# Measurement and mathematical simulation of acoustic characteristics of an artificially lengthened vocal tract

Vojtěch Radolf [a,*], Jaromír Horáček [a], Pavel Dlask [b], Zdeněk Otčenášek [b], Ahmed Geneid [c], Anne-Maria Laukkanen [d]

[a] Institute of Thermomechanics AS CR, v. v. i., Dolejškova 1402/5, 182 00 Prague, Czech Republic
[b] Musical Acoustic Research Center, HAMU, Malostranské nám. 13, 118 01 Prague, Czech Republic
[c] Department of Otolaryngology and Phoniatrics – Head and Neck Surgery, Helsinki University Hospital and University of Helsinki, PL 220, 00029, HUS, Helsinki, Finland
[d] Speech and Voice Research Laboratory, School of Education, Åkerlundink. 5, University of Tampere, 33100 Tampere, Finland

## ARTICLE INFO

## ABSTRACT

Phonation into tubes is used for voice training and therapy. In the present study, the formant frequencies were estimated from measurements of the acoustic pressure and the acoustic input impedance for a plexiglass model of the vocal tract (VT) prolonged by a glass tube. Similar transfer function measurements were performed with a human VT in vivo. The experimental results matched the mathematical modelling and confirmed the legitimacy of assuming rigid walls in mathematical simulations of the acoustic characteristics of an artificial VT model prolonged by a tube. However, this study also proved a considerable influence from soft tissues in the yielding walls of human VT cavities on the first formant frequency, $F_1$. The measured $F_1$ for the VT model corresponded to the computed value of 78 Hz. The experiments in a human instead resulted in a much higher value of $F_1$: about 200 Hz. The results confirm that a VT model with yielding walls must be considered for mathematical modelling of the occluded or semi-occluded human vocal tract, e.g. prolonged by tubes or straws. This is explained by an acoustic-structural interaction of the vocal tract cavities with a mechanical low-frequency resonance of the soft tissue in the larynx.

## 1. Introduction

The human voice plays an important role in society, serving as the major carrier for human speech interpersonal communication. Understanding the basic principles of voice production is important for better interpretation of clinical findings, detection of laryngeal cancers or other pathologies, and for treatment of voice disorders.

Phonation is a complex physiological process that involves several basic factors such as airflow coming from the lungs, vocal-fold vibration and acoustic resonances of the cavities of the vocal tract. The vocal folds, excited by the air flow, generate a primary laryngeal tone whose fundamental frequency corresponds to the vibration frequency of the vocal folds. In the airways above the vocal folds, i.e. in the vocal tract, the acoustic resonant phenomena modify the spectrum of the

* Corresponding author. Tel.: +420 26605 3121.
E-mail addresses: radolf@it.cas.cz (V. Radolf), jaromirh@it.cas.cz (J. Horáček), pavel.dlask@hamu.cz (P. Dlask), zdenek.otcenasek@hamu.cz (Z. Otčenášek), Ahmed.Geneid@hus.fi (A. Geneid), Anne-Maria.Laukkanen@uta.fi (A.-M. Laukkanen).

primary laryngeal tone, especially the higher harmonics. The so-called formant is a result of the acoustic resonance of the vocal tract, which occurs as a peak in the voice spectrum. The formants in human voice define vowels and cause differences in the voice quality. The formants have three characteristics: frequency; amplitude and bandwidth. The formant frequencies are given by the size and shape of the acoustic resonance cavities that can be varied, for instance, by changing the position of the tongue or by mouth opening. The amplitude of the formant is related to (1) the proximity of the harmonic frequency to the resonance frequency, (2) strength of the harmonics of the voice source, and (3) the proximity of the adjacent formants. The first two formants, which are necessary for distinguishing the individual vowels, are mainly determined by the position of the tongue and the mouth opening. The higher formants cause differences in the voice timbre; see [1–3].

Voice disorders negatively influence quality of life, and patients generally see a laryngologist or phoniatrician to find the cause of a disorder and receive treatment. Phoniatricians often see patients with various voice disorders of a functional or organic origin. An integral part of voice therapy is "voice re-education," which involves utilization of a proper, healthy voice technique. In addition to an optimal breathing and vocal fold adjustment in relation to the amount of subglottic pressure, a very important factor is a correct utilization of the resonant spaces above the vocal folds. Knowledge of the exact space relations and of the ways to influence the acoustic voice quality by small shape changes in the vocal tract can improve the re-education techniques and contribute to voice therapy.

Voice training and therapy often involve semi-occlusions, i.e. sounds during which the vocal tract is either narrowed, as in the production of closed vowels [i, y, u]; constricted, as in voiced fricatives, e.g. [β, z], or nasals, e.g. [m, n]; completely closed in a rapid succession, as in a lip or tongue trill; or artificially lengthened, as when one phonates into a tube or a straw; e.g. see [4]. Phonation into a small diameter hard-walled "resonance tube" has been used especially in Scandinavia; see [5]. For voice therapy purposes, the outer end of this tube is placed in water ("water resistance therapy"), although, in voice training of normally voiced subjects, the outer end of the tube is kept in the air. During the semi-occlusion exercises, the trainee/patient is instructed to aim at ease of phonation and maximal vibratory sensations in the front part of the vocal tract and the facial structures. The goal of the semi-occlusion exercises is an effortlessly produced, well-resonating voice; see [4,6], for example.

One reason for the use of semi-occlusions in voice training and therapy lies in the fact that they slow down the airflow in the vocal tract and increase the supraglottal air pressure. Consequently, the intraglottal air pressure also increases. This reduces the vocal fold collision pressure and, hence, the mechanical load related to voice production. Aiming at maximal vibratory sensations in the vocal tract is supposed to help in optimizing the glottal and vocal tract setting for maximal sound energy transfer from the glottis to the outer space. Increased supraglottal air pressure aids the trainee in this optimization as it intensifies the sensations of resonatory vibrations; see [7,8]. Furthermore, semi-occlusions lower the formant frequencies, especially $F_1$ [9]. If $F_1$ comes sufficiently close to $f_0$ of phonation, the inertive reactance of the vocal tract may mechanically assist vocal fold vibration by reducing the phonation threshold pressure, i.e. the lowest subglottic pressure needed to start and maintain vocal fold vibration; e.g. see [10].

The effect of a particular semi-occlusion is, thus, related to two things: the amount of flow resistance it offers and the amount of inertive reactance within the fundamental frequency range of phonation. Obviously, the narrower a constriction in the vocal tract is, the longer and narrower the tube one phonates into, or the deeper in water the outer end of the tube is inserted, the higher the flow resistance will be. The amount of inertive reactance, in turn, is related to the resonances that will be created in the vocal tract of a subject phonating into a particular tube. To the best of our knowledge, no information on formant frequencies measured directly during phonation into tubes can be found in the literature. Such information would be needed in order to shed further light on the basis for the exercise and to be able to choose the best suited tube for the trainee/patient.

Applying a mathematical model of voice production, Story et al. [9] estimated the effects of various vocal tract configurations on the acoustic input impedance of the vocal tract, which is defined as the ratio of the input pressure (seen from the glottal end) to the acoustic volume velocity. According to their results, the frequency of the lowest maximum of the impedance, i.e. the first acoustic resonance (formant frequency, $F_1$) lowers with a semi-occlusion of the vocal tract (VT). The acoustic input reactance (imaginary part of the impedance) of the vocal tract increases in the fundamental frequency range of speech, which could explain the beneficial experience of using semi-occlusions in vocal exercises, since a higher reactance of the vocal tract lowers the phonation threshold pressure and may alter the voice source waveform in such a way that a slightly higher sound pressure level (SPL) and stronger voice overtones (louder and brighter voice quality) are obtained [11].

According to the calculations by Story et al. [9], $F_1$ was approximately in the range of 200–300 Hz for phonation into hard-walled tubes of 10 cm and 30 cm in length (inner diameter 8 mm). The vocal tract wall was considered to be yielding, having a low mechanical resonance frequency of its own. The parameters of such a VT model were set to the values suggested by Sondhi and Schroeter [12]. The present paper aims to experimentally verify the existence of this low-frequency eigenmode, which originates from the vibration of a soft tissue structure in the human vocal tract, as proposed by Sondhi [13]. The motivation to write this paper resulted from our numerical experiment. When applying the model described in [14], with the assumption of a hard-walled vocal tract, we obtained the result that $F_1$ decreases even below 100 Hz in phonation into a tube of 26.4 cm in length (inner diameter 6.8 mm). The main purpose of the paper is to prove that a hard-walled model of VT with a tube lowers the first formant frequency as compared to a soft-walled model.

The present study investigated the influence of phonation into a resonance tube on the lowest vocal tract resonances. The study consisted of the following: (1) acoustic analysis of recordings of a female subject phonating into the tube, (2) measurement of the transfer function (the frequency-dependent ratio of the acoustic pressure at the outlet of the tube to

the acoustic volume velocity at the glottis) when the vocal tract was externally excited by using a random noise in vivo for the female subject and in vitro for a plexiglass model of the vocal tract, (3) measurement of the acoustic input impedance for the vocal tract model using a commercial instrument analysis system, and (4) comparison of the experimental results (from (1) to (3)) with the theoretical results obtained by applying a mathematical model of VT acoustics.

## 2. Methods

### 2.1. Investigation of human speech phonation into a tube

A female subject with no voice or hearing problems who had extensive experience in voice training and in using resonance tubes volunteered as a subject for the experiment. The subject was one of the authors (A-M L) and fully aware of the aims of the study and acquainted with the methodology applied. No other medical research methods were used except for the ordinary videoendoscopy used in phoniatric clinics. The investigation has thus been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki).

The subject phonated the vowel [u] first without the tube and then into the resonance tube (made of glass, 26.4 cm in length, 6.8 mm in inner diameter, 10 mm in outer diameter) using a normal (neither breathy nor pressed) speaking voice at a comfortable pitch and loudness. The frequency spectrum of the signal was measured outside the tube using a sound-level metre (B&K type 2239, frequency range 8 Hz to 16 kHz). An acoustic analysis was done in Matlab by averaging the Fast Fourier Transform (FFT) frequency spectra computed by using a Hanning time window of 0.5 s with a 75 percent overlap. The resulting averaged (filtered) spectra were obtained by using frequency bands (windows) that were equal to the fundamental phonation frequency $f_0$ and shifting them by a frequency step $\Delta f$. The maxima of these "filtered spectra" can be considered formants.

### 2.2. Measurement of the transfer functions

#### 2.2.1. Experimental setup

A transfer function of the vocal tract prolonged by a tube was measured using pressure signals recorded at two different positions. The acoustic pressure at the lips' position was measured by a special B&K 4182 microphone probe (range 1 Hz to 20 kHz), and the acoustic pressure at the open outlet of the tube was registered by a B&K 4138 miniature microphone (range 6.5 Hz to 140 kHz); see the schema in Fig. 1. Signals were recorded by a PC-controlled B&K PULSE 10 measurement system using a 32.8 kHz sampling frequency. The acoustic system was excited by a white noise signal (range 0–6.4 kHz) using a loudspeaker (an MTC 5 1/4″, 4 Ω woofer) located outside. The glass resonance tube was connected to the vocal tract at the lips. There was no special impedance matching between the speaker and the tube.

The same measurement setup was used for both the subject and for a physical model of the vocal tract. For in vivo experiments, the subject kept her vocal folds tightly together, i.e. the glottis closed, while her vocal tract simulating phonation of the vowel [u:] was excited by the external loudspeaker. The subject was examined using a videoendoscopy system during phonation and when the glottis was closed without phonation (see Fig. 2): an Olympus videoendoscopy system that included a Xenon light source (Elvis Exera III, CLV 190, Olympus, Finland) and a fiberscope ENF-VH HD Distal chip fiberscope (ENF-VH, Olympus, Finland). The picture of the glottal closure while holding breath without phonation shows that the glottis was fully closed and the false vocal folds were somewhat adducted.

A vocal tract model used for in vitro measurement was made of plexiglass (see Fig. 3). Its geometrical configuration was based on a 3D volume model obtained from magnetic resonance images [15] for the vowel [u:]. This vowel was chosen because in previous investigations the subject used the articulation of [u:] during tube phonation. The channel of the VT model was variable in its height, whereas the width (thickness) was kept constant (see Figs. 3 and 4b). The input of the VT model at the glottis end was closed with a rigid plexiglass plate. Rectangular cross-sectional areas of the plexiglass model from the glottis to the lips were set to be the same as in the study [15], where it was shown that the difference between the first four resonance frequencies of the 1D and 3D mathematical models of the VT for the vowel [u:] is negligible (less than 2 percent).
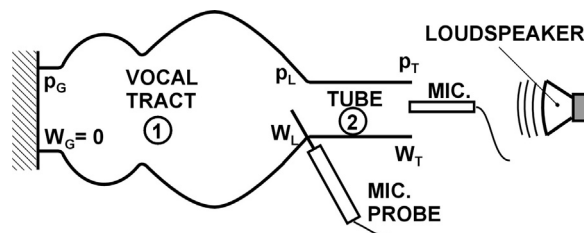


**Fig. 1.** A layout of the experimental setup for measuring transfer functions in vivo and in vitro.
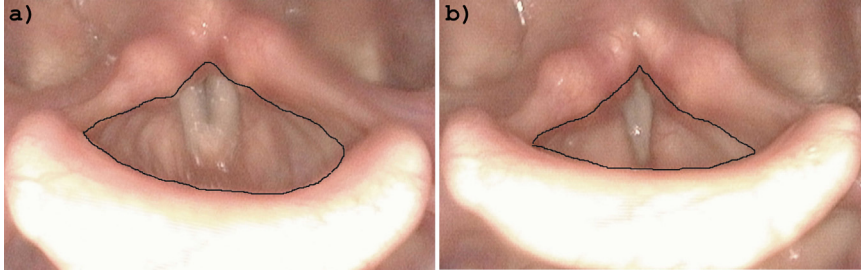
**Fig. 2.** Videoendoscopic investigation of the subject's glottis: (a) during phonation into the resonance tube and (b) while holding the tube between lips without phonation. The outlet of the laryngeal tube is marked with a solid black line.
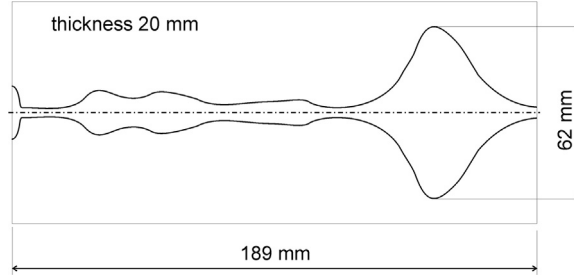


**Fig. 3.** Geometry of the plexiglass VT model.

### 2.2.2. Processing and interpretation of experimental data from transfer function measurements

The data recorded both in vivo and in vitro experiments were processed using the following matrix equations in a frequency domain (see the derivation in Appendix A):

$$\begin{bmatrix} p_L \\ U_L \end{bmatrix} = \mathbf{T_{VT}} \cdot \begin{bmatrix} p_G \\ U_G \end{bmatrix} = \begin{bmatrix} A_{VT} & B_{VT} \\ C_{VT} & D_{VT} \end{bmatrix} \cdot \begin{bmatrix} p_G \\ U_G \end{bmatrix}, \tag{1}$$

and

$$\begin{bmatrix} p_T \\ U_T \end{bmatrix} = \mathbf{T_{TB}} \cdot \begin{bmatrix} p_L \\ U_L \end{bmatrix} = \begin{bmatrix} A_{TB} & B_{TB} \\ C_{TB} & D_{TB} \end{bmatrix} \cdot \begin{bmatrix} p_L \\ U_L \end{bmatrix}, \tag{2}$$

where $p$ and $U$ are the acoustic pressure and volume velocity respectively, $\mathbf{T_{VT}}$ is a transfer matrix of the vocal tract, $\mathbf{T_{TB}}$ is a transfer matrix of the tube, and the indices $G$, $L$, $T$ respectively mean the position of glottis, lips and open tube end.

The volume velocity is zero ($U_G=0$) at the closed VT end at the glottis position, and thus we can eliminate $U_L$ from Eq. (1):

$$U_L = C_{VT}/A_{VT} \cdot p_L, \tag{3}$$

and putting Eq. (3) into Eq. (2) yields the transfer function for the pressure between the input and the end of the tube:

$$p_L(\omega)/p_T(\omega) = A_{VT}/(A_{TB}A_{VT} + B_{TB}C_{VT}). \tag{4}$$

Generally, the transfer function $U_T/U_G$ of the complete system (VT+tube) between the acoustic volume velocity at the glottis position and the tube end can be expressed as (see Eq. (A.13) in Appendix A):

$$U_T(\omega)/U_G(\omega) = 1/(A_{TB}A_{VT} + B_{TB}C_{VT}). \tag{5}$$

This equation is valid regardless of the boundary conditions at the glottis when the radiation impedance $Z_{T,RAD}$ at the tube output, defined as the ratio of output pressure to output volume velocity, is omitted. Because the denominator in Eq. (4) is the same as in Eq. (5), resonance frequencies (peaks) of the measured transfer function $p_L(\omega)/p_T(\omega)$ should be approximately equal to the peaks of the transfer function $U_T(\omega)/U_G(\omega)$ of the complete system.

The numerator $A_{VT}(\omega)$ in Eq. (4) is the same as a denominator in the transfer function $U_L U_G$ of the vocal tract alone (see Eq. (A.16) in Appendix A):

$$U_L(\omega)/U_G(\omega) = 1/A_{VT}. \tag{6}$$

Again, when omitting the radiation impedance $Z_{L,RAD}$ at the lips, Eq. (6) is generally valid regardless of the boundary conditions at the glottis. Antiresonance frequencies (spectral dips) of the transfer function (4) then should be very close to the peaks of the transfer function (6) of the vocal tract without a tube.
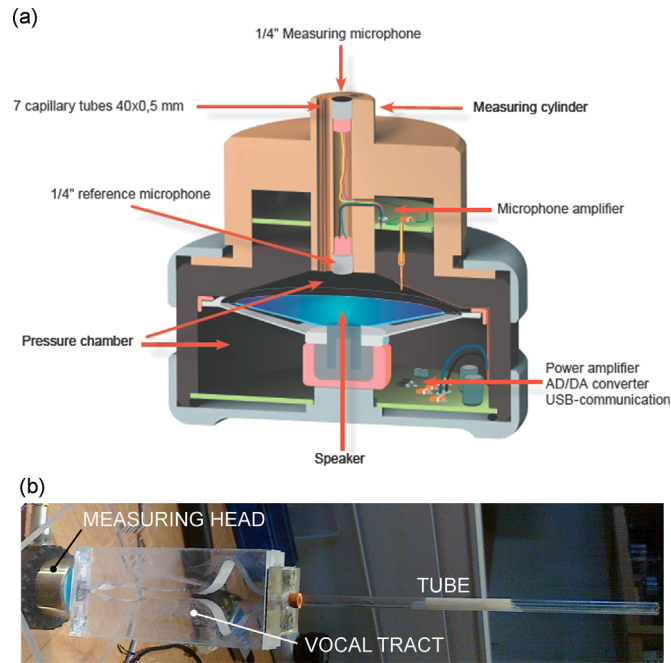
(a)



(b)



**Fig. 4.** (a) A layout of the measuring head of the BIAS 6 system [16] (reproduced with permission of Prof. Gregor Widholm) and (b) experimental setup for measuring the acoustic input impedance on a plexiglass model of the vocal tract prolonged by a tube and installed in the measuring head of the system BIAS 6 at the position of the vocal folds.

## 2.3. Measurement of the acoustic input impedance

### 2.3.1. Experimental setup

A measurement of the acoustic input impedance was only carried out on the plexiglass model of the vocal tract using the Brass Instrument Analysis System (BIAS 6) controlled by a PC. A layout of the measuring head of the system is shown in Fig. 4a. The input of the VT model at the position of vocal folds was fixed to the measuring head and sealed (see Fig. 4b). The VT model was excited by a sweep pulse in the frequency range of 20–4000 Hz. The acoustic input impedance, $Z_{IN}$, was then evaluated by the BIAS software, and the data that consisted of the frequency-dependent magnitude and phase of $Z_{IN}$ were recorded by a PC at the frequency step of 0.5 Hz.

### 2.3.2. Processing and interpretation of experimental data from input impedance measurements

The VT input impedance $Z_{IN}$, calculated as the ratio of the input pressure to the input acoustic volume velocity, is given by the expression (see Eq. (A.17) in Appendix A):

$$Z_{IN} = \frac{p_G}{U_G} = \frac{D \cdot Z_{T,RAD} - B}{A - C \cdot Z_{T,RAD}}, \tag{7}$$

where $A$, $B$, $C$, $D$ are the components of the transfer matrix of the complete system (VT+tube) and $Z_{T,RAD}$ is the radiation impedance at the open end of the tube. The transfer matrix of the complete system relates the input variables with the output variables by the equation (see Eqs. (A.9) and (A.10) in Appendix A):

$$\begin{bmatrix} p_T \\ U_T \end{bmatrix} = \mathbf{T_{TB}} \cdot \mathbf{T_{VT}} \cdot \begin{bmatrix} p_G \\ U_G \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \begin{bmatrix} p_G \\ U_G \end{bmatrix}. \tag{8}$$

Let us consider the situation that the input is closed by a rigid wall ($U_G=0$) and the output loaded by the impedance $Z_{T,RAD}$. After substituting these boundary conditions, denoted as C-O$Z_{RAD}$ (closed input, open loaded output) in Eq. (8), it can easily be derived that the denominator of the input impedance in Expression (7) is equal to the frequency equation of the complete system (VT+tube) with the boundary conditions C-O$Z_{RAD}$. Thus the peaks (resonance frequencies) of $|Z_{IN}(\omega)|$ represent resonances or natural acoustic frequencies of the system with boundary conditions C-O$Z_{RAD}$. These resonances are identical to resonance frequencies obtainable from a commonly used transfer function (A.13) of the complete system.

Considering the input is open ($p_G=0$), and the output is loaded by the impedance $Z_{T,RAD}$, the substitution of these boundary conditions denoted as O-O$Z_{RAD}$ (open input, open loaded output) in Eq. (8) indicates that the numerator of Eq. (7) is the same as the frequency equation of the system with the O-O$Z_{RAD}$ boundary conditions. It means that the spectral dips (antiresonances) of $|Z_{IN}(\omega)|$ represent natural acoustic frequencies of the complete system with the boundary conditions O-O$Z_{RAD}$.

### 2.4. Input parameters of mathematical models

A vocal tract channel with yielding walls was modelled according to [9,12] as a system of 43 cylindrical elements of 4 mm in length except for the first and last element, whose length was 2 mm, giving the total length of the present subject's vocal tract of 16.8 cm measured during previous experiments [17]. The following formant frequencies estimated from the acoustic recording of the vowel [u:] for the female subject (without a tube) were prescribed to the VT mathematical model: $F_1 = 367$ Hz, $F_2 = 730$ Hz, $F_3 = 2518$ Hz, $F_4 = 3735$ Hz (see [17]). Through a tuning procedure [14] carried out by changing the vocal tract shape, i.e. the area of the cylindrical cross-sections, the best-fitting vocal tract configuration was obtained. When connecting the tube to the vocal tract, the diameter of the last element of the VT model was increased to 10 mm (the outer diameter of the tube) and the penultimate element's diameter was recalculated as an average of the two of its neighbouring elements. The resulting model is shown in Fig. 5.

A mathematical model of the vocal tract acoustics with rigid walls is described in Appendix A. Its length as well as the length of the plexiglass VT model was 18.9 cm. The tube was modelled as one element with rigid walls.

The following values of fluid density, dynamic air viscosity, and speed of sound were used as input for the models: $\rho_0 = 1.2$ kgm$^{-3}$, $\mu = 1.8 \cdot 10^{-5}$ kgm$^{-1}$s$^{-1}$, $c_0 = 343$ ms$^{-1}$ (considering the temperature in the laboratory) for the artificial VT and $c_0 = 353$ ms$^{-1}$ for the human VT.

## 3. Results

### 3.1. Measurements compared to mathematical modelling of human phonation into the tube

Fig. 6 shows results from the acoustic analysis of samples recorded of our female subject phonating into the tube. Average FFT spectra were computed for $f_0 = 152$ Hz, with a frequency step of $\Delta f = 10$ Hz. No formant frequency (maxima in the filtered spectrum) could be detected in the frequency range of 0–500 Hz. A peak at 28 Hz in the FFT spectrum may have been caused by a mechanical resonance of the yielding vocal tract walls.

A transfer function was calculated from the glottis to the point inside the tube located 2 mm from its open outlet, both considering the radiation impedance $Z_{T,RAD}$ (thick line) and omitting $Z_{T,RAD}$ (thin line). The results can be seen in Fig. 7a. The results for the input impedance, obtained following the same constraints, are shown in Fig. 7b. The curves in Fig. 7b are nearly identical, suggesting that $Z_{T,RAD}$ had no effect on the magnitude of the input impedance.

The resonance frequencies obtained by mathematical modelling were $F_1 = 208$ Hz, $F_2 = 602$ Hz, $F_3 = 694$ Hz, $F_4 = 1304$ Hz, $F_5 = 1945$ Hz, $F_6 = 2473$ Hz, $F_7 = 2588$ Hz and $F_8 = 3223$ Hz. These values correspond well to the resonances measured (see thick lines in Figs. 6 and 7a), except for the first resonance frequency, $F_1$, as mentioned above. If $F_1$ was really 208 Hz, it would suggest that phonation into the resonance tube is assisted by reactance of the vocal tract at fundamental frequencies generally used in speech (i.e. $f_0$ below 200 Hz).

Fig. 7b shows the computed results for the input impedance of the human vocal tract. The three lowest resonances are clearly visible: $F_1 = 208$ Hz, $F_2 = 602$ Hz and $F_3 = 694$ Hz; however, the resonances $F_4$, $F_5$, $F_7$ and $F_8$ are not identifiable in
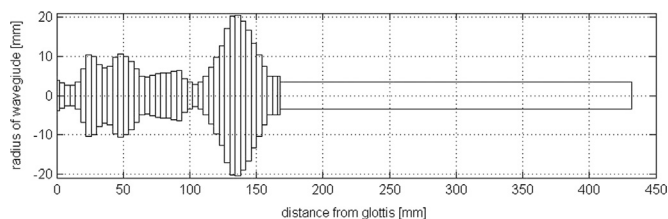


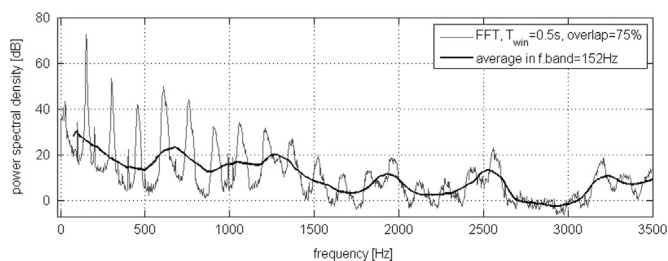**Fig. 5.** Geometry of the VT model prolonged by the tube.



**Fig. 6.** Spectra of the acoustic signal measured outside the human vocal tract during phonation into the tube. Thick and thin lines respectively denote the "filtered spectrum" and the FFT spectrum.
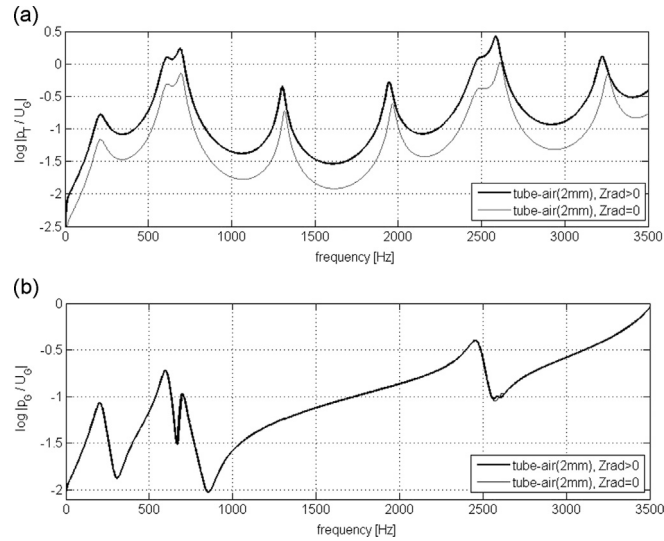
**Fig. 7.** Results of mathematical modelling of the human vocal tract: (a) the magnitude of the transfer function $p_T/U_G$ between the glottis position and the cross-section of the tube, located 2 mm from its open outlet, both considering the radiation impedance $Z_{T,RAD}$ (thick line) and omitting it (thin line) and (b) the magnitude of the acoustic input impedance $p_G/U_G$ considering the radiation impedance $Z_{T,RAD}$ (thick line) and omitting it (thin line).
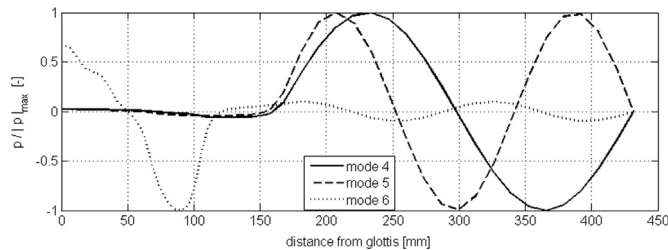


**Fig. 8.** Computed acoustic mode shapes of the human vocal tract prolonged by the tube for the resonance frequencies $F_4 = 1318$ Hz (solid line), $F_5 = 1965$ Hz (dashed line) and $F_6 = 2479$ Hz (dotted line) for the geometry of the VT model shown in Fig. 5. (The wall compliance and radiation impedance were omitted.).
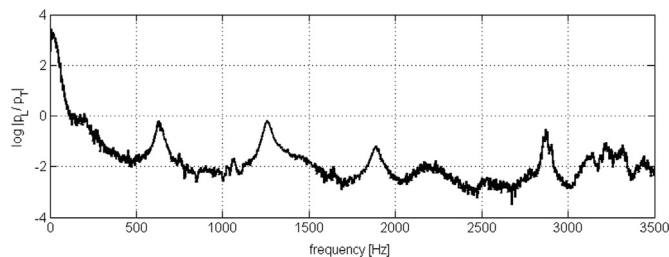


**Fig. 9.** A measured transfer function between the lips and the open tube end for the human vocal tract prolonged by the tube and excited by random noise from the loudspeaker.

Fig. 7b. The reason consists of the fourth, fifth and sixth acoustic mode shapes shown in Fig. 8. The resonances $F_4$ and $F_5$ are associated with the acoustic mode shapes in which the dominant vibration amplitudes are inside the resonance tube, and the pressure at the glottis is very small: i.e. the input impedance in Fig. 7b does not have clear resonance peaks at the frequencies of $F_4$ and $F_5$. In contrast, the resonance $F_6$ is clearly visible in the computed input impedance in Fig. 7b because the associated acoustic mode has a high amplitude at the glottis; see Fig. 8.

### 3.2. Measurements compared to mathematical modelling of the human and artificial vocal tracts prolonged by the tube and excited by the loudspeaker

The transfer function $|p_L(\omega)/p_T(\omega)|$ measured in vivo for the human vocal tract prolonged by the tube is shown in Fig. 9. The first resonance frequency can be identified at about $F_1 \cong 190$ Hz. However, due to a strong damping effect of the soft tissues, no sharp peak appears in the spectrum. Most likely, this is also the reason why there are no recognizable

antiresonances in the measured spectrum. Higher resonances, which are clearly visible in Fig. 9 at frequencies $F_2 \cong 630$ Hz, $F_3 \cong 1262$ Hz and $F_4 \cong 1889$ Hz, approximately correspond to the formant frequencies obtained from human phonation (see Fig. 6), even though the glottis was permanently closed during the external excitation from the loudspeaker. A steep increase of the transfer function near the zero frequency can be caused by a more prominent acoustic structural interaction of the vocal tract with vibrating yielding walls.

The measured and the computed magnitudes of the transfer function $|p_L(\omega)/p_T(\omega)|$ for the artificial VT model prolonged by the tube can be seen in Fig. 10. The computation was performed according to Eq. (4) assuming hard walls of the vocal tract. First resonance peaks occur at frequencies 73 Hz and 78 Hz for the measured and computed results, respectively. The first two antiresonances, which according to Eq. (6) are the resonance frequencies of the vocal tract alone (without a tube), can be well recognized in the spectra at about 400–500 Hz and 900–1000 Hz.

Fig. 11 shows the measured and the computed magnitudes of the acoustic input impedance for the artificial VT model prolonged by the tube. The computation was performed according to Eq. (7), assuming hard walls of the vocal tract. The radiation impedance, $Z_{T,RAD}$, at the tube output was omitted in this mathematical model. The first resonance peak obtained from the measurement occurs at the frequency of $F_1 = 99$ Hz, whereas the computed value $F_1 = 78$ Hz is equal to the first resonance frequency of the computed transfer function (see Fig. 10), as expected by comparing Eqs. (4) and (7).

### 3.3. Summary of the results

Tables 1 and 2 summarize the measured and computed resonance frequencies and their relative differences both for the transfer function and the input impedance of the artificial VT. Similarly, Table 3 compares the measured and computed acoustic resonance frequencies obtained for the human vocal tract. The values marked by "x" could not be identified in the measurements of the human VT. Table 4 compares formant bandwidths obtained from the measured transfer functions (TF) $|p_L(\omega)/p_T(\omega)|$ of the artificial vocal tract and of the human vocal tract, both prolonged by the tube. The formant bandwidth was measured as the frequency difference between the two points 3 dB below the peak level of the solid lines in Figs. 9 and 10. The formant bandwidth values were not identifiable in the case of $F_1$ (190 Hz) for the human vocal tract (Fig. 9) and in the case of the two close formants (resonances) measured in the model (560 and 672 Hz; see Fig. 10).

## 4. Discussion

### 4.1. Discussion of the results

The first six acoustic resonance frequencies resulting from the computed transfer function $|p_L(\omega)/p_T(\omega)|$ for the artificial VT differed only slightly from the measured results (see Fig. 10 and Table 1). The maximum relative difference was 6.8 percent for the first resonance (formant), $F_1$. A relatively large difference of 21.2 percent in $F_1$ between the measured and
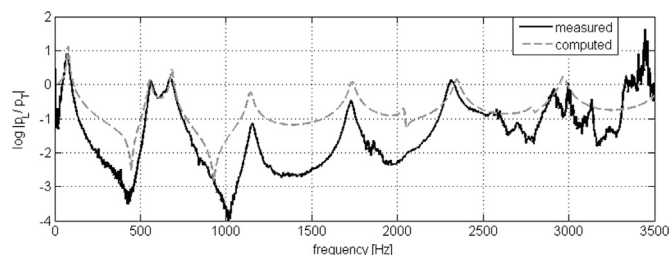


**Fig. 10.** A comparison of measured (solid line) and computed (dashed line) transfer functions $p_L/p_T$ between the lips and the open tube end for the plexiglass vocal tract prolonged by the tube. Resonances were excited by random noise using the external loudspeaker, see Fig. 1.
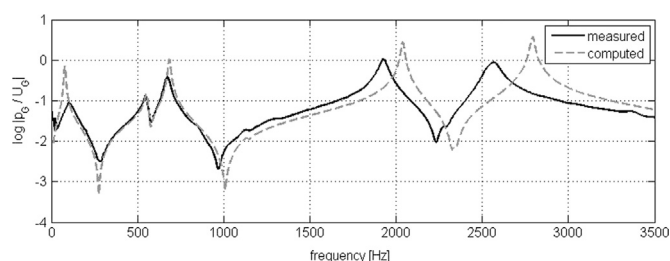


**Fig. 11.** A comparison of measured (solid line) and computed (dashed line) magnitudes of the input impedance $p_G/U_G$ for the artificial plexiglass vocal tract model prolonged by the tube.

**Table 1**
Measured and computed resonance frequencies (Hz) evaluated from the transfer functions (TF) of the artificial vocal tract prolonged by the resonance tube.

| | $F_1$ | $F_2$ | $F_3$ | $F_4$ | $F_5$ | $F_6$ |
|---|---|---|---|---|---|---|
| **VT model:** measured TF | 73 | 560 | 672 | 1153 | 1729 | 2313 |
| **VT model:** computed TF | 78 | 555 | 684 | 1142 | 1735 | 2347 |
| Difference (%) | 6.8 | − 0.9 | 1.8 | − 1.0 | 0.3 | 1.5 |

**Table 2**
Measured and computed resonance frequencies (Hz) evaluated from the input impedance $Z_{IN}$ of the artificial vocal tract prolonged by the resonance tube.

| | $F_1$ | $F_2$ | $F_3$ | $F_4$ | $F_5$ |
|---|---|---|---|---|---|
| **VT model:** measured $Z_{IN}$ | 99 | 546 | 673 | 1926 | 2564 |
| **VT model:** computed $Z_{IN}$ | 78 | 555 | 684 | 2039 | 2792 |
| Difference (%) | − 21.2 | 1.6 | 1.6 | 5.9 | 8.9 |

**Table 3**
Formant frequencies (Hz) obtained from the acoustic spectra of phonation into the resonance tube and from the measured and computed transfer functions (TF) of the human vocal tract prolonged by the tube (x: unidentified values).

| | $F_1$ | $F_2$ | $F_3$ | $F_4$ | $F_5$ | $F_6$ |
|---|---|---|---|---|---|---|
| **Human VT:** phonation | x | 680 | x | 1280 | 1930 | 2530 |
| **Human VT:** measured TF | 190 | 630 | x | 1262 | 1889 | x |
| Difference (%) | x | − 7.4 | x | − 1.4 | − 2.1 | x |
| **Human VT:** computed TF | 208 | 602 | 694 | 1304 | 1945 | 2473 |

**Table 4**
Formant bandwidths (Hz) obtained from the measured transfer functions of artificial and human vocal tracts prolonged by the resonance tube (x: unidentified values).

| | $B_1$ | $B_2$ | $B_3$ | $B_4$ | $B_5$ |
|---|---|---|---|---|---|
| **VT model:** measured TF | 12 | 17 | 26 | 21 | 21 |
| **Human VT:** measured TF | x | 16 | x | 22 | 23 |
| Difference (%) | x | − 5.9 | x | 4.8 | 9.5 |

computed input acoustic impedance $Z_{IN}$ (Fig. 11 and Table 2) was probably caused by porosity of a microphone and capillary tubes of the measuring head (see Fig. 4), contrary to the rigid wall boundary condition ($U_G=0$) applied to the mathematical model. The first resonance frequency, $F_1$, for the artificial VT model with hard walls was found to be between 73 and 99 Hz, which is much lower than the frequency $F_1 \cong$ 190–208 Hz resulting from measurements of transfer functions for the human vocal tract and confirmed by computations (see Figs. 7, 9 and Table 3). Fig. 2 shows that the size of the epilaryngeal tube of the subject was approximately 50 percent smaller when holding the breath without phonation than during phonation. The effect of such a change on $F_1$ can be estimated to be only about 3–4 percent, based on the results of Titze and Story [11], who calculated that $F_1$ was about 10 percent higher when the cross-sectional area of the laryngeal tube was six times smaller.

The second resonance frequency $F_2$ for the artificial VT model with hard walls was found to be between 5and 560 Hz, and the agreement between the measured and computed resonance was very good, i.e. less than 1.6 percent (see Tables 1 and 2). The measured formant frequency $F_2$ for the human vocal tract was found to be higher, i.e. between 630 and 680 Hz (see Figs. 6 and 9 and Table 3). We should note that the artificial VT was 2.1 cm longer than the model of the human VT. It means that the relative difference in the total length (VT+tube) of these models was about 5 percent. Therefore, we can expect that resonance frequencies of the artificial VT are approximately 5 percent lower than resonances of the human VT. Moreover, the geometry (area function) of the two models was not exactly the same, which could also cause shifts of formant frequencies. Yet, another factor affecting the results was the difference in the speed of sound, which was 343 m/s for the temperature in the laboratory and 353 m/s in the human VT. The higher speed of sound increases resonances of the human VT by 2.9 percent.

The two close resonance frequencies between 500 and 1000 Hz, $F_2$ and $F_3$, were clearly distinguishable in the results of physical and mathematical modelling (see Figs. 7, 10 and 11). In contrast, they were not identified as two peaks in the measured characteristics of the human vocal tract (see Figs. 6 and 9). This discrepancy between the results from modelling and those obtained from the human may stem in the fact that the frequency resolution of the mathematical models was quite high, given by the frequency step of 1 Hz. Another reason could be a higher damping of the acoustic modes in the human vocal tract due to the interaction of acoustic waves with soft tissues of the VT walls.

A potential error in the input impedance introduced by omitting the radiation impedance $Z_{T,RAD}$ at the tube output is insignificant. This is demonstrated in Fig. 7b, where there is practically no difference between the curves for $Z_{T,RAD} > 0$ and $Z_{T,RAD} = 0$. Even though the transfer function is affected by the omission of $Z_{T,RAD}$, the effect mainly concerns the level of the curve, and just minor shifts in the resonance frequencies can be detected (see Fig. 7a).

The comparison between Figs. 10 and 11 shows that the first three clearly visible resonance frequencies of the artificial vocal tract that resulted from the measured and computed transfer functions fit well with the corresponding resonances that were detectable in the input impedances. However, the higher peaks in the computed $Z_{IN}$ spectrum above 1900 Hz in Fig. 11 were suppressed in the computed transfer function curve by adjacent antiresonances around 2000 Hz and 2800 Hz; see the dashed line in Fig. 10.

Similarly, resonances around 1100 Hz and 1700 Hz that can be seen in Fig. 10 practically vanished in the $Z_{IN}$ spectrum (see Fig. 11) as a result of very tightly adjacent antiresonances. These antiresonances that are around 275 Hz, 574 Hz, 1008 Hz (cf. Fig. 11), 1141 Hz, 1730 Hz, 2326 Hz and 2348 Hz could be computed exactly from the frequency equation given by the numerator of Eq. (7), which is equal to zero as explained in Section 2.3.2. These values represent natural frequencies of the system (VT+tube) when the glottis is not closed ($p_G = 0$).

We should note that the frequency range studied here (0–3.5 kHz) fits the domain of validity of 1D mathematical model of the VT for the vowel [u:]. As the maximum width of the mouth cavity was 62 mm for the artificial VT, and the maximum diameter of the mathematical model of the human VT was 41 mm, it can be calculated that the first transversal acoustic mode could exist above 2.77–2.85 kHz.

Since the estimated formant bandwidths $B_2$, $B_4$, and $B_5$ for the formant frequencies $F_2$ to $F_5$ are practically the same for the VT model and for the human VT (the difference is less than 10 percent: see Table 4), it can be concluded that the compliance of the human walls for higher formant frequencies does not influence the acoustic damping of the human VT cavity prolonged by the tube. The values of the measured bandwidths appear to be narrower than the values found in the literature for ordinary phonation without a tube: e.g. see [18]. An explanation can be found in the different methods of measurement. Here we used the transfer function measured between the lips and the open tube end, whereas the spectrum envelope is usually used as it is described, e.g. in [18]. We should also note that the glass tube used in this study has less damping than the human vocal tract itself and, depending on the acoustic mode shape, it can influence the resulting damping of the coupled system (VT+tube).

The computed results shown in Fig. 12 suggest that, for phonation with a resonance tube in the air, $F_1$ does not go lower than roughly 200 Hz. Thus, it would offer the beneficial effect of increased reactance below this level, i.e. at the fundamental frequencies typically found in speech. To optimize the effect of vocal tract reactance on phonation, males would probably need to use longer or narrower tubes, as suggested by the Story's results [9]. When the tube is long or narrow enough, the effective mechanism eventually changes from acoustic reactance to the magnitude of flow resistance.

## 4.2. Physical background for the influence of yielding walls on the acoustic resonances of the human vocal tract

Following our previous studies of acoustic structural interaction (see, for example, [19,20]), we can explain the influence of yielding walls in the vocal tract cavity on the acoustic frequencies by using a simplified model of a coupled mechanical–acoustical system. Similarly, Rabiner and Schaffer [21] considered the effect of yielding walls on acoustical resonances in a simple cylindrical cavity (compliant shell) simulating the vocal tract. However, they did not include the mechanical resonance of the structure in their model.

Let us consider the coupled system shown in Fig. 13, consisting of a simplified vocal tract cavity (1) and a tube (2) with cross-section areas $S_1$ and $S_2$ and lengths $L_1$, $L_2$. The glottis is closed by a yielding wall having a mass $M$ and vibrating with a displacement $w(t)$ on a spring of stiffness $K$. The resonance frequencies of such a coupled system are given in the solution of
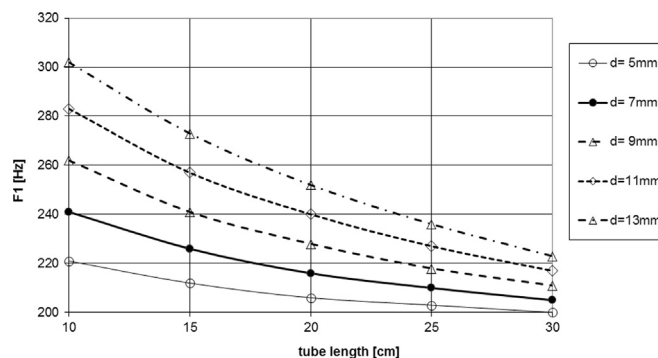


**Fig. 12.** First resonance frequency of the human VT prolonged by tubes of different lengths and diameters ($d$), calculated according to Story [9,12].
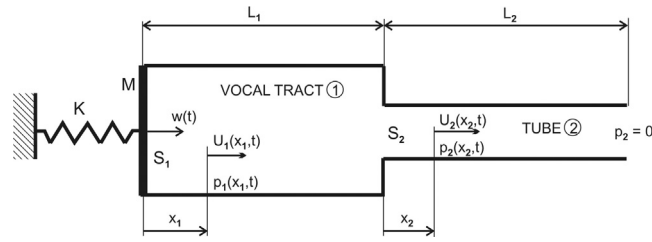
**Fig. 13.** Simplified model of the vocal tract prolonged by a tube and closed with a yielding wall.
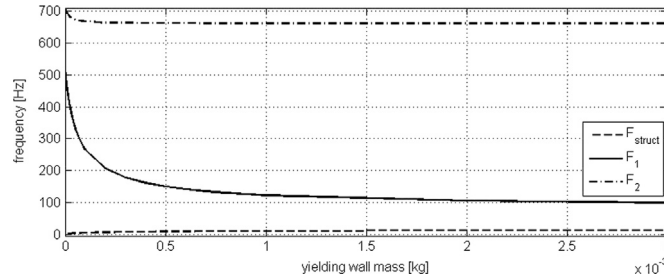


**Fig. 14.** Natural frequencies of the simplified model of the vocal tract prolonged by the tube and closed with the yielding wall.

the following transcendent frequency equation derived in Appendix B:

$$\omega^2 + \frac{\omega \rho_0 c_0 S_1}{M} \cdot \frac{S_2 \, \sin(kL_1)\cos(kL_2) + S_1 \, \cos(kL_1)\sin(kL_2)}{S_2 \, \cos(kL_1)\cos(kL_2) - S_1 \, \sin(kL_1)\sin(kL_2)} - \omega_0^2 = 0, \tag{9}$$

where $k = \omega/c_0$ is the wavenumber, $\omega$ is the angular frequency and $\omega_0 = \sqrt{K/M}$ is the eigenfrequency of the mechanical system.

The numerical solution of the frequency Eq. (9) was found for the parameters $\rho_0 = 1.2$ kgm$^{-3}$, $c_0 = 353$ ms$^{-1}$, $L_1 = 16.2$ cm, $L_2 = 26.4$ cm, $S_1 = 3.52$ cm$^2$, $S_2 = 0.36$ cm$^2$ in correspondence with the parameters used above for the acoustical system. The mechanical resonance frequency was kept constant at 15 Hz as in [12], $\omega_0 = 2\pi \cdot 15$. The resulting first three natural frequencies are shown in Fig. 14 for the mass $M$ varying from 0 to 3 g.

The results are in agreement with a general analysis of the solution of Eq. (9) presented in Appendix B for two extreme cases in which the mass $M$ goes to zero or to infinity. The lowest natural frequency $F_{struct}$ corresponds to the mechanical resonance $\omega_0$ and varies from $F_{struct} \rightarrow 0$ Hz for $M \rightarrow 0$ to $F_{struct} \rightarrow 15$ Hz for $M \cong 3$ g. The second natural frequency of the coupled system corresponds to the first acoustic resonance $F_1$, which is strongly influenced by coupling with the vibrating wall when the mass $M$ decreases below about two grams. The higher acoustic resonance $F_2$ is influenced by the vibrating wall in a much smaller range of the mass $M = 0$–0.2 g. Similarly, a small influence of the yielding wall on the second acoustic resonance, $F_2$, was found by Hanna et al. [22].

The acoustic resonances of the system for $M \rightarrow 0$ are equal to frequencies $F_1 = 507$ Hz and $F_2 = 710$ Hz, corresponding to the acoustic system with both ends opened: see Eq. (B.16) in Appendix B. The acoustic resonances for $M > 3$ g are equal to the acoustical resonances $F_1 = 84$ Hz and $F_2 = 661$ Hz, corresponding to the case when the vocal tract is closed by a rigid wall at the glottis: see Eq. (B.17). The frequency $F_1$ in this case corresponds to the artificial model of the vocal tract used in the experiments, in which the first resonance frequency was found to be in the range of 73–99 Hz; see Tables 1 and 2.

Titze et al. [23] presented an empirical relation for the vibrational thickness of a female vocal fold as

$$T_{VF} = \frac{10.63 \text{ mm}}{1 + 1.69 \, (f_0/190 \text{ Hz})}, \tag{10}$$

which results in $T_{VF} = 4.2$ mm for speaking frequency range $f_0 = 150$–160 Hz of our subject. Larsson and Hertegård [24] measured the length and width of the vocal folds at the speaking fundamental frequency for mezzo-soprano as $L_{VF} = 8.4$–11.3 mm and $W_{VF} = 3.4$–4.6 mm. Considering the simplicity of the triangular shape of female vocal folds [25] and density of the tissue $\rho_{VF} = 1020$ kgm$^{-3}$, we can calculate that the vocal folds' mass $M \cong 0.2$ g. Following the solid line in Fig. 14, we find that the first acoustical resonance frequency corresponding to 0.2 g is $F_1 = 206$ Hz. This value corresponds to $F_1 = 190$–208 Hz found for the human VT; see Table 3. Similarly, the second acoustical resonance frequency $F_2 = 665$ Hz corresponds to the $F_2$ measured in humans; see Table 3, in which $F_2 = 630$–680 Hz.

We can conclude that the substantial difference in $F_1$ caused by the inclusion of soft walls in the vocal tract model is in principle given by the acoustic-structural interaction of the acoustic cavity, semioccluded by the tube, with the yielding wall that can, for example, be created by the soft tissue in the larynx.

## 5. Conclusions

The in vitro experiments confirmed the legitimacy of assuming rigid walls in mathematical modelling of the acoustical resonance properties of the artificial models of a human VT with hard walls. However, similar in vivo experiments carried out in a real human VT showed a substantial difference between the first resonance (formant) frequency $F_1$ of the rigid wall VT model and the $F_1$ of real human VT acoustic cavities covered by soft tissues. The differences between the human data and the data from the rigid-wall VT model were negligible in the case of higher formants. The results of the present study confirmed the approach of Story et al. [9] by showing that a VT model with yielding walls is necessary when considering mathematical modelling of human vocal tracts prolonged by tubes. Finally, it was shown that the physical background for the system behaviour is in principle given by the acoustic-structural interaction of the semi-occluded vocal tract with a dynamical system originating from a low-frequency mechanical resonance of soft tissue at the glottis, for example the vocal fold tissue in the larynx.

## Acknowledgements

## Appendix A. A mathematical model of the vocal tract acoustics

A mathematical model of the vocal tract acoustics is based on an analytical solution of a 1D wave equation for acoustic wave propagation in the vocal tract cavity [9,12,26]:

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{1}{S} \cdot \frac{\partial S}{\partial x} \cdot \frac{\partial \phi}{\partial x} - \frac{1}{c_0^2} \cdot \left( \frac{\partial^2 \phi}{\partial t^2} + c_0 \cdot r_N \cdot \frac{\partial \phi}{\partial t} \right) = 0 \tag{A.1}$$

where $\phi$ is the flow velocity potential related to the acoustic pressure $p$ and the acoustic volume velocity $U$ by equations

$$p = -\rho_0 \partial \phi / \partial t - c_0 \rho_0 r_N \phi, \quad U = S \partial \phi / \partial x \tag{A.2}$$

$x$ is the longitudinal coordinate along the vocal tract measured from the vocal folds to lips; $t$ is time; $r_N$ is the specific acoustic resistance per a unit length; $S(x)$ is the cross-sectional area of the cavity; $c_0$ is the speed of sound; and $\rho_0$ is the fluid density.

The relationship between the acoustic pressure $p$ and the volume velocity $U$ at the input and output of each conical acoustic element can be described by the transfer matrix as (see [26])

$$\begin{bmatrix} p_{OUT} \\ U_{OUT} \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} p_{IN} \\ U_{IN} \end{bmatrix}, \tag{A.3}$$

where the elements of the transfer matrix are

$$a = \frac{\xi_0}{\xi_0 + L_e} \cdot \left( \cosh(\gamma L_e) + \frac{1}{\gamma \xi_0} \cdot \sinh(\gamma L_e) \right),$$

$$b = -\frac{c_0 \rho_0 (r_N + j k) \cdot \xi_0}{S_{IN} \cdot \gamma (\xi_0 + L_e)} \cdot \sinh(\gamma L_e),$$

$$c = S_{OUT} \cdot \frac{(1 - \gamma^2 \xi_0 (\xi_0 + L_e)) \cdot \sinh(\gamma L_e) - \gamma L \cdot \cosh(\gamma L_e)}{\gamma (\xi_0 + L_e)^2 \cdot c_0 \rho_0 (r_N + j k)},$$

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

$L_e$ is the length of the conical acoustic element; $S_{IN}$ and $S_{OUT}$ are the respective cross-sectional areas of the element input and output; $\gamma$ is a complex exponent given by the formulas

$$\gamma = \overline{\alpha} + j \overline{\beta}, \tag{A.4}$$

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

where $k = \omega / c_0$ is the wavenumber; $\omega$ is the angular frequency of the harmonic signal and $j = \sqrt{-1}$ is the imaginary unit.

The coefficient $\xi_0$ is defined by the input ($R_{IN}$) and output ($R_{OUT}$) radii of the element

$$\xi_0 = \frac{R_{IN}}{R_{OUT} - R_{IN}} \cdot L, \tag{A.5}$$

and the frequency-dependent viscous losses were considered as

$$r_N = \frac{2}{R_{OUT} - R_{IN}} \cdot \sqrt{2k\mu/c_0\rho_0}, \tag{A.6}$$

where $\mu$ is the dynamic air viscosity.

The acoustic properties of the vocal tract can be described in matrix form as

$$\begin{bmatrix} p_L \\ U_L \end{bmatrix} = \mathbf{T_{VT}} \cdot \begin{bmatrix} p_G \\ U_G \end{bmatrix}, \tag{A.7}$$

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

$$\mathbf{T_{VT}} = \begin{bmatrix} A_{VT} & B_{VT} \\ C_{VT} & D_{VT} \end{bmatrix} = \mathbf{T_N} \cdot \mathbf{T_{N-1}} \cdot \ldots \cdot \mathbf{T_2} \cdot \mathbf{T_1}, \tag{A.8}$$

and $N$ is the number of conical elements. The transfer matrix of the tube assumed to be one cylindrical element with rigid walls is given by Eq. (A.3) when $\xi_0 \to \infty$. The complete system (VT+tube) is described by the transfer matrix $\mathbf{T}$ given by the multiplication of transfer matrices of the VT and the tube:

$$\mathbf{T} = \mathbf{T_{TB}} \cdot \mathbf{T_{VT}} = \begin{bmatrix} A_{TB}A_{VT} + B_{TB}C_{VT} & A_{TB}B_{VT} + B_{TB}D_{VT} \\ C_{TB}A_{VT} + D_{TB}C_{VT} & C_{TB}B_{VT} + D_{TB}B_{VT} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}. \tag{A.9}$$

The matrix $\mathbf{T}$ describes the relationship between the variables at the glottis and the open tube end:

$$\begin{bmatrix} p_T \\ U_T \end{bmatrix} = \mathbf{T} \cdot \begin{bmatrix} p_G \\ U_G \end{bmatrix}. \tag{A.10}$$

The output pressure at the open tube end can be written as the output volume velocity by introducing the radiation impedance (see [7], for example):

$$p_T = Z_{T,RAD} \cdot U_T. \tag{A.11}$$

Putting Eq. (A.11) into Eq. (A.10), eliminating $p_G$ and taking into account the known property of the transfer matrices (see e.g. [26]), i.e.

$$\det(\mathbf{T}) = AD - BC = 1, \tag{A.12}$$

yields the expression for the transfer function of the complete system,

$$U_T(\omega)/U_G(\omega) = 1/(A - C \cdot Z_{T,RAD}), \tag{A.13}$$

or a more frequently used transfer function:

$$p_T(\omega)/U_G(\omega) = Z_{T,RAD}/(A - C \cdot Z_{T,RAD}). \tag{A.14}$$

Similarly, considering the vocal tract without a tube and loaded by the radiation impedance at the lips,

$$p_L = Z_{L,RAD} \cdot U_L, \tag{A.15}$$

yields the expression for the transfer function of the vocal tract

$$U_L(\omega)/U_G(\omega) = 1/(A_{VT} - C_{VT} \cdot Z_{L,RAD}). \tag{A.16}$$

To derive the acoustic input impedance, let us insert again Eq. (A.11) into Eq. (A.10). After eliminating $U_T$, we can express the ratio of the input pressure to input acoustic volume velocity as

$$Z_{IN} = \frac{p_G}{U_G} = \frac{D \cdot Z_{T,RAD} - B}{A - C \cdot Z_{T,RAD}}. \tag{A.17}$$

If the vocal folds were open during the measurement of the transfer function in vivo, the pressure at the glottis would be approximately zero ($p_G = 0$). Then we can eliminate $U_L$ from Eq. (1),

$$U_L = D_{VT}/B_{VT} \cdot p_L, \tag{A.18}$$

and putting Eq. (A.18) into Eq. (2) yields the measured transfer function

$$p_L(\omega)/p_T(\omega) = B_{VT}/(A_{TB}B_{VT} + B_{TB}D_{VT}). \tag{A.19}$$

The denominator in Eq. (A.19) is identical to the frequency equation of the system with O-O$Z_{RAD}$ boundary conditions as mentioned in Section 2.3.2 when omitting $Z_{T,RAD}$. Thus, the resonances of the measured transfer function $p_L(\omega)/p_T(\omega)$ would be approximately equal to the resonances of the complete system with the vocal folds open.

We assumed the output loaded by the acoustic radiation impedance of a vibrating circular plate with a radius, $R$, placed in an infinite wall (see e.g. [14]) to be

$$Z_{RAD} = \frac{c_0 \rho_0}{\pi R^2} \cdot \left[ 1 - \frac{J_1(2kR)}{kR} + j \frac{H_1(2kR)}{kR} \right], \tag{A.20}$$

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

## Appendix B. A mathematical model of acoustic-structural interaction

Considering the coupled mechanical–acoustical system shown in Fig. 13, the equation of motion for the mass $M$ vibrating on a spring of stiffness $K$ reads

$$M \ddot{w}(t) + K w(t) = F(t), \tag{B.1}$$

where $w$ is the translation of the mass, $\ddot{w}$ denotes the second derivative of $w$ with respect to time $t$ and $F$ is the force loading the mass by the pressure in the vocal tract:

$$F(t) = -S_1 \cdot p_1(x_1 = 0, t). \tag{B.2}$$

In the simplest case, when omitting the losses, wave Eq. (A.1) can be written as

$$\frac{\partial^2 \phi}{\partial x^2} - \frac{1}{c_0^2} \cdot \frac{\partial^2 \phi}{\partial t^2} = 0. \tag{B.3}$$

Considering a harmonic signal of an angular frequency $\omega$, the solution for the velocity potential can be found in the form

$$\phi_1(x_1, t) = \phi_{01}(x_1) \cdot e^{j\omega t}, \quad \phi_{01}(x_1) = \alpha_1 e^{jkx_1} + \beta_1 e^{-jkx_1}, \tag{B.4}$$

$$\phi_2(x_2, t) = \phi_{02}(x_2) \cdot e^{j\omega t}, \quad \phi_{02}(x_2) = \alpha_2 e^{jkx_2} + \beta_2 e^{-jkx_2}, \tag{B.5}$$

where indices 1 and 2 correspond respectively to the vocal tract and the tube.

The displacement, $w$, of the wall is given by

$$w(t) = w_0 \cdot e^{j\omega t}. \tag{B.6}$$

The unknown constants $\alpha_1, \beta_1, \alpha_2, \beta_2$ can be obtained from the following boundary and continuity conditions for the pressure, $p$, and the acoustic volume velocity $U$:

$$\dot{w} = U_1(x_1 = 0, t)/S_1, \tag{B.7}$$

$$p_2(x_2 = L_2, t) = 0, \tag{B.8}$$

$$p_1(x_1 = L_1, t) = p_2(x_2 = 0, t), \tag{B.9}$$

$$U_1(x_1 = L_1, t) = U_2(x_2 = 0, t). \tag{B.10}$$

Using Eq. (A.2), i.e. $p = -\rho_0 \partial \phi / \partial t$ and $U = S \partial \phi / \partial x$, conditions (B.7)–(B.10) yield a system of equations:

$$\alpha_1 - \beta_1 = c_0 w_0 \tag{B.11}$$

$$\alpha_2 e^{jkL_2} + \beta_2 e^{-jkL_2} = 0 \tag{B.12}$$

$$\alpha_1 e^{jkL_1} + \beta_1 e^{-jkL_1} = \alpha_2 + \beta_2 \tag{B.13}$$

$$S_1 \left( \alpha_1 e^{jkL_1} - \beta_1 e^{-jkL_1} \right) = S_2 (\alpha_2 - \beta_2) \tag{B.14}$$

The solution of Eqs. (B.11)–(B.14) results in complicated relations for constants $\alpha_1, \beta_1, \alpha_2, \beta_2$. After putting them into Eqs. (B.4) and (B.5), deriving the acoustic pressure $p_1(x_1, t)$ and inserting it into (B.2) and (B.1), we finally get the frequency equation

$$\omega^2 + \frac{\omega \rho_0 c_0 S_1}{M} \cdot \frac{S_2 \sin(kL_1) \cos(kL_2) + S_1 \cos(kL_1) \sin(kL_2)}{S_2 \cos(kL_1) \cos(kL_2) - S_1 \sin(kL_1) \sin(kL_2)} - \omega_0^2 = 0, \tag{B.15}$$

where $\omega_0^2 = K/M$ is a squared angular mechanical resonance frequency.

A brief analysis of frequency Eq. (B.15) can be done for two extreme cases. When the mass, $M$, goes to zero (and so does the stiffness, $K$, because we assume $\omega_0$ to be constant), then the numerator of the second element must be zero and thus, we obtain the solution $\omega = 0$ and the frequency equation for the open–open system:

$$S_2 \sin(kL_1) \cos(kL_2) + S_1 \cos(kL_1) \sin(kL_2) = 0. \tag{B.16}$$

If the mass, $M$, goes to infinity (and so does $K$), then $\omega=\omega_0$ or the denominator must be zero, which yields the frequency equation for the closed–open system:

$$S_2 \cos(kL_1)\cos(kL_2) - S_1 \sin(kL_1)\sin(kL_2) = 0. \tag{B.17}$$

# References

[1] G. Fant, *Acoustic Theory of Speech Production*, second edition, Mouton, S'Gravenage, 1960.
[2] J. Sundberg, *The Science of the Singing Voice*, N. Illinois University Press, DeKalb, Illinois, 1987.
[3] I.R. Titze, *Principles of Voice Production*, National Center for Voice and Speech, second printing, Iowa City, IA, 2000.
[4] A.M. Laukkanen, On speaking voice exercises. A study on the acoustic and physiological effects of speaking voice exercises applying manipulation of the acousticaerodynamical state of the supraglottic space and artificially modified auditory feedback [doctoral dissertation], *Acta Universitatis Tamperensis ser A* 445 (1995).
[5] S. Simberg, A. Laine, The resonance tube method in voice therapy: description and practical implementations, *Logopedics Phoniatrics Vocology* 32 (2007) 165–170.
[6] I.R. Titze, A.K. Verdolini, *Vocology. The Science and Practice of Voice Habilitation*, The National Center for Voice and Speech, Salt Lake City, 2012.
[7] I.R. Titze, 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.
[8] I.R. Titze, A.M. Laukkanen, Can vocal economy in phonation be increased with an artificially lengthened vocal tract? A computer modeling study, *Logopedics Phoniatrics Vocology* 32 (2007) 147–156.
[9] B.H. Story, A.M. Laukkanen, I.R. Titze, Acoustic impedance of an artificially lengthened and constricted vocal tract, *Journal of Voice* 14 (2000) 455–469.
[10] I.R. Titze, The physics of small-amplitude oscillation of the vocal folds, *Journal of the Acoustical Society of America* 83 (1988) 1536–1552.
[11] I.R. Titze, B.H. Story, Acoustic interactions of the voice source with the lower vocal tract, *Journal of the Acoustical Society of America* 101 (1997) 2234–2243.
[12] M.M. Sondhi, J. Schroeter, A hybrid time-frequency domain articulatory speech synthesizer, *IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP* 35 (1987) 955–967.
[13] M.M. Sondhi, Model for wave propagation in a lossy vocal tract, *Journal of the Acoustical Society of America* 55 (1974) 1070–1075.
[14] T. Leino, A.M. Laukkanen, V. Radolf, Formation of the actor's/speaker's formant: a study applying spectrum analysis and computer modeling, *Journal of Voice* 25 (2011) 150–158.
[15] T. Vampola, J. Horáček, J.G. Švec, FE modeling of human vocal tract acoustics. Part I: production of Czech vowels, *Acta Acustica united with Acustica* 94 (2008) 433–447.
[16] G. Widholm, BIAS 6 MANUAL Version 1, ART – Acoustics Research Team, 2008, ⟨http://www.bias.at/userfiles/files/000809artim.pdf⟩ (accessed 06.10.15).
[17] A.M. Laukkanen, J. Horáček, R. Havlík, Case-study magnetic resonance imaging and acoustic investigation of the effects of vocal warm-up on two voice professionals, *Logopedics Phoniatrics Vocology* 37 (2012) 75–82.
[18] G.M. Fant, On the predictability of formant levels and spectrum envelopes from formant frequencies, in: M. Halle, H. Lunt, M. MacLean (Eds.), *For Roman Jakobson*, Mouton and Company, Gravenhage, The Netherlands1956, pp. 109–120.
[19] J. Horáček, I. Zolotarev, Influence of the acoustic-structural couplings upon free vibration of mechanical systems, *Archives of Acoustics* 18 (1993) 297–309.
[20] D.G. Gorman, J. Horáček, Analysis of the free vibration of a coupled plate/fluid interacting system and interpretation using sub-system modal energy, *Engineeering Structures* 29 (2007) 754–762.
[21] L.R. Rabiner, R.W. Schafer, *Theory and Applications of Digital Speech Processing*, Prentice Hall, 2011.
[22] N. Hanna, J. Smith, J. Wolfe, Low frequency response of the vocal tract: acoustic and mechanical resonances and their losses, in: T. McGinn (Ed.), *Acoustics 2012 Fremantle: Proceedings of the Annual Conference of Australian Acoustical Society*, CD ROM. ISBN: 978-0-646-59039-4, 2012.
[23] I.R. Titze, J.G. Švec, P.S. Popolo, Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues, *Journal of Speech, Language, and Hearing Research* 46 (2003) 919–932.
[24] H. Larsson, S. Hertegård, Vocal fold dimensions in professional opera singers as measured by means of laser triangulation, *Journal of Voice* 22 (2008) 734–739.
[25] I.R. Titze, Physiologic and acoustic differences between male and female voices, *Journal of the Acoustical Society of America* 85 (1989) 1699–1707.
[26] V. Radolf, *Direct and Inverse Task in Acoustics of the Human Vocal Tract (Ph.D. thesis)*, Czech Technical University in Prague, 2010.