Objective evaluation and quantitative analysis of high speed recordings in unilateral vocal fold paralysis

Jörg Lohscheller, Raphael Schwarz, Tobias Wurzbacher, Hikmet Toy, and Ulrich Eysholdt
Department of Phoniatrics and Pediatric Audiology, D-91054 Erlangen, Bohlenplatz 21, Germany, e-mail:
joerg.lohscheller@phoni.imed.uni-erlangen.de

Hoarseness in unilateral vocal fold paralysis is mainly due to irregular vocal fold vibrations caused by asymmetries within the larynx physiology. In order to derive quantitative parameters which describe characteristics of disturbed vocal fold oscillations an image processing algorithm had been developed which was applied to extract the contours of the vocal fold edges from digital high speed recordings. The two-dimensional oscillation pattern were used to determine quantitative measures as a left-right and longitudinal vocal fold asymmetry factors which describe the level of asymmetries of the vocal fold oscillations. Multiple trajectories of vocal fold oscillations are extracted which describe the characteristics of vocal fold vibrations at different position alongside the glottal axis. Besides the assessment and description of vocal fold dynamics, oscillations of vocal folds are modelled using an modified version of a two-mass-model of vocal folds. The vibrations of the model are automatically adapted to the extracted vocal fold trajectories using a developed optimization procedure. The distribution of the optimized model parameters reflect the degree of disturbances of vocal fold oscillations. The algorithm had been applied to 15 subjects suffering from unilateral vocal fold paralysis and 15 healthy subjects. The obtained model parameter values show an interrelation with vocal fold perturbance and thus can serve as a basis for therapy selection and quantification of therapy outcome in case of unilateral vocal fold paralysis.

1 Introduction

Human voice arises from vibrations of two opposing vocal folds within the larynx. In normal voice production vocal folds show a mainly symmetric vibration pattern. An injury of the recurrent laryngeal nerve causes a so-called recurrent laryngeal nerve paralysis (RLNP) which inhibits a properly symmetric adjustment of vocal folds. Asymmetries within the larynx morphology and physiology due to the unilateral paralysis entail perturbations of the vocal fold oscillations which result into hoarse voice [1]. The predominant symptom of RLNP is paralytic dysphonia. Besides etiologies such as neurological diseases and viral infections, thyroid gland surgery is the major cause of RLNP [2]. The detection and quantification of RLNP is an important issue in medical voice assessment. This paper addresses the approach of visualizing and analyzing the vibration pattern of vocal folds in case of RLNP.

Therefor vocal fold vibration patterns are investigated using a high-speed camera system. By applying an image processing algorithm the movements of the vocal fold edges are extracted and visualized within a 2D-image which enables to detect asymmetries in vocal fold vibrations. By adapting a bio-mechanical model of vocal folds the extracted vibrations pattern can be reproduced and asymmetries can be quantified.

2 Methods

2.1 Data acquisition and image processing

High-speed glottography (HGG) is the recording of vocal fold vibrations during phonation using an endoscope coupled with a digital high-speed camera. In order to get a reasonable time resolution the frame rate of the camera has to exceed the fundamental frequency of the vocal folds by several times. Hence, the frame rate of the high-speed system was set to 4000 Hz. As the vocal folds can be observed in real time during motion, i.e. phonation, HGG allows the detection of vibration irregularities. The spatial resolution of the CCD-array amounts 256×256 pixel. A 250 W Xe-light lamp serves as a light source. HGG and image processing techniques are used to obtain the vocal fold oscillations for 15 healthy voice subjects and for 15 subjects suffering from unilateral RLNP. For the performance of HGG each subject is instructed to phonate a vowel /ae/ as in 'hat' in a "‘comfortable"’ way. Fig. 1 shows the recording situation (left) and a single image extracted from a high-speed recording (right). Details concerning the recording procedure are described in [3] and [4].

2.2 Visualization of vocal folds dynamics

The extraction of vocal fold edges from high speed sequences enables to describe quantitatively laryngeal dy-

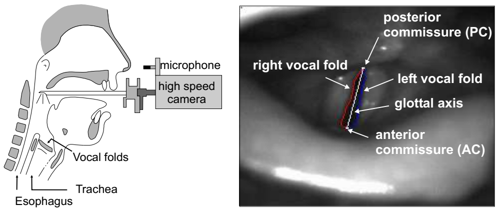

Figure 1: Examination situation: A rigid endoscope coupled to a high speed camera with a 90^o optic records vocal fold oscillations with a sampling rate of 4,000 frames per second (left). A single frame of a high speed sequence is shown on the right. The colored lines mark the edges of the left (blue) and right (red) vocal fold.

namics during phonation. However, an visual assessment of the nature of the extracted vocal fold movements is challenging. The expert has to assess both the actual geometry of vocal fold in each frame and the changes of the geometries between consecutive frames of an entire high speed sequence. Due to the successive nature of the visual assessment characteristic features of vocal fold oscillations are hardly detectable.

In the following a visualization technique is described which captures the oscillation pattern of vocal folds within one two-dimensional image. Firstly, as illustrated in Fig. 2, from a high speed sequence the opening between the vocal folds (glottis) is segmented using an adapted region growing algorithm. Thereby the parts of the vocal folds are determined which enclose the glottis. From the frames with the maximum glottal area (time-steps t and $t+m$) the location of the anterior commissure (AC) and the posterior commissure (PC) are estimated. For all intermediate frames the positions of the ACs and PCs can be derived by linear interpolation. The line between AC and PC is regarded as mean axis of the vocal folds. Subsequently, for each segmented image the vocal fold axis is sampled with a constant increment ΔL and the distances $d_{i,t}^{R/L}$ from the left and right vocal fold to the axis is calculated. The distances $d_{i,t}^{R/L}$ are regarded to be positiv if the contour of the right vocal fold is on the right hand side of the axis and negative if the right vocal fold is on the legt hand side (vice versa for the left vocal fold), see Fig. 2. Since the positions of the anterior and posterior commissures are known, information about the vocal fold edges can even be derived from this parts of the axis where no glottal area is segmented.

For the purpose of visualization the distances $d^{L/R}(i,t)$ are transferred into a two-dimensional matrix D. Fig. 3 shows schematically the registration of the distances within the matrix.

Firstly, for each frame the position of the posterior commissure is mapped as midline of the matrix. Following, beginning from the midline the values of the distances

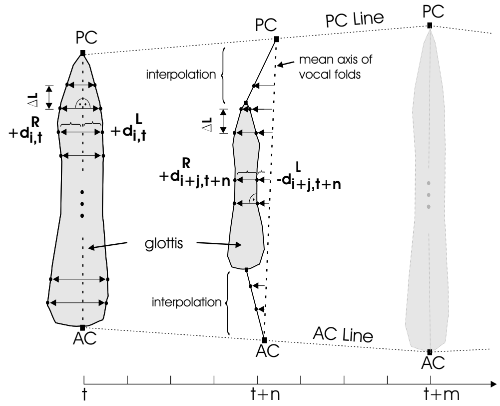

Figure 2: Calculation of the distances $d^{L/R}(i,t)$ between the anterior commissure (A) and the posterior commissure (PC) to the vocal fold axis.

$d_{i,t}^L$ of the left vocal fold are inserted upwardly (for the right vocal fold vice versa). The matrix can be visualized be color-coding the matrix entries. The intensity encodes the absolute value of the matrix entries and the color the algebraic sign (green: positive distance, red: negative distance). Fig. 4 shows an example of visualizing matrix D for normal voice production. A clear regular and symmetric v-shaped vibration pattern of both vocal folds can be seen. The image summarizes the temporal vibration pattern within all parts of the vocal folds in lateral direction. Thus, following we refer this kind of visualization as 'Lateral Displacement Pattern'-Plot (LDP-Plot).

Using LDP-Plots regular and non-regular vocal fold vibrations can be detected. Furthermore, from the two-dimensional representation trajectories $\delta(t)$ can easily be derived which describe the oscillation pattern at one specific position of the vocal folds (see lower chart in Fig. 4).

For the purpose of quantifying oscillation asymmetries which are captured within the trajectories a biomechanical model of vocal folds is introduced which is capable to reproduce characteristic oscillation patterns of vocal folds.

2.3 Classification of vocal fold vibrations

The most widely used biomechanical model of the vocal folds, the 2MM, was presented by Ishizaka & Flanagan [5]. It was simplified by Steineke & Herzel [6], see Fig. 5.

Each vocal fold is represented by two coupled oscillators according to the body-cover model [7]. The subglottal pressure P_s causes aerodynamic forces which are the driving forces to enable vocal fold oscillations. In a laminar approach, the forces can be described by Bernoulli's law [8]. The tissue properties of the vocal folds are represented by the masses $m_{i\alpha}$, stiffness coefficients $k_{i\alpha}$, and coupling coefficients $k_{c\alpha}$. The indices i, α represent the

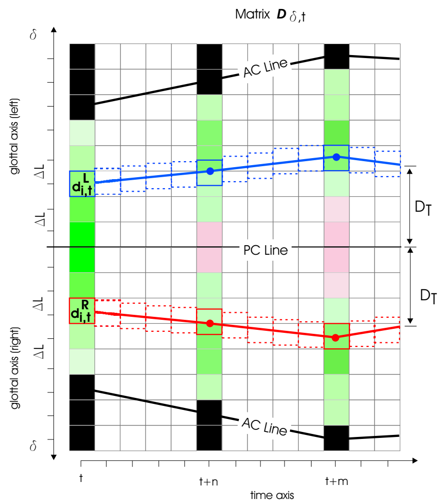

Figure 3: Visualization of vocal fold oscillations using the Lateral Distance Pattern (LDP)-Plot. The distances $d^{L/R}(i,t)$ are summarized within the 2-D matrix $D_{\alpha,t}$. For visualization purpose the values of the matrix are color-coded. Thereby the intensity encodes the absolute values of the matrix entries and the color the algebraic sign (green: positive distance, red: negative distance)

lower ($i = 1$) and upper ($i = 2$) position, as well as the left ($\alpha = l$) and right ($\alpha = r$) side. The damping coefficients of each mass, for the sake of clarity not depicted within Fig. 5, are set to the standard model parameter values as described in [5]. The masses are coupled to a rigid wall, representing the laryngeal cartilaginous framework. For the masses only a lateral displacement is allowed. Nonlinearities of the elastic forces and interactions with the vocal tract are neglected [6]. Hence, the only nonlinearities result from the driving Bernoulli force which is assumed to act only on the lower masses and from the impact forces which occur in case of collision of opposing masses [8]. The Bernoulli force depends on the subglottal pressure P_s, the geometry dimensions of the lower masses, and the positions of the upper and lower masses relative to each other [8]. The impact forces act as additional restoring forces when the vocal folds get into contact.

The differential equations describing the oscillations of the 2MM are solved with a fourth-order Runge-Kutta method [8]. According to the vocal fold oscillations the model dynamic is described by the minimum opening formed by the vibrating masses. In the following these 2MM oscillations are called *theoretical curves*.

In order to impose asymmetry between the vocal folds two factors are introduced unilaterally and defined as follows:

$$Q_{k1} = \frac{k_{1\beta}}{k_{10}}, \qquad (1)$$

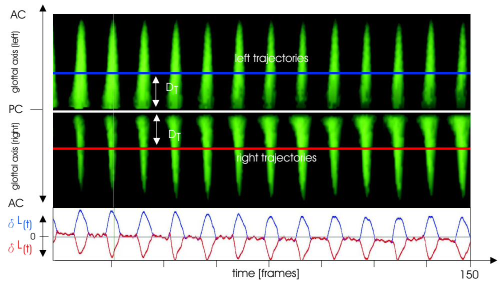

Figure 4: Visualization of vocal fold oscillations within a Lateral Discplacement Pattern (LDP)-Plot for a subject with normal voice production (top). The LDP exhibits a regular v-shaped The regular oscillation pattern is also visible within the trajecotries of the left and right vocal fold which are extracted at the blue and red lines wihtin the LDP-plot.

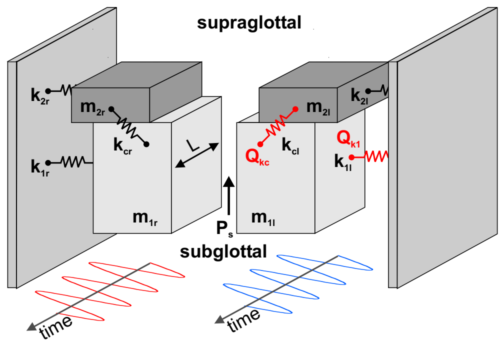

Figure 5: Two-Mass-Model (2MM) of vocal folds. For the purpose of adapting the model to experimental vocal fold oscillations the value of the red springs are modified by the scaling factors Q_{kc} and Q_{k1}.

$$Q_{kc} = \frac{k_{c\beta}}{k_{c0}}. \qquad (2)$$

The index β denotes the paralyzed side, i.e. $\beta = l$ ($\beta = r$) in case of a left-sided (right-sided) paralysis. Again k_{10} and k_{c0} denote the standard parameters [5].

Adapting the dynamic behavior of the 2MM to the extracted vocal fold oscillations can be regarded as an inverse problem. Thereby parameter values of the 2MM have to be determined in such a manner that the difference between computed and extracted vocal fold oscillation vanishes. For the inversion a real-valued genetic algorithm is applied which is capable to determine parameter values of the 2MM which reproduce vocal fold oscillations even in the case of RLNP.

3 Results

3.1 Lateral distance, velocity and acceleration pattern of vocal folds

The image processing algorithm and the parameter optimization had been applied to 15 subjects with normal voice production and 15 subjects suffering from RLNP. In the following the results are presented in detail for just one subject with normal (A) and one subject with paralytic voice (B). Fig. 6 summarizes the results of the image processing for both subjects.

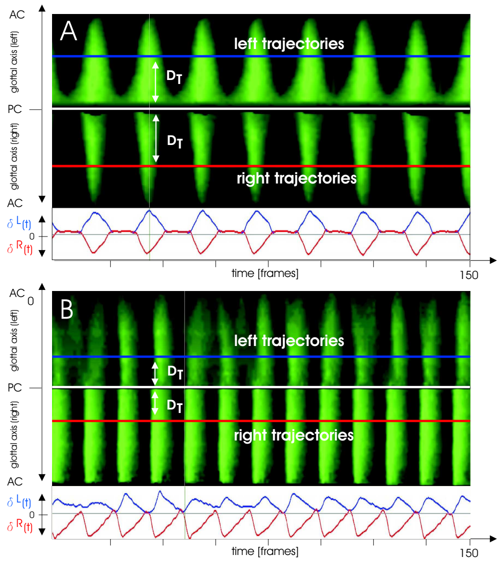

Figure 6: (A) Lateral Displacement Pattern (LDP) of vocal folds and corresponding displacement trajectories $\delta^{L/R}$ for normal voice production. (B) LDP-Plot and corresponding displacement trajectories $\delta^{L/R}$ for a subject suffering from unilateral vocal fold paralysis.

The upper figure shows the LDP-Plot and one representative trajectory for subject A. A regular and v-shaped vibration pattern can be identified with a slightly minor oscillation amplitude of the right vocal fold. Within the lower figure the LDP-Plot of subject B is shown who suffers from a left sided vocal fold parylisis. While the right vocal fold shows a highly periodic oscillation pattern the vibrations of the paralytic vocal fold are clearly disturbed. Besides a asymmetries within the oscillation amplitudes there is a distinct phase delay between left and right vocal fold.

From the LDP-Plot which results from the Matrix D also the velocities and acceleration patterns of the vocal folds can be derived by calculating the first and second derivative with respect to time. The vocal fold velocities are summarized within the Lateral Velocity Pattern (LVP)-Plot in Fig. 7.

For subject A both vocal folds show a very conform ve-

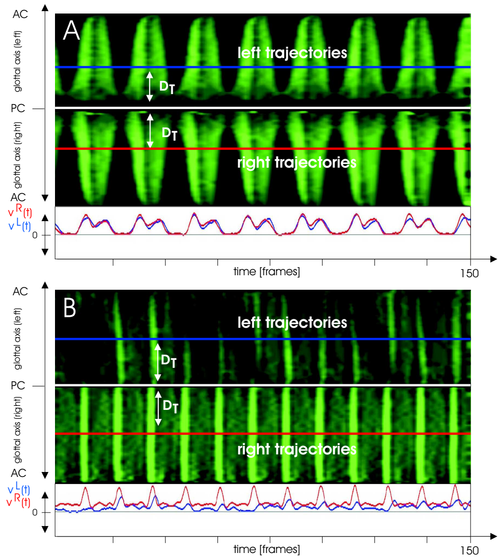

Figure 7: (A) Lateral Velocity Pattern (LVP) of vocal folds and corresponding velocity trajectories $v^{L/R}$ for normal voice production. (B) LVP-Plot and corresponding velocity trajectories $v^{L/R}$ for a subject suffering from unilateral vocal fold paralysis.

locity pattern which is also visible within the trajectories $v^L(t)$ and $v^R(t)$. Contrarily, in subject B the velocity of the paralytic vocal fold is much lower and lacks behind the right vocal fold.

The accelerations alongside the vocal folds are depicted within the Lateral Acceleration Pattern (LAP)- Plot in Fig. 8. The matrix entries coded with green pixel denote an acceleration of the vocal folds while red pixels means a deceleration.

The LAP-plot for subject A shows that both vocal folds accelerate and decelerate simultaneously two times within each oscillation cycle (see trajectories $a^L(t)$ and $a^R(t)$). Contrarily, the vocal fold accelerations in subject B are again non-symmetric and non-simultaneous. The regular and disturbed oscillation patterns could be found in all 15 normal and 15 paralytic voices.

3.2 Parameter optimization

Fig. 9 shows the results of the parameter optimization for 15 normal voice subjects (green points) and 15 subjects suffering from RLNP. For classification purpose, just the obtained results for the parameter combinations (Q_{k1}, Q_{kc}) are used, since the other optimization parameters do not provide information about model asymmetries.

The numbers $P1$ - $P15$ indicate the corresponding high-speed recordings of the paralytic voice subjects. Due to the symmetry of vocal fold oscillation for normal voice,

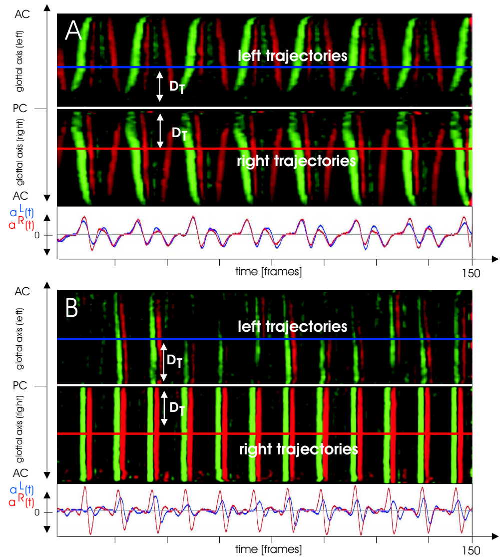

Figure 8: (A) Lateral Acceleration Pattern (LAP) of vocal folds and corresponding acceleration trajectories $a^{L/R}$ for normal voice production. (B) LAP-Plot and corresponding acceleration trajectories $a^{L/R}$ for a subject suffering from unilateral vocal fold paralysis.

$Q_{k1} \approx 1$ and $Q_{kc} \approx 1$ is obtained for the 15 healthy voice subjects, i.e. the results are located near the point of symmetry (1,1) within the (Q_{k1}, Q_{kc}) plane. The circle in Fig. 9 specifies the region where the deviations for the parameter combinations (Q_{k1}, Q_{kc}) from (1,1) are less than 0.2. Within this region the parameters of all healthy voice subjects are located. Contrarily, all parameter combinations of the pathologic subjects are distributed around the circle.

4 Discussion

Within this work we introduced a visualization technique which can be used for investigating asymmetries in vocal fold oscillations derived from high-speed recordings. The visualization of vocal folds using LDP-plots can be regarded as enhancements of the spatio-temporal vocal fold visualization (Contour-plots) described in [9]. In contrary to Contour-plots the LDP-plots use spatial information about the location of the anterior commissure (AC) and posterior commissure (PC) and can thus derive vocal fold oscillation pattern alongside the entire axis between AC and PC. The conventional Contour-plots just extracts information of vocal fold oscillations at the parts of the vocal folds where the glottis is open. Furthermore, the Contour-plots refers the displacement of the vocal folds to the glottal axis. Horizontal shifts of the vocal folds within the oscillation cycles can thus not be detected. The LDP-plots overcome the principal drawback of the Contour-plots and contain novel information

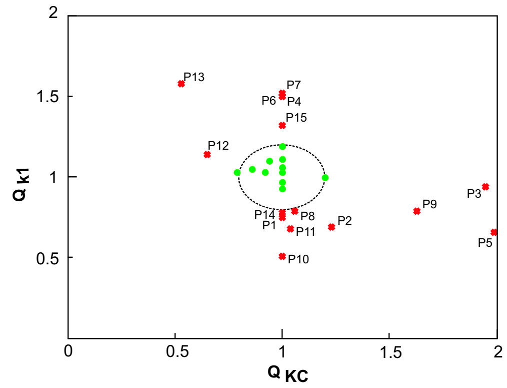

Figure 9: Result of adapting the 2MM to experimental trajectories in 15 subjects suffering from RLNP (red crosses) and 15 healthy voice subjects(green points) within the parameter plane Q_{k1}, Q_{kc}. The parameter combinations of non-pathologic subjects are closely gathered near the point of symmetry (1,1) while the parameter combinations of the pathologic subjects are scattered almost throughout the entire parameter plane.

of vocal fold oscillations.

From the LDP-plots the velocity and the acceleration pattern of vocal folds can be derived and be visualized. The LVP and LAP-plots are highly sensitive for vocal fold asymmetries and can thus detect even slight nonregularities within the vibration patterns. Combining the three plots the specific features of vocal fold oscillations in case of RLNP can be visualized.

For classifying healthy and paralytic voice from the LDP-plots trajectories are extracted which are reproduced by adapting the biomechanical 2MM using parameter optimization. The resulting position of the optimization parameters in the plane (Q_{k1}, Q_{kc}) can be used for classification purpose. According to Fig. 9 the parameters of the healthy voice subjects are located close to the point (1,1) within the parameter space (Q_{k1}, Q_{kc}). The depicted circle with a radius of 0.2 comprises all parameter combinations calculated for the healthy subjects. In contrast, the resulting parameter values for the paralytic voice subjects are scattered around the circle. The distance to the point (1,1) is clearly increased compared to the healthy subjects. Hence, a healthy voice can be classified by a parameter location close to the "‘normal"’ circle. The radius defines the bandwidth for healthy voice within the parameter space. The degree of the voice disorder due to unilateral recurrent laryngeal nerve paralysis influences the position of the optimization parameters within the parameter plane (Q_{k1}, Q_{kc}). The parameter values can be used as a measure for the degree of laryngeal asymmetry. They may allow to classify more specifically the different degrees of voice disorders. The clinical use of the inversion procedure can support to quantify a voice disorder which may be necessary for therapy selection as well as for a quantification of therapy outcome.

5 Acknowledgment

The work was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG), EY15/11 and the Sonderforschungsbereich SFB 603.

References

[1] N. Isshiki and M. Tanabe and K. Ishizaka and D. Broad, 'Clinical significance of asymmetrical vocal cord tension', *Ann. Otol. Rhinol. Laryngol.*, Vol. 86. pp. 58-66. (1977)

[2] E. Yumoto and R. Minoda and M. Hyodo and T. Yamagata, 'Causes of recurrent laryngeal nerve paralysis', *Auris. Nasus. Larynx*, Vol. 29. pp. 41-45. (2002)

[3] T. Wittenberg and M. Moser and M. Tigges and U. Eysholdt, 'Recording, proccessing and analysis of digital highspeed sequences in glottography', *Machine Vision and Applications*, Vol. 8. No. 12. pp. 399-404. (1995)

[4] M. Tigges and T. Wittenberg and F. Rosanowski and U. Eysholdt, 'High-speed imaging and image processing in voice disorders', *Proc. SPIE 2927* 1996, Wien, pp. 209-215

[5] K. Ishizaka and J. L. Flanagan, 'Synthesis of voiced sounds from a two-mass model of the vocal cords', *Bell Syst. Techn. J.*, Vol. 51. pp. 1233-1268. (1972)

[6] I. Steinecke and H. Herzel, 'Bifurcations in an asymmetric vocal fold model', *J. Acoust. Soc. Am.*, Vol. 97. pp. 1571-1578. (1995)

[7] M. Hirano, 'Morphological structure of the vocal cord as a vibrator and its variations', *Folia Phoniatr. Logop.*, Vol. 26. (1974)

[8] M. Doellinger and U. Hoppe and F. Hettlich and J. Lohscheller and S. Schubert and U. Eysholdt, 'Vibration parameter extraction from endoscopic image series of the vocal folds', *IEEE Trans. Biomed. Eng.*, Vol. 49. pp. 773-781. (2002)

[9] J. Neubauer, P. Mergell, U. Eysholdt and H. Herzel , 'Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial modes', *IEEE Trans. Biomed. Eng.*, Vol. 49. pp. 773-781. (2002)