# Computerized Analysis of Vocal Folds Vibration From Laryngeal Videostroboscopy

***Silvia Gora, *Noy Yavin, *David Elad, †,‡Michael Wolf, and †,‡Adi Primov-Fever,** *‡*Tel-Aviv, and* †*Tel Hashomer, Israel*

**Summary: Objectives.** To develop an objective analysis of laryngeal videostroboscopy (VSS) movies in the space-time domain for quantitative determination of the true vocal folds (TVFs) vibratory pattern to allow for detection of local pathologies at early stages of development.

**Methods.** Contours of the TVF and false vocal folds (FVFs) were tracked on each frame of a VSS movie. A registration algorithm was used with respect to the centerline of the FVF to eliminate movements not related to TVF vibration. The registered contours of the TVF were analyzed in time and frequency domains.

**Results.** The TVF vibration demonstrated a sinusoidal pattern with the same fundamental frequency at every section along the folds of healthy subjects, as well as detection of an abnormal area with a different fundamental frequency in TVF with local pathologies. Analysis of the TVF vibration time delay of healthy subject revealed a posterior-to-anterior longitudinal wave that was not detected by visual observation.

**Conclusions.** An objective analysis of laryngeal VSS movies was developed for quantitative determination of the TVF vibration. This analysis was able to detect and quantify TVF characteristics in normal subjects as well as in patients with pathologies beyond the ability of examinee's naked eyes.

**Key Words:** Vocal folds vibration–Active contours–Videostroboscopy.

## INTRODUCTION

Laryngeal videostroboscopy (VSS) is the preferable diagnostic method of true vocal fold (TVF) vibration and serves as the gold standard for clinical evaluation of vocal folds (VFs) pathologies.[1] The acquired data are a sequence of images, each taken from a successive cycle in a slightly different phase, which yields the illusion of a complete glottal cycle called "the stroboscopic glottal cycle."[2,3] It enables visual exploration of the closure of the TVF, the mucosal wave movement, and the symmetry and periodicity of vibrations.[4] However, the VSS interpretation is a subjective physician-dependent observation, and thus, development of an objective and precise evaluation is of paramount importance.

Many studies were conducted to develop objective tools for processing and optimization of recorded images of laryngeal VSS for improvement of the diagnosis and evaluation of VF disorders. The first step for any analysis was segmentation of the TVF, which required overcoming illumination, movement, and resolution problems. The methods that were implemented included region growing,[5] thresholding and edge detection,[6] level set segmentation,[7] motion estimation,[8] morphology,[9] and different variations of active contours.[10–15] However, tracking of the TVF contours is still problematic at TVF closure, when the glottal area between the TVF is minimal.

Various post-segmentation methods were proposed for analysis of the TVF performance that was acquired either via VSS or high-speed videoendoscopy (HSVSS). These methods include fitting an affine model to the TVF displacements,[10] Nyquist plot to study opening and closing pattern of the TVFs,[6,16] comparison of contour displacements, connected component analysis of the glottal area,[12] TVF opening angle,[13] deviation of the TVF boundaries from the medial axis in time and space, and the mean deviation per each point along the boundary.[15] These analyses were used to explore the different patterns of TVF opening and closing in different populations of healthy subjects, as well as in pathologic cases. Most of these evaluations were used on the data from the glottal area.

In this work, we used a new method for time and space analysis of VSS movies for exploration of the dynamics of the TVF. The space analysis may be used to objectively explore development of irregularities and abnormal structures that may allow for more precise clinical diagnosis and then improved medical monitoring.

## METHODS

### Study design

Videostroboscopic movies of the vibrating TVF were recorded in the Department of Otolaryngology of the Chaim Sheba Medical Center using a digital EndoSTROB system and a rigid endoscope (model DX Ls6035; XION Medical GmbH, Germany). The recordings were acquired in audio video interleave (AVI) format at a rate of 25 frames/s. The study was composed of four subjects: two healthy volunteers and two patients with a pseudocyst on one of their TVFs. The study was approved by the hospital ethical committee (#0415-13-SMC) and the subjects signed an informed consent form.

A sequence of at least three periods of vibratory cycles in the same pitch was selected from the VSS movies from each subject. The frames of the selected footage were sampled into a sequence of bitmap (BMP) files of resolution 576 × 720 for further processing using MATLAB by MathWorks. An example

Accepted for publication May 29, 2015.

S.G. and N.Y. have contributed equally to this work.

From the *Department of Biomedical Engineering, Faculty of Engineering, Tel Aviv University, Tel Aviv, Israel; †Department of Otorhinolaryngology, Chaim Sheba Medical Center, Tel Hashomer, Israel; and the ‡Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel.

Address correspondence and reprint requests to David Elad, Department of Biomedical Engineering, Faculty of Engineering, Tel Aviv University, 69978 Tel Aviv, Israel. E-mail: elad@post.tau.ac.il

Journal of Voice, Vol. 30, No. 4, pp. 478-484
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2015.05.021

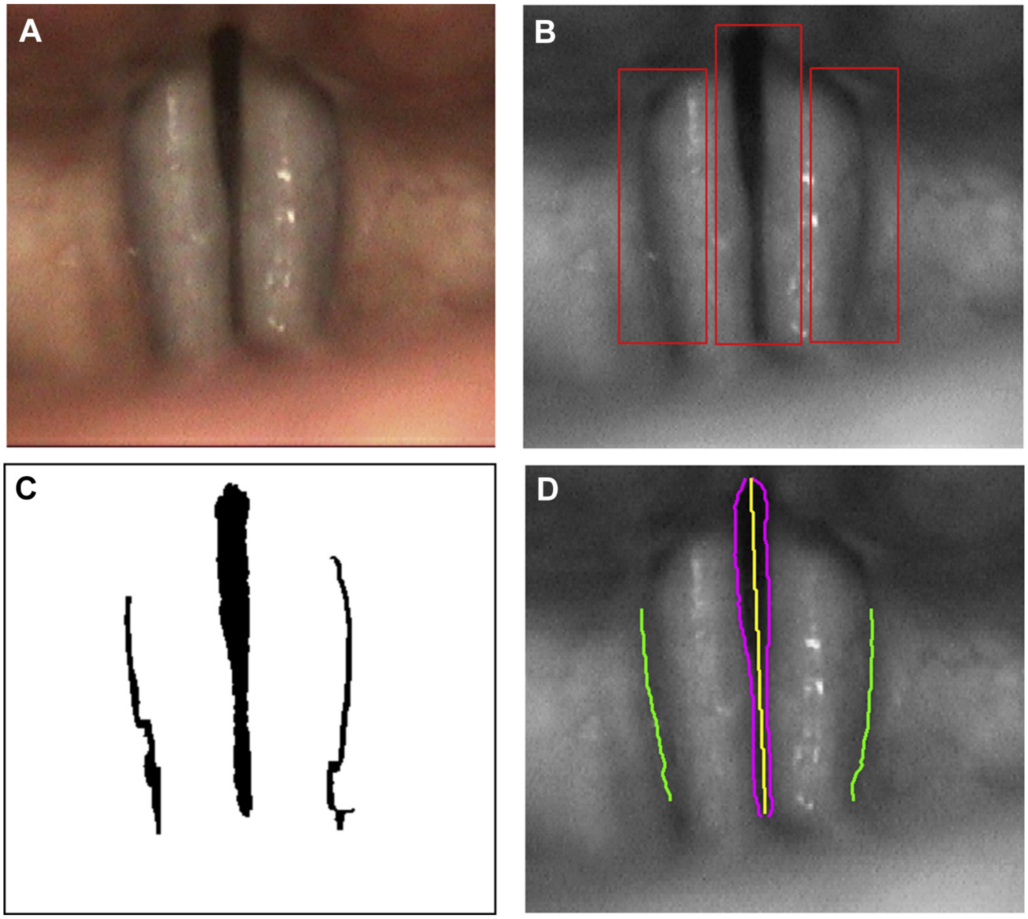


**FIGURE 1.** Description of the procedure for detection of the outline contours of the true vocal folds (TVFs) and the false vocal folds (FVFs) on each frame. (A) The region of interest of the original image. (B) The gray level image with specified regions for the TVFs and the left and right FVFs. (C) The regions of the glottal area and the left and right FVFs after filtration and thresholding. (D) The detected outlines of the TVF (*purple*) and the FVF (*green*) with the centerline of the FVF (*yellow*). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

is shown in Figure 1A. The analysis of the dynamic characteristics of the TVF was composed of (1) tracking the contours of the TVF in each frame; (2) registration of the data to filter noise not related to TFV vibration; and (3) dynamic analysis of the TVF vibration.

### Contour tracking

Image processing for contour tracking of the TVF started with selection of a region of interest (ROI) that included the TVF and FVF on all frames from a sequence of a single subject (Figure 1A). Then, the regions of the TVF and of left and right FVFs were manually marked on the first frame (Figure 1B). The following steps of the processing included texture classification using Gabor filter to emphasize the differences between the TVF, the glottal area and the FVF,[17] local thresholding (based on the preselected border points), and connected components. The outcome at this stage was a binary image of the glottal area surrounded by the left and right FVF bands (Figure 1C).

In the final stage of the tracking analysis, we implemented the greedy active contour algorithm[18] to identify the contours which trace the TVF and FVF outlines. The algorithm was modified for open contours by limiting the movement of the contour end points. The required initialization for the snake algorithm was done automatically by sampling points at the binary boundaries of the segmented image (Figure 1C). This avoided, at this stage, the need to take care of big movements between frames. The contours that outline the TVF (purple curves) and the FVF (green curves) were obtained within six iterations (Figure 1D). Examples of the tracked contours on different frames of the VSS are demonstrated in Figure 2.

The acquired images contained movements not related to the TVF functional vibrations due to patient's breathing and examiner's hand instability. These movements were filtered before the dynamic analysis by registration of all images with respect to a quasistationary feature on the acquired images. Further examination of the video recordings revealed that during phonation, the FVF movements between frames are negligible and appear to be symmetrical. Accordingly, we also assumed in this study that the FVF contours were symmetric and calculated their centerline (see the *yellow line* in Figure 1D), which will be used for registration of the extracted contours. The TVF and FVF contours from all the frames of the video recording, as