Non-stationary Modeling of Vocal Fold Vibrations During a Pitch Raise

T. Wurzbacher[1], R. Schwarz[1], U. Hoppe[2], U. Eysholdt[1], and J. Lohscheller[1]

[1]Department of Phoniatrics and Pediatric Audiology
Erlangen University Hospital, Germany
`Tobias.Wurzbacher@phoni.imed.uni-erlangen.de`

[2]Institute of Biomedical Engineering and Informatics
Ilmenau Technical University, Germany
`ulrich.hoppe@tu-ilmenau.de`

Abstract

In clinical examination, vocal fold vibrations are usually visualized during a stationary sustained phonation. This experimental condition is easy to handle, but does not reflect the voice in every day life. The objective of this study is to investigate the laryngeal mechanism during a non-stationary phonation. A monotonous pitch raise paradigm (MPRP) as an example of a non-stationary phonation is considered. An extension of the biomechanical two-mass-model (2MM) with time-dependent parameters is developed to model vocal fold vibrations and belonging physiological parameters. The procedure iteratively adapts the extended 2MM via frequency and amplitude matching to real vibrations of vocal fold edges, obtained from an endoscopic investigation with a high-speed camera system (high-speed glottography, HGG). In simulations the proposed fitting scheme is capable to approximate the non-stationary oscillation of the vocal folds. The verification of the algorithm is exemplarily demonstrated with a synthetically generated data set and recordings of a female subject with normal voice. The time-dependent model parameters adjusted by the proposed inversion algorithm are estimates of the adaptation of the vocal folds during a non-stationary phonation.

1. Introduction

Voice examination in clinical routine is most commonly based on stationary processes as sustained phonation, either for acoustical analysis or for stroboscopy. Periodic signals allow simpler and easier procedures than aperiodic events would do. However, during running speech, the human voice is driven by continuous non-stationary conditions. Consequently, the voice examination under stationary circumstances is an approximation of order zero.

In the past, mainly the vocal onset was used as a non-stationary task: it was expected that instabilities were easier to observe during the onset period, when the dynamic equilibrium between muscular tension and subglottal pressure was regulated to a balance [2].

Real-time recording techniques, the most advanced being the high-speed glottography (HGG), allow to observe the vibrating vocal folds during non-stationary phonation. An overview over the actual status of the HGG recording technique is given in [1].

Vocal fold dynamics can be extracted from the HGG recordings [2, 3]. With the aim of deducing internal states of the vocal folds, biomechanical models were introduced. They fulfilled successfully their experimental test phase and are
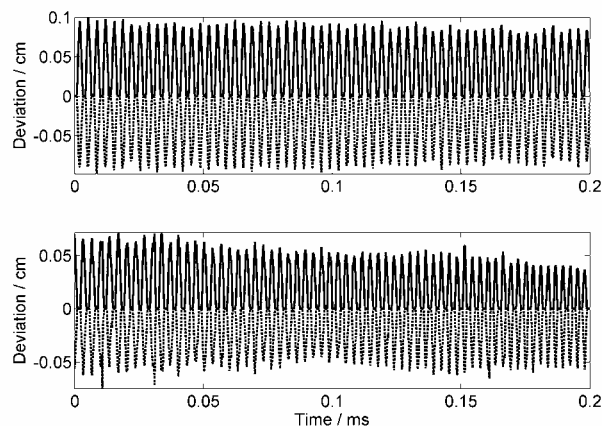

Figure 1: *Movements of the left (solid) and right (dotted) vocal fold edges computed from a digital image series (medial glottal third). Top: Sustained phonation with constant pitch. Bottom: Section of a monotonous pitch raise.*

currently under clinical trial. They restrict themselves on the steady state section after the vocal onset, i.e. on a stationary vibration with constant fundamental frequency as it is shown in the upper part of Fig. 1 with time- and space-related symmetry.

The main disadvantage of steady state examination is, that it is unpredictable whether there exist parameter constellations, which produce instability of the vibration. It could be possible, that an "instable" parameter combination is situated very close to a stable one within the phase space, although only a stable vibration is observed.

To produce a continuous parameter change during the examination, we tested several types signal paradigms. They had to be easy to perform for both the proband, who generally is untrained in musical terms, and the examiner. A "monotonous pitch raise paradigm" (MPRP) seemed to be an acceptable compromise (Fig. 1, lower trace) [3]. The proband is free to choose the starting and the end pitch.

In this paper, we present a biomechanical model for interpretation of MPRP-based HGG recordings, which have prior to be evaluated by image processing algorithms (for more details, see [2]). The model is derived from an extension of the two-mass model (2MM) of Ishizaka and Flanagan [4] with time-dependent parameters. The solution of the 2MM

differential equation is inverted in the sense that a parameter set is numerically optimized, which adapts the oscillations of the biomechanical model to vocal fold vibrations. As a first approximation step, the inversion algorithm is verified with a synthetically generated data set, and in a second step is applied to the record of a single healthy female subject.

2. Methods

2.1. High-speed glottography

A black and white HGG camera is used, 128 x 256 pixel resolution, with a recording rate of 3704 Hz. The frame rate is equivalent to the sampling frequency of the pixel series.

A young (22 y) female subject with a normal voice and no history or clinical signs for voice disorder was instructed to phonate the vowel /a/. She was instructed to increase the pitch during the phonation monotonically from a comfortable frequency up to an arbitrary higher one (MPRP). No more instructions about pitch shift duration or frequency range were given. Two HGG sequences were recorded: 1) under sustained phonation, 2) under MPRP condition.

2.2. Image Processing

The motion of the free vocal fold edge is extracted from the HGG series by image processing [2]. The actual procedure uses an adaptive contour algorithm [5]. The time-varying distances of two opposing vocal fold edge points from glottal midline are regarded as experimental trajectories for the left and right side. The movements are evaluated only in the medial glottal third, as the oscillation amplitude of the vocal folds is largest in this region, which increases the accuracy of the extraction.

A laser projection system is used to calibrate the images in order to determine the size of the vocal folds in metrical units [6].

2.3. Time-dependent two-mass model

A basic version [7] of the 2MM of Ishizaka and Flanagan [2] is used in this work to model the vocal folds. With considerable simplifications the lumped 2MM describes the main features of the vibration pattern of the vocal folds and can be regarded as the simplest model with a physiological relevance [7, 8, 9]. Unlike other complex biomechanical models with many parameters, the 2MM can be interpreted more easily along with reduced computational costs.

Within the 2MM one vocal fold is represented by two coupled oscillators, which are set into vibrations by a subglottal air pressure P_s. The deviations of the model's lower masses are regarded as theoretical trajectories. Interactions with the vocal tract as well as nonlinearities of the elastic forces are neglected. Hence, the only incorporated nonlinearities are the driving Bernoulli force, which is assumed to act solely on the lower masses, and the two impact forces. These impact forces act as additional restoring forces, which are represented by spring constants, when the vocal folds get into contact. The vibrating masses, vocal fold tensions, and damping coefficients of the vocal folds are denoted by the parameters m, k, and r. In the following, a symmetrical setup between the left and right side is assumed which is typical for normal voices [7]. To simplify matters, corresponding indices for the upper and lower as well as the left and right sides are neglected.

Originally, the model has fixed model parameters, which results in stationary oscillation patterns. In order to describe a pitch raise with the 2MM, an extension of this model with time-dependent versions of the parameters $k(t)$, $m(t)$, and $P_s(t)$ is introduced. These time-varying model parameters are considered as the most physiologically relevant components for a non-stationary phonation [10].

The extended 2MM is characterized by a set of differential equations, which are solved by a fourth-order Runge-Kutta method. While dealing with time-dependent model parameters, in particular with mass $m(t)$, the general form of Newton's second law

$$F = \frac{d}{dt}(mv) = m\frac{d}{dt}v + v\frac{d}{dt}m \qquad (1)$$

has to be taken into account for the equations of motion. Hence, the time derivative of mass m is incorporated in the model equations.

2.4. Numerical inversion scheme

Solving the differential equations of the 2MM with a given parameter set is called the 'direct problem'. As a result the deviations of the lower masses of the model are obtained. They are called theoretical trajectories as mentioned above. The 'inverse problem' is to find a parameter set, which produces the observed motion curves, starting from given trajectories. For clinical purpose the inversion of the 2MM is used to determine physiologically relevant parameters, which are not directly observable. The model parameters must be determined in such a manner that the experimental trajectories as retrieved from the HGG recording and those computed from the 2MM are similar. For healthy voices and unilateral vocal fold paralysis the inversion of vocal fold oscillations of a stationary phonation has previously succeeded [7, 9]. However, these inversion schemes are not applicable for a non-stationary phonation, since they were developed for constant model parameters. Hence, the inversion of a pitch raise is a new task and demands novel adoption algorithms.

The most important parameter influencing the vibratory patterns of the model is the subglottal pressure $P_s(t)$ that mainly affects the amplitude of the theoretical trajectory. Furthermore, the masses $m(t)$ and tensions $k(t)$ predominantly influence the frequency of the oscillation as well as the amplitude. To adjust oscillation frequency $f_o(t)$, spring constants $k(t)$ are increased, while the masses $m(t)$ are reciprocally scaled with a scalar weighting factor $Q(t)$ [7]. This reflects the fact that the vibrating portions of the vocal folds diminish with an increase of muscle tension. This is expressed by

$$\begin{aligned} k(t) &= k_o \cdot Q(t) \\ m(t) &= m_o / Q(t), \end{aligned} \qquad (2)$$

where k_0 and m_0 denote the standard parameters of the 2MM [4]. Here, the ancillary condition $Q(t)$ reduces the dimensionality for optimization by one.

In the following the numerical approach is presented that adapts the extended time-dependent 2MM to the experimental trajectories during MPRP. The algorithm is divided into a starting and a main phase. A schematic draft of the numerical inversion scheme is illustrated in Fig. 2.

In the starting phase of the algorithm the experimental trajectories serve as input data. First of all, the fundamental

frequencies $f_0^e(t)$ of the experimental trajectories are determined via a short term Fourier analysis. Overlapping windows are used and short term stationary is assumed which is well fulfilled for the case of a monotonous pitch raise. In another step, the experimental envelopes $env^e(t)$ of the trajectories' amplitudes are computed. After the preparation, the model parameters are initialized, which is important to assure the convergence of the algorithm. The initialization is done in the following manner. Taking equation (2) into account, the fundamental frequencies of the 2MM are approximated by the simple mass-spring oscillator equation [7]

$$f_0(t) = \frac{1}{2\pi}\sqrt{\frac{k(t)}{m(t)}} = \frac{1}{2\pi}Q(t)\sqrt{\frac{k_0}{m_0}} \ . \qquad (3)$$

Solving (3) for $Q(t)$ yields

$$Q(t) = 2\pi f_0(t)\sqrt{\frac{m_0}{k_0}} \ . \qquad (4)$$

Setting $f_0(t) = f_0^e(t)$ and taking $Q(t)$ after (4) results in appropriate initial values for $k(t)$ and $m(t)$. Subglottal pressure $P_s(t)$ is set up to two times of the standard pressure to $16\,\mathrm{cm}\ \mathrm{H_2O}$. This is done to obtain a self sustained-phonation in the 2MM, due to the higher stiffnesses, which occur at high frequencies.

In the main phase of the numerical inversion scheme the frequencies of the theoretical trajectories are matched to the experimental ones. If the frequencies $f_0^t(t)$ of the theoretical trajectories are lower than the experimental ones $f_0^e(t)$, the scalar weighting factor $Q(t)$ of equation (2) is increased and vice versa, until the differences between both frequencies fall below a desired threshold. Thereafter, the differences of the envelopes between experimental and theoretical trajectories are compared. If the theoretical envelopes are below the experimental ones subglottal pressure is increased and vice versa. The order of magnitude for the increment of $Q(t)$ and $P_s(t)$ is about 10^{-3}. Since subglottal pressure $P_s(t)$ has an influence on the fundamental frequency, the pitch is re-matched again after the envelope matching. This procedure is iterated until the envelope and frequency characteristics of the theoretical trajectories match the experimental ones. Here, thresholds for both matching parts are applied. The outputs of the inversion scheme $m(t)$, $k(t)$, and $P_s(t)$ are estimates of the laryngeal mechanisms during the MPRP.

As a measure of a possible error, the inversion scheme is exemplarily verified with a synthetically generated pitch raise. For this purpose, the parameters of the extended 2MM are set to predefined values over time and the deflections of the lower masses serve as input trajectories. Next, the inversion algorithm is applied to the synthetically generated trajectories and the output parameters, shown in Fig. 3, are compared to the predefined ones.

A first application to real data, the inversion algorithm is run to experimental trajectories, depicted in the upper Fig. 4, of a female subject with normal voice.
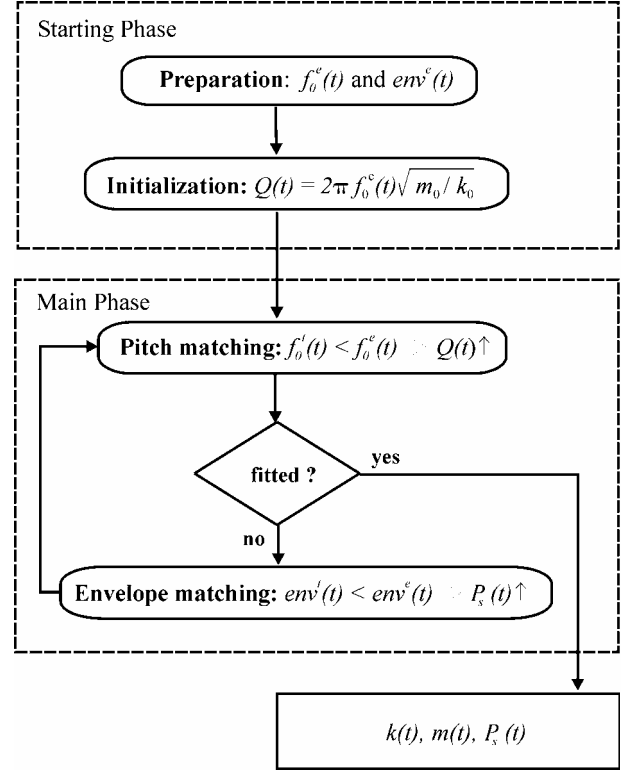

Figure 2: *Numerical inversion scheme to estimate biomechanical parameter sets describing the laryngeal mechanisms during MPRP.*

3.   Results and Discussion

In Fig. 3 the approximation results of the extended 2MM to a synthetically generated trajectory of a pitch raise are shown. A close match between the simulated frequency increase (+) and the adapted frequency (-) is visible in Fig. 3a). Considering the predefined parameters of the pitch raise a good approximation could be achieved. Differences between the parameters are hardly detectable. Even the predefined subglottal pressure $P_s(t)$ is adequately estimated, which rises from 8 to 15 cm H_2O. The correlation coefficient between experimental and theoretical trajectory amounts 0.96. This result demonstrates the applicability of the inversion scheme for parameter extraction during a pitch raise.

The inversion scheme is also applied to real vocal fold vibrations of a female subject. From the HGG recording 1000 consecutive frames were processed. This corresponds to a time interval of 270 ms. The selected frames include the pitch raise whereas the phonation onset is excluded. The time where the vocal folds are in contact diminishes until no closure is visible at high frequencies at all. The trajectories for the left and right side exhibit symmetry. Thus, only the left side is exemplarily chosen to discuss the results in the following. The envelopes $env^e(t)$ as determined by the inversion scheme are plotted in the upper Fig. 4, too. The oscillation frequencies $f_0^e(t)$ of the vocal folds, which increase from 200 Hz to 500 Hz, are displayed in the lower graph of Fig. 4. Applying the inversion algorithm to the experimental trajectories given in Fig. 4, results in the outputs
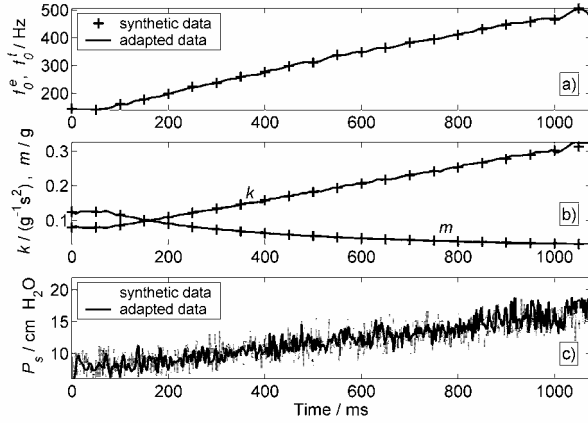
Figure 3: *Simulated pitch raise (MPRP) with predefined parameters (+ and dashed) and the output of the inversion scheme (solid).*
a) Frequencies (left and right) $f_0^e(t)$ *and* $f_o^t(t)$.
b) Time-dependent model parameters mass m(t), spring k(t).
c) Time-dependent subglottal pressure $P_s(t)$.

Figure 5: *Results obtained by the inversion algorithm during a pitch raise of female subject with normal voice. Graph a) illustrates the pitch match between the experimental (dotted) and the theoretical (solid) curve. Fitting output of time-dependent model parameters are plotted in b) mass m(t), spring k(t), and c) subglottal pressure $P_s(t)$ with its sliding mean value.*
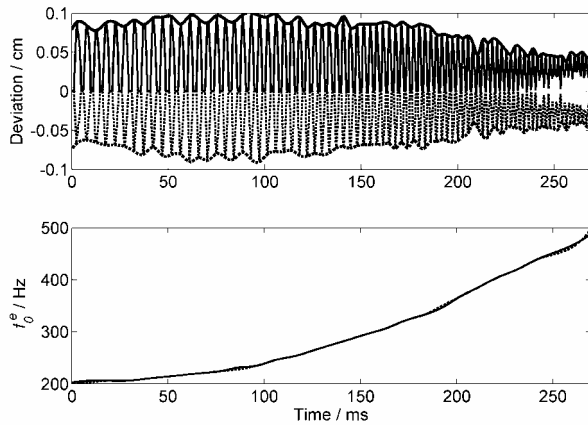

Figure 4: *Top: Experimental trajectories for the left (solid) and right (dotted) vocal fold of a female subject with normal voice during MPRP.*
Bottom: Corresponding fundamental frequencies $f_0^e(t)$ *(left and right).*

Figure 6: *Theoretical trajectory (solid) plotted over the experimental (dotted) one. The correlation coefficient between both curves is 0.91.*
Top: Entire MPRP time interval of 270 ms.
Bottom: Zoomed into a 38 ms time interval.

that are depicted in Fig. 5. In Fig. 5a) the pitch raise $f_0^e(t)$ and the approximated frequency $f_0^t(t)$ are shown. Chart 5b) displays the changing of tension $k(t)$ and of mass $m(t)$. Tension increases and mass diminishes over time. At the beginning of the pitch raise $t_0 = 0$, the estimated value for tension $k(t_0)$ is increased and for the mass $m(t_0)$ decreased respectively, compared to the standard values k_0 and m_0. This reflects the high pitch of female voice. The estimated subglottal pressure fluctuates around its sliding mean value illustrated in 5c). Like expected $P_s(t)$ increases during pitch raise, since from a physiological point of view the subglottal
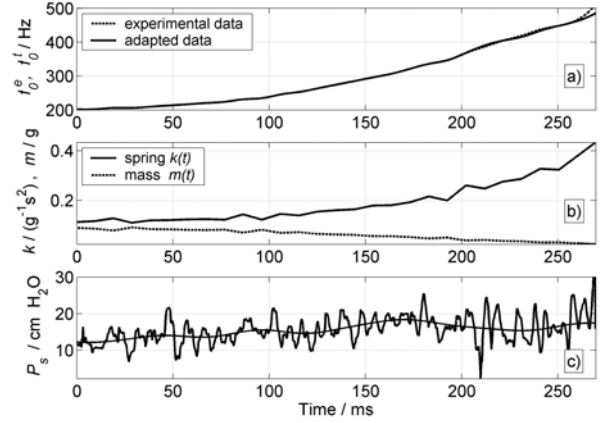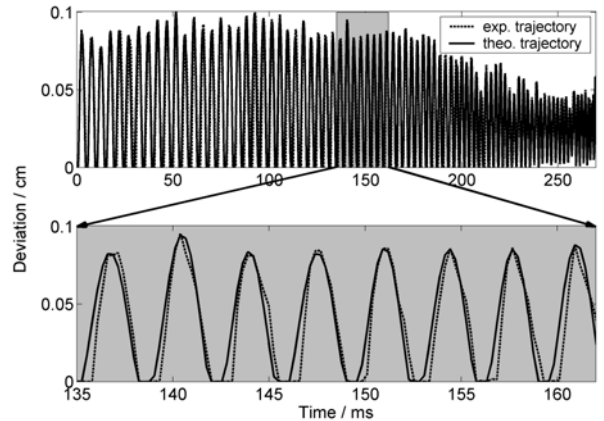
pressure has to sustain the vocal folds oscillating, even if they are stiffened [10]. All parameters found by the inversion scheme are within a physiological relevant region.

Plotting the adapted theoretical trajectory over the experimental one, as displayed in Fig. 6, reflects the close match of the trajectories. The general shape, like oscillation amplitude and frequency, is reproduced. This is also visible in a shorter time interval of the pitch raise and by the correlation coefficient between both curves that amounts 0.91. Minor differences appear during the impact of the vocal folds. Further on, there are some deviations of the theoretical trajectories from the experimental ones at high frequencies, where no glottis closure occurs.

The differences between the trajectories can be ascribed to the following aspects. Firstly, some simplifications were made by the computation of the acting forces. Also the insufficient emphasis of the vocal fold length within the 2MM for a pitch raise has to be mentioned. Secondly, the segmentation of vocal fold edges can be disturbed by the limited spatial resolution of the CCD chip of the camera. This becomes more important for the small amplitudes of vocal fold oscillations at high frequencies.

As seen in Figs. 5 and 6, the inversion algorithm adapted the 2MM to non-stationary vocal fold vibrations. The adjusted time-dependent model parameters estimate physiological parameters that cannot directly be measured. They provide an insight into the phonation mechanism during a non-stationary phonation. The clinical significance for the quantification of voice disorders of the obtained parameters has to be evaluated in an ongoing study. Further investigations will include analyzes of pathologic voices during a pitch raise, which demands more complex models of the vocal folds. Vocal fold length and its influence on the fundamental frequency as well as a more independent adjustment of muscle tension $k(t)$ and mass $m(t)$ will be incorporated in more detail.

4. Conclusions

The fitting scheme is capable to adapt a time-dependent version of the 2MM to vocal fold vibrations during a MPRP. The 2MM models, on the one hand, the oscillations of a synthetically generated pitch raise, on the other hand, experimental trajectories obtained from endoscopic investigations of a female subject with high accuracy. The extended 2MM supported by the developed inversion scheme enables the extraction of model parameters, which provide information about the adaptation of the vocal folds during a pitch raise.

5. Acknowledgement

6. References

[1] Eysholdt U; Rosanowski F; Hoppe U, (2003): Irregular vocal fold vibrations caused by different types of laryngeal asymmetry, *Eur Arch Otorhinolaryngol* 260: 412 – 417.

[2] Eysholdt U.; Tigges M.; Wittenberg T.; Pröschel U., 1996. Direct Evaluation of High-Speed Recordings of Vocal Fold Vibrations. *Folia Phoniatr. Logop.* (48), 163-170.

[3] Hoppe U; Rosanowski F; Döllinger M; Lohscheller J; Eysholdt U, (2003): Visualization of the Laryngeal Motorics during a Glissando, *J. Voice* 17: 370 – 376.

[4] Ishizaka K.; Flanagan J.L., 1972. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Syst. Techn. J.* 51, 1233-1268.

[5] Lohscheller J., 2004. Dynamics of the Laryngectomee Substitute Voice Production. In *Kommunikations-störungen: Berichte aus Phoniarie und Pädaudiologie,* Eysholdt U. (ed.). Aachen: Shaker-Verlag.

[6] Schuberth S.; Hoppe U.; Döllinger M.; Lohscheller J.; Eysholdt U., 2002. High-Precision Measurement of the Vocal Fold Length and Vibratory Amplitude. *Laryngoscope* 112, 1043-1049.

[7] Döllinger M.; Hoppe U.; Hettlich F.; Lohscheller J.; Schuberth S.; Eysholdt U., 2002. Vibration Parameter Extraction From Endoscopic Image Series of the Vocal Folds. *IEEE Trans. Biomed. Eng.* 49(8), 773-781.

[8] Hoppe U., 2001. Mechanisms of Hoarseness – Visualisation and Interpretation by Means of Nonlinear Dynamics. In *Kommunikationsstörungen: Berichte aus Phoniarie und Pädaudiologie,* Eysholdt U. (ed.). Aachen: Shaker-Verlag.

[9] Schwarz R.; Lohscheller J.; Wurzbacher T.; Eysholdt U.; Hoppe U., 2004. Modeling Vocal Fold Vibrations in Case of Unilateral Vocal Fold Paralysis. *Proc. 2nd IASTED Conference on Biomedical Engineering.* Innsbruck: Acta Press, 417-059.

[10] Hoppe U.; Rosanowski F.; Lohscheller J.; Schuster M.; Eysholdt U., 2003. Endoscopic and acoustical investigation of a glissando. *Advances in Quantitative Laryngoscopy.* Hamburg.