

2.23 — akustyka mowy: akustyczne modelowanie
organu mowy

Janusz Kacprowski

AKUSTYCZNE MODELOWANIE
ORGANU MOWY

14 / 1982

P.269



WARSZAWA 1982

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 29 marca 1982 r.

Zarejestrowana pod nr 14/1982



57060



Na prawach rękopisu

Instytut Podstawowych Problemów Techniki PAN

Nakład 150 egz. Ark. wyd. 1,7. Ark. druk. 2,5

Oddano do drukarni w kwietniu 1982 r.

Nr zamówienia 18470/82

Z - 87.

Warszawska Drukarnia Naukowa, Warszawa,
ul. Śniadeckich 8

Janusz Kacprowski
Zakład Akustyki Cybernetycznej
IPPT - PAN

AKUSTYCZNE MODELOWANIE ORGANU MOWY

1. Przedmiot i zakres pracy

Mową jako dźwiękową formą języka zajmują się w sensie badawczym dwie ściśle ze sobą związane dziedziny nauki: akustyka mowy i fonetyka akustyczna. Pierwsza rozpatruje sygnał mowy przede wszystkim z fizycznego punktu widzenia, traktując w sposób drugoplanowy jego aspekt lingwistyczny. Fonetyka natomiast, posługując się odpowiednio rozszerzonym i ukierunkowanym aparatem pojęciowym i metodycznym akustyki, traktuje ten sygnał głównie jako dźwiękową realizację fonetycznego kodu językowego, służącego do wzajemnego porozumiewania się ludzi w określonych obszarach etnograficznych. W obecnej pracy stosowany jest ten pierwszy, to jest fizyczny sposób interpretacji sygnału mowy.

Przedmiotem pracy jest modelowanie podstawowych procesów i czynności głosotwórczych, fonacyjnych i artykulacyjnych, organu mowy, których efektem finalnym jest akustyczny sygnał mowy. W parametrach fonetyczno-akustycznych tego sygnału zakodowane są informacje, określające zarówno treść fonetyczną i lingwistyczną przekazywanej nim wiadomości, jak i inwariantne cechy osobnicze głosu jej nadawcy. Informacje te stanowią przedmiot zainteresowań i badań fonetyki akustycznej, lingwistyki stosowanej i kilku innych dziedzin nauki i techniki, m.in. telekomunikacji, informatyki oraz medycyny w zakresie laryngologii i foniatryi.

Z fizycznego punktu widzenia mowę należy traktować jako przebieg czasowy ciśnienia akustycznego $p(t)$ fali dźwiękowej, wytwarzanej przez układ oddechowy i efekторы artykulacyjne

człowieka, których działanie kinematyczne jest sterowane przez centralny układ nerwowy, a korygowane przez akustyczne sprzężenie zwrotne, jakie istnieje między narządami mowy i słuchu. Poprawność przekazywania informacji za pomocą mowy zależy zatem od sprawności i współdziałania trzech systemów biocybernetycznych:

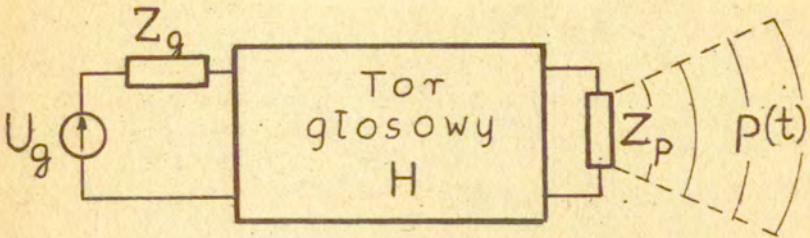
- centralnego układu nerwowego,
- mięśniowego układu oddechowo-artykulacyjnego,
- percepcyjnego układu słuchowego.

Ograniczenie się do modelowania działania wyizolowanego układu oddechowo-artykulacyjnego jest już zatem z natury rzeczy poważnym uproszczeniem, które jest konsekwencją założonej z góry fizycznej, a ściśle biorąc akustycznej metody analizy rozpatrywanych zjawisk i procesów.

2. Fizyczna interpretacja procesu wytwarzania dźwięków mowy

Zagadnieniem wytwarzania w sposób sztuczny dźwięków mowopodobnych, a więc prymitywnie pojmowaną syntezą mowy, zajmowano się od dawna. Obszerny przegląd historyczny stosowanych do tego celu bardziej lub mniej złożonych urządzeń, głównie mechanicznych, przedstawił KÖSTER [27]. Naukowe podstawy modelowania procesów i czynności głosotwórczych organu mowy powstały jednak dopiero na początku lat 60-tych tego wieku, kiedy FANT [3] opracował akustyczną teorię wytwarzania mowy, a FLANAGAN [7] zebrał i usystematyzował w swoim dziele o analizie, syntezie i percepcji mowy wyniki dotychczasowych, w tym również własnych, prac badawczych w tej dziedzinie, z uwzględnieniem dostępnych wówczas metod i środków techniki komputerowej.

Od tej pory sygnał mowy, wyrażony przebiegiem czasowym $p(t)$ fali głosowej, zaczęto traktować, w ujęciu klasycznej teorii układów linearnych, jako odpowiedź biernego akustycznego układu akustycznego o właściwościach rezonansowych, utworzonego przez odpowiednio ukształtowane wnęki toru głosowego, w ogólnym przypadku gardłowo-ustno-nosowego, pobudzanego ze źródła lub źródła energii akustycznej i obciążonego impedancją promieniowania otworu ust oraz ewentualnie nosa, jak to przedstawiono schematycznie na rys.1 dla szczególnego przypadku artykulacji ustnej i wzbudzenia krtaniowego.



Rys.1. Schemat zastępczy organu mowy przy artykulacji ustnej i wzbudzeniu krtaniowym

Ogólne wyrażenie na ciśnienie akustyczne fali głosowej przy tych założeniach jest określone w notacji operatorowej Laplace'a, a więc w domenie częstotliwości zespolonej $s = \sigma + j\omega$, wzorem

$$P(s) = U_g(s) \cdot H(s) \cdot Z_p(s) \quad /1/$$

w którym $U_g(s)$ jest funkcją wzbudzającą źródła krtaniowego o impedancji wewnętrznej $Z_g(s)$, $H(s)$ oznacza transmitancję toru głosowego, wyrażoną stosunkiem prędkości objętościowej $U_o(s)$ powietrza w otworze ust do prędkości objętościowej $U_g(s)$ u wlotu toru głosowego, t.j. w otworze głośni, a $Z_p(s)$ jest impedancją promieniowania otworu ust, traktowanego jako akustyczne źródło fali głosowej. Ponieważ transformacja Laplace'a jest matematyczną operacją odwracalną, przebieg czasowy ciśnienia akustycznego fali głosowej $p(t)$ można wyznaczyć analitycznie ze wzoru /1/ za pomocą przekształcenia odwrotnego

$$p(t) = \mathcal{L}^{-1} \{ P(s) \} \quad /2/$$

Formalną i bardzo dogodną do celów badawczych konsekwencją odwracalności tej operacji matematycznej jest możliwość stosowania dwójstej i równoważnej prezentacji sygnału mowy, mianowicie w domenie czasu lub w domenie częstotliwości.

Przy artykulacji głosek bezdźwięcznych o wzbudzeniu ponadkrtaniowym, zwłaszcza spółgłosek trących i zwarto-trących, np. typu o b s t r u e n t, zmienia się sens fizyczny i lokalizacja źródła wzbudzającego drgania w torze głosowym, które w tym przypadku ma charakter fizyczny źródła szumowego, działającego

na zasadzie turbulencyjnego przepływu powietrza przez przewężenie, utworzone w określonym przekroju toru głosowego.

Ze wzoru wyjściowego /1/ wynika, że akustyczny sygnał mowy można przedstawić w postaci iloczynu trzech czynników, opisujących odpowiednio:

- funkcję wzbudzącą źródła krtaniowego U_g ,
- funkcję transmitancji toru głosowego H ,
- funkcję impedancji promieniowania ust Z_p .

Zagadnienie akustycznego modelowania organu mowy sprowadza się zatem do wyrażenia poszczególnych funkcji oddzielnie w postaci analitycznej, opisującej ich cechy fizyczne /akustyczne/, istotne z punktu widzenia czynności głosotwórczych organu mowy odpowiednio w procesie fonacji / U_g /, artykulacji / H / i emisji / Z_p / sygnału mowy. Dodatkowym problemem jest uwzględnienie w postaci każdej z tych funkcji oddziaływania i wpływu na nią funkcji pozostałych, gdyż wszystkie trzy biorą nierozłączny udział w procesie wytwarzania mowy i opisują elementy składowe realizującego ten proces złożonego układu biocybernetycznego, jakim jest organ mowy.

3. Impedancja promieniowania ust

Trzecim czynnikiem we wzorze ogólnym /1/ jest impedancja promieniowania $Z_p(s)$, zdefiniowana jako stosunek operatorowy ciśnienia akustycznego $P(s)$ w płaszczyźnie warg do prędkości objętościowej $U_o(s)$ w otworze ust:

$$Z_p(s) = \frac{P(s)}{U_o(s)} \quad /3/$$

Można wykazać, że funkcja ta tylko w niewielkim stopniu zależy od warunków artykulacji dźwięków mowy i w pierwszym przybliżeniu mogłaby być traktowana jako czynnik stały. Niemniej jednak analityczne wyznaczenie jej postaci do celów modelowania czynności głosotwórczych organu mowy jest niezbędne co najmniej z dwóch względów, gdyż impedancja promieniowania:

- opisuje akustyczne warunki promieniowania fali głosowej w przestrzeń otaczającą,
- jest miarą obciążenia akustycznego układu przesyłowego jakim jest tor głosowy i określa wpływ tego obciążenia na właściwości transmisyjne toru, wyrażone funkcją transmitancji $H(s)$.

Zagadnienie wyznaczania impedancji promieniowania otworu ust, aproksymowanego do postaci tłoka kołowego o promieniu \underline{r} , można rozwiązać analitycznie przyjmując, że uproszczonym modelem akustycznym źródła fali głosowej jest fizyczne źródło dźwięku w jednej z pozostających do wyboru klasycznych postaci, znanych z teorii akustyki, mianowicie:

- /a/ kuli pulsującej o promieniu \underline{r} , a więc źródła quasi-punktowego,
- /b/ tłoka drgającego w nieskończenie wielkiej odgradzie płaskiej,
- /c/ tłoka drgającego w obudowie kulistej o promieniu \underline{R} , symulującej głowę ludzką.

Ten ostatni model wydaje się najbardziej realistyczny z anatomicznego punktu widzenia.

Analiza porównawcza tych modeli /PLANAGAN [7], KACPROWSKI [20,21]/ wykazała jednak, że wybór optymalnego modelu nie jest krytyczny i może być dokonany w drodze racjonalnego kompromisu, w oparciu o weryfikację doświadczalną.

Przyjmując za punkt wyjścia przypadek /b/ tłoka o promieniu \underline{r} w odgradzie płaskiej, otrzymuje się znane wyrażenie na jego impedancję promieniowania w postaci [31]:

$$Z_p = \frac{\rho c}{A} \left[\left(1 - \frac{J_1(2\beta r)}{\beta r} \right) + j \frac{S_1(2\beta r)}{\beta r} \right] \quad /4/$$

gdzie: $J_1(x)$ i $S_1(x)$ są odpowiednio funkcją Bessela 1-go rodzaju i funkcją Struvego 1-go rzędu,

$A = \pi r^2$ jest powierzchnią tłoka o promieniu \underline{r} ,

$\beta = \omega/c = 2\pi/\lambda$ jest współczynnikiem długości fali.

Postać funkcji $J_1(x)$ i $S_1(x)$ jest trudna do odtworzenia modelowego i wymaga ich rozwinięcia na szereg potęgowe. Sprawa upraszcza się jednak znacznie w zakresie częstotliwości, w którym $\beta r = 2\pi r/\lambda \ll 1$. Łatwo wykazać, że ograniczenie to jest spełnione w rzeczywistych warunkach artykulacji. Maksymalna powierzchnia otworu ust w przypadku samogłoski otwartej /a/ jest rzędu $A \approx 5 \text{ cm}^2$, a w przypadku samogłoski zaokrąglonej /u/ maleje do wartości $A \approx 0.9 \text{ cm}^2$, czemu odpowiadają zastępcze promienie tłoka $r_a \approx 1.3 \text{ cm}$ i $r_u \approx 0.5 \text{ cm}$, czyli warunek ten obowiązuje w zakresie częstotliwości $f < 5000 \text{ Hz}$. W rozwinięciu funkcji Bes-

sela i Struvego wystarczy w tym przypadku uwzględnić tylko pierwsze wyrazy szeregu, co prowadzi do uproszczenia wyrażenia /4/ do postaci:

$$Z_p = \frac{\rho c}{A} \left[\frac{(\beta r)^2}{2} + j \frac{8(\beta r)}{39\pi} \right] \quad /5/$$

MORSE i INGARD [29] wykazali, że uproszczoną postacią funkcji impedancji promieniowania Z_p modelu /c/, t.j. tłoka kołowego o promieniu r w obudowie kulistej o promieniu R , symulującej głowę ludzką, jest wyrażenie:

$$Z_p = \frac{\rho c}{A} \left[\frac{(\beta r)^2}{2} \cdot K(r, R, \lambda) + j \frac{8(\beta r)}{39\pi} \cdot S(r, R, \lambda) \right] \quad /6/$$

w którym współczynniki K i S są funkcjami stosunku r/R oraz rozmiarów obudowy kulistej wyrażonych stosunkiem $\beta R = 2\pi R/\lambda$. W przypadku gdy $r/R \ll 1$, co jest spełnione w warunkach rzeczywistych, gdyż $R \approx 9$ cm, czyli $r/R \approx 0.1$, wartości współczynników K i S dążą w granicy do 1, gdy $\beta R = 2\pi R/\lambda \rightarrow \infty$ i wyrażenie /6/ przybiera postać /5/ zgodnie z interpretacją fizyczną, gdyż w miarę wzrostu częstotliwości obudowa kulista nabiera cech akustycznych odgrody płaskiej.

Model /a/, t.j. źródło quasi-punktowe w postaci kuli pulsującej o promieniu r , ma znaczenie raczej teoretyczne niż praktyczne ze względu na jego sferyczną charakterystykę kierunkowości promieniowania, dość znacznie odbiegającą od rzeczywistych warunków emisji fali głosowej przez naturalny organ mowy, natomiast postacią funkcji impedancji promieniowania Z_p niewiele różni się od modeli /b/ i /c/. Impedancja promieniowania kuli pulsującej o promieniu r i powierzchni $A = 4\pi r^2$

$$Z_p = \frac{\rho c}{A} \left[\frac{(\beta r)^2}{1 + (\beta r)^2} + j \frac{\beta r}{1 + (\beta r)^2} \right] \quad /7/$$

przy czym, gdy $\beta r \ll 1$

$$Z_p = \frac{\rho c}{A} \left[(\beta r)^2 + j (\beta r) \right] \quad /8/$$

a więc ma postać jak dla modeli /b/ i /c/, różniąc się od nich jedynie współczynnikami liczbowymi składowych: rzeczywistej i urojonej, jak to widać z porównania wzorów /5/ i /8/.

Impedancja promieniowania wszystkich trzech modeli fizycznych ma zatem charakter bezwładnościowy i wyraża się wzorem

ogólnym

$$Z_p = R_p + j\omega M_p$$

/9/

gdzie:

$$R_p = k_1 \frac{\rho \omega^2}{2\pi c} \doteq \omega^2 \quad \text{i nie zależy od } r,$$

$$M_p = k_2 \frac{8\rho}{3\pi^2 r} \doteq \frac{1}{r} \quad \text{i nie zależy od } \omega,$$

a współczynniki liczbowe k_1 i k_2 są stałymi charakterystycznymi dla przyjętego typu modelu. Fizycznym modelem impedancji promieniowania ust jest zatem, zgodnie ze wzorem /9/, szeregowe połączenie rezystancji R_p i inertancji /masy akustycznej, indukcyjności/ M_p .

Alternatywnym rozwiązaniem modelu impedancji promieniowania jest równoległe połączenie masy akustycznej /indukcyjności/ M_p o tej samej wartości, co w układzie szeregowym, to jest

$$M_p = k_2 \frac{8\rho}{3\pi^2 r} \doteq \frac{1}{r}$$

z rezystancją R_p^*

$$R_p^* = k_3 \frac{c\rho}{r^2} \doteq \frac{1}{r^2}$$

przy czym oba te parametry nie zależą od częstotliwości. Szczegółową analizę porównawczą różnych postaci funkcji impedancji promieniowania ust znaleźć można w pracach [20,21].

Przy akustycznym modelowaniu organu mowy, wielkością fizyczną istotną z fonetyczno-akustycznego punktu widzenia jest ciśnienie akustyczne fali głosowej w punkcie pomiarowym położonym w przestrzeni w określonej odległości od płaszczyzny warg, zazwyczaj, choć nie zawsze, na osi symetrii ust. Wymagało to zweryfikowania przestrzennych charakterystyk promieniowania modeli fizycznych pod względem ich zgodności z odpowiednimi charakterystykami modelowanego układu biologicznego. Pomiar przestrzennego rozkładu ciśnienia akustycznego wokół głowy ludzkiej przeprowadzone na makietach naturalnej wielkości [6] wykazały nieznaczne odchylenia w stosunku do modeli teoretycznych, nie przekraczające ± 3 dB co najmniej w granicach kąta przestrzennego 60° względem osi symetrii ust.

W celu zoptymalizowania i znormalizowania funkcji impedancji promieniowania wprowadzono współczynnik korekcyjny $K_g(\omega)$

zależny od częstotliwości, który nadawałby wyrażeniu na Z_p charakter uniwersalny, niezależny - w wyniku racjonalnie uzasadnionego kompromisu - od zastosowanego modelu fizycznego. FANT w klasycznej pracy [3] stwierdził, że wartości tego współczynnika wzrastają w zakresie częstotliwości od 300 do 5000 Hz o około 7 dB. W nowszych pracach obserwuje się tendencję do upraszczania postaci tego współczynnika, bez istotnego wpływu na dokładność aproksymacji. Uniwersalna i stosowana obecnie w większości modeli impedancja promieniowania Z_p ma postać [36]

$$Z_p = R_p + j\omega M_p = \frac{\rho \omega^2}{4\pi c} \cdot K_s(\omega) + j\omega \frac{8\rho}{3\pi^2 r} \quad /10/$$

gdzie

$$K_s(\omega) = \begin{cases} \frac{0.6\omega}{2\pi \cdot 1600} + 1 & \longrightarrow \text{gd } 0 \leq \omega \leq 2\pi \cdot 1600 \text{ Hz} \\ 1.6 = \text{const.} & \longrightarrow \text{gd } \omega > 2\pi \cdot 1600 \text{ Hz} \end{cases} \quad /11/$$

a więc w zakresie małych częstotliwości rośnie prawie proporcjonalnie do częstotliwości i przybiera wartość stałą dla $f > 1600$ Hz.

Przytoczone rozważania, dotyczące akustycznego modelowania impedancji promieniowania otworu ust, są oczywiście słuszne, i to w szerszym zakresie częstotliwości, w przypadku impedancji promieniowania otworów nosa, których średnica jest odpowiednio mniejsza. W przypadku nazalizacji, zarówno naturalnej jak i wymuszonej, ze względu na małą w stosunku do długości fali λ wzajemną odległość przestrzenną otworów ust i nosa, wypadkowa fala głosowa w punkcie pomiarowym może być traktowana jako linearna superpozycja ciśnień akustycznych, pochodzących od obu źródeł dźwięku: ustnego i nosowego. Potwierdziły to badania eksperymentalne efektu nazalizacji, dokonane na modelach analogowych [26].

4. Funkcja wzbudzająca źródła krtaniowego

4.1. Mechanizm działania źródła krtaniowego.

Z fizycznego punktu widzenia źródło krtaniowe jest generatorem aerodynamicznym, którego drgania są określone przez parametry mechano-akustyczne fałdów głosowych i przez ciśnienie podgłośniowe, a w pewnym stopniu zależą też od jego obciążenia impedancją wejściową toru głosowego.

Fałdy głosowe o masie \underline{m} , sprężystości \underline{g} i rezystancji strat \underline{r} są wprawiane w quasi-periodyczne drgania wymuszone przez lokalne zmiany ciśnienia w otworze głośni, wywołane przepływem powietrza wypychanego z płuc przez kanały oskrzeli i tchawicę pod wpływem siły nacisku mięśni klatki piersiowej na zbiorniki płuc. Drgania fałdów głosowych powodują zmiany powierzchni otworu głośni A_g , a więc jej impedancji akustycznej Z_g , modulując tym samym strumień przepływającego przez nią powietrza, który u wlotu do kanału głosowego ma postać ciągu dyskretnych impulsów, które pełnią rolę funkcji źródła wzbudzającego drgania powietrza w kanale głosowym o częstotliwościach, odpowiadających rezonansom własnym, zależnym od jego chwilowej konfiguracji geometrycznej w określonych warunkach artykulacji.

4.2. Akustyczna impedancja głośni.

Skrajnie uproszczony model fizyczny głośni ma postać prostopadłościennej szczeliny o długości \underline{l} , głębokości \underline{d} i czynnej powierzchni A_g , prostopadłej do kierunku przepływu powietrza, która jest funkcją zmieniającej się quasi-okresowo szerokości szczeliny $\underline{w}(t)$. Klasycy teorii źródła krtaniowego, m.in. van den BERG [1] i WEGEL [38], wykazali, że składowa rzeczywista R_g impedancji głośni Z_g

$$Z_g = R_g + j\omega L_g = (R_v + R_k) + j\omega L_g \quad /12/$$

jest sumą dwóch składników, z których:

$R_v \doteq A_g(t)^{-3}$ - jest rezystancją strat tarciovych przepływu,

$R_k \doteq A_g(t)^{-1}$ - jest rezystancją kinetyczną, związaną z procesem przemiany epadku ciśnienia podgłośniowego P_g w głośni na energię kinetyczną przepływu, t.j. prędkości objętościowej $U_g(t)$.

Reaktancja $j\omega L_g = j\omega g d \cdot A_g(t)$ jest reaktancją bezwładnościową masy powietrza, drgającego w szczelinie głośni.

Z porównania wartości bezwzględnych i wzajemnych stosunków składowych R_v , R_k i $j\omega L_g$ impedancji Z_g wynikają następujące wnioski, istotne przy opracowaniu modelu funkcji źródła krtaniowego [23]:

/a/ W zakresie częstotliwości tonu krtaniowego $F_0 < 1000$ Hz i przy ciśnieniu podgłośniowym $P_g = 4 \div 16$ cm H_2O , co odpowia-

da normalnemu wysiłkowi głosowemu, $\omega L_g \ll R_g$, dominującą składową impedancji Z_g jest składowa rzeczywista, czyli $Z_g \approx R_g$.

/b/ W zakresie małych wartości powierzchni otworu głośni, gdy $A_g < 0.2 A_{g \max}$, a więc w stanach bliskich zwarcia fałdów głosowych, $R_v > R_k$ i dominującą składową rezystancji głośni jest oporność przepływu, czyli $R_g \approx R_v$. Natomiast w przeważającej części okresu tonu krtaniowego, gdy $A_g > 0.2 A_{g \max}$, $R_k > R_v$ i o rezystancji głośni decyduje oporność kinetyczna, czyli można przyjąć, że $R_g \approx R_k$.

/c/ Stała czasowa $\tau = L_g / R_g$ impedancji Z_g nie przekracza wartości $\tau_{\max} = 0.25$ ms, a więc jest o rząd wielkości mniejsza od najkrótszego okresu drgań fałdów głosowych $T_0 = 2.5$ ms głosów męskich / $F_0 = 400$ Hz/, co dowodzi, że okresowe zmiany impedancji Z_g "nadążają" za powodującymi je zmianami powierzchni otworu głośni wskutek drgań fałdów głosowych.

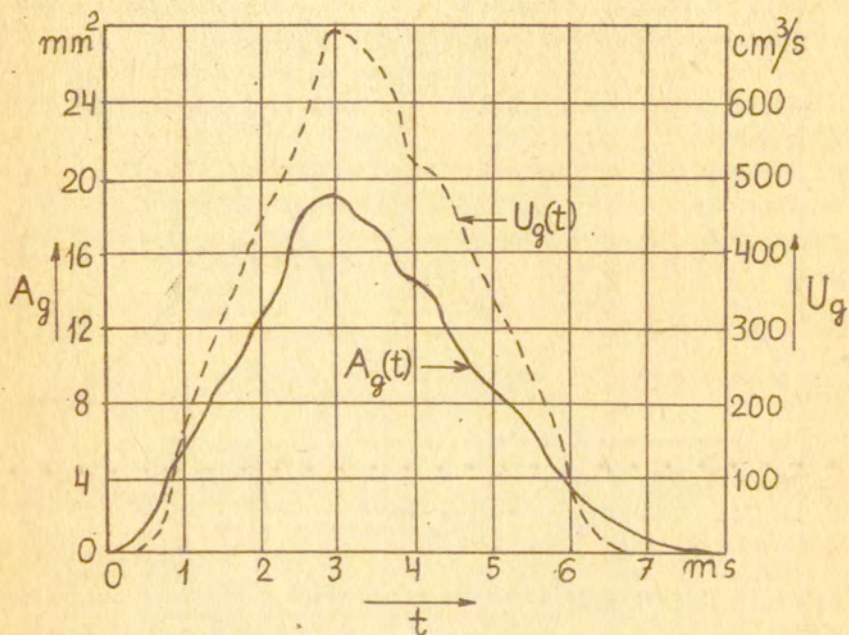
/d/ W przeciętnych warunkach anatomicznych dorosłego mężczyzny $A_{g \text{ śr}} \approx 5 \text{ mm}^2$ i moduł impedancji głośni $|Z_g| \approx 100 \cdot 10^5$ omów akustycznych MKS, a więc jest o rząd wielkości większy od impedancji wejściowej Z_t toru głosowego, którego wartości maksymalne $Z_{t \max}$, odpowiadające rezonansom 1/4-falowym toru, maleją w funkcji częstotliwości asymptotycznie do wartości impedancji falowej toru $Z_0 \approx 8.5 \cdot 10^5$ omów akustycznych MKS, osiągając wartości porównywalne z impedancją Z_g źródła krtaniowego jedynie w okolicach pierwszego formantu samogłoskowego F1.

Źródło krtaniowe można zatem traktować w układzie zastępczym jako źródło prądowe o stałej wydajności prędkości objętościowej U_g , niezależnej od chwilowej konfiguracji geometrycznej /artykulacyjnej/ toru głosowego, której wpływ na wielkość i kształt funkcji przepływu, t.j. prędkości objętościowej $U_g(t)$ daje się zauważyć jedynie w okolicach F1 samogłosek. Kształt fali przepływu ośrodka przez głośnię, t.j. postać funkcji wzbudzającej $U_g(t)$, można wyznaczyć w oparciu o prawa rządzące przepływem stałym, ze wzoru:

$$U_g(t) = \frac{P_g}{Z_g(t)} \approx \frac{P_g}{R_g(t)} = \frac{P_g}{\underbrace{A_g(t)^{-3}}_{R_v} + \underbrace{A_g(t)^{-1}}_{R_k}} \quad /13/$$

4.3. Kształt i widmo impulsów krtaniowych.

Na rys.2 podany jest przykład kształtu impulsów krtaniowych, reprezentowanych przez jeden okres funkcji powierzchni otworu głośni $A_g(t)$ w czasie fonacji /linia ciągła/ i odpowiadający mu okres funkcji przepływu $U_g(t)$ /linia przerywana/



Rys.2. Przykład kształtu impulsów krtaniowych [7]

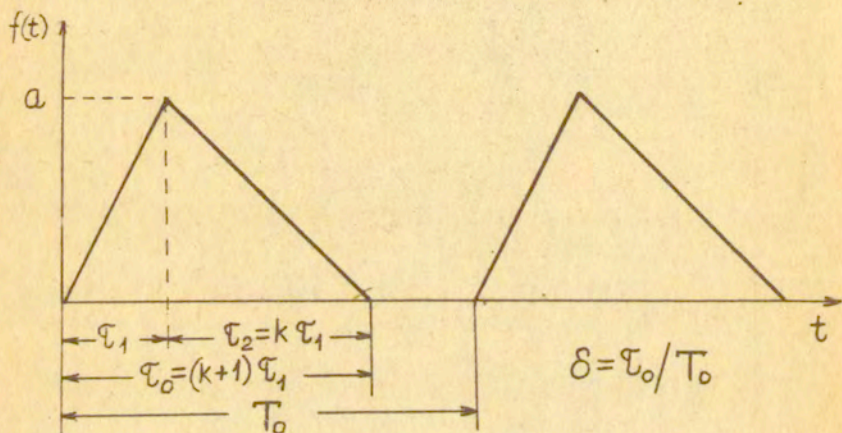
Ogólnie można stwierdzić, że:

/a/ Obie funkcje mają kształt zbliżony do trójkątnego, co jest ogólną cechą impulsów krtaniowych, przy czym różnią się stromością zboczy, co wynika z faktu, że, zgodnie ze wzorem /13/, przy stałej wartości ciśnienie podgłośniowego P_g , $U_g \dot{=} A_g^3$ przy małych wartościach powierzchni A_g , natomiast $U_g \dot{=} A_g$ przy dużych wartościach A_g .

/b/ Zwiększenie wysiłku głosowego, t.j. intensywności fonacji, objawia się, poza oczywistym zwiększeniem amplitud obu przebiegów, skróceniem czasu otwarcia głośni τ_o , a więc zmniejsz-

szaniem wartości współczynnika wypełnienia okresu tonu krztaniowego $\delta = \tau_0/T_0$, od $\delta \approx 1.0$ przy $P_g = 4 \text{ cm H}_2\text{O}$ do $\delta = 0.5 \div 0.6$ przy $P_g = 24 \text{ cm H}_2\text{O}$.

Na rys.3 przedstawiono aproksymację impulsów krztaniowych do postaci trójkątów o określonych parametrach geometrycznych kształtu, wyrażonych przez ich amplitudę a , czasy narastania τ_1 i zaniku $\tau_2 = k \tau_1$, długość $\tau_0 = \tau_1 + \tau_2$ i współczynnik wypełnienia okresu tonu krztaniowego T_0 : $\delta = \tau_0/T_0$.



Rys.3. Aproksymacja impulsów krztaniowych do postaci trójkątów

Transformacja Laplace'a takiego przebiegu $\mathcal{L} = \left\{ \begin{matrix} f(t) \\ -s\tau_1 \\ -s\tau_0 \end{matrix} \right\}$ ma postać

$$\mathcal{L} \{ f(t) \} = F(s) = \frac{a}{s^2} \left[\frac{1}{\tau_1} - \left(\frac{1}{\tau_1} + \frac{1}{\tau_2} \right) \cdot e^{-s\tau_1} + \frac{1}{\tau_2} \cdot e^{-s\tau_0} \right] \quad /14/$$

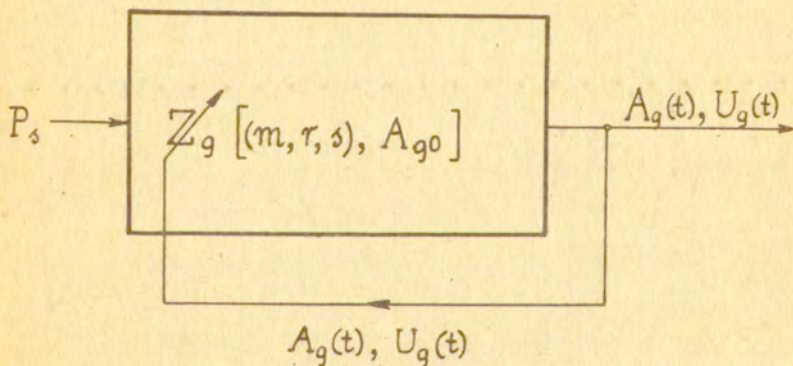
a jej zera na płaszczyźnie częstotliwości zespolonej $s = \sigma + j\omega$ wyznaczają minima obwiedni widma impulsu. Bliższe szczegóły dotyczące rozkładu zer funkcji /14/ można znaleźć w pracach [20, 21].

W szczególnym przypadku impulsu symetrycznego, gdy $\tau_1 = \tau_2$, minima obwiedni występują przy częstotliwościach harmonicznych, odpowiadających parzystym wielokrotnościom długości impulsu, t.j. $f_x = 2n/\tau_0$, gdzie $n = 1, 2, 3, \dots$

Wspólną cechą trójkątnych impulsów krtaniowych jest średnie nachylenie obwiedni ich widma proporcjonalne do f^{-2} , czyli równe -12 dB/oktawę. Ten fakt pozwala określić postać analityczną funkcji wzbudzającej źródła krtaniowego $U_g(s)$ we wzorze ogólnym /1/, której aproksymacją jest dyskretne /prążkowe/ widmo harmoniczne o częstotliwości podstawowej F_0 i obwiedni opadającej ze stromością -12 dB/oktawę.

5. Fizyczne modele źródła krtaniowego

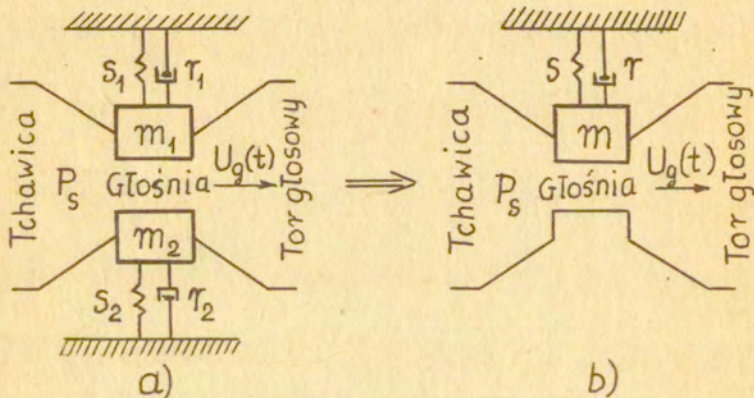
Z fizycznej zasady działania źródła krtaniowego wynika, że jest ono relaksacyjnym układem generacyjnym o sprzężeniu zwrotnym, zasilanym stałym ciśnieniem podgłośniowym P_g , którego produkt wyjściowy w postaci funkcji powierzchni otworu głośni $A_g(t)$ i funkcji przepływu, t.j. prędkości objętościowej powietrza $U_g(t)$ oddziałuje zwrotnie na parametry wewnętrzne układu, określające impedancję akustyczną głośni, jakto przedstawiono na rys.4.



Rys.4. Schemat blokowy generatora krtaniowego w układzie o sprzężeniu zwrotnym

Istnieje kilka sposobów realizacji technicznej takiego układu, z możliwością jego symulacji komputerowej. Najprostszym rozwiązaniem, którego autorami są FLANAGAN i LANDGRAF [8], jest model mechaniczny pokazany na rys.5. W modelu tym fałdy głosowe są przedstawione w postaci prostego mechanicznego układu drga-

jącego o stałych skupionych i jednym stopniu swobody, złożonego z dwóch nieodkształcalnych mas m_1 i m_2 o sprężystościach zawieszzeń s_1 i s_2 oraz wiskotycznych rezystancjach strat r_1 i r_2 , reprezentujących odpowiednio lewy i prawy fałd głosowy, które drgają przeciwsośnie i synfazowo ruchem prostoliniowym, przesu- wając się ku sobie i od siebie, i zmieniają w ten sposób powie- rzchnię otworu głośni $A_g(t)$, a w konsekwencji jej impedancję akustyczną $Z_g(t) = f[A_g(t)]$, patrz rys.5a. Ze względu na założo- ną symetrię obu układów drgających, w realizacji technicznej lub symulacji komputerowej obie masy można sprowadzić do pojedyncze- go układu drgającego, tzw. modelu jednomasowego, o parametrach wynikowych m , s , r równych sumom odpowiednich parametrów częst- kowych, patrz rys.5b.

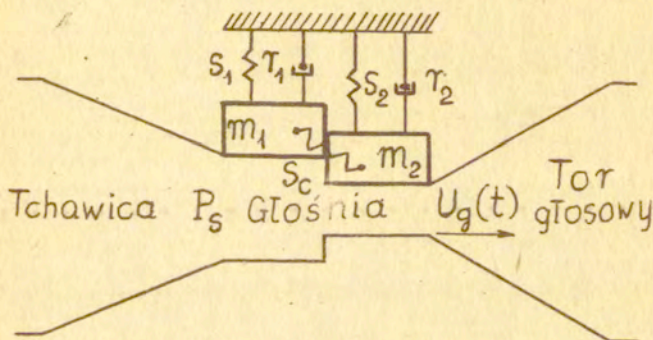


Rys.5. Uproszczony schemat zastępczy jednomasowego modelu źródła krtaniowego

Wadą tego stosunkowo prostego modelu jest jednak fakt, że nie uwzględnia on istotnej cechy fizjologicznej naturalnego źródła krtaniowego, jaką jest odkształcalność postaciowa fałdów głosowych, objawiająca się asymetrią drgań i przesunięciami fazowymi między wychyleniami ich dolnych i górnych krawędzi w ruchu drgającym prostym.

ISHIZAKA [17] podjął próbę opracowania skrajnie rozbudowa- nego modelu o stałych rozłożonych i wielu stopniach swobody,

złożonego z określonej liczby elementarnych mas, sprężystości i rezystancji strat. Tak skomplikowane rozwiązanie nie okazało się jednak praktycznie przydatne nie tylko do celów fonetyki akustycznej i syntezy mowy, ale nawet do diagnostyki foniatrycznej, gdyż najważniejsze cechy fizjologiczne źródła krztaniowego, istotne przy modelowaniu procesu fonacji, można z dostateczną dokładnością odtworzyć już za pomocą modelu dwumasowego, patrz np. ISHIZAKA i FLANAGAN [18]. W modelu tym /patrz rys.6/, każdy fałd głosowy, lewy i prawy, przedstawiony jest w postaci dwóch mas \underline{m}_1 i \underline{m}_2 , które reprezentują odpowiednio ich części dolne i górne. Masy te są połączone sprężystością sprzęgającą \underline{s}_c , która symbolizuje sztywność fałdów głosowych na zginanie w płaszczyźnie pionowej, równoległej do kierunku ich drgań.



Rys.6. Uproszczony schemat zastępczy dwumasowego modelu źródła krztaniowego

Znaczenie fizyczne parametrów modelu jest następujące:

/a/ Podział masy \underline{m} na dwie masy cząstkowe \underline{m}_1 i \underline{m}_2 symuluje odkształcalność postaciową fałdów głosowych.

/b/ Sprężystości zawiesznień \underline{s}_1 i \underline{s}_2 reprezentują sztywności mechaniczne obu części fałdu głosowego, przeciwstawiające się :

- chwilowym wychyleniom z położenia spoczynkowego,
- odkształceniom lepko-sprężystym w fazie zwarcia, kiedy fałdy przeciwległe stykają się ze sobą, ulegając chwilowej i odwracalnej deformacji.

/c/ Rezystancje strat \underline{r}_1 i \underline{r}_2 pochodzą od:

- tarcia wiskotycznego tworzywa /tkanki mięśniowej i łącznej/
- zjawiska przyczepności wzajemnej /lepkości/ fałdów głosowych, stykających się ze sobą w fazie zwarcia podczas ich ruchu ku sobie i od siebie.

Stosunkowo duża dowolność w wyborze wartości parametrów \underline{m} , \underline{g} , \underline{r} fałdów głosowych pozwala nie tylko na jakościowe, ale i ilościowe modelowanie czynności fonacyjnych krtani zarówno w stanach fizjologicznie normalnych, jak i patologicznych, co ma szczególne znaczenie w akustycznych metodach diagnostyki foniatrycznej, opartych na pomiarze przebiegu i analizie rozkładów statystycznych częstotliwości podstawowej F_0 lub okresu T_0 tonu krtaniowego, patrz np. [11, 12].

Parametry masy \underline{m} i sztywności \underline{g} fałdów głosowych określają częstotliwość ich drgań własnych, t.j. częstotliwość podstawową F_0 tonu krtaniowego, według wzoru

$$F_0 = \frac{1}{2\pi} \sqrt{\frac{\underline{g}}{\underline{m}}} \quad /15/$$

W rzeczywistych warunkach fonacji, pod wpływem skurczu mięśni głosowych zmienia się naciąg /napięcie/ fałdów głosowych, które ulegają odpowiednio skróceniu lub wydłużeniu. Wskutek tego zmieniają się, w odwrotnych stosunkach, ich masa \underline{m} i sztywność \underline{g} , co powoduje zmianę częstotliwości tonu krtaniowego od wartości F_0 do F'_0

$$F'_0 = \frac{1}{2\pi} \sqrt{\frac{\underline{g} \cdot Q}{\underline{m}/Q}} = Q \cdot F_0 \quad /16/$$

gdzie Q jest parametrem pomocniczym, stosowanym w badaniach modelowych przy kształtowaniu charakterystyki intonacyjnej konkretnej frazy lub segmentu sygnału mowy.

6. Funkcja transmitancji toru głosowego

6.1. Tor głosowy jako układ akustyczny.

Funkcja transmitancji toru głosowego $H(s) = U_o(s)/U_g(s)$ opisuje cechy segmentalne, przede wszystkim strukturę widmowoformantową dźwięków mowy, która w głównej mierze określa ich znaczenie fonetyczne. Wyznaczenie postaci analitycznej tej funkcji i jej aproksymacja w odpowiednim układzie fizycznym stanowi

zatem zagadnienie centralne przy modelowaniu czynności głosotwórczych organu mowy. Ze względu na bardzo złożoną strukturę anatomiczną toru głosowego, określoną przez wzajemną konfigurację geometryczną efektorów artykulacyjnych, zadanie to można rozwiązać jedynie na drodze daleko idących, ale jednocześnie racjonalnie uzasadnionych uproszczeń, i przedstawienie toru głosowego jako układu akustycznego o wyznaczalnych parametrach fizycznych.

Najbardziej ogólne rozwiązanie, polegające na potraktowaniu toru głosowego jako bezstratnej rury o przekroju kołowym, opisanym ciągłą funkcją powierzchni $A(x)$ i zastosowanie do takiego modelu równania falowego /WEBSTER [37] / o postaci

$$\frac{1}{A(x)} \cdot \frac{\partial}{\partial x} \left[A(x) \frac{\partial p}{\partial x} \right] = \frac{1}{c^2} \cdot \frac{\partial^2 p}{\partial t^2} \quad /17/$$

okazało się zbyt skomplikowane analitycznie i, poza nielicznymi wyjątkami /patrz np. [2, 14, 35] / nie znalazło zastosowań praktycznych do celów modelowania.

6.2. Aproksymacja toru głosowego w postaci układu rezonansowego o stałych skupionych.

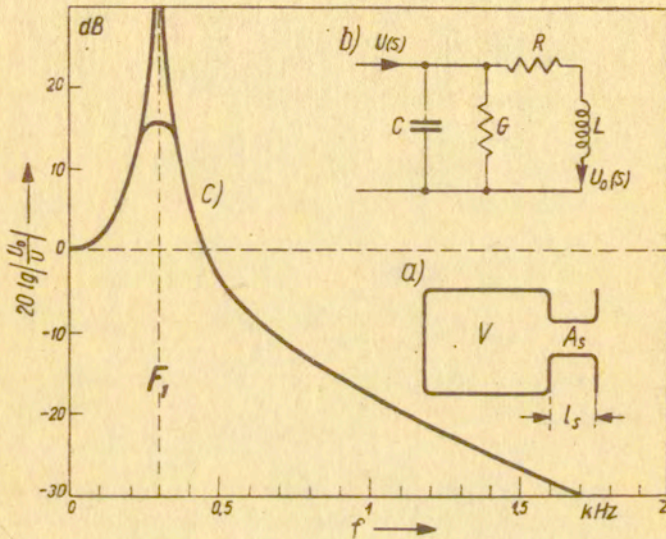
Najbardziej radykalnym uproszczeniem jest potraktowanie toru głosowego jako układu o stałych skupionych, którego rozmiary są mniejsze od długości fali dźwiękowej λ , co jest słuszne jedynie w ograniczonym od góry zakresie częstotliwości mowy $f < 1000$ Hz. W najprostszym przypadku toru głosowego gardłowo-ustnego modelem takim może być klasyczny rezonator Helmholtza, przedstawiony na rys.7 wraz z jego elektrycznym układem zastępczym, w którym pojemność $C = V/\rho c^2$ odpowiada akustycznej podatności wnęki rezonatora, indukcyjność $L = \rho l_g/A_g$ - masie akustycznej powietrza w szyjce przy pominięciu efektów brzegowych i promieniowania, a rezystancja R i konduktancja G uwzględniają straty układu. Transmitancja prądowa układu jest klasyczną charakterystyką rezonansu równoległego.

Przeciętna objętość jamy gardłowo-ustnej mężczyzny $V \approx 75$ cm³. Przyjmując orientacyjnie długość otworu warg $l_g = 3$ cm i jego przekrój $A_g = 1.0$ cm², co w przybliżeniu określa geometryczną konfigurację układu warg przy wymawianiu samogłoski /u/, otrzymuje się, przy pominięciu strat, częstotliwość rezonansową mo-

delu

$$F_1 = \frac{1}{2\pi\sqrt{LC}} = \frac{c}{2\pi} \sqrt{\frac{A_E}{l_E V}} \approx 300 \text{ Hz} \quad /18/$$

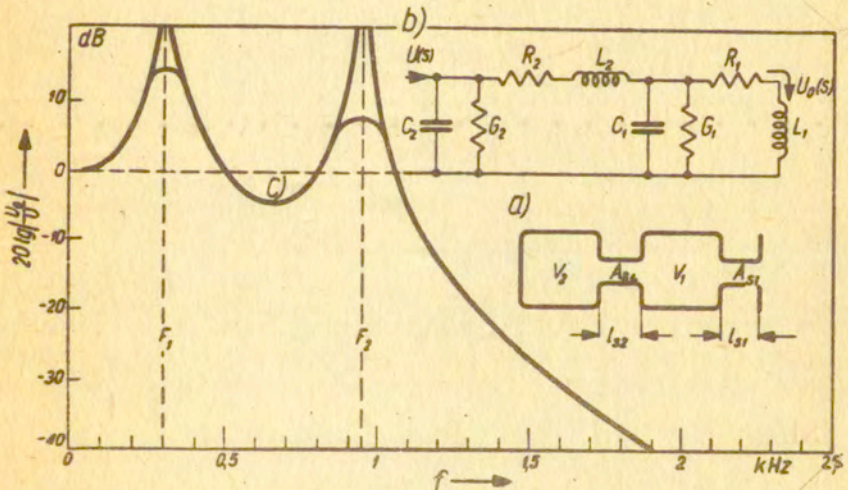
która odpowiada przeciętnej częstotliwości pierwszego formantu F_1 samogłoski /u/ wymawianej głosem męskim.



Rys.7. Akustyczny model toru głosowego w postaci pojedynczego rezonatora Helmholtza

Taka aproksymacja funkcji transmitancji $H(\omega)$ jest jednak uproszczeniem zbyt daleko idącym, choćby z tego powodu, że uwzględnia tylko jeden zakres formantowy samogłoski, co nie wystarcza do jej identyfikacji fonetycznej. Bardziej realistycznym rozwiązaniem jest model dwurezonansowy, złożony z dwóch sprzężonych ze sobą rezonatorów Helmholtza, przedstawiony na rys.8, który jest słuszny w przypadku pewnych dźwięków mowy, zwłaszcza samogłoskowych, kiedy tor głosowy jest wyraźnie podzielony położeniem języka na dwie wnęki: przednią V_1 , zakończoną otworem wylotowym warg, i tylną V_2 , sprzężoną z przednią za pośrednictwem przewężenia między grzbietem języka i podniebieniem. Funkcja transmitancji takiego układu ma dwa wierzchołki rezonansowe

F_1 i F_2 , które odpowiadają częstotliwościom własnym F_{01} i F_{02} obu rezonatorów traktowanych oddzielnie jedynie wtedy, gdy ich współczynnik sprzężenia akustycznego jest pomijalnie mały. W miarę wzrostu sprzężenia częstotliwość rezonansowa F_1 ulega obniżeniu, a F_2 podwyższeniu, z tym, że teoretycznie spełniony jest warunek $F_1 \cdot F_2 = F_{01} \cdot F_{02} = \text{const}$. Przy stałej wartości współczynnika sprzężenia obu rezonatorów, częstotliwości F_1 i F_2 , określające wierzchołki funkcji transmitancji $H(\omega)$, t.j. formanty F_1 i F_2 modelu, zależą w pierwszym przybliżeniu od stosunku $\frac{k_c}{k} = V_1/V_2 = C_1/C_2$, którego wzrost powoduje malenie stosunku F_2/F_1 . Jest to zgodne ze znanym w fonetyce akustycznej stwierdzeniem, że uszeregowaniu samogłosek podstawowych w kolejności od najbardziej przedniej do najbardziej tylnej, to jest /i, y, e, a, o, u/, towarzyszy konsekwentne malenie częstotliwości drugiego formantu F_2 , jak również, z wyjątkiem samogłosek /o, u/, stosunku F_2/F_1 , por. JASSEM [19], s.193.



Rys.8. Akustyczny model toru głosowego w postaci dwóch sprzężonych rezonatorów Holmholtza

Ogólnie biorąc jednak, przedstawienie toru głosowego w postaci zespołu n układów rezonansowych o stałych skupionych jest zbyt prymitywne do celów modelowania jego struktury akustycznej

w rzeczywistych warunkach artykulacji. Wynika to przede wszystkim, choć nie tylko, z oczywistego faktu, że utożsamianie cech struktury widmowo-formantowej poszczególnych głosek z częstotliwościami rezonansowymi określonych wyodrębnionych obszarów toru głosowego jest sztuczne i nieuzasadnione zarówno z akustycznego, jak i artykulacyjnego punktu widzenia.

Dla ścisłości warto jednak wspomnieć, że metoda prezentacji cech widmowych /formantowych/ głosek za pomocą odpowiednio zrealizowanego zespołu dyskretnych, odseparowanych od siebie i niezależnie przestrajanych elektrycznych układów rezonansowych analogowych lub cyfrowych, znalazła zastosowanie praktyczne w widmowo-parametrycznych synteźatorach mowy, których zadaniem jest symulacja sygnału mowy jako finalnego produktu artykulacji, a nie czynności głosotwórczych i struktury artykulacyjnej organu mowy, patrz np: [22].

6.3. Aproksymacja toru głosowego w postaci akustycznego układu transmisyjnego o stałych rozłożonych.

Pierwszym krokiem zmierzającym do modelowania toru głosowego z uwzględnieniem jego konfiguracji artykulacyjnej i wyznaczenia funkcji transmitancji $H(\omega)$ jest przedstawienie go w postaci kilku, w zasadzie nie więcej niż czterech, rur jednorodnych o różnych długościach i przekrojach, z których każda odtwarza pewien odcinek toru głosowego o stałym przekroju i z punktu widzenia analogii elektro-akustycznych może być traktowana jako odcinek elektrycznej linii długiej o określonych i fizycznie wyznaczalnych parametrach jednostkowych, rozłożonych w sposób ciągły. Takie podejście do modelowania ma zasadnicze zalety, gdyż:

/a/ Bardziej realistycznie, w porównaniu z układami o stałych skupionych, odtwarza akustyczną strukturę toru głosowego i pozwala wiązać ją z fonetycznymi cechami dźwięków mowy w określonych warunkach artykulacji.

/b/ Ze względu na stosunkowo proste i łatwe do stosowania w tym przypadku metody analizy układu, analogiczne do używanych w teletransmisji i opartych głównie na rozwiązywaniu równań trygonometrycznych, zarówno metodami graficznymi, jak i numerycznymi, pozwala na wyznaczenie funkcji transmitancji toru głosowego.

wego jako całości, a zwłaszcza rozkładu jej biegunów i zer, t.j. struktury formantowo-antyformantowej.

/c/ Nie wprowadza w zasadzie ograniczeń częstotliwościowych, z wyjątkiem promienia \underline{r} poprzecznego przekroju rury $\underline{r} < \lambda/4$, fala płaska/.

Skrajnym uproszczeniem jest przedstawienie toru głosowego w postaci jednorodnej rury bezstratnej o długości $L = 17$ cm, zamkniętej na początku $Z_g = \infty$ / i otwartej na końcu $Z_p = 0$, symulującej konfigurację geometryczną toru przy artykulacji samogłoski neutralnej [a] . Rura taka jest układem wielorezonansowym, którego kolejne częstotliwości rezonansowe $F_n = (2n-1) \frac{c}{4 \cdot L}$, gdzie $n = 1, 2, 3, \dots$, odpowiadają nieparzystym wielokrotnościom pierwszego rezonansu ćwierćfalowego i określają częstotliwości formantów tej samogłoski $F_n = 500; 1500; 2500; 3500$ itd. Hz. Mimo swojej prostoty, a może właśnie z tego powodu, tak skrajnie uproszczony model jest często stosowany jako układ odniesienia przy teoretycznym i eksperymentalnym badaniu wpływu różnych czynników, zwłaszcza warunków brzegowych, np. impedancji źródła krtaniowego, impedancji promieniowania ust, strat w ściankach toru głosowego itd., na postać funkcji transmitancji toru $H(s)$ modeli bardziej złożonych i skomplikowanych.¹⁾

O przydatności posługiwania się do celów badawczych modelami złożonymi z kilku, w konkretnym przypadku trzech, rur jednorodnych o różnych długościach i przekrojach świadczyć może przeprowadzona w ZAC IPPT-PAN analiza wpływu wymuszonej nazalizacji wskutek rozszczepu podniebienia na właściwości transmisyjne /strukturę formantową/ samogłosek ustnych [26]. Weryfikację doświadczalną analizy teoretycznej przeprowadzono na modelu analogowym, odtwarzającym w skali rozmiarów 5:1, t.j. w skali częstotliwości 1:5, konfigurację toru gardłowo-ustno-nosowego. Użytkano dobrą zgodność wyników analizy teoretycznej z wynikami pomiarów na modelu i - co jest istotne - również z wynikami pomiarów fonospektroskopowych, przeprowadzonych na pacjentach w stanach patologicznych i pooperacyjnych rozszczepu podniebienia [34].

¹⁾ W Zakładzie Akustyki Cybernetycznej IPPT-PAN model taki, zrealizowany w symulacji komputerowej, był użyty do wyznaczenia optymalnego stopnia segmentacji wzdłużnej /kwantyzacji przestrzennej/ artykulacyjnego modelu toru głosowego [24,25].

7. Model artykulacyjny toru głosowego

7.1. Wprowadzenie.

Najbardziej uniwersalnym modelem akustycznym toru głosowego jest model konfiguracyjny, który odtwarza czynności głosotwórcze organu mowy w oparciu o jego uproszczoną, ale zbliżoną do rzeczywistej strukturę artykulacyjną. Opracowanie optymalnego i w pewnym sensie znormalizowanego modelu artykulacyjnego jest jeszcze dalekie od stadium finalizacji [36] i musi być poprzedzone pracami badawczymi o charakterze podstawowym, które od kilku lat są prowadzone równoległe w wielu ośrodkach badawczych, przede wszystkim w Szwecji /Department of Speech Communication, Speech Transmission Laboratory, KTH, Sztokholm/, Francji /Laboratoire de la Communication Parlée, E.N.S.E.R.G., Grenoble; Institut de Phonétique, Université de Grenoble/, USA /Acoustics Research Department, Bell Laboratories; Research Laboratory of Electronics, Massachusetts Institute of Technology, M.I.T., Cambridge, Mass./ oraz ZSRR /Instytut Problemów Transmisji Informacji, Akademia Nauk ZSRR, Moskwa/.

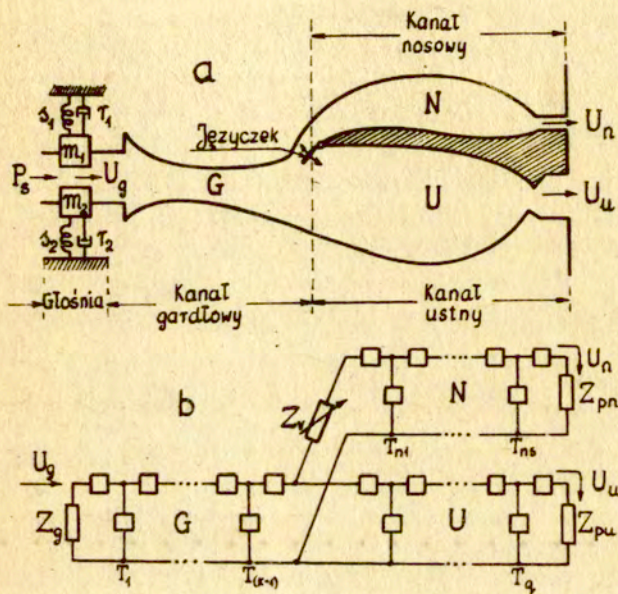
* Obserwowany w ostatnich latach wzrost zainteresowania modelem konfiguracyjnym organu mowy wytkumaczyć można m.in. wielostronnością zastosowań takiego modelu, realizowanego w symulacji komputerowej, który okazuje się przydatny i dogodny nie tylko do badań podstawowych i stosowanych w fonetyce akustycznej, ale również do zastosowań medycznych w diagnostyce foniatrycznej [11,12] oraz programowanej syntezy mowy, patrz np. [9].

Koncepcja i założenia akustyczne modelu artykulacyjnego, opracowanego w Pracowni Akustyki Mowy ZAC IPPT-PAN, nie odbiegają wprawdzie w sposób zasadniczy od innych rozwiązań, ale zmierzają do modelowego ujęcia efektu nazalizacji, co ma istotne znaczenie dla wszystkich wymienionych wyżej dziedzin zastosowań [24,25].

7.2. Założenia metodyczne modelu artykulacyjnego.

Na rys.9 przedstawiony jest model akustyczny toru głosowego gardłowo-ustno-nosowego /G + U + N/. Kanał gardłowo-ustny /G + U/, który stanowi wyłączną drogę przepływu energii akustycznej przy artykulacji ustnej, ma postać rury o nieregularnym

przekroju poprzecznym, określonym ciągłą funkcją powierzchni $A(x)$, gdzie x jest odległością od początku układu współrzędnych, t.j. od otworu głośni.



Rys.9. Model akustyczny toru głosowego $G + U + N$ /a/ i jego elektryczny schemat zastępczy /b/

W przeciętnych warunkach anatomicznych łączna długość kanału $G + U$, t.j. $L_{GU} = 16 \div 19$ cm, a jego przekrój poprzeczny, zależnie od warunków artykulacji, przybiera wartości od części cm^2 do około 15 cm^2 . W punkcie $x = 0$ tor $G + U$ jest pobudzany funkcją źródła krztaniowego U_g , a na końcu obciążony impedancją promieniowania ust Z_{pu} . Mniej więcej w połowie długości tego toru $x \approx 8$ cm/ dołączony jest do niego kanał nosowy N , który również można traktować jako rurę akustyczną o określonym kształcie przekroju i długości około $12 \div 13$ cm, obciążony impedancją promieniowania otworów nosa Z_{pn} . Rolę zaworu, kontrolującego stopień sprzężenia akustycznego obu kanałów, a więc tym samym stopień nazalizacji pierwotnie nienosowych dźwięków

mowy, pełni w przypadkach fizjologicznie normalnych ruchome zakończenie podniebienia miękkiego /tzw. języczek tylny, łac. velum/, a w przypadkach patologicznych - szczelina, wywołana rozszczepem lub ubytkiem podniebienia miękkiego.

W ogólnym przypadku artykulacji ustno-nosowej, ciśnienie akustyczne fali głosowej w punkcie pomiarowym, położonym na osi symetrii układu w odległości r , wyraża się, zgodnie z zasadą superpozycji ciśnień akustycznych, wzorem

$$p(r) = \frac{j\omega\rho}{4\pi r} \cdot (\vec{U}_u + \vec{U}_n) \cdot \exp(-j2\pi r/\lambda) \quad /19/$$

w którym \vec{U}_u i \vec{U}_n oznaczają odpowiednio prędkości objętościowe w otworach ust i nosa. Przy tych założeniach modelowanie funkcji transmitancji toru głosowego w ogólnym przypadku artykulacji ustno-nosowej, a więc przy uwzględnieniu nazalizacji bądź naturalnej, bądź wymuszonej, t.j. wywołanej anomaliami anatomicznymi, sprowadza się do wyznaczenia postaci analitycznych, w formie dogodnej do symulacji komputerowej, dwóch funkcji transmitancji:

/a/ $H_u(\omega) = U_u(\omega) / U_g(\omega)$ - toru gardłowo-ustnego z uwzględnieniem wpływu bocznikującej ten tor impedancji wejściowej Z_n kanału nosowego,

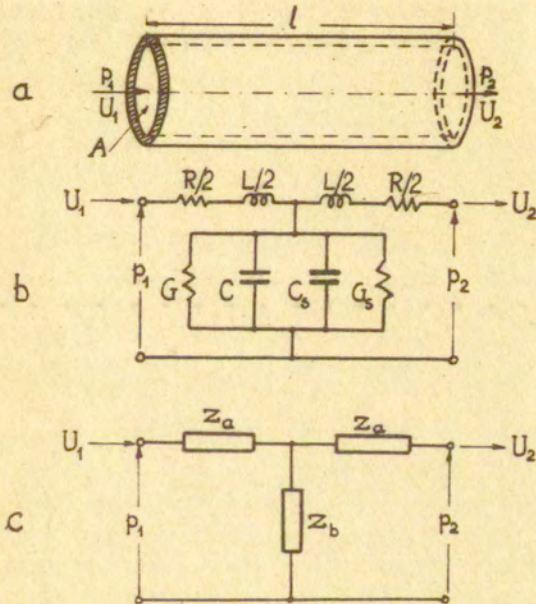
/b/ $H_n(\omega) = U_n(\omega) / U_g(\omega)$ - toru gardłowo-nosowego z uwzględnieniem wpływu bocznikującej ten tor impedancji wejściowej Z_u kanału ustnego.

7.3. Funkcja transmitancji toru gardłowo-ustnego.

Ogólną i najczęściej stosowaną zasadą konfiguracyjnego modelowania toru głosowego jest dyskretyzacja ciągłej funkcji powierzchni przekroju $A(x)$ przez jej sprowadzenie do postaci skokowej i aproksymowanie za pomocą skończonej liczby q jednorodnych segmentów elementarnych, z których każdy reprezentuje odcinek toru o długości $l = L_{GU}/q = \text{const.}$ w postaci rury cylindrycznej o przekroju A_i , gdzie $i = 1, 2, 3, \dots, q$. Dokładność aproksymacji jest tym większa, im mniejsza jest długość l segmentu elementarnego, przy czym warunkiem granicznym jest $l \leq \frac{1}{8} \lambda_{\text{min}}$, co w warunkach rzeczywistych, gdy $L_{GU \text{ max}} = 20 \text{ cm}$, $f < 5000 \text{ Hz}$, sprowadza się do warunku $q = 20$, $l = 1 \text{ cm}$. Warunek ten został zweryfikowany doświadczalnie na przykładzie

segmentacji toru głosowego w postaci jednorodnej rury cylindrycznej o długości $L = 17.5$ cm i stałym przekroju $A(x) = \text{const.}$, symulowanej komputerowo i odpowiadającej artykulacji samogłoski neutralnej E .

Zgodnie z teorią układów akustycznych, schemat zastępczy elementarnego segmentu toru głosowego, t.j. odcinka jednorodnej rury cylindrycznej o długości l i powierzchni przekroju poprzecznego A /rys.10/ można przedstawić w postaci czwórnika np. typu T .



Rys.10. Elementarny segment toru głosowego w postaci rury cylindrycznej /a/ oraz jego czwórnikowe schematy zastępcze /b/ i /c/

Akustyczne impedancje Z_a ramion wzdłużnych i Z_b ramienia poprzecznego czwórnika T wyrażone są wzorami /20/ i /21/, dobrze znanymi z teorii układów akustycznych:

$$Z_a = \frac{1}{2} [R + j\omega L] \quad /20/$$

$$Z_b = \left[(G + G_g) + j\omega (C + C_g) \right] \quad /21/$$

w których wielkości L , C , R , G , mają określone znaczenia fizyczne jako klasyczne parametry akustyczne rury stratnej o ściankach sztywnych, mianowicie:

- L - masa akustyczna powietrza w rurze,
- C - podatność akustyczna powietrza w rurze,
- R - rezystancja akustyczna strat na tarcie wiskotyczne,
- G - akustyczna konduktancja strat na przewodnictwo cieplne.

Parametry C_g i G_g uwzględniają specyfikę struktury materiałowej tworzywa /tkanki mięśniowej i łącznej/ ścian kanału głosowego, które nie są idealnie sztywne, jak to się zakłada w rozważaniach uproszczonych, ale mają skończoną impedancję akustyczną powierzchniową o charakterze bezwładnościowym

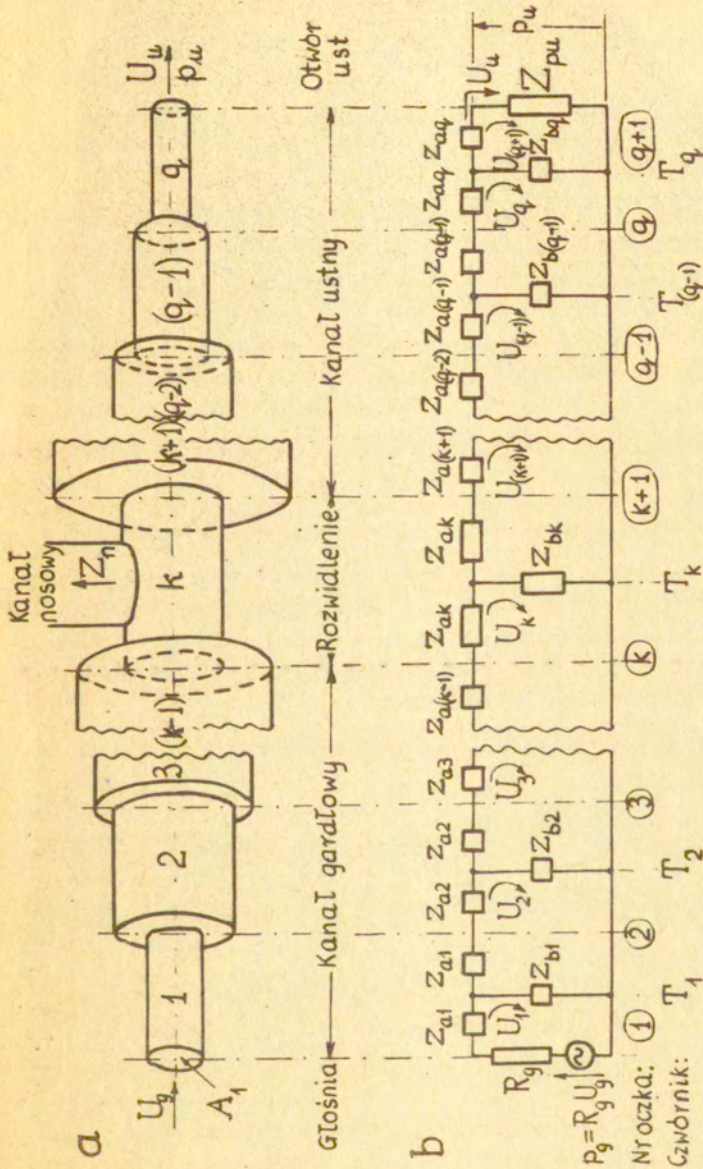
$$z_g = r_g + j\omega m_g \quad /22/$$

i, poddane działaniu lokalnego ciśnienia akustycznego w kanale głosowym, zostają wprowadzone w drgania wymuszone, powodując dodatkowe straty energii. Elementy C_g , G_g , symbolizujące te straty, mają następujące znaczenie fizyczne:

- C_g - podatność akustyczna, równoważna odwrotności masy akustycznej ścian kanału,
- G_g - odwrotność akustycznej rezystancji, t.j. konduktancja akustyczna ścian kanału.

Zagadnieniu wyznaczania wartości liczbowych parametrów jednostkowych m_g i r_g tworzywa ścian kanału głosowego poświęcono sporo uwagi i badań w pracach badawczych kilku ośrodków, przyjmując za punkt wyjścia wyniki pomiarów bezpośrednich, przeprowadzonych na powierzchni ciała ludzkiego /patrz np. FRANKE [10], a następnie weryfikując je metodami fonetyczno-akustycznymi /FANT i in. [4,5], SONDHI [32], MRAYATI i GUERIN [30]/ w rzeczywistych warunkach artykulacji.

Na rys.11 przedstawiono schemat strukturalny modelu toru gardłowo-ustnego i jego elektryczny schemat zastępczy w postaci kaskadowego połączenia q czwórników typu \underline{T} , z których każdy reprezentuje kolejny segment elementarny toru.



Rys. 11. Model akustyczny toru głosowego gardłowo-ustnego z uwzględnieniem boczniującego wpływu kanału nosowego /a/ i jego elektryczny schemat zastępczy /b/

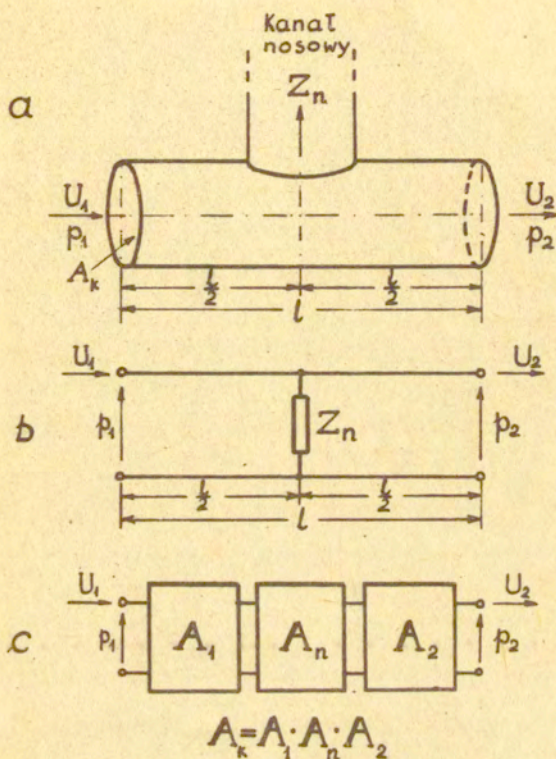
Parametrami zmiennymi, które opisują konfigurację geometryczną toru G + U w konkretnych warunkach artykulacji są: liczba segmentów q określająca długość toru oraz przekroje A_1 poszczególnych segmentów, odtwarzające w postaci dyskretnej ciągłą funkcję powierzchni $A(x)$. Układ jest zasilany z generatora tonu krtaniowego o impedancji wewnętrznej R_g , a obciążony impedancją promieniowania ust Z_{pu} .

Na uwagę zasługuje segment k -ty i odpowiadający mu czwórnik T_k , który reprezentuje rozgałęzienie toru głosowego /med.: nasopharynx/, t.j. obszar, w którym impedancja wejściowa Z_n kanału nosowego, w przypadku nazalizacji naturalnej lub wymuszonej, boczkuje tor G + U, stanowiący wyłączną drogę przepływu energii akustycznej przy artykulacji nienosowej. Segment ten różni się swoją strukturą od pozostałych segmentów modelu i może być przedstawiony, jak to pokazano na rys.12, w postaci iloczynu trzech macierzy łańcuchowych $A_k = A_1 \cdot A_n \cdot A_2$, z którego można wyznaczyć wartości impedancji wzdłużnych $Z_{a(k)}$ i poprzecznej $Z_{b(k)}$ czwornika T_k . Parametrem zmiennym jest w tym przypadku impedancja wejściowa Z_n kanału nosowego, która jest ilościową miarą stopnia nazalizacji. Przy braku nazalizacji, t.j. w warunkach artykulacji wyłącznie ustnej, $Z_n = \infty$ i czwórnik T_k sprowadza się do postaci, opisującej typowy segment cylindryczny.

Funkcję transmitancji toru G + U, wyrażoną stosunkiem prędkości objętościowej $U_u = U_{(q+1)}$ w otworze ust, t.j. w ostatnim segmencie układu zastępczego, do prędkości objętościowej $U_g = U_1$ źródła krtaniowego w pierwszym segmencie układu, można wyznaczyć dowolną z klasycznych metod analizy układów elektrycznych, np. dogodną do obliczeń numerycznych metodą wyznacznikową. W tym przypadku

$$H_u(\omega) = \frac{U_u}{U_g} = \frac{U_{(q+1)}}{U_1} = \frac{\Delta_{1(q+1)}}{\Delta_{11}} \quad /23/$$

gdzie $\Delta_{1(q+1)}$ oraz Δ_{11} są podwyznacznikami /minorami/ elementów $z_{1(q+1)}$ oraz z_{11} macierzy impedancyjnej Z_{GU} układu równań Kirchhoffa dla $(q + 1)$ oczek układu zastępczego.



Rys.12. Model akustyczny rozgałęzienia toru głosowego /a/ i jego czwórnikowe schematy zastępcze /b/ oraz /c/

Do celów testowania i weryfikacji założeń strukturalnych modelu niezbędna jest znajomość danych liczbowych, określających funkcję powierzchni $A(x)$ toru gardłowo-ustnego w rzeczywistych i różnych warunkach artykulacji. Wobec braku reprezentatywnych danych, wyznaczających postać tej funkcji choćby w najprostszym przypadku samogłosek polskich, zastosowano metodę pośrednią. Wykorzystując dane liczbowe, dotyczące konfiguracji toru głosowego przy artykulacji 12 samogłosek francuskich [30] i wybierając z nich te, które w artykulacyjnym czworoboku samogłoskowym zbliżone są do sześciu polskich samogłosek podsta-

wowych, drogą kolejnych prób i przybliżeń wyznaczono dyskretne wartości hipotetycznych funkcji powierzchni $A(x)$ toru głosowego w przeciętnych warunkach ich artykulacji głosem męskim. Jako kryterium dokładności aproksymacji przyjęto zgodność częstotliwości biegunów funkcji transmitancji $H_u(\omega)$ toru głosowego ze średnimi wartościami formantów $F1 \div F4$ dla głosów męskich, które podał JASSEM [19].

7.4. Funkcja transmitancji toru gardłowo-nosowego.

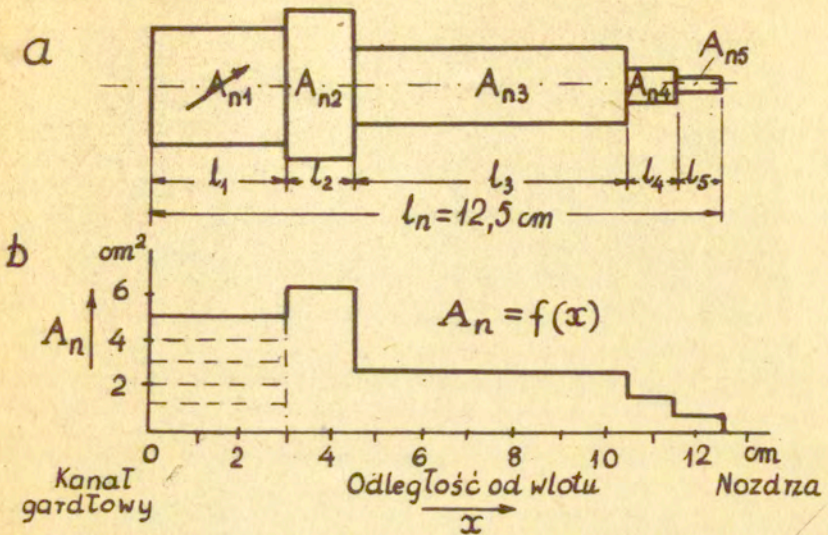
Metoda postępowania przy wyznaczaniu funkcji transmitancji toru gardłowo-nosowego $G + N$ jest analogiczna jak dla toru $G + U$. Ponieważ jednak struktura anatomiczna kanału nosowego jest stosunkowo prosta i - co jest istotne - nie zależy od warunków artykulacji, wykazując jedynie zróżnicowania osobnicze, można zastosować daleko idące uproszczenia jego funkcji powierzchni $A_n(x)$, jak to z reguły czyniono w analogowych syntezatorach mowy, konstruowanych w latach 50-tych i 60-tych.

Wykorzystując dane dostępne w literaturze [13, 15, 16, 33], w obecnym rozwiązaniu zaproponowano, przynajmniej do celów badań pilotażowych, aproksymację kanału nosowego w postaci pięciu segmentów cylindrycznych o różnych długościach, przy czym tylko pierwszy z nich ma zmienną powierzchnię przekroju A_{n1} , która jest ilościową miarą stopnia nazalizacji, patrz rys.13. Impedancję wejściową Z_n kanału nosowego, która jest parametrem funkcji transmitancji /23/ toru gardłowo-ustnego, określa wzór

$$Z_n = \frac{\Delta_N}{\Delta_{11}} \quad /24/$$

w którym Δ_N jest wyznacznikiem charakterystycznym, a Δ_{11} - podwyznacznikiem elementu z_{11} macierzy impedancyjnej Z_N kanału nosowego, opisanego układem równań Kirchhoffa dla $n = 6$ oczek układu zastępczego.

Czwórnikowy schemat zastępczy toru gardłowo-nosowego $G + N$ jest konkatenacją $(k + 5)$ czwórników elementarnych, przy czym czwórniki $T_1 \div T_{(k-1)}$ reprezentują strefę gardłową G toru i mają to samo znaczenie, jak w przypadku toru gardłowo-ustnego $G + U$ na rys.11, natomiast czwórniki $T_{(k+1)} \div T_{(k+5)}$ symulują tor nosowy N , zgodnie z jego konfiguracją geometryczną, patrz rys.13.



Rys. 13. Model akustyczny kanału nosowego /a/ i funkcja powierzchni jego przekroju $A_n(x)$ /b/

Czwórnik T_k^* , reprezentujący rozwidlenie toru głosowego, jest strukturalnym odpowiednikiem czwórnika T_k , symulującego to rozwidlenie w przypadku toru gardłowo-ustnego $G + U$ z nazalizacją, z tą jedynie różnicą, że impedancją bocznikującą w tym punkcie tor gardłowo-nosowy jest teraz impedancja wejściowa Z_u kanału ustnego, reprezentowanego przez czwórniki $T_{(k+1)} \hat{=} T_q$ na schemacie zastępczym toru $G + U$ na rys. 11. W pełnej analogii do wzoru /24/, impedancja wejściowa Z_u kanału ustnego wyraża się wzorem

$$Z_u = \frac{\Delta_U}{\Delta_{(k+1)(k+1)}} \quad /25/$$

gdzie Δ_U jest wyznacznikiem charakterystycznym, a $\Delta_{(k+1)(k+1)}$ - podwyznacznikiem elementu $\underline{z}_{(k+1)(k+1)}$ macierzy impedancyjnej Z_U kanału ustnego w określonych warunkach artykulacji.

Przy artykulacji wyłącznie nosowej, jak to ma miejsce w przypadku spółgłosek nosowych, kanał ustny jest zamknięty na końcu i impedancja promieniowania ust $Z_{pu} = \infty$. Parametrem

zmiennym jest wtedy nie tylko kształt funkcji powierzchni $A(x)$ wnętrza kanału ustnego, ale i jej długość, zależna od miejsca zwarcia, która różni np. dwuwargową głoskę /m/ od językowo-zębowego /n/ lub językowo-przednio-podniebiennego /ń/.

Funkcja transmitancji toru gardłowo-nosowego $G + N$, wyrażona stosunkiem prędkości objętościowej U_n w otworze nosa do prędkości objętościowej U_g w otworze głośni, przy uwzględnieniu bo-
cznikującego wpływu impedancji wejściowej Z_u kanału ustnego, wyraża się wzorem

$$H_n(\omega) = \frac{U_n}{U_g} = \frac{\Delta_{1(k+6)}}{\Delta_{11}} \quad /26/$$

w którym Δ_{11} i $\Delta_{1(k+6)}$ są podwyznacznikami elementów z_{11} oraz $z_{1(k+6)}$ macierzy impedancyjnej Z_{GU} stopnia $n = (k + 6)$ toru gardłowo-nosowego.

8. Podsumowanie

Przedmiotem, a zarazem celem pracy było syntetyczne przedstawienie aktualnego stanu badań nad modelowaniem czynności głosotwórczych, t.j. fonacyjnych i artykulacyjnych, organu mowy, których produktem finalnym jest akustyczny sygnał mowy, traktowany jako nośnik informacji, określających zarówno treść fonetyczną i lingwistyczną przekazywanej nim wiadomości, jak i inwariantne cechy osobnicze głosu jej nadawcy.

Zgodnie z fizyczną interpretacją procesu wytwarzania mowy, ciśnienie akustyczne fali głosowej wyrażono jako iloczyn trzech czynników: funkcji wzbudzającej źródła krtaniowego, funkcji transmitancji toru głosowego oraz impedancji promieniowania ust lub/i nosa. Każdy z tych czynników przedstawiono w postaci uproszczonego opisu matematycznego, który daje podstawy teoretyczne tworzenia fizycznych modeli odpowiednich członów funkcjonalnych złożonego układu biocybernetycznego, jakim jest organ mowy. Praca ma charakter przeglądu, uwzględniający m.in. własne osiągnięcia badawcze Zakładu Akustyki Cybernetycznej IPPT-PAN w tej dziedzinie. Cytowane pozycje bibliograficzne kierują czytelnika do rozwiązań szczegółowych, opisanych w krajowej i zagranicznej literaturze przedmiotu.

Bibliografia

1. BERG van den, J.W., ZANTEMA, J.T., DOORNENBAL, P.: On the resistance and the Bernoulli effect of the human larynx. Journ. Acoust. Soc. Amer., 1957, 29, s.626-631.
2. CHIBA, T., KAJIYAMA, M.: The vowel, its nature and structure. Tokyo Kaiseikan Pub. Co., 1941.
3. FANT, G.: Acoustic Theory of Speech Production. Mouton and Co., s Gravenhage, 1960.
4. FANT, G.: Vocal tract wall effects, losses and resonance bandwidths. STL-QPSR, 1972, 2-3, s.28-52.
5. FANT, G., NORD, L., BRANDERUD, P.: A note on the vocal tract wall impedance. STL-QPSR, 1976, 4, s.13-20.
6. FLANAGAN, J.L.: Analog measurements of sound radiation from the mouth. Journ. Acoust. Soc. Amer., 1960, 32, s.1613-1620.
7. FLANAGAN, J.L.: Speech Analysis, Synthesis and Perception. Springer Verlag, Berlin-Heidelberg-New York, wyd.1, 1965, wyd.2, 1972.
8. FLANAGAN, J.L., LANDGRAF, L.L.: Self-oscillating source for vocal-tract synthesizers. IEEE Trans. on Audio- and Electroacoustics, 1968, AU-16, s.57-64.
9. FLANAGAN, J.L., ISHIZAKA, K., SHIPLEY, K.L.: Synthesis of speech from a dynamic model of the vocal cords and vocal tract. Bell. Syst. Techn. Journ., 1975, 54, s.485-506.
10. FRANKE, E.K.: Mechanical impedance measurements of the human body surface. Rpt. No. 6469, Wright Air Development Centre, Dayton, Ohio, 1951.
11. GUBRYNOWICZ, R., MIKIEL, W., ŻARNECKI, P.: Akustyczna metoda oceny stanu źródła krtaniowego w przypadkach zmian patologicznych fałdów głosowych. Arch. Akust., 1980, 15, 1, s. 3-30.
12. GUBRYNOWICZ, R., MIKIEL, W., ŻARNECKI, P.: Acoustical analysis for the evaluation of laryngeal dysfunction in the case of vocal cord paralysis. [W:] Speech Analysis and Synthesis, t.5, s.139-163, IPPT-PAN (red. W.Jassem), PWN, Warszawa, 1981.
13. HECKER, M.H.L.: Studies of nasal consonants with an articulatory speech synthesizer. Journ. Acoust. Soc. Amer., 1962, 34, s. 179-188.
14. HEINZ, J.M.: Reduction of speech spectra to description in terms of vocal tract area functions. D. Sc. Thesis, Mass. Inst. of Techn., Cambridge, Mass., 1962.

15. HOUSE, A.S.: Analog studies of nasal consonants. Journ. Speech and Hearing Disorders, 1957, 22, s.190-204.
16. HOUSE, A.S., STEVENS, K.N.: Analog studies of the nasalization of vowels. Journ. Speech and Hearing Disorders, 1956, 21, s.218-232.
17. ISHIZAKA, K.: On models of the larynx. Journ. Acoust. Soc. Japan, 1966, 22, 293-294.
18. ISHIZAKA, K., FLANAGAN, J.L.: Synthesis of voiced sounds from a two-mass model of the vocal cords. Bell System Techn. Journ., 1972, 51, s.1233-1268.
19. JASSEM, W.: Podstawy Fonetyki Akustycznej, IPPT-PAN, PWN, Warszawa, 1973.
20. KACPROWSKI, J.: Teoretyczne podstawy syntezy samogłosek polskich w rezonansowych układach formantowych. Rozprawy Elektrotechn., 1962, 8, 1, s.127-203.
21. KACPROWSKI, J.: Theoretical bases of the synthesis of Polish vowels in resonance circuits. [W:] Speech Analysis and Synthesis, t.1, s.219-287, IPPT-PAN (red. W.Jassem), PWN, Warszawa, 1968.
22. KACPROWSKI, J.: Podstawy teoretyczne i realizacja techniczna formantowego syntezyzatora mowy SYNFOR II. Arch. Akust., 1969, 4, 2, s.199-220.
23. KACPROWSKI, J.: Fizyczne modele źródła krtaniowego. Arch. Akust., 1977, 12, 1, s.47-70.
24. KACPROWSKI, J.: Model symulacyjny kanału głosowego z uwzględnieniem zjawiska nazalizacji. Arch. Akust., 1977, 12, 4, s.281-302.
25. KACPROWSKI, J.: An acoustical model of the vocal tract for the diagnostics of cleft palate. [W:] Speech Analysis and Synthesis, t.5, s.165-183, IPPT-PAN (red. W.Jassem), PWN, Warszawa, 1981.
26. KACPROWSKI, J., MIKIEL, W., SZEWCZYK, A.: Akustyczne badania modelowe rozszczepu podniebienia. Arch. Akust., 1976, 11, 2, s.167-187.
27. KÖSTER, J.P.: Abriss historischer Ansätze der Sprachsymthese. [W:] Speech Analysis and Synthesis, t.4, s.41-104, IPPT PAN (red. W.Jassem), PWN, Warszawa, 1976.
28. MORSE, P.M.: Vibration and Sound. Mc Graw-Hill, New York 1948.
29. MORSE, P.M., INGARD, K.U.: Theoretical Acoustics, Mc Graw-Hill, New York, 1968.

30. MRAYATI, M., GUERIN, B.: Etude des caractéristiques acoustiques des voyelles orales françaises par simulation du conduit vocal avec pertes. Revue d'Acoustique, 1976, 2, 36, s.18-31.
31. RAYLEIGH, J.W.S.: The Theory of Sound. Macmillan, New York, 1945.
32. SONDEHI, M.M.: Model of wave propagation in a lossy vocal tract. Journ. Acoust. Soc. Amer., 1974, 55, s.1070-1075.
33. STEVENS, K.N., KASOWSKI, S., PANT, G.: An electrical analogue of the vocal tract. Journ. Acoust. Soc. Amer., 1953, 25, s.734-742.
34. TŁUCHOWSKI, W., MIKIEL, W., KOMOROWSKA, A., WIDLICKA, E. Próba korelacji bioelektrycznych i fonospektroskopowych w różnych typach operacji rozszczepów podniebienia u dzieci. Otolaryngologia Polska, 1974, 8, 6a, s.70-77.
35. UNGEHEUER, G.: Elemente einer Akustischen Theorie der Vocal-Artikulation. Springer Verlag, Berlin-Heidelberg-Göttingen, 1962.
36. WAKITA, H., PANT, G.: Toward a better vocal tract model. STL-QPSR, 1978, 1, s.9-29.
37. WEBSTER, A.G.: Acoustical impedance and the theory of horns. Proc. Nat. Acad. Sci. U.S., 1919, 2, s. 275-282.
38. WEGEL, R.L.: Theory of vibration of the larynx. Bell System Techn. Journ., 1930, 2, s.207-227.

Pracę wykonano w ramach problemu węzłowego
06.9: Podstawowe Problemy Elementów i Systemów
Inżynierii Biomedycznej