ARCHIVES OF ACOUSTICS
2, 4, 235 (1977)

A SIMULATIVE MODEL OF THE VOCAL TRACT INCLUDING THE EFFECT OF NASALIZATION*)

JANUSZ KACPROWSKI

Department of Cybernetic Acoustics, Institute of Fundamental Technological Research, Polish Academy of Sciences (Warszawa)

The subject and the aim of this work is the development of an articulatory model of the speech organ, adapted to computer simulation, for investigation of the influence of nasalization due to cleft palate upon the formant structure of originally non-nasal speech sounds, e.g. oral vowels, under realistic operating conditions from the acoustical, physiological and clinical points of view. This was achieved by reproducing the anatomy of the vocal tract under different articulatory conditions including all the possible individual variants, which depend on personal voice characteristics in both normal physiological states and in various pathological conditions. An essential part of the model was the taking into account of the losses in the vocal tract, the radiation impedance of the mouth and/or nose orifice, and the introduction of an additional parameter, which expresses in a quantitative manner the degree of nasalization, depending on the extent of the cleft palate. The model is intended for clinical applications in computer-aided acoustic diagnostics in phoniatry.

1. Introduction

The subject of an earlier work of the present author and his colleagues [12] was a graphical analysis of the influence of cleft palate (med.: *palatoschisis molle*) upon the transmission characteristics of the vocal tract, which determine the spectral structure of the oral vowels. Experimental verification was performed on an analogue model simulating, in an extremely simplified manner, the geometrical configuration of the vocal tract during the articulation of the nautral vowel Ǝ. The model had a scale of 5 : 1 dimensionally and, consequently, 1 : 5 in frequency. Good agreement was obtained between the analytical and experimental results. These preliminary investigations confirmed the possibility of objective evaluating the influence of cleft palate, which results in the forced nasalization of originally non-nasal speech sounds, upon the formant

*) This work was performed within the framework of the nodal problem 10.4.3.

structure of the oral vowels. This influence proved to be measurable and amenable to analytical treatment. It establishes many possibilities for the clinical application of acoustical methods in phoniatric diagnostics for the objective evaluation of pathological and postoperative states in cleft palate and for the objective control and documentation of the rehabilitation process.

The results of preliminary experiments, although encouraging from the cognitive point of view, proved to have limited clinical usefulness because of the simplifications introduced. The aim of the present work is to extend the previous results and to increase their accuracy by elaborating a universal articulatory model of the speech organ, which is adapted to computer simulation. This aim may be achieved by:

– reproducing the anatomy of the vocal tract in the articulation of different oral vowels (firstly for the Polish basic vowels [i, y, e, a, o, u]) under steady state conditions with all the possible individual variants, determined by the individual features of the patient's voice in both normal physiological states and pathological conditions;

– taking into account the losses existing in the vocal tract and the radiation impedance of the mouth and/or nose orifice;

– introducing an additional parameter, which expresses in a quantitative manner the degree of nasalization, which depends on the extent of cleft palate or on its actual advance.

2. An acoustical model of the vocal tract in the articulation of voiced sounds with glottal excitation

A simplified acoustical model of the supraglottal part of the speech organ is shown in Fig. 1a. The vocal pharynx-mouth tract $(G+U)$ may be approximated as a tube of irregular cross-sectional area given by the function $A(x)$, where x is the distance from the glottis. Under the normal anatomical conditions of an adult male the length $l_k = l_g + l_u$ of the pharynx-mouth tract, ending at the point $x = l_k$ by the mouth orifice, is of the order 16-19 cm, and its cross-sectional area ranges from some fraction of a square centimetre to about 15 cm^2. The pharynx tract is excited at the point $x = 0$ by the larynx generator with an internal impedance Z_g, which produces a volume velocity U_g in the glottal orifice. At the point $x \approx 8$ cm the nasal tract N is attached to the pharynx-mouth tract. The nasal tract has the form of an irregular tube of the length $l_n \approx 12.5$ cm and total volume equal to about 50 cm^3. The movable soft palate (med.: *epiglottis*) acts as a valve which controls the degree of acoustic coupling between the nasal and mouth tracts, i.e. the amount of nasalization of originally non-nasal speech sounds.

The acoustic energy of the speech wave is emitted through the radiation impedance Z_{pu} of the mouth and/or Z_{pn} of the nose. The corresponding volume velocities are U_u and U_n. These volume velocities produce, at a distance r

on the axis of symmetry of the radiating system, an acoustic pressure $p(r)$ in the speech wave, which is a linear superposition of two waves radiated through the mouth and nasal orifices[1]). It may be expressed as

$$p(r) = \frac{j\omega\varrho}{4\pi r}(U_u + U_n)e^{-j2\pi r/\lambda} \tag{1}$$

(for a spherical wave, $2\pi r \leqslant \lambda$), as is explained in [12].

The acoustic model of the articulatory effectors shown in Fig. 1a may be presented in the form of the equivalent electrical circuit shown in Fig. 1b, in which the continuous cross sectional area function $A(x)$ of the pharynx-mouth tract is approximated by a definite number q of uniform elementary segments, e.g. T-type four-poles, each of which represents a segment of the tract of length $l = l_k/q$ in the form of a cylindrical tube of cross-section A_i ($i = 1, 2, \ldots, q$). The accuracy of the approximation increases as the length l

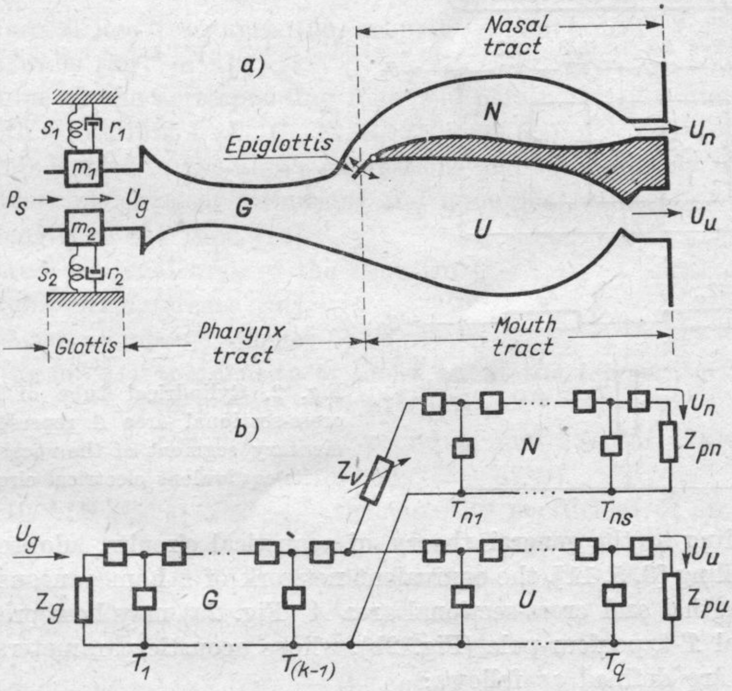

Fig. 1. Simplified acoustical model of the supraglottal part of the speech organ (a) and its electrical equivalent circuit (b)

P_S – subglottal pressure, (m_1, s_1, r_1) and (m_2, s_2, r_2) – vocal folds, U_g – volume velocity of the larynx source, Z_g – acoustic impedance of the glottis (larynx source), G – pharynx tract, U – mouth tract, N – nasal tract, U_u – volume velocity through the mouth orifice, U_n – volume velocity through the nostrils, Z_{pu} – radiation impedance of the mouth orifice, Z_{pn} – radiation impedance of the nose orifice, Z_v – acoustic impedance of the nose-to-mouth coupling

[1]) The component originating from the acoustic pressure wave, radiated by the external surface of the vibrating walls of the vocaltract, has been neglected.

of each cylindrical elementary segment decreases. The limiting condition is $l = \lambda_{\min}/8$, where $\lambda_{\min}$ is the wave-length at the highest frequency in the considered range, i.e. $f_{\max} \approx 5000$ Hz. It follows that the representation of the area function $A(x)$ of the pharynx-mouth tract in terms of $q \approx 20$ discrete elementary uniform segments of length $l = 1$ cm is quite satisfactory for modelling purposes, since the resulting «quantization noise» may be neglected. The electrical equivalent circuit in Fig. 1b constitutes the theoretical basis and starting point for the design and construction of many articulatory speech synthesizers, both static [2, 3, 17] and dynamic, i.e. controlled by electrical signals, initially analogue [6, 16] and subsequently digital [1].

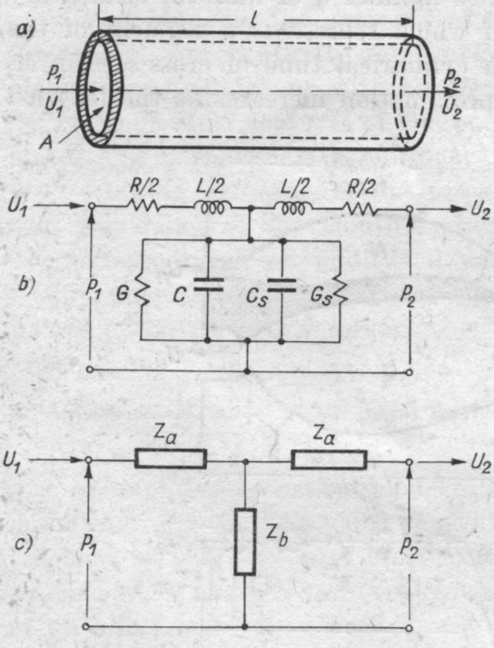

Fig. 2. Cylindrical tube of length l and cross-sectional area A representing an elementary segment of the vocal tract (a) and its equivalent electrical circuits (b, c)

According to the general theory of acoustical circuits, adopted for vocal tract modelling [3, 5, 11], the equivalent network of a homogeneous cylindrical tube of length l and cross-sectional area A (Fig. 2a) may be represented as a symmetrical T-type four-pole (Fig. 2b), whose acoustic parameters L, C, R, G, C_s and G_s are defined as follows:

$$L = L_j l = \varrho l/A$$

— the acoustic mass of the air in the tube [$\mathrm{kg\,m^{-4}}$];

$$C = C_j l = Al/\varrho c^2$$

— the acoustic compliance of the air in the tube [$\mathrm{kg^{-1}\,m^4\,s^2}$];

$$R = R_j l = \frac{S}{A^2}\sqrt{\omega\varrho\mu/2}\,l$$

– the acoustic viscous loss resistance at the tube walls [kg m^{-4} s^{-1}];

$$G = G_j l = S\frac{\eta-1}{\varrho c^2}\sqrt{\frac{\lambda\omega}{2c_p^3\varrho}}\, l$$

– the acoustic loss conductance due to heat conduction at the tube walls [kg^{-1} m^4 s];

$$C_s = C_{sj} l = -\frac{m_s S}{r_s^2+\omega^2 m_s^2}\, l$$

– the negative acoustic compliance equivalent to the reciprocal acoustic mass of the vibrating vocal tract walls [$kg^{-1}m^4$ s^2];

$$G_s = G_{sj} l = \frac{r_s S}{r_s^2+\omega^2 m_s^2}\, l$$

– the reciprocal loss resistance (i.e. acoustic conductance) of the vibrating vocal tract walls [kg^{-1} m^4 s].

The value of the corresponding four-pole parameters per unit length are labelled with subscripts j (L_j, C_j, R_j, G_j, C_{sj} and G_{sj}).

The symbols used in the above expressions and subsequently in this paper have the following physical definitions and numerical values:

l – the length of the tube [m],
A – the cross-sectional area of the tube [m^2],
S – the tube circumference [m],
$\omega = 2\pi f$ – the angular frequency [s^{-1}],
$\varrho = 1.14$ [kg m^{-3}] – the density of moist air at the temperature of the human body ($t = 37°$C),
$c = 350$ [m s^{-1}] – the velocity of sound in moist air at the temperature of the human body ($t = 37°$C),
$\mu = 1.86\cdot10^{-5}$ [N m^{-2} s] – the viscosity coefficient of air ($t = 20$ °C, $P_0 = 0.76$ m Hg),
$\lambda = 55\cdot10^{-4}$ [cal m^{-1} s^{-1} $°C^{-1}$] – the thermal conductivity of air,
$c_p = 0.24\cdot10^3$ [calkg^{-1} $°C^{-1}$] – the specific heat of air at constant pressure $P_0 = 0.76$ m Hg,
$\eta = 1.4$ – the adiabatic constant of air,
$r_s \approx 16\cdot10^3$ [kg m^{-2} s^{-1}] – the loss resistance of the tissue of the vocal-tract walls per unit area[2]),
$m_s \approx 15$ [kg m^{-2}] – the mass of the tissue of the vocal-tract walls per unit area[2]).

[2]) The numerical values of r_s and m_s were determined by direct measurements of the surface impedance of the vocal-tract wall tissue [9], and then experimentally verified in model investigations [15].

The physical meaning of the parameters C_s and G_s calls for a brief explanation. The vocal-tract walls are not infinitely stiff, as was assumed in simplified considerations (see e.g. [12]), but have a finite acoustic impedance per unit area z_s. When exposed to local acoustic pressure in the vocal tract, they are forced to vibrate. The reactive component of the impedance z_s at frequencies above the resonance frequency of the walls, i.e. when $f > 100$ Hz, has an inertial character. The impedance z_s may thus be represented by a series connection of the loss resistance r_s and the mass m_s,

$$z_s = r_s + j\omega m_s, \tag{2}$$

where z_s, r_s and m_s are quantities per unit area. The impedance z_s, which corresponds to the energy dissipated by the wall vibrations, ought (on formal grounds) to be included in the transverse branch of the four-pole T in Fig. 2b. It is thus convenient to present z_s as the parallel connection of a real (resistive) and an imaginary (reactive) component in the form of an admittance per unit area $y_s = z_s^{-1}$:

$$y_s = \frac{r_s}{r_s^2 + \omega^2 m_s^2} - j\omega \frac{m_s}{r_s^2 + \omega^2 m_s^2}. \tag{3}$$

Multiplying y_s by the tube circumference S, we obtain the acoustic admittance per unit length of the walls,

$$Y_{sj} = \frac{r_s S}{r_s^2 + \omega^2 m_s^2} - j\omega \frac{m_s S}{r_s^2 + \omega^2 m_s^2} = G_{sj} + j\omega C_{sj}, \tag{4}$$

where G_{sj} is the acoustic conductance per unit length of the walls, and C_{sj} is the negative acoustic compliance, equivalent to the reciprocal mass, per unit length of the tube walls.

If the walls of the tube are infinitely hard, $m_s = \infty$ and $r_s = \infty$; thus $Y_{sj} = 0$ and the transverse admittance of the four-pole T in Fig. 2b is simply $Y = G + j\omega C$.

Using the general transmission line theory, we can transmute the equivalent circuit of a tube of length l, shown in Fig. 2b as a T-type four-pole, to the form presented in Fig. 2c. Expressing the acoustic impedance per unit length of the longitudinal branches of the four-pole as

$$Z_j = R_j + j\omega L_j = \frac{S}{A^2}\sqrt{\omega\varrho\mu/2} + j\omega\frac{\varrho}{A} \tag{5}$$

and the acoustic admittance per unit length of the transverse branch of the four-pole as

$$\begin{aligned} Y_j &= (G_j + G_{sj}) + j\omega(C_j + C_{sj}) \\ &= \left(S\frac{\eta - 1}{\varrho c^2}\sqrt{\frac{\lambda\omega}{2c_p\varrho}} + \frac{r_s S}{r_s^2 + \omega^2 m_s^2}\right) + j\omega\left(\frac{A}{\varrho c^2} - \frac{m_s S}{r_s^2 + \omega^2 m_s^2}\right), \end{aligned} \tag{6}$$

we obtain the following expressions for the characteristic impedance Z_0 and the propagation constant γ of the tube[3]):

$$Z_0 = \sqrt{Z_j/Y_j} = \sqrt{(R_j+j\omega L_j)/[(G_j+G_{sj})+j\omega(C_j+C_{sj})]}, \qquad (7)$$

$$\gamma = a+j\beta = \sqrt{Z_j Y_j} = \sqrt{(R_j+j\omega L_j)[(G_j+G_{sj})+j\omega(C_j+C_{sj})]}. \qquad (8)$$

From expressions (7) and (8) for the impedance parameters z_a and z_b of the four-pole T in Fig. 1c we get

$$z_a = Z_0 \operatorname{tgh}(\gamma l/2) \approx Z_0 \gamma l/2 = Z_j \frac{l}{2} = (R_j+j\omega L_j)\frac{l}{2} = \frac{1}{2}(R+j\omega L) = \frac{1}{2} Z, \qquad (9)$$

$$\begin{aligned} z_b &= \frac{Z_0}{\sinh(\gamma l)} \approx \frac{Z_0}{\gamma l} = \frac{1}{[(G_j+G_{sj})+j\omega(C_j+C_{sj})]l} \\ &= [(G+G_s)+j\omega(C+C_s)]^{-1} = Y^{-1}, \end{aligned} \qquad (10)$$

where $Z = Z_j l$ and $Y = Y_j l$ (cf. (5) and (6)).

The approximations applied in expressions (9) and (10), which consist in replacing the hyperbolic functions tgh x and sinh x by the values of the argument x, that is by the first term of their series expansions, are valid for small values of the argument x. This condition is fulfilled in the case considered of a short section of tube, where $l \leqslant \frac{1}{8}\lambda_{\min}$.

3. The pharynx-mouth tract including the shunting effect of the nasal tract

According to the earlier assumptions, the pharynx-mouth tract can be approximated by a cascade connection (or chain) of q T-type four-poles (cf. Fig. 1b) each of which represents an elementary segment of the vocal tract in the form of a cylindrical tube of length $l = 1$ cm and cross-sectional area A_i ($i = 1, 2, \ldots, q$). The quantities q and A_i are considered as variable parameters which describe the actual geometric configuration of the vocal tract depending on the personal physio-pathological features of the patient's speech organ and on the temporary articulation conditions expressed in terms of the vocal-tract cross-sectional area function $A(x)$.

Using the scarce data which may be found in the literature (see e.g. [3, 15, 17] and adapting it to the articulatory conditions of Polish vowels, we have shown in Table 1 the discrete values of the $A(x)$ area function of the basic vowels [i, y, e, a, o, u] for different values of x in $\Delta x = 1$ cm steps. The data should be treated as preliminary information which corresponds to the average normal physiological conditions of an adult male and which calls for individual verification. The eventual correction should be based on an experimental comparison of the pole frequencies of the vocal-tract transfer function $H(\omega)$

[3]) Depending on the relative values of the parameters $R_j, L_j, G_j, G_{sj}, C_j$ and C_{sj}, some of them may be neglected in numerical calculations. In the particular case of a lossless tube, the parameters R_j, G_j, G_{sj} and C_{sj} are equal to zero, whence

$$Z_0 = \sqrt{L_j/C_j} = \varrho c/A, \quad \gamma = j\omega\sqrt{L_j C_j} = j\omega/c = j\beta, \quad a = 0.$$

Table 1. Discrete values of the cross-sectional area function $A(x)$ of the vocal tract in the articulation of Polish oral vowels [i, y, e, a, o, u] (approximate data).

Vowel		[i]	[y]	[e]	[a]	[o]	[u]
Distance from glottis x [cm]		Cross-sectional area of the vocal tract $A(x)$ [cm²]					
Pharynx tract	1	3.0	3.0	2.3	2.0	2.5	2.6
	2	2.0	3.0	1.4	1.0	1.5	2.0
	3	10.0	12.0	6.5	2.3	6.0	8.0
	4	13.0	12.0	6.5	3.0	5.0	10.0
	5	13.0	11.0	8.0	4.3	3.2	5.3
	6	12.8	8.5	10.5	5.0	1.4	2.5
	7	12.8	7.5	10.5	1.7	0.7	3.5
Bifurcation	8	11.7	2.0	9.5	2.0	1.4	2.5
Mouth tract	9	7.3	1.0	7.3	3.0	2.3	2.5
	10	3.6	1.0	5.7	5.0	2.6	2.2
	11	2.2	2.0	4.5	6.5	4.0	2.2
	12	0.6	6.0	4.0	8.0	5.7	2.5
	13	0.6	8.0	3.3	9.0	8.0	4.0
	14	0.5	6.0	3.3	9.5	10.3	7.5
	15	0.6	3.0	3.3	9.0	10.3	13.0
	16	1.5	5.0	5.0	8.8	10.3	13.0
	17	4.0	8.0	7.0	8.0	8.0	4.0
	18	—	—	—	—	4.8	0.3
	19	—	—	—	—	—	0.7

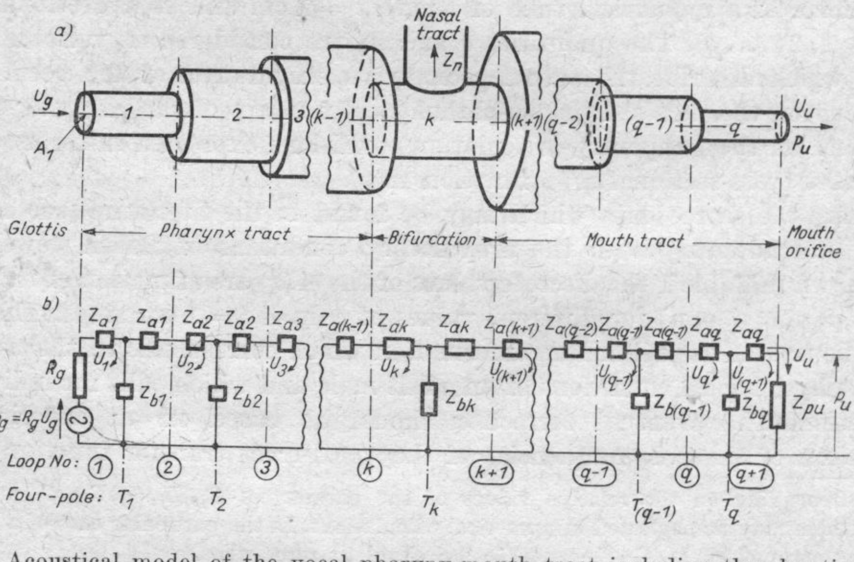

Fig. 3. Acoustical model of the vocal pharynx-mouth tract including the shunting effect of the nasal tract (a) and its equivalent electrical circuit (b)

with the formant frequencies of the different natural isolated vowels spoken by the subject under examination. The average values of the formant frequencies of Polish basic vowels measured by JASSEM may be found in [10].

An accustic model of the vocal pharynx-mouth tract (including the shunting effect of the nasal tract) and the corresponding electrical equivalent circuit are shown in Fig. 3. The k-th[4]) segment of the model and the corresponding equivalent T_k-section represent the bifurcation of the vocal tract (med.: *nasopharynx*), i.e. the space in which the input impedance Z_n of the nasal tract shunts the pharynx-mouth tract constituting the only path of energy flow in purely oral articulation. The section T_k differs with regard to its structure from other sections of the vocal tract and may be represented in the form shown in Fig. 4 which may be used to determine the longitudinal impendances z_{ak}

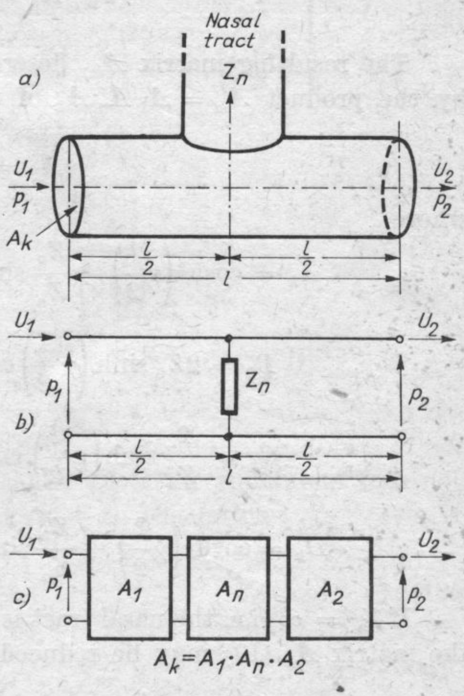

Fig. 4. Acoustical model of the vocal-tract's bifurcation (a) and its equivalent electrical circuits in the form of a transmission line shunted by the impedance Z_n (b) and in the form of a cascade connection of three four-poles described by their chain matrices $\boldsymbol{A}_1$, $\boldsymbol{A}_n$, $\boldsymbol{A}_2$ (c)

and the transverse impedance z_{bk} of the T_k four-pole, depending on the actual value of the shunting impedance Z_n. For this reason the k-th segment of length l and cross-sectional area A_k were replaced by two identical subsegments, each of length $l/2$, between which the shunting impedance Z_n, representing the input impedance of the nasal tract, is located (see Fig. 4b). In this way the equivalent electrical circuit of the bifurcation has been shown in the form of a cascade connection of three four-poles described by their chain matrices $\boldsymbol{A}$.

[4]) Provided that under normal physiological conditions the vocal tract bifurcates into the mouth and nasal tract at a distance $x \approx 8$ cm from the glottis, $k = 8$.

Two extreme four-poles represent the left and the right halves of the bifurcation segment, and their matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are equal,

$$\boldsymbol{A}_1 = \boldsymbol{A}_2 = \begin{bmatrix} \cosh\left(\gamma \frac{l}{2}\right) & Z_0 \sinh\left(\gamma \frac{l}{2}\right) \\ \frac{\sinh\left(\gamma \frac{l}{2}\right)}{Z_0} & \cosh\left(\gamma \frac{l}{2}\right) \end{bmatrix}, \tag{11}$$

while the middle four-pole, which represents the shunting impedance Z_n, is described by a matrix $\boldsymbol{A}_n$ of the form

$$\boldsymbol{A}_n = \begin{bmatrix} 1 & 0 \\ 1/Z_n & 1 \end{bmatrix}. \tag{12}$$

The resulting matrix $\boldsymbol{A}_k$, describing the k-th segment, may be expressed by the product $\boldsymbol{A}_k = \boldsymbol{A}_1 \boldsymbol{A}_n \boldsymbol{A}_2$ of the three matrices as

$$\boldsymbol{A}_k = \begin{bmatrix} A_k & B_k \\ C_k & D_k \end{bmatrix}, \tag{13}$$

where

$$A_k = \cosh^2\left(\gamma \frac{l}{2}\right) + \frac{Z_0}{Z_n} \sinh\left(\gamma \frac{l}{2}\right) \cosh\left(\gamma \frac{l}{2}\right) + \sinh^2\left(\gamma \frac{l}{2}\right),$$

$$B_k = 2Z_0 \sinh\left(\gamma \frac{l}{2}\right) \cosh\left(\gamma \frac{l}{2}\right) + \frac{Z_0^2}{Z_n} \sinh^2\left(\gamma \frac{l}{2}\right),$$

$$C_k = \frac{2}{Z_0} \sinh\left(\gamma \frac{l}{2}\right) \cosh\left(\gamma \frac{l}{2}\right) + \frac{1}{Z_n} \cosh^2\left(\gamma \frac{l}{2}\right),$$

$$D_k = \cosh^2\left(\gamma \frac{l}{2}\right) + \frac{Z_0}{Z_n} \sinh\left(\gamma \frac{l}{2}\right) \cosh\left(\gamma \frac{l}{2}\right) + \sinh^2\left(\gamma \frac{l}{2}\right).$$

If $Z_n = \infty$ (i.e. the nasal tract is disconnected and there is no nasalization), the matrix $\boldsymbol{A}_k$ (13) may be reduced to the simple form

$$\boldsymbol{A}_k = \begin{bmatrix} \cosh(\gamma l) & Z_0 \sinh(\gamma l) \\ \frac{1}{Z_0} \sinh(\gamma l) & \cosh(\gamma l) \end{bmatrix}, \tag{14}$$

which describes the four-pole network of a homogeneous tube segment of length l. This may have been expected and confirms the theoretical assumptions and the correctness of mathematical manipulations.

Using the well-known formulae of matrix algebra applied in the theory of electrical four-poles (see e.g. [14]), we can determine the impedances z_{ak} of the longitudinal branches and the impedance z_{bk} of the transverse branch of the four-pole T_k by formulae:

$$z_{ak} = \frac{A_k - 1}{C_k}, \tag{15a}$$

$$z_{bk} = \frac{1}{C_k}, \tag{15b}$$

where A_k and C_k are elements of the matrix $\boldsymbol{A}_k$ (formula (13)).

Using the previously justified approximations: $\sinh(\gamma l/2) \approx \gamma l/2$ and $\cosh(\gamma l/2) \approx 1$ which are valid provided $l \leqslant \lambda/8$, we can reduce the matrix $\boldsymbol{A}_k$ (formula (13)) to a simpler form,

$$\boldsymbol{A}_k = \begin{bmatrix} 1+\left(\gamma\frac{l}{2}\right)^2 + \frac{Z_0}{Z_n}\left(\gamma\frac{l}{2}\right), & 2Z_0\left(\gamma\frac{l}{2}\right) + \frac{Z_0^2}{Z_n}\left(\gamma\frac{l}{2}\right)^2 \\ \frac{2}{Z_0}\left(\gamma\frac{l}{2}\right) + \frac{1}{Z_n}, & 1+\left(\gamma\frac{l}{2}\right)^2 + \frac{Z_0}{Z_n}\left(\gamma\frac{l}{2}\right) \end{bmatrix}, \tag{16}$$

from which, according to formulae (15), we obtain

$$z_{ak} = Z_0\left(\gamma\frac{l}{2}\right)\frac{Z_0 + Z_n\left(\gamma\frac{l}{2}\right)}{Z_0 + 2Z_n\left(\gamma\frac{l}{2}\right)} \tag{17a}$$

and

$$z_{bk} = \frac{Z_0}{\gamma l}\,\frac{Z_n}{Z_n + Z_0/\gamma l}. \tag{17b}$$

As before, if $Z_n\ \infty$, the impedances z_{ak} and z_{bk} of the equivalent four-pole representing the bifurcation of the tract reduce to the forms

$$z_{ak} = Z_0\frac{\gamma l}{2} \tag{18a}$$

and

$$z_{bk} = \frac{Z_0}{\gamma l} \tag{18b}$$

which express the longitudinal impedances z_a and the transverse impedance z_b of a symmetrical T-type four-pole replacing a homogeneous cylindrical tube of length l (cf. formulae (9) and (10)).

Without jeopardizing the accuracy of the numerical calculations, the k-th segment and the corresponding four-pole T_k, which represent the bifurcation of the vocal tract, may be considered as lossless circuits with $Z_0 = \varrho c/A_k$ and $\gamma = j\omega/c$. This implies further simplification of the structure of the four-pole

T_k, whose impedance elements z_{ak} (17a) and z_{bk} (17b) reduce to

$$z_{ak} = j\frac{\omega l\varrho}{2A_k}\,\frac{\varrho c/A_k + jZ_n\omega l/2c}{\varrho c/A_k + jZ_n\omega l/c} \tag{19a}$$

and

$$z_{bk} = -j\frac{\varrho c^2}{A_k\omega l}\,\frac{Z_n}{Z_n - j\varrho c^2/A_k\omega l}, \tag{19b}$$

where c is the velocity of sound in moist air at 37 °C (the temperature of the human body) and $\omega = 2\pi f$ is the angular frequency.

The larynx tone generator is shown in Fig. 3b as a voltage (pressure) source of electromotive force (acoustic pressure) $p_g = R_g U_g$ and internal resistance R_g. The source parameters U_g and R_g may be determined using the equivalent (a.c.) circuit of the larynx generator after FLANAGAN [4]. This takes into account only the a.c. components $U_g'(t)$, $A_g'(t)$ and $P_s'(t)$, of the flow (volume velocity) function

$$U_g(t) = U_{g0} + U_g'(t),$$

of the glottis area function

$$A_g(t) = A_{g0} + A_g'(t)$$

and of the subglottal pressure

$$P_s(t) = P_{s0} + P_s'(t)$$

for small signal amplitudes $\big(U_g'(t) \ll U_{g0},\ A_g'(t) \ll A_{g0},$ and $P_s'(t) \ll P_{s0}\big)$. Under these conditions

$$U_g \approx \left(\frac{2P_{s0}}{\varrho}\right)^{1/2} A_g'(t) \tag{20}$$

and

$$R_g \approx (R_v + 2R_k)_{P_{s0},A_{g0}}. \tag{21}$$

In equation (21)

$$R_v = \frac{12\mu d l^2}{A_{g0}^3}$$

is the classical viscous loss resistance in the glottal slit of depth d, length l and rest area A_{g0}, whereas

$$R_k = 0.875\frac{\varrho U_{g0}}{2A_{g0}^2} = 0.875\frac{(2\varrho P_{s0})^{1/2}}{2A_{g0}}$$

is the kinetic loss resistance due to the transformation of the pressure drop P_{s0} in the glottis into the kinetic energy of the air flow (cf. [13]).

Under the normal anatomical conditions of an adult male ($d \approx 3$ mm,

$l \approx 18$ mm, $A_{g0} \approx 5$ mm^2) and at average voice effort ($P_{s0} \approx 10$ cm H_2O) the source resistance R_g (21) is of the order $100 \cdot 10^5$ MKS acoustic ohms and is much greater than the vocal-tract input impedance Z_t [5]).

4. The transmission function of the pharynx-mouth tract

The electrical equivalent circuit of the vocal pharynx-mouth tract in Fig. 3b may be described by Kirchhoff's equations for $(q+1)$ independent loops of the circuit:

$$\text{Loop No. 1:} \quad Z_{11} U_1 + Z_{12} U_2 = R_g U_g, \tag{22a}$$

$$\text{Loop No. 2:} \quad Z_{21} U_1 + Z_{22} U_2 + Z_{23} U_3 = 0, \tag{22b}$$

$$\text{Loop No. 3:} \quad Z_{32} U_2 + Z_{33} U_3 + Z_{34} U_4 = 0, \tag{22c}$$

. . . .

$$\text{Loop No. } i: \quad Z_{i(i-1)} U_{(i-1)} + Z_{ii} U_i + Z_{i(i+1)} U_{(i+1)} = 0, \tag{22d}$$

. . . .

$$\text{Loop No. } (q+1): \quad Z_{(q+1)q} U_q + Z_{(q+1)(q+1)} U_{(q+1)} = 0, \tag{22e}$$

where

$$Z_{ii} = (\underbrace{z_{a(i-1)} + z_{b(i-1)}}_{z_{d(i-1)}}) + (\underbrace{z_{ai} + z_{bi}}_{z_{di}}) = z_{d(i-1)} + z_{di} \tag{23}$$

is the *self impedance* of the i-th loop, while

$$Z_{(i-1)i} = Z_{i(i-1)} = -z_{b(i-1)} \tag{24a}$$

and

$$Z_{i(i+1)} = Z_{(i+1)i} = -z_{bi} \tag{24b}$$

are the *mutual impedances* of the adjacent loops $(i-1)i$ (formula (24a)) or $i(i+1)$ (formula (24b)), respectively. The defining equations (23) and (24) are valid for loops numbered from 2 to q inclusively, i.e. for $2 \leqslant i \leqslant q$. On account of the assymetry of loops No 1 and $(q+1)$, the following notation was introduced for their self and mutual impedances:

$$Z_{11} = z_{a1} + z_{b1} + R_g = z_{d1} + R_g, \tag{25a}$$

$$Z_{12} = Z_{21} = -z_{b1} \tag{25b}$$

and

$$Z_{(q+1)(q+1)} = z_{aq} + z_{bq} + Z_{pu} = z_{dq} + z_{pu}, \tag{26a}$$

$$Z_{q(q+1)} = Z_{(q+1)q} = -z_{bq}, \tag{26b}$$

where Z_{pu} is the radiation impedance of the mouth orifice.

---

[5]) Except in the neighbourhood of the first ($F1$) and the second ($F2$) formant range of vowels (see e.g. [13]).

The loop equations (22) may be represented in the form of the matrix equation $\boldsymbol{Z}_{GU}\boldsymbol{U} = \boldsymbol{P}$, in which

$$\boldsymbol{Z}_{GU} = \left[\begin{array}{cccc|cc} Z_{11} & Z_{12} & 0 & 0 & 0 & 0 \\ Z_{21} & Z_{22} & Z_{23} & 0 & 0 & 0 \\ 0 & Z_{32} & Z_{33} & Z_{34} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & Z_{(q+1)q} & Z_{(q+1)(q+1)} \end{array}\right] \tag{27}$$

is the loop-impedance matrix of the $(q+1)$-th order. The elements lying on the main diagonal are the self impedances of the particular loops, and elements beyond the main diagonal are the mutual impedances of adjacent loops. The characteristic feature of this matrix is that in each i-th row $(1 \leqslant i \leqslant q+1)$, only the elements $Z_{i(i-1)}$, Z_{ii} and $Z_{i(i+1)}$ differ from zero.

From equations (22) the volume velocity U_i in an arbitrary i-th loop of the equivalent circuit in Fig. 3b may be directly determined and calculated according to the general formula

$$U_i = \frac{\Delta_{1i}}{\Delta} R_g U_g, \tag{28}$$

where $\Delta = \det \boldsymbol{Z}_{GU}$ is the characteristic determinant of the impedence matrix (27), and Δ_{1i} is the cofactor of the element Z_{1i}. The transmission function $H_u(\omega)$ of the pharynx-mouth tract (including nasalization due to the shunting effect of the nasal tract), expressed as the ratio of the volume velocity $U_{(q+1)}$ through the radiation impedance Z_{pu} of the mouth orifice to the volume velocity $U_g \approx U_1$ through the glottis, is thus equal to

$$H_u(\omega) = \frac{U_{(q+1)}}{U_1} = \frac{\Delta_{1(q+1)}}{\Delta_{11}}. \tag{29}$$

Similarly, the transmission function $H'_u(\omega)$, expressed as the ratio of the acoustic pressure $p_u = Z_{pu} U_{(q+1)}$ in the mouth orifice to the volume velocity $U_g \approx U_1$ through the glottis, is equal to

$$H'_u(\omega) = \frac{p_u}{U_g} = \frac{Z_{pu}\, U_{(q+1)}}{U_1} = \frac{\Delta_{i(q+1)}}{\Delta_{11}} Z_{pu}, \tag{30}$$

where Z_{pu} is the radiation impedance of the mouth orifice, simulated by a circular piston of area A_q vibrating in a infinitely large flat baffle. The impedance Z_{pu} is given by the approximate expression [5, 12]

$$Z_{pu} \approx \frac{\omega^2 \varrho}{2\pi c} + j\omega \frac{8\varrho}{3\pi\sqrt{\pi A_q}}, \tag{31}$$

where ϱ is the density of air, c – the velocity of sound, and $\omega = 2\pi f$ – the angular frequency.

The use of formulae (29) and (30) in numerical calculations and in computer simulation must be based on a knowledge of the numerical data for both the cross-sectional area function $A(x)$ of the vocal tract under definite articulation conditions (cf. Table 1) and the input impedance Z_n of the nasal tract shunting the vocal tract at the point of bifurcation.

5. The input impedance of the nasal tract

In view of the scarcity of sufficiently representative data for determining the anatomical structure of the nasal tract, e.g. from X-ray pictures of the human head (cf. [3], Fig. 2.4-2), its cross-sectional area function $A_n(x)$, as simulated in most analogue speech synthesizers, is usually extremely simplified (see [6, 7, 8, 17]). The simplification is reasonable since the shape of the area function $A_n(x)$ does not depend on the articulatory conditions and is constant except for differences between individuals.

Using data from the medical and phonetic literature, we approximated the nasal tract for the present work in the form of a cascade connection of five homogeneous cylindrical tubes of lengths l_i and cross-sectional areas A_{ni}, where $i = 1, 2, \ldots, 5$ (Fig. 5). In Table 2 the numerical values of the parameters

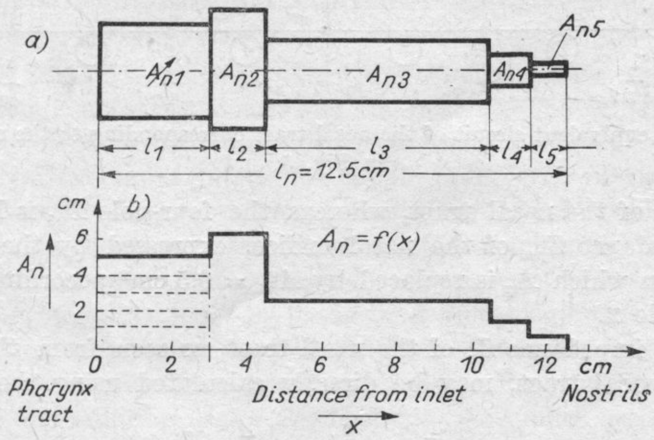

Fig. 5. Acoustical model of the nasal tract (a) and the area function $A_n(x)$ of its cross-section (b)

l_i and A_{ni} are presented and describe the geometric configuration of the model which corresponds to the average anatomical conditions of an adult male. The only variable parameter is the cross-sectional area A_{n1} of the first cylindrical segment, whose value determines the actual degree of nasalization of the vowel considered, which depends on the extent of cleft palate. In this way, the input impedance Z_n of the nasal channel, shunting the pharynx-mouth tract, is determined – ceteris paribus – by the value of A_{n1}.

Table 2. Numerical values of the parameters l_i, A_{ni} ($i = 1, 2, \ldots, 5$) determining the geometrical configuration of the nasal tract in Fig. 5

Segment No	1	2	3	4	5
Length l_i [cm]	3.0	1.5	6.0	1.0	1.0
Cross-sectional area A_{ni} [cm²]	variable	6.0	2.0	1.2	0.5

As for the pharynx-mouth tract considered above, the electrical equivalent circuit of the nasal tract, shown in Fig. 5, may be presented in the form of a cascade connection of five four-poles T_{ni} ($i = 1, 2, \ldots, 5$) which simulate the successive segments of the acoustical model, as is shown in Fig. 6. The four-pole T_{ni} is supplied from a voltage source which represents the acoustic pressure

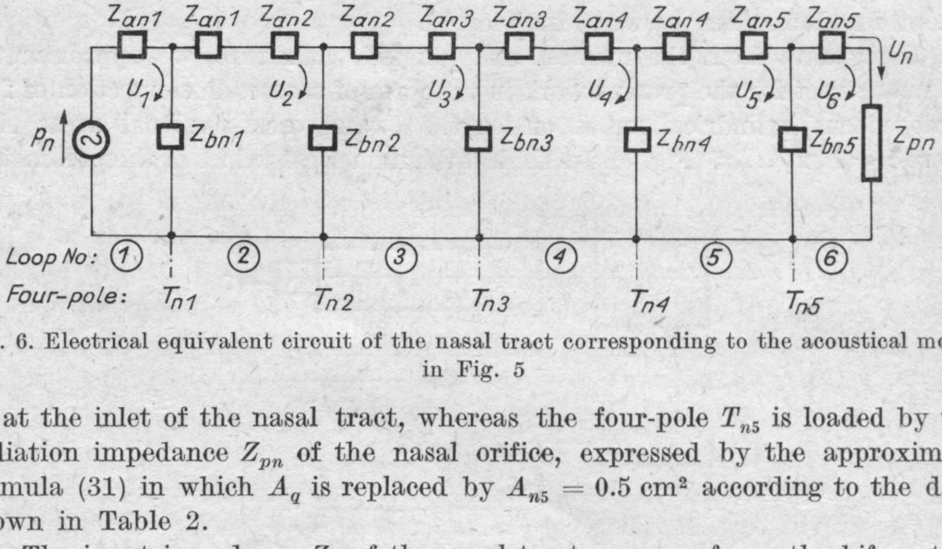

Fig. 6. Electrical equivalent circuit of the nasal tract corresponding to the acoustical model in Fig. 5

p_n at the inlet of the nasal tract, whereas the four-pole T_{n5} is loaded by the radiation impedance Z_{pn} of the nasal orifice, expressed by the approximate formula (31) in which A_q is replaced by $A_{n5} = 0.5$ cm² according to the data shown in Table 2.

The input impedance Z_n of the nasal tract, as seen from the bifurcation point of the vocal tract, may be directly calculated using the formula

$$Z_n = \frac{p_n}{U_1} = \frac{\Delta}{\Delta_{11}}, \tag{32}$$

in which Δ is the characteristic determinant, and Δ_{11} is the cofactor of the element Z_{11} of the loop-impedance matrix $\boldsymbol{Z}_N$,

$$\boldsymbol{Z}_N = \begin{bmatrix} Z_{11} & Z_{12} & 0 & 0 & 0 & 0 \\ Z_{21} & Z_{22} & Z_{23} & 0 & 0 & 0 \\ 0 & Z_{32} & Z_{33} & Z_{34} & 0 & 0 \\ 0 & 0 & Z_{43} & Z_{44} & Z_{45} & 0 \\ 0 & 0 & 0 & Z_{54} & Z_{55} & Z_{56} \\ 0 & 0 & 0 & 0 & Z_{65} & Z_{66} \end{bmatrix}, \tag{33}$$

which describes the model of the nasal tract in Fig. 6. The elements Z_{ii} ($i = 1, 2, \ldots, 6$), lying on the main diagonal of the matrix $\boldsymbol{Z}_N$, are, by analogy with matrix $\boldsymbol{Z}_{GU}$ (27), the self-impedances of the respective loops of the equivalent network in Fig. 6, whereas the elements Z_{ij} ($i, j = 1, 2, \ldots, 6$; $i \neq j$), beyond the main diagonal, are the mutual impedances of the adjacent loops, according to definitions (23) and (24). The physical meaning of the impedances z_{ani} of the longitudinal branches and the impedances z_{bni} of the transverse branches of the four-poles T_{ni} ($i = 1, 2, \ldots, 5$) is determined by the general formulae (9) and (10) derived earlier for the pharynx-mouth tract. The only difference is that in the case of the nasal tract the effect of energy dissipation due to vibration of the tract wall may be disregarded by putting $G_{sj} = C_{sj} = 0$ in equation (10).

The input impedance Z_n of the nasal tract, given by equation (32) for different values of the inlet area A_{ni} depending on the degree of nasalization, is a variable parameter occurring in equations (12), (13), (16), (17), (19) and subsequent ones. It is used for the computation of the transmission functions $H_u(\omega)$ (formula (29)) and $H'_u(\omega)$ (formula (30)) of the pharynx-mouth tract including nasalization.

6. The transmission function of the pharynx-nasal tract

As a result of the nasalization effect due to cleft palate, the sound wave is radiated through both the mouth and the nostrils. The contribution of the sound energy emitted through the nostrils to the total energy of the speech wave depends on the input impedance Z_n (formula (32)) of the nasal tract, and decreases to zero as $Z_n \to \infty$. In most pathological cases and in postoperative states this contribution is significant. In this connection, the application of the superposition theorem to acoustic pressure, $p(r) = p_u + p_n$, according to the general formula (1), must be based on a knowledge of the transmission function $H_n(\omega)$ of the pharynx-nasal tract ($G+N$), including the parallel impedance of the mouth-tract input impedance Z_u (see Fig. 1). The methodology and procedure of computation are analogous to those ones used in sections 3 and 4 for the pharynx-mouth tract.

The equivalent electrical circuit for the pharynx-nasal tract including the shunting effect of the mouth tract is shown in Fig. 7. The four-poles $T_1, T_2, \ldots, \ldots, T_{(k-1)}$, where $k = 8$[6]), represent, as in Fig. 3, the pharynx tract, whose cross-sectional area function $A(x)$ is determined — depending on the conditions of articulation — by the numerical values in the upper $(k-1)$ rows of Table 1. The four-pole T'_k simulates the bifurcation of the vocal tract, i.e.

[6]) Provided that the bifurcation of the vocal tract is located at a distance $x = 8$ cm from the glottis.

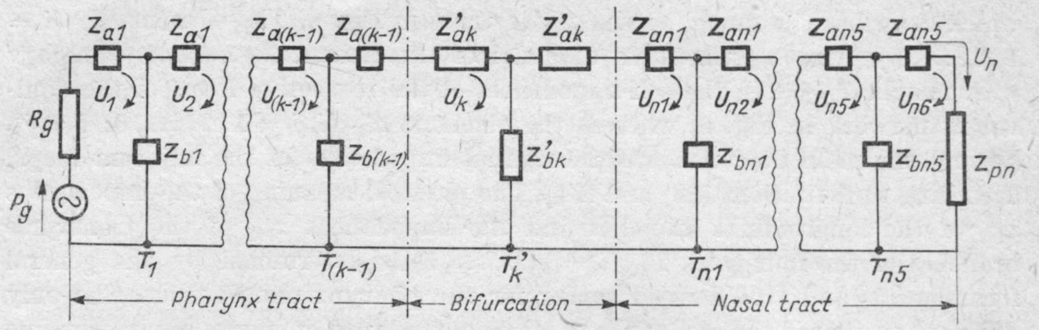

Fig. 7. Electrical equivalent circuit of the pharynx-nasal vocal tract including the shunting effect of the mouth tract

the place where the input impedance Z_u of the mouth tract shunts the pharynx-nasal tract. The configuration of the four-pole T_k' is identical with that of the four-pole T_k previously shown in Fig. 4, the only difference consisting in the shunting impedance whose role is now fulfilled by Z_u instead of Z_n. In full analogy to formulae (17) and (19), the longitudinal impedances z_{ak}' and the transverse impedance z_{bk}' of the four-pole T_k' may be expressed as

$$z_{ak}' = Z_0\left(\gamma\frac{l}{2}\right)\frac{Z_0+Z_u\left(\gamma\frac{l}{2}\right)}{Z_0+2Z_u\left(\gamma\frac{l}{2}\right)} \tag{34a}$$

and

$$z_{bk}' = \frac{Z_0}{\gamma l}\,\frac{Z_u}{Z_u+Z_0/\gamma l}. \tag{34b}$$

Disregarding the losses of the four-pole T_k', they may be written as

$$z_{ak}' = j\,\frac{\varrho\omega l}{2A_k}\,\frac{\varrho c/A_k+jZ_u\,\omega l/2c}{\varrho c/A_k+j\,Z_u\omega l/c} \tag{35a}$$

and

$$z_{bk}' = -j\,\frac{\varrho c^2}{A_k\,\omega l}\,\frac{Z_u}{Z_u-j\varrho c^2/A_k\omega l}, \tag{35b}$$

where Z_0 and γ denote, as before, the characteristic impedance (7) and the equivalent propagation constant (8) of the four-pole. Z_u is the input impedance of the mouth tract, measured or calculated at the input terminals of the four-pole $T_{(k+1)}$, provided the four-pole T_q is loaded with the radiation impedance Z_{pu} (formula (31)) of the mouth orifice (cf. Fig. 3). As for the input impedance Z_n of the nasal-tract (32), the impedance Z_u may be computed from the formula

$$Z_u = \frac{p_u}{U_{(k+1)}} = \frac{\Delta_u}{\Delta_{(k+1)(k+1)}}, \tag{36}$$

where Δ_u is the characteristic determinant of the loop-impedance matrix $\boldsymbol{Z}_u$ of the mouth tract, which is presented in Fig. 3 as a cascade connection of the four-poles $T_{(k+1)}, T_{(k+2)}, \ldots, T_q$ under definite articulation conditions, and $\Delta_{(k+1)(k+1)}$ is the cofactor of the element $Z_{(k+1)(k+1)}$ of this matrix.

Finally, the transmission function $H_n(\omega)$ of the pharynx-nasal tract (including the shunting effect of the mouth tract), defined as the ratio of the volume velocity U_n through the radiation impedance Z_{pn} of the nostrils to the volume velocity U_g through the glottis, is given by the expression

$$H_n(\omega) = \frac{U_n}{U_g} = \frac{\Delta_{1n}}{\Delta_{11}}, \tag{37}$$

where Δ_{11} and Δ_{1n} are the cofactors of the elements Z_{11} and Z_{1n} of the loop-impedance matrix $\boldsymbol{Z}_{GN}$ of order $n = k+6$, which corresponds to the Kirchhoff equations for $k+6$ independent loops of the network, shown in Fig. 7, which represents the pharynx-nasal tract. Similarly, the transmission function $H_n'(\omega)$ of the tract, defined as the ratio of the acoustic pressure $p_n = Z_{pn} U_n$ in the nose outlet orifice to the volume velocity U_g through the glottis, is equal to

$$H_n'(\omega) = \frac{p_n}{U_g} = \frac{Z_{pn} U_n}{U_g} = \frac{\Delta_{1n}}{\Delta_{11}} Z_{pn}. \tag{38}$$

7. Conclusions

The resulting transmission function $H(\omega)$ of the complex, i.e. pharynx-mouth-nasal vocal tract, defined as the ratio of the volume velocity or acoustic pressure at the measuring point on the axis of symmetry of the system, which radiates the sound wave through the mouth and nose orifices, to the volume velocity of the larynx generator, is given, in accordance with the superposition theorem (1), by the expressions

$$|H(\omega)| = |H_u(\omega)| + |H_n(\omega)| \tag{39}$$

or

$$|H'(\omega)| = |H_u'(\omega)| + |H_n'(\omega)|, \tag{40}$$

where $H_u(\omega)$, $H_u'(\omega)$, and $H_n(\omega)$, $H_n'(\omega)$ describe radiation from the mouth (29), (30), and the nose (37), (38), respectively.

The pole-zero distribution in the frequency domain of the model transmission function $|H(\omega)|$ describes the formant-antiformant structure of the particular speech sound, e.g. an oral vowel, whose phonetic features are uniquely determined by the cross-sectional area function $A(x)$ of the vocal tract, and the degree of nasalization Z_n. In this way, using the model and its mathematical description, one can compute the spectral characteristic of an arbitrary speech sound under definite articulatory conditions (which depends on the vocal-tract geometry), thus simulating the anatomy of the real (i.e. biological) system.

The simulative model of the vocal tract presented is both universal and versatile. It can simulate the articulatory conditions of arbitrary speech sounds with glottal excitation, taking into account any personal differences which result from either the individual anatomical features of the speech organ or the pathological anomalies, including the forced nasalization of oral vowels due to cleft palate.

The superiority of the simulative model, over analogue models, arises from its high accuracy in reproducing the real physio-pathological conditions of the biological system. This is a consequence of taking into account both the losses in the vocal tract and the radiation impedances of the mouth and nose orifices, which cannot be taken into consideration in analogue systems [12]. An additional advantage of the model is the variable parameter Z_n which corresponds to the nasal-tract input impedance, constituting a quantitative measure of the nasalization effect which depends on the extent of the cleft palate. It enables us to determine the quantitative relations between the spectral structure of the directly measured diagnostic signal (i.e. the acoustic pressure of the speech wave) and the physio-pathology of the vocal tract. These relations not only facilitate and make phoniatric diagnosis objective, but also are useful in the control and documentation of the rehabilitation process.

The clinical verification of the simulative model presented constitutes the subject of active research being conducted at the Department of Cybernetic Acoustics, Institue of Fundamental Technological Research, Polish Academy of Sciences, in close cooperation with the Otolaryngological Clinic of the Institute of Surgery of the Medical Academy in Warsaw. This will be preceded by the elaboration of the model in a form convenient to digital presentation and computer simulation, in order to obtain a rational compromise between the accuracy of reproducing the physio-pathological features of the biological system and the operative capabilities of moderately-sized computers currently available (e.g. the ODRA 1300).

References

[1] J. B. Dennis, *Computer control of an analog vocal tract*, Proc. Stockholm Speech Comm. Seminar, Royal Inst. Techn., Stockholm 1962.

[2] H. K. Dunn, *The calculation of the vowel resonances and an electrical vocal tract*, J. Acoust. Soc. Am., **22**, 740–753 (1950).

[3] G. Fant, *Acoustic theory of speech production*, s'-Gravenhage, Mouton and Co., 1960.

[4] J. L. Flanagan, *Some properties of the glottal sound source*, J. Speech and Hearing Res., **1**, 99–116 (1968).

[5] J. L. Flanagan, *Speech analysis, synthesis and perception*, Springer Verlag, 2nd Edition, Berlin-Heidelberg-New York 1972.

[6] M. H. L. Hecker, *Studies of nasal consonants with an articulatory speech synthesizer*, J. Acoust. Soc. Am., **34**, 179–188 (1962).

[7] A. S. House, *Analog studies of nasal consonants*, J. Speech and Hearing Disorders, **22**, 190–204 (1957).

[8] A. S. House, K. N. Stevens, *Analog studies of the nasalization of vowels*, J. Speech and Hearing Disorders, **21**, 218–232 (1956).

[9] K. Ishizaka, J. C. French, J. L. Flanagan, *Direct determination of vocal-tract wall impedance*, J. Acoust. Soc. Am., **55**, 879 (A) (1974).

[10] W. Jassem, *Fundamentals of acoustic phonetics*, Polish Scientific Publishers, PWN, Warsaw 1973 [in Polish].

[11] J. Kacprowski, *Theoretical bases of the synthesis of Polish vowels in resonance circuits*, Speech Analysis and Synthesis, Vol. 1, 219–287, Polish Scientific Publishers, PWN, Warsaw 1968.

[12] J. Kacprowski, W. Mikiel, A. Szewczyk, *Acoustical modelling of cleft palate*, Archives of Acoustics, **1**, 2, 137–158 (1976).

[13] J. Kacprowski, *Physical models of the larynx source*, Archives of Acoustics, **2**, 1, 47–70 (1977).

[14] J. Lagasse, *Étude des circuits électriques*, Editions Eyrolles, Paris 1962.

[15] M. Mrayati, B. Guerin, *Études des characteristiques acoustiques des voyelles françaises par simulation du conduit vocal avec pertes*, Revue d'Acoustique, **9**, 36, 18–32 (1976).

[16] G. Rosen, *Dynamic analog speech synthesizer*, J. Acoust. Soc. Am., **30**, 201–209 (1958).

[17] K. N. Stevens, S. Kasowski, G. Fant, *An electrical analog of the vocal tract*, J. Acoust. Soc. Am., **25**, 734–742 (1953).

Received on 14th May 1977