Numerical Modelling of Vocal Fold Dynamics by a 3D Multi-Mass-Model

Anxiong Yang[1], Jörg Lohscheller[1], Michael Stingl[2], Daniel Voigt[1], Ulrich Eysholdt[1], Michael Döllinger[1]

[1] *Medical School Department of Phoniatrics and Pediatric Audiology, University Hospital Erlangen, Bohlenplatz 21, 91054 Erlangen, Germany*

[2] *Department of Applied Mathematics II, University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany*

Abstract

The emitted human voice signal originates due to oscillations of both vocal folds being located in the larynx. So called functional dysphonia being observable during phonation due to disturbing the vocal dynamics yields voice disorders, i.e. hoarseness.

In historical voice research, endoscopic recorded vocal fold dynamics were mostly simulated via 1D Two- or 2D Multi-Mass-Models. However, by using these models only vocal fold movements in lateral and longitudinal directions are taken into account.

This work presents an enhancement of these numerical models: a Multi-Mass-Model extended from 2D to 3D displacements. It consists of five mass planes arranged in vertical direction. Each plane contains five longitudinal coupled mass-spring oscillators. Via the new proposed model not only the vertical movements but also the diffusion of the mucosal wave from inferior to superior can be simulated.

In order to better simulate vocal fold vibrations, the time-invariant model parameters are adjusted in a hybrid optimization procedure. The evaluation of the optimization procedure is performed by using lateral symmetric data sets synthetically generated by the model. Different occurring glottis closure-types are also considered.

1. Introduction

Human voice is generated by oscillations of the two lateral opposing vocal folds [1]. A disturbed or hoarse voice arises from irregular vocal fold vibrations [2]. Mostly, these irregularities are caused by anatomical asymmetries between left and right vocal fold [3]. Asymmetries of the larynx such as unilateral vocal fold polyps or paralysis can be directly observed using endoscopes [1]. However, vocal fold pathology such as functional dysphonia presenting no visual evidences for morphological laryngeal asymmetries is only visible during phonation [1]. The vocal folds oscillate between 100 Hz and 300 Hz. Owing to the vocal fold oscillations in continual rapid motion, endoscopic High-speed (HS) digital imaging applying 2000 – 4000 fps allow to observe laryngeal dynamics, Figure 1 [2]. Investigation of vocal fold dynamics requires a quantitative analysis of HS video data. Two main approaches have been applied to realize the objective evaluation. One approach is to direct quantitatively analyze laryngeal dynamic characteristics like time-varying glottal area, vocal fold length and displacements [4]-[6]. By analyzing the instantaneous frequency and amplitude as well as phase measures the glottal perturbation can be detected and measured [7]-[9].

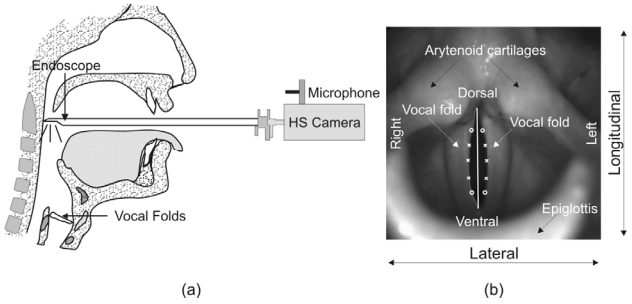

Figure 1: (a) Endoscopic High-speed digital imaging in clinical investigation. The vocal fold dynamics are recorded in real-time. (b) A single frame in open glottis state is shown. The typical extraction points for motion curves (i.e. trajectories) are indicated with "**x**". In the new biomechanical model the trajectories at positions "**o**" are also considered.

Another main approach to analyze endoscopic recorded vocal fold dynamics is the adaption of biomechanical models to vocal fold vibrations extracted from HS video data. By biomechanical models the trajectories at specific positions on the vocal fold edges are taken into account for adaption, Figure 1. Due to the increasing computational power, the technique has been gradually improved:

First, only one lateral trajectory for each vocal fold at medial position was recreated by using a 2-Mass-Model [10], [11]. The corresponding trajectory fit was manually performed [12]. The first fully automatic optimization procedure was realized by Döllinger *et al.* [1] applying the Nelder-Mead algorithm. Normal voices and a functional dysphonia were successfully adapted. To differentiate normal from unilateral vocal fold paralysis classification schemes were proposed [13]. Since several laryngeal diseases occur only at certain frequencies, optimization for non-stationary vocal fold dynamics was performed by Wurzbacher *et al.* [14].

Later, vocal fold movements at three specific positions (marked with "**x**" in Figure 1) distributed along the longitudinal direction were adapted by enhanced 2D Multi-Mass-Models [15]. The corresponding optimization procedure for stationary and non-stationary phonation was successfully accomplished through verification on clinical data sets [15], [16].

So far, low dimensional biomechanical models consider mostly 1D or 2D movements. Optimization only allowed to analyzing superior dynamics of the larynx [15]. Important interactions along the medial parts of the vocal folds as well as the vertical movements were not investigated. However

these topics are essential for quantification of vocal fold 3D dynamics.

2. Methods

2.1 3D Laryngeal Dynamics and Different Glottis Closure-Types

In clinics, endoscopic digital high-speed (HS) imaging is used to observe human vocal fold vibrations during phonation. However, the corresponding observation is performed from top-view position, where only the superior vocal fold dynamics along the longitudinal and lateral directions can be recorded, Figure 1. Nevertheless, under the action of subglottal pressure the vocal fold mucosa movement from inferior to superior is issued in. The movements wavelike diffuse from the lower part to the upper part of the vocal folds [17]. These dynamics look like a wave travelling along the surface (i.e. mucosa). Hence, this event is called mucosal wave [17]. The most critical region of the mucosal wave propagation occurs at the medial surface of vocal folds [18]. Additionally, vocal fold dynamics are nonlinearly combined actions between both vocal folds as well as sub- and supraglottal systems [18], [19].

In order to obtain 3D standard dynamics of human vocal folds, the most realistic model was chosen, i.e. the human hemilarynx model [17], [20], where only one vocal fold was retained to expose and observe the dynamics occurring at the entire surface of the opposite vocal fold (Figure 2) [20], [21], [22]. This setup was chosen as main template for the new 3D-Multi-Mass-Model.

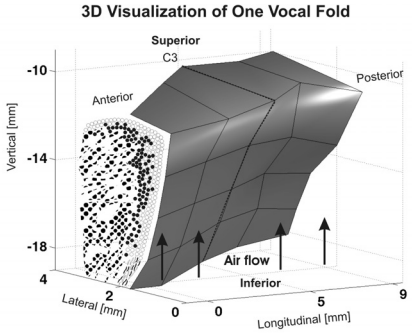

Figure 2: 3D reconstruction and visualization of one vocal fold surface as gathered from hemilarynx experiments **[17]**. 30 black grid points serving as actual positions of tracking markers are homogeneously distributed on the vocal fold surface.

Using endoscopic recording techniques five common glottis closure-types during normal phonation can be observed, which are rectangle, hourglass, triangular-pointed dorsal, triangular-pointed ventral and convex [23]-[25], see Figure 3. These 5 glottis closure-types are also considered within the 3DM.

2.2 3D-Multi-Mass-Model

For modelling of 3D vocal fold dynamics a 3D-Multi-Mass-Model (3DM, Figure 4) based on the models presented by Schwarz et al. [13] and Wurzbacher et al. [14], is used.

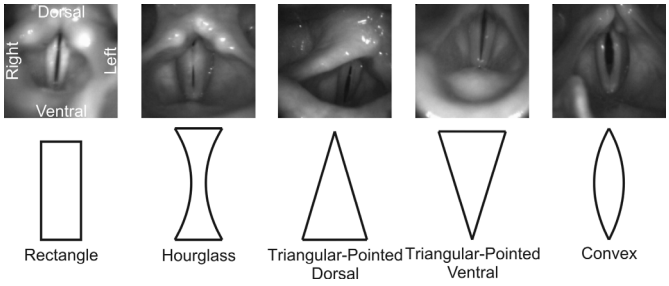

Figure 3: Classification of 5 different glottis closure-types.

Within the 3DM each vocal fold consists of five planes arranged from inferior to superior. Each plane contains five longitudinal coupled mass-spring-oscillators at each side, Figure 4. During phonation the subglottal pressure P^{sub} serves as driving force. It is in compliance with the Bernoulli law [1]. Under the action of subglottal pressure the masses located on both sides (left and right) are set into vibration. The 3DM is a simplified model, so that the viscous losses inside the tissue of the vocal folds are neglected. However, these kinds of models proved to simulate dynamics of the vocal folds during phonation in earlier studies [8], [10], [11].

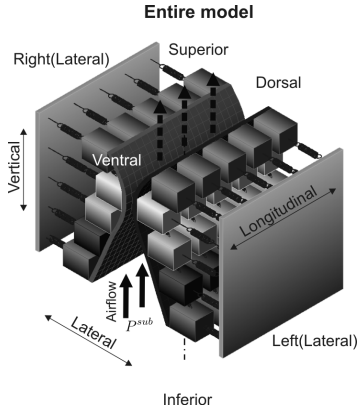

Figure 4: Schematic of the 3D-Multi-Mass-Model. Using anchor springs the masses are elastically connected to a rigid body. Additionally, the masses are connected to each other through springs in vertical and longitudinal directions.

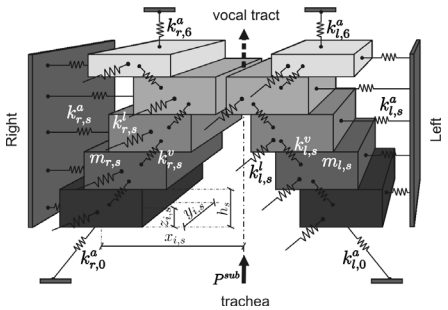

Figure 5: Schematic representation of the model parameters. Stiffness of anchor springs, longitudinal springs and vertical springs are indicated with $k^a_{r,s}$, $k^l_{r,s}$, $k^v_{r,s}$ at right side (i = 1, … ,5) on each plane (s = 1, … ,5 from inferior to superior). Subscript l represents the left side (i = 6, …, 10). $h_{i,s}$ is the thickness of mass elements. The 3D positions of mass elements are denoted with ($x_{i,s}$, $y_{i,s}$, $z_{i,s}$).

The 3DM dynamics are modified by changing model parameters as spring stiffness ($k_{i,s}$), mass ($m_{i,s}$) rest position ($\mathbf{x}_{i,s}$) as well as subglottal pressure (P^{sub}), Figure 5. The initial values are on the basis of the 2-Mass-Model [10] and on the properties of mucosal wave dynamics obtained from Hemilarynx experiments [17], [20].

2.3 Optimization Procedure

In order to adapt the 3DM dynamics to human vocal fold dynamics, an optimization procedure for adjusting the model parameters is essential. Modification can be imposed by expressing the model parameters $m_{i,s}$, $k_{i,s}$, … in terms of the standard model parameters $\tilde{m}_{i,s}$, $\tilde{k}_{i,s}$, … by introducing optimization parameters $Q_{i,s}$, Q_p, Q_r, which respectively influence the constants of spring, mass, subglottal pressure and rest position. They are in accordance with [1], [14] as follows:

$$\begin{aligned} k_{i,s} &= \tilde{k}_{i,s} \cdot Q_{i,s} \quad & m_{i,s} &= \tilde{m}_{i,s} / Q_{i,s} \\ \mathbf{x}_{i,s} &= \tilde{\mathbf{x}}_{i,s} \cdot Q_r \quad & P^{sub} &= \tilde{P}^{sub} \cdot Q_p \end{aligned} \tag{1}$$

Due to 5 planes and 10 mass elements at each plane, there are altogether 50 $Q_{i,s}$ to be adjusted within the 3DM. Using the above representation the vibratory pattern of the vocal fold oscillations can be modified towards symmetric/asymmetric dynamics. By using these optimization parameters the masses, elastic modulus as well as subglottal pressure and rest positions are varied.

As objective function, minimizing the Euclidean distance between trajectories is chosen, i.e. minimize the distances between adapted 3D trajectories $c^{M}_{i,s}[n]$ and experimental 3D trajectories $c_{i,s}[n]$.

The optimization procedure is implemented in each plane, each cross-section and each side of the 3DM. A combination of two optimization algorithms (*Simulated Annealing (SA)* [26], [27] + *Powell Method (Powell)* [28]) is applied. As a global optimization algorithm *SA* is used to seek the region close to the global minimum of the objective function. Then using a local optimization algorithm (*Powell)* the results can be refined.

The similarity of the adapted and synthetic 3D trajectories is judged with two criterions: objective function and correlation coefficient between corresponding 3D trajectories. In this work, two reasonable heuristic stop criterions have been chosen: the value for the objective function must be less than 0.2 with correlation coefficient higher than 0.75.

In order to measure and verify the applicability of the optimization procedure, we used synthetic data of 5 different symmetric glottis closure-types. For each glottis closure-type 5 optimization parameter sets were randomly selected. Therefore, the synthetic 3D trajectories generated with these predefined parameters represent various oscillation patterns being symmetric in lateral direction.

3. Results

Table 1 shows achieved similarities between synthetic generated 3D trajectories and fitted 3D trajectories. The values reveal high accuracy and demonstrate the applicability of the optimization procedure.

Closure-type	Subjects	Accuracy	Correlation	Obj. func.
Rectangle	5	92%±3%	95%±3%	0.04±0.02
Hourglass	5	89%±3%	95%±3%	0.06±0.03
Triangular-pointed dorsal	5	92%±5%	97%±2%	0.04±0.03
Triangular-pointed ventral	5	92%±3%	97%±2%	0.04±0.01
Convex form	5	88%±7%	94%±5%	0.05±0.03

Table 1: Validation with average accuracy regarding the optimization parameter values, correlation coefficient, and value of objective function (*Obj. func.*). The accuracy value indicates the similarity between predefined optimization parameters and model parameters gained by optimization.

4. Discussion

A numerical biomechanical modelling approach for human vocal fold 3D dynamics as well as a corresponding optimization procedure was introduced. Using synthetic data of different symmetric glottis closure-types the optimization procedure was proven to be applicable. In next work, pathological synthetic data will be investigated and adapted. For obtaining human vocal fold 3D dynamics from endoscopic recordings, a laser spots projection system is in development. In future, studying of model parameters of the 3DM will help to quantify voice disorders and to extract physiological laryngeal parameters.

5. Acknowledgements

This work was funded by Deutsche Forschungsgemeinschaft DFG, Grant no. FOR-894/1, Strömungsphysikalische Grundlagen der Menschlichen Stimmgebung.

6. References

[1] M. Doellinger, U. Hoppe, F. Hettlich, J. Lohscheller, S. Schubert, and U. Eysholdt, .Vibration parameter extraction from endoscopic imageseries of the vocal folds, *IEEE Trans. Biomed. Eng.*, vol. 49, pp. 773-781, 2002.

[2] U. Eysholdt, F. Rosanowski, and U. Hoppe, “Vocal fold vibration irregularities caused by different types of laryngeal asymmetry,” *Eur Arch Otorhinolaryngol*, vol. 260(8), pp. 412–417, 2003.

[3] L. Sulica and A. Blitzer, “Vocal fold paresis: evidence and controversies” *Curr Opin Otolaryngol Head Neck Surg*, vol. 15, pp. 159-162, 2007.

[4] Q. Qiu, H. K. Schutte, L. Gu, and Q. Yu, “An automatic method to quantify the vibration properties of human vocal folds via videokymography,” *Folia Phoniatr Logop*, vol. 55(3), pp. 128–136, 2003.

[5] Y. Yan, E. Damrose, and D. Bless, “Functional analysis of voice using simultaneous high-speed imaging and

acoustic recordings," *J Voice*, vol. 21, pp. 604–616, 2007.

[6] T. Braunschweig, J. Flaschke, P. Schelhorn-Neise, and M. Doellinger, "High-speed video analysis of the phonation onset, with an application to the diagnosis of functional dysphonia," *Med Phys Eng*, vol. 30(1), pp. 59-66, 2008.

[7] Y. Yan, K. Ahmad, M. Kunduk, and D. Bless, "Analysis of vocal-fold vibrations from high-speed laryngeal images using a Hilbert transform-based methodology," *J Voice*, vol. 19(2), pp. 161–175, 2005.

[8] I. T. Tokuda, J. Horacek, J. G. Svec, and H. P. Herzel, "Bifurcations and chaos in register transitions of excised larynx experiments,"Chaos vol. 18, 013102, 2008.

[9] M. Doellinger, J. Lohscheller, A. McWhorter, and M. Kunduk, "Variability of Normal Vocal Fold Dynamics for Different Vocal Loading in One Healthy Subject Investigated by Phonovibrograms," *J Voice*, In Press 2008.

[10] K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Syst. Tech. J.*, vol. 51, pp. 1233–1268, 1972.

[11] I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocal fold model," *J. Acoust. Soc. Am.*, vol. 97, pp. 1874–1884, 1995.

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

[13] R. Schwarz, U. Hoppe, M. Schuster, T. Wurzbacher, U. Eysholdt, and J. Lohscheller, "Classification of unilateral vocal fold paralysis by endoscopic digital high-speed recordings and inversion of a biomechanical model," *IEEE Trans. Biomed. Eng.*, vol. 53, pp. 1099–1108, 2006.

[14] T. Wurzbacher, R. Schwarz, M. Doellinger, U. Hoppe, U. Eysholdt, and J. Lohscheller ,"Model based classification of non stationary vocal fold vibrations," *J. Acoust. Soc. Am.*, vol. 120, pp. 1012–1027, 2006.

[15] R. Schwarz, M. Döllinger, T. Wurzbacher, U. Eysholdt, and J. Lohscheller, "Spatio-temporal Quantification of Vocal Fold Vibrations Using High-Speed Videoendoscopy and a Biomechanical Model," *J. Acoust. Soc Am.*, vol. 123, pp. 2717–2732, 2008.

[16] T. Wurzbacher, M. Doellinger, R. Schwarz, U. Hoppe, U. Eysholdt, and J. Lohscheller, "Spatiotemporal Classification of Vocal Fold Dynamics By a Multi Mass Model Comprising Time-Dependent Parameters," *J. Acoust. Soc. Am.*, vol. 123, pp. 2324–2334, 2008.

[17] A. Boessenecker, D. A. Berry, J. Lohscheller, U. Eysholdt, and M. Doellinger "Mucosal Wave Properties of a Human Vocal Fold," *Acta Acustica United with Acustica*, vol. 93(5), 815-823, 2007.

[18] R. Wilhemls-Tricarico, ''Physiological modeling of speech production: methods for modeling soft tissue articulators,'' *J. Acoust. Soc. Am.*, vol. 97, pp. 3085–3098, 1995.

[19] S. Becker, S. Kniesburges, S. Mueller, A. Delgado, G. Link, M. Kaltenbacher, and M. Doellinger, "Flow-structure-acoustic interaction in a human voice model," *J. Acoust. Soc. Am.*, vol. 125(3), pp. 1351–1361, 2009.

[20] M. Doellinger and D. A. Berry, "Visualization and Quantification of the Medial Surface Dynamics of an Excised Human Vocal Fold during Phonation," *J Voice*, vol. 20(3), pp. 401-413, 2006.

[21] M. Doellinger and D. A. Berry, "Computation of the three-dimensional medial surface dynamics of the vocal folds," *J. Biomech.*, vol. 39, pp. 369–374, 2006.

[22] M. Doellinger, D. A. Berry, and D. W. Montequin. "The Influence of Epilarynx Area on Vocal Fold Dynamics," *Otolaryng Head Neck*, vol. 135(5), 724-729, 2006.

[23] O. Rasp, J. Lohscheller, M. Doellinger, U. Eysholdt, and U. Hoppe, "The pitch rise paradigm: A new task for real-time endoscopy of non-stationary phonation," *Folia Phoniatr Logop*, vol. 58, pp. 175-185, 2006.

[24] A. M. Sulter, H. K. Schutte, and D. G. Miller, "Standardized laryngeal videostroboscopic rating: Differences between untrained and trained male and female subjects, and effects of varying sound intensity," *J Voice*, vol. 10(2), pp. 175–189, 1996.

[25] P. Dejonckere, P. Bradley, P. Clemente, G. Cornut, L. Crevier-Buchman, G. Friedrich, P. V. D. Heyning, M. Remacle, V. Woisard, and Committee on Phoniatrics of the European Laryngological Society (ELS), "A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques," *Eur Arch Otorhinolaryngol* 258(2), 77–82, 2001.

[26] L. Ingber and B. Rosen, "Genetic algorithms and very fast simulated annealing," *J. Mathematical Computer Modelling*, 16(11), 87–100, 1992.

[27] L. Ingber, "Adaptive simulated annealing," *J. Control and Cybernetics*, 25(1), 33–54, 1996.

[28] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes in C," *Cambridge University Press*, 1995.