Published in final edited form as:

Biomech Model Mechanobiol. 2015 January ; 14(1): 169–184. doi:10.1007/s10237-014-0597-1.

Development of a time-dependent numerical model for the assessment of non-stationary pharyngoesophageal tissue vibrations after total laryngectomy

Björn Hüttner,

Department of Phoniatics and Pediatric Audiology, Medical School, University Hospital Erlangen, Bohlenplatz 21, 91054 Erlangen, Germany

Georg Luegmair,

Department of Phoniatics and Pediatric Audiology, Medical School, University Hospital Erlangen, Bohlenplatz 21, 91054 Erlangen, Germany

Rita R. Patel,

Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue, 47405 Bloomington, USA

Anke Ziethe,

Department of Phoniatics and Pediatric Audiology, Medical School, University Hospital Erlangen, Bohlenplatz 21, 91054 Erlangen, Germany

Ulrich Eysholdt,

Department of Phoniatics and Pediatric Audiology, Medical School, University Hospital Erlangen, Bohlenplatz 21, 91054 Erlangen, Germany

Christopher Bohr,

Department of Otorhinolaryngology Head and Neck Surgery, Medical School, University Hospital Erlangen, Waldstrasse 1, 91054 Erlangen, Germany

Irina Sebova,

Department of Phoniatics and Pediatric Audiology, Medical School, University Hospital Erlangen, Bohlenplatz 21, 91054 Erlangen, Germany

Marion Semmler,

Department of Phoniatics and Pediatric Audiology, Medical School, University Hospital Erlangen, Bohlenplatz 21, 91054 Erlangen, Germany

Michael Döllinger

Department of Phoniatics and Pediatric Audiology, Medical School, University Hospital Erlangen, Bohlenplatz 21, 91054 Erlangen, Germany

Abstract

Laryngeal cancer due to, e.g., extensive smoking and/or alcohol consumption can necessitate the excision of the entire larynx. After such a total laryngectomy, the voice generating structures are

lost and with that the quality of life of the concerning patients is drastically reduced. However, the vibrations of the remaining tissue in the so called pharyngoesophageal (PE) segment can be applied as alternative sound generator. Tissue, scar, and geometric aspects of the PE-segment determine the postoperative substitute voice characteristic, being highly important for the future live of the patient. So far, PE-dynamics are simulated by a biomechanical model which is restricted to stationary vibrations, i.e., variations in pitch and amplitude cannot be handled. In order to investigate the dynamical range of PE-vibrations, knowledge about the temporal processes during substitute voice production is of crucial interest. Thus, time-dependent model parameters are suggested in order to quantify non-stationary PE-vibrations and drawing conclusions on the temporal characteristics of tissue stiffness, oscillating mass, pressure, and geometric distributions within the PE-segment. To adapt the numerical model to the PE-vibrations, an automatic, block-based optimization procedure is applied, comprising a combined global and local optimization approach. The suggested optimization procedure is validated with 75 synthetic data sets, simulating non-stationary oscillations of differently shaped PE-segments. The application to four high-speed recordings is shown and discussed. The correlation between model and PE-dynamics is 97 %.

Keywords

Substitute voice; PE-segment; Multi-mass model; High-speed recordings; Non-stationary optimization

1 Introduction

One of the most important instruments of human interaction is voiced communication (Titze 2000). Its information carrier is the acoustic signal that is generated by two oscillating vocal folds which are located in the larynx. A healthy voice is characterized by regular and symmetric vibration patterns (Döllinger et al. 2002). Laryngeal cancer due to, e.g., excessive smoking and/or alcohol consumption (Papadas et al. 2010) disturbs the generation of the acoustic signal and often requires the removal of the complete larynx (laryngectomy) and therewith of the two vocal folds (Shah 2001). Figure 1a shows the postsurgical situation: The esophagus and pharynx are directly connected, at which the mucosa tissue in the changeover is called pharyngoesophageal (PE) segment. The volume in the changeover is denoted as pseudoglottis. The trachea is sutured into the frontal neck, where a hole (tracheostoma) preserves breathing. By removing the vocal folds, the natural source for voiced communication is lost. However, tissue vibrations in the PE-segment can be applied as alternative sound generator. For this purpose, the trachea and the esophagus are connected by a shunt so that closing the tracheostoma during exhalation forces the air to stream through the shunt and to pass the PE-segment where the surrounding tissue is excited to oscillate what generates a substitute sound signal (tracheo-esophageal voice) (Singer et al. 1983). However, patients often suffer from a reduced quality of the substitute voice (Schuster et al. 2003). Studies showed that anatomic and morphologic characteristics of the PE-segment, i.e., shape and tissue stiffness are crucial to the quality of the substitute voice, and thus for the intelligibility (van As et al. 1999). Correlations between the geometry and the dynamics of the PE-segment during the substitute voice production and the quality of the

resulting acoustic signal have not been identified yet. Since the viscoelastic properties of the oscillating tissue might be manipulated during the surgical intervention, these interrelations are of great interest with regard to voice rehabilitation. In a first step, it is necessary to quantify the tissue vibrations in the PE-segment and investigate their relation to the viscoelastic tissue characteristics. Then, correlations to the quality of the substitute voice can be determined. The gained knowledge might be incorporated into the pharyngeal reconstructive surgery during the laryngectomy, e.g., by purposely shaping the pseudoglottis or by constructively modifying the viscoelastic tissue properties.

A convenient method for the visual assessment of PE-vibrations is high-speed (HS) endoscopy (Schuster et al. 2005) (see Fig. 1). This method allows to investigate oscillations in real time and therefore to analyze irregular vibration patterns (Lohscheller et al. 2004). In order to represent the conditions during spontaneous speech realistically and enable an extensive understanding of the biomechanical details of the alternative voice production, a non-stationary phonation paradigm, e.g., the modulation of pitch, is used (Rasp et al. 2006).

Commonly, the recorded PE-vibrations are subjectively evaluated, implicating disadvantages concerning the duration of the data analysis and the reproduction of the findings (Döllinger 2009). Since the biomechanical properties of the tissue in the vocal folds or in the PE-segment cannot be assessed in such an evaluation, numerical models, and simulations including optimization procedures have already been applied successfully (Schwarz et al. 2011). Yang et al. (2012) showed that the optimized model parameters are able to uncover the biomechanical properties of vocal fold tissue, i.e., the distribution of the local effective mass and stiffness as well as the pressure, in a quite good manner. The basic myoelastic principles of tracheo-esophageal and laryngeal voice production are identical (Moon and Weinberg 1987), which facilitates the adaptation of multi-mass models that were originally invented for the vocal folds. A first approach for an objective description of PE-vibrations was given by Schwarz et al. (2011) who applied a biomechanical multi-mass model (PE-MMM) to simulate the oscillations of the PE-segment during a sustained phonation. The dynamics of the PE-MMM are adapted to the PE-dynamics by means of an automatic optimization procedure, allowing for a quantitative analysis. However, the PE-MMM is limited to stationary PE-oscillations, i.e., vibrations with near constant fundamental frequency and oscillation amplitude (e.g., phonation of a sustained vowel) (Schwarz et al. 2011).

In this work, an extension of the PE-MMM by time-dependent model parameters is suggested (PEM(t)) in order to analyze and quantify non-stationary tissue vibrations in the PE-segment. Dynamics of the PEM(t) are adapted to PE-dynamics that were extracted from HS recordings. This is performed by an automatic optimization procedure that incorporates a global and a local optimization algorithm. A block-based optimization is executed, similar to that applied for the time-dependent vocal fold model (TMM) (Wurzbacher et al. 2008), allowing for handling variations in amplitude and frequency. The variations in time of the optimized model parameters objectively describe the temporal alterations of the pseudoglottal pressure, the oscillating mass, and the tissue stiffness that occur in the PE-segment during non-stationary phonation. The proposed optimization procedure is validated with synthetic data sets. The error between the predefined model parameters and the

optimized model parameters is a measure for the quality of the optimization procedure. To demonstrate the capability of the PEM(t) of handling different shaped PE-segments, synthetic data is generated with circular, elliptic, and triangular shaped model contours. The potential to simulate the dynamics of real PE-vibrations is demonstrated for HS recordings of four male laryngectomees that perform a non-stationary phonation, i.e., variations in pitch and amplitude.

2 Methods

2.1 Endoscopic high-speed recordings

HS recordings of the PE-dynamics of four laryngectomized males (R1 to R4, 59 ± 3 years old) are taken. All patients underwent concomitant larynx removal and neck dissection. Postoperatively, R1 and R4 received radiotherapy while R2 and R3 received radiochemotherapy. Within the range of individual variations, all patients display the same precondition. The subjects were asked to phonate the vowel /a/ with increasing pitch. R1 was recorded at the College of Health Sciences at the University of Kentucky (USA) and R2 to R4 at the Department of Phoniatics and Pediatric Audiology at the University Hospital Erlangen (Germany). The vibrations of the PE-segments are observed from a top view with rigid 70° and 90° endoscopes for R1 and R2 to R4, respectively. The applied HS cameras are a Photron FASTCAM MC2 (512×512 pixels, 4000 fps) for R1 and a Wolf HRES ENDOCAM 5562 (256×256 pixels, 4000 fps) for R2 to R4. A sketch-up of the recording situation can be seen in Fig. 1a.

The PE-dynamics are extracted from the HS recordings by means of a knowledge-based image processing algorithm using active contours (Lohscheller et al. 2003, 2004; Stiglmayr et al. 2008). They are described by the area function $a^{\text{PE}}[n]$, i.e., the time signal of the pseudoglottal area and the segmented contours $c^{\text{PE}}[z, n]$. Each contour in frame $n = 1, \dots, N$ is parameterized by $z \in [0, 1[$ (Schwarz et al. 2011).

2.2 Time-dependent multi-mass model: PEM(t)

For the analysis of the extracted time signals of the pseudoglottal opening $a^{\text{PE}}[n]$ and the contours $c^{\text{PE}}[z, n]$, a time-dependent multi-mass model (PEM(t)) is applied. It consists of eight horizontally coupled mass-spring oscillator-elements that are arranged in the transversal plane. Sketches of the entire model and a close up of a mass-spring oscillator-element can be seen in Fig. 2. The model allows for two-dimensional oscillations in the transversal plane. Vertical movements are disregarded. The masses $m_{i,s}$ are arranged in a closed shape with the index s indicating the lower ($s = 1$) and the upper plane ($s = 2$), respectively. The index i ($i = 1, \dots, 8$) gives the position along the contour. The physics of the PEM(t) are similar to the PE-MMM, i.e., the stationary multi-mass model for the PE-segment (Schwarz et al. 2011), but are expanded for time-dependent model parameters. These are the pseudoglottal pressure $P^s[n]$, the masses $m_{i,s}[n]$, the anchor, longitudinal, vertical, and coupling springs $k_{i,s}^a[n]$, $k_{i,s}^l[n]$, $k_i^v[n]$, $k_i^c[n]$ and the corresponding dampers $r_{i,s}^a[n]$, $r_{i,s}^l[n]$, and $r_i^v[n]$. Hence, variations of frequency and amplitude of $a^{\text{PE}}[n]$ and $c^{\text{PE}}[z, n]$ can be simulated. These model parameters have been used in previous works to support time

dependency for both two-mass and multi-mass models for the vocal folds (Lucero and Koenig 2005; Wurzbacher et al. 2006, 2008).

Because of the non-stationary masses, the equation of motion as known from the PE-MMM (Schwarz et al. 2011) has to be expanded to incorporate the time derivative of mass-element $m_{i,s}$ (Wurzbacher et al. 2008):

$$m_{i,s}[n] \cdot \ddot{\mathbf{x}}_{i,s}[n] = -\dot{m}_{i,s}[n] \cdot \dot{\mathbf{x}}_{i,s}[n] + \mathbf{F}_{i,s}^a[n] + \bar{\mathbf{F}}_{i,s}^a[n] + \mathbf{F}_{i,s}^v[n] + \mathbf{F}_{i,s}^l[n] + \mathbf{F}_{i,s}^c[n] + \mathbf{F}_{i,s}^d[n]. \quad (1)$$

In general, the forces are the same as for the stationary PE-MMM (Schwarz et al. 2011):

- $\mathbf{F}_{i,s}^a[n]$ acts due to the anchor springs ($k_{i,s}^a[n]$) and dampers ($r_{i,s}^a[n]$).
- $\bar{\mathbf{F}}_{i,s}^a[n]$ is an imaginary anchor force. Because of the missing symmetry of the PEM(t), circular vibration modes of the masses may occur for large oscillation amplitudes. To avoid this, additional anchors $\bar{a}_i[n]$, anchor springs $\bar{k}_{i,s}^a[n]$, and dampers $\bar{r}_{i,s}^a[n]$ are introduced. They lie in the transversal plane and stand perpendicular to the original ones. They are just for the purpose of computation and have no physiological background (see Fig. 2b).
- $\mathbf{F}_{i,s}^v[n]$ acts due to the connection of the upper and lower masses by springs ($k_i^v[n]$).
- $\mathbf{F}_{i,s}^l[n]$ originates from the connection of adjacent masses by springs ($k_{i,s}^l[n]$) and dampers ($r_{i,s}^l[n]$).
- During the collision of two masses or a mass and a longitudinal connection spring, a restoring force $\mathbf{F}_{i,s}^c[n]$ is generated which is realized by an additional spring $k_{i,s}^c[n]$.
- $\mathbf{F}_{i,s}^d[n]$ is the force due to the volume flow through the pseudoglottis and is created by the pseudoglottal pressure $P[n]$.

The position of the mass $m_{i,s}$ at time step n is denoted as $\mathbf{x}_{i,s}[n] = (x_{i,s}[n], y_{i,s}[n])^T$. The model dynamics are defined by the smallest model contour, i.e., the polygon formed by the mass positions ($\mathbf{x}_1[n], \dots, \mathbf{x}_8[n]$), and the opening area $a^{\text{PEM}(t)}[n]$ that is defined as the polygon area. We declare $\mathbf{x}_l[n] = \mathbf{x}_{i,s}[n]$ and $a^{\text{PEM}(t)}[n] = a_s^{\text{PEM}(t)}[n]$ with $s = 1$ if $a_1^{\text{PEM}(t)}[n] < a_2^{\text{PEM}(t)}[n]$ and $s = 2$ otherwise. The model contours are defined as $c^{\text{PEM}(t)}(z, n)$ with $z = (\mathbf{x}_1[n], \dots, \mathbf{x}_8[n])^T$.

$\mathbf{x}_{i,s}^r[n]$ and $\mathbf{x}_{i,s}^{r,a}[n]$ are the rest positions of mass $m_{i,s}$ and the corresponding anchor a_i and are dependent of time, too. By adapting $\mathbf{x}_{i,s}^r[n]$ and $\mathbf{x}_{i,s}^{r,a}[n]$, PE-shapes as observed in HS

recordings (van As et al. 1999), and thus differing from the circular arrangement in Fig. 2a, can be realized. $\dot{\mathbf{x}}_{i,s}[n]$ and $\ddot{\mathbf{x}}_{i,s}[n]$ [see Eq. (1)] are the velocity and acceleration of mass $m_{i,s}$ at time step n . The mass trajectories $\mathbf{x}_{i,s}[n]$ are gained from Eq. (1) by numerical integration by a fourth-order Runge-Kutta algorithm with step size 0.25 ms (Döllinger et al. 2002; Yang et al. 2011).

2.3 Automatic parameter optimization

In order to simulate non-stationary PE-vibrations, the dynamics of the PEM(t) have to be adapted to the dynamics as extracted from the HS recordings. For that, the parameters of the PEM(t) are varied in order to approximate the oscillations of the PE-segment. In a first step, the PEM(t) is initialized by the extracted PE-contour. Afterward, appropriate model parameters are determined by an automatic optimization procedure. In the following, the different steps of the optimization are introduced:

2.3.1 Block-based optimization—The time-dependent optimization is handled in blocks, similar to Wurzbacher et al. for the TMM (Wurzbacher et al. 2008). The pseudoglottal area signal $a^{\text{PE}}[n]$ and contours $c^{\text{PE}}[z, n]$ are divided into $b = 1, \dots, B$ blocks, each containing four oscillation cycles starting with a maximum opened pseudoglottis. To assure smoothness, consecutive blocks overlap by 50 %. The time steps within block b are denoted as $n_b = 1, \dots, N_b$ and the signals of area and contour as $a_b^{\text{PE}}[n_b]$ and $c_b^{\text{PE}}[z, n_b]$, respectively. The adaptation of the model dynamics to the PE-dynamics is executed successively for each block beginning with the first one.

2.3.2 Model initialization—To reproduce different model shapes, the masses' initial rest positions $\hat{\mathbf{x}}_{i,s}^r$ differ from the circular arrangement as shown in Fig. 2a. In a first step, the masses are placed along the PE-contour $c_1^{\text{PE}}[z, n_1^*]$ with n_1^* being the time step of the maximum PE-opening within the first block:

$$n_1^* = \underset{k \in n_1}{\operatorname{argmax}}(a_1^{\text{PE}}[k]). \quad (2)$$

The arrangement $\mathbf{x}_{i,s}^{r,*}(i = 1, \dots, 8)$ along the contour $c_1^{\text{PE}}[z, n_1^*]$ is determined by two conditions:

- The difference between the area $a_1^{\text{PE}}[n_1^*]$ and the area of the polygon spanned by the rest positions is minimized.
- The distance between adjacent rest positions on the contour is maximized, i.e., equidistant rest positions.

The determined positions are radially projected onto the minimum PE-opening contour $c_1^{\text{PE}}[z, n_1^{*,*}]$

$$n_1^{**} = \underset{k \in n_1}{\operatorname{argmin}}(a_1^{\text{PE}}[k]), \quad (3)$$

where they define the positions $\mathbf{x}_{i,s}^{**}$ ($i = 1, \dots, 8$). The masses' initial rest positions $\hat{\mathbf{x}}_{i,s}^r$ are determined by the following condition:

$$\hat{\mathbf{x}}_{i,s}^r = \mathbf{x}_{i,s}^{**} + 0.375 \cdot (\mathbf{x}_{i,s}^r - \mathbf{x}_{i,s}^{**}). \quad (4)$$

In order to prevent an overestimation of $\hat{\mathbf{x}}_{i,s}^r$ due to collisions of the masses during closure, the factor 0.375 is applied instead of the mean distance between the minimum and maximum opening (Wurzbacher et al. 2008).

Finally, the rest positions of the upper plane are radially shifted by factor 0.05 toward the center of mass of the contour in order to start with a convergent pseudoglottal profile (Schwarz 2007). The anchors are placed in a radial distance of 1.2 cm from the masses' initial positions (Schwarz 2007).

2.3.3 Time-dependent scaling—Time-dependent model dynamics are achieved by modifying the constant initialization values $\hat{m}_{i,s}, \hat{k}_{i,s}^a, \hat{k}_i^v, \hat{P}^s$ (see Table 1) by time-dependent parameters $Q_1[n], \dots, Q_8[n]$, and $Q_P[n]$ (Wurzbacher et al. 2006, 2008; Schwarz et al. 2008, 2011):

$$m_{i,s}[n] = \hat{m}_{i,s} / Q_i[n], \quad (5)$$

$$k_{i,s}^a[n] = \hat{k}_{i,s}^a \cdot Q_i[n], \quad (6)$$

$$k_i^v[n] = \hat{k}_i^v \cdot Q_i[n], \quad (7)$$

$$P^s[n] = \hat{P}^s \cdot Q_P[n]. \quad (8)$$

The masses and anchor springs of the oscillator-element i are adapted by the same value Q_i . The longitudinal spring stiffness $k_{i,s}^l[n]$ and the coupling spring $k_{i,s}^c[n]$ are not varied by a Q-factor as they are dependent of $k_{i,s}^a$, and thus are dependent of time anyway (Schwarz et al. 2008). The nine Q-factors are summarized in the parameter set

$$P[n] = \{Q_1[n], \dots, Q_8[n], Q_P[n]\} \quad (9)$$

for each time step.

The applied initialization values for the masses $\hat{m}_{i,s}$, anchor springs $\hat{k}_{i,s}^a$, and vertical springs $\hat{k}_{i,s}^v$ are derived from those of Ishizaka and Flanagan divided by the number of masses (Ishizaka and Flanagan 1972; Schwarz et al. 2011). However, first applications showed that the necessary pseudoglottal pressure was beyond physiological condition. To correct this, the initialization values for the masses and springs are reduced by 30%. The pseudoglottal pressure $\hat{P}^s$ is initialized with 25 cmH₂O, see Table 1. The remaining model parameters are adopted from the PE-MMM (Schwarz et al. 2011).

To account for time-variant amplitudes of the PE-oscillations, the rest positions for the masses and anchors are heuristically modified over time. Therefore, the constant initialization values $\hat{\mathbf{x}}_{i,s}^r$ and $\hat{\mathbf{x}}_{i,s}^{r,a}$ are modified by a time-dependent parameter:

$$\mathbf{x}_{i,s}^r[n] = q_i^r[n] \cdot \hat{\mathbf{x}}_{i,s}^r, \quad (10)$$

$$\mathbf{x}_{i,s}^{r,a}[n] = q_i^{r,a}[n] \cdot \hat{\mathbf{x}}_{i,s}^{r,a}. \quad (11)$$

For each block b , an individual rest position $\mathbf{x}_{i,s,b}^r$ is determined after sec. 2.3.2 and replacing n_1 by $n_b \cdot q_i^{r,b}$ is the factor that radially projects $\hat{\mathbf{x}}_{i,s}^r$ onto $\mathbf{x}_{i,s,b}^r$. $q_i^{r,b}$ is obtained by linear interpolation between $q_i^{r,b}$ and $q_i^{r,b+1}$ for each block b and all time steps n .

2.3.4 Objective function—To adapt the model dynamics to those of the PE-segment, an adequate set of optimization parameters $P[n]$ has to be found for every time step n . The adaptation quality is quantified by the objective function Γ that measures the error between the model dynamics and those of the PE-segment. The applied objective function Γ is a combination of three criteria that describe different aspects of the vibration:

1. *Area* The difference signal between $a^{\text{PE}}[n]$ and $a^{\text{PEM}(t)}[n]$ normalized to the area signal of $a^{\text{PE}}[n]$.

$$\Delta a = \frac{\sum_{n=1}^N (a^{\text{PEM}(t)}[n] - a^{\text{PE}}[n])^2}{\sum_{n=1}^N (a^{\text{PE}}[n])^2}. \quad (12)$$

2. *Intersection* The intersection $a^{\text{sec}}[n]$ between $a^{\text{PE}}[n]$ and $a^{\text{PEM}(t)}[n]$ normalized to the sum of the area signals.

$$\Delta s = \frac{\sum_{n=1}^N (|a^{\text{PEM}(t)}[n] - a^{\text{sec}}[n]| + |a^{\text{PE}}[n] - a^{\text{sec}}[n]|)^2}{\sum_{n=1}^N (a^{\text{PEM}(t)}[n] + a^{\text{PE}}[n])^2}. \quad (13)$$

3. *Contour distance* The distance $d_i[n]$ of mass $m_i[n]$ to the PE-contour normalized to the radius $r[n]$ of a circle with the same area as the PE-segment in frame n .

$$d := \frac{\sum_{n=1}^N \left(\frac{1}{8} \sum_{i=1}^8 d_i[n] \right)^2}{\sum_{n=1}^N (r[n])^2} \quad \text{with} \quad r[n] = \sqrt{\frac{a^{\text{PE}}[n]}{\pi}}. \quad (14)$$

The criteria (1) to (3) are inherited from the PE-MMM (Schwarz et al. 2011). In the following, Eqs. (12–14) are combined in the objective function

$$\Gamma := \Delta a + \Delta s + d \quad (15)$$

that approximates the area (Eqs. (12), (13)) and the shape (Eqs. (13), (14)) of the PEM(t) to the PE-segment.

2.3.5 Initialization of optimization parameters—For solving the optimization problem, the objective function Γ has to be minimized. In a first step, an appropriate start parameter set $P^{\text{start}} = \{Q_1^{\text{start}}, \dots, Q_8^{\text{start}}, Q_P^{\text{start}}\}$ has to be found. The initialization of the parameters Q_i is based on the mass-spring oscillator system and the oscillation frequency f^{PE} of the pseudoglottal opening $a_1^{\text{PE}}[n]$ of the first block (Wurzbacher et al. 2008):

$$Q_i^{\text{start}} = 2\pi f^{\text{PE}} \sqrt{\frac{\hat{m}_{i,s}}{\hat{k}_{i,s}}}, \quad \text{for } i = 1, \dots, 8. \quad (16)$$

The initial value for Q_P is found by setting Q_i after Eq. (16) and varying Q_P from low to high values, such that the resulting pressure is 12 cmH₂O P_s 63 cmH₂O with step size $\Delta P^s = 1$ cmH₂O. Q_P^{start} best approximates $a^{\text{PEM}(t)}[n]$ to the area function $a^{\text{PE}}[n]$ within the first block, see Eq. (12).

2.3.6 Optimization procedure—For the minimization of the objective function Γ , an adequate optimization parameter set $P[n]$ is required that fits the model dynamics to the PE-dynamics. Because of the block-based optimization, the objective function Γ_b is minimized successively for each block, beginning with the first one. The optimization of the first block is initialized with P^{start} (see Sect. 2.3.5). Because of the 50% overlap, the first and last optimization parameter set $P_b[1]$ and $P_b[N_b]$ of block b are initialized with the parameter set at half length of the previous block ($P_{b-1}[N_{b-1}/2]$) (Wurzbacher et al. 2008).

Since the optimization problem is non-convex (Döllinger et al. 2002), an alternation of a global and a local optimization algorithm is applied to the problem of minimizing Γ_b . Loops of *Simulated Annealing* (SA) (Ingber 1996) and *Powell's direction set method* (PDSM) (Brent 1973) are executed up to five times or until the improvement of Γ is below an empirical break off of $\Gamma = 0.0001$. The SA algorithm randomly runs through the parameter space and is thus used to find preliminary parameter sets $P_b[n_b]$. With these, a local search by PDSM is applied to find parameter sets closer to the global minimum. A schematic of the optimization procedure for block b is depicted in Fig. 3a. The minimization of the objective function by SA and PDSM is initialized with the current optimization quality Γ_b , the first

and last parameter set $P_b[1]$ and $P_b[N_b]$ of the actual block and the PE-dynamics $a_b^{\text{PE}}[n_b]$, $c_b^{\text{PE}}[z, n_b]$ (see Fig. 3b). At first, the optimization parameter sets are generated for all time steps n_b by linear interpolation between $P_b[1]$ and $P_b[N_b]$. Afterward, the PEM(t) is scaled with the parameter sets $P_b[n_b]$ and the dynamics are computed. The model dynamics and PE-dynamics are then evaluated by the objective function according to Eq. (15). Only the last parameter set $P_b[N_b]$ is modified by SA or PDSM. The first parameter set $P_b[1]$, being the result of the previous optimized block, is kept constant. When optimizing the first block, the first and last parameter set $P_1[1]$ and $P_1[N_1]$ are chosen to be modified. The minimization of the objective function is iterated until a stop criterion is achieved. More precisely, the SA and PDSM algorithms both stop after 600 iterations or if $\epsilon < 0.0001$, i.e., the difference between the objective functions in two consecutive iteration steps.

2.4 Validation and application

2.4.1 Validation with synthetic data sets—The capability and reliability of the proposed optimization procedure of finding appropriate optimization parameter sets $P[n]$ are tested on synthetic data sets. For this, dynamics of the PEM(t) are generated with predefined parameter sets $\tilde{P}[n]$ with $N = 1,000$ time steps. Herein, a time step n has a duration of $t = 0.25$ ms. To simulate a wide variety of PE-vibrations, the parameter sets $\tilde{P}[n]$ are generated randomly under the following assumptions:

- The parameters $\tilde{Q}_i[n]$ ($i = 1, \dots, 8$) and $\tilde{Q}_P[n]$ present a linear decrease or increase between the samples $n = 1$ and $n = N$. This simulates an increase/decrease of frequency and amplitude of the PE-dynamics and demonstrates the capability of the PEM(t) to handle both.
- The parameters of the first and last time step are randomly chosen from predefined parameter spaces that allow 12 cm H₂O P_s 63 cm H₂O for the pseudoglottal pressure (Deshmane et al. 1995; Takeshita et al. 2013). For the frequency of the area signal, the condition 60 Hz f 180 Hz (Lundström et al. 2008) has to be fulfilled. To avoid chaotic vibrations, the oscillation frequencies of the eight mass-spring-elements (see Eq. (16)) are supposed to differ in a range of at most ± 10 Hz.

To consider variable PE-shapes, the PEM(t) is configured with three different model configurations, namely circular, elliptic, and triangular opening shapes (Schuster et al. 2005). Furthermore, to account for irregular PE-shapes, the initial model contours are varied by multiplying the initial rest position of the masses m_i by a random factor $r \in [0.8, 1.2]$.

For the circular, the elliptic, and the triangular model shape, 25 parameter sets $\tilde{P}[n]$ are generated in each case, i.e., 75 runs in total. The resulting dynamics are then remodeled by finding appropriate optimization parameter sets $P^*[n]$ by means of the presented optimization procedure. The optimized parameter sets $P^*[n]$ are then compared to the predefined parameter sets $\tilde{P}[n]$ by computing the mean relative error:

$$\bar{\epsilon}_{rel} = \frac{1}{N} \sum_{n=1}^N \frac{|\tilde{P}[n] - P^*[n]|}{\tilde{P}[n]}. \quad (17)$$

2.4.2 Application to real PE-vibrations—The capability of the suggested optimization procedure to adapt the dynamics of the PEM(t) to those of real PE-vibrations is demonstrated for four HS recordings which show variations in pitch and amplitude. The PE-dynamics are extracted from the HS recordings and are modeled by the PEM(t) and the introduced procedure.

3 Results

3.1 Validation with synthetic data sets

The mean relative errors [Eq. (17)] and the appendant standard deviations are listed in Table 2. The results are separated for the three model configurations and the averaged error over all three configurations. The errors for $\bar{Q}_i = \text{mean}(Q_1, \dots, Q_8)$ and Q_P are given as well as the configurations' overall errors $\bar{Q} = \text{mean}(Q_P, Q_1, \dots, Q_8)$. The grand error over all parameters and model configurations amounts to $3.2 \pm 3.8\%$. The errors for $\bar{Q}_i$ are smaller than the errors for Q_P for all configurations, i.e., the masses and springs are better adopted than the pressure. The best optimization is achieved for the triangular model configuration with an error of $2.9 \pm 3\%$ for $\bar{Q}$ followed by the elliptic one ($3.0 \pm 3.8\%$). With an error of $3.5 \pm 4.2\%$, the circular configuration presents the worst optimization results. Concerning the overall errors $\bar{Q}$, the standard deviations are larger than the errors themselves.

Figure 4 visualizes the achieved adaptation results, represented by the objective function Γ in dependence of the mean relative errors. The results are separated for triangular (triangles), elliptic (squares), and circular (circles) model contours. Here, Γ was computed for the complete dynamics, i.e., over all time steps after the optimization. ϵ_{rel} corresponds to $\bar{Q}$ of Table 2. Γ takes values between 0.0022 and 0.0329 and ϵ_{rel} between 1.51 and 8.0 %.

3.2 Application to real PE-vibrations

Non-stationary phonations of four male laryngectomized subjects were recorded by a HS-camera. The PE-dynamics were extracted from the HS recordings R1 to R4 and were simulated by the PEM(t). The optimization results are summarized in Table 3. The value Γ represents the results of the objective function while evaluating the dynamics of PE-segment and PEM(t) for all time steps. Γ takes values between 0.12 and 0.27. Γ_a , Γ_s and Γ_d are the objective functions for the area, the intersection and the contour distance criteria that in sum result in Γ [see Eq. (15)]. Clearly, the adaptation of the area functions (Γ_a) is the best, followed by the intersection area criterion (Γ_s). The distance of the masses to the PE-contour (Γ_d) presents the worst results. κ describes the correlation between the area functions of the PE-segment and the PEM(t). The value is at least 97 % with a mean of 98 %. The high correlations are in good agreement with the small values for Γ_a .

Figures 5, 6, 7 and 8 display the adaptation results for R1 to R4. The subfigures (a) depict the area functions $a^{\text{PE}}[n]$ and $a^{\text{PEM}(t)}[n]$ of the PE-segment (red solid line) and the PEM(t) (black dashed line), respectively, for a short sequence of non-stationary phonation. The frequency development over time for the PE- and the PEM(t)-oscillations, determined for the area signals, is visualized in subfigures (b). Subfigures (c) show the extracted PE-segment (red area and black solid line) and the PEM(t) contour (dashed line) for five discrete time steps which are marked with black dots in subfigure (a). The black squares indicate the positions of the mass-elements of the PEM(t) numbered from 1 to 8. The five time steps describe an oscillation cycle from open to open state. The time developments of the optimized parameters $Q_1[n], \dots, Q_8[n], Q_P[n]$ are depicted in subfigures (d).

The asked pitch raise could not be performed by all examined subjects. However, they all managed to perform a variation of pitch. The phonated pitch ranges in R1 to R4 comprise 79 – 102, 105 – 143, 148 – 165 and 80 – 92 Hz, respectively (red solid lines in Fig. 5b, 6b, 7b, 8b). Only recordings R1 and R3 present phonations with almost continuous pitch raise. In R2, the pitch decreases by time and abruptly increases at the end. The pitch in R4 presents an irregular development over time.

The non-stationary phonation presents either variations in frequency [subfigures (b)], fluctuations of the amplitudes of the area functions [subfigures (a)], or both. The PE-area signal of R1 shows an increasing amplitude over time. In contrast, the amplitudes of R3 and R4 decrease. The area signal of R2 does not demonstrate a clear trend. It presents a tremulous behavior which explains the worst adaptation results (Table 3). R3 and R4 present a complete closure, R2 only for the time steps $t = 239$ ms, and R1 does not close completely.

4 Discussion

An objective quantification of the PE-dynamics was formerly performed with the PE-MMM that is limited to oscillations that do not vary in amplitude and frequency (Schwarz et al. 2011). For the understanding of the biomechanical processes in the PE-segment for the complete spectrum of tracheo-esophageal sound production, it is of interest to investigate non-stationary phonations with variable frequency and amplitude. To overcome this limitation of the PE-MMM, the suggested PEM(t) applies time-dependent model parameters to objectively quantify and analyze the temporal behaviors of the pseudoglottal pressure, the oscillating mass, and the tissue tension during a non-stationary phonation task.

4.1 Validation and comparison to other models

The predefined parameter values of 75 synthetic data sets were identified with a mean error of 3.2 ± 3.8 %. Table 4 compares the synthetic validation of the PEM(t) to that of three other multi-mass models (stationary MMM (Schwarz et al. 2008) and time-dependent TMM (Wurzbacher et al. 2008) for vocal folds, PE-MMM (Schwarz et al. 2011) for PE-segment). Among all compared models and parameters, the PEM(t) achieves the smallest errors in identifying predefined, synthetic parameter values. Concerning the vocal folds, the relative error increases significantly for Q_P and slightly for Q_i from the stationary to the non-stationary model. This cannot be seen for the PE-segment models, as the errors decrease for both parameters. Additionally, the errors are worse for the stationary PE-model compared to

the MMM. Against this trend, the PEM(t) achieves significant better optimization results than the TMM, emphasizing the quality of the applied optimization procedure for the PEM(t).

Though the MMM, TMM, PE-MMM, and PEM(t) are related to each other, they differ in some essential points. Differently to the PEM(t), optimize the MMM, the TMM, and the PE-MMM the masses' rest positions $\mathbf{x}_{i,s}^r$. The optimization of the rest positions requires an individual optimization parameter for each mass. This would result in additional eight parameters for the optimization process and mean additional temporal expense. Thus, we decided to determine the temporal behavior of the rest positions by means of a heuristic approach (see Sect. 2.3.3), which is justified by the quality of the optimization results.

The presented PEM(t) applies eight coupled mass-spring oscillator-elements, whereas the MMM, TMM, and PE-MMM only consist of six elements. The additional elements allow for a better replication of the PE-contour, which positively influences the applied objective function Γ (see Eq. (15)) that consists of criteria that are derived from contour properties (Γ_s and Γ_d).

For the multi-mass models in Table 4, different optimization algorithms are applied to adapt the model dynamics. The MMM and PE-MMM use a genetic algorithm (GA) and the TMM applies the PDSM. The optimization of the presented PEM(t) stands out as it applies a combination of SA and PDSM. PDSM is a local optimization procedure, i.e., it looks for the next relative minimum. However, GA and SA are global optimization procedures, i.e., they are applied to find the global minimum of the objective function. Hence, the worst validation for the TMM (Table 4) may result of the fact that only a local optimization procedure is applied that runs into the next minimum, regardless if it is a local or global one. The better results for the MMM and PE-MMM, compared to the TMM, may be a consequence of the application of global optimization procedures. The PEM(t) combines a global and a local optimization, improving the consistency and stability in finding the global minimum (Yang et al. 2011).

4.2 Application to real PE-vibrations

Optimization quality—The general capability of the PEM(t) to simulate the dynamics of real PE-vibrations is demonstrated for four HS recordings R1 to R4 that all present non-stationary phonations, i.e., variations in pitch and amplitude. The extracted area functions were simulated with a mean correlation of 98 % (Table 3). For comparison, the MMM and PE-MMM achieve correlations of the area functions of 94 % (Schwarz et al. 2008) and 82 % (Schwarz et al. 2011), respectively. The high achieved correlations for the PEM(t) are in agreement with the good accordance of the PE and the model pitch (Figs. 5b, 6b, 7b, 8b and the small values for Γ_a (Table 3). The objective functions for R1, R3, and R4 are approximately in the same order of magnitude, see Table 3. For R2, the objective functions are about twice as large. The worse adaptation for R2 is on the one hand due to the irregular oscillation behavior of the PE-segment that results in a tremulous area function, see Fig. 6a. On the other hand, mass m_2 demonstrates a bad adaptation to the PE-contour for the closing phase, see Fig. 6c. From image one to two, the PE-segment closes, whereas mass m_2

presents an antipodal movement, i.e., it moves away from the point of origin. From figure two to three, mass m_2 follows the closing pseudoglottis and moves toward the origin. Concerning the open phase of the pseudoglottis (images four and five), mass m_2 demonstrates a good adaptation to the PE-contour.

The objective function Γ is a combination of three criteria, namely Γ_a , Γ_s , and Γ_d (see Eq. (15)). These are necessary to evaluate both the area and the contour adaptation of the PEM(t). The interaction of the three criteria is exemplarily demonstrated for the second image of Fig. 5c. Here, the PE-contour presents a concave shape, that is not simulated by the model contour. Moreover, the PEM(t) shows a clear deviation of mass m_6 from the PE-contour. The computed objective functions amount to $\Gamma_a = 0.0081$, $\Gamma_s = 0.0443$, and $\Gamma_d = 0.0368$ for this time step. The best adaptation result is obtained for the area. This is also visible in Fig. 5a at $t = 158$ ms (second black dot), where the areas of the PE-segment and the PEM(t) are approximately equal. Thus, the failure in fitting the concave contour is adjusted by the deviation of mass m_6 , such that both areas are similar. However, due to the deviation of m_6 and the non-concave PEM(t) contour, Γ_s is worse compared to Γ_a . The aberration of m_6 is clearly visible in Γ_d that also presents a worse adaptation result than Γ_a . This example demonstrates that a good fitting of the areas alone is not a sufficient attribute for a good adaptation. However, combined with Γ_s and Γ_d , the objective function Γ is a meaningful measure of the adaptation quality.

On the whole, the presented optimization results for R1 to R4 demonstrate good success in fitting the dynamics of the PEM(t) to the dynamics of the real PE-vibrations. The area functions in Figs. 5a, 6a, 7a and 8a agree with each other and the model contours replicates the PE-contour in an acceptable manner Figs. 5c, 6c, 7c and 8c. Additionally, the given pitch changes in Figs. 5b, 6b, 7b and 8b are modeled by the PEM(t), even though demonstrating in part complex behaviors, e.g., Fig. 6c. However, the adaptation quality of the PEM(t) was only demonstrated on four recordings. In future, more PE-dynamics will be simulated in order to gain more reliable data.

Optimized parameters—The time dependency of the PEM(t) is achieved by time-dependent Q -values that are adapted by the optimization procedure and that modify the constant model parameters according to Eqs. (5)–(8). Thus, the time development of the optimized Q -values objectively quantifies the temporal behavior of the tissue stiffness and the oscillating mass ($Q_1[n] \dots Q_8[n]$) and the temporal behavior of the pseudoglottal pressure ($Q_P[n]$). Figs. 5d, 6d, 7d and 8d show the time development of the textitQ-values after the optimization that all present significant temporal characteristics, what justifies the incorporation of time-dependent model parameters in the PEM(t). The optimization results in Table 2 confirm the adaptation procedure in having found adequate Q -values.

In recording R1, the optimization parameters $Q_1[n]$ to $Q_8[n]$ present a general increase over time (Fig. 5d). This behavior is equivalent to an increase of frequency in the mass-spring oscillator system $f \propto \sqrt{\frac{k}{m}}$, see Eq. (16). Thus, the pitch rise of R1 (Fig. 5b) is quantified by the optimized model parameters. Within the first 60 ms, R1 presents a decreasing frequency what is quantified by a decrease of $Q_2[n]$, $Q_3[n]$, $Q_4[n]$, $Q_6[n]$, and $Q_8[n]$ in the

corresponding time segment. The following pitch rise is the most distinctively quantified by $Q_1[n]$ and $Q_6[n]$. In the time interval of the demonstrated area function in Fig. 5a, the amplitude of the PE-segment presents a slight increase that is quantified by an increase of Q_P , i.e., the pseudoglottal pressure. A numerical study with a multi-mass model for the vocal folds demonstrated an increasing pressure with increasing phonation frequency (Yang et al. 2012). Thus, the increasing $Q_P[n]$ is also consistent with the phonated pitch rise. Consequently, the non-stationary dynamics of the PE-segment in R1, i.e., the increasing amplitude of the area signal and the increasing vibration frequency over time, are objectively quantified by the optimized Q -values. Moreover, the phonated pitch rise can be correlated to an increase of the tissue stiffness and/or a decrease of the oscillating tissue mass, as well as an increase of the pseudoglottal pressure.

Similar behavior can be seen for the pitch rise of R3 (Fig. 7b). Here, $Q_1[n]$ to $Q_8[n]$ generally increase over time, with a distinctive increase for $Q_4[n]$ and $Q_8[n]$ (7d). $Q_P[n]$ also increases over time, what is consistent with the increased frequency. The decreasing amplitude of the area signal within the first 100 ms (Fig. 7a) is a result of the tissue stiffening (i.e., increase of the Q -values) that is not balanced by the increased pressure. Here again, the pitch rise is a consequence of an increased tissue tension, a decrease of the tissue mass participating in the vibration, and an increase of the pseudoglottal pressure.

The pitch in R2 decreases over time with an abrupt increase starting from approximately 225 ms (Fig. 6b). This behavior is exactly quantified by $Q_2[n]$, even though the adaptation of mass m_2 to the PE-contour is expandable for the closing phase (see Fig. 6c). Within the first 150 ms $Q_P[n]$ increases, which would actually increase the frequency and were contradictory to the pitch fall. However, in this interval all Q_i except $Q_2[n]$ and $Q_6[n]$ increase, resulting in tissue stiffening. This stiffening counteracts the increasing pressure. From 150–200 ms $Q_1[n]$ – $Q_8[n]$ and $Q_P[n]$ generally decrease in correspondence to the falling pitch. The steep increasing pitch starting from 250 ms is quantified by all Q_i except Q_8 . The results of R2 show that not all regions of the PE-segment contribute in the same manner to the non-stationary phonation. Moreover, the responsible tissue regions may change over time, more precisely with phonated pitch. For example, $Q_6[n]$ does not quantify the falling pitch up to 225 ms, however, it quantifies the pitch rise at the end of the recording. $Q_8[n]$ in turn quantifies the pitch fall but does not fit the pitch rise.

Recording R4 presents a pitch rise with a falling frequency at the end (Fig. 8b). The increasing pitch is reflected in all $Q_i[n]$ that all present a slight increase over time (Fig. 8d). The decreasing pitch is quantified by $Q_4[n]$ to $Q_6[n]$. The amplitude of the area function decreases within the depicted time interval from 50–150 ms (Fig. 8a). This is quantified by Q_P that also decreases within this period (Fig. 8d).

The results in R1 to R4 demonstrate that the optimized model parameters are capable to objectively quantify the non-stationary dynamics of the PE-segment. Time-dependency of the model masses, springs, and pressure has been applied in numerical vocal folds models before (Lucero and Koenig 2005; Wurzbacher et al. 2006, 2008). By investigations with a time-dependent two-mass model, Wurzbacher et al. (2006) demonstrated an increasing Q -value for the springs and masses with increasing pitch. This finding is in accordance to ours.

Yang et al.(2012) simulated vocal fold vibrations at different excitation pressures with a three-dimensional multi-mass model. They found that the model pressure is larger for higher oscillation frequencies. We noticed the same effect, namely an increase of the model pressure over time in order to fit the phonated pitch rise.

Physiologic interpretation—The sound productions for both laryngeal and alaryngeal voice base on the myoelastic principle, and thus rely on the same physics (Moon and Weinberg 1987; Liu et al. 2004). Consequently, it is justified to interpret the optimized model parameters of the PEM(t) in a physiological manner: The optimization parameter $Q_P[n]$ describes the time variation of the pressure in the PE-segment. The parameters $Q_1[n] \dots Q_8[n]$ modify the springs and the masses in the PEM(t), and thus describe the stiffening and softening of the PE-tissue during the phonation process as well as the oscillating mass, i.e., the tissue that participates in the vibration. The total mass distribution in the PE-segment does not change. However, the mass which is entrained in the vibration may change. A muscle tension will stiffen the PE-tissue and consequently less tissue may actively vibrate. Thus, both the tissue stiffness and the oscillating mass change which is described by the PEM(t) by applying the same value $Q_i[n]$ for the springs and the masses, see Eqs. (5)–(7). The results show that the vibration frequency is mainly modified by varying the local tissue stiffness over time and/or by modifying the oscillating mass. The amplitude of the pseudoglottal opening is defined by an interaction of the viscoelastic tissue properties and the pressure caused by the bypassing air stream. Moreover, the vibration characteristics of the PE-segment locally differ along the contour. That means, an increase of the oscillation frequency of the PE-segment does not consider a change of the oscillation behavior of the complete pseudoglottis. The tissue regions along the PE-contour are more or less involved in the non-stationary phonation process. These regions are not characterized by the local shape of the PE-segment, what is demonstrated by the positions of mass m_1 and mass m_6 in R1. Both present similar temporal behavior (Fig. 5c, d) and represent regions of the PE-segment where the contour demonstrates different curvature. While the location of m_1 is favorably curved, m_6 correlates rather to a non-curved contour. Thus, the local oscillation behavior is rather influenced by the tissue properties, e.g., the tissue stiffness, than the local shape of the PE-segment.

For a better understanding of the biomechanical processes in the PE-segment during a non-stationary phonation, more recordings of the PE-segment will be performed and analyzed.

4.3 Limitations and outlook

Even though the validation with synthetic data sets and the application to real data revealed a sufficient accuracy, an enhancement of the suggested optimization procedure is possible from the computational point of view. The objective function is non-convex, i.e., it contains various local minima, complicating the optimization process (Döllinger et al. 2002). An analytic solution for the ordinary differential equation system does not exist, which requires the application of a global optimization procedure. The utilized *Simulated Annealing* method is a stochastic search algorithm, i.e., the optimization results achieved in this work are neither unique nor do they per se represent the global minimum. However, the achieved results are consistent with findings of previous studies, thus justifying the suggested

optimization procedure. The physiological significance of model parameters obtained from an optimization process was demonstrated by Yang et al. (2012).

A limitation of the suggested PEM(t) is the constricted consideration of the vertical volume of PE-tissue which participates in the vibration. Because of the top view during the endoscopic examination, only the upper part of the oscillating tissue can be registered. The heights of the mass-elements in the PEM(t) are 0.25 cm and 0.05 cm for the lower and the upper plane, respectively, similar to multi-mass models for the vocal folds (Schwarz et al. 2008; Wurzbacher et al. 2006). The influence of more caudally placed tissue on the PE-dynamics is thus not captured by the simulation.

The PE-dynamics are described by the time signals of the pseudoglottal area and contours. Due to the fact that the vibration patterns of the PE-segment have an irregular shape with neither a symmetry axis nor a point of symmetry, no specific contour points can be tracked over time. On the contrary, for numerical vocal fold models the masses' rest positions are well defined by projections of specific contour points onto the glottal mid line (Schwarz et al. 2008; Wurzbacher et al. 2008; Döllinger et al. 2002). However, a comparable advance is not portable for the PE-segment. Instead, the masses' initial rest positions are determined by adjusting them between the maximum and the minimum PE-contour, see Sect. 2.3.2. Due to the irregular contours and vibration patterns, this approach may result in a suboptimal setting. The influence of the initialization to the optimization results has not been investigated yet.

Adopted from the PE-MMM, the masses and springs are initialized with the values of Ishizaka and Flanagan, divided by the number of the applied masses (Schwarz et al. 2011). Additionally, the initialization values are reduced by a heuristic factor of 30 %. Otherwise, the necessary pseudoglottal pressures were beyond physiological conditions. The initialization values are adopted from numerical vocal fold models as no comparable values have been published for the PE-segment yet.

The PE-dynamics extracted from the HS recordings are given in pixel units as there is no access to metric information. However, the HS recordings of the PE-dynamics were acquired with two kinds of cameras with resolutions at 512×512 and 256×256 pixels. For the purpose of comparability, and because the numerical model needs metric input data, the

pixel units are mapped into metric units by the factor $\Gamma = \frac{d^{\text{PE}}}{d^{\text{HS}}}$ that relates the physical size

of the PE-segment (anterior-posterior distance) with the size in the HS recording. In this work, we assume $d^{\text{PE}} = 1.07$ cm for the physical size, according to Lundström et al. (2008). Thus, the mapping into metric dimensions is a first approximation. For more precise metric measurements in the future, the geometry and dynamics of the PE-vibrations will be investigated by a laser projection system (Luegmair et al. 2010). For the in vivo application, the laser system will be miniaturized and coupled to the endoscope. The deformations of the projected laser dots on the surface can be used to determine the exact metric dimensions of the PE-segment. The mapping factor Γ is also the limiting factor for the accuracy of the input data. For the extraction of the PE-area out of the HS videos, a pixel-based approach is applied whose accuracy depends on the camera resolution. Thus, the accuracies are equal to

Γ and Γ^2 for the input PE-contour and the input PE-area, respectively. The mapping factors amount to 0.0108, 0.0626, 0.0577, and 0.0214 for R1 to R4. The output of the model is only limited by the numerical accuracy of the applied computer.

The results show that the optimized model parameters objectively describe the non-stationary PE-vibrations. The time-dependent parameters of the TMM were applied to classify healthy and pathological vocal fold vibrations (Wurzbacher et al. 2008). This undertaking is not applicable for tracheo-esophageal sound production that is per se a pathological voice. The here performed objective quantification is rather a first step to the investigation of the interactions between the vibration characteristics of the PE-segment and the resulting substitute sound signal. For that purpose, the acoustic signal will be objectively assessed by an automatic speech recognition system (Haderlein et al. 2009) and will be compared to the optimized model parameters describing the dynamic range of the PE-vibrations. The aim will be the definition of ranges and/or combinations of Q-values that correlate with a good/poor substitute voice. The knowledge, how certain types of tissue vibration patterns or specially shaped PE-segments benefit the voice quality might influence the pharyngeal reconstructive surgery during laryngectomy. Thus, the PE-tissue can be stiffened by controlling tissue scaring or a specific shaped pseudoglottis can be reconstructed. Another application of the gained information might be the treatment of aphone tracheo-esophageal speakers due to muscle hypertonicity. Knowing more about how certain PE-regions are involved into sound production can identify qualified regions for injecting botulinum toxin (Bartolomei and Zambito Marsala 2011).

5 Conclusion

For the analysis of irregular tissue vibrations, the existing stationary multi-mass model for the PE-segment was augmented by time-dependent model parameters. The resulting PEM(t) is a step toward the objective description of non-stationary tissue vibrations in the PE-segment. Hence, non-stationary phonation paradigms in terms of pitch and amplitude modulations can be applied. These are closer to the running alternative speech in everyday life, and thus allow for more realistic investigations than the analysis of a sustained phonation.

By a time-dependent optimization procedure, the dynamics of the PEM(t) are adapted to the dynamics of the PE-segment. The irregular behavior of the PE-vibrations is reflected in the temporal characteristics of the optimized model parameters. The modulations of pitch and amplitude are the results of spatial as well as temporal alterations of the stiffness and effective mass distribution along the PE-contour and of the pressure in the PE-segment. However, further investigations are necessary and will be performed to correlate the optimized model parameters to the quality of the substitute voice. Moreover, regions in the PE-segment will have to be identified that more or less influence the substitute sound signal.

Acknowledgments

The work was supported by Deutsche Krebshilfe Grant No. 109204 “Analyse und Modellierung der pharyngo-ösophagealen Schleimhautdynamik nach krankheitsbedingter Kehlkopfentfernung”. Dr. Patel’s contributions to this investigation were supported by NIH/NIDCD grant no. R03DC011360-01.

References

- van As CJ, Tigges M, Wittenberg T, de Coul BMO, Eysholdt U, Hilgers FJ (1999) High-speed digital imaging of neoglottic vibration after total laryngectomy. *Arch Otolaryngol Head Neck Surg* 125(8):891–897 [PubMed: 10448737]
- Bartolomei L, Zambito Marsala S, Pighi GP, Cristofori V, Pagano G, Pontarin M, Gioulis M, Marchini C (2011) Botulinum toxin type a: an effective treatment to restore phonation in laryngectomized patients unable to voice. *Neurol Sci* 32(3):443–447 [PubMed: 21479612]
- Brent RP (1973) Algorithms for minimization without derivatives. Prentice-Hall, Englewood Cliffs
- Deshmane VH, Rao RS, Parikh HK, Divatia JV, Parikh DM, Suktharnkar PS, Nikam MT (1995) Pharyngoesophageal segment manometry: its role in determining post-laryngectomy speech. *IJO and HNS* 47(3):185–190
- Döllinger M (2009) The next step in voice assessment: High-speed digital endoscopy and objective evaluation. *Curr Bioinform* 4(2):101–111
- 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 [PubMed: 12148815]
- Haderlein T, Riedhammer K, Nöth E, Toy H, Schuster M, Eysholdt U, Hornegger J, Rosanowski F (2009) Application of automatic speech recognition to quantitative assessment of tracheoesophageal speech with different signal quality. *Folia Phoniater Logop* 61(1):12–17 [PubMed: 19122460]
- Ingber L (1996) Adaptive simulated annealing. *J Control Cybern* 25(1):33–54
- Ishizaka K, Flanagan JL (1972) Synthesis of a voiced sound from a two-mass model of the vocal cords. *bell syst tech J* 51(6):1233–1268
- Liu H, Wan M, Wang S, Niu H (2004) Aerodynamic characteristics of laryngectomees breathing quietly and speaking with the electrolarynx. *J Voice* 18(4):567–577 [PubMed: 15567058]
- Lohscheller J, Döllinger M, Schuster M, Eysholdt U, Hoppe U (2003) The laryngectomee substitute voice: image processing of endoscopic recordings by fusion with acoustic signals. *Methods Inf Med* 42(3):277–281 [PubMed: 12874662]
- Lohscheller J, Döllinger M, Schuster M, Schwarz R, Eysholdt U, Hoppe U (2004) Quantitative investigation of the vibration pattern of the substitute voice generator. *IEEE Trans Biomed Eng* 51(8):1394–1400 [PubMed: 15311824]
- Lucero JC, Koenig LL (2005) Simulations of temporal patterns of oral airflow in men and women using a two-mass model of the vocal folds under dynamic control. *J Acoust Soc Am* 117(3 Pt 1):1362–1372 [PubMed: 15807024]
- Luegmair G, Kniesburges S, Zimmermann M, Sutor A, Eysholdt U, Döllinger M (2010) Optical reconstruction of high-speed surface dynamics in an uncontrollable environment. *IEEE Trans Med Imag-ing* 29(12):1979–1991
- Lundström E, Hammarberg B, Munck-Wikland E, Edsberg N (2008) The pharyngoesophageal segment in laryngectomees-videoradiographic, acoustic, and voice quality perceptual data. *Logoped Phoniater Vocol* 33(3):115–125 [PubMed: 18608876]
- Moon JB, Weinberg B (1987) Aerodynamic and myoelastic contributions to tracheoesophageal voice production. *J Speech Hear Res* 30(3):387–395 [PubMed: 3669645]
- Papadas TA, Alexopoulos EC, Mallis A, Jelastopulu E, Mastronikolis NS, Goumas P (2010) Survival after laryngectomy: a review of 133 patients with laryngeal carcinoma. *Eur Arch Otorhinolaryngol* 267(7):1095–1101 [PubMed: 19921233]
- Rasp O, Lohscheller J, Döllinger M, Eysholdt U, Hoppe U (2006) The pitch rise paradigm: A new task for real-time endoscopy of non-stationary phonation. *Folia Phoniater Logop* 58:175–185 [PubMed: 16636565]
- Schuster M, Lohscheller J, Kummer P, Hoppe U, Eysholdt U, Rosanowski F (2003) Quality of life in laryngectomees after prosthetic voice restoration. *Folia Phoniater Logop* 55(5):211–219 [PubMed: 12931055]
- Schuster M, Rosanowski F, Schwarz R, Eysholdt U, Lohscheller J (2005) Quantitative detection of substitute voice generator during phonation in patients undergoing laryngectomy. *Arch Otolaryngol Head Neck Surg* 131(11):945–952 [PubMed: 16301364]

- Schwarz R (2007) Model-based quantification of pathological voice production. PhD thesis, University Erlangen-Nuremberg, shaker (2007)
- Schwarz R, Döllinger M, Wurzbacher T, Eysholdt U, Lohscheller J (2008) Spatio-temporal quantification of vocal fold vibrations using high-speed videoendoscopy and a biomechanical model. *J Acoust Soc Am* 123(5):2717–2732 [PubMed: 18529190]
- Schwarz R, Hüttner B, Döllinger M, Luegmair G, Eysholdt U, Schuster M, Lohscheller J, Gürlek E (2011) Substitute voice production: Quantification of pe segment vibrations using a biomechanical model. *IEEE Trans Biomed Eng* 58(10):2767–2776 [PubMed: 21558056]
- Shah J (2001) Cancer of the head and neck American Cancer Society Atlas of Clinical Oncology, BC Decker Inc, Hamilton
- Singer MI, Blom ED, Hamaker RC (1983) Voice rehabilitation after total laryngectomy. *J Otolaryngol* 12(5):329–334 [PubMed: 6644863]
- Stiglmayr M, Schwarz R, Klamroth K, Leugering G, Lohscheller J (2008) Registration of pe segment contour deformations in digital high-speed videos. *Med Image Anal* 12(3):318–334 [PubMed: 18304855]
- Takeshita TK, Zozolotto HC, Ribeiro EA, Ricz H, de Azevedo-Marques PM, Dantas RO, Aguiar-Ricz L (2013) Relation between the dimensions and intraluminal pressure of the pharyngoesophageal segment and tracheoesophageal voice and speech proficiency. *Head Neck* 35(4):500–504 [PubMed: 22266980]
- Titze IR (2000), Principles of voice, Production, vol 2 NCVS
- Wurzbacher T, Schwarz R, Döllinger M, Hoppe U, Eysholdt U, Lohscheller J (2006) Model-based classification of nonstationary vocal fold vibrations. *J Acoust Soc Am* 120(2):1012–1027 [PubMed: 16938988]
- Wurzbacher T, Döllinger M, Schwarz R, Hoppe U, Eysholdt U, Lohscheller J (2008) Spatiotemporal classification of vocal fold dynamics by a multimass model comprising time-dependent parameters. *J Acoust Soc Am* 123(4):2324–2334 [PubMed: 18397036]
- Yang A, Stingl M, Berry DA, Lohscheller J, Voigt D, Eysholdt U, Döllinger M (2011) Computation of physiological human vocal fold parameters by mathematical optimization of a biomechanical model. *J Acoust Soc Am* 130(2):948–964 [PubMed: 21877808]
- Yang A, Berry DA, Kaltenbacher M, Döllinger M (2012) Three-dimensional biomechanical properties of human vocal folds: parameter optimization of a numerical model to match in vitro dynamics. *J Acoust Soc Am* 131(2):1378–1390 [PubMed: 22352511]

Fig. 1.

a After removing the larynx, the trachea is connected to the frontal neck (tracheostoma) to preserve breathing. The tissue in the changeover of esophagus and pharynx is called PE-segment. Closing the tracheostoma while exhaling forces the air to pass the voice prosthesis valve that connects trachea and esophagus. The mucosa tissue in the PE-segment is forced to oscillate what generates the acoustic signal for substitute voice. The dynamics of the PE-segment are registered by a high-speed camera. **b** Extracted frames of a HS recording of the PE-segment, representing one oscillation cycle. Visible is the pseudoglottis (*dark area*) and the surrounding mucosa tissue. For the analysis of the PE-dynamics, the PE-contour is segmented (*white lines*) and extracted in each frame

Fig. 2.

The PEM(t) **a** consists of eight horizontally coupled two-mass oscillators (**b**). The pseudoglottis is represented by the minimum opening spanned by either the masses $m_{i,s}$ ($i = 1, \dots, 8$) of the *lower* ($s = 1$) or *upper* ($s = 2$) plane. $k_{i,s}^a$, $k_{i,s}^l$, and k_i^v are the anchor, the longitudinal, and the vertical coupling spring, a_i are the anchors corresponding to the masses $m_{i,s}$. $\bar{a}_i$ and $\bar{k}_{i,s}^a$ are imaginary anchors and corresponding anchor springs. For the purpose of clarity, dampers are not visualized. **a** PEM(t). **b** Two-mass oscillator-element. *Left: Top view. Right: Front view*

Fig. 3.

Flow charts of adapting the dynamics of the PEM(t) to those of the PE-segment. The dynamics are separated into blocks with 4 oscillation cycles and 50 % overlap. **a** Each block is initialized with the optimized parameter set at half width of the previous block. The adaptation within a block is performed by loops of *Simulated Annealing* and *Powell's direction set method*. The adaptation is completed after a predefined number of iterations or if the improvement between two consecutive iterations is smaller than a predefined threshold Γ . **b** For minimizing the objective function (SA and PDSM), the optimization parameters are linearly interpolated between the parameter sets of the first and last time step. The PEM(t) is scaled with the parameter sets, and the resulting dynamics are compared to those of the PE-segment. If the adaptation quality is not sufficient, the parameter set of the last time step is varied

Fig. 4. J versus ε_{rel} for the optimizations of the 75 synthetic data sets with *triangular* (triangles), *elliptic* (squares), and *circular* (circles) model contours. The objective function was computed for the complete dynamics, i.e., after the optimization process. The relative error describes the mean over $Q_1[n], \dots, Q_8[n], Q_A[n]$

Fig. 5.

Adaptation results after fitting the PEM(t) to the dynamics of R1. **a** Area functions $a^{\text{PE}}[n]$ (solid line) and $a^{\text{PEM}(t)}[n]$ (dashed line) for a short sequence of non-stationary phonation. **b** Frequency over time for the PE-segment and the PEM(t) for the complete phonation process. **c** Extracted PE contour (black solid line), pseudoglottis (red area), and state of the PEM(t) (black dashed line). The squares indicate the positions of the mass-elements m_1 to m_8 . Mass m_1 is color coded in gray, the remaining masses are arranged in clockwise order. Demonstrated is one oscillation cycle at the time steps marked with circles in subfigure (a).

d Time development of the parameter set $P[n]$, containing the nine optimization parameters of the PEM(t), over the complete phonation process. The symbols guide the eye and do not mark specific time steps

Fig. 6. Adaptation results after fitting the PEM(t) to the dynamics of R2. **a** Area functions $a^{\text{PE}}[n]$ (solid line) and $a^{\text{PEM}(t)}[n]$ (dashed line) for a short sequence of non-stationary phonation. **b** Frequency over time for the PE-segment and the PEM(t) for the complete phonation process. **c** Extracted PE contour (black solid line), pseudoglottis (red area), and state of the PEM(t) (black dashed line). The squares indicate the positions of the mass-elements m_1 to m_8 . Mass m_1 is color coded in gray, the remaining masses are arranged in clockwise order. Demonstrated is one oscillation cycle at the time steps marked with circles in subfigure (a).

d Time development of the parameter set $P[n]$, containing the nine optimization parameters of the PEM(t), over the complete phonation process. The symbols guide the eye and do not mark specific time steps

Fig. 7. Adaptation results after fitting the PEM(t) to the dynamics of R3. **a** Area functions $a^{\text{PE}}[n]$ (solid line) and $a^{\text{PEM}(t)}[n]$ (dashed line) for a short sequence of non-stationary phonation. **b** Frequency over time for the PE-segment and the PEM(t) for the complete phonation process. **c** Extracted PE contour (black solid line), pseudoglottis (red area), and state of the PEM(t) (black dashed line). The squares indicate the positions of the mass-elements m_1 to m_8 . Mass m_1 is color coded in gray, the remaining masses are arranged in clockwise order. Demonstrated is one oscillation cycle at the time steps marked with circles in subfigure (a).

d Time development of the parameter set $P[n]$, containing the nine optimization parameters of the PEM(t), over the complete phonation process. The symbols guide the eye and do not mark specific time steps

Fig. 8.

Adaptation results after fitting the PEM(t) to the dynamics of R4. **a** Area functions $a^{\text{PE}}[n]$ (solid line) and $a^{\text{PEM}(t)}[n]$ (dashed line) for a short sequence of non-stationary phonation. **b** Frequency over time for the PE-segment and the PEM(t) for the complete phonation process. **c** Extracted PE contour (black solid line), pseudoglottis (red area), and state of the PEM(t) (black dashed line). The squares indicate the positions of the mass-elements m_1 to m_8 . Mass m_1 is color coded in gray, the remaining masses are arranged in clockwise order. Demonstrated is one oscillation cycle at the time steps marked with circles in subfigure (a).

d Time development of the parameter set $P[n]$, containing the nine optimization parameters of the PEM(t), over the complete phonation process. The symbols guide the eye and do not mark specific time steps

Table 1

Initialization values for the PEM(t) for the masses ($\hat{m}_{i,s}$) anchor ($\hat{k}_{i,s}^a$) and vertical springs ($\hat{k}_i^v$), and pseudoglottal pressure ($\hat{P}^s$). The units are g, N/mm and cm H₂O. $i = 1, \dots, 8$ is the index of the mass-element in plane $s = 1, 2$

Plane s	$\hat{m}_{i,s}$	$\hat{k}_{i,s}^a$	$\hat{k}_i^v$	$\hat{P}^s$
1	0.0109	0.007	0.0022	25
2	0.0022	0.0007	0.0022	25

Table 2

Mean relative errors $\bar{\epsilon}_{rel}$ for the optimized Q -values (in %). The data is averaged over 25 randomly generated synthetic data sets for triangular, elliptic, and circular model configurations in each case. $\bar{Q}_i$ and $\bar{Q}$ are the average over Q_1 to Q_8 and over all nine Q -values, respectively

	$\bar{Q}_i$	Q_P	$\bar{Q}$
$\bar{\epsilon}_{rel}^{triangular}(\%)$	2.9 ± 3.3	3.2 ± 3.3	2.9 ± 3.3
$\bar{\epsilon}_{rel}^{elliptic}(\%)$	3.0 ± 3.9	3.2 ± 2.7	3.0 ± 3.8
$\bar{\epsilon}_{rel}^{circular}(\%)$	3.5 ± 4.2	3.9 ± 3.4	3.5 ± 4.2
$\bar{\epsilon}_{rel}^{all}(\%)$	3.1 ± 3.8	3.4 ± 3.2	3.2 ± 3.8

Table 3

Optimization results after fitting the PEM(t) to the extracted PE-dynamics of R1 to R4. Γ is the value of the objective function for the complete dynamics. Γ_a , Γ_s , and Γ_d are the objective functions for the area, the intersection, and the contour distance criteria, see Eq. (15). κ describes the correlation between the area signals of the PE-segment and the PEM(t)

	Γ	Γ_a	Γ_s	Γ_d	$\kappa(\%)$
R1	0.13	0.03	0.04	0.06	98
R2	0.27	0.06	0.08	0.13	97
R3	0.12	0.03	0.04	0.05	99
R4	0.15	0.04	0.04	0.07	98

Table 4

Summary of the mean relative errors of identifying predefined values for Q_P and Q_i for the stationary MMM (Schwarz et al. 2008, the non-stationary TMM (Wurzbacher et al. 2008, the PE-MMM (Schwarz et al. 2011 and the PEM(t)

$\bar{\epsilon}_{rel}$ in %	vocal folds		PE-segment	
	MMM	TMM	PE-MMM	PEM(t)
Q_P	3.5	15.1	6.3	3.4
Q_i	5.1	6.6	9.1	3.1