OBJECTIVE VOCAL FOLD VIBRATION ASSESSMENT FROM VIDEOKYMOGRAPHIC IMAGES

S. Bianchi, L. Bocchi, C. Manfredi, G. Cantarella*, N. Migali*

Department of Electronics and Telecommunications, Università degli Studi di Firenze
Via S. Marta 3, 50139 Firenze, Italy

* Otolaryngology Department, University of Milan, Ospedale Maggiore IRCCS, Via F. Sforza 35, 20122 Milano, Italy

Abstract - Vocal folds oscillation crucially influences all the basic qualities of voice, such as pitch and loudness, as well as the spectrum. Stroboscopy provides the standard view of the larynx. Videokymography is a new diagnostic tool developed to overcome specific limitations of stroboscopy in severely dysphonic patients with an aperiodic signal. It registers the movements of the vocal folds with a high time resolution on a line perpendicular to the glottis.

The main focus of this paper is on measuring and tracking quantitative parameters for objective vocal fold function assessment from videokymographic (VKG) examinations of subjects with normal and pathological laryngeal function. Active contour search is realised, by properly adjusted snake algorithm. Examples are reported, showing the robustness and reliability of the proposed technique.

I. INTRODUCTION

Vibration of the vocal folds is a highly relevant aspect of voice production, both in normal and in pathological voices. The periodicity, or lack of periodicity, critically determines the quality of voice. It is typically described in terms of jitter and/or shimmer, period to-period correlation, or by spectral characteristics. Another method for acquisition of physiological data is direct visual inspection of the vocal folds vibration by means of stroboscopy. An admitted limitation of the stroboscopic image is that vocal fold vibration must be relatively periodic to visualize a slow motion representation of the phontonic cycle. In fact, aperiodicities associated with some voice qualities make stroboscopy inappropriate, since any disturbance of the vibration distorts the resulting stroboscopic image.

The kymographic concept introduced by [1], [2] seems to be an optimal solution, since each vibratory cycle is documented in terms of a sequence of several images, which can be acquired directly from a single-line camera [1] or by extraction from high-speed image sequences[3],[4]. Videokymography allows isolation of specific portions of the glottic image (taken at up to 7812 images per second) to be analyzed for closure. Such kymographic images give a good view of the movements of the vocal folds, periodic or nonperiodic, but only for part of the image, i.e., the single line. This study aims at offering an automatic quantitative method to obtain vibration properties of human vocal folds via videokymography, by developing a digital image processing algorithm optimized for the analysis of videokymography (VKG) recordings, such as intensity adjustment, noise removal and glottis identification. The presented method extends previous work [5] and combines an active contour model with a parameter extraction algorithm that can accurately track the vibrational wave in videokymograms and automatically quantify its properties in terms of few parameters, useful for clinicians. Tracking parameters, other than simply measuring their mean value and std, is in fact considered of utmost importance by clinicians, as irregular patterns can be found at their instant of occurrence during phonation and put into relation with images and acoustic signal analysis. Specifically, the amplitude and period ratios between right and left vocal fold, as well as the ratio between opening and closing phase are considered [6]. When required, more parameters could be added, on analogy to [7]. Examples are given concerning pathological subjects, that show the robustness of the contour detection algorithm.

II. MATERIALS AND METHODS

Videokymography (VKG) is based on a special camera, which can operate in two different modes: standard and high-speed. In the standard mode, the camera provides standard images displaying the whole vocal folds at standard video frame rate (30/25 frames/s, with 720x486/768x576 pixels of resolution). In the high-speed mode, the video camera delivers images from only a single line selected from the whole image, at the speed of approximately 7875/7812.5 line-images/s and 720x1/768x1 pixels resolution. The resulting high-speed image, called “videokymogram”, displays the vibratory pattern of the selected part of the vocal folds. Kymographic recording is divided into video frames, i.e. segments of approx. 15/18 ms duration. Images are not in colour, and continuous high-intensity light is desirable.
The vibratory pattern displayed in kymographic images is dependent on the measuring position. There are two factors which influence the resulting image:

1- position along the glottal axis
2- angle with respect to the glottal axis.

In normal cases, the position in the middle of the vocal folds is usually considered to provide the representative vibratory pattern of the whole vocal folds. In case of vocal fold lesions, however, the vibration characteristics generally differ along the glottal axis.

The angle of the measuring line is, as a standard, adjusted to be perpendicular to the glottal axis. When using VKG, the measuring position is adjusted prior to the examination.

Usually, there is only limited time available for the examination. Therefore, phonation at comfortable pitch and loudness is mostly targeted for kymographic imaging.

Despite its usefulness, in our knowledge, until now no quantitative analysis of VKG images is commercially available, and only few work has been made towards its fulfillment [6], [7]. At present, physicians perform only qualitative evaluation of VKG parameters, basically by visual inspection of subsequent frames, or by manual measures from printed images. Such analysis prevents from reliable comparison among wide sets of data, and hence from finding and defining standard reference values for classification and assessment of treatment effectiveness. Based on such requirements, this work aims at providing first results that would allow filling this gap.

Parameter extraction is obtained here by means of two subsequent stages: image analysis, for vocal folds contours detection, followed by signal analysis, for parameter evaluation from data sets representing vocal folds edges. Specifically, and with reference to Fig.2, the parameters to be measured and tracked are:

- $R_{amp}$, the ratio between the right and left vocal fold amplitudes, related to possible asymmetries between the two folds;
- $R_{per}$, the ratio between the right and the left vibratory periods, inversely related to possible frequency variations due to pathology;
- $R_{oc}$, the ratio between opening and closing phase (Open and Closed, respectively), basically related to glottal insufficiency.

For healthy voices, such parameters should be equal or close to one, and almost constant during all phonatory cycles. Any asymmetry due to pathology can thus be quantified by evaluating and tracking the above-mentioned parameters.

### A. First stage: Edge detection

The correct detection of the vocal folds contour is carried out in a two-step process, the first one aiming at finding an initial contour to which active snakes are applied in the second step.

![Figure 2: Main parameters for objective videokymographic image analysis (from [1]).](image)

#### A.1 First step: initial contour

The following routine, that handles some basic settings, is executed before the snake. It normalises grey levels, initialises contour, takes notice of black lines, to be disregarded by the algorithm, thus finding the significant rows in the image. Then, it sets to 0 the level of the image outside the right and left edges, defined by the user. This allows avoiding noisy fluctuations of the grey levels not overlapped with the vocal folds. The routine scans the significant rows and, for each of them, determines the largest interval of pixels with a grey level lower than a pre-specified threshold. A first contour is thus obtained, by storing the interval coordinates for each line in two separate arrays.

#### A.2 Second step: snake active contour

Snakes are planar deformable contours that are useful in several image analysis tasks. They are often used to approximate the locations and shapes of object boundaries on the basis of the reasonable assumption that boundaries are piecewise continuous or smooth [8]. Representing the position of a *snake* parametrically by $v(s) = (x(s), y(s))$ with $s$ in $[0, 1]$, its energy can be written as:

$$E_{snake} = \int E_{int}[v(s)] ds + \int E_{ext}[v(s)] ds$$

where $E_{int}$ represents the internal energy of the snake, due to bending and it is associated with *a priori* constraints, $E_{ext}$ is an external potential energy which depends on the image and accounts for *a posteriori* information. The final shape of the contour corresponds to the minimum of this energy.

In the original technique [9] the internal energy is defined as:

$$E_{int}[v(s)] = \frac{1}{2} a(s) \left( \frac{dv}{ds} \right)^2 + b(s) \left( \frac{\partial^2 v}{\partial s^2} \right)^2$$

This energy is composed of a first order term controlled by $a(s)$ (*tension* of the contour) and a second order term controlled by $b(s)$ (*rigidity* of the contour).
The external energy couples the snake to the image. It is defined as a scalar potential function whose local minima coincide with intensity extremes, edges, and other image features of interest:

$$E_{\text{ext}}[v(s)] = -c \nabla G_s \ast I(x, y)$$

(3)

where $I(x, y)$ is the image intensity, $G_s$ is a Gaussian of standard deviation $s$, $\nabla$ is the gradient operator and $c$ a weight associated with image energies [8], [9], [10].

As concerns the energy minimization, the original model employs the variational calculus to iteratively minimize the energy. There may be a number of problems associated with this approach such as algorithm initialization, existence of local minima, and selection of model parameters. Among existing methods, the greedy algorithm [10] exhibits a low computational cost, provided the initial position of the snake is relatively close to the desired contour. In our application, the technique described in step 1 provides a fairly good approximation of the real contour, therefore allowing us to utilize the active contour and the snake algorithm to perform a fine tuning of the contour on the image. Each vocal fold has been modelled as an independent snake, having its extreme point constrained to belong to the first and the last scan line of the image, respectively. Notice that, differently from [7], the snake is applied on the whole contour and not on sequential rows. This makes the search particularly efficient and robust.

B Second stage: parameter extraction and tracking

In this stage, data consist of (time, edge value) pairs, for each fold, obtained as described in the previous steps. As already said, three clinically relevant parameters are extracted from data.

$R_{\text{amp}} =$ ratio between the average amplitude of the left vocal fold and that of the right vocal fold. The amplitude is defined as the distance between each point and a fictitious closed-fold point, chosen to be halfway between the minimum values of the folds.

$R_{\text{per}} =$ ratio between the right vocal fold period and the left vocal fold period. The period is defined as the mean value among all periods in the frame. Each period is obtained by evaluating the distance between consecutive maximum edge values, determined relatively to the closed-fold point axis.

$R_{\text{oc}} =$ ratio between the opening and closing phase of the folds, determined searching consistent, non noise-generated interruptions of the time coordinate.

Following [7] as well as future clinicians suggestions, more parameters could be easily added and extracted.

III. EXPERIMENTAL RESULTS

Algorithms were applied to a set of VKG recordings (Kay Elemetrics VKG Camera, Model 8900©), ORL Dept., Ospedale Maggiore, Milano, Italy, belonging to both normal and pathological patients. Specifically, 11 patients (6 male, 5 female, age 24-81, mean 52 years) were analysed, affected by: leucoplachia, granuloma, polyp, dysphonia, and possible vocal fold paralysis. The work was carried out under C++ development environment.

Each image has been processed and visually inspected to qualitatively assess the contour identification. Both the results of the first step (before the application of the snake algorithm), and of the second one, as optimized through the active contour, are considered. Notice that the first step works reasonably well in about 80% of the test cases, although there is a considerable amount of noise which reduces the reliability of the measured parameters. In the remaining 20% of cases, the images present few dark zones, which cause the detection of artefacts appearing as anomalous contours. The application of the active contour method, however, greatly reduces the presence of both noise and artefacts, achieving an accurate contour detection.

Fig.3 shows the results obtained with the first step (Fig.3a) and the second step (Fig.3b) for one patient: male, 75 years old, affected by leucoplachia on the left vocal fold. The figure is relative to a single VKG frame out of about 450, for about 2 min. total duration of the whole visit, which comprises laryngoscopic, VKG and simultaneous audio recording of sustained /a/. Notice that the first step gives almost irregular initial contours, while the second one smoothes the lines and successfully removes outliers.

![Figure 3 – Edge contour deetection. (a) first step: preliminary contour; (b) second step: final contour.](image)

Fig.4 shows the tacking of the three parameters $R_{\text{oc}}$, $R_{\text{amp}}$, $R_{\text{per}}$ on a set of about 90 subsequent frames. Notice that, due to the length of the exam, the emission slightly changed with time, ranging from /a/ to /ae/ and /ao/ ($F_0$ varied in the range 140Hz-230Hz).
This fact, in conjunction with pathology, and the difficulty of the operator to keep the endoscope fixed on the same line through the whole analysis, caused possible changes in the VKG parameters, as pointed out in fig.4 as far as Roc is concerned, which shows a mean value Roc-mean=1.96, with 0.8 std. Instead, Ramp and Rper are quite stable: Ramp-mean=0.99, std=0.2, Rper-mean=1, std=0.02, according to the pathology under study, that does not causes strong irregularities in the vocal folds oscillation.

Similar results were obtained with the other recordings. Both visual inspection of contours and objective parameters tracking have provided clinicians with useful details and information, also in cases not clearly distinguishable with stroboscopy alone.

IV. FINAL REMARKS

Kymographic imaging provides valuable information on the dynamic behaviour of the laryngeal tissues that is not so clearly distinct in the classical stroboscopic viewing, especially in case of early or non-specific lesions, irregular closure patterns, vocal fold weakness, paralysis, that may also allow for the differentiation of weakness due to overuse, aging, paresis, or early stages of neurological conditions. The information can be used in basic research, vocal fold modelling, as well as in clinical practice, as, for instance, in evaluating the results achieved by phonosurgery.

Hence, the need for automatic evaluation and tracking of objective parameters, extracted from VKG images, is of great concern. This paper aims at providing a basic set of such parameters, by means of active contour techniques for edge detection. Specific adjustments were made in a first step, in order to deal with high varying and noisy images as those under study.

Current research focuses on refinements of the proposed technique, as well as on the estimation of a wider set of parameters. A user-friendly interface is also under construction, with the aim of making the analysis fully automatic and allowing easily storing and retrieving patient’s data.

V. REFERENCES