MODELLING OF NON-STATIONARY PHONATION FOR CLASSIFICATION OF VOCAL FOLD VIBRATIONS
4TH INTERNATIONAL MAVEBA WORKSHOP

T. Wurzbacher, R. Schwarz, H. Toy, U. Eysholdt, J. Lohscheller
Department of Phoniatrics and Pediatric Audiology, University Hospital Erlangen, Medical School, Erlangen, Germany

In contrast to the endoscopic examination of a stationary phonation, the examination of a non-stationary phonation screens a broader spectrum of vocal fold oscillations. For this reason, vocal fold vibrations are investigated and recorded with a high-speed camera system in eight normal and in eight pathological voices during a pitch raise. A quantitative analysis of the observed vocal fold dynamics in terms of symmetry and regularity is done with a time-dependent two-mass model of the vocal folds. The model's parameters are numerically optimized to emulate the observed non-stationary vocal fold vibrations. These parameters permit an objective interpretation of vocal fold oscillating asymmetries and allow a classification in normal and pathological cases. The practicability of the optimization algorithm is demonstrated with a set of 242 synthetically generated data sets. By applying the optimization procedure to the recordings of the 16 subjects a correct classification into groups of normal and pathological cases was achieved.

I. INTRODUCTION

Vocal fold examination in clinical routine is based on laryngeal endoscopic techniques. Conventionally, the examination condition for the assessment of vocal fold oscillation is a stationary sustained phonation, i.e. phonating a vowel at a constant pitch and intensity. Only fractions of the voice disorder that appear at this pitch and intensity level can be registered. This limits the diagnosis of voice disorders. To overcome this drawback a non-stationary phonation is investigated that screens a broad phonation range. A "monotonous pitch raise" (MPR) paradigm is explored, where the proband is free to choose the starting and the end pitch of phonation [1].

High-speed glottography (HGG) is the state of the art real-time recording technique, which allows to observe the non-stationary vocal fold vibrations [2]. Objective medical diagnosis demands to extract parameters from the observed HGG vibrations that describe symmetry and regularity of vocal fold oscillations. These parameters enable a classification of vocal fold oscillation patterns. Quantitative parameters can be derived by adapting a biomechanical two-mass model (2MM) of vocal folds to the HGG vibration patterns. Recently, the adaptation of the vibration behavior of a 2MM succeeded in stationary vibration patterns of normal voices and in cases of unilateral recurrent laryngeal nerve paralysis [3, 4]. However, the 2MM is restricted to the steady state section after the vocal onset i.e. on a stationary phonation. In order to model a pitch raise a time-dependent 2MM for the interpretation of MPR-based high-speed recordings is presented. The model's dynamic is adjusted with five time-dependent parameters, which are used to determine asymmetries in vocal fold vibrations. These parameters are numerically optimized in the sense that the vibration behavior of the 2MM is adapted to the HGG observed vocal fold vibrations. The performance of the proposed algorithm is tested with 242 predefined data sets and is applied to eight normal and to eight pathological cases.

II. METHODOLOGY

A. Examination Conditions

Eight young female subjects (average age of 20.3 years with standard deviation of 3.7 years) served as subjects for the endoscopic investigations. The subjects had a normal voice and exhibited no history or clinical signs for voice disorders. Eight further subjects (38.3 ± 13.4 years) suffered from different voice disorders as functional dysphonia (3), Reinke's Edema (1), unilateral vocal fold paralysis (3), and vocal fold polyp (1). All subjects were instructed to increase monotonically the pitch during the phonation of the vowel /a/ from a comfortable frequency up to an arbitrary higher one (MPR). No more instructions about pitch shift duration or frequency range were given.

For each subject a MPR-HGG sequence was recorded with a frame rate of 4.000 frames per second (High-Speed Endocam, Wolf corp., Knittlingen, Germany) [2].

The dynamics of the vocal folds are represented by the motions of the vocal fold edges. Their movements are extracted from the HGG sequences by image processing at the medial glottal third, as the oscillation amplitude of the vocal folds is largest in this region [5]. The time-varying deflections of two opposing vocal fold edge points from glottal midline are regarded as experimental trajectories $d_a^{HGG}(t)$ for the left (a = l) and right (a = r) side. In the following all characteristics that are marked by superscript HGG refer to the experimental trajectories.

Models and analysis of vocal emissions for biomedical applications. 4th international workshop. October 29-31, 2005 – Firenze, Italy.
Edited by C. Manfredi. ISBN 88-8453-320-1 (online)

B. *Two-Mass Model, its Time-Dependent Extensions, and Symmetry Factors*

To emulate experimental trajectories $d_a^{HGG}(t)$ a 2MM is chosen, since the model is able to represent the dominant modes of vocal fold vibrations [3], [6]. In the following a brief summary of the time-dependent extensions of the 2MM is depicted. Within the 2MM one vocal fold is represented by two coupled oscillators, which are set into vibrations by a subglottal air pressure $P_s(t)$. Incorporated nonlinearities are the driving Bernoulli force and the impact forces. These impact forces, which are represented by spring constants c_{ia} (t), act as additional restoring forces when the model's masses get into contact. Indices used for the parameters indicate the lower (i = 1) and upper (i=2) plane of the model. The vibrating masses and spring tensions of the model are denoted by the parameters $m_{ia}(t)$ and $k_{ia}(t)$. The coupling between the lower and upper masses is represented by the spring constants $k_{ca}(t)$. The rest positions $x_{0,a}(t)$ of the upper and lower spring constants are assumed to be equal and the glottal length of the model is signed as l(t). The model is described by a system of differential equations:

$$d/dt(\mathbf{x}) = \mathbf{A}(t)\,\mathbf{x} + \mathbf{b}(\mathbf{x},t)\,. \qquad (1)$$

Matrix $\mathbf{A}(t)$ contains the aforementioned tissue properties while the non-linear parts are captured in vector $\mathbf{b}(\mathbf{x},t)$. The vector $\mathbf{x}$ contains positions and velocities of the left, right, lower and upper masses (T transposed vector)

$$\mathbf{x}^T = [x_{1l}\; v_{1l}\; x_{2l}\; v_{2l}\; x_{1r}\; v_{1r}\; x_{2r}\; v_{2r}]\,. \qquad (2)$$

In accordance with the vocal fold oscillations the model dynamic is described by the minimum opening formed by the vibrating masses. These 2MM oscillations are called theoretical trajectories $d_a(t)$. In equation (1) the more general form of Newton's second law

$$F_{ia} = m_{ia}(t)\; d/dt(\, v_{ia}(t)\,) + v_{ia}(t)\; d/dt(\, m_{ia}(t)\,) \qquad (3)$$

has been accounted for. Hence, the time derivatives of masses $m_{ia}(t)$ are incorporated in the equations of motion. Equation (1) is numerically solved by a fourth order Runge-Kutta method with a step size h fixed to the frame rate of the HGG recordings h=1/4000.

The time-dependent parameters summarized in $\mathbf{A}(t)$ and $\mathbf{b}(\mathbf{x},t)$ influence the oscillation behavior of the model. The subglottal pressure $P_s(t)$ mainly affects the amplitude of the model's oscillation and to some minor extend the oscillation frequency [1]. $P_s(t)$ can be seen as a measure for the energy flowing into the system [3]. Furthermore, the masses $m_{ia}(t)$ and tensions $k_{ia}(t)$ predominantly influence the frequency as well as the amplitude of the oscillation [1]. In the model the parameters $m_{ia}(t)$, $k_{ia}(t)$, ..., are expressed in terms of Ishizaka's and Flanagan's [8] standard parameters $k_{0,ia}$, $m_{0,ia}$, ..., by introducing factors $Q_a(t)$, $R_a(t)$, and U(t):

$$\begin{aligned} k_{ia}(t) &= k_{0,ia}\; Q_a(t), & k_{ca}(t) &= k_{0,ca}(t)\; Q_a(t), \\ m_{ia}(t) &= m_{0,ia}/Q_a(t), & c_{ia}(t) &= c_{0,ia}(t)\; Q_a(t), \\ x_{0,a}(t) &= x_{0,a}\; R_a(t), & P_s(t) &= P_{s0}\; U(t). \end{aligned} \qquad (4)$$

As a measure of asymmetry between the oscillations of the left and right side of the model the ratios

$$Q(t) = Q_l(t) / Q_r(t) \quad \text{and} \quad R(t) = R_l(t) / R_r(t) \qquad (5)$$

are introduced [7]. If $Q_l(t)$ and $Q_r(t)$, respectively $R_l(t)$ and $R_r(t)$ differ from each other the model's oscillation become asymmetric. Time-varying irregularities within non-stationary vocal fold vibrations are captured in time variations of the two symmetry factors. These irregularities can be described and visualized by a curve

$$C(t) := C(\, Q(t), R(t)\,) \qquad (6)$$

in the Q(t)—R(t) plane. A curve C(t) that represents symmetric and regular vibrations is close to the point of perfect symmetry (1,1), while a wide dithering of the curve indicates oscillating irregularities. Thus, from this curve C(t) two characteristics can be derived that describe the degree of asymmetries in vocal fold vibrations:

- The distance d_g of the center of gravity of the curve C(t) to the point (1,1) describes the mean degree of oscillation symmetry.
- The radius r_c of a circle that encloses 90 % of C(t) is a measure of the oscillation stability over time.

A rating value R_v for vocal fold oscillation asymmetries is defined by the combination of both criteria:

$$R_v := d_v + r_c\,. \qquad (7)$$

C. *Parameter Optimization of Time-Dependent Two-Mass Model*

In order to derive the curve C(t) a parameter set of the 2MM $S(t) = [Q_l(t)\; Q_r(t)\; R_l(t)\; R_r(t)\; U(t)]$ has to be determined by an optimization procedure. Here, optimization means to find proper values of the parameter set S(t) that adapts the model dynamics $d_a(t)$ to experimental trajectories $d_a^{HGG}(t)$. Due to the time-dependency of non-stationary vibrations, for each sampling point the parameter set S(t) has to be optimized. The optimization of an entire non-stationary trajectory of about one second (number of samples N = 4.000) spans an optimization space of 20.000 parameters. The change of the parameters S(t) is continuous over time.

In the following an algorithm is introduced that takes advantage of the continuous parameter variation and reduces the dimensionality of the solution space. A schematic overview of the algorithm is depicted in Fig. 1. The algorithm divides the trajectories $d_a^{HGG}(t)$ in K blocks. Each block κ represents one period of oscillation with $N\kappa$ samples. The time samples of the parameters and trajectories are indexed as $t_{n,\kappa}$, where κ represents the period and n the time sample within the period κ. The time constants of the parameter set S(t) can be assumed to be beyond the period length. Thus, it is sufficient to optimize just the first $S(t_{0,\kappa})$ and the last $S(t_{N\kappa,\kappa})$ values of the parameter set of one period. The parameter values in between are obtained by linear interpolation. The continuity of the parameters of consecutive periods is ensured by an overlap of one sample, see Fig. 1.

The optimization of the parameter set S(t) starts at period κ=0 of the experimental trajectories $d_a^{HGG}(t)$. Initial values for the parameters $Q_a(t_{n,0})$ with n={0, N_0} are derived by using a simple mass-spring oscillator equation [3]

$$Q_a(t_{0,0}) = Q_a(t_{N0,0}) = 2\pi f_a (m_{0,1a} / k_{0,1a})^{1/2} \quad (8)$$

where f_a is the frequency of period κ=0. The factors $R_a(t)$ and U(t) are set to one. Following initialization the parameter optimization is performed. Therefor, the error, also called objective function,

$$e_\kappa = 1/N_\kappa \sum |d_a(t_{n,\kappa}) - d_a^{HGG}(t_{n,\kappa})|^2, \ \kappa = 0, ..., K\text{-}1 \quad (9)$$

between experimental and theoretical trajectories is minimized. Since the behavior of the 2MM is non-linear and the objective function is non-convex, a stochastic optimization procedure, ASA [8], is used to find the best fit. ASA adjusts the ten optimization parameters so that the objective function e_0 gets minimal.

After the optimization of the first period κ=0 all the consecutive periods are processed. Due to the periods' overlap of one sample, the end state $S(t_{N\kappa-1,\kappa-1})$ of period κ-1 is identical to the start state $S(t_{0,\kappa})$ of period κ. Only one parameter set $S(t_{N\kappa,\kappa})$ at the last time index of the following periods is needed for optimization. Hence, the optimization space has just five dimensions compared to the first period. Finally, a lowpass filter is applied to the elements of S(t) to remove artifacts caused by the period by period processing.

D. Verification and Accuracy of the Optimization

In order to estimate the performance of the proposed optimization algorithm, 242 synthetically non-stationary trajectories with the 2MM where produced. For this, different predefined parameter sets $S^*(t)$ with varying slopes of pitch increase, subglottal pressure levels, and rest positions were generated. These sets were compared to the outcome of the ASA period by period optimization S(t), The objective function is more sensitive to variations of $Q_a(t)$ than to variations of $R_a(t)$. Hence, the factors $Q_a(t)$ match very closely with a relative error of about 2.7%, whereas the factors $R_a(t)$ show a relative error of about 17.9%.

III. RESULTS

The optimization algorithm was applied to the experimental trajectories $d_a^{HGG}(t)$ of 16 subjects. For each subject the curve C(t) (solid line) within the Q(t)—R(t) plane is depicted in Fig. 2. The curves C(t) of the normal voices are located closer to the symmetry point (1,1) and they spread less across the plane than for the pathological cases. The resulting rating value R_v is shown for the normal and for the pathological cases in Fig. 3. The mean rating value of the normal voices is 0.25 and 0.51 for the pathological ones, respectively. Furthermore, the mean values of d_g and r_c are increased for the pathological vocal fold vibrations.

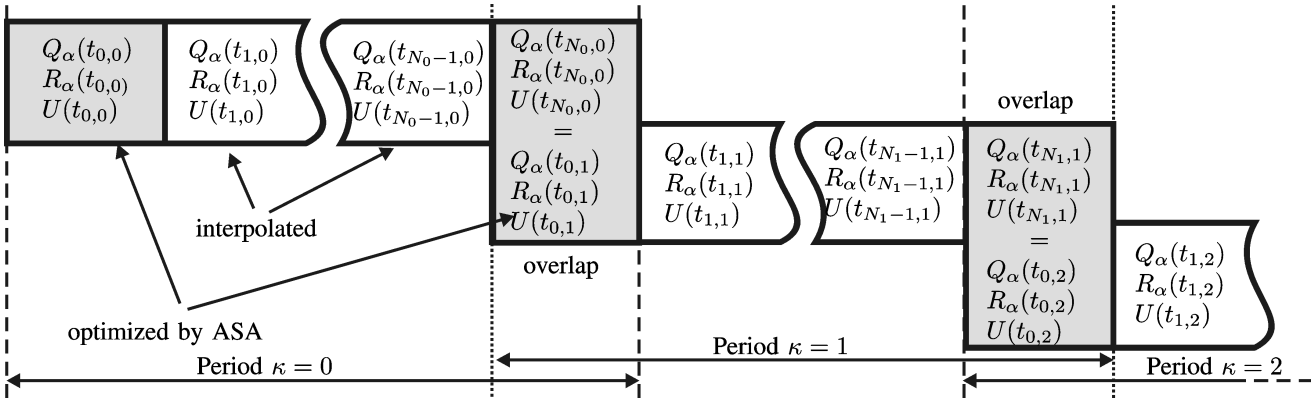

Fig. 1. Schematic overview of the period by period ASA (Adaptive Simulated Annealing [8]) optimization. The periods overlap by one sample. The number of samples in period κ is $N\kappa$. The parameters in gray-filled boxes are optimized, while the others are obtained by linear interpolation.

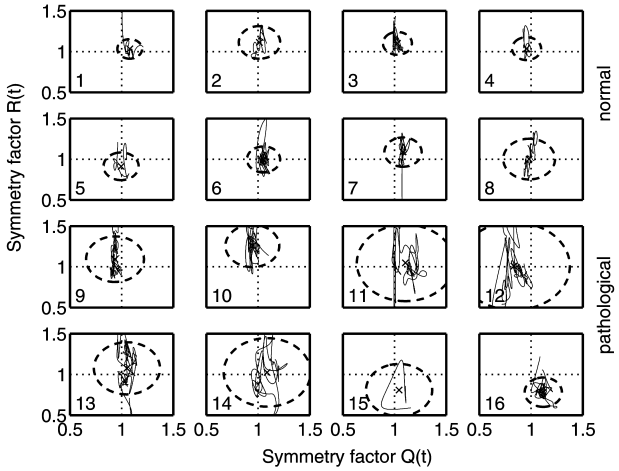

Fig. 2. Plots of the curve C(t) for normal and pathological cases in the Q(t)–R(t) plane. Perfect symmetry is located at the point (1,1). A circle is drawn around the center of gravity for each curve. The radius r_c is determined so that 90% of the curve C(t) lies inside the circle. The subject number is printed in the lower left corner.

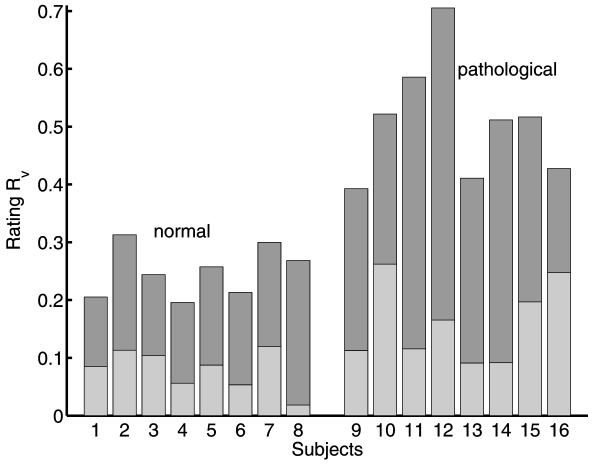

Fig. 3. Rating R_v of the normal and pathological subjects. Small values indicate a symmetrical and regular vibration. The rating is composed out of the distance d_g (light gray) from the center of gravity from the point (1,1) and of the 90% radius r_c (dark gray), see Fig. 2.

IV. Discussion

The adaptation of a time-dependent 2MM to non-stationary vocal fold oscillations enables to derive a two dimensional parameter curve C(t) = C(Q(t), R(t)). From the curve C(t) a rating R_v is calculated that quantifies different degrees of vocal fold asymmetries. Within this rating the mean degree of asymmetry is captured by the distance d_g of the point of gravity to the symmetry point (1,1). The distance d_g can be regarded as counterpart of the asymmetries that are observable in stationary phonation. In contrast, the value r_c describes irregularities resulting from non-stationary vocal fold vibrations. In Fig. 3 a clear distinction between normal and pathological vocal fold vibrations is only possible by the combined evaluation of d_g and r_c. The rating value r_c — that can not be revealed in case of a stationary phonation — makes an important contribution to describe irregularities in non-stationary vocal fold vibrations. The rating R_v can be used as an objective measurement of voice quality in terms of vocal fold oscillation symmetries and regularities. It enables the classification of healthy and pathological voices in non-stationary phonation.

Further investigations will focus on the potential to classify even different kinds of voice disorders, the severity of dysphonia, and to quantify the outcome of voice therapy.

V. Conclusion

Dynamical changes of the vocal folds can be accessed by phonating a pitch increase during an endoscopic high-speed recording. Voice quality in terms of vocal folds' oscillation symmetry and regularity is expressed by a two dimensional parameter curve. This time-dependent curve gives quantitative information on the dynamical state changes of the vocal folds. A classification of vocal fold vibrations into a healthy and a pathological group is possible.

References

[1] U. Hoppe, F. Rosanowski, M. Döllinger, J. Lohscheller, U. Eysholdt, "Visualization of the laryngeal motorics during a glissando," *J. Voice*, vol. 17, pp. 370–376, 2003.

[2] T. Wittenberg, M. Moser, M. Tigges, U. Eysholdt, "Recording, processing and analysis of digital high speed sequences in glottography," *Mach. Vision. Appl.*, vol. 8, pp. 399–404, 1995.

[3] M. Döllinger, U. Hoppe, F. Hettlich, J. Lohscheller, S. Schuberth, U.Eysholdt, "Vibration parameter extraction from endoscopic image series of the vocal folds," *IEEE Trans. Biomed. Eng.*, vol. 49, pp. 773–781, 2002.

[4] R. Schwarz, J. Lohscheller, T. Wurzbacher, U. Eysholdt, U. Hoppe, *Modeling vocal fold vibrations in case of unilatereal vocal fold paralysis*, IASTED Biomed. Eng. Innsbruck, ACTA PRESS, Feb 2004.

[5] U. Eysholdt, M. Tigges, T. Wittenberg, U. Pröschel, "Direct evaluation of high-speed recordings of vocal fold vibrations," *Folia Phoniatr. Logop.*, vol. 48, pp. 163–170, 1996.

[6] K. Ishizaka, J. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal coords," *Bell Syst. Techn. J.*, vol. 51, pp. 1233–1268, 1972.

[7] P. Mergell, I. R. Titze, and H. Herzel, "Irregular vocal fold vibration - high speed observation and modeling," *J. Acoust. Soc. Am.*, vol. 108, pp. 2996–3002, 2000.

[8] L. Ingber. Adaptive Simulated Annealing. [Online]. Available: http://www.ingber.com/