ARCHIVES OF ACOUSTICS
14, 1-2, 67-96 (1989)

DYNAMICAL MODEL OF THE VOCAL TRACT IN CONSONANT AND NASALIZED ARTICULATION

WANDA NOWAKOWSKA, PIOTR ŻARNECKI

Laboratory of Speech Acoustics
Institute of Fundamental Technological Research, Polish Academy of Sciences
(00-049 Warszawa, ul. Świętokrzyska 21)

In the paper a simulation model of the vocal organ is presented. The model has been programmed in Turbo Pascal for an IBM PC computer. With the use of this model as a research tool the possibilities of analysing the formant structure of the speech signal has been demonstrated in the static states for the cases of oral and nasalized vowels and nasal consonants. For analysed speech sounds the approximate vocal tract cross-sections were determined. An example of simulation of continuous changes of the articulatory tract geometry which take place in natural speech has been presented and the accompanying changes of the phonetical and acoustical structure of the signal has been described. Also potential possibilities of applying the described model to the studies on the phenomenoús of coarticulation and articulation in the conditions of forced nasalization has been demonstrated.

W artykule przedstawiono symulacyjny model narządu mowy, zaprogramowany w języku Turbo Pascal na komputerze IBM PC. Za pomocą opracowanego modelu jako narzędzia badawczego, pokazano możliwości analizowania struktury formantowej sygnału mowy w stanach statycznych w przypadku samogłosek ustnych, samogłosek nazalizowanych oraz spółgłosek nosowych. Dla badanych głosek wyznaczono przybliżone przekroje artykulacyjne toru głosowego. Przedstawiono przykład symulowania zmieniającej się w sposób ciągły geometrii toru artykulacyjnego, zachodzącej w mowie naturalnej oraz podano wyniki obserwacji towarzyszących temu zjawisku zmian fonetyczno-akustycznej struktury sygnału. Wykazano także potencjalne możliwości zastosowania opracowanego modelu do badania zjawiska koartykulacji oraz procesu artykulacji w warunkach nazalizacji wymuszonej.

1. Object and aim of the study

For a number of years the investigations of acoustical and phonetic features of a signal have been carried out on the basis of modelling the human articulatory organ. Modelling of the biological vocal tract, pharyngeal-oral-nasal in general case, consists in representing as a physical system its anatomical structure, and mainly its geometrical layout which changes in course of articulation of various speech sounds.

In such, it is also possible, to some extent, to take into account the individual features of the speech organ. In the case of voiced speech sounds the vocal tract model is excited by the simulated quasi-periodical laryngeal source, whereas in the case of other phones, e.g. fricatives and affricates, it is necessary to take into account the noise excitation or both kinds of excitation together. The vocal tract configuration is represented with linear equivalent systems and the nonlinear acoustic quantities which refer mainly to the source characteristics and acoustic properties of the resonance systems formed by the vocal channel are described by the linear equations system according to the generally accepted methodology (VAN DEN BERG [1], FANT [11]).

The computer simulation model of the pharyngeal-nasal-oral vocal tract which is now being worked upon turns back to the earlier works (NOWAKOWSKA [31], NOWAKOWSKA, ŻARNECKI [32, 33]). The model makes it possible to study the phonetical and acoustical features of the speech signal, in particular, the spectral structure of the oral and nasalized vowels and the nasal consonants. This is accomplished by an analysis of the transmittance of the modelled vocal tract system. The software allows for dynamical modelling of the articulatory structures which change in time, as e.g. nasal consonant – vowel syllables. It also makes it possible to model the co-articulation phenomena. These possibilities can occur useful in the automatic system for recognizing isolated words which is currently being elaborated in the Laboratory of Speech Acoustics, IFTR PAS (WIĘŹLAK [39]). The significant advantage of the present model is the possibility of investigating the articulation process in the conditions of forced nasalization that takes place in the case of cleft soft palate, which is the relatively frequent development defect of the speech organ.

2. Development of the analog modelling methods of the vocal tract

The discrete models of the vocal tract proposed in literature have taken diverse forms. The earliest model was the classical Helmholtz resonator which, being an uniresonance system, reproduced the spectral structure of a vowel only within the range of the first formant, what was not sufficient for its phonetical identification (CRANDALL [6]). The two-resonance model consisting of two coupled Helmholtz resonators (FANT [11], KACPROWSKI [20]) also appeared to be insufficient, mainly because of technical difficulties concerned with retuning systems, necessary for precise choice of such acoustical parameters as volume of the resonant cavity and dimensions of the neck, i.e. the acoustic compliance and the acoustic inertance, which describe the resonance frequencies of the system. Therefore, the acoustic models of the form of chained sectors of homogeneous tubes appeared to be far more convenient and susceptible of more precise modelling of the geometrical configuration of the vocal tract in the conditions of articulation that approached reality (DUNN [9], STEVENS and HOUSE [35], FANT [11], COKER [5]). The intense development of acoustic methods of modelling of the speech organ which emerged in the 50-ies and 60-ies was directed at improving the modelling accuracy of the articulatory tract, what corresponded to an amelioration of acoustical structure of the speech signal and to the elevation of the limit frequency of applicability of the models. This was achieved by increasing the number of elementary cylindrical segments into which the vocal tract has been spatially partitioned by convention. Together with the increase of the number of segments the analog model of the vocal tract becomes an acoustic transmission system with unit parameters continuously distributed in space, the electrical analogue of which is a long line. The first analog voice tract worked out on this basis was a system formed by connecting 25 *T*-type four-terminal networks, each of them representing a cylindrical tract segment of length $l = 0.5$ cm (DUNN [9]). The authors of subsequent gradually improved and now already classical versions of analog configuration models, consisting usually of 17–30 segments, were STEVENS, KASSOWSKI, FANT [35] and FANT [11], STEVENS and HOUSE [36] have developed the articulatory model of vocal tract by adding a simple model of laryngeal source and VAN DEN BERG [1] has worked out an analog subglottal system which co-operated with the channel model. For the first time the numerical control of an analog configuration model was applied by DENNIS [7].

Presently, regarding the development of applications of computer systems, the analog vocal tract model have been replaced by numerical models realized by computer simulation. The essence of the computer simulation process consists in building an algorithm of the mathematical description of a physical structure of a chosen analog model and in finding the values of selected phonetical-acoustic parameters of the speech signal. According to the accessible data the first simulation model of a vocal tract was worked out by KELLY and LOCHBAUM [25]. Since then a number of simulation models have been worked out in diverse scientific centres in the world, mainly in France, Sweden, Japan and USA, the representatives of which are, among others, MRAYATI and GUERIN [29], FLANAGAN, ISCHIZAKA and SHIPLEY [13], MAEDA [27], WAKITA and FANT [37].

The present work is the first attempt in Poland to put forward a simulation model of the vocal tract and to carry out its verification with the phonetical data of Polish language, with reference to the previously developed analog articulatory model (KACPROWSKI [23, 24]).

3. Vocal tract model with variable geometry

The speech organ is anatomically complex in respect of the structure as well as the function. Nevertheless, in order to illustrate the way of modelling the speech signal the anatomical description of this biological system can be greatly simplified by limiting the considerations only to these segments of the articulatory system which are significant from the point of view of their voice generation and articulation functions.

The vocal tract consists of three cavities: pharyngeal, oral and nasal, and their contribution in shaping the output signal depends on their current configuration; in particular it depends on:

a) the position of the soft palate (Fig. 1),

b) the position and the type of the exciting source (quasi-periodical and/or noise-type).

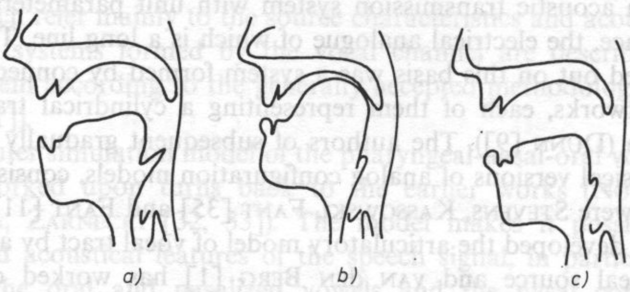


FIG. 1. Scheme of articulation with nasal cavity disengaged (a), with oral cavity disengaged (b) and scheme of oral-nasal articulation (c)

The soft palate can adhere to the back wall of the pharyngeal cavity, thus closing the air passage to the nasal cavity. In this case the air flow coming from the lungs to the larynx passes through the pharyngeal and oral cavities, subsequently (Fig. 1 a). If the soft palate forms the velar closure with the back of the tongue then the oral cavity is blocked off and the air flows through the pharyngeal-nasal cavity and comes out through the nasal appertures (Fig. 1 b). The intermediate position of the soft palate, in which it does not form the closure neither with the back wall of the pharynx nor with the back of the tongue, makes it possible for the air flow to enter both cavities, the oral and the nasal one (Fig. 1 c). In this case air can come out through the mouth and through the nostrils or, if there is an occlusion in some place in the oral cavity, only through the nostrils, as it is in the case of articulation of nasal consonants. Considerable part of the surface of pharyngeal-oral tract walls is formed by the moving articulators of the speech organ: tongue, soft palate, lips and the jaw. Hence, the geometrical configuration, i.e. the shape, length and the volume of the pharyngeal-oral tract is subject to considerable variations during the articulation process; on the contrary, the structure and the geometry of the nasal canal is constant with the exception of the inlet part formed between the uvula and the back wall of the nasal cavity. Shape of the nasal canal does not depend on the articulation conditions and exhibits only individual variations. Therefore, large simplifications can be done considering the shape function of the cross-sectional area of this tract by treating the parallel segments of the nasal canal as one tube of the cross-section equal to the sum of their cross-sections and by approximating the irregular shapes of the cross-sections by an appropriate circular shape. Similarly, the pharyngeal-oral tract has been represented as a tube with the stepwise variable circular cross-section, constant within the range of subsequent segments. Such an approach is generally known in the literature. The usefulness and rightness of application of the circular approximation, and not, for example an ellyptical one, has been justified by WAKITA and FANT [37], who demonstrated that application of an ellyptical cross-section for the vocal tract yields the comparable results to those obtained with the circular approximation, but considerably increases the complexity of calculations what in consequence prolongs the calculations. In effect these considerations settled the question of the widespread use of models with circular cross-sections.

The basic problem in modelling the voice generating and articulating functions of the speech organ is finding the transmittance function $T(f)$ of the vocal tract, i.e. the ratio of the volume velocities or acoustic pressures in the mouth opening and/or nasal openings and in the glottal opening. The widely applied principle of analog modelling of the vocal tract consists in representing the system as a tube with stepwise variable cross-sectional area, partitioned into a finite number of elementary segments. Every elementary segment of a cylindrical tube of the length l and cross-section A is treated as a lumped-parameter system in a form of an equivalent electrical system, e.g. a four-terminal network. This four-terminal network is described by unit parameters L, C, R and G which are equivalent to the unit parameters of an acoustic tube, i.e. the acoustic inertance, acoustic compliance, friction loss resistance and heat loss conductance. Acoustic impedance of a laryngeal source Z_l (J. KACPROWSKI [22]) and acoustic radiation impedance Z_0 of this openings of mouth and nostrils are replaced by equivalent two-terminal networks (J. KACPROWSKI [20]). In the presented model the pharyngeal-oral tract has been described as a chain-type connection of seventeen ($n = 17$) elementary segments (cylindrical tubes), each of them representing a 1 cm long, tract segment. Assumption of this length of an elementary segment results from the relation of geometrical dimensions of a tract segment and length of the wave which propagates along the tract in the considered frequency range. In the Fig. 2 an elementary segment in the form of a cylindrical tube of length l (a), and it's equivalent four-terminal networks (b), (c) have been presented (KACPROWSKI [22], [24]). Similarly, the pharyngeal-nasal tract has been represented as a chain of twelve ($n = 12$) elementary segments. In both cases, the last segment of the vocal tract is loaded with the radiation impedance of a piston placed in an infinitely large flat baffle board. The radius of the nasal outlet opening is constant and equal to the radius r of the last segment, and on the basis of anatomical data it has been assumed $r = 0.27$ cm.

The equivalent system for a uniform cylindrical tube segment of the length l and cross-sectional area A for both channels is in the considered case a symmetrical T-type four-terminal network the acoustic parameters of which are described by the formulae

$$L = \frac{\varrho l}{A} \tag{1}$$

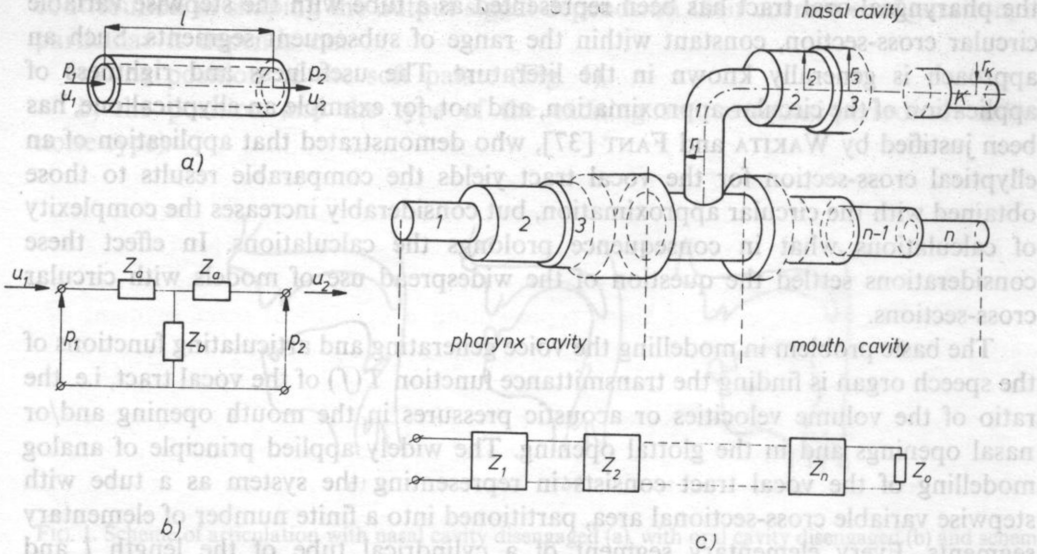


FIG. 2. Cylindrical tube of length l and cross-section A which represents an elementary vocal tract segment (a), its four-terminal network equivalent elements (P — acoustic pressure, U — volume velocity) (b, c)

acoustic inertance of the air in the tube $[\mathrm{kg \cdot m^{-4}}]$,

$$C = \frac{A \cdot l}{\varrho c^2} \tag{2}$$

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

$$R = \frac{l \cdot S \sqrt{\varrho \mu \omega}}{\sqrt{2 A^2}} \tag{3}$$

acoustic resistance of loss caused by viscous friction of the air near tube walls $[\mathrm{kg \cdot m^{-4} \cdot s^{-1}}]$,

$$G = \frac{(\eta - 1) l S}{\varrho c^2} \sqrt{\frac{\lambda \omega}{2 \xi \varrho}} \tag{4}$$

acoustic conductance of loss due to heat conductivity near the tube walls $[\mathrm{kg^{-1} \cdot m^4 \cdot s}]$.

The letter symbols used in Eqs. (1)–(4) and applied in the further part of this paper have the following physical meaning and numerical values:

$\varrho = 1.14$ $[\mathrm{kg \cdot m^{-3}}]$ — air density,

$\mu = 1.86 \cdot 10^{-5}$ $[\mathrm{N \cdot s \cdot m^{-2}}]$ — air viscosity coefficient in the temperature 37°C,

$\eta = 1.4$ — adiabatic constant of the air,

$\lambda = 2.3 \cdot 10^{-2}$ $[\mathrm{J \cdot deg^{-1} \cdot m^{-1} \cdot s^{-1}}]$ — thermal conductivity of the air,

$\xi = 10^3$ $[\mathrm{J \cdot deg^{-1} \cdot kg^{-1}}]$ — specific heat of the air.

Furthermore:
l – length of a single segment [m]
A – area of the tube opening [m^2]
S – parameter of the tube opening [m]
$\omega = 2\pi F$ – pulsation, angular frequency [s^{-1}]

The walls of the biological vocal tract formed of the connective tissue are not ideally rigid and smooth, as it is assumed in simplified considerations, but exhibit some loss and vibrational properties expressed by the surface impedance z_s having an initial character. In the frequency range $f > 100$ Hz this impedance can be represented as a series connection of the resistance r_s and mass m_s

$$z_s = r_s + j\omega m_s, \tag{5}$$

where r_s[kg·m^{-2}·s^{-1}] and m_s[kg·m^{-2}] denote the mechanical loss resistance and the co-vibrating mass of wall material of the vocal tract per unit area, respectively. In the literature of the subject a lot of attention has been paid to the problem of finding the numerical values of the unit parameters r_s and m_s. The results of direct measurements carried out "in vivo" and "in vitro" on the connective tissue of human body have been taken as a starting point (cf. e.g. Flanagan, Ishizaka, Shigley [13], Nord, Fant, Branberud [30]) or the vibrational properties of the appropriately damped physical resonance systems have been studied (van den Berg [1], Dunn [10], Fujimura and Lindqvist, [14]). The mean values obtained in the measurements and verified in model tests are $r_s \approx 16\cdot 10^3$ [kg·m^{-2}·s^{-1}] and $m_s \approx 15$ [kg·m^{-2}]. These values have been introduced into the calculations in the present work. Because of numerical reasons and the reasons concerning the design of the scheme, it is convenient to place the impedance z_s (5) in the transversal branch of the T-type four-terminal network (Fig. 2 b), in the form of the equivalent loss admittance of the vocal tract walls related to the surface S of an elementary segment

$$Y_s = \frac{r_s S l}{r_s^2 + \omega^2 m_s^2} - j\omega \frac{m_s S l}{r_s^2 + \omega^2 m_s^2} = G_s + j\omega C_s, \tag{6}$$

where

$$G_s = \frac{r_s S l}{r_s^2 + \omega^2 m_s^2} \quad [\mathrm{kg}^{-1}\cdot\mathrm{m}^4\cdot\mathrm{s}] \tag{7}$$

is the acoustic loss conductance in the voice tract walls, and

$$C_s = \frac{m_s S l}{r_s^2 + \omega^2 m_s^2} \quad [\mathrm{kg}^{-1}\cdot\mathrm{m}^4\cdot\mathrm{s}^2] \tag{8}$$

is the negative acoustic compliance, equivalent to the co-vibrating acoustic mass of the vocal tract walls.

Applying the above formulae, the impedance Z_a of the longitudinal branches of the four-terminal network (according to notations in Fig. 2 b and c) and the impedance Z_b of the transversal branch have been found as

$$Z_a = \frac{1}{2}(R + j\omega L), \tag{9}$$

$$Z_b = [(G_p + G_s) + j\omega(C_p + C_s)]^{-1}. \tag{10}$$

The analog model of the vocal tract is the system formed as a chain on n such networks (Fig. 2d), each of them representing a tract segment of length l. Every network is described by a chain matrix $\mathscr{A}_i$, $i = 1, 2, \ldots, n$ of the form

$$\mathscr{A}_i = \begin{bmatrix} 1+\dfrac{Z_a}{Z_b}; & 2Z_a + \dfrac{Z_a^2}{Z_b} \\ \dfrac{1}{Z_b}; & 1+\dfrac{Z_a}{Z_b} \end{bmatrix}. \tag{11}$$

The whole system is described by the matrix $\mathbf{A}$ which is the product of n such matrices A_i (Fig. 2e)

$$\mathbf{A} = \prod_{i=1}^{n} \mathscr{A}_i = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}. \tag{12}$$

The relations between the acoustic input values P_1, U_1 and output values P_2, U_2 of the four-terminal-network model of the vocal tract can be calculated from the matrix equation

$$\begin{bmatrix} P_1 \\ U_1 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} Z_0 \\ 1 \end{bmatrix} U_2, \tag{13}$$

where A_{21} and A_{22} are the elements of the chain matrix $\mathbf{A}$ (12), and opening, U_2 volume velocity in the nasal or mouth opening, Z_0 radiation impedance of the mouth or nose, P_2 acoustic pressure in the mouth or nasal opening.

The transmittance function $T(F)$ calculated for the whole tract is expressed by the relation

$$T = \frac{U_2}{U_1} = \frac{1}{A_{22} + A_{21} Z_0}, \tag{14}$$

where A_{21} and A_{22} are the elements of the chain matrix $\mathbf{A}$ (12), and

$$P_2 = U_2(f) \cdot Z_0(f) = U_1(f) \cdot T(f) \cdot Z_0(f) \tag{15}$$

where Z_0 acoustic pressure in the mouth opening.

The values of acoustic pressure have been found in relative units by referencing them to the pressure $P_2(F_0)$ where F_0 is the laryngeal tone frequency.

Hence

$$P(F) = \frac{P_2(f)}{P_2(F_0)} = \frac{U_1(f)}{U_1(F_0)} \frac{Z_0(f)}{Z_0(F_0)} \frac{T(f)}{T(F_0)}. \tag{16}$$

4. Implementation of the vocal tract simulation model on an IBM PC/XT computer

The model of the speech organ has been programmed in the Turbo Pascal language and implemented on an IBM PC/XT computer. This simulation vocal tract model serves to find the following data:

1. The frequency spectrum of acoustic pressure at the nasal or mouth openings or their sum, according to the Eq. (16), for a given geometrical configuration of the pharyngeal-oral and pharyngeal-nasal tract;
2. The instantaneous values of acoustic signal with the use of the inverse Fourier transform, which serve as input data to a *D/A* converter and an acoustic amplifier which generates the speech signal.

The earlier works of the authors (NOWAKOWSKA [31, 33]; NOWAKOWSKA, ŻARNECKI [32]) which concerned the development and realization of the vocal tract model allowed for finding an acoustic signal in the mouth and nasal openings for a given static articulatory configuration and for calculating the frequencies of the $F1$, $F2$, $F3$ formants as well as the formant's bands widths $B1$, $B2$, and $B3$. Utilizing the results of simulations and the results of former experiments, a vocal tract model has been constructed which makes it possible to simulate the time-variable articulatory structures. The transition from a static model to a dynamical one necessitates time description of configuration changes of the vocal tract. The structure of the simulation model of the vocal tract and the organization scheme of the program which realizes the model have been shown in the Figs. 3 and 4. The input data for the vocal tract simulation are the labels of the modelled speech sounds. The articulatory description applied here contains the information on the configuration of articulatory organs during fonation, i.e. the articulation tract geometry which corresponds to the given speech sound. Thus, the geometrical description resolves itself to the specification of the lengths l [cm] and the radii of the cross-sections r [cm] of every segment. These data are achieved and can be changed during an experiment in conversation mode by the operator. In the case of dynamical modelling, the signal segment, for one laryngeal tone period, which is stored in the computer memory, corresponds to every stationary part of the sound. Connecting these parts additionally necessitates for modelling the intermediate states resulting from the smooth change of the articulatory system. The smooth changes of the instantaneous values of the acoustic signal during transition from a stationary segment to the next one are obtained by appropriately changing the shape of the amplitude envelope modifying function in every segment, what is done after every period of the laryngeal tone. Up till now the shape of the envelope modifying function has been chosen in an informal way. The calculated frequency characteristic with the basic frequency and the envelope modifying function taken into account and with the use of the inverse Fourier transform allows for calculating a series of instantaneous values of the acoustic speech signal at the mouth opening, nasal openings or both at the same time. The series of instantaneous values are displayed on the printout as a plot. The

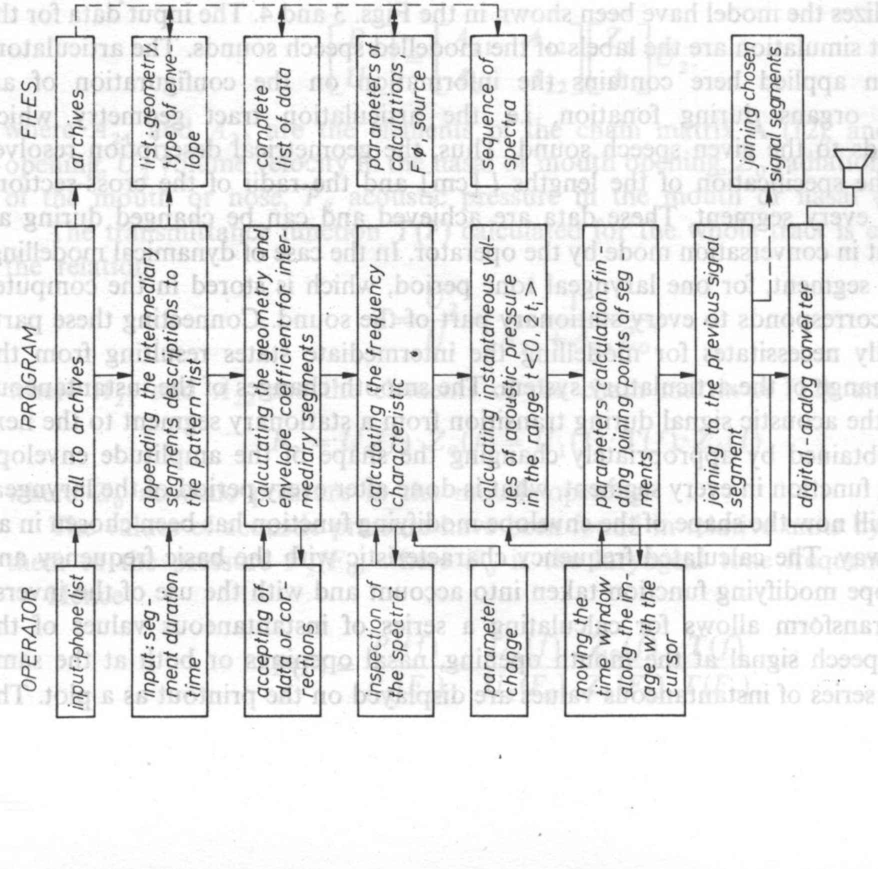


FIG. 3. Operation principle of the program realizing the vocal tract model. Broken lines represent the links which can be used in the case of utilizing the intermediate results stored in memory

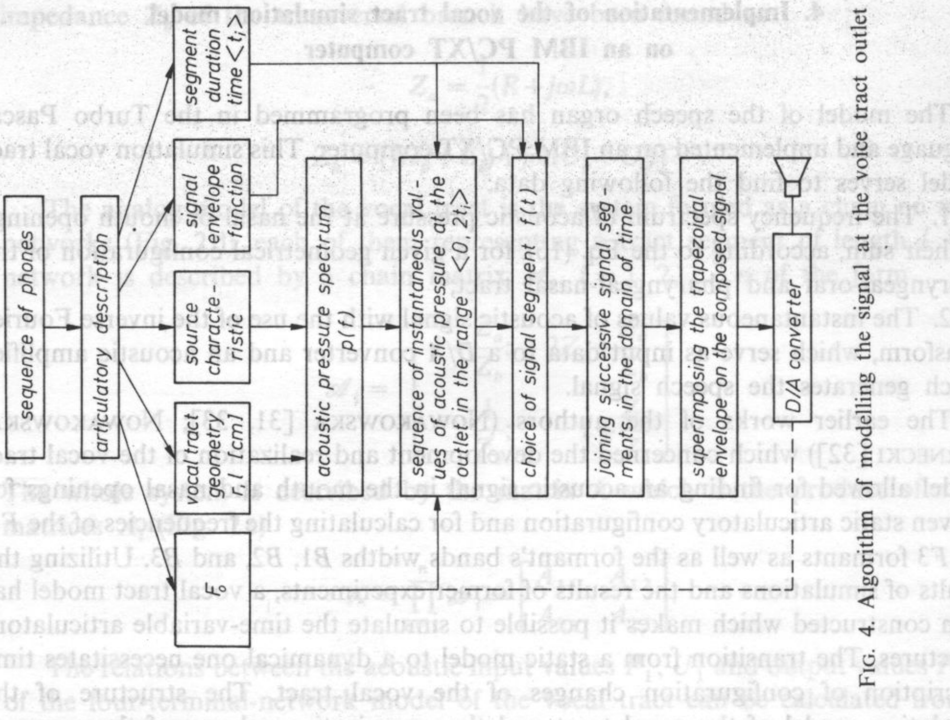


FIG. 4. Algorithm of modelling the signal at the voice tract outlet

calculated series of these values serves as a basis for choice of a time segment to be joined with the next segment. The part of a series is chosen which allows for a smooth signal change on the boundary between the segments when the current segment is chained with the preceding one in a specified time interval. The subsequent stage of signal processing needs the trapezoidal envelope to be imposed, to eliminate the impulse components in the monitored signal. Finally, the obtained signal is sent to an acoustic amplifier through the *D*/*A* converter. Calculations of the acoustic pressure complex form, at the mouth and nasal openings are carried out according to the algorithm showed in the Fig. 5 for subsequent vocal tract excitation

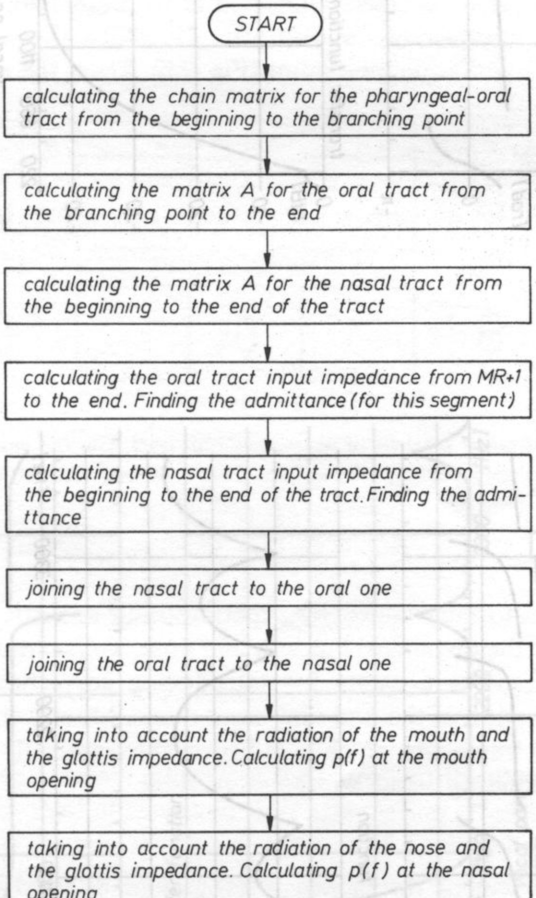
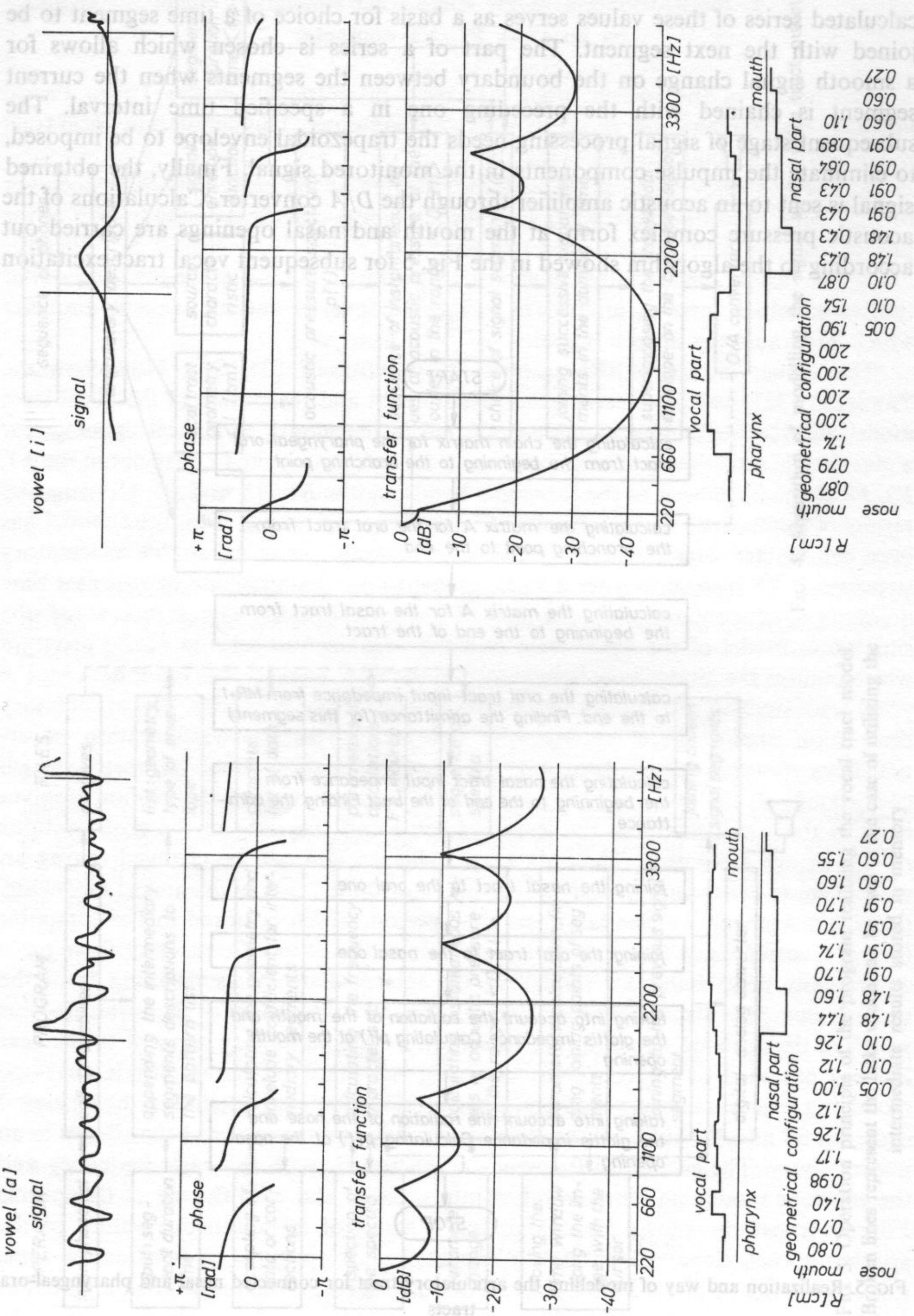


FIG. 5. Realization and way of modelling the articulatory tract for connected nasal and pharyngeal-oral tracts


vowel [i]
signal
phase
+π
[rad]
0
-π
transfer function
0
[dB]
-10
-20
-30
-40
220
660
1100
2200
3300 [Hz]
vocal part
nasal part
mouth
pharynx
nose
R [cm]
geometrical configuration
0.87 0.79 1.74 2.00 2.00 2.00 2.00 1.90 1.54 0.87 0.43 0.43 0.43 0.43 0.64 0.89 1.10 0.27
0.05 0.10 0.10 1.48 1.48 0.91 0.91 0.91 0.91 0.60 0.60
vowel [a]
signal
phase
+π
[rad]
0
-π
transfer function
0
[dB]
-10
-20
-30
-40
220
660
1100
2200
3300
[Hz]
vocal part
nasal part
mouth
pharynx
nose
R[cm]
geometrical configuration
0.80 0.70 1.40 0.98 1.17 1.26 1.12 1.00 1.12 1.26 1.44 1.60 1.70 1.74 1.70 1.70 1.60 1.55
0.05 0.10 0.10 1.48 1.48 0.91 0.91 0.91 0.91 0.60 0.60 0.27
]

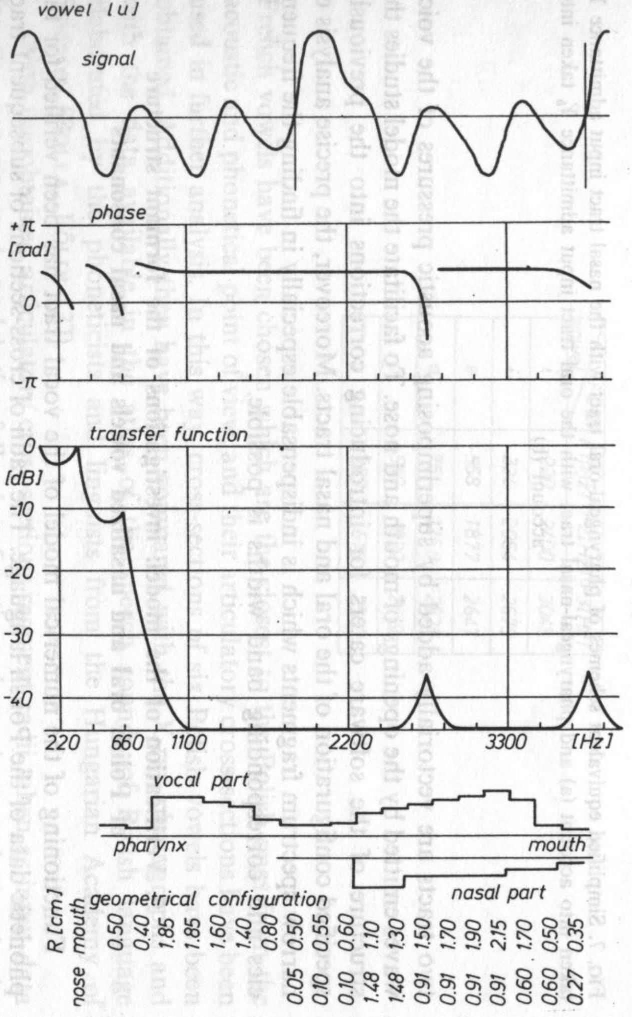


FIG. 6. Frequency characteristic corresponding to the vowels [a, i, u]

frequencies in the range between the initial frequency F_i and the final one F_f, with the specified step F_d. The results are normalized with respect to the acoustic pressure calculated for the reference frequency F_0. The spectrum is updated following the calculation of every current frequency, what gives the possibility of analysing the results before the simulation is finished and allows for corrections. In the printouts the values of the acoustic pressure modulus and phase at the mouth opening are separately plotted. The example printouts with the data concerning the [a], [i] and [u] vowels are presented in Figs. 6 a, b, c.

Calculation of the spectrum of an acoustic signal in the case of two tracts being active is carried out as follows. First the acoustic pressure is calculated at the output of the tracts in two cases: a) with the input admittance of the nasal tract Y_n, which shunts the pharyngeal-oral tract, connected in the branching point (Fig. 7 a), and b) with the input admittance of the oral tract Y_b, which shunts the pharyngeal-nasal tract, connected in the branching point (Fig. 7 b). Then, the output signals from the two tracts are vectorially added by superimposing acoustic pressures of the voice waves emitted by the openings of mouth and nose. To facilitate the model studies the structure of the software caters for introducing corrections into the previously specified configuration of the oral and nasal tracts. Moreover, the precise analysis of narrow spectrum fragments which is indispensable especially in finding the frequencies and corresponding band widths is possible.

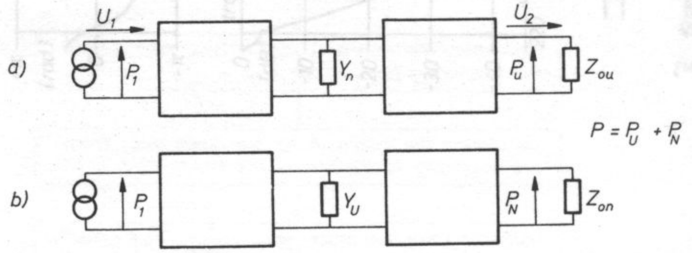


FIG. 7. Simplified equivalent schemes of pharyngeal-oral tract with the nasal tract input admittance Y_n taken into account (a) and pharyngeal-nasal tract with the oral tract input admittance Y_b taken into account (b)

5. Verification of the model: investigations of the formant structure of Polish oral and nasalized vowels and nasal consonants

Functioning of the numerical model of the vocal tract has been verified for the phonetic data of the Polish language. The radii of cross-sections of subsequent tract segments were the input data for the program. For Polish vowels the detailed descriptions of articulatory cross-sections of the pharyngeal-nasal tract during fonation do not exist. Those which can be found in various phonetic and phonological works can provide only approximate informations because the articulatory system is considered there as a static object. Because there was no possibility of taking tomographic images of vocal tract during the pronouncement of Polish phones it has been desired to use the indirect method of finding the articulatory cross-sections which has been applied (among others) by MRAYATI [28]. As a starting point the articulatory cross-sections of a pharyngeal-oral tract given in the cited paper have been taken. Comparative analysis of Polish and French vowels in respect of their place of articulation, described by their mutual position in the vowel diagram and their phonetical description based on the frequencies of first three formants *F*1, *F*2, *F*3, has been carried out (Tables 1 and 2). These among the eleven French vowels have been chosen which are the closest to the Polish ones from the acoustic and phonetical point of view, and their articulatory cross-sections have been used in further analysis. In this way, cross-sections of six Polish vowels have been obtained. Additionally they have been compared with the X-ray photographs and with the data available in the atlas of the speech sounds of Polish language elaborated by the phoneticists and linguists from the Hungarian Academy of Sciences (BOLLA, FÖLDI [3]).

Table 1. Formant frequencies for French vowels (MRAYATI [28]).

[Hz]	u	o	ɔ	ɑ	a	ɛ	e	i	Y	ø	œ
F1	240	350	510	650	725	510	350	240	240	350	510
F2	750	865	1000	1200	1300	1950	2200	2500	1850	1600	1400
F3	2200	2450	2250	2200	2640	2300	2500	3140	2110	2500	2300

Table 2. Formant frequencies for Polish vowels (JASSEM [19])

F [Hz] vowel	F1[Hz]	F2[Hz]	F3[Hz]
i	220	2400	3045
ɨ	324	2064	2679
e	568	1877	2647
ɑ	751	1268	2526
o	540	922	2619
u	285	675	2580

With the obtained articulatory cross-sections segment radii taken as input data for the programmed model the following functions have been calculated: the transmittance functions, frequencies of the formants $F1$, $F2$, and $F3$ and the corresponding formant band widths $B1$, $B2$, and $B3$ of the considered vowels. In the Table 3 the formant frequencies obtained from the model have been compared with the results obtained by other authors.

Table 3. Formant frequencies and formant band widths obtained with the model for Polish vowels (column A), the same data according to Jassem [19] (column B) and according to Wakita and Fant [37] (column C).

vowel	F1 [Hz]			B1 [Hz]	
	A	B	C	A	C
ɜ	549		529	15	20
i	255	220	289	23	36
ɨ	320	324		21	
e	441	568	462	16	20
a	647	750	739	26	26
o	600	540	575	20	29
u	291	285	290	24	37

vowel	F2 [Hz]			B2 [Hz]	
	A	B	C	A	C
ɜ	1518		1475	30	41
i	2420	2400	2354	30	29
ɨ	2121	2064		41	
e	1779	1877	2011	30	65
a	1280	1268	1142	40	40
o	1070	922	894	32	30
u	621	675	636	17	24

vowel	F3 [Hz]			B3 [Hz]	
	A	B	C	A	C
ɜ	2490		2443	80	85
i	2890	3045	2888	210	266
ɨ	2530	2679		210	
e	2650	2647	2777	95	234
a	2510	2526	2491	80	64
o	2530	2619	2428	60	32
u	2720	2580	2441	20	26

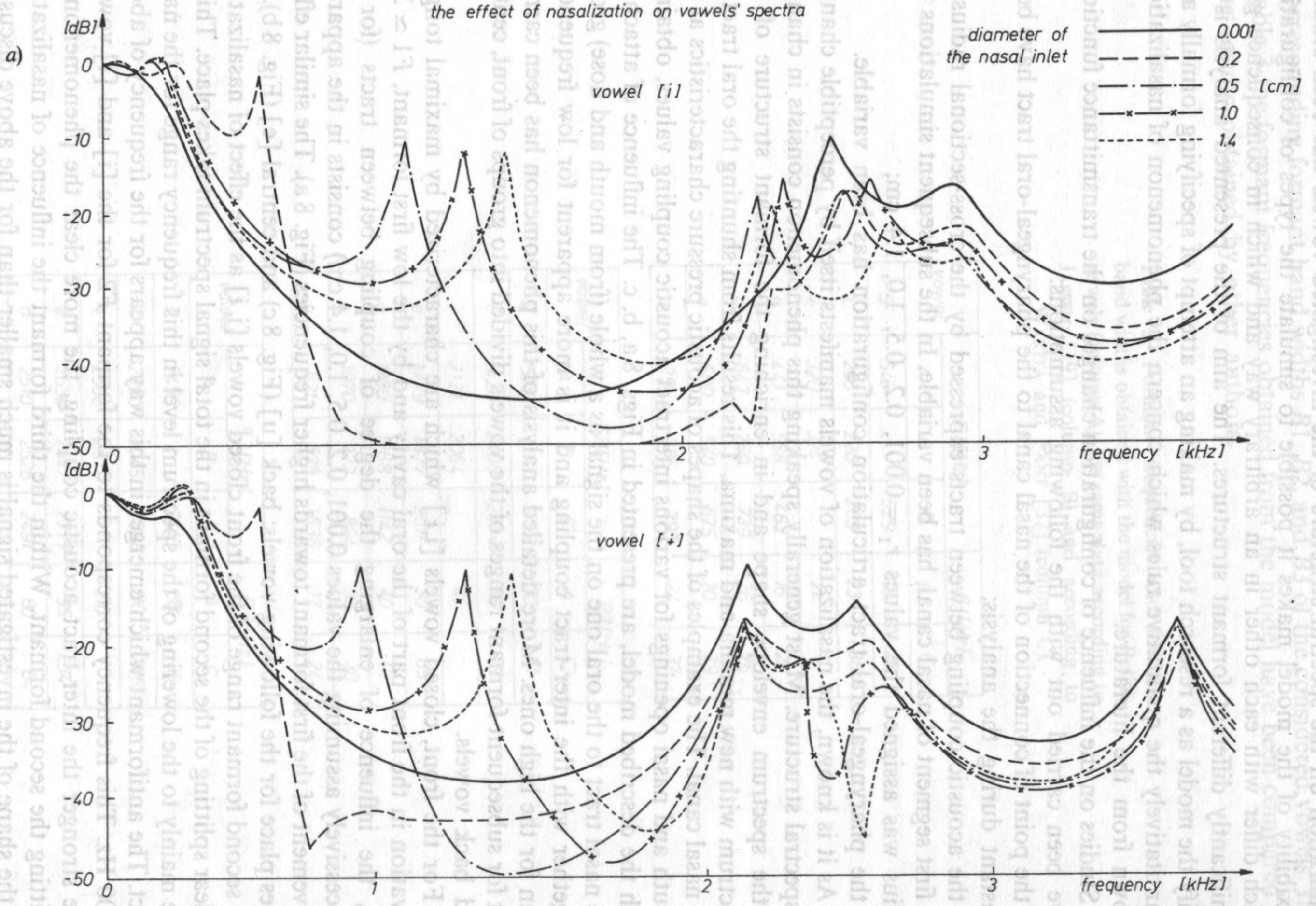
With the use of the developed model the influence of the phenomenon of nasalization on the spectral structure of the investigated vowels has been studied. Flexibility of the model makes it possible to simulate the types of configuration which differ with each other in an arbitrary way and which in consequence give significantly different formant structures. The aim of the presented analysis is to verify the model as a research tool, by making an attempt of specifying formally and quantitatively the qualitative rules which concern the phenomenon of nasalization, known from the literature.

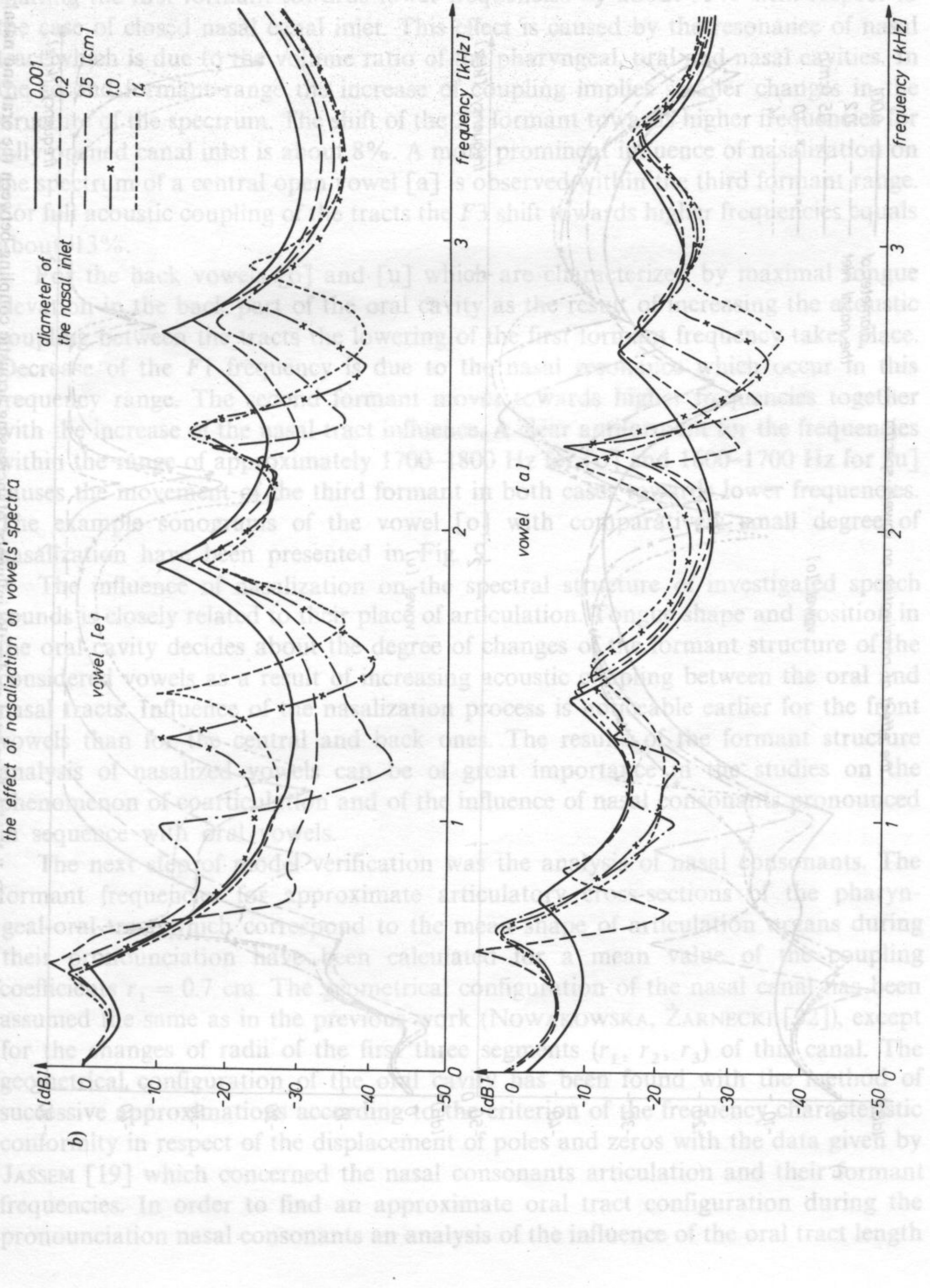
Studies on the influence of configuration changes on the transmittance functions have been carried out with the following assumptions:

- the point of connection of the nasal canal to the pharyngeal-oral tract have been constant during the analysis;
- the acoustic coupling between tracts expressed by the cross-sectional radius of the first segment of nasal canal, has been variable. In the subsequent simulations the radius was assigned the values $r_1 = 0.001$, 0.2, 0.5, 1.0, 1.4 cm;
- the pharyngeal-oral tract articulation configuration has been variable.

As it is known, the nasalization of vowels manifests itself by perceptible changes in spectral structure. Most generally speaking this phenomenon consists in changes of the spectrum envelope shape and in enriching the formant structure of the spectrum with new minima and maxima. This results from shunting the oral tract by the nasal canal. The examples of the shapes of acoustic pressure characteristics at the mouth and nasal openings for various intertract acoustic coupling values, obtained with the described model, are presented in Figs. 8 a, b, c. The influence of attaching the nasal tract to the oral one on the signal as a whole (from mouth and nose) grows together with the inter-tract coupling and it is more apparent for low frequencies than for the high ones. More detailed analysis of this phenomenon has been carried out for subsequent-formant ranges of the vowels divided into groups of front, central and back vowels.

For the front, closed vowels [i,ɨ] which are characterized by maximal tongue elevation in the front part of the oral cavity and by the low first formant, $F1 \simeq 300$ Hz, the influence of enlarging the degree of coupling between tracts (for r_1 successively assuming the values 0.001, 0.2, 0.5, 1.0, 1.4 cm) consists in the apparent movement of the first formant towards higher frequencies (Fig. 8 a). The similar effect takes place for the following vowels: back [u] (Fig. 8 c) and central [e] (Fig. 8 b). In the second formant range of the front closed vowels [i, ɨ] as the effect of nasalization a clear splitting of the second formant in the total signal spectrum takes place. This is due mainly to the lowering of the spectrum level in this frequency range for the nasal tract. The antiformant which emerges in this way appears for the frequency of about 2500 Hz. This frequency corresponds to the formant $F2$ for the [i] and [ɨ] vowels. The stronger the inter-tract acoustic coupling, the more clear the phenomenon of splitting the second formant. Within the third formant the influence of nasalization on the shape of the investigated signal is much smaller than for the above discussed formant ranges.


a)
the effect of nasalization on vawels' spectra
diameter of the nasal inlet
0.001
0.2
0.5 [cm]
1.0
1.4
vowel [i]
vowel [ɨ]
[dB]
0
-10
-20
-30
-40
-50
0
1
2
3
frequency [kHz]


b)
the effect of nasalization on vowels' spectra
diameter of the nasal inlet
0.001
0.2
0.5 [cm]
1.0
1.4
vowel [e]
vowel [a]
[dB]
0
-10
-20
-30
-40
-50
0
1
2
3
frequency [kHz]

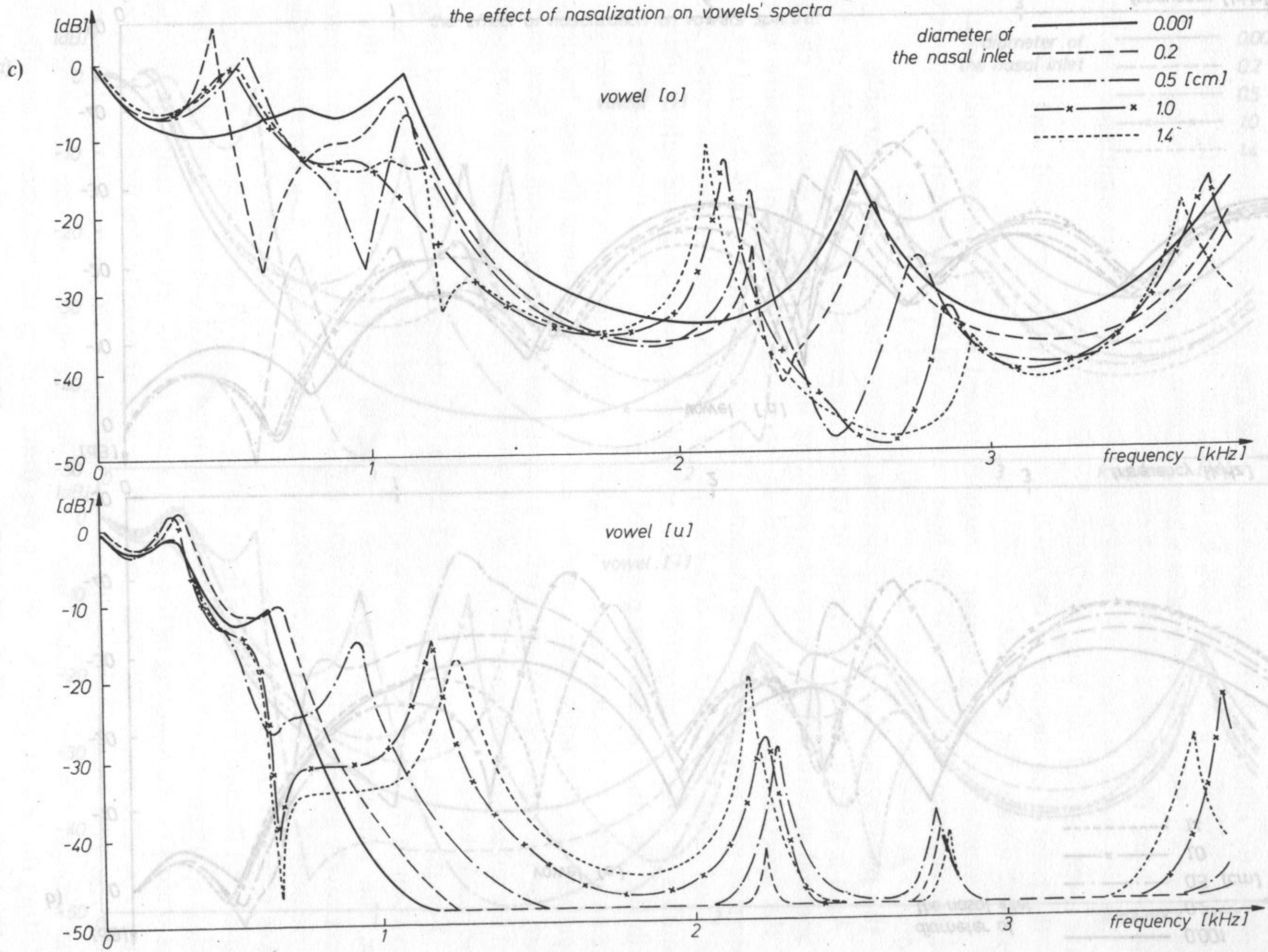


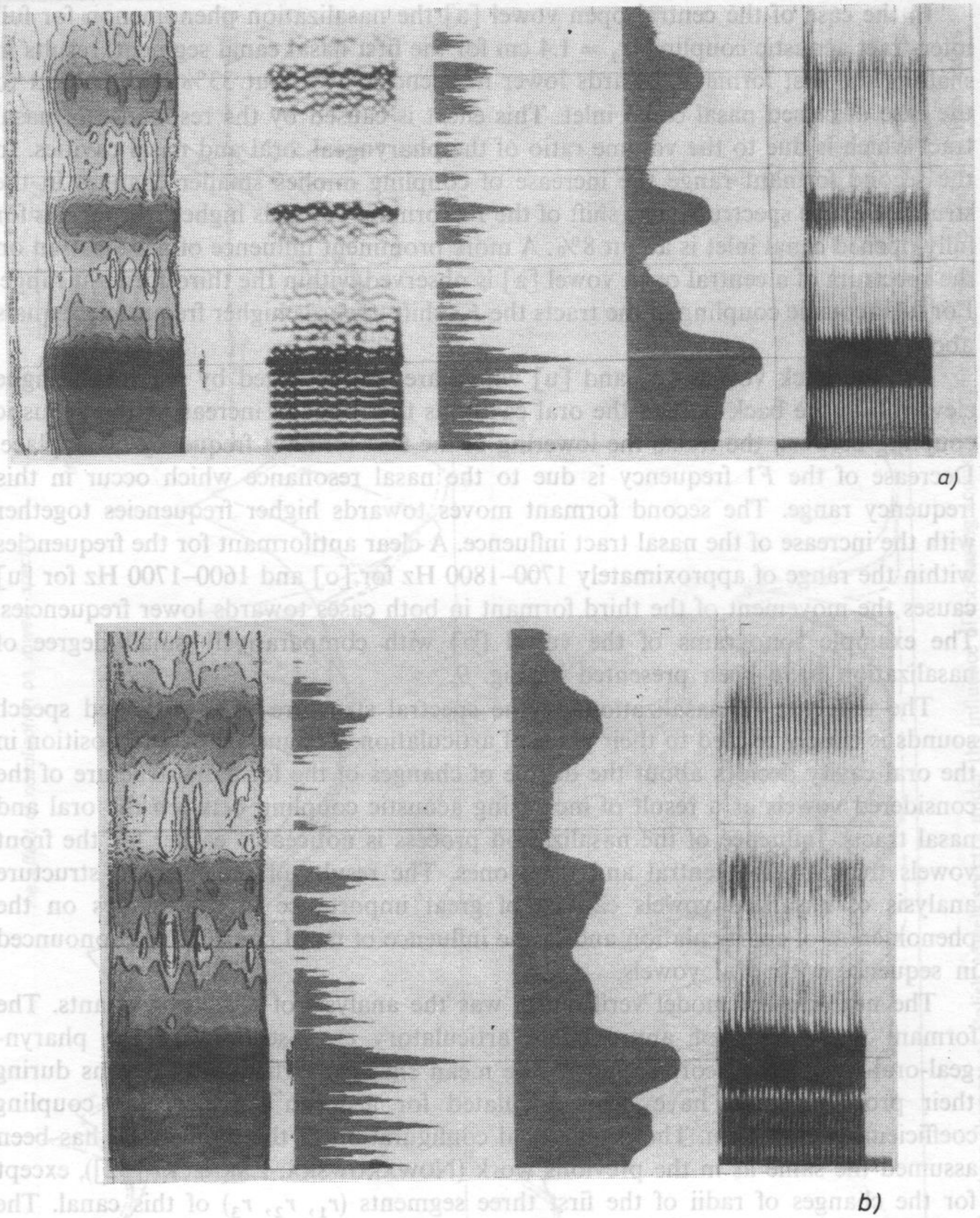
FIG. 8 a, b, c. Formant structure changes of Polish vowels spectra under the influence of increasing acoustic coupling between the oral and nasal tracts

In the case of the central open vowel [a] the nasalization phenomenon for full inter-tract acoustic coupling ($r_1 = 1.4$ cm for the first nasal canal segment) results in shafting the first formant towards lower frequencies by about 33% with respect to the case of closed nasal canal inlet. This effect is caused by the resonance of nasal tract which is due to the volume ratio of the pharyngeal, oral and nasal cavities. In the second formant range the increase of coupling implies smaller changes in the structure of the spectrum. The shift of the $F2$ formant towards higher frequencies for fully opened canal inlet is about 8%. A more prominent influence of nasalization on the spectrum of a central open vowel [a] is observed within the third formant range. For full acoustic coupling of the tracts the $F3$ shift towards higher frequencies equals about 13%.

For the back vowels [o] and [u] which are characterized by maximal tongue elevation in the back part of the oral cavity as the result of increasing the acoustic coupling between the tracts the lowering of the first formant frequency takes place. Decrease of the $F1$ frequency is due to the nasal resonance which occur in this frequency range. The second formant moves towards higher frequencies together with the increase of the nasal tract influence. A clear antiformant for the frequencies within the range of approximately 1700–1800 Hz for [o] and 1600–1700 Hz for [u] causes the movement of the third formant in both cases towards lower frequencies. The example sonograms of the vowel [o] with comparatively small degree of nasalization have been presented in Fig. 9.

The influence of nasalization on the spectral structure of investigated speech sounds is closely related to their place of articulation. Tongue shape and position in the oral cavity decides about the degree of changes of the formant structure of the considered vowels as a result of increasing acoustic coupling between the oral and nasal tracts. Influence of the nasalization process is noticeable earlier for the front vowels than for the central and back ones. The results of the formant structure analysis of nasalized vowels can be of great importance in the studies on the phenomenon of coarticulation and of the influence of nasal consonants pronounced in sequence with oral vowels.

The next step of model verification was the analysis of nasal consonants. The formant frequencies for approximate articulatory cross-sections of the pharyngeal-oral tract which correspond to the mean shape of articulation organs during their pronounciation have been calculated for a mean value of the coupling coefficients $r_1 = 0.7$ cm. The geometrical configuration of the nasal canal has been assumed the same as in the previous work (NOWAKOWSKA, ŻARNECKI [32]), except for the changes of radii of the first three segments (r_1, r_2, r_3) of this canal. The geometrical configuration of the oral cavity has been found with the method of successive approximations according to the criterion of the frequency characteristic conformity in respect of the displacement of poles and zeros with the data given by JASSEM [19] which concerned the nasal consonants articulation and their formant frequencies. In order to find an approximate oral tract configuration during the pronounciation nasal consonants an analysis of the influence of the oral tract length


a)
b)

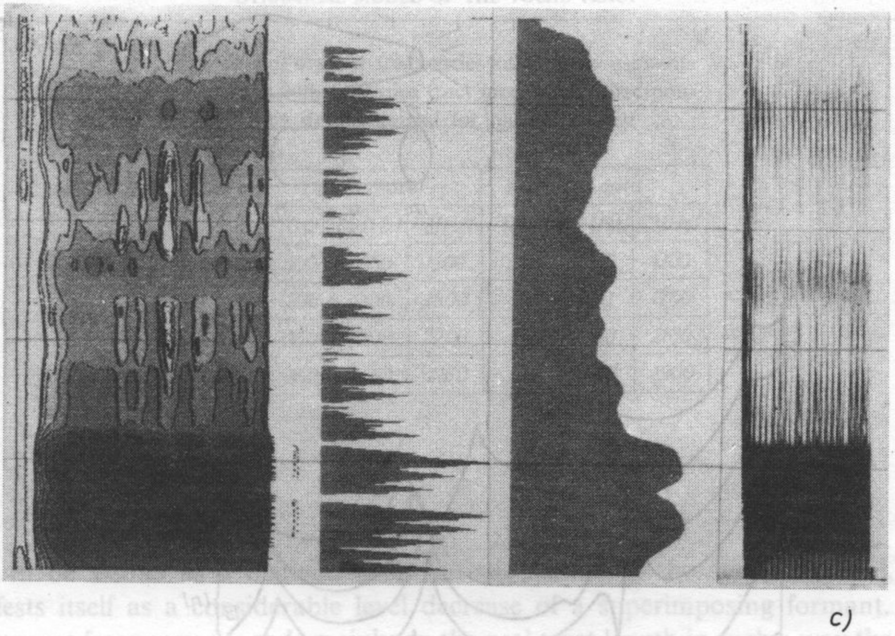


FIG. 9. Sonogram of the vowel [o] in the case of oral (a) and nasalized articulation $r_1 = 0.1$ cm (b) and $r_2 = 0.2$ cm (c). From left to right there are presented: three-dimensional spectrogram with 5 dB level contours, spectral cross-section with a band-pass filter of width 45 Hz, spectral section with a band-pass filter of width 300 Hz and three-dimensional spectrogram with the 300 Hz filter

on the displacement of poles and zeros for a constant, average acoustic inter-tract coupling has been carried out. The oral tract length was being reduced from the seventeenth segment to the eleventh one, for constant cross-section along the whole tract. The frequency characteristics corresponding to various oral tract lengths have been shown in Fig. 10. In the first approximation it can be assumed that during the articulation of a nasal consonant [m] this length equals 17 segments; for the consonant [n] – 14–15 segments, for [ɳ] about 12–13 segments and for [ŋ] – 10 11 segments. Basing on the above analysis and the data presented in the paper BOLLA, FÖLDI [23] the approximate articulatory cross-sections of the pharyngeal-oral tract for nasal consonants articulation have been found and the formant frequencies have been calculated (Table 4). The acoustic pressure courses at the mouth and nasal openings for hypothetical articulation cross-section of nasal consonants obtained with the computer model are displayed in Figs. 11a, b, c, d.

In the nasal consonants spectra the formant caused by the pharyngeal cavity resonance with frequency about 300 Hz dominates. The formant resulting from the oral resonant cavity partly coincides with the nasal formants. The formant frequencies of the given speech sounds depend on the length of the oral resonant

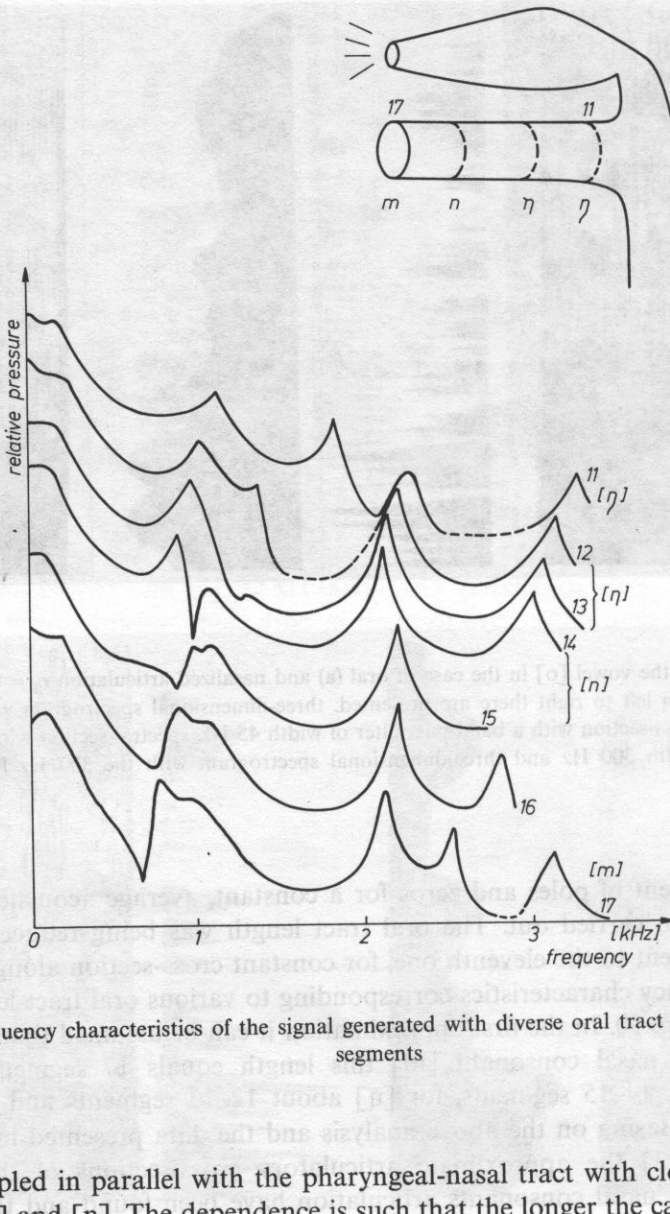


FIG. 10. Frequency characteristics of the signal generated with diverse oral tract lengths from 11 to 17 segments

cavity coupled in parallel with the pharyngeal-nasal tract with closed outlet, in the case of [m] and [n]. The dependence is such that the longer the cavity the lower the formant frequencies. In the spectrum of the nasal consonant an antiformant appears which usually manifests itself as a clear amplitude level decrease in a certain frequency range, characteristic for a given speech sound. This phenomenon takes place if an antiformant appears between the formants, whereas in the case of an antiformant which is near one of the formant frequencies the effect can be blurred to some extent. In general, the influence of a formant is stronger and an antiformant

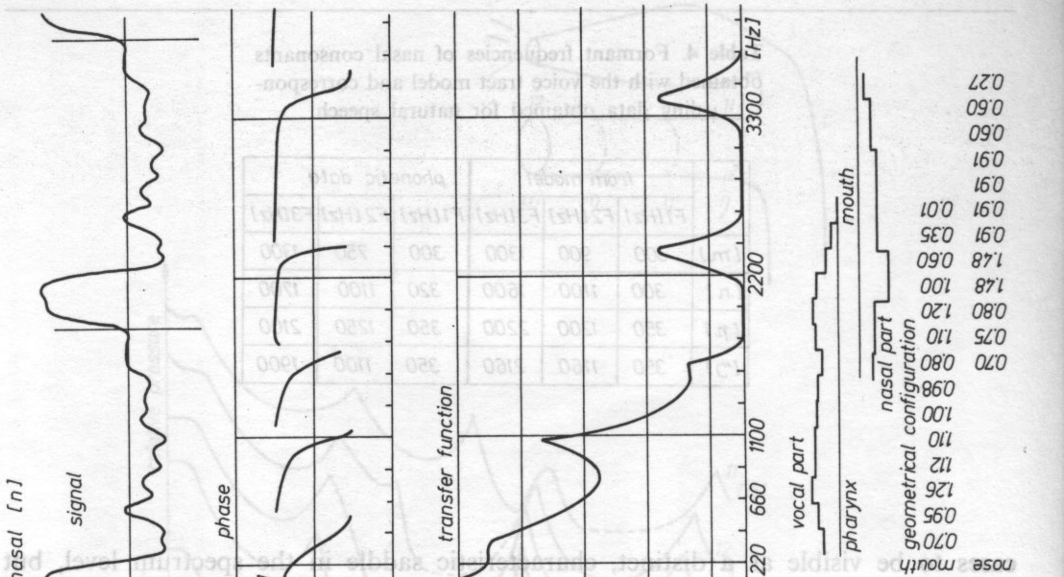
Table 4. Formant frequencies of nasal consonants obtained with the voice tract model and corresponding data obtained for natural speech

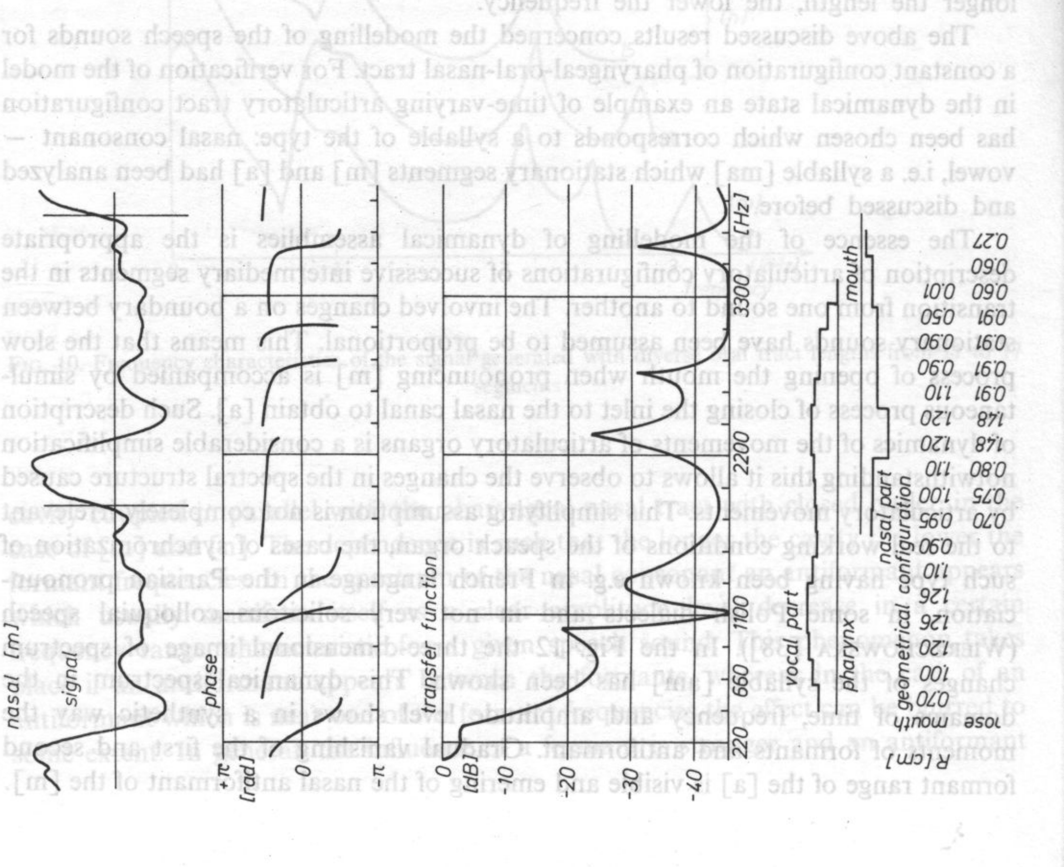
	from model			phonetic data		
	F1[Hz]	F2[Hz]	F3[Hz]	F1[Hz]	F2[Hz]	F3[Hz]
[m]	300	900	1300	300	750	1300
[n]	300	1100	1600	320	1100	1700
[ɲ]	350	1200	2200	350	1250	2100
[ɔ̃]	350	1160	2160	350	1100	1900

cases to be visible as a distinct, characteristic saddle in the spectrum level, but manifests itself as a considerable level decrease of a superimposing formant. The antiformant frequency depends mainly on the oral tract length in such a way that the longer the length, the lower the frequency.

The above discussed results concerned the modelling of the speech sounds for a constant configuration of pharyngeal-oral-nasal tract. For verification of the model in the dynamical state an example of time-varying articulatory tract configuration has been chosen which corresponds to a syllable of the type: nasal consonant – vowel, i.e. a syllable [ma] which stationary segments [m] and [a] had been analyzed and discussed before.

The essence of the modelling of dynamical assemblies is the appropriate description of articulatory configurations of successive intermediary segments in the transition from one sound to another. The involved changes on a boundary between stationary sounds have been assumed to be proportional. This means that the slow process of opening the mouth when pronouncing [m] is accompanied by simultaneous process of closing the inlet to the nasal canal to obtain [a]. Such description of dynamics of the movements of articulatory organs is a considerable simplification notwithstanding this it allows to observe the changes in the spectral structure caused by articulatory movements. This simplifying assumption is not completely irrelevant to the real working conditions of the speach organ, the cases of synchronization of such type having been known e.g. in French language in the Parisian pronounciation, in some Polish dialects and in not very solicitous colloquial speech (WIERZCHOWSKA [38]). In the Fig. 12 the three-dimensional image of spectrum changes of the syllable [am] has been shown. This dynamical spectrum in the domains of time, frequency and amplitude level shows in a synthetic way the moments of formants and antiformant. Gradual vanishing of the first and second formant range of the [a] is visible and emering of the nasal antiformant of the [m].


nasal [n]
signal
phase
+π
[rad]
0
-π
transfer function
0
[dB]
-10
-20
-30
-40
220
660
1100
2200
3300
[Hz]
vocal part
nasal part
mouth
pharynx
geometrical configuration
R [cm]
nose
mouth
0.70 0.95 1.26 1.12 1.10 1.00 0.98 0.80 1.10 1.20 1.00 0.60 0.35 0.01
0.70 0.75 0.80 1.48 1.48 0.91 0.91 0.91 0.91 0.60 0.60 0.27


nasal [m]
signal
phase
+π
[rad]
0
-π
transfer function
0
[dB]
-10
-20
-30
-40
220
660
1100
2200
3300
[Hz]
vocal part
nasal part
mouth
pharynx
geometrical configuration
R [cm]
nose
mouth
0.70 1.00 1.20 1.26 1.26 1.10 0.90 0.95 1.00 1.10 1.20 1.20 1.10 0.90 0.90 0.50 0.01
0.70 0.75 0.80 1.48 1.48 0.91 0.91 0.91 0.91 0.60 0.60 0.27

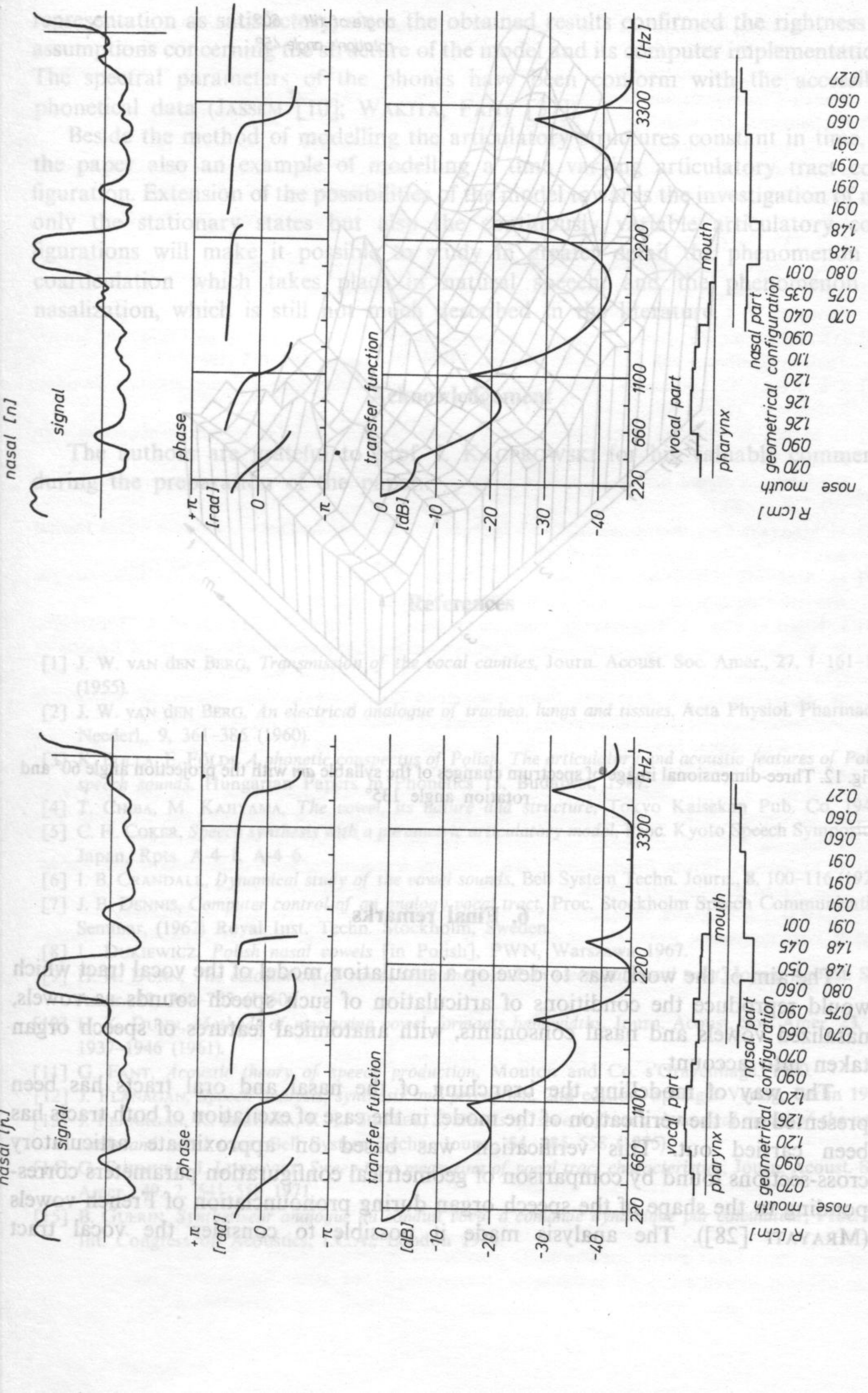


FIG. 11 a, b, c, d. Frequency characteristics of Polish nasal consonants obtained with the model

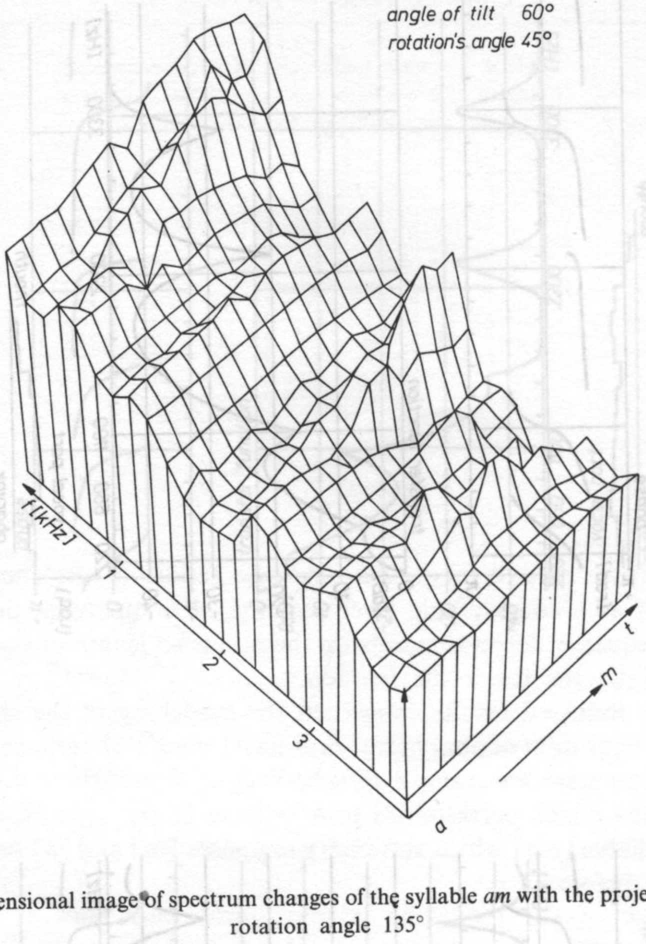


Fig. 12. Three-dimensional image of spectrum changes of the syllable *am* with the projection angle 60° and rotation angle 135°

6. Final remarks

The aim of the work was to develop a simulation model of the vocal tract which would reproduce the conditions of articulation of such speech sounds as vowels, nasalized vowels and nasal consonants, with anatomical features of speech organ taken into account.

The way of modelling the branching of the nasal and oral tracts has been presented and the verification of the model in the case of excitation of both tracts has been carried out. This verification was based on approximate articulatory cross-sections found by comparison of geometrical configuration parameters corresponding to the shape of the speech organ during pronounciation of French vowels (MRAYATI [28]). The analysis made it possible to consider the vocal tract representation as satisfactory, since the obtained results confirmed the rightness of assumptions concerning the structure of the model and its computer implementation. The spectral parameters of the phones have been conform with the accessible phonetical data (JASSEM [10]; WAKITA, FANT [37]).

Beside the method of modelling the articulatory structures constant in time, in the paper also an example of modelling a time varying articulatory tract configuration. Extension of the possibilities of the model towards the investigation of not only the stationary states but also the continously variable articulatory configurations will make it possible to study in greater detail the phenomenon of coarticulation which takes place in natural speech, and the phenomenon of nasalization, which is still not much described in the literature.

Acknowledgement

The authors are grateful to prof. J. KACPROWSKI for his valuable comments during the preparation of the paper.

References

[1] J. W. VAN DEN BERG, *Transmission of the vocal cavities*, Journ. Acoust. Soc. Amer., **27**, 1–161–168 (1955).

[2] J. W. VAN DEN BERG, *An electrical analogue of trachea, lungs and tissues*, Acta Physiol. Pharmacol. Neederl., **9**, 361–385 (1960).

[3] K. BOLLA, E. FÖLDI, *A phonetic conspectus of Polish. The articulatory and acoustic features of Polish speech sounds.* Hungarian Papers in Phonetics 18, Budapest, 1987.

[4] T. CHIBA, M. KAJIYAMA, *The vowel, its nature and structure*, Tokyo Kaisekan Pub. Co, 1941.

[5] C. H. COKER, *Speech synthesis with a parametric articulatory model*, Proc. Kyoto Speech Symposium, Japan, Rpts. A-4–1, A-4–6.

[6] I. B. CRANDALL, *Dynamical study of the vowel sounds*, Bell System Techn. Journ., **8**, 100–116 (1927).

[7] J. B. DENNIS, *Computer control of an analogy vocal tract*, Proc. Stockholm Speech Communication Seminar, (1962) Royal Inst. Techn. Stockholm, Sweden.

[8] L. DUKIEWICZ, *Polish nasal vowels* [in Polish], PWN, Warszawa 1967.

[9] H. K. DUNN, *The calculation of vowel resonances and an electrical vocal tract*, Journ. Acoust. Soc. Amer. **22**, 740–753 (1950).

[10] H. K. DUNN, *Methods of measuring vowel formants bandwidths*, Journ. Acoust. Soc. Amer. **23**, 12, 1937–1946 (1961).

[11] G. FANT, *Acoustic theory of speech production*, Mouton and Co. s'Gravenhage, 1960.

[12] J. FLANAGAN, *Speech analysis, synthesis and perception*, 2nd edition, Springer-Verlag, Berlin 1972.

[13] J. FLANAGAN, K. ISHIZAKA, K. L. SHIPLEY, *Synthesis of speech from a dynamical model of the vocal cords and vocal tract*, Bell System Techn. Journ., **54**, 485–558 (1975).

[14] G. FUJIMURA, J. LINDQVIST, *Sweep-tone measuring of vocal tract characteristics*, Journ. Acoust. Soc. Amer., **49**, 2, 541–558 1971.

[15] B. GUERIN, *Synthetiseur analogue du conduit vocal a comande dynamique par calculateur*, Proc. 8th Int. Congress on Acoustics, I.C.A., London 1974.

[16] H. L. HECKER, *Studies of nasal consonants with an articulatory speech synthesizer*, Journ. Acoust. Soc. Amer., 34, 179–188, 1962.

[17] J. M. HEINZ, *An analysis of speech spectra in terms of a model of articulation*, Proc. Stockholm Speech Communication Seminar, Royal Inst. Techn., Stockholm, 1962.

[18] K. ISHIZAKA, *On model of the larynx*, Journ. Acoust. Soc. Amer. **22**, 293–294 (1968).

[19] W. JASSEM, *Foundations of acoustic phonetics* [in Polish] PWN, Warszawa 1979.

[20] J. KACPROWSKI, *Theoretical bases of the synthesis of Polish vowels in resonance circuits*, in: Speech Analysis and Synthesis vol. 1 [ed.] W. Jassem, PWN, Warszawa, 219–287 (1968).

[21] J. KACPROWSKI, W. MIKIEL, A. SZEWCZYK, *Acoustical modelling of cleft palate*, Archives of Acoustics, **1**, 2, 137–158 (1976).

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

[23] J. KACPROWSKI, *A simulative model of the vocal tract including the effect of nasalization*, Archives of Acoustics, **2**, 4, 236–256 (1977).

[24] J. KACPROWSKI, *An acoustical model of the vocal tract for the diagnostics of cleft palate*, in: Speech Analysis Synthesis, vol. 5 [ed.] W. Jassem, PWN, Warszawa, 165–183 (1981).

[25] J. L. KELLY, C. LOCHBAUM, *Speech synthesis*, Proc. Stockholm Speech Communication Seminar, Royal Inst. Techn., Stockholm, (1962).

[26] D. MADEJ, W. NOWAKOWSKA, *Realisation of a microcomputer model of speech signal generator* (in Polish), IFTR Reports 3 (1986).

[27] S. MAEDA, *A digital simulation method of the vocal tract system*, Speech Communication, **1**, 3–4, 199–229 (1982).

[28] M. MRAYATI, *Contribution aux études sur la production de la parole*. These de Docteur d'Etat. Institut National Polytèchnique de Grenoble.

[29] M. MRAYATI, B. GUERIN, *Étude des caracteristiques acoustiques des voyelles orales françaises par simulation du conduit vocal avec pertes*, Revue d'Acoustique, **9**, 36, 18–32 (1976).

[30] L. NORD, G. FANT, P. BRANDERUD, *A note on the vocal tract impedance*, STL-QPSR, **4**, 13–20 (1976).

[31] W. NOWAKOWSKA, *A simulation model of pharyngeal-oral voice tract* (in Polish), IFTR Reports 41 (1983).

[32] W. NOWAKOWSKA, P. ŻARNECKI, *Preliminary model investigations on the influence of nasalization phenomenon on the spectral structure of vowels* (in Polish), IFTR Reports 39 (1984).

[33] W. NOWAKOWSKA, *Investigations of the influence of nasalization on the spectral structure of Polish vowels and consonants* (in Polish), IFTR Reports 46 (1985).

[34] G. ROSEN, *Dynamic analog speech synthesizer*, Journ. Acoustic. Soc. Amer., **30**, 201–209 (1958).

[35] K. N. STEVENS, S. KOSSOWSKI, G. FANT, *An electrical analog of the vocal tract*, Journ. Acoust. Soc. Amer., **25**, 734–742 (1953).

[36] K. STEVENS, A. HOUSE, *Development of a quantitative description of vowel articulation*, Journ. Acoust. Soc. Amer. **27**, 484–493 (1955).

[37] H. WAKITA, G. FANT, *Toward a better vocal tract model*, STL-QPSR, **1**, 9–29 (1978).

[38] B. WIERZCHOWSKA, *Phonetics and phonology of the Polish language* (in Polish) Ossolineum Warsaw (1980).

[39] W. WIĘŹLAK, *Application of the simplified articulatory description of speach signal to the recognition of a limited set of isolated words* (in Polish), Ph.D. thesis, IFTR PAS Warsaw (1987).

Received on February 2, 1988.