18th International Conference
ENGINEERING MECHANICS 2012
Svratka, Czech Republic, May 14 – 17, 2012

pp. 1087–1102
Paper #184

# COMPUTER SIMULATION OF MUSICAL SINGER'S VOICE BASED ON MRI AND ACOUSTIC MEASUREMENTS

**V. Radolf[*], A. Nissinen[****], A. M. Laukkanen[**], R. Havlík[***], J. Horáček[*]**

**Abstract:** *An inverse method was used to estimate the vocal tract geometry as a 1 D model on the basis of acoustical characteristics of a professional musical singer before and after vocal exercising. The basic geometrical data for the model were obtained from magnetic resonance images (MRI) registered during sustained phonation of vowels [a:], [i:], [u:] produced in naive and professional ways (before and after exercising respectively). The model was used for numerical simulations of the voice signals. The results of simulation were compared to the acoustic recordings. According to the results, a singer's formant cluster was accomplished after exercising. It seemed to be due to lowering of the larynx and lengthening and narrowing of the epilarynx. The area ratio between the low pharynx and the epilarynx increased for [a:] but decreased for [i:] and [u:] after exercising being between 3.33 and 4.39. There was a qualitative agreement between 3D measurements of MRI and the results of modeling. The results suggest that for a singer's formant cluster a relatively low pharynx over epilarynx ratio may be sufficient, at least if the larynx lowers.*

**Keywords:** ***Biomechanics of voice, singer's and speaker's formant cluster, acoustic effects of vocal exercises.***

## 1. Introduction

In operatic singing, singers make use of a special voice quality in order to be heard over the orchestra without a microphone. The important acoustic component which determines the operatic quality of the voice, especially in male singers, is the so-called "singer's formant" (Sundberg, 1974, 2003; Titze, 2000). Similarly the voice quality of speakers, especially actors, is improved by the so-called "speaker's formant" (Leino, 1994; Leino et al., 2011).

Sundberg (1974) formulated an acoustic interpretation of singer´s formant as clustering of formants F3 – F5 in the vicinity of 3 kHz. According to the calculations by Fant (1960) such a situation can occur if the ratio of the cross-sectional area of the lower pharynx over the outlet of the epilaryngeal tube is six or higher. The physical modeling results by Sundberg supported the findings of Fant. Different results were obtained by Detweiler (1994) who investigated by MRI scanning the vowel phonation of tenor and baritone singers phonating vowel [a:] in modal register. According to her results, the subjects laryngopharyngeal/laryngeal outlet cross-sectional area ratios varied from 2.9:1 to 3.7:1, which wasn´t consistent with the hypotheses by Sundberg and Fant. Later, Sundberg (2003) criticized the Detweiler's way of measuring the areas, which was accomplished from the sagittal and transversal distances measured in 2D slices. The epilarynx tube, located just above the vocal folds, was also theoretically considered as a dominant source of clustering of formant frequencies especially when the cross-sectional area of this tube is approximately six times less than the area of the lower pharynx by Titze and Story (1997), and Story (2003). However, the physiological adjustment used in singers or in actors has not been known in sufficient details yet.

[*] Ing. Vojtěch Radolf, Ph.D. and Ing. Jaromír Horáček, DrSc.: Institute of Thermomechanics, Academy of Sciences of the Czech Republic; Dolejškova 1402/5; 182 00, Prague; CZ, e-mail: radolf@it.cas.cz and e-mail: jaromirh@it.cas.cz

[****] Antti Nissinen Ph.D, Dept. of Applied Physics, University of Eastern Finland, FI-70211, Kuopio, Finland, e-mail: antti.nissinen@uef.fi

[**] Prof. Anne-Maria Laukkanen, Ph.D.: Speech and Voice Research Laboratory, School of Education, University of Tampere; FIN-33014, Tampere; Finland, e-mail: Anne-Maria.Laukkanen@uta.fi

[***] MUDr. Radan Havlík, Ph.D.: AUDIO – Fon Centr. s.r.o.; Obilní trh 4; 602 00, Brno; CZ, e-mail: radan.ha@seznam.cz

The rationale of the present paper was to estimate the anatomical/geometrical adjustments and to model the acoustic changes that occur in the voice of a male professional musical actor after vocal warming up by vocal exercises.

## 2. MRI measurement and 3D vocal tract modeling

A Czech male musical actor (60 years, baritone) served as a subject in the magnetic resonance imaging investigation. Lying supine in the MRI machine he first produced the vowels [a:], [i:], [u:] in a naive technique and after ca 5 minute vocal exercising again in a professional "musical actor's" manner, aiming at the best voice quality. Each vowel sample was produced for at least 20 s to enable the MRI scanning. The samples were produced on a comfortable pitch, at approximately the same fundamental frequency. Being vocally trained, the subject was able to keep the articulation and phonation constant through a sustained vowel phonation. The subject's head position was stabilized with a support. Before starting the MRI measurement the subject phonated each sample with normal auditory feedback (i.e. without the MRI device on) and sustained the same phonation setting during MRI scanning. MRI scanning was performed at the Dept. of Medical Imaging, St. Anne's Faculty Hospital in Brno, using the 1.5 Tesla MRI device (Symphony Magnetom, Siemens). The imaging parameters were as follows: Field of view 236 x 270 mm, slice thickness 1.5 mm, acquisition time 20.07 s, number of averages 1, repetition time 5.49 s, echo time 2.88 s, number of sagittal images 44, resolution 512 x 448 pixels (cf. Laukkanen et.al., 2012).

For viewing the MR images (see Figure 1) and for measuring the changes of the vocal tract a Syngo FastView software (Siemens AG) and ITK Snap 2.0 were used. Lengths of different parts of the vocal tract were measured from midsagittal images as follows. Horizontal length (HL) was measured as the distance from the lips to the anterior lower edge of the anterior arch of Atlas. Vertical length (VL) was measured from the anterior lower edge of the anterior arch of Atlas down to the vocal folds. Total length (TL) of the vocal tract was calculated by adding HL and VL. Length of the epilaryngeal tube (EL) was measured from the collar of the epiglottis down to the vocal folds.

The 3D model of the vocal tract was reconstructed from the MRI images using Mimics (version 13.1, Materialise, Belgium). Especially the areas of the outlet of the epilarynx and the inlet of the low pharynx were studied from transversal slices of the 3D volume models of this part of the vocal tract, see Figure 2. The area of the inlet to the pharynx was studied just above the collar of epiglottis, while the area of the outlet of the epilarynx was studied just below the collar of epiglottis, at the point where the epilaryngeal tube and the sinus piriformes are separated. The region of interest (the region of the lower pharynx and upper epilarynx) of the vocal tract was segmented and the surface mesh of the focal tract was constructed. The surface meshes consisted of triangular elements. The cross sectional areas of the meshes were computed using Matlab. The cross sections were taken so that they were perpendicular with the midline of the mesh (vocal tract). In other words, the tangent of the midline determines the normal vector of the cross section. The area of each cross section was determined by modeling the cross section with a 2D mesh and summing the areas of the elements in the mesh. The ratio of the inlet to the pharynx over the outlet of the epilarynx was calculated from the areas.

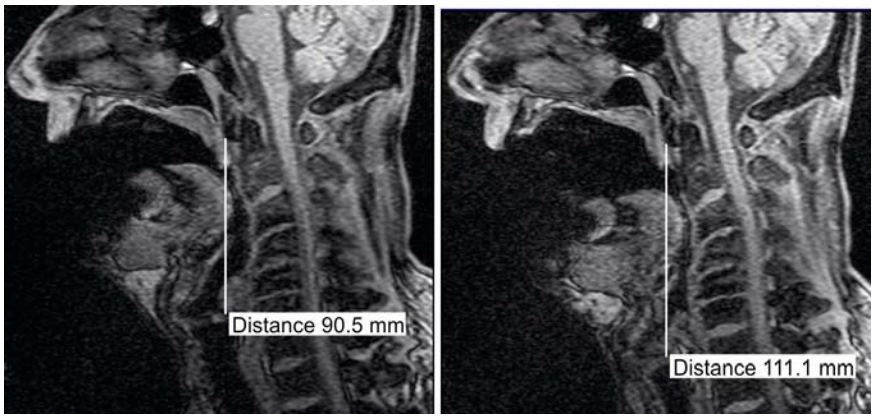
## 3. Acoustic measurements

Due to noise and magnetic field no acoustic recording was possible during the MR imaging. For acoustic measurements the subject's voice was recorded during the same tasks afterwards in a sound-treated studio using the microphone Center 322 DATA LOGGER with software lingWAVES version 2.5 (sensitivity 30-130dB , sampling frequency of 44.1 kHz).

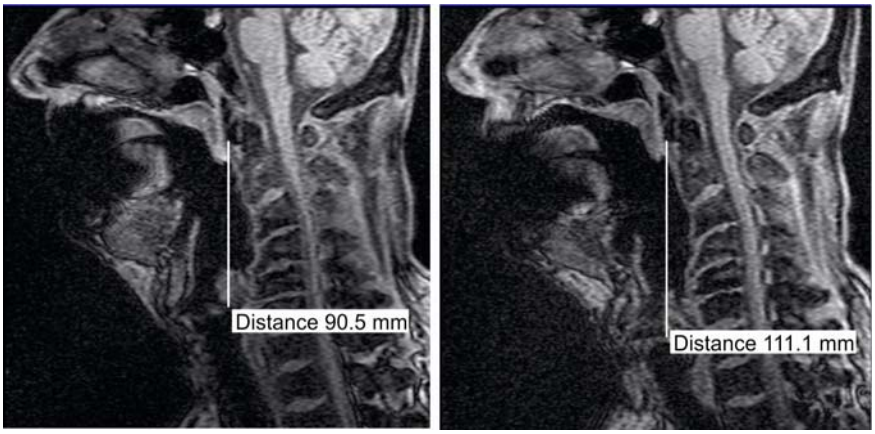
Acoustic analysis was done in Matlab by averaging frequency spectra calculated by FFT using 1s time windows with 75% overlap (see thin lines in Figure 4). Sound pressure level SPL was computed for each harmonics. Then the resulting spectra were averaged in the frequency bands (windows) equal to the fundamental frequency F0 with overlap of F0-10 Hz. Thus the new curves of "filtered spectra" were obtained (see thick lines in Figure 4) and the maxima of these curves were considered as formants.

## 4. MRI results

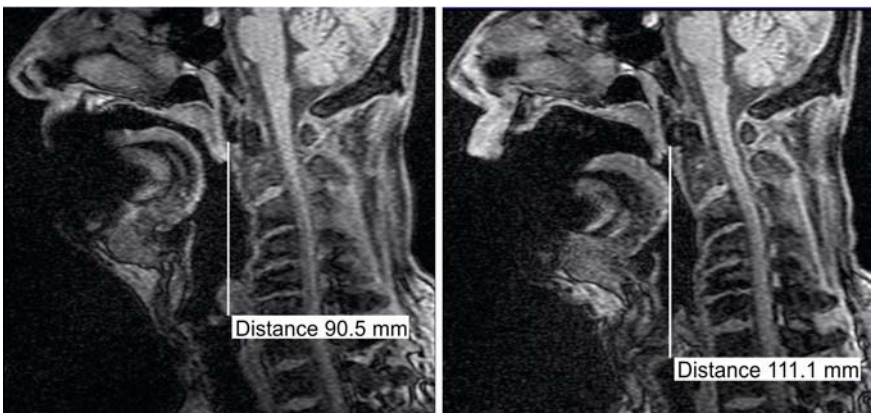
The MR images (see Figure 1) revealed lowering of the larynx from the fourth to the fifth vertebra, rising of the soft palate and thus a tighter closure of the velar port, prolongation of the vocal tract and of the epilaryngeal tube, widening of the mouth cavity due to a lower position of the tongue, narrowing and lengthening of the epilaryngeal and lower pharyngeal region and widening of the higher pharynx for [i:], and narrowing of it for [a:] and [u:] in phonation after the vocal exercising.




[a:] before and after warm-up




[i:] before and after warm-up




[u:] before and after warm-up

*Fig. 1: Midsagittal MR images for the vowels [a:], [i:], [u:] before and after the vocal exercising.*

The measured data for vocal tract length and its changes after warm up are summarized in Table 1 for all three vowels investigated.

The areas measured in the cross-sections at the end of the lower pharynx $A_{ph}$ and at the upper end of the epilaryngeal tube $A_{ep}$ and their ratios $R = A_{ph} / A_{ep}$ are summarized in Table 2. For all three vowels [a:], [i:] and [u:] after warm up the ratio was between $R$=3.33 and $R$=4.39, i.e. all ratios were lower than the value $R$=6 considered in the previous studies.

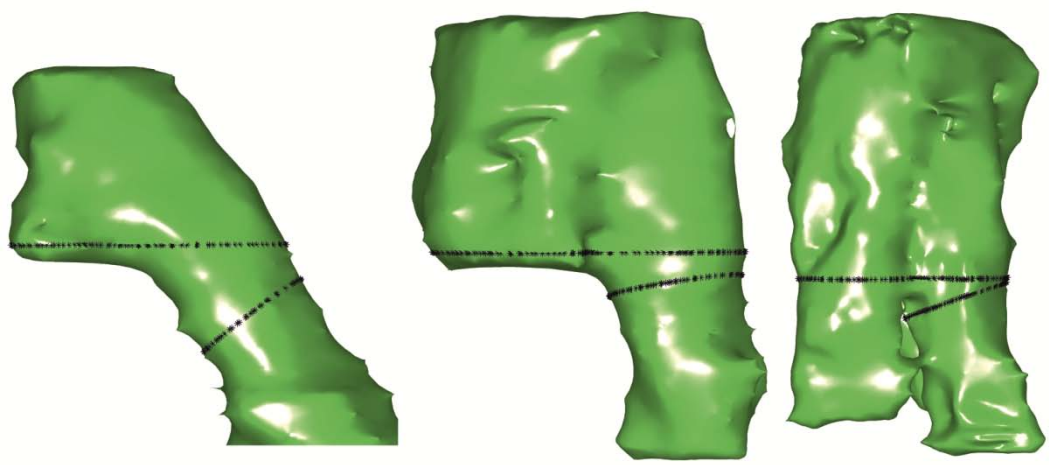


*Fig. 2: The 3D volume models of the vocal tract in the region of the lower pharynx and the upper epilaryngeal tube for the vowels [a:], [i:] and [u:] with the marked cross-sections where the areas were measured (from left to right).*

## 5. Acoustic results

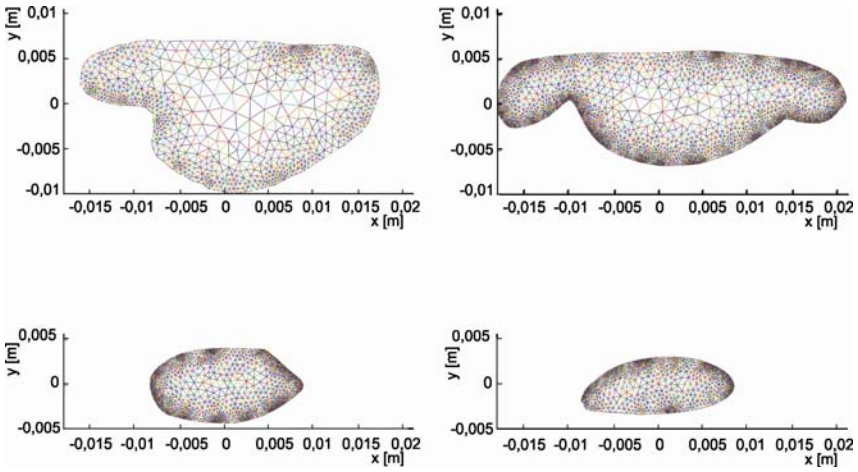
The acoustic analyses showed that after exercising a cluster of two or three formants was formed for the vowel [a:] in the range of F3-F5 between 2 and 4 kHz, and similarly in the range of F2-F4 between 1.4 and 3.4 kHz for the vowel [i:] and in the range of F3-F4 between 1.8 and 3 kHz for the vowel [u:] (see Figure 4 and Table 3). Such clustering of the higher formants leads to a stronger speaker's/singer's formant. A weak resonance in the frequency region at about 1.6 kHz for the vowels [a:] and [u:] and at about 1.2 kHz for the vowel [i:] can be caused by a slight nasality, which corresponds well to the study by Vampola et al. (2008b), where the first oro/nasal formants were found in the frequency region above the formants $F_1$ and $F_2$ for the vowels [a:] and [u:] and between the formants $F_1$ and $F_2$ for the vowel [i:].

*Tab. 1: Horizontal length of the vocal tract HL, vertical length of vocal tract VL, total length of vocal tract TL, length of the epilaryngeal tube EP and their changes after the warm-up for the vowels [a:], [i:] and [u:]. (MRI results)*

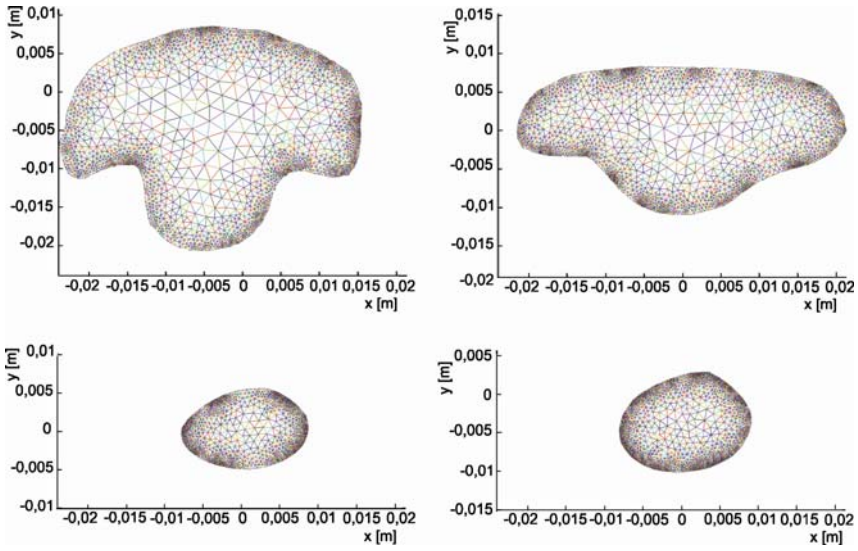
| distance [mm] | **[a:]** | | | **[i:]** | | | **[u:]** | | |
|---|---|---|---|---|---|---|---|---|---|
| | before | after | change | before | after | change | before | after | change |
| HL | 89 | 86.7 | -2.3 % | 89.7 | 90.5 | 0.8 % | 89.4 | 94.3 | 4.9 % |
| VL | 92.5 | 112.3 | 19.8 | 87.8 | 102.7 | 14.9 | 93.8 | 111.2 | 17.4 |
| TL | 181.5 | 199 | 17.5 | 177.5 | 193.2 | 15.7 | 183.2 | 205.5 | 22.3 |
| EP | 21 | 23 | 2 | 21 | 23 | 2 | 21 | 23 | 2 |

*Tab. 2: Areas measured in cross-sections of the lower pharynx and of the upper epilarynx and their ratios before and after warm-up for the vowels [a:], [i:] and [u:]. (MRI results)*

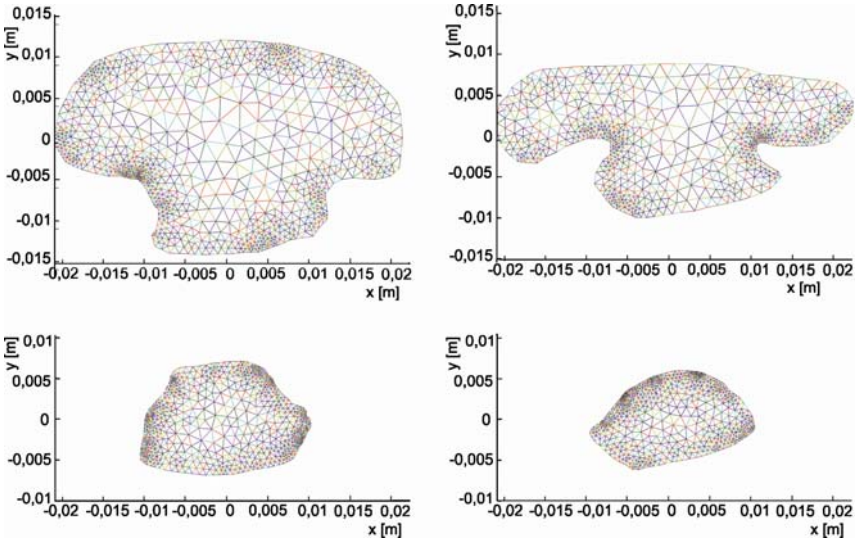
| area [$cm^2$] | **[a:]** | | **[i:]** | | **[u:]** | |
|---|---|---|---|---|---|---|
| | before | after | before | after | before | after |
| $A_{ph}$ (pharynx) | 3.959 | 3.430 | 7.736 | 5.761 | 7.966 | 5.440 |
| $A_{ep}$ (epilarynx) | 1.099 | 0.833 | 1.696 | 1.312 | 2.201 | 1.636 |
| ratio ($R=A_{ph}/A_{ep}$) | 3.60 | 4.12 | 4.56 | 4.39 | 3.62 | 3.33 |




vowel [a:]




vowel [i:]




vowel [u:]

*Fig. 3: Areas of the inlet of the low pharynx (upper panels) and the outlet of the epilarynx (lower panels) before (left) and after (right) voice exercising for phonation of vowels [a:], [i:], [u:].*

## 6. Modeling approach

The possible vocal tract changes resulting in the formation of a speaker's (or singer's) formant cluster were also studied using a 1D mathematical model of voice production (Radolf, 2010). The 1D vocal tract model was developed from the 3D volume model obtained from the MR images (Vampola et al., 2008). The formant frequencies measured from the vowels [a:], [i:] and [u:] recorded from the subject of the present study before and after exercising were prescribed to the model and by a tuning procedure (changing the vocal tract shape, i.e. the size of area cross-sections) the best fitting vocal tract configurations were obtained. The length of the real vocal tract of the subject was measured using MRI data and the values given in Table 1 were used for both the epilaryngeal and total length of the vocal tract model before and after vocal exercising. Vocal tract channel was modeled as a system of conical elements of 4 mm in length except for the first and the last element, whose length was modified according to the total length of vocal tract given by Table 1.

The model is based on an analytical solution of 1D wave equation for acoustic wave propagation in the vocal tract cavity (Radolf, 2010):

$$\frac{\partial^2\varphi}{\partial x^2}+\frac{1}{A}\frac{\partial A}{\partial x}\frac{\partial\varphi}{\partial x}-\frac{1}{c_0^2}\left(\frac{\partial^2\varphi}{\partial t^2}+c_0\,r_N\,\frac{\partial\varphi}{\partial t}\right)=0 \tag{1}$$

where ($\varphi$ is the flow velocity potential, $x$ is longitudinal coordinate along the vocal tract measured from the vocal folds to the lips, $t$ is time, $r_N$ is specific acoustic resistance per a unite length, $A(x)$ is the cross-sectional area of the cavity and $c_0$ is speed of sound.

Relation between the acoustic pressure $p$ and the volume velocity $W$ at the input and output of each conical acoustic element can be described by the transfer matrix as

$$\begin{bmatrix} p_{OUT} \\ W_{OUT} \end{bmatrix}=\begin{bmatrix} a & b \\ c & d \end{bmatrix}\cdot\begin{bmatrix} p_{IN} \\ W_{IN} \end{bmatrix}, \tag{2}$$

where the elements of the transfer matrix are

$$a=\frac{\xi_0}{\xi_0+L}\cdot\left(\cosh(\gamma L)+\frac{1}{\gamma\,\xi_0}\cdot\sinh(\gamma L)\right),\quad b=-\frac{z_0(r_N+j\,k)\cdot\xi_0}{A_{IN}\cdot\gamma(\xi_0+L)}\cdot\sinh(\gamma L),$$

$$c=A_{OUT}\cdot\frac{\left(1-\gamma^2\xi_0(\xi_0+L)\right)\cdot\sinh(\gamma L)-\gamma L\cdot\cosh(\gamma L)}{\gamma(\xi_0+L)^2\cdot z_0(r_N+j\,k)},$$

$$d=\frac{A_{OUT}}{A_{IN}}\frac{\xi_0}{\xi_0+L}\cdot\left(\cosh(\gamma L)-\frac{1}{\gamma(\xi_0+L)}\cdot\sinh(\gamma L)\right),$$

$L$ is length of the element, $A_{IN}$ and $A_{OUT}$ are the cross-sectional areas of the element input and output, respectively, $\gamma$ is a complex exponent given by the formulas:

$$\gamma=\alpha+j\beta,\quad \alpha=\frac{r_N}{\sqrt{2+2\cdot\sqrt{1+(r_N/k)^2}}},\quad \beta=\frac{k}{2}\cdot\sqrt{2+2\cdot\sqrt{1+(r_N/k)^2}}, \tag{3}$$

$k=\omega/c_0$ is the wave number, $\omega$ is angular frequency of harmonic signal, $j$ is imaginary unit: $j=\sqrt{-1}$. The coefficient $\xi_0$ is defined by input and output radius of the element $R_{IN}$ and $R_{OUT}$, respectively,

$$\xi_0=\frac{R_{IN}}{R_{OUT}-R_{IN}}\cdot L. \tag{4}$$

Frequency dependent viscous losses were considered as

$$r_N=\frac{1}{R}\cdot\sqrt{2k\mu/c_0\rho_0}, \tag{5}$$

where $\mu$ is dynamic air viscosity.

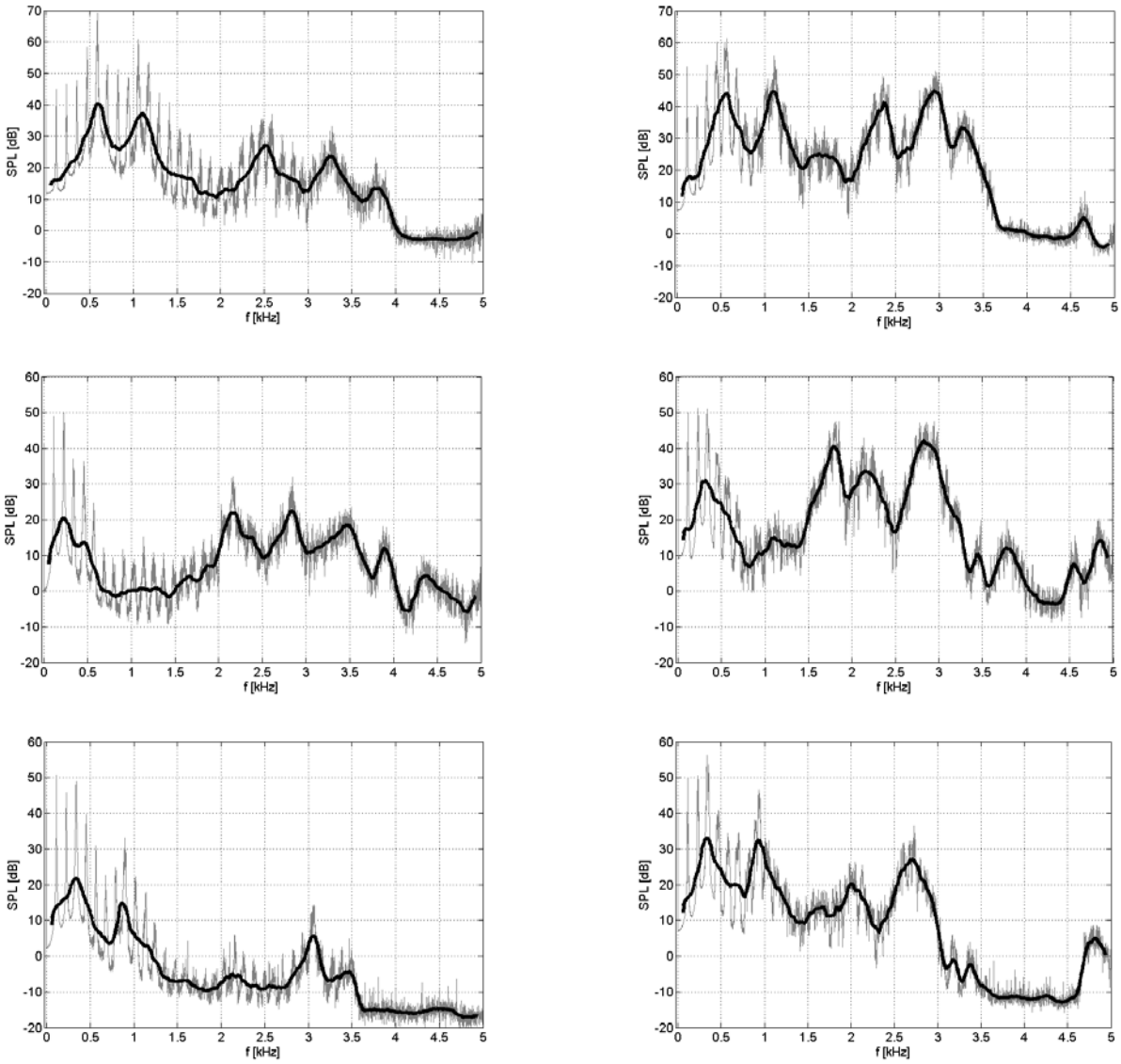



*Fig. 4: Spectra (LTAS – Rectangular window) of the acoustic signals for the vowels [a:] (top panels) , [i:] (middle) and [u:] (bootom) before (left) and after (right) the vocal exercise.*

Acoustic properties of the whole vocal tract can be described by matrix form

$$\begin{bmatrix} p_{LIP} \\ W_{LIP} \end{bmatrix} = \mathbf{T}_{VT} \cdot \begin{bmatrix} p_{GLOT} \\ W_{GLOT} \end{bmatrix}, \tag{6}$$

where $\mathbf{T}_{VT}$ is a transfer matrix obtained by multiplication of transfer matrices of all elements from the vocal folds to the lips

$$\mathbf{T}_{VT} = \begin{bmatrix} a_{VT} & b_{VT} \\ c_{VT} & d_{VT} \end{bmatrix} = \mathbf{T}_{N_e+1,N_e} \cdot \mathbf{T}_{N_e,N_e-1} \cdot \ldots \cdot \mathbf{T}_{3,2} \cdot \mathbf{T}_{2,1}, \tag{7}$$

and $N_e$ is number of conical elements.

Both for the calculation of the eigenfrequency and for the numerical simulation of the acoustic pressure at the lips we assume the output at the lips loaded by the acoustic radiation impedance of vibrating circular plate with radius *R* placed in an infinite wall (see e.g. Vampola *et al*., 2008)

$$Z_{A\,rad} = \frac{c_0 \rho_0}{\pi R^2} \cdot \left[ 1 - \frac{J_1(2kR)}{kR} + j \frac{H_1(2kR)}{kR} \right], \tag{8}$$

where $J_1$ is the Bessel function of the first kind of order 1 and $H_1$ is the Struve function of order 1 and $\rho_0$ is fluid density.

Since the radiation impedance gives the ratio of acoustic pressure and volume velocity $Z_{A\,rad} = p_{LIP}/W_{LIP}$, we get from (6) and (7)

$$\begin{bmatrix} p_{LIP} \\ p_{LIP}/Z_{A\,rad} \end{bmatrix} = \begin{bmatrix} a_{VT} & b_{VT} \\ c_{VT} & d_{VT} \end{bmatrix} \cdot \begin{bmatrix} p_{GLOT} \\ W_{GLOT} \end{bmatrix}. \tag{9}$$

In the eigenfrequency calculations, the boundary condition: $W_{GLOT} = 0$ is assumed at the vocal folds, i.e. a closed input by a rigid wall, which gives the following frequency equation:

$$a_{VT} - Z_{A\,rad} \cdot c_{VT} = 0\,. \tag{10}$$

Directly from the system of equations (9), we obtain the acoustic pressure at the lips linearly dependent on the excitation velocity signal at the vocal folds:

$$p_{LIP} = \frac{a_{VT} \cdot d_{VT} - b_{VT} \cdot c_{VT}}{a_{VT} - Z_{A\,rad} \cdot c_{VT}} \cdot Z_{A\,rad} \cdot W_{GLOT}\,. \tag{11}$$

*Tab. 3: Fundamental frequency (F0) and the formant frequencies (F1 - F5) evaluated from the acoustic recordings and obtained from the modeling for the vowels [a:], [i:] and [u:] before and after the vocal exercise.*

| **[a:]** | $F_0$ [Hz] | $F_1$ [Hz] | $F_2$ [Hz] | $F_3$ [Hz] | $F_4$ [Hz] | $F_5$ [Hz] |
|---|---|---|---|---|---|---|
| Acoustic signal – before exercising | 118 | 600 | 1100 | 2510 | 3260 | 3780 |
| Model – before exercising | random | 600 | 1099 | 2507 | 3254 | 3787 |
| Acoustic signal - after exercising | 112 | 560 | 1100 | 2370 | 2940 | 3260 |
| Model - after exercising | random | 560 | 1099 | 2367 | 2935 | 3268 |
| **[i:]** | | | | | | |
| Acoustic signal – before exercising | 114 | 230 | 2170 | 2840 | 3470 | 3890 |
| Model – before exercising | random | 230 | 2171 | 2831 | 3473 | 4543* |
| Acoustic signal-after exercising | 114 | 310 | 1790 | 2160 | 2820 | 3770 |
| Model - after exercising | random | 310 | 1783 | 2162 | 2826 | 4166* |
| **[u:]** | | | | | | |
| Acoustic signal – before exercising | 114 | 340 | 860 | 2150 | 3060 | - |
| Model – before exercising | random | 343 | 856 | 2149 | 3052 | 4783* |
| Acoustic signal - after exercising | 115 | 340 | 940 | 2010 | 2700 | - |
| Model - after exercising | random | 340 | 940 | 2003 | 2707 | 4160* |

* Formants resulted from modeling, but were not prescribed for tuning where only $F_1$-$F_4$ were used.

The sensitivity of a particular eigenfrequency $f_n$ to a change in cross-sectional area $A_i$ in dimensionless form:

$$ {}^{A}S_{i,n} = \frac{\Delta f_n / f_n}{\Delta A_i / A_i} \tag{12} $$

can be substituted by the following relation (see Radolf (2010)):

$$ {}^{A}\tilde{S}_{i,n} = \frac{1}{2}\left\{ L_{i-1}\partial_i^{i-1} \cdot \left[ \frac{\rho_0}{2A_{i-1\,red}^2}\left( W_{i-1,n}^{\,2} + W_{i,n}^{\,2} \right) - \frac{1}{\rho_0 c_0^2} p_{i-1,n}^2 \right] \right. $$

$$ \left. + L_i \partial_i^{i} \cdot \left[ \frac{\rho_0}{2A_{i\,red}^2}\left( W_{i,n}^{\,2} + W_{i+1,n}^{\,2} \right) - \frac{1}{\rho_0 c_0^2} p_{i,n}^2 \right] \right\} \cdot \frac{A_i}{\sum_{e=2}^{N_e+1} m_e W_{e,n}^{\,2}}, \tag{13} $$

where $A_{i\,red}$ is a reduced cross-sectional area of an $i$-th conical element of length $L_i$

$$ A_{i\,red} = \frac{1}{3}\left( A_i + \sqrt{A_i A_{i+1}} + A_{i+1} \right), \tag{14} $$

$m_i$ is an acoustic mass lumped into $i$-th cross-section:

$$ m_i = \frac{\rho_0}{2}\left( \frac{L_{i-1}}{A_{i-1\,red}} + \frac{L_i}{A_{i\,red}} \right), \tag{15} $$

$\partial_i^{i-1}$ and $\partial_i^{i}$ are the partial derivatives of the reduced cross-section

$$ \partial_i^{i-1} = \frac{\partial A_{i-1\,red}}{\partial A_i} = \frac{1}{3}\cdot\left( 1 + \frac{1}{2}\sqrt{\frac{A_{i-1}}{A_i}} \right), \qquad \partial_i^{i} = \frac{\partial A_{i\,red}}{\partial A_i} = \frac{1}{3}\cdot\left( 1 + \frac{1}{2}\sqrt{\frac{A_{i+1}}{A_i}} \right), \tag{16} $$

$N_e$ is number of conical elements, $p_{i,n}$ resp. $W_{i,n}$ is an amplitude of acoustic pressure resp. volume velocity in the *i*-th cross-section for the *n*-th eigenfrequency $f_n$.

Radolf (2010) derived that the dimensionless sensitivity of a particular eigenfrequency $f_n$ to a change of the length $L_i$ of the i-th element:

$$ {}^{L}S_{i,n} = \frac{\Delta f_n / f_n}{\Delta L_i / L_i}, \tag{17} $$

can be substituted by the following relation

$$ {}^{L}\tilde{S}_{i,n} = -\frac{1}{2}\left( \frac{A_{i\,red}}{\rho_0 c_0^2} p_{i,n}^2 + \frac{\rho_0}{2A_{i\,red}}\left( W_{i,n}^2 + W_{i+1,n}^2 \right) \right) \cdot \frac{L_i}{\sum_{e=2}^{N_e+1} m_e W_{e,n}^2}. \tag{18} $$

Based on the results of Story (2006) and using the sensitivity functions (13) and (18) an iterative process was developed for computation of *i*-th cross-sectional area by the formula:

$$ \left( A_i \right)_{k+1} = \left( A_i \right)_k \cdot \left( 1 + \sum_{n=1}^{N_F} \left( z_n \cdot {}^{A}\tilde{S}_{i,n} \right)_k \right), \tag{19} $$

and similarly for computation of *i*-th element length the following formula was derived

$$ \left( L_i \right)_{k+1} = \left( L_i \right)_k \cdot \left( 1 + \sum_{n=1}^{N_F} \left( z_n \cdot {}^{L}\tilde{S}_{i,n} \right)_k \right), \tag{20} $$

where $N_F$ is number of the prescribed formant frequencies, $z_n$ is a function of the difference between the desired eigenfrequency $f_{n,d}$ and the instantaneous *n*-th eigenfrequency $f_{n,k}$ in the *k*- th step

$$z_n = \alpha \cdot \frac{f_{n,d} - f_{n,k}}{f_{n,k}} , \tag{21}$$

$\alpha$ is an additional coefficient that can accelerate the iterative process and it was typically set to $\alpha = 10$.

The iteration procedure stopped when the root of the sum of the squared differences between desired and instantaneous eigenfrequencies

$$\delta = \sqrt{\sum \left( f_{nd} - f_{nk} \right)^2} \tag{22}$$

was less than a desired tolerance value. This value was set in the computations to 10 Hz in all cases.

First five formants F1 – F5 were used for modeling of the vowel [a:] and first four formants F1 – F4 for modeling the vowels [i:] and [u:]. The speed of sound, the density, and the dynamic viscosity of the air were considered as follows: $c_o$= 353 ms$^{-1}$; $\rho_o$ = 1.2 kgm$^{-3}$, $\mu$ = 1.8·10$^{-5}$ kgm$^{-1}$s$^{-1}$. The number of iteration steps needed for computation of the geometry of the vocal tract for the vowel [a:] was 65 for before and 187 for after the warm up, for the vowel [i:] 34 iterations for before and 35 for after, and for the vowel [u:] it was 24 iterations for before and 26 iterations for after warm up. The numerical simulation of the acoustic pressure at the lips was performed in the frequency domain by random excitation of the vocal tract at the level of the vocal folds. Computational time was from 12s to 31s on PC Intel Core i5-670, 3.47 GHz, 3.49 GB RAM.

Figures 5-7 show the computed vocal tract shapes, the cross-sectional areas and the spectra of the numerically simulated acoustic signals at the lips for all three vowels before and after the warm-up.

## 7. Comparison of the audio and MRI results with modeling

The formant frequencies resulting from the optimization procedure are compared with the prescribed formants obtained from the acoustic analyses in Table 3. All formant frequencies after the optimization procedures are within the prescribed limit of 10 Hz.

The SPL values resulted from modeling are compared with the measured acoustic data in Table 4. The SPL values in the total frequency region 0–5 kHz analyzed have the same tendency in the measurements and in modeling, in both cases the $SPL_{\text{total}}$ values decreased slightly after the vocal exercising for vowel [a:] and increased for the vowels [i:] and [u:]. The SPL values ($SPL_{\Delta sf}$) measured within the singer´s formant clustering frequency range $\Delta_{sf}$ increased for vowel [a:] by 15.2 dB and for the vowels [i:] and [u:] by 18.0 dB and 21.5 dB, respectively. Corresponding computed $SPL_{\Delta sf}$ values increased after the warm up by 0.6 dB, 2.2 dB and 6.7 dB.

The singer's formant clustering in the intervals $\Delta_{SF}$ between the formants F3 and F5 for vowel [a:], between the formants F2 and F4 for vowel [i:] and between F3 and F4 for vowel [u:] corresponds well with the measured frequency bands (compare the measured acoustic spectra in Fig. 4 with the spectra in Figs. 5–7 for modeling). After the vocal exercising the formants $F_1$ and $F_2$ of [a:] remained nearly unchanged while the formants $F_3$–$F_5$ were clustering, thereby increasing the SPL level in the frequency band of 2-4 kHz. The distance between $F_3$ and $F_5$ in the acoustic measurement decreased from 1270 Hz down to 890 Hz, and in the modeling the distance between these formants decreased from 1280 Hz to 901 Hz (see Table 4). Similarly for vowel [i:] the distance between $F_2$ and $F_4$ decreased from 1300 Hz to 1030 Hz in the measurement and from 1303 Hz to 1041 Hz in modeling. For vowel [u:] the distance between F3 and F4 decreased from 910 Hz down to 690 Hz in the measurement and from 903 Hz to 703 Hz in modeling. Thus, the singer's formant clustering frequency bandwidths $\Delta_{SF}$ in modeling decreased substantially and in good agreement with the acoustic measurements, the decrease is about 380 Hz for the vowel [a:], 270 Hz for vowel [i:] and 220 Hz for vowel [u:] (see Table 4).

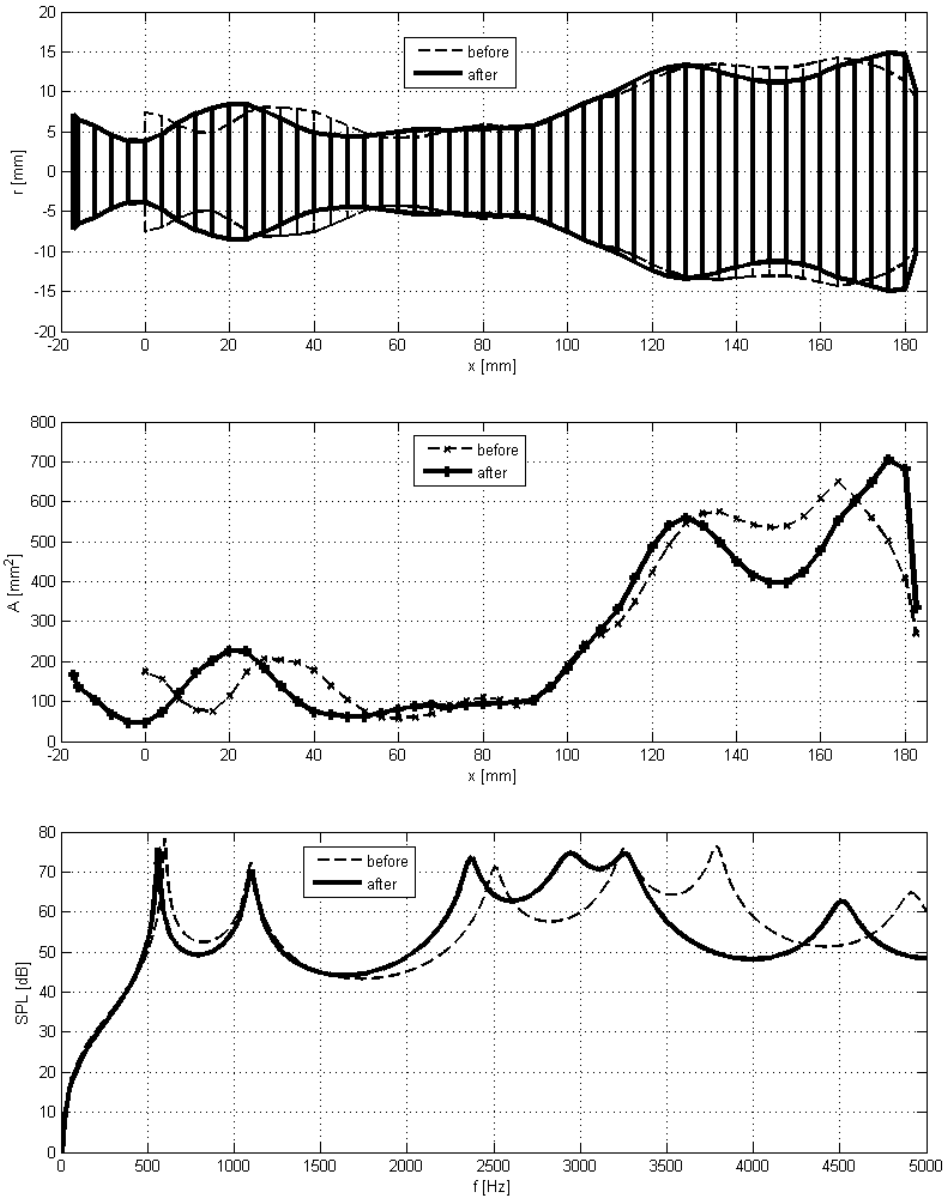



*Fig. 5: Vocal tract geometry (top panel - from the vocal folds on the left to the lips on the right), cross-sectional areas along the vocal tract (middle panel) and acoustic pressure at the lips (lower panel) – results of modeling for vowel [a:] before and after the vocal exercise.*

The geometrical changes resulting from modeling (see Figs. 5–7) can be compared with the MRI midsagittal slices shown in Fig. 1. The most important change in the vocal tract geometry for vowel [a:] was the prolongation of the vocal tract by lowering of the larynx. The changes in the oral cavity that can be realized by changes of the tongue position and/or by the lips opening are not well visible in the MR images, because the subject had some small metal piece in his lower jaw and therefore the images near the lower lip are of poor quality and the comparison between the modeling and MRI in this vocal tract region is problematic. For the vowels [i:] and [u:], the most important changes were also the prolongation of the vocal tract at the vocal folds end due to a downward shifting of the larynx and the lower pharynx. The main change in the vocal tract geometry for the vowel [i:] was the increase of the vocal tract volume in the upper pharynx (compare Figs. 1 and 6).

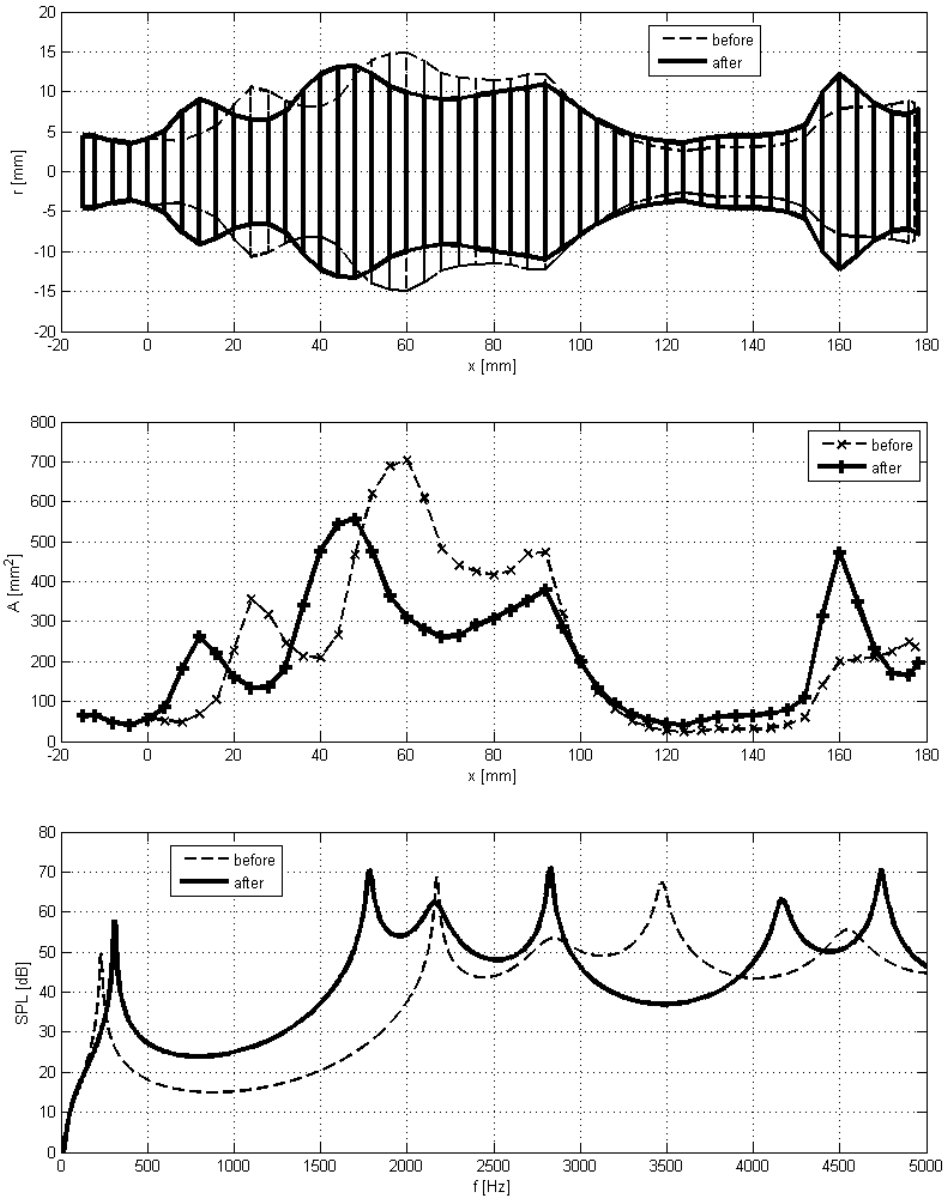



*Fig. 6: Vocal tract geometry (top panel - from the vocal folds on the left to the lips on the right), cross-sectional areas along the vocal tract (middle panel) and acoustic pressure at the lips (lower panel) – results of modeling for vowel [i:] before and after the vocal exercise.*

The area ratio $R=A_{ph}/A_{ep}$ of the vocal tract cross-sections ( the inlet of the lower pharynx over the outlet of the epilarynx) for the model were measured from the computed areas shown in Figs. 5 – 7 taking into account the epilarynx length (EP) specified in Table 1. The area ratios are summarized in Table 5 for all three vowels and compared with the ratios obtained from the volume model of the vocal tract (see Fig. 2 and Table 2). Results from modeling were from 1.82 to 5.02 for phonation before and from 2.35 to 4.10 for phonation after the exercising. The area ratios measured from the MRI, where from 3.60 to 4.56 for phonation before and from 3.33 to 4.39 for phonation after exercising. It should be noted that an exact quantitative comparison of the area ratios measured from the MRI and modeling is problematic, because the ratio value is sensitive to the specifically chosen cross-sections. However, the tendency in the changes of the ratio before and after the vocal exercising is the same for modeling and MRI measurements, the increase of the area ratio for the vowel [a:] and the decrease for the vowels [i:] and [u:] – see $R_{after}/R_{before}$ in Table 2. The area ratios measured and modeled for all vowels after exercising were found to be between 2.35 and 4.39 (see Table 5). These values are lower than the hypothetical ratio 6 considered being optimal for the establishment of a singer's formant cluster by

Sundberg and Fant (Sundberg, 2003), however the present results are in agreement with the results by Detweiler (1994), where the ratio was found between 2.9 and 3.7.

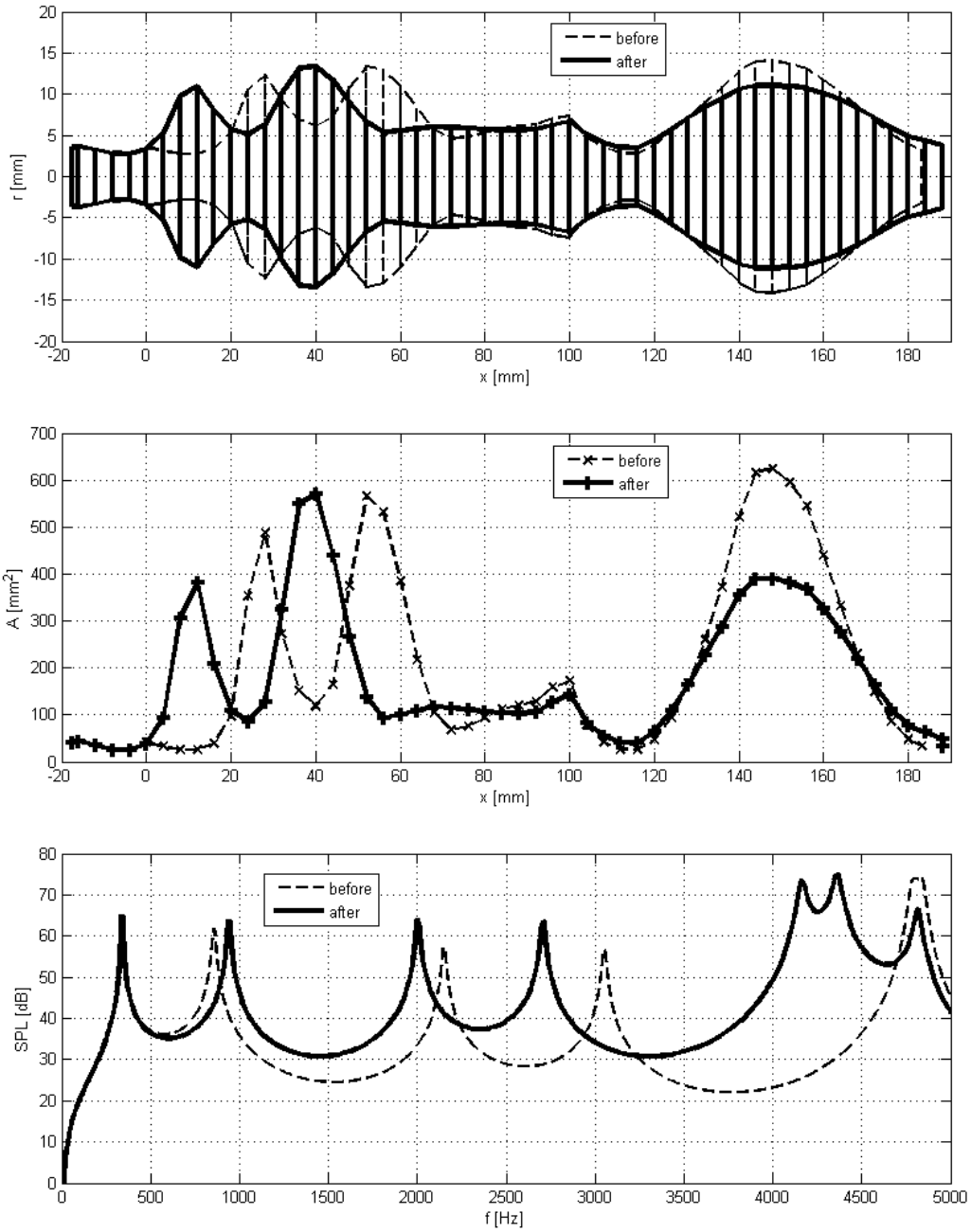



*Fig. 7: Vocal tract geometry (top panel - from the vocal folds on the left to the lips on the right), cross-sectional areas along the vocal tract (middle panel) and acoustic pressure at the lips (lower panel) – results of modeling for vowel [u:] before and after the vocal exercise.*

*Table 4: Comparison of the measured and modeled results for the vowels [a:], [i:] and [u:] before and after the vocal exercise: (a) the frequency distance $\Delta_{SF}$ between the formants important for creation the singer's formant cluster, (b) the SPL computed in the frequency bandwidth $\Delta_{sf}$ of the singer's formant cluster ( $SPL_{\Delta_{sf}}$ ) and (c) the $SPL_{total}$ computed in the total frequency range considered.*

| **[a:]** | | $\Delta_{SF} = F_5 - F_3$ [Hz] | $SPL_{\Delta_{sf}}$ [dB] $\Delta_{sf}$ = 2-4 kHz | $SPL_{total}$ [dB] 0 – 5 kHz |
|---|---|---|---|---|
| before exercising | acoustic signal | 1270 | 55 | 77.8 |
| | modeling | 1280 | 128.2 | 131.5 |
| after exercising | acoustic signal | 890 | 70.2 | 76.9 |
| | modeling | 901 | 128.8 | 131.4 |
| differences | acoustic signal | -380 | 15.20 | -0.90 |
| after - before | modeling | -379 | 0.60 | -0.10 |

| **[i:]** | | $\Delta_{SF} = F_4 - F_2$ [Hz] | $SPL_{\Delta_{sf}}$ [dB] $\Delta_{sf}$ = 1.4-4 kHz | $SPL_{total}$ [dB] 0 – 5 kHz |
|---|---|---|---|---|
| before exercising | acoustic signal | 1300 | 50.8 | 59.3 |
| | modeling | 1302 | 115.3 | 118.0 |
| after exercising | acoustic signal | 1030 | 68.8 | 70.2 |
| | modeling | 1043 | 117.5 | 122.0 |
| differences | acoustic signal | -270 | 18.0 | 10.9 |
| after - before | modeling | -259 | 2.2 | 4.0 |

| **[u:]** | | $\Delta_{SF} = F_4 - F_3$ [Hz] | $SPL_{\Delta_{sf}}$ [dB] $\Delta_{sf}$ = 1.8-3.5 kHz | $SPL_{total}$ [dB] 0 – 5 kHz |
|---|---|---|---|---|
| before exercising | acoustic signal | 910 | 31.3 | 60.3 |
| | modeling | 903 | 99.1 | 116.6 |
| after exercising | acoustic signal | 690 | 52.8 | 68.4 |
| | modeling | 704 | 105.8 | 122.7 |
| differences | acoustic signal | -220 | 21.5 | 8.1 |
| after - before | modeling | -199 | 6.7 | 6.1 |

*Table 5: Comparison between the MRI measurement and the modeled cross-sectional area ratios measured inside the vocal tract at the lower pharynx ($A_{ph}$) and upper epilarynx ($A_{ep}$) for the vowels [a:], [i:] and [u:] before and after the vocal exercise.*

| **Area ratios** $R = A_{ph}$ [mm$^2$] / $A_{ep}$ [mm$^2$] | | **[a:]** | **[i:]** | **[u:]** |
|---|---|---|---|---|
| before exercising - $R_{before}$ | modeling | 207/114=1.82 | 356/104=3.42 | 487/97=5.02 |
| | MRI | 3.60 | 4.56 | 3.62 |
| after exercising - $R_{after}$ | modeling | 228/97=2.35 | 262/86=3.05 | 381/93=4.10 |
| | MRI | 4.12 | 4.39 | 3.33 |
| $R_{after}/R_{before}$ | modeling | 1.29 | 0.89 | 0.82 |
| | MRI | 1.14 | 0.96 | 0.92 |

## 8. Conclusions

The area ratios of the lower pharynx over the epilaryngeal tube were lower than earlier hypothesized for the male singers' voice. The results suggest that the origin of the singer's formant is not necessarily associated with an increase of the pharyngeal over the epilaryngeal cross-sectional area ratio but can also be obtained by other geometrical changes of the vocal tract cavity. The conclusions should be considered with caution because all results are only for one subject investigated.

**Acknowledgement**

The research was supported by the projects GAČR P101/12/P579 (modeling part) and GAČR P101/12/1306 (experimental part). The authors are also very grateful to Doc. MUDr. Petr Krupa from the Hospital U Svaté Anny in Brno for enabling the MRI measurements.

## References

Detweiler, R.F. (1994) An investigation of the laryngeal system as the resonance source of the singer's formant. *Journal of Voice,* 8, 4, pp.303-313

Fant, G. (1960) *Acoustic theory of speech production*. Mouton, S'Gravenage, 2nd ed.

Laukkanen, A-M., Horáček, J., Krupa, P. & Švec, J.G. (2011) The effect of phonation into a straw on the vocal tract adjustments and formant frequencies. A preliminary MRI study on a single subject completed with acoustic results. *Biomedical Signal Processing and Control*. DOI 10.1016/j.bspc.2011.02.004.

Leino, T. (1994) Long-term average spectrum study on speaking voice quality in male actors, in: SMAC93, *Proceedings of the Stockholm Music Acoustics Conference*, (A. Friberg, J. Iwarsson, E. Jansson & J. Sundberg eds) The Royal Swedish Academy of Music, No 79, Stockholm, pp.206-210.

Leino, T., Laukkanen, A.M. & Radolf, V. (2011) Formation of the Actor's/Speaker's Formant: A Study Applying Spectrum Analysis and Computer Modeling. *Journal of Voice*, 25(2), pp. 150-158.

Radolf, V. (2010) Direct and inverse task in acoustics of the human vocal tract. *PhD . thesis*, Czech Technical University in Prague, 95 p.

Story, B.H. (2003) Using imaging and modeling techniques to understand the relation between vocal tract shape to acoustic characteristics, in: *Proc. of the Stockholm Music Acoustics Conf.* (SMAC 03), Stockholm, Sweden, pp. 435-438.

Story, B.H. (2006) Technique for "tuning" vocal tract area functions based on acoustic sensitivity functions, *Journal of the Acoustical Society of America,* 119(2), pp. 715-718.

Sundberg, J. (1974) Articulatory interpretation of the "singing formant". *Journal of the Acoustical Society of America,* 55 (4) pp. 838-844.

Sundberg, J. (2003) Research on the singing voice in retrospect. *Speech, Music and Hearing*, *TMH-QPSR,* Vol. 45, KTH Stockholm, pp. 11-22.

Titze, I.R. (2000) *Principles of voice production*. National Center for Voice and Speech, Iowa City, IA

Titze, I.R. & Story, B.H. (1997) Acoustic interactions of the voice source with the lower vocal tract, *Journal of the Acoustical Society of America,* 101, 4, pp. 2234–2243.

Vampola, T., Horáček, J. & Švec, J.G. (2008) FE modeling of human vocal tract acoustic. Part I: Production of Czech vowels. *Acta Acoustica united with Acoustica*, 94, pp. 433-447.

Vampola, T., Horáček, J., Vokřál, J. & Černý, L. (2008) FE modeling of human vocal tract acoustics. Part II. Influence of velopharyngeal insufficiency on phonation of vowels. *Acustica United with Acta Acustica*, 94, pp. 448-460.