A non-invasive procedure for early stage discrimination of malignant and precancerous vocal fold lesions based on laryngeal dynamics analysis

Unger J¹, Lohscheller J¹, Reiter M², Eder K², Betz Christian², Schuster M²

1 Dept. of Computer Science, Trier University of Applied Sciences, Germany
2 Dept. of Otorhinolaryngology, Head and Neck Surgery, University of Munich, Germany

Running title
Early stage detection of malignant vocal fold lesions

Keywords
Larynx carcinoma, precancerous lesions, high-speed endoscopy, phonovibrogram, wavelet-based analysis

Financial support
The German Research Foundation (DFG) supports this work. Grant no. LO-1413/2-2.

Correspondence to
Dipl. Inf. Jakob Unger
Trier University of Applied Sciences
Medical Informatics
Schneidershof
54293 Trier
Germany
email: unger@hochschule-trier.de
phone: +49 (0)651/8103-575
Disclose of interest

The authors report no declaration of interest. Maria Schuster received remuneration for lectures in 2010 by Richard Wolf GmbH, Tuttlingen, Germany.

Other notes

Word count: 4793
Total number of figures: 7
Total number of tables: 1
Abstract

About two thirds of laryngeal cancers originate at the vocal cords. Early stage detection of malignant vocal fold alterations, including a discrimination of premalignant lesions, represents a major challenge in laryngology as precancerous vocal fold lesions and small carcinomas are difficult to distinguish by means of regular endoscopy only. We report a procedure to discriminate between malignant and precancerous lesions by measuring the characteristics of vocal fold dynamics by means of a computerized analysis of laryngeal high-speed videos. Ten patients with squamous cell T1a carcinoma, ten with precancerous lesions with hyperkeratosis, and ten subjects without laryngeal disease underwent high-speed laryngoscopy yielding 4,000 images per second. By means of wavelet-based phonovibrographic analysis, a set of three clinically meaningful vibratory measures were extracted from the videos comprising a total number of 15,000 video frames. Statistical analysis (ANOVA with post-hoc two-sided t-tests, P <0.05) revealed that vocal fold dynamics is significantly affected in presence of precancerous lesions and T1a carcinoma. On the basis of the three measures a discriminating pattern was extracted using a support vector machine-learning algorithm performing an individual classification in respect to the different clinical groups. By applying a leave-one-out cross-validation strategy, we could show that the proposed measures discriminate with a very high performance between precancerous lesions and T1a carcinoma (sensitivity: 100%, specificity: 100%). Although a large-scale study will be necessary to confirm clinical significance, the set of vibratory measures derived in this study may be applicable to improve the accuracy and reliability of non-invasive diagnostics of vocal fold lesions.
Introduction

Upper aerodigestive tract cancer ranks 7th amongst worldwide cancer incidences, with laryngeal malignancies being one of the most common representative [1]. About two thirds of laryngeal cancers originate at the vocal cords, and more than 90% are squamous cell carcinomas [2]. An early diagnosis is directly correlated with a favorable prognosis; more than 90% of patients with early laryngeal cancer can be cured without loosing laryngeal function [3]. In laryngology, one of the greatest challenges is thus the early detection of malignant alterations of the vocal folds and their distinction from premalignant lesions. Both resemble in appearance, showing irregular or thickened mucosa due to structural changes with invasion of the sub-epithelial space in case of malignant lesions [4] or hyperkeratosis and hyperplasia to severe dysplasia [3,5] in case of premalignant lesions.

Dysphonia initially caused by defective mucosal vibration is the first presenting symptom for malignant processes. Thus, the assessment of anomalous properties of the laryngeal dynamics induced by infiltrative processes of the deep submucosal structures of the vocal folds serves as predictor of glottic cancer. Clinically, laryngostroboscopic imaging constitutes the most important tool for functional investigation of the larynx. Since the diagnostic evaluation of the stroboscopic videos or fiber-endoscopy bases just on a subjective visual inspection slight changes of the vocal fold dynamics induced by early infiltrative processes may not be observable for the human eye. Furthermore, small carcinomas and precancerous alterations can hardly be distinguished. As a consequence, biopsy or total excision of the suspicious lesion is currently necessary, including histopathologic examination of the excised tissue for reliable diagnostics. However, clear distinction before invasive diagnostic would be essential for therapy planning. On the one hand, as malignant transformation in low dysplasia degrees is rather rare [6,7], unnecessarily augmented excision and the risk of unfavorable scarring could be prevented when pre-surgical diagnostics would be sufficiently reliable [8]. On the other hand, knowing about the malignant nature of a lesion could accelerate the adequate therapy and lead to better functional outcome.
either by excision or radiotherapy. Thus, pre-treatment diagnostics should be optimized for adequate surgical or non-surgical therapies. Several pre-surgical diagnostic methods referring to the vibratory function or the structure of a lesion have been introduced for this purpose. The vibratory function of the vocal folds including the mucosal wave propagation is commonly judged visually using stroboscopy with rigid laryngoscopes [9]. As malignant lesions infiltrate subepithelial structures mucosal wave propagation can’t be observed anymore. However, precancerous lesions and carcinomas could considerably disturb the temporal structure of affected voices. Hence, the visually assessed virtual stroboscopic slow motion video is distorted due to a desynchronized triggering of the light source flashes. Moreover, this method is based on perceptual assessment with limited reliability [10].

Other optical methods such as autofluorescence or narrow-band-imaging strive after a more detailed 2D-imaging of the superficial structure of mucosal alterations [11,12,13]. While these two methods are both highly sensitive in detecting tissue abnormalities as well as quite reliable in showing the exact superficial extension of the lesions, they are somewhat unspecific with regards to a differentiation of non-invasive versus invasive lesions. In contrast, both optical coherence tomography and confocal laser endomicroscopy provide insights into sub-surface tissue architecture at the microscopic level in vivo. It was reported that optical coherence tomography can differentiate well between precancerous lesions and early invasive cancer, whereas confocal laser endomicroscopy serves to monitor malignant alterations at the cellular level [14,15]. As both methods are usually applied during direct laryngoscopy under general anesthesia, however, they have not yet found their way into clinical routine.

In this article, we present a new method that examines the vibratory function of the vocal folds with high temporal resolution. In normal larynges, vocal fold vibrations show a three-dimensional pattern depending on the myoelastic characteristics of the vocal folds and aerodynamics during expiration [16]. Using endoscopy from above, a two-dimensional regular opening and closing
pattern of the glottis can be observed. Symmetry of the vocal fold vibrations and glottal closure is a precondition of a normal voice [17]. Nowadays, high-speed video (HSV) laryngoscopy allows for new insight into physiologic and pathologic mechanisms of the function of vocal folds [18]. HSV captures vocal fold dynamics in realtime and is suited for a quantitative analysis. Due to its objective nature HSV is proven to be more reliable than stroboscopy [19]. Supplemented with appropriate computerized analysis procedures HSV allows for the detection of even slight disturbances of vocal folds [20].

Several image processing approaches have been developed for the analysis of the vibratory characteristics of vocal folds. Most of current approaches analyze the time-varying glottal area [21] enclosed by the vocal folds or the lateral deflection given by the distance between both vocal folds along the glottal axis [22]. For phase asymmetries concerning left and right vocal fold and the anterior-posterior (AP) dimension, a high diagnostical relevance was shown [23,24,25]. Lateral and anterior-posterior asymmetries are seen to characterize changes of mass or tension of the vocal folds during vibration [17, 25, 26], and therefore need to be considered for a full and detailed quantitative analysis.

Lohscheller et al. introduced a comprehensive analysis approach that allows the quantitative description of the entire two-dimensional vibration pattern of both vocal folds along the visible anterior-posterior dimension (glottal axis) [27]. The procedure extracts and describes the time-varying contours of the two medial vocal fold edges as a function of the glottal axis. Following the extracted spatio-temporal information of vocal fold vibration is mapped into a clinically meaningful 2D image that encodes the time-varying distances from both vocal folds in relation to the glottal midline axis as color information. The resulting graph is termed phonovibrogram (PVG) [27]. It was demonstrated that PVGs transform the relevant motion information into characteristic static geometric structures that comprehensively describe vocal fold vibration patterns [28].
Besides providing a powerful diagnostic tool for visual assessment, PVGs further form the basis for a computerized analysis of vocal fold vibrations. For this purpose, the geometric shape of the PVG structures are extracted and quantitatively described using image processing and machine-learning approaches. Voigt et al. demonstrated the applicability of PVGs for the discrimination of healthy and paralytic larynges [29], as well as healthy and functional voice disorders [30]. In this first approach a high number of more than 300 features were extracted from the PVG that are, however, rather difficult to interpret in the clinical context. To overcome this problem of high dimensionality and low interpretative power, a wavelet-based approach was recently proposed [31]. This method is capable to condense the entire information about the opening and closing mechanism of the vocal folds encoded within a PVG within a very low number of parameters, which are interpretable under clinical aspects as they are related to the glottal closure types defined in the basic protocol for functional assessment of voice pathology elaborated by the European Laryngological Society [32]. The wavelet-based PVG analysis approach showed promising results for the classification of functional and organic voice disorders [33].

As distinction between precancerous diseases and malignant alterations of the vocal folds relies primarily on endoscopy, in this study, we examine whether high-speed laryngoscopy combined with PVG analysis might contribute to differentiate between healthy vocal folds, precancerous lesions and carcinomas of the vocal folds. By extending the PVG-wavelet approach, we investigated anomalies of vocal fold vibratory characteristics in the presence of precancerous lesions and T1a-carcinoma, and compared the results to data obtained from healthy subjects.

**Materials and Methods**

Laryngeal high-speed video recordings were performed on 30 subjects. Ten subjects (H1-H10, 2 females, 8 males, age: 58.0±15.1 yrs) showed no signs of voice disorders and were used as control group. Ten subjects (C1-C10, 2 females, 8 males, age: 62.7±10.0 yrs) were diagnosed with a unilateral squamous cell T1a-carcinoma (6 left-sided, 4 right-sided). The remaining ten
subjects suffered from precancerous lesions with hyperkeratosis (P1-P10, 1 female, 9 males, age: 62.9±11.9 yrs). In the precancerous lesions group, histological examinations showed hyperkeratosis to severe dysplasia.

All subjects were examined while sitting on a chair with straight but slightly back tilted head (to facilitate endoscopy). They were instructed to phonate a sustained vowel /ae/ at a comfortable pitch and loudness during the examination procedure. Recordings of subjects with vocal fold precancerous lesions or carcinomas were performed before surgical interventions. All diagnoses were made after surgery according to histological report of excised tissue of the vocal folds.

For the high-speed endoscopy, the HRES Endocam 5562 High-Speed Video System (Richard Wolf GmbH, Tuttlingen, Germany) was used. All laryngeal recordings were captured at 4,000 frames per second in color with a spatial resolution of 256 x 256 pixels. For each subject, a sequence of 500 frames (125 ms) was considered. Images were captured with a rigid 70° endoscope (Model HRES laryngoscope, Richard Wolf GmbH) and a 300-watt Xenon light source (LP 5132, Richard Wolf GmbH). The rigid laryngoscope was coupled to the high-speed digital camera head.

**Phonovibrogram computation**

In order to compare objectively the specific characteristics of the vibration patterns between the different groups, all high-speed recordings were quantitatively analyzed using phonovibrograms [28]. The approach includes the following steps:

Initially, the vibrating medial vocal fold edges were extracted from all high-speed recordings as shown in Fig. 1a). In this study, a total of 15,000 high-speed images were successfully segmented using a specially designed and clinically evaluated segmentation procedure [34]. Within a subsequent transformation step (Fig. 1b) the distances between the vocal fold edges and the glottal midline were computed and the left vocal fold was virtually turned around the posterior glottal ending represented by the point P. For visualization purposes the computed
distance values were color-coded resulting into a corresponding color strip; here a gray-scale representation was chosen. Iterating the described procedure for an entire high-speed sequence and concatenating the resulting gray-scale strips results into a two-dimensional image that is denoted as phonovibrogram (PVG). A PVG contains the entire information about the full spatio-temporal vibration pattern of the left (upper part) and right (lower part) vocal fold along the glottal axis. In Fig. 1c) three PVG oscillation cycles of the male subject H8 are depicted. For each oscillation cycle the shape of the geometric pattern, which is accented exemplarily by the white dotted contour line within the second oscillation cycle, contains precise information about the opening and closing process of the respective vocal fold. To get a detailed description of the PVG construction process, refer to Lohscheller et al. [28].

**Analysis of Phonovibrograms**

As shown before the entire information about the two-dimensional vocal folds dynamics can be captured within a PVG. Therefore, the analysis of vocal fold vibrations is consistent with the quantitative analysis of the geometric patterns within a PVG. In the following the procedure is presented that shows how the relevant information about the vocal fold vibrations are extractable from PVGs and can be further condensed into a set of distinct clinically useful measures.

Initially from the PVGs the glottal area waveform (GAW), as well as the hemi-glottal area waveforms (H-GAW\(_{L/R}\)) were derived reflecting each vocal fold side individually. As shown in Fig. 1a) the H-GAW\(_{L/R}\) are defined as the areas spanned between the left/right vocal fold and the glottal main axis, representing the proportional change of the glottal area induced by the particular vocal fold movement. Exemplarily, the time varying H-GAW\(_{L/R}\), which were derived from healthy subject H8, are shown in Fig. 1d). PVGs and H-GAWs provided the basis for the further objective analysis of vocal fold vibrations.
Measure M1- Lateral phase delay: In case of unilateral alterations of the elasticity of vocal fold tissue, it can be assumed that temporal shifts between the vibration of the left and right vocal fold occur. Fig. 2 shows for subject P3 the variations in time of the left and right hemi-GAWs in respect to the total GAW signal. For each side the horizontal arrows represent the temporal displacement between the appropriate H-GAW and the GAW signal. The left vocal fold is in advance while the right vocal fold drags behind.

As H-GAW and GAW signals can be seen as narrow-band oscillations, phasing of each signal can be expressed as a function of time by the corresponding phase signals \( \varphi_{H-GAW_L}(n) \), \( \varphi_{H-GAW_R}(n) \) and \( \varphi_{GAW}(n) \) using a complex wavelet analysis [31]. As shown in Fig. 2b) for each point in time the angle difference

\[
(1) \quad \Theta_{L,R}(n) = \varphi_{H-GAW_{L,R}}(n) - \varphi_{GAW}(n)
\]

represents the phase shift between the particular H-GAW and the GAW signal. In order to avoid phase differences greater than \( \pi \) and less than \( -\pi \) the first measure \( \Theta_{L,R} \) is derived by transforming the angles on the complex unit circle via \( z = e^{i \varphi} \) (Fig. 2c) and computing for an entire high-speed sequence the mean phase difference in the complex plain as:

\[
(2) \quad \Theta_{L,R} = \frac{1}{N} \sum_{n=1}^{N} \text{arg} \left[ e^{i \varphi_{H-GAW_{L,R}}(n)} - i \varphi_{GAW}(n) \right] \quad (\text{Measure M1})
\]

where \( N \) denotes the total number of video frames. For an entire high-speed sequence the measure \( \Theta_{L,R} \) reaches zero for perfect synchronism. For negative values the corresponding vocal fold is in advance to the GAW signal whereas a positive sign indicates a delay. In this
study the measure $\bar{\theta}_{L,R}$ is used to describe the degree of lateral phase delay occurring in the different clinical groups.

**Measure M2 - Lateral asymmetry of oscillation modes:** In order to derive quantitative information about the degree of lateral asymmetry in respect of the vocal fold oscillation modes we apply the recently introduced PVG wavelet analysis approach introduced by Unger et al. [31]. The wavelet-based approach objectively describes vocal fold dynamics by quantifying the geometrical patterns occurring within a PVG for each vocal fold independently. Fig. 3a) shows two PVG examples; one for the healthy subject H9 (upper row) and the subject C4 (lower row) suffering from a unilateral T1a carcinoma. The goal of the analysis is to extract for each vocal fold automatically the geometric structure from the PVG, which is here accented by the white dotted lines within the third oscillation cycle of the PVGs. These contours represent the spatio-temporal information about the particular vocal fold oscillation pattern. In a former work it was shown that these geometrical patterns within the PVGs are extractable by performing a PVG-wavelet decomposition [31]. The results of applying the procedure to a PVG are shown for the two subjects in Fig. 3b). In order to quantify the extracted curve progressions quantitatively a principle component analysis (PCA) was applied according to [31]. The PCA condenses the main information about the curve characteristics within the first three dominant eigenvalues (see Fig. 3c) for the left ($\lambda_{L1}$, $\lambda_{L2}$, $\lambda_{L3}$) and the right vocal fold ($\lambda_{R1}$, $\lambda_{R2}$, $\lambda_{R3}$). Due to the direct relation between PVG and vocal fold dynamics these eigenvalues represent the information about the spatio-temporal characteristics of the vibratory patterns for each vocal fold independently.

Unger et al. showed [31] that the eigenvalues $\lambda_{1-3}$ constitute a quantitative representation of the formerly only subjectively rateable glottal closure types that are defined in the basic protocol for functional assessment of voice pathology elaborated by the European Laryngological Society.
[32]. Using the quantifiable eigenvalues the degree of the lateral differences of the oscillation modes can thus easily be determined as follows:

\[
\Delta \lambda = \sqrt{\sum_{j=1}^{3} (\lambda_{Lj} - \lambda_{Rj})^2}
\]  

(Measure M2)

In Fig. 3 on the right, the second measure \( \Delta \lambda \) is exemplarily given for the physiologic and the pathologic case. In case of an increased lateral vibration asymmetry the proposed measure \( \Delta \lambda \) is considerably raised compared to the physiologic case. In this study we used this second measure to quantify the vibration asymmetry of the vocal folds induced by precancerous lesions or T1a carcinoma.

**Measure M3 - Lateral asymmetry of anterior-posterior phase delay progression:**

Depending on the particular oscillation mode, vocal folds start opening first either at their anterior, posterior or medial parts. Fig. 4a) shows a PVG extracted from subject H5, which incorporates two oscillation cycles. Here, vocal folds started opening first at the anterior third at position \( k_2 = 90\% \) from the most posterior to the anterior ending of the vocal folds. The opening process continues over time from anterior to posterior (indicated by the white arrows) resulting into a characteristic anterior-posterior phase delay progression.

For each position \( k \) alongside the glottal axis and for each vocal fold the AP phase delay can be analyzed in detail by evaluating the temporal shifts between the trajectories \( T_{L,R}(n,k) \) and the corresponding \( H\text{-GAW}_{L,R}(n) \) (see Fig. 4b). According to the measure M1 the phase delay between the trajectories \( T_{L,R}(n,k) \) and \( H\text{-GAW}_{L,R} \) can be expressed via the complex wavelet phase denoted as \( \varphi_{T_{L,R}}(n,k) \). In contrast to M1, for each vocal fold phase the delays are
computed between each line of a PVG and the appropriate hemi-GAW resulting into an anterior-posterior phase delay

\[ \phi_{L,R}(n, k) = \arg \left( \exp \left( i \phi_{H-GAW_{L,R}}(n) - i \phi_{T_{L,R}}(n, k) \right) \right) \]

along the glottal axis as shown in Fig. 4c). Here, \( \phi_{L,R}(k, n) \) quantifies a phase difference along the anterior-posterior direction. Consequently, the absolute distance

\[ \Delta \phi(n, k) = |\phi_L(n, k) - \phi_R(n, k)| \]

represents the degree of lateral AP phase asymmetry for a single trajectory. By averaging over all N frames of a high-speed sequence and over all K trajectories the measure

\[ \bar{\Delta \phi} = \frac{1}{N \cdot K} \sum_{n=1}^{N} \sum_{k=0}^{K} \Delta \phi(n, k) \quad \text{(Measure M3)} \]

constitutes the mean AP phase asymmetry that reaches zero when anterior-posterior phase displacements were perfectly identical for the left and right vocal fold.

**Analysis procedure:** To identify potential differences within the above described three measures between the groups, statistical tests were performed using Matlab R2013b. Shapiro-Wilk test was applied to test of normality. Bonferroni-Holm corrected ANOVA were subsequently used with post-hoc two-sided t-tests. The significance level was taken as \( p<0.05 \).

Besides the identification of group differences using statistics, we further investigated if the proposed measures were sufficient for a personalized classification of the subjects into the
different clinical groups. For this purpose, we used a machine-learning approach. A support vector machine (SVM) with a radial basis function (RBF) kernel was trained on those measures having the most distinctive power according to the results of the statistical analysis. For training and classification a leave-one-out strategy was applied.

Results

The three measures M1, M2 and M3 were evaluated for all 30 subjects in order to demonstrate disparities between vocal fold dynamics between the three groups.

Measure M1 - Lateral phase delay: Figure 5a) shows the phase delays $\bar{\Theta}_{L,R}$ between the H-GAWs and the GAW signals for each clinical group. As expected, for the control group the values of $\bar{\Theta}_{L,R}$ are close to zero representing just minor lateral phase delays. The value range of the right vocal fold can thus be regarded as normal range, which is accented by the grey shaded band within the graph.

For the precancerous lesions group the measure $\bar{\Theta}_{L,R}$ is shown for the subgroups “unilateral left”, “unilateral right”, and “bilateral”. For all subgroups the values $\bar{\Theta}_{L,R}$ are slightly enlarged and reveal further an increased variance. Despite the alteration, the median values of $\bar{\Theta}_{L,R}$ are however in most cases within the normal range.

For the T1a carcinoma group the measure $\bar{\Theta}_{L,R}$ is presented in Fig. 5a) and is subdivided for the subgroups “unilateral left” and “unilateral right”. For both subgroups, the affected vocal fold clearly vibrates temporally ahead of the contralateral side indicated by the negative sign of the measure $\bar{\Theta}_{L,R}$, while the phase delays of the non-affected vocal folds are within the normal range. The consistent temporal leading of the cancerous vocal fold side is a clear distinct
vibration characteristic selectively present in T1a carcinoma and not visible in the precancerous lesions group.

For clinical interpretation the lateral difference between the left and right phases $\Delta \bar{\theta} = |\bar{\theta}_L - \bar{\theta}_R|$ is of particular interest. For the three clinical groups, the values of $\Delta \bar{\theta}$ are displayed as boxplots in Fig. 5b). The phase differences were lowest for the control group. Both for the precancerous lesions groups and for the T1a carcinoma group, the distances $\Delta \bar{\theta}$ were increased in comparison to the control group. ANOVA revealed significant differences between the groups. The post-hoc tests disclosed that the phase difference $\Delta \bar{\theta}$ of the T1a carcinoma group is significantly increased in respect to the control group. The results of the statistical tests including the p-values are summarized in Table 1.

**Measure M2 - Lateral asymmetry of oscillation modes:** For each clinical group, the results of the measure $\Delta \lambda$ are presented in Fig. 6a), representing the degree of lateral asymmetry of vocal fold dynamics. As expected, the control group was characterized by low values indicating a high level of symmetry. ANOVA revealed significant differences between the groups (Table 1). The pair-wise post-hoc tests further showed that all groups exhibit significantly different asymmetry values $\Delta \lambda$. The tests prove, that significant differences exist even between the precancerous lesions and the T1a cancer group.

**Measure M3 - Lateral asymmetry of anterior-posterior phase delay progression:** For each clinical group the results concerning the degree of lateral anterior-posterior phase delay asymmetry ($\Delta \phi$) are displayed in Fig. 6b). The control group was characterized by very low asymmetry values and a small within-group variance. ANOVA test revealed significant differences between all groups (Table 1). The post-hoc tests further corroborated significantly
increased asymmetry values of the precancerous lesions group and the T1a carcinoma group in respect to the control group.

The results of the statistical evaluation proved that the measures $\Delta \lambda$ and $\Delta \phi$ exhibited the most distinctive power to discriminate efficiently between the groups. In order to investigate if a personalized classification of the subjects to the different clinical groups was feasible, a SVM (RBF kernel) was trained on the basis of $\Delta \lambda$ and $\Delta \phi$ spanning a two-dimensional parameter space. Based on a leave-one-out strategy, the parameter space could be automatically subdivided into three different parameter regions that distinguish the healthy group, the precancerous lesions and the T1a carcinoma group. Overall, all subjects except one could be classified correctly. The misclassified subject originally belonging to the “precancerous lesions” group was classified as healthy (see Fig. 7).

Considering the automated classification of the groups “T1a carcinoma” vs. “control group”, a sensitivity (true positive rate, TPR) of 100% as well as the specificity (true negative rate, TNR) of 100% could be achieved. In case of the two-class problem “precancerous lesions” vs. “control group” sensitivity and specificity are TPR=100% and TNR=90%. For the intermediate two-class problem “precancerous lesions” vs. “T1a carcinoma”, all subjects were classified correctly.

Discussion
In this article, we describe the first objective approach to distinguish squamous cell T1a-carcinoma from precancerous lesions of the vocal folds with reference to normal vocal folds based on a computerized analysis of laryngoscopic high-speed videos. The here presented analysis of vocal fold dynamics is performed by extracting in a first step the vibrating vocal fold edges from the high-speed video-stream [33] and condensing the extracted vocal fold dynamics into phonovibrograms (PVGs) [27, 28]. Former studies already showed that PVGs are principally
suitable for detecting even slight inter- and intra-individual changes of vocal fold dynamics [29, 35].

In this study, we applied a wavelet-based approach for the quantitative analysis of PVGs in order to describe objectively the vibratory characteristics alongside the entire glottal axis with a limited set of clinically meaningful parameters [31, 33]. Since carcinoma as well as precancerous lesions change the internal structure of the vocal folds, we propose three parameters that are designed to measure the medio-lateral (coronar) and anterior-posterior (sagittal) asymmetry and phasing of the vocal folds for the squamous cell T1a carcinoma, the precancerous lesions and the healthy group.

The first measure $\Theta_{L,R}$ quantifies the medio-lateral phase displacement in respect to the glottal area waveform (GAW). Due to the inertia of a unilateral mass augmentation in the T1a group one would expect a slight delay of the affected vocal fold. However, the results show that the affected vocal folds exhibit a distinct negative phase-delay in respect to the GAW signal. This means that the affected vocal fold’s movement tends to be predominantly in advance of the contralateral side. The reason can be seen in the characteristic stiffness of the vocal fold that was found to be significantly greater than in normal vocal fold tissue [36]. For the precancerous lesions group, the sign and amount of the phasing is distributed in a broader range. However, statistical testing proved that for carcinoma the absolute lateral phase delay between the vocal folds is significantly increased whereas the increment for the precancerous lesions provides no significant difference.

Likewise, significantly increased values of the lateral vibration asymmetry of oscillation modes $\Delta \lambda$ and the lateral asymmetry of the AP phase delay $\Delta \phi$ are confirmed by statistical analysis from healthy subjects over the precancerous lesions group to the T1a group. In contrast to conventional approaches, which analyze frequently the lateral asymmetry of amplitude values, the here applied measures $\Delta \lambda$ give information about the lateral asymmetry with regard to the entire spatio-temporal oscillation pattern. It therefore comprises much more information about
the entire laryngeal dynamics than conventional asymmetry measures. The asymmetry of the lateral AP phase delay $\Delta \phi$ is described for the first time. It shows minimal variance for the healthy group and is thus very selective for the differentiation between the physiological and pathological cases. Furthermore, in combination with $\Delta \lambda$ it allows a correct classification of each subject to its appropriate group as shown by the machine learning approach. The different clinical groups showed clear delimitable regions within the parameter space.

In this first evaluation, high-speed laryngoscopy in combination with the PVG analysis is a promising approach to distinguish malignant from precancerous lesions and from healthy vocal folds. The parameters received by the wavelet-based analysis were highly sensitive to differentiate the diverse vocal fold vibration patterns. The results suggested that the presented method has the potential to improve and complement the assessment of specific properties of vocal fold dynamics in a quantitative way. It needs however to be mentioned that the here presented results refer to the identification of squamous cell carcinomas, which is the most frequent vocal fold cancer. Thus, a direct transfer to other cancer as to the Ackerman’s tumor is not possible since it grows more superficially. A multi-center large-scale study is to be pursued to confirm and generalize the results of this research study.

The computing time for the entire procedure is in the range of just a few minutes so that it can be directly applied subsequent to the laryngeal examination. The only preconditions are the availability of a laryngeal high-speed video system and a sufficient image quality of the video data (illumination, fogging of the optic, and no vocal fold occlusion). Concerning the high-speed system, it can be expected that due to the technological change in imaging systems high-speed laryngoscopes will potentially replace the currently available stroboscopes in the foreseeable future. Due to its high sensitivity and the direct clinical applicability the described procedure has the potential to improve the diagnostic process flow by lowering the number of unnecessary biopsies and therewith the risk of unfavorable scarring and persistent hoarseness after a unnecessary removal of non-malignant tissue.
By combining this method with further new imaging technologies such as auto-fluorescence, narrow-band imaging, optical coherence tomography coupled to endoscopy, or confocal laser endoscopy, a further step towards an improved objective diagnostics may be feasible [15]. In particular, the combination with optical coherence tomography or narrow band imaging might on the one hand identify the nature of a lesion and on the other hand also its boarders [37]. We aim at a more precise indication and extent of surgical intervention in vocal fold pathology resulting in a better outcome for the patients' voice.

Acknowledgments

This work is supported by the German Research Foundation (DFG). Grant no. LO-1413/2-2.
References


Table 1

Results of statistical analysis concerning the identification of potential differences between the healthy control group (H), precancerous lesions group (P), and T1a carcinoma group (C). Bonferroni-Holm corrected ANOVA and two-sided post-hoc t-tests were performed on the measures M1-M3. Six of nine post-hoc test showed significant differences between the groups.

<table>
<thead>
<tr>
<th></th>
<th>ANOVA</th>
<th>Post-hoc t-tests (p&lt;0.05)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Groups H, P, C</td>
<td>Group H vs. Group P</td>
</tr>
<tr>
<td>M1: ΔΘ</td>
<td>0.0149*</td>
<td>5.9·10^-2*</td>
</tr>
<tr>
<td>M2: Δλ</td>
<td>9.5·10^-7*</td>
<td>6.5·10^-3*</td>
</tr>
<tr>
<td>M3: Δϕ</td>
<td>6.8·10^-4*</td>
<td>1.6·10^-3*</td>
</tr>
</tbody>
</table>
Figure Legends

Figure 1
(a) Sequence of endoscopic HS recordings of the larynx. The glottal area, enclosed by both vocal folds, is divided by the glottal axis into left and right hemi-glottal area. (b) Transforming and intensity-coding the segmented vocal fold contours for a single high-speed image. White regions visualize large distances from the glottal axis and dark sections indicate little distances to the midline. (c) PVG representation of vocal fold dynamics obtained from a segmented high-speed video containing three oscillation cycles. (d) Temporal progression of left and right hemi-glottal area extracted from a PVG.

Figure 2
(a) GAW and Hemi-GAWs of subject P3 with a precancerous lesion on the right side. (b) Phase signals of GAW, H-GAW_L and H-GAW_R. The phase is estimated using a wavelet-based analysis of phonovibrograms. The phase difference provides information about the relationship of the temporal phasing: the left vocal fold waveform is in advance of the GAW whereas the right vocal fold is slightly delayed. (c) To avoid potential phase jumps, phases are transformed to the complex unit circle shown in.
Figure 3
(a) PVGs shown from the healthy subject H9 (top) and the subject C4 with a T1 carcinoma on the left vocal fold (bottom). (b) A wavelet-based analysis was applied [29] to extract the recurring geometric PVG pattern for left and right vocal fold separately. (c) The geometric contour patterns can be quantified adequately with merely three coefficients for each vocal fold side obtained from principal component analysis as described in [29]. The Euclidian distance between the eigenvalues of left and right side provides a measure of lateral asymmetry of oscillation modes, which is considerably increased for the subject C4.

Figure 4
(a) PVG composed of two oscillation cycles showing an anterior-posterior (AP) phase displacement. The vocal folds first open at the anterior part shown at the position $k_2 = 90\%$. The opening process gradually continues at posterior parts of the glottis indicated by the white arrows. (b) Trajectories extracted at the $k_1 = 10\%$ and $k_2 = 90\%$ positions show different phase delays in respect to glottal area waveform (GAW). (c) The phase delay progression at frame no. 35 is shown. The anterior-posterior phase displacement is reflected by the continuous progression from negative to positive trajectory angles $\phi_{L,R}$. The lateral asymmetry of AP phase delay progression $\Delta \Phi$ is finally given by the absolute phase difference between the left and right trajectory phases.
Figure 5
(a) Lateral phase delays between the H-GAWs and the GAW signals (measure M1). The control group exhibits relatively small phase displacements with just a minor spreading representing the normal range (emphasized by the grey shaded band). For the carcinoma group the affected vocal folds tend to vibrate in advance of the contralateral side whereas no preferential phase direction is seen for the precancerous lesions. (b) The absolute phase between the left and right side. Statistics revealed a significantly increased phase delay of the carcinoma group in respect to the control group. The increase for the precancerous lesions group was not significant.

Figure 6
(a) Measures M2: lateral asymmetry of oscillation modes. (b) M3: lateral asymmetry of AP phase delay progression. Both measures are significantly higher for the pathologies than the control group. Moreover, lateral asymmetry of oscillation modes is significantly different for precancerous lesions compared to the carcinoma group.

Figure 7
Parameter space of the two most distinct measures M2 and M2. The space is divided into three regions using a Support Vector Machine (SVM) with RBF kernel. For separating the control group and precancerous lesion the boundary causing the misclassification of one subject with precancerous lesions is shown. Carcinoma and precancerous lesions are separated correctly.
Figure 2

(a) GAW & H-GAWs

The left hemi-glottal area is delayed

The phase of the right hemi-glottal area is in advance of GAW

(b) Complex wavelet Phase

frame $n_1 = 4$

$\varphi_{H-GAW_L}$

$\varphi_{H-GAW_R}$

$\varphi_{GAW}$

$c)$ Phase difference

frame #4

$Z = e^{i\varphi}$

$z_L$

$z_R$

$z_{GAW}$

$\theta_L(n_1) = 0.65$

$\theta_R(n_1) = -0.42$
Figure 3

a) PVG

female, subject H9

b) extracted PVG pattern

c) quantified PVG pattern

measure M2
Lateral difference

Δλ = 6.85

female, subject C4

Δλ = 73.61
Figure 5

a) M1: Lateral phase delay $H - GAW_{L,R}$ vs. $GAW$

b) Difference of lateral phase delay $H - GAW_{L}$ vs. $H - GAW_{R}$
Figure 6

a) **M2: Lateral asymmetry of oscillation modes**

b) **M3: Lateral asymmetry of AP phase delay progression**
Figure 7

M2: Lateral asymmetry of oscillation modes $\Delta \lambda$

M3: Lateral asymmetry of AP Phase delay progression $\Delta \phi$

carcinoma
precancerous lesions
control group
misclassified subject

controls
precancerous lesions
carcinoma
Cancer Research

A non-invasive procedure for early stage discrimination of malignant and precancerous vocal fold lesions based on laryngeal dynamics analysis

Jakob Unger, Jorg Lohscheller, Maximilian Reiter, et al.

Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-14-1458

Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2014/11/04/0008-5472.CAN-14-1458.DC1

Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.