Finite element modelling of vocal tract changes after voice therapy

T. Vampola[a,*], A. M. Laukkanen[b], J. Horáček[c], J. G. Švec[d]

[a]*Department of Mechanics, Biomechanics and Mechatronics, Faculty of Mechanical Engineering, Czech Technical University in Prague, Karlovo nám. 13, 121 35 Prague 2, Czech Republic*

[b]*Department of Speech Communication and Voice Research, University of Tampere, FIN 33014 Tampere, Finland*

[c]*Institute of Thermomechanics, Academy of Sciences of the Czech Republic, Dolejškova 5, 182 00 Praha 8, Czech Republic*

[d]*Department of Experimental Physics, Faculty of Sciences Palacký University Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic*

Abstract

Two 3D finite element (FE) models were constructed, based on CT measurements of a subject phonating on [a:] before and after phonation into a tube. Acoustic analysis was performed by exciting the models with acoustic flow velocity at the vocal folds. The generated acoustic pressure of the response was computed in front of the mouth and inside the vocal tract for both FE models. Average amplitudes of the pressure oscillations inside the vocal tract and in front of the mouth were compared to display the cost-efficiency of sound energy transfer at different formant frequencies. The formants F1–F3 correspond to classical vibration modes also solvable by 1D vocal tract model. However, for higher formants, there occur more complicated transversal modes which require 3D modelling. A special attention is given to the higher frequency range (above 3.5 Hz) where transversal modes exist between piriform sinuses and valleculae. Comparison of the pressure oscillation inside and outside the vocal tract showed that formants differ in their efficiency, F4 (at about 3.5 kHz, i.e. at the speaker's or singer's formant region) being the most effective. The higher formants created a clear formant cluster around 4 kHz after the vocal exercise with the tube. Since the human ear is most sensitive to frequencies between 2 and 4 kHz concentration of sound energy in this frequency region (F4–F5) is effective for communication. The results suggest that exercising using phonation into tubes help in improving the vocal economy.

1. Introduction

Earlier observations have shown that sound pressure level (SPL) tends to increase after vocal exercising on semi-occlusions like voiced fricatives and phonation into a tube [1]. Applying computer simulation with a self-oscillating vocal fold model and an interactive vocal tract model, Titze [4] concluded that a semi-occlusion at the lips strengthens the interaction between the voice source and the vocal tract by raising the mean supraglottal and intraglottal pressures. Impedance matching by a sufficient vocal fold adduction and narrowing of the epilaryngeal tube may improve the efficiency and economy of voice production. Based on modelling results it was hypothesized [5] that an artificial extension of the vocal tract leads to a more efficient and economic phonation, especially if narrowing of the epilaryngeal region occurs at the same time. This in turn is also prone to lead to formant clustering around 3–3.4 kHz [6].

---

*Corresponding author. Tel.: +420 224 357 244, e-mail: tomas.vampola@fs.cvut.cz.

These and other studies suggest that the phonation into a tube leads to beneficial effects in voice production. What has not been clear, however, is what changes remain in the vocal tract after the tube has been removed. To address these questions, the Computerized Tomography (CT) methods was recently used to visualize the shape of the vocal tract of a female subject before, during and after phonation into a tube [8]. The CT results showed clear changes of the vocal tract after phonation into a tube. These changes include widening the frontal part of the oral cavity and of the lower pharynx (just above the epiglottis), and narrowing in the region between the lower part of the tongue body and the back wall of the pharynx. Acoustic recordings showed slightly lowered formant frequencies F1, F2, F4 and F5 and slightly raised formant F3. The overall sound pressure level (SPL) of voice increased by 3 dB.

The present study investigates the effects of the vocal tract changes on the sound energy transfer through the vocal tract. Three-dimensional (3D) finite-element (FE) models of the vocal tract created from the CT measurements were used to investigate the changes in the transfer function of the vocal tract and in the acoustic pressure radiated out of the mouth. The investigations aimed at determining the acoustic consequences of the changes of the supra-glottal spaces before and after phonation into the tube.

2. CT data and acoustic data acquisition

The CT data were obtained using the Light Speed VCT GE – 64 (General Electric) device. A female subject volunteered to be the experimental subject. She had an extensive experience in voice training and in using the 'resonance tubes'. She signed a consent form allowing the CT examination to be performed. For the measurement details, see [8]. The midsagittal and corresponding coronal slices obtained from the CT measurements are shown in Fig. 1 for phonation the vowel [a:] before and after phonation into the resonance tube (made of glass, length 28 cm, inner diameter 7 mm). The airways of the vocal tract are clearly distinguishable from the surrounding tissues here.
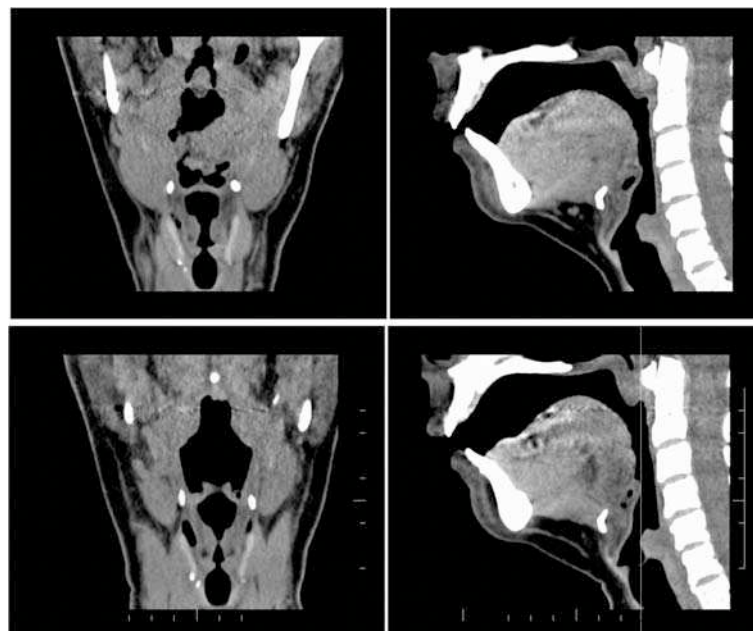

Fig. 1. Coronal (on the left) and midsagittal (on the right) images of the vocal tract for the vowel [a:] obtained from the CT scannings performed before (upper panel) and after (lower panel) phonating into the tube

When performing the examination, the subject was given a sign through the CT intercom system to start phonating. First, the subject produced the vowel [a:] for c. 4 seconds, then took a breath and produced the vowel [a:] again for c. 4 seconds. The first phonation was used as a sign of the subject's readiness. The CT scanning occurred immediately after the subject started the second phonation (see Fig. 2).
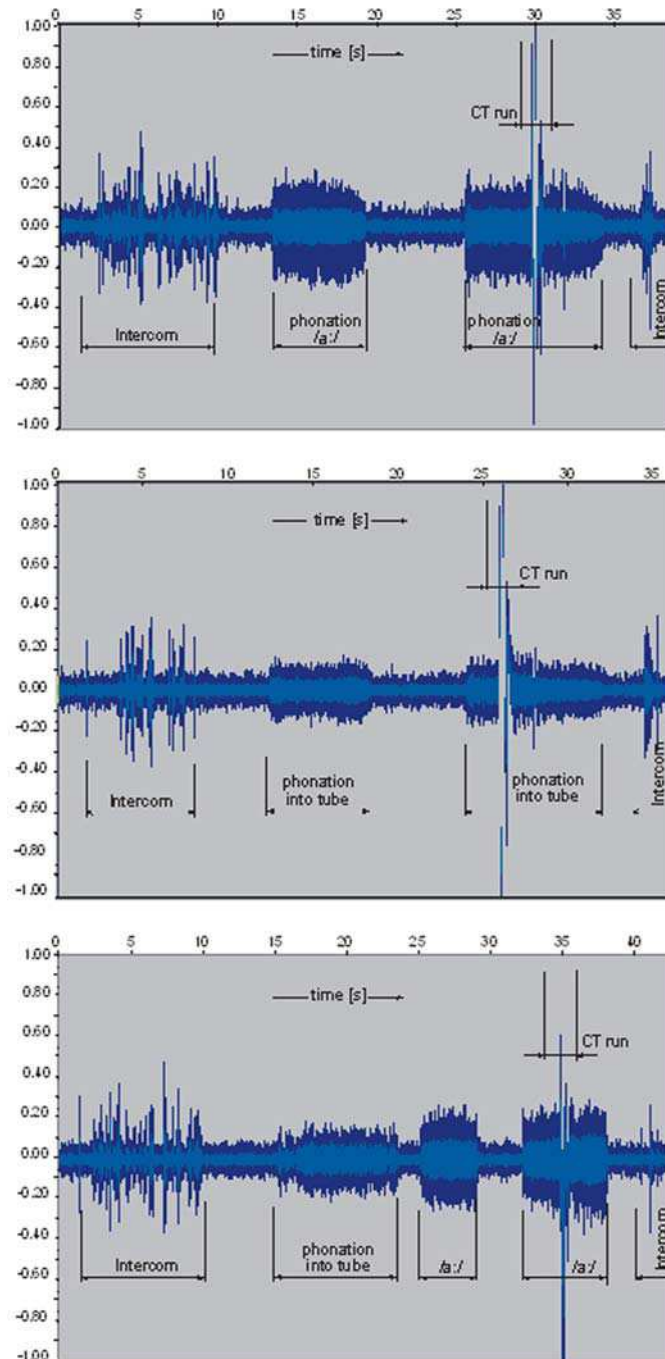

Fig. 2. Audio signal samples measured during phonation in the CT device: phonation /a:/ before phonation into the tube (upper panel), phonation into the tube (middle panel), phonation /a:/ after the tube (bottom panel)

A relatively high background noise level in the room with the CT apparatus is also visible in the figure. This procedure was repeated twice before phonation into the tube, twice during phonation into the tube and twice after phonation into the tube. For the after-tube examination, the subject performed the phonation into the tube first and then immediately continued by phonating without the tube.

3. Analysis of acoustic data

The analysis of the formant frequencies was done using the LPC analysis that yielded formant values, fundamental frequency (F0) and sound pressure level (SPL). The histograms from the LPC-obtained formant frequencies for the acoustic recordings before and after the phonation into the tube are shown in Fig. 3. The figure reveals that the largest changes occurred in the 3rd formant, which increased by c. 150 Hz after the tube-phonation. The sound pressure level (SPL) was c. 2.4 dB higher after than before the phonation into the tube. The fundamental frequency F0$\cong$171 Hz decreased on average by c. 9.5 % after the phonation into the tube, the third formant increased by c. 7.5 % and the first formant by c. 3 %. The changes in the other formants were rather small.
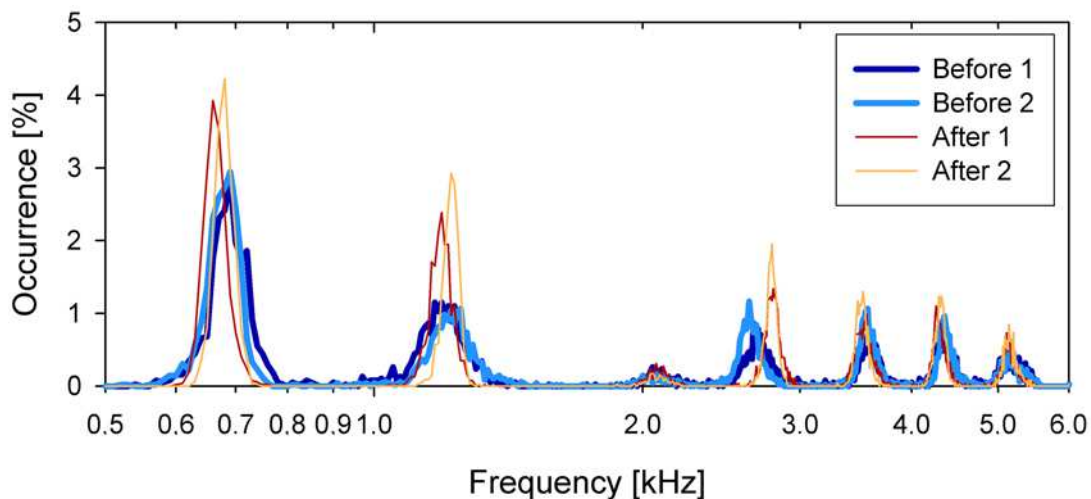

Fig. 3. LPC formant histogram for the audio recordings before and after the phonation into the tube

4. Volume and finite element models

The recorded CT images were first segmented and processed into 3D volume models of the vocal tract. The developed volume models of the vocal tract for the vowel /a:/ positioned into the backbones of the skeleton and mandible are shown in Fig. 4 for two cases, i.e. before and after phonation into the tube. No large changes in the position of the larynx related to the vertebrae before and after phonation into the tube can be visible, but a wider mouth opening after phonation into the tube can be recognized here.

The volume models of the vocal tract for phonation before and after tube-phonation are compared in Fig. 5, where the epilaryngeal airways, the piriform sinuses and the valleculae are clearly visible. The shape of the vocal tract appears to be visually similar before and after tube-phonation with one major difference: before the tube-phonation the velar passage port to the
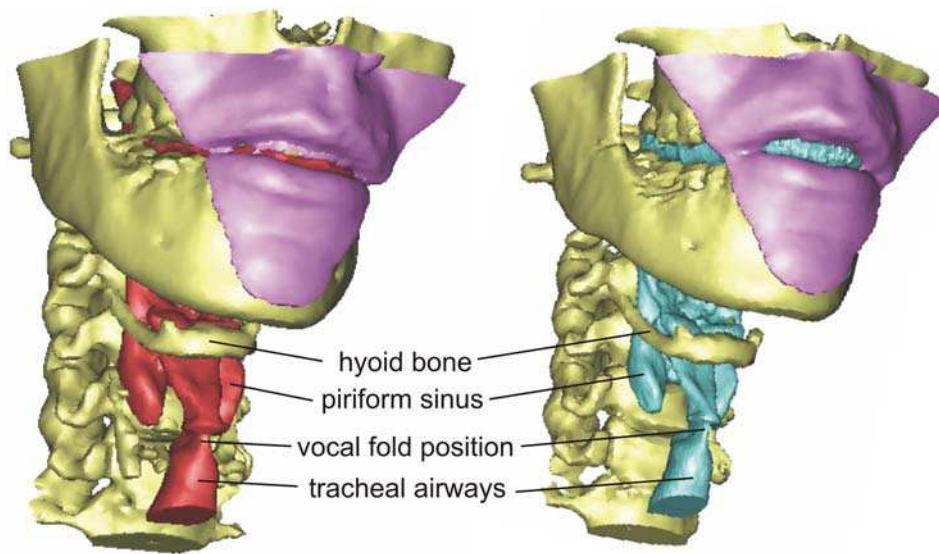
Fig. 4. Volume models of the vocal tract positioned in the mouth and neck bones during phonation of vowel /a:/ before (on the left) and after (on the right) phonation into the tube
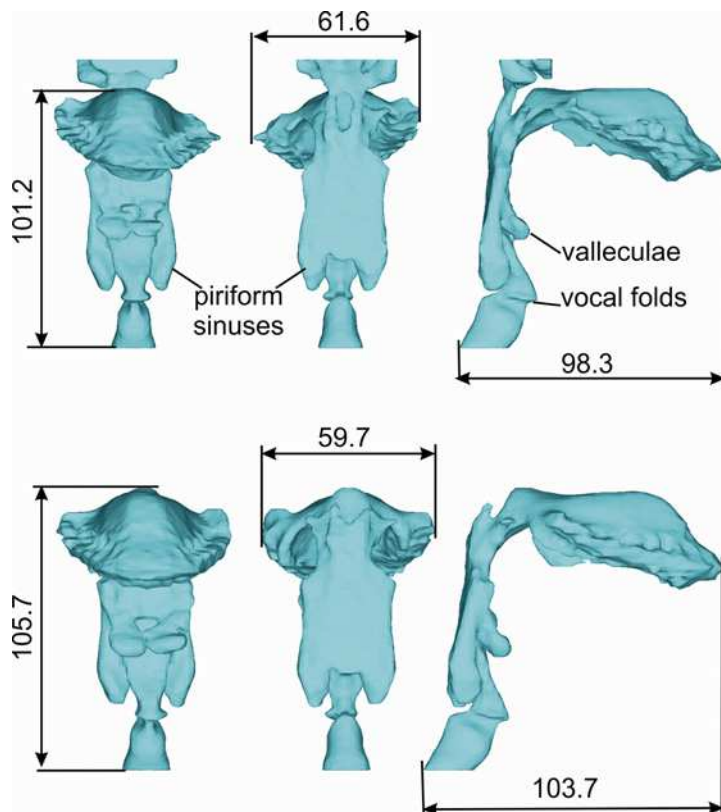

Fig. 5. Volume models of the vocal tract for phonation on [a:] before and after phonation into the tube

nasopharynx was open, whereas after the tube-phonation it was closed. There are no remarkable changes of the airways in the glottal region near the vocal folds.

The total volume of the acoustic vocal tract spaces was considerably larger after the tube-phonation; it increased by 38.5 %. The total length of the vocal tract after the tube-phonation was increased by 4.2 %. The volume of the valleculae and piriform sinuses increased on average

by c. 64 % and 7 %, respectively. The volume of the epilaryngeal tube did not show much difference after the phonation into the tube.

After meshing the volume models, two 3D finite element (FE) models of the vocal tract were constructed for phonation on [a:] before and after tube-phonation. The models were created within the framework of the code ANSYS. Since the CT scans did not provide complete information on the nasal tract, in the case of velopharyngeal opening (recall Fig. 5), the connection of the supraglottal and nasal spaces had to be disregarded and the passage was artificially closed. The FE models consisted of about 125 000 and 175 000 acoustic finite elements for phonation before and after the tube, respectively.

In order to simulate the radiation of the sound out of the mouth, the head was modelled approximately as a sphere with a diameter of about 15 cm and the surrounding air was modelled as a set of finite elements filling in the space between the head and an outer sphere at the distance of c. 10 cm from the head (Fig. 6). The outer sphere was modelled with infinite, non-reflecting elements FLUID130. The following material parameters of the acoustic tetrahedral finite elements FLUID30 were considered: speed of sound 350 m/s, air density 1.2 kg/m3 and the dimensionless sound absorption coefficient 0.0005 at the walls of the acoustic spaces. The acoustic space between the spheres was modelled by about 515 000 finite elements with an average element size of c. 5 mm.
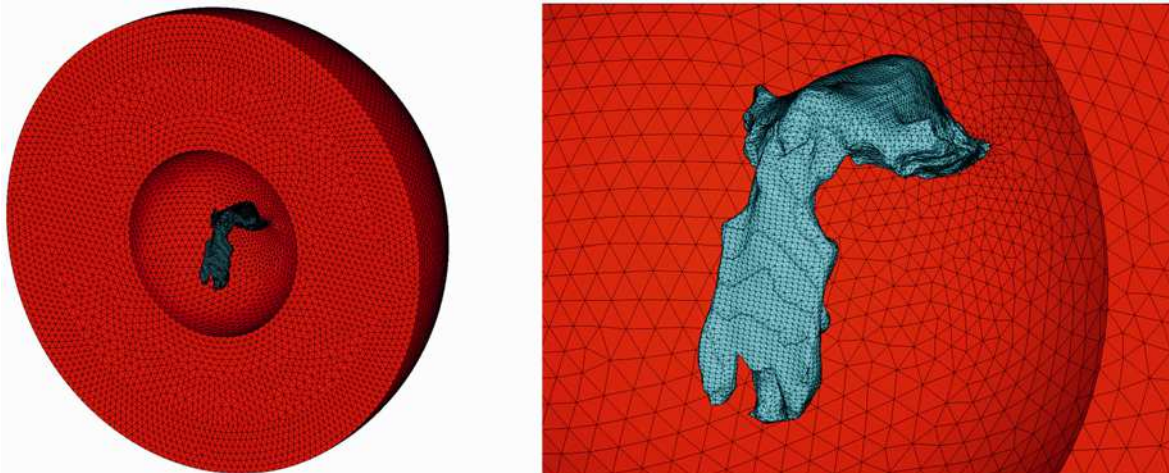

Fig. 6. FE model of the vocal tract for phonation on the vowel [a:], which includes also the surrounding infinite acoustic space allowing simulation of the sound propagation out of the mouth

5. Acoustic modal analysis of the FE models of the vocal tract

Acoustic modal analysis was performed on the FE models while considering, for simplicity, the following boundary conditions: zero acoustic pressure ($p = 0$) at the lips, closed vocal tract at the level of the vocal folds and no acoustic damping ($\mu = 0$) on the walls of the acoustic spaces (see Fig. 7). Nine formant frequencies were observed within the frequency range from zero up to c. 5 200 Hz. The first three eigenmodes for the frequencies F1–F3 are typical fundamental modes. The acoustic characteristics of these eigenmodes are similar like in a simple tube when the propagation of the longitudinal waves in the vocal tract is dominant and when the acoustics of the vocal tract can be described by a simple 1D acoustic theory (see, e.g. [7]). The higher eigenmodes for the formant frequencies F4–F9, however, are associated with cross-sectional
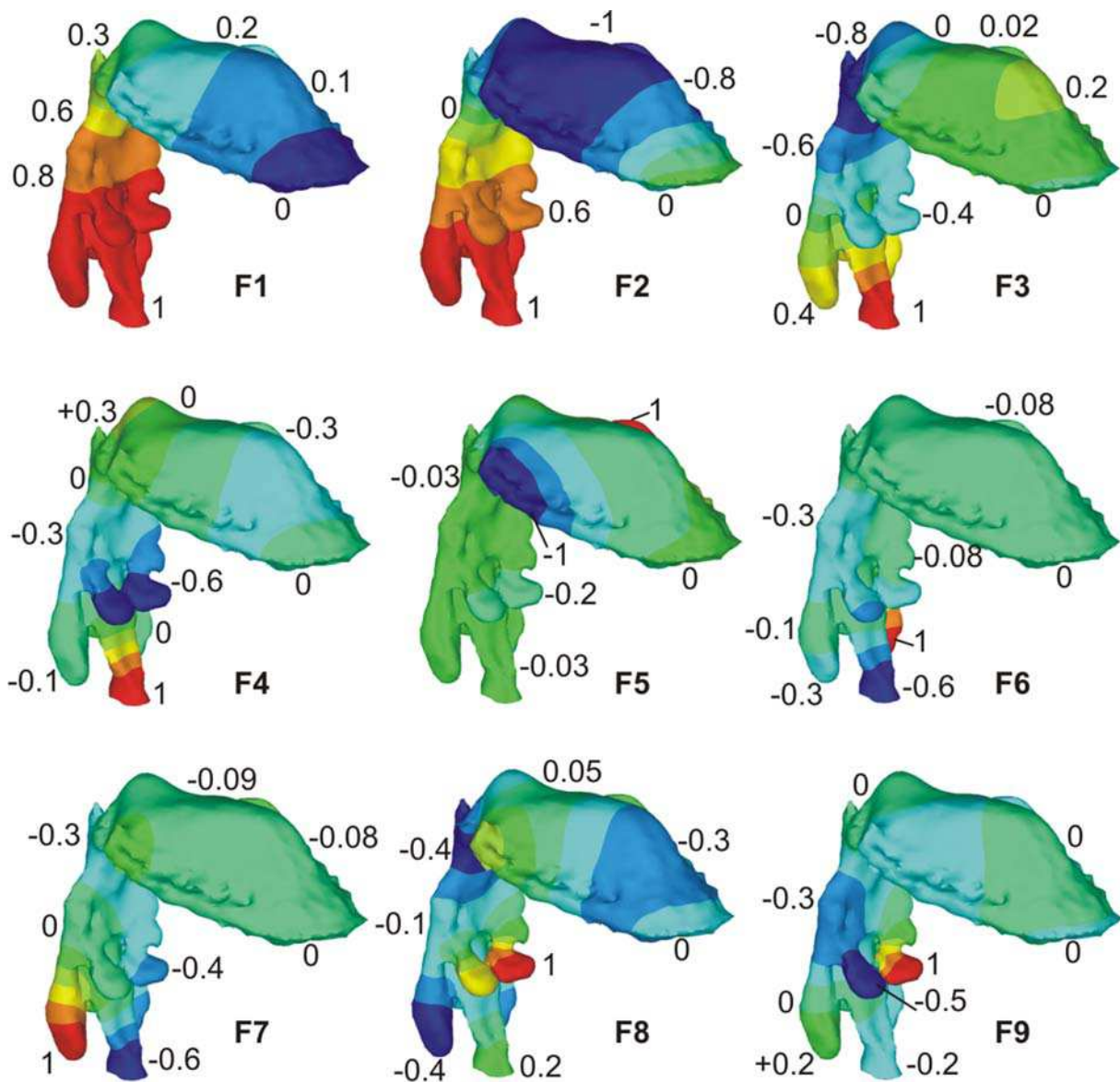
Fig. 7. Computed acoustic mode shapes of vibrations for formant frequencies F1–F9 of the vocal tract model of vowel [a:] after phonation into the tube

waves (see, e.g., the mode shape for the frequency F5) or notably 3D internal resonances in the laryngeal-pharyngeal part where the piriform sinuses, valleculae, laryngeal ventricles and the space between the false vocal folds just above the vocal folds play an important role (see the mode shape for the frequencies F4 and F6–F9, where the dominant antinodes for the pressure are situated in the above mentioned places). The acoustic mode shapes of vibrations did not essentially between the models for the vowel [a:] before and after the tube phonation.

6. Transient analysis of the FE models of the vocal tract

The supraglottal spaces of the FE models were excited at the level of the vocal folds by a very short impulse of acoustic flow velocity covering a broad band frequency range from zero up to about 5 kHz (see Fig. 8). The transient analysis was performed with the time step $\Delta t = 0.5 \cdot 10^{-4}$ s and the total simulated time was $T = 0.1$ s, resulting in the frequency
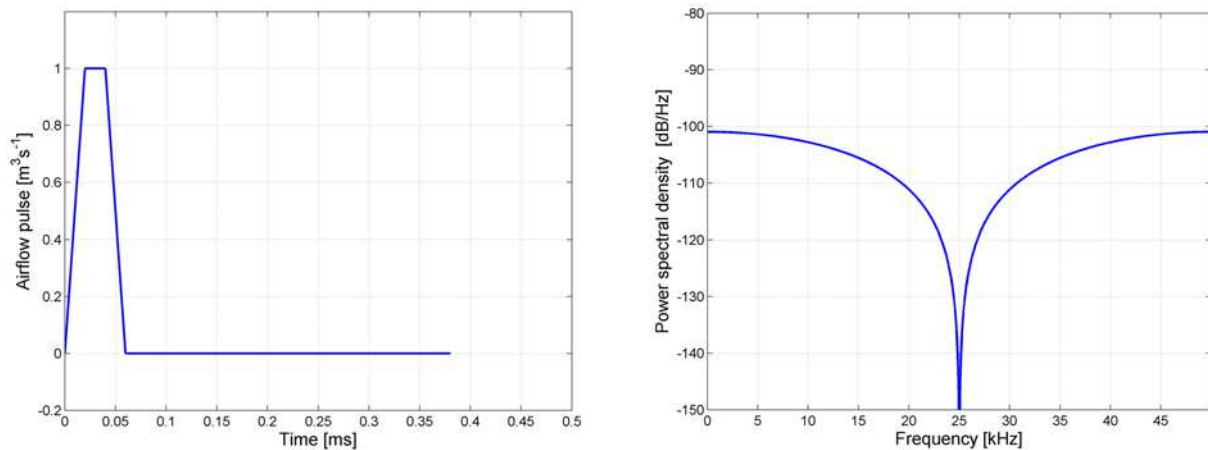
Fig. 8. The airflow velocity pulse used for the excitation of the FE models shown in the time and frequency domain
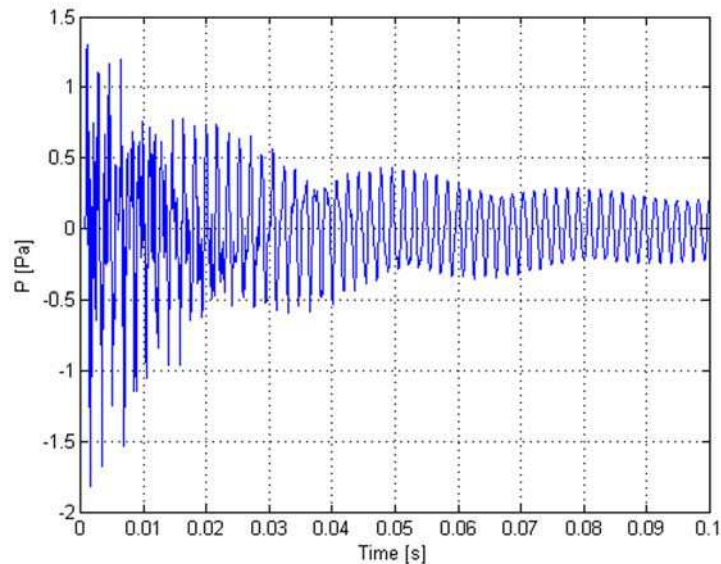

Fig. 9. Transient response in time domain of the acoustic pressure computed at the distance of 10 cm in front of the mouth excited by a short airflow velocity pulse at the vocal folds

resolution 10 Hz in the spectra. The acoustic pressure and the spectrograms of the response computed at the distance of 10 cm in front of the mouth are presented in Figs. 9 and 10. The transient time response of the pressure (see Fig. 9) shows a more sinusoidal waveform with increasing time, which is due to higher radiation losses for higher frequencies.

The numerically simulated spectrograms for the vowel [a:] before and after tube phonation are compared in Fig. 10. Formant F3 increased after phonation into the tube while the formants F1, F2, F4 decreased. Formation of a stronger cluster of formants (corresponding to the speaker's formant, see [2, 3]) is clearly visible around 4 kHz for phonation after the tube.

The power spectral densities computed at the distance of 10 cm in front of the mouth for the cases of before- and after-tube phonations are compared in Fig. 11. The computed resonance frequencies R1, R2 and R4 decreased down after phonation into the tube (about 1 % for R1, 5.1 % for R2 and 2.9 % for R4) and only the R3 increased by c. 5.1 %. While the frequencies of
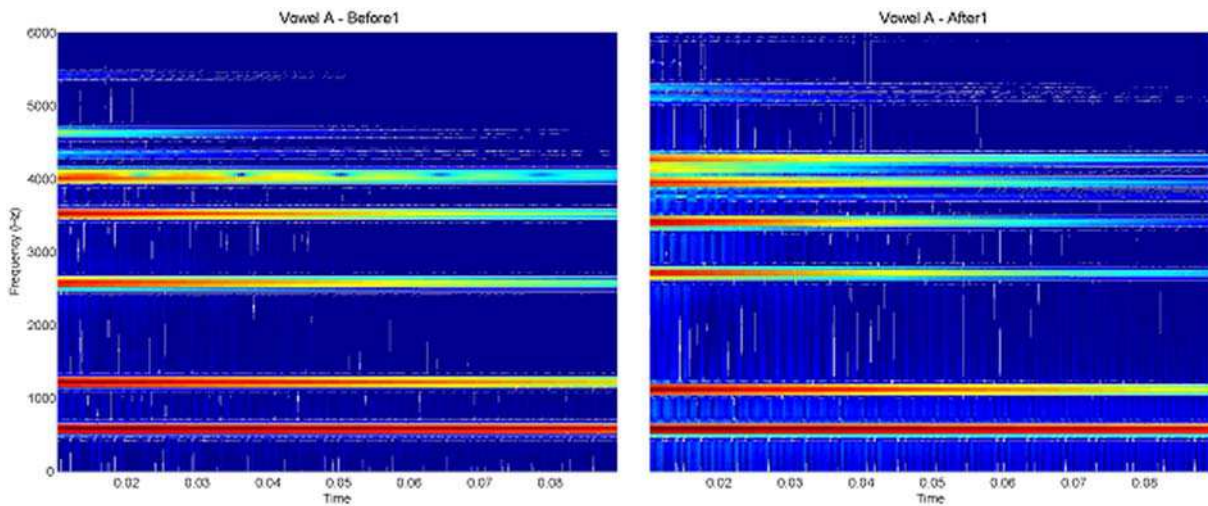
Fig. 10. Numerically simulated spectrograms of the excited sound in front of the mouth for the vocal tract models before (left) and after the tube-phonation (right)
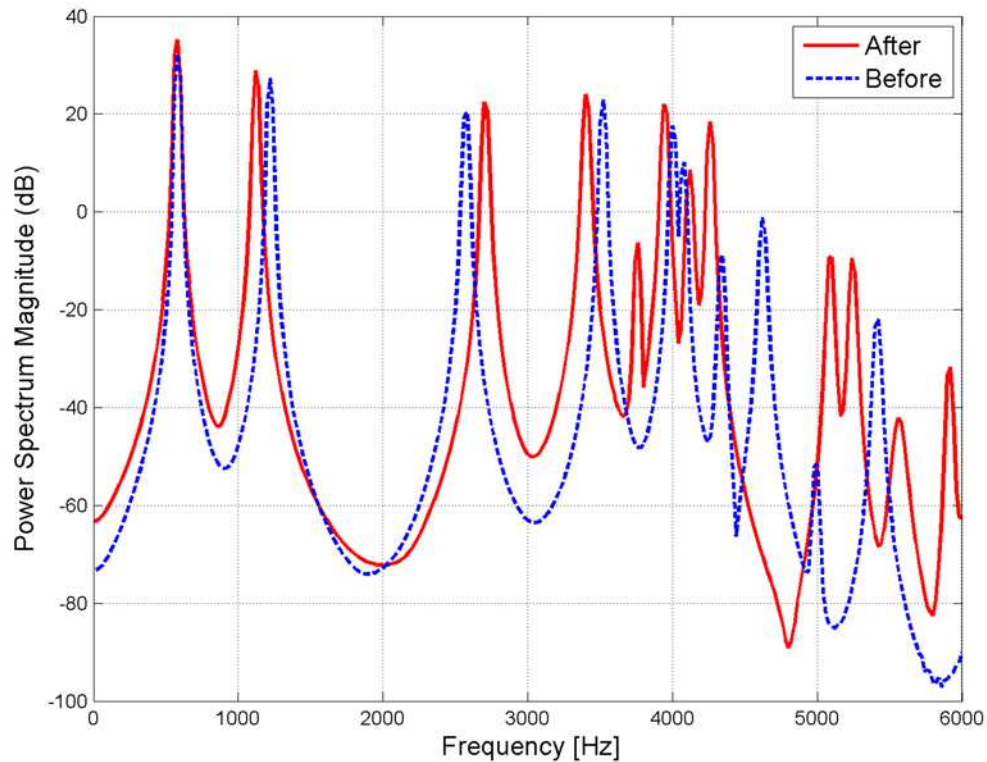

Fig. 11. Power spectral densities of the acoustic pressure computed in the distance of 10 cm in front of the mouth excited by a short airflow velocity pulse at the vocal folds

these resonances do not match perfectly the LPC-derived formant frequencies from the recorded acoustic signals, the general tendency of the resonance changes resembles the earlier observed tendency of the formant F3 to increase while the formants F1, F2 and F4 decrease.

Apart from the changes of the resonance frequencies, important differences can be seen in the spectral amplitudes of the resonance peaks. The peaks for the first four resonances (R1–R4) reach values, which are c. 3–4 dB higher for the case of the phonation after the tube. Even larger differences can be seen for the higher formants in the frequency region of 4–5 kHz, where some

of the resonant peak levels are up to c. 20 dB higher after than before the tube phonation. However, in this frequency region is difficult to determine corresponding resonances before and after the tube, because of a complete restructuring of the spectra.

7. Modelling of vocal tract transfer efficiency

For comparison of the efficiency of the acoustic pressure transfer through the vocal tract from the vocal folds to the distance 10 cm in front of the mouth we introduced the quantity EF (transfer efficiency) that characterizes the effectiveness of the sound pressure transfer from the vocal tract to the surroundings:

$$EF(f) = SPL_{out}(f) - SPL_{in}(f), \tag{1}$$

where SPL_{out} [dB] is the sound pressure level of the acoustic pressure *in front of the mouth* given as

$$SPL_{out}(f) = 20 * \log(|^{out}p(f)|/20 * 10^{-6}), \qquad {}^{out}p(f) = \mathrm{Re}\,{}^{out}p(f) + \mathrm{i}\,\mathrm{Im}\,{}^{out}p(f), \tag{2}$$

and $^{out}p(f)$ is the acoustic pressure in pascals in front of the mouth at the frequency f, SPL_{in} [dB] is the sound pressure level of the average acoustic pressure *inside the vocal tract* given as

$$\begin{aligned} SPL_{in}(f) &= 20 * \log\left(\tfrac{1}{N}\sum_{j=1,N}|^{in}p_j(f)|/20 * 10^{-6}\right), \\ {}^{in}p_j(f) &= \mathrm{Re}\,{}^{in}p_j(f) + \mathrm{i}\,\mathrm{Im}\,{}^{in}p_j(f), \end{aligned} \tag{3}$$

where $^{in}p_j(f)$ is the acoustic pressure in the j-node inside the FE model and N is the number of nodes inside the FE model of the vocal tract.

The acoustic pressures $^{out}p(f)$ in front of the mouth and $^{in}p_j(f)$ inside the vocal tract were computed by harmonic analysis with the frequency step $\Delta = 1$ Hz in selected narrow frequency regions around the detected formants (recall Fig. 11). The computed transfer efficiency (EF) of sound energy transfer at different formant frequencies is shown in Fig. 12. The maximum efficiency is possible to detect after the tube phonation in the frequency region 3.5–4 kHz, i.e. in the region of the so-called speaker's formant.
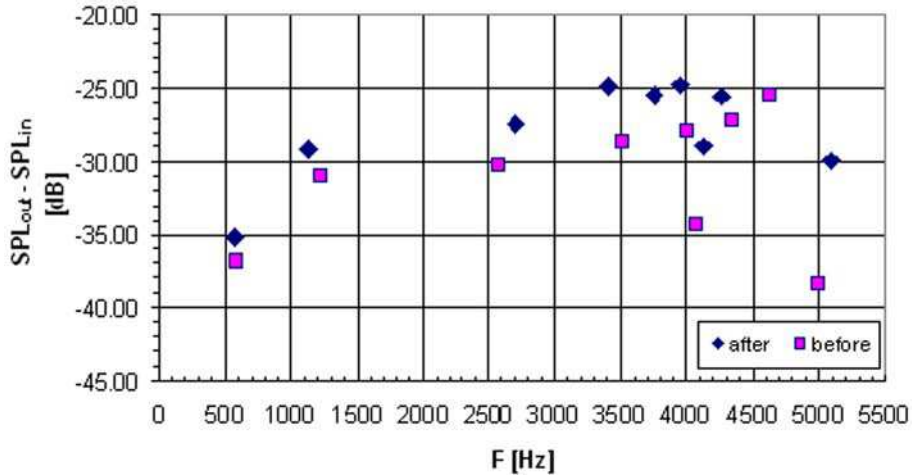

Fig. 12. Computed transfer efficiency of sound energy transfer at the formant frequencies for numerically simulated phonation before and after the tube-phonation

8. Conclusion

The lower formants F1–F3 represent classical vibration modes, which can be obtained also with 1D vocal tract models. At higher formants, however, more complicated transversal 3D modes of vibration are prominent, which require a 3D modelling approach. Comparison of the relative amplitudes of the pressure oscillations inside the vocal tract with the acoustic pressure in front of the mouth showed that formants differ in their efficiency; the formants (F4–F6) between 3.5–4.5 kHz, i.e. at the speaker's or singer's formant region were found to be the most effective. The human hearing threshold is also relatively low between 2 and 4 kHz. Consequently, sound energy concentration around 3.5 kHz (F4) region is useful to assure maximum audibility of the produced sound and can thus play a role in communication both from the point of view of production and perception.

The c. 3 dB-increase in SPL after the phonation into a tube observed here qualitatively corresponds to previous experimental findings. A tendency for SPL increase has been observed in earlier studies, e.g. measurements by Laukkanen [1]. An important finding of this study is that the c. 3 dB increase in SPL was possible to obtain through the vocal tract modelling, even without considering the interaction between the vocal tract and the vocal fold vibration.

The results suggest that exercising using a resonance tube can help in optimizing the vocal tract configuration for improved energy transfer from the vocal tract and thus improved vocal economy.

Acknowledgements

References

[1] Laukkanen, A. M., On speaking voice exercises. A study on the acoustic and physiological effects of speaking voice exercises applying manipulation of the acoustic-aerodynamical state of the supraglottic space and artificially modified auditory feedback, Ph.D. thesis, Acta Universitatis Tamperensis ser A, Vol. 445, University of Tampere, 1995.

[2] Leino, T., Long-term average spectrum study on speaking voice quality in male actors, Proceedings of the Stockholm Music Acoustics Conference 1993, Stockholm, The Royal Swedish Academy of Music, 1994, pp. 206–210.

[3] Nawka, T., Anders, L. C., Cebulla, M., Zurakowski, D., The Speaker's formant in male voices, Journal of Voice 11(4) (1997) 422–428.

[4] Titze, I. R., Voice training and therapy with a semi-occluded vocal tract: Rationale and scientific underpinnings, Journal of Speech Language and Hearing Research 49 (2006) 448–459.

[5] Titze, I. R., Laukkanen, A. M., Can vocal economy in phonation be increased with an artificially lengthened vocal tract? A computer modeling study, Logopedics Phoniatrics Vocology 32 (2007) 147–156.

[6] Titze, I. R., Story, B. H., Acoustic interactions of the voice source with the lower vocal tract, Journal of the Acoustical Society of America 101(4) (1997) 2 234–2 243.

[7] Vampola, T., Horáček, J., Švec, J. G., FE modeling of human vocal tract acoustic. Part I: Production of Czech vowels, Acta Acoustica United with Acta Acustica 94 (2008) 433–447.

[8] Vampola, T., Laukkanen, A. M., Horáček, J., Švec, J. G., Vocal tract changes caused by phonation into a tube: A case study using computer tomography and finite element modeling, Journal of the Acoustical Society of America (2010), DOI: 10.1121/1.3506347. (in print)