IMAGE ANALYSIS OF VIDEOKYMOGRAPHIC DATA

*Adam Novozámský[a], Jiři Sedlář[a], Aleš Zita[a], Filip Šroubek[a], Jan Flusser[a], Jan G. Švec[b], Jitka Vydrová[c], Barbara Zitová[a] **

[a]Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic, Prague, Czech Republic
{novozamsky,sedlar,zita,zitova,sroubekf, flusser}@utia.cas.cz
[b]Voice Research Lab, Department of Biophysics
Faculty of Sciences, Palacký University, Olomouc, Czech Republic
svecjang@gmail.com
Voice Centre Prague, Medical Healthcom, Ltd., Czech Republic
vydrova@medico.cz

ABSTRACT

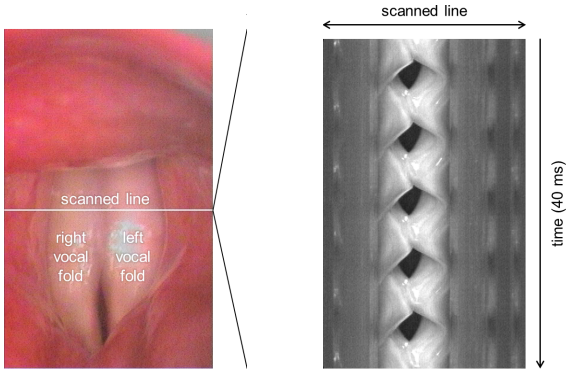
Videokymography (VKG) is a high-speed medical imaging technique used in laryngology and phoniatrics for examination of vocal fold vibrations, it offers important characteristics for diagnosis and treatment of voice disorders. VKG repeatedly scans only a single line from the scene and captures movements of vocal folds in this region of interest. This paper proposes methods for computer assisted evaluation of diagnostically important vibration features, related to movements of vocal folds and their surroundings. They are derived from existing as well as newly developed methods of digital image processing, mainly based on data segmentation and morphological operations. Performance of the developed methods is compared to expert manual assessments and it proves to be comparable with clinicians conclusions.

Index Terms— videokymography, medical imaging, data segmentation

1. INTRODUCTION

Digital image processing methods form an integral part of medical data analysis and evaluation. Our paper addresses an analysis of videokymographic data (videokymograms), which are collected using videokymography (VKG) - a high-speed imaging technique convenient for observation of vocal fold vibrations. VKG is used in laryngology and phoniatrics for diagnosis of vibration parameters of vocal folds. Our aim is to complement visual evaluation of videokymograms, which can be tedious and clinician-dependent. We proposed automatic software tools for VKG preprocessing and detection of important features.

*The work was supported by the Technology Agency of the Czech Republic under the project no TA04010877 and by GACR agency project GA13-29225S.



(a) standard mode (b) VKG mode

Fig. 1. Two modes of videokymographic camera data acquisition: (a) standard and (b) videokymographic. The videokymogram (b) is composed of successively acquired scanned lines at the location indicated in (a).

There are several techniques of capturing human vocal fold vibrations for assessment of their functionality. The most commonly used are the videostroboscopy, high-speed videoendoscopy and the latest, videokymography (VKG). The VKG is an original Czech-Dutch method, developed in 1994 in Groningen (NL) as an alternative to high-speed video recording [1]. The system consists of specially adapted CCD camera, which operates in two modes - standard (50 fps - interlaced, Figure 1(a)) and in high speed (currently 7200 lines per second), when the system records images of a single horizontal line of the selected camera row and stacks them below each other (Figure 1 (b)). The method allows efficient recording of vibration patterns of vocal folds. An application of digital image processing methods for analysis of vocal fold

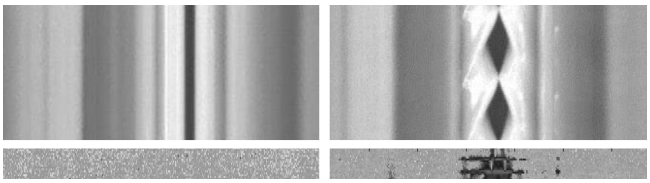


Fig. 2. The VKG data without (top left) and with (top right) an activity. Respective representations in the column-wise Fourier spectral domain close-ups are shown in the bottom line.

data have been attracting an attention for some time. Most of them have been oriented on high speed videoendoscopy [2], while VKG attracted lesser interest [3], eventhough this modality offers many benefits due to its efficient data and vocal fold characteristics representation [4]. Our approach broaden proposed methodology for VKG data and make them more robust to VKG data variability. To facilitate the evaluation of VKG data we focus ourselves on three phases of VKG analysis: (I) data preprocessing, (II) vocal folds characteristics extraction and (III) auxilliary features extraction.

2. VKG PREPROCESSING

Typical videokymogram is a gray-scale image capturing several openings and closings of vocal folds (see Figure 1(b)). The data can be noisy, with low contrast and with reflections caused by present mucus. All these factors can negatively influence the performance of further software analysis tools. Moreover, due to the manual examination procedure of the data acquisition, the laryngoscope can be randomly shifted from the optimal position. An important factor which influences the data examination is the patient's discontinued phonation, when the vibrations are missing in VKG data at all.

To ensure the best possible outcome of the automatic analysis of VKG data we apply data preprocessing steps such as median denoising, locally-adaptive contrast enhancement and, if needed, mucus reflection removal using adaptive thresholding followed by diffusion inpainting. The effect of unexpected patient movements and his discontinued phonation is handled by selection of meaningful data subsequences only. This method was developed to select only these parts of VKG recording where the vocal folds are approximately in the center of the VKG image and are active. The Fourier transform of fixed width columns is analyzed and these VKGs with the spectral response under the given threshold are omitted from further processing (see Figure 2).

The attention is paid also to VKG data taken in the standard mode. They can be blurred due to the wrong camera focus and movements of the patient. We proposed to apply multichannel blind deconvolution method [6] to improve the sharpness of the data, even that they are only for visual inspection and are not used in the further automatic analysis.

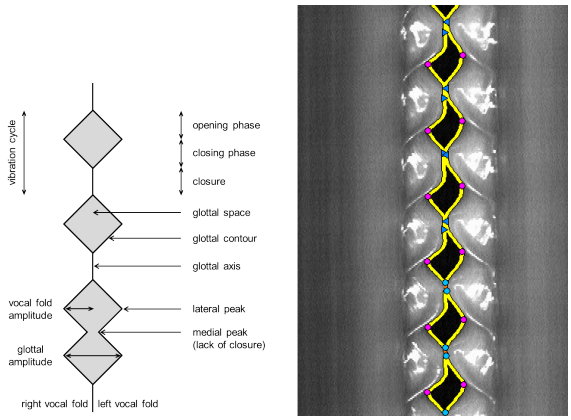


Fig. 3. (left) - Vibration features in videokymograms. (right) - The shape of glottal space and detected base glottal features – opening and closing points (light blue), medial peaks (dark blue), and lateral peaks (magenta).

feature	notation
opening points	O_i
closing points	C_i
lateral peaks	A_i^R, A_i^L
medial peaks	M_i^R, M_i^L

Table 2. Base glottal features in videokymograms (see Figure 3); upper indices R and L denote the right and left vocal folds, respectively, and lower index i denotes the number of the vibration cycle in the videokymogram.

3. VOCAL FOLD CHARACTERISTICS

Proposed methods for analysis of VKG data are based on vocal folds / glottal contours and detected base features (see Figure 3) which are key elements for computation of established vibration parameters [5]. The primary step for all further VKG evaluation is the detection of glottal contour (Figure 3 - (right), yellow curves), which is realized by means of an thresholding segmentation with an optimized threshold estimated by normalized graph cuts [7, 8]. This approach maximizes dissimilarity between two parts of the scene according to both spatial and gray-level relations of their pixels. The detected glottal space is then used for estimation of elementary glottal features - opening and closing points, lateral and medial peaks, vibration cycles and their opening, closing, open and closed phases, and glottal and vocal fold amplitudes (see Figure 3). The opening and closing points, and the lateral and medial peaks (see Table 2) are the base features and are used for derivation of the other aforementioned features called derived glottal features. Their derivation from the base features is listed in Table 1. They all are used for computation of es-

feature	notation and definition
generalized opening points	$\tilde{O}_i^j = \{O_i, M_i^j\}$
generalized closing points	$\tilde{C}_i^j = \{C_i, M_i^j\}$
opening phase duration	$t_i^{oj} = A_i^j(y) - \tilde{O}_i^j(y)$
closing phase duration	$t_i^{cj} = \tilde{C}_i^j(y) - A_i^j(y)$
open phase duration	$T_i^{oj} = t_i^{oj} + t_i^{cj} = \tilde{C}_i^j(y) - \tilde{O}_i^j(y)$
closed phase duration	$T_i^{cj} = \tilde{O}_{i+1}^j(y) - \tilde{C}_i^j(y)$
vibration cycle duration	$T_i^j = T_i^{oj} + T_i^{cj} = t_i^{oj} + t_i^{cj} + T_i^{cj} = \tilde{O}_{i+1}^j(y) - \tilde{O}_i^j(y)$
vocal fold amplitudes	$a_i^j = \max(\lvert A_i^j(x) - \tilde{O}_i^j(x)\rvert, \lvert A_i^j(x) - \tilde{C}_i^j(x)\rvert)$
glottal amplitudes	$a_i = A_i^L(x) - A_i^R(x)$

Table 1. Derived glottal features in videokymograms [5]; upper index $j \in \{R, L\}$ denotes the right and left vocal folds, respectively, and lower index i denotes the number of the vibration cycle in the videokymogram.

tablished vocal fold vibration parameters [5]. Their detailed definition and discussion can be found in [9].

The proposed base and derived glottal features and set of vocal fold vibration parameters [9, 5] were evaluated on the testing dataset of 50 videokymograms and compared to manual evaluations [10], done by evaluators with different level of experience and resulting in 18 assessment sets in total (18 $\times$ 50 videokymograms). The comparison was realized in an automatic–visual and visual–visual manner. The method introduces the following notation. Let P denotes the set of both automatically and visually evaluated parameters, n the number of evaluated videokymograms, and m the number of visual evaluations. Let $E_A(p;i)(p \in P; i = 1;...;n)$ denotes the result of automatic evaluation of parameter p in videokymogram i, and $E_V(p;i;j)(p \in P; i = 1;...;n; j = 1;...;m)$ the result of j^{th} visual evaluation of parameter p in videokymogram i. Let $V^+(p;i)$ denote set of indices of visual evaluations of parameter p in videokymogram i with defined result (non-NA)

$$V^+(p;i) = \{j \mid j \in \{1...m\} \wedge E_V(p;i;j) > 0\}$$

Then the consensus result is defined as NA if and only if the result of at least half of corresponding visual evaluations was NA; otherwise, the definition estimates it by the most frequent non-NA result. For each parameter $p \in P$ and videokymogram $i \in \{1, ..., n\}$ the proposed method compares the consensus result of visual evaluations $E_V(p;i)$ with the result of automatic evaluation $E_A(p;i)$ (automatic–visual match) and with the results of visual evaluations $E_V(p;i;j), (j = 1, ..., n)$ (visual–visual match).

The automatic–visual match compared the automatically estimated parameter categories with the category most frequently selected by the visual evaluators whereas the visual–visual match estimated reliability of the visual evaluations (how often the assessments of the visual evaluators were in agreement). In all cases two evaluations are set to be matching if their respective results fall into the same category or into directly neighboring non-NA categories. The results can be seen in Table 3. Figure 5 illustrates the variability of the

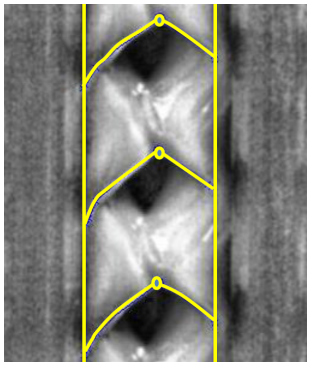


Fig. 4. Detected lateral mucosal waves with the starting points (circles) and their detected extent.

VKG data that the proposed algorithms must to be able to cope.

The experiments showed consistency between automatic and visual evaluations. The similarity in comparative statistics demonstrates that the performance of the automatic evaluation is comparable with visual evaluations and thus the proposed approach in the computer-aided evaluation is found applicable in clinical practice.

4. LATERAL MUCOSAL WAVES EXTRACTION

Besides of the established set of vibratory features [5] the research was focused on the auxiliary characteristics of vocal chords and their vibrations – laterally travelling mucosal waves. Mucosal waves are tissue waves propagating across located on the upper surface of vocal folds. They propagate laterally across the surface until they disappear or reaches the lateral border of the vocal fold. Their presence and extent can indicate how pliable a vocal fold is and can may indicate problems with stiffness of vibrating tissue.

In videokymograms, mucosal waves are demonstrated as diagonal, sometimes slightly bended lines on vocal folds running in the direction of the opening movement. Their detec-

vibration parameter	automatic–visual match	visual–visual match
NumberOfCyclesR	98%	95%
NumberOfCyclesL	98%	96%
VariabilityR	92%	93%
VariabilityL	88%	93%
ClosureDuration	98%	93%
AmplitudeDifferences	100%	88%
FrequencyDifferences	100%	95%
PhaseDifferences	88%	85%
AxisShift	88%	79%
SkewingR	86%	87%
SkewingL	90%	85%

Table 3. Comparison of results of automatic and visual evaluations on a set of 50 videokymograms by the automatic–visual and visual–visual match with tolerance between closely neighboring classes. The similarity in comparative statistics for each parameter indicates that the performance of the automatic evaluation is comparable and often better than visual evaluations.

tion can be complicated by reflections, low contrast and their small extent. To solve the problem we introduced the *iterated masked cross-correlation* method. It is based on the detection of self-similarity of the data around the expected position of the wave, which starts at lateral peaks and runs in the direction of the connector of the opening points and the lateral peaks.

The respective cross-correlation kernel is established, positioned at the beginning of the wave and shaped as a tilted rectangle with its size proportional to the glottal space. The kernel is then iteratively updated as the cross-correlation is processed in the given direction, till any progress is made or the steps are shorten below certain level. In this way an approximate path of a glottal wave is constructed and its main direction is then estimated using Fourier spectral analysis of the detected wave path. In Figure 4 there is an illustration of the VKG data with the detected directions and extent of the lateral mucosal waves (circles are representing starting points of the method).

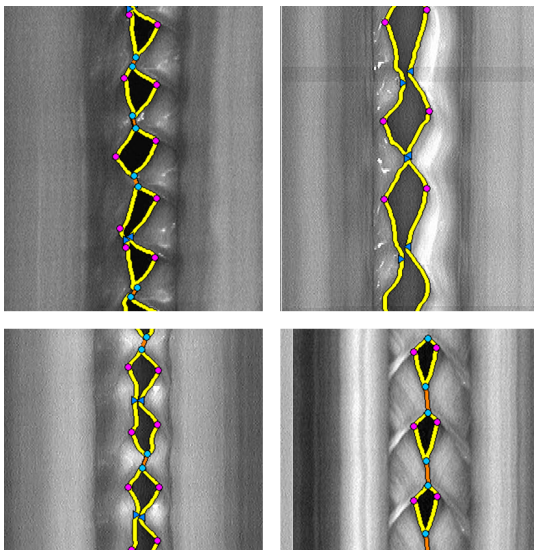


Fig. 5. The variability of the VKG data with detected features.

5. CONCLUSION

New image processing methods for analysis of vocal fold vibrations in videokymograms were developed. The motivation was to create the tools for the computer-aided evaluation of vibratory patterns to be used by laryngologists and phoniatricians in clinical practice for more detailed diagnosis of voice disorders. The introduced algorithms provide data preprocessing, detection of the glottal space by normalized graph cuts thresholding and extraction of glottal vibration features and parameters proposed in [5]. Comparison of the performance of the developed methods with subjective visual evaluations indicated good match making them promising for future use in clinical practice. In addition to the glottal features an original methodology was also developed for detecting laterally travelling mucosal waves. These features which are revealing on the pliability of the vocal folds and their surrounding are detected using *iterated masked cross-correlation* method. In the future the research will be focused on new sets of auxiliary features describing the close surrounding of vocal folds and their interpretability with respect to the diagnosis and treatment of voice disorders.

6. REFERENCES

[1] J. G. Svec and H. K. Schutte, "Videokymography: high-speed line scanning of vocal fold vibration," *J Voice*, vol. 10, no. 2, pp. 201–205, 1996.

[2] J. Lohscheller, U. Eysholdt, H. Toy, and M. Dollinger, "Phonovibrography: Mapping high-speed movies of vocal fold vibrations into 2-d diagrams for visualizing and analyzing the underlying laryngeal dynamics," *Medi-*

cal Imaging, IEEE Transactions on, vol. 27, no. 3, pp. 300–309, March 2008.

[3] Q. Qiu, H. K. Schutte, Gu L., and Q. Yu, "An automatic method to quantify the vibration properties of human vocal folds via videokymography," *Folia Phoniatr Logop*, vol. 55, no. 3, pp. 128–136, 2003.

[4] J. G. Svec and H. K. Schutte, "Kymographic imaging of laryngeal vibrations. current opinion," *Otolaryngology & Head and Neck Surgery*, vol. 20, no. 6, pp. 458–465, 2012.

[5] J. G. Svec, F. Sram, and H. K Schutte, "Videokymography in voice disorders: what to look for?," *The Annals of otology, rhinology, and laryngology*, vol. 116, no. 3, pp. 172180, March 2007.

[6] F. Sroubek and J. Flusser, "Multichannel blind deconvolution of spatially misaligned images," *Image Processing, IEEE Transactions on*, vol. 14, no. 7, pp. 874–883, July 2005.

[7] W. Tao, H. Jin, Y. Zhang, L. Liu, and D. Wang, "Image thresholding using graph cuts," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 38, no. 5, pp. 1181–1195, Sept 2008.

[8] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[9] J. Sedlar, "Image analysis in microscopy and videokymography," *Ph.D. thesis, Charles University, Prague, Czech Republic*, 2012.

[10] V. Hampala, "Visual evaluation of videokymographic features in voice disorders (in czech)," *Master's thesis, Palacky University, Olomouc, Czech Republic*, 2011.