



Objective Parameters from Videokymographic Images: A User-Friendly Interface

C. Manfredi¹, L. Bocchi¹, G. Cantarella², G. Peretti³, G. Guidi¹, V. Mezzatesta¹

¹Department of Electronics and Telecommunications, Univ. of Firenze, Firenze, Italy

²Otolaryngology Department, Univ. of Milano, Ospedale Maggiore Policlinico Mangiagalli e Regina Elena, Fondazione IRCCS, Milano, Italy

³Otolaryngology Clinic, Univ. of Brescia, Spedali Civili di Brescia, Brescia, Italy

manfredi@det.unifi.it, leo@asp.det.unifi.it, giovanna.cantarella@policlinico.mi.it, g.peretti@tin.it

Abstract

Videolaryngostroboscopy (VLS) is a first choice examination for diagnosis of several laryngeal pathologies. However, in case of strong a-periodicity of the vocal sound, it becomes ineffective in describing subsequent phases of the vocal cycle. To overcome such limitation, a new technique, named videokymography (VKG) has been developed. VKG delivers images from a single line selected from the whole VLS image, at the speed of approximately 8000 line-images/s. However, despite its usefulness, until now no quantitative analysis of VKG images is commercially available. This paper presents a new tool for measuring and tracking quantitative parameters from VKG images. It performs evaluation of left-to-right period, amplitude and phase ratios, as well as of phase symmetry index. Robust techniques for edge detection have been implemented, to reduce both noise and artefacts. The new tool has been provided with a user-friendly interface for managing patients' data and image analysis, according to a set of parameters that can be easily adjusted by the user. VKG images from one non-dysphonic and seven dysphonic subjects were analysed, providing objective parameters useful in diagnosis support.

Index Terms: voice quality assessment, objective voice parameters, VKG, image analysis

1. Introduction

Along with the refinement of diagnostic and therapeutic techniques in phoniatrics and phonosurgery, the need for objective methods of vocal cycle evaluation, along with its pathological or post-treatment changes, has gained more and more relevance. Videolaryngoscopy (VLS), while being nowadays considered as a first-choice test for the diagnosis of most laryngeal functional pathologies, bears some intrinsic limitations that limit its clinical application. In fact, the stroboscopic image of the vocal fold vibration is basically an optical illusion made by the human eye, coming from the virtual reconstruction of adjacent phases of different vocal cycles, kept by the stroboscopic flashes in subsequent time instants. Hence, in case of strong intensity deficiency or a-periodicity of the emitted vocal sound, the lightening mechanism, settled at the phonatory frequency, results ineffective in representing subsequent phases of the vocal cycle, losing details concerning the mucosal wave characteristics. Moreover, inter-subject variability, related to the clinician's skillfulness, always influences results. To

overcome such limitations, a new high-speed device has been developed, named Videokymography (VKG) [1], [2]. 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, depending on video coding standards, 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. Kymographic recording is divided into video frames, i.e. segments of approx. 15/18 ms duration. Line selection is obtained by moving from the VLS image to the high-speed VKG one by means of a footswitch at the time in which the upper edge of the video image corresponds to the selected point. (Fig. 1).

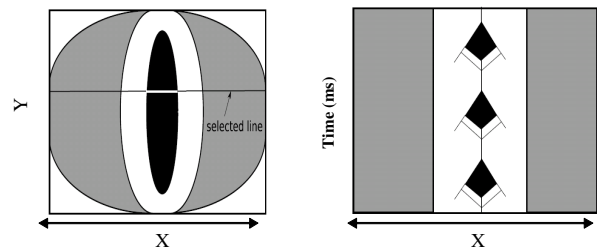


Figure 1 – From VLS (left) to VKG (right) images

At present, the evaluation of VKG parameters by clinicians is basically qualitative. It relies on subjective perceiving of vocal folds symmetry, presence or absence of constant glottis deficiency in any cycle phase, amplitude of vibration of one fold with respect to the other. Such evaluation, while of utmost importance for a single subject, makes it difficult any comparison among different subjects, different pathological conditions and multicentric databases.

2. Materials and Methods

The discussion made above give the reason for a software tool devoted to the automatic evaluation of such measures. First results can be found in [3], [4]. The Kay Elemetrics VKG Camera, Model 8900® used in our study [5], provides frames (256 grey levels) made up of 144 lines, corresponding to a time duration of about 18.4 ms, with a size of 320 x 240 pixels each. Half of the lines are black, and must be disregarded

before elaboration. Moreover, a black strip is present on each side of the images. It is therefore necessary to exclude such areas, and restrict the elaboration to the meaningful part of the image, made up by 120 lines, whose width varies between 35% and 60% of the original one. Hence, a new digital image processing algorithm has been developed by the authors, optimized for the analysis of VKG recordings.

2.1 Parameters

The following parameters have been implemented:

R_{amp} = ratio between the amplitude of one vocal fold vibration and the opposite one. R_{amp} is an index of asymmetry (Fig.2a).

R_{per} = ratio between the vibration wavelength of vocal folds. It is inversely related to frequency variation due to pathological structural alterations of the vocal fold. (Fig. 2b).

R_{oc} = ratio between the measure of the length of the opening and closing phase within a single glottis cycle, related to the degree of glottis incompetence (fig.2c).

PSI, Phase Symmetry Index = ratio between the difference of the two vocal folds phases (t_1 , t_2) and the period T (fig.2d).

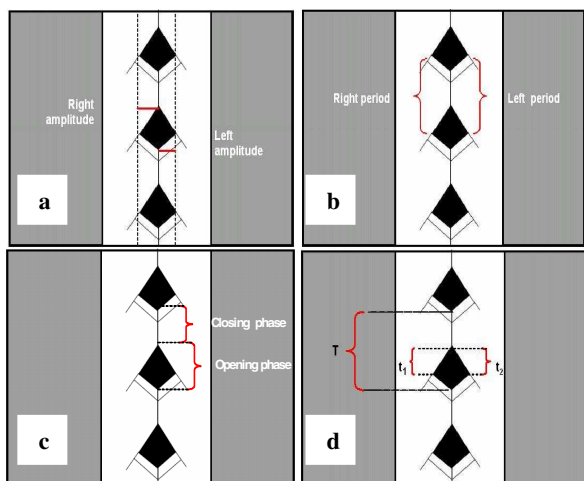


Figure 2 – (a) R_{amp} ; (b) R_{per} ; (c) R_{oc} ; (d) PSI

Moreover, in case of glottis incompetence, the mean distance between folds, M_{op} , has been evaluated. The parameter extraction procedure is based on a robust contour detection algorithm, made up of two steps. The first one, which handles some basic settings, defines an initial contour of the glottis area opening, to which a refining iterative procedure based on snake “surface” active contour, is applied [6]-[7]. Snakes are deformable planar contours, depending on energy deformation functions that increase as the snake deformation increases. Snake energy, E_{snake} , can be viewed as the sum of the internal energy of the snake, associated with a priori constraints, and 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 E_{snake} . Two internal energies are defined here, continuity (E_{cont}) and bending (E_{bend}). E_{cont} contributes to keep the distance between adjacent points as regular as possible, opposing to too sharp displacements. E_{bend} approximates the bending radius of the contour section and increases as the bending radius increases. The external energy (E_{ext}) takes into account the grey level intensity of the points inside the surface. When the contour approaches the edge to

be identified, E_{ext} opposes to the expanding trend of the contour, and forces it to bend, and to adapt to the sharp variation of light intensity.

2.2 User-Friendly Interface

To facilitate parameter extraction from non-expert users, a simple interface has been developed, made up by two windows: one for visualization and the other for control purposes. The interface loads the first image of the available set and produces a first active contour. The visualization window shows in subsequent images the active contour as modified in each iteration, and the tracking of the parameters R_{per} , R_{amp} , R_{oc} , and PSI. Examples are given in Sect.3.

The second window shows control buttons for both patient and snake options management (Fig.3). The first section is relative to the choice of items to be displayed. “Image” button allows displaying the VKG frame; “Snake” button superimposes on the selected frame the initialization snake. “Plot” button displays the plots of the objective parameters, showing their time evolution (dotted or solid line).

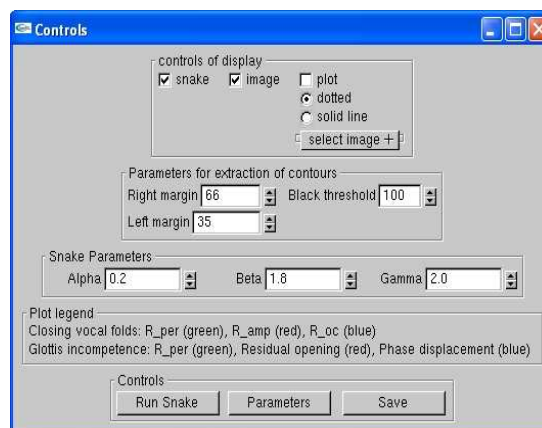


Figure 3 – User interface for VKG image analysis

Notice that VKG frames are affected by “noise”, i.e. darker or brighter spots comparable in intensity to the glottal opening that should be carefully disregarded as they could false the edge detection. In most cases, narrowing the image enhances the results. In a previous version of the tool [8], narrowing was made manually by the user, acting on corresponding buttons for the right and the left margin setting (see Fig.3). In this new version, narrowing is performed automatically by the program. It searches for vocal folds margins, finds the black threshold and sets all pixels above it to white colour. However, manual adjustment is still allowed, in case of extremely noisy images. Figure 4 compares the original frame (above) and the same one after automatic margin selection (below) (patient n. 7). Initial contour is shown, pointing out enhanced results with the automatic setting. The same result can be obtained with manual setting, acting on appropriate buttons (right margin, in this example), but this procedure can be time consuming, as it is difficult to a-priori know the exact setting on each side.

Moreover, if required, three parameters: α , β , γ can be manually adjusted in the range 0-3. Specifically, high values of α , β , γ should be selected for: non-continuous edges; irregular edges, far from the circular shape; faded images, respectively. At the end of the session, the user can ask for mean and standard deviation (STD) of parameters to be displayed and saved in a text document for future use.

The software tool has been developed in C++, both under UNIX and Windows (Microsoft Visual C++ 6.0) environment platforms. Thus, its portability characteristics allow running the program on most common existing platforms.

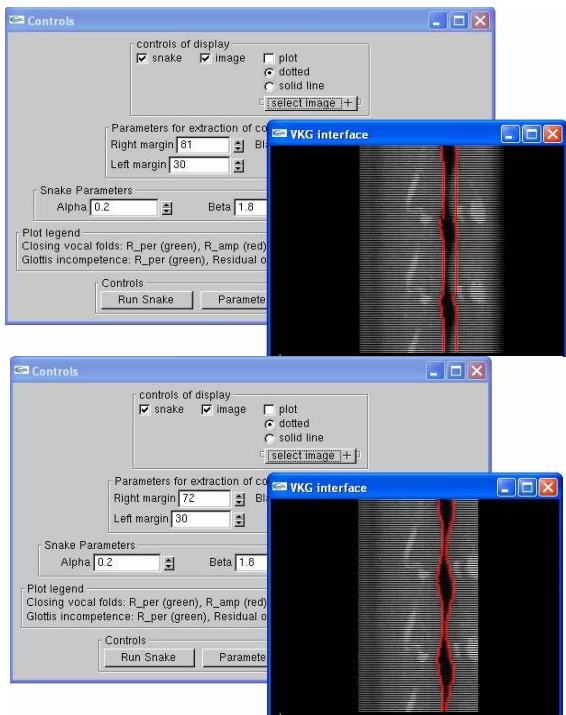


Figure 4 – Upper: without setting of margins; Lower: automatic setting of margins.

3. Experimental results

The proposed approach was applied to a set of VKG recordings from one normal subject and a set of dysphonic patients. Specifically, 7 patients (5 male, 2 female, age 30-81, mean 62 years) were analysed, affected by different pathologies: leukoplachia, granuloma, polyp, functional dysphonia, and hypo motility of the left vocal fold. All subjects were asked to emit the Italian sustained vowel /a/ at comfortable pitch. Typically, from each VKG video recording, a set of about 700-3000 frames is obtained, depending on the recording time. However, only the 30% of such images are VKG ones, the other being VLS images. Furthermore, commonly sets of no more than 15-20 subsequent frames are available, due to recording problems during the clinical visit.

The next three figures show results obtained for three different subjects: healthy, leukoplachia, functional dysphonia. Each figure shows the results relative to one frame out of the available set for each subject.

Figure 5 shows results for the healthy case (male, 45 years old): both the VKG image and parameters tracking confirm regular vibration of vocal folds, with R_amp (red line), R_per (green line) and R_oc (blue line) close to unity and PSI (yellow line) close to 0 on the whole emission. A set of 28 frames was analysed for this case. See Table 1, subject n.8.

The second case, depicted in Fig. 6, comes from a set of 18 subsequent frames, relative to a male patient affected by leukoplachia on the left vocal fold. R_oc (blue line) is highly varying and larger than 1 over the whole phonation, pointing out irregular closing/opening phases. R_amp (red line) is clearly unstable and far from regularity condition ($R_amp < 1$),

pointing out this pathology. R_per (green line) is close to 1, confirming period regularity. PSI index (yellow line) is slightly larger than 0, as expected. Table 1, subject n.3.

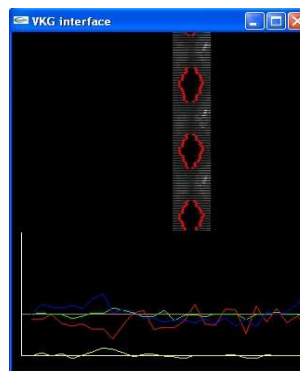


Figure 5 – Healthy subject: VKG image with contours and parameters tracking (1st frame)

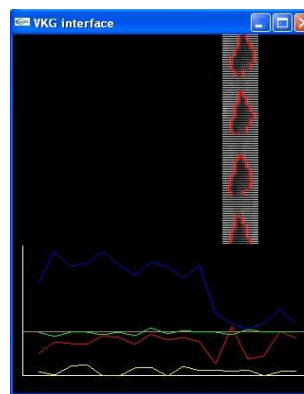


Figure 6 – Leukoplachia of the left vocal fold: VKG image with contours and parameters tracking (1st frame)

Finally, a subject with functional dysphonia is considered. Fig. 7 shows one frame coming from a set of 20 subsequent frames. Irregular vocal folds vibration is depicted and quantified especially as far as R_amp is concerned (red line), which strongly oscillates around and above 1. R_per (green line) and R_oc (blue line) are close to 1 and PSI (yellow line) is close to 0, as for healthy subjects. Table 1, patient n.6.

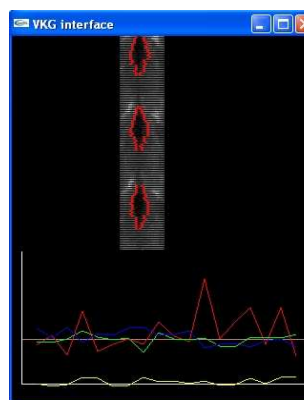


Figure 7 – Dysphonic subject: VKG image with contours and parameters tracking (18th frame)

Table 1 summarises the results obtained for all cases. The mean value and STD (in parentheses) for all parameters and all patients is reported. About 20 subsequent frames were considered for each patient.

Table 1 – Mean and STD values for VKG parameters

n.	R_oc	R_per	R_amp	PSI
1.	2,08 (0,62)	0,99 (0,05)	0,77 (0,2)	0,12 (0,1)
2.	0,99 (0,14)	0,99 (0,07)	1,37 (0,18)	0,26 (0,09)
3.	1,06 (0,07)	1,024 (0,126)	1,97 (0,5)	-0,03 (0,1)
4.	0,93 (0,05)	0,99 (0,02)	1,19 (0,15)	0,08 (0,04)
5.	1,11 (0,18)	0,93 (0,23)	0,47 (0,16)	-0,04 (0,2)
6.	1,06 (0,15)	1,03 (0,09)	1,18 (0,44)	0,06 (0,08)
7.	-	1,17 (0,4)	M_op 11,4 (1,23)	
8.	1,05 (0,22)	1,01 (0,06)	0,86 (0,2)	0,03 (0,07)

The first three cases concern elder male patients affected by leukoplachia. All of them show abnormal R_amp values, well below or above the normative value. Patient n. 1 also shows very high opening phase as related to closing phase. Patient n.2 shows high PSI index, pointing out non-symmetrical vocal folds vibration. For patient n.3, irregular amplitude ratio is evidenced.

Patient n.4 (adult female, nodule exeresis) has quite good parameters, showing good recovering after surgery.

Patient n.5 (elder female, right vocal fold paralysis) shows highly irregular vocal folds vibration, as confirmed by low R_amp and R_per.

Patient n.6 (adult male, functional dysphonia), shows quite regular vibration of the vocal folds, all parameters and relative std being close to the range for healthy subjects. However, large STD is measured for R_amp and R_oc, pointing out irregular vibration.

Case 7 (adult female, glottis incompetence due to polyp on the right vocal fold). In this case, as previously said, only R_per out of the four parameters is still meaningful and is evaluated. The parameter considered here is the mean opening distance, M_op, that reaches very high values (in pixels), as expected.

Finally, case 8 (adult male, healthy) confirms that, in case of regular vocal fold vibration, during whole phonation the first three parameters are close to unity while PSI varies around 0, with reasonably low std across these values.

Notice that, in most of the analysed cases, images are strongly corrupted by noise, thus influencing results. With better quality images, such those obtainable with an advanced VKG device [9], more reliable results would be achieved.

4. Discussion

A digital image processing algorithm has been developed, optimized for the analysis of VKG recordings that require intensity adjustment and noise removal for exact glottis identification. Robust techniques for edge detection have been implemented, greatly reducing the presence of both noise and artefacts. A set of parameters, easily adjustable by the user, allows refining the image analysis procedure in case of strong

noise. The new tool, designed according to clinician's suggestions, has been provided with a user-friendly interface that allows storage and retrieval of patients' data, and real-time analysis of VKG images. Plots and pictures allow the clinician objective evaluation of results and give reliable and reproducible measures. When applied to a large set of data, the new tool would allow defining reference values for normal and pathological cases, providing a valid support to diagnosis and surgical effectiveness evaluation.

Future work would concern adding further parameters, as well as further simplifying the interface and refining the contour detection algorithm. Specifically, work is in progress as far as initial contour detection, starting interface, menu options and graphical display of parameters tracking are concerned.

As already pointed out, the new VKG device as described in [9] will allow simplifying the contour detection procedure, as well as the parameters extraction techniques, thanks to its enhanced characteristics with respect to the first one.

5. Conclusions

A new tool for measuring and tracking quantitative parameters from VKG images has been developed and is proposed here. With this tool, VKG images can be analysed both in case of complete and incomplete glottis closure. It performs evaluation of left-to-right period, amplitude and phase ratios, phase symmetry index, and mean distance between folds, also providing images and plots for inspection of results.

Based on these first results, and on [3], [9], and clinicians suggestions, other parameters are under study, as well as methodological and interface refinements.

6. References

1. Schutte HK, Svec JG, Sram F. "First results of clinical application of videokymography", *Laryngoscope*, 108:1206-1210, 1998.
2. Svec JG, Schutte HK." Videokymography: High-speed line scanning of vocal fold vibration", *J Voice*, 10:201-205, 1996.
3. Qiu Q., Schutte HK, Gu L., Yu O., "An Automatic Method to Quantify the Vibration Properties of Human Vocal Folds via Videokymography", *Folia Phoniatrica et Logopaedica*, 55:128-136, 2003.
4. Manfredi C., Bocchi L., Bianchi S., Migali N., Cantarella G., "Objective vocal fold vibration assessment from videokymographic images", *Biomedical Signal Processing and Control*, 1: 129-136, 2006.
5. Kay Elemetrics, 2002. High-Speed Video (HSV) Model 9700: Instruction Manual, Issue A. Kay Elemetrics Corp., Lincoln Park, NJ.
6. Kass, M., Witkin, A., Terzopoulos, D., "Snakes: Active contour models", *Int. J. Computer Vision*, 1: 321-331, 1988.
7. McInerney, T., Terzopoulos, D., "Deformable models in medical image analysis: a survey", *Medical Image Analysis*, 1: 91-108, 1996.
8. C. Manfredi, L. Bocchi, G.Cantarella, G.Peretti, "A User-Friendly Interface for Objective Parameters Extraction from Videokymographic Images", 128th ALA Meeting, San Diego, CA, April 26-27, 2007.
9. Q. Qiu, H. K. Schutte, "A New Generation Videokymography for Routine Clinical Vocal Fold Examination", *Laryngoscope*, 116: 1824-1828, 2006.