermore, there are two menu items to load and save the segmentation data (\*.seg). The segmentation data are the user-specified threshold and seed point settings. They are saved in a simple text format, where each line represents the seed points and threshold points of one frame.

Finally, there are menu items to export the glottal area waveform and the glottis contour waveform, both as text files (\*.txt). The text file for the glottal area waveform simply consists of one glottal area value (number of pixels segmented as the glottis) for each frame in the film. The text file for the glottis contour waveform contains two lines for each frame that specify the left and right part of the glottis contour. Each contour consists of the  $x$ -values of the pixels in the pixel rows 0 (top) to 255 (bottom). For the pixel rows above and below the glottis, the  $x$ -values for the contours in the file are set to zero. The glottal contours written to the file are meant for further processing or plotting in other programs like MATLAB.

## 2.2 Control panel

The upper part of the control panel contains radio buttons to specify the way the frames of the film are displayed: either as color images (original data), grayscale images, edge pictures or "aligned gradient" pictures. The grayscale images are used for the actual glottis segmentation, i. e., as basis for the region growing method. The edge and gradient pictures emphasize the edges in an image. This can help you to adjust the region growing thresholds (see below) such that the contours of the segmented glottis correspond well to the real edges of the glottis.

Below the radio buttons there are three checkboxes. When the checkbox "Seed-threshold coupling" is checked, the  $y$ -coordinates of the threshold points automatically take the  $y$ -coordinates of the seed points, i. e., they cannot be changed independently. In this case, the threshold values for the region growing algorithms are always specified at the vertical positions of the seed points. When the checkbox "Show glottis area" is checked, the segmented glottal area is drawn with purple color in the images. When the checkbox "Show glottis edge" is checked, the contour of the segmented glottis is drawn in yellow.

Below the checkboxes are the buttons "Clear segmentation data" and "Calc. glottal area curve". The first button clears all user defined adjustments of seed points and threshold points. The second button (re-)calculates the glottal area waveform based on the current seeded region growing settings. After the calculation, the area waveform is shown in the bottom part of the program window.

## 2.3 Images

Three frames of the film are shown as images in the main part of the program window. The left (main) image shows the *current* frame, where the user can modify the position of the seed points and the threshold points. The two images at the right side (side images) show two frames that can be selected by the user to facilitate visual comparisons of frames at different positions in the film. The current frame can be selected to be shown in one of the side images with the context menu (right-click) in the main image. The current frame shown at the left can be set using the scrollbar below the images or changed to the previous or the next frame with  $\text{Ctrl} + \leftarrow$  or  $\text{Ctrl} + \rightarrow$ . For every frame you can choose to specify the segmentation data manually or let the software interpolate these data between the neighboring frames with manually specified data. For the current frame, you can select manual segmentation with the item "User defined segmentation" in the context menu of the main image. If "User defined segmentation" is selected, the main image gets a red border and the seed points and threshold points can be dragged with the mouse. The three seed points are labeled with "S0", "S1", and "S2", and must be moved into the area of the glottis. Thereby, S0 is always the uppermost and S2 the lowermost of the three points.

There are three threshold points that define the threshold progression for the seeded region growing from the top to the bottom of the frame. Each of these points specifies the threshold value at a certain  $y$ -

coordinate of the frame. The threshold value associated with a point is represented by its  $x$ -coordinate, i.e., the further left a threshold point the lower the threshold value, and reverse. The threshold values are written next to the points. The threshold progression is defined by the linear interpolation between the threshold points and indicated as a white line in the image. If the threshold points are coupled to the seed points, only the threshold *values* can be changed ( $x$ -coordinate), but their  $y$ -coordinates are forced equal to the seed points'  $y$ -coordinates. In this case, seed points and threshold points are connected by dashed horizontal lines.

If selected in the control panel, the segmented region is shown as a purple area and the glottal contour is shown in yellow.



Below the scrollbar for the current frame are buttons to play the sequence of frames as a movie. The button "Play" starts to play the sequence from the current frame on, and the button "To start" sets the first frame as the current frame. The "Playback speed" can be set to any number between 1 and 100, where 1 corresponds to very slow playback and 100 to the fastest possible playback of the frame sequence. The actual number of frames per seconds depends on the speed of your computer.

## 2.4 Area waveform

The glottal area waveform shows the time function of the glottal area according to the segmentation for a certain range of frames around the current frame. Note that **you must manually trigger the calculation of the glottal area** in all the frames with the button "Calc. glottal area curve" in the control panel, because the calculation takes a couple of seconds. The current frame is indicated by a black vertical line, and the frames with user-defined segmentation data are shown as red vertical lines. Using the context menu of the area waveform panel you can zoom in and out and scale the amplitude of the waveform.

## 3 Segmentation guide

This is a brief step-by-step guide to obtain the glottal area waveform of a laryngoscopic film. The software comes with a short example film (example.avi) with 3910 frames recorded at a frame rate of 4000 frames/s. Start the program and load this film by selecting "Load AVI film" from the file menu. You should now see the first frame (frame 0) of the film in all three images in the main part of the program window. Right-click into the main image and choose "User defined segmentation" from the context menu. The image has now a red frame and it is possible to drag the seed points (S0-S2) and the threshold points in the image with the mouse. Place the seed points along the centerline of the glottis with S0 near the upper end, S1 in the middle, and S2 near the lower end, similar to Figure 1. Then move the threshold points (those connected by the vertical white line) further to the right to increase the threshold values. Move them so far that the yellow contour appearing around the glottis closely traces the glottal edges in the image. If you increase the threshold values too much, the structures around the glottis will also be wrongly segmented as the glottis. Your threshold progression and glottal contour should look similar to Figure 1. Note that the  $y$ -coordinate of the threshold points can not be changed, because the checkbox "Seed-threshold coupling" on the control panel is checked. If you uncheck this box, you are free to move the threshold points also vertically. You can choose "Edge picture" on the control panel to see how well the yellow contour actually corresponds to the glottal edges and re-adjust the thresholds for a better fit. To see the region that is actually segmented with the selected seed and threshold settings, check the box "Show glottis area" on the control panel.

After the segmentation for frame 0, you can step through the following frames using the scrollbar or  +  and see, how well the segmentation settings work for the other frames. When you encounter a frame where the segmentation fails, for example because the segmented area extends beyond the real glottis, you can manually specify the seed points and threshold points for this frame again (select "User-defined segmentation" in the context menu of the main image). For the frames between frame 0 and the new frame, the seed point and threshold point positions are automatically interpolated. You should now check the other frames and make user-defined segmentation settings whenever the segmented glottis

seems incorrect.

When you press the button "Calc. glottal area curve", the segmented glottis will be calculated for all frames based on your settings, and the area waveform will be displayed as a graph at the bottom. Use the context menu of the graph to select the range of frames to show and the amplitude scaling of the graph. When you are happy with the results, don't forget to save the segmentation data, which you can later load to continue or improve the segmentation. For the film "example.avi" there is a corresponding example segmentation file provided with the software.

Finally, you can export the glottal area waveform or the glottal contours in the frames as text files using the File menu.

## Acknowledgments

I thank Simon Preuß for proof-reading this manual.

## References

- Birkholz, Peter (2016). "GlottalImageExplorer - An open source tool for glottis segmentation in endoscopic high-speed videos of the vocal folds". In: *Studentexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2016*. Ed. by Oliver Jokisch. TUDPress, Dresden.
- Lohscheller, J. et al. (2007). "Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos". In: *Medical Image Analysis* 11.4, pp. 400–413.