

Akustyka Mowy

Ryszard Gubrynowicz

20250  
KOMPUTEROWE MODELOWANIE ARTYKULACJI  
GŁOSEK JĘZYKA POLSKIEGO

(PRACA HABILITAC.)

4/2000



P.269

WARSZAWA 2000

<http://rcin.org.pl>

INSTYTUT PODSTAWOWYCH PROBLEMÓW TECHNIKI PAN

BIBLIOTEKA

02-106 Warszawa, ul. Pawińskiego 5B

Tel. 22-826-74-10

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 13 kwietnia 2000 r.

recenzent - Prof.dr hab. Andrzej Rakowski



56520

(PRACA HABILITAC.)



9259



Praca habilitacyjna

---

Instytut Podstawowych Problemów Techniki PAN  
Nakład 100 egz. Ark. wyd.11.0 Ark. druk.13,5  
Oddano do drukarni w grudniu 2000 r.

---

ATOS - Poligrafia-Reklama, W-wa, ul. Jana Kazimierza 35/37

<http://rcin.org.pl>

## Spis treści:

1. Wstęp .....	7
2. Narząd artykulacyjny człowieka jako złożony układ akustyczny generacji dźwięku .....	18
2.1 Budowa anatomiczna organu mowy .....	18
2.2 Układ oddechowy .....	21
2.3 Układ fonacyjny .....	21
2.4 Szumowe pobudzenie toru głosowego .....	24
2.5 Mechanizm generacji sygnału mowy .....	25
3. Artykulacja dźwięków mowy .....	28
4. Dźwięki mowy języka polskiego i ich cechy fonetyczno-akustyczne .....	40
4.1 Opis artykulacyjny dźwięków mowy polskiej .....	40
4.2 Samogłoski .....	53
4.3 Głoski szczelinowe (trące) .....	57
4.4 Głoski zwarto-szczelinowe .....	62
4.5 Głoski zwarto-wybuchowe .....	64
4.6 Koartykulacja .....	66
4.6.1 Upodobnienia .....	67
4.6.2 Palatalizacja .....	68
5. Zależność między konfiguracją geometryczną toru głosowego i sygnałem akustycznym mowy .....	70
5.1 Wstęp .....	70
5.2 Modelowanie artykulacji samogłoskowej .....	73
5.3 Podstawy fizyczne modelowania toru głosowego jako układu akustycznego .....	76
6. Modelowanie pobudzenia krtaniowego .....	98
6.1 Wstęp .....	98
6.2 Budowa krtani .....	98
6.3 Podstawy działania źródła krtaniowego .....	100
6.4 Model jednomasowy .....	107
6.5 Opis ruchu fałdów głosowych .....	112
6.6 Dwumasowy model źródła krtaniowego .....	117
6.7 Nieliniowy model źródła krtaniowego dla małej amplitudy drgań .....	125

6.7.1 Model uwzględniający rozchodzenie się w fałdach głosowych fal powierzchniowych .....	126
6.7.2 Analiza ruchu fałdów przy braku sprzężenia głośni z torem głosowym .....	132
6.7.3 Warunki oscylacji przy bezwładnościowym sprzężeniu źródła z torem głosowym .....	142
7. Źródło pobudzenia bezdźwięcznego .....	153
8. Komputerowy model układu artykulacyjnego .....	163
8.1 Wstęp .....	163
8.2 Ogólny schemat realizacji modelu artykulacyjnego .....	164
8.3 Opis programu <i>MODEL</i> .....	172
8.4 Przyjęta metoda modelowania rezonantów i ich połączeń w oparciu o dane artykulacyjne głosek polskich .....	179
8.5 Wyniki modelowania połączeń rezonantów .....	198
9. Zakończenie .....	206
10. Literatura .....	211

## Wykaz symboli uproszczonej transkrypcji fonetycznej stosowanej w pracy.

W pracach o charakterze fonetycznym stosowany jest na ogół międzynarodowy system transkrypcji fonetycznej zalecany przez International Phonetic Association, który opiera się na klasyfikacji dźwięków mowy według ich artykulacji. Taka transkrypcja była stosowana w niniejszej pracy. Sterownik czcionki tego alfabetu jest swobodnie dostępny w Internecie, np. pod adresem <http://www.sil.org/fonts/encore-ipa.html>

Obecnie coraz częściej jest stosowana transkrypcja kodu fonetycznego w kodzie ASCII, w tzw. kodzie SAMPA, opracowana przez J. Wellsa z University College of London (<http://www.phon.ucl.ac.uk/home/sampa/home.htm>).

Zaletą kodu SAMPA jest jego uniwersalność umożliwiająca uniezależnienie się od stosowanego edytora tekstu, systemu operacyjnego i możliwość bezpośredniej transmisji transkrypcji w trybie znakowym poprzez sieć Internetu. Kod SAMPA został użyty przy realizacji m.in. językowych baz danych EUROM1 [CHAN i in., 1995], BABEL [ROACH i in., 1996; GUBRYNOWICZ, 1998].

### Samogłoski ustne

Symbol IPA	Symbol SAMPA	Np. w wyrazie
i	i	tik
ɪ	I	typ
e	e	test
a	a	pat
o	o	pot
u	u	puk

### Samogłoski nosowe

Symbol IPA	Symbol SAMPA	Np. w wyrazie
ẽ	e~	tę
ã	a~	tą

## Spółgłoski

### Spółgłoski zwarte

Symbol IPA	Symbol SAMPA	W wyrazie (transkr. SAMPA)
p	p	puk /puk/
b	b	bat /bat/
t	t	test /test/
d	d	dym /dĭm/
k	k	kat /kat/
g	g	gen /gen/

### Spółgłoski trące

Symbol IPA	Symbol SAMPA	w wyrazie (transkr. SAMPA)
f	f	fakt /fakt/
v	v	waga /vaga/
s	s	syk /sĭk/
z	z	zbir /zbir/
ʃ	S	szyk /Sik/
ʒ	Z	żyto /Zito/
ɕ	s'	świt /s'vit/
ʒ	z'	źle /z'le/
x	x	hak /hak/

### Spółgłoski zwarto-trące

Symbol IPA	Symbol SAMPA	w wyrazie (transkr. SAMPA)
ʦ	ts	coś /tsos'/
ʣ	dz	dzwon /dzvon/
tʃ	tS	czapka /tSapka/
dʒ	dZ	dżem /dZem/
tɕ	ts'	ćwicz /ts'vitS/
ʣ	dz'	dźwig /dz'vig/

## Sonoranty

Symbol IPA	Symbol SAMPA	w wyrazie (transkr. SAMPA)
m	m	mysz /mIS/
n	n	nasz /naS/
ɲ	n'	koń /kon'/
ŋ	N	bank /baNk/*
w	w	łyk /wIk/
j	j	jak /jak/
l	l	luk /luk/
r	r	ryk /rIk/

\*Spółgłoska nosowa /ɲ/ występuje w języku polskim tylko przed spółgłoskami /k,g/.

## I. Wstęp

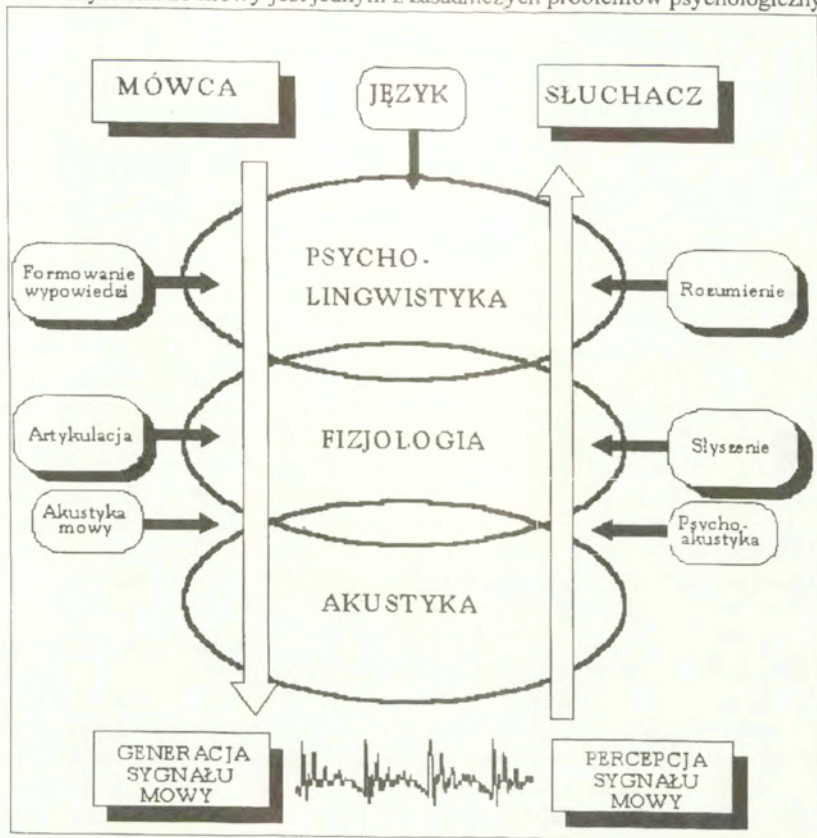
Mowa jest jednym z najbardziej powszechnych i efektywnych środków porozumiewania się między ludźmi. Ostatnio coraz bardziej widoczna staje się tendencja stosowania jej również jako jednego z najbardziej naturalnych środków do przekazywania informacji i komunikatów między człowiekiem i maszyną, zwłaszcza w tzw. wszechstronnych systemach dialogowych, w których przekazywany jest równocześnie obraz wizualny wraz z dźwiękiem [COLE i in., 1996].

Aby móc w pełni zrozumieć dlaczego mowa jest tak efektywnym nośnikiem informacji należy poznać podstawowe fakty związane z generacją sygnału mowy, formowaniem komunikatów i poleceń, oraz z ich percepcją i rozumieniem. Zarówno formowanie sygnału po stronie nadawczej jak i jego przetwarzanie i dekodowanie zawartej w nim informacji semantycznej po stronie odbiorczej, są złożonymi, wielopoziomowymi procesami intelektualnymi, a nie zwykłymi procesami kodowania i dekodowania informacji. Na przykład, wypowiedziane wyrazy, czy zdania są tylko stosowanym do przekazywania informacji semantycznej środkiem, a nie celem samym w sobie. Warto zwrócić uwagę na fakt, że już w okresie wczesnego dzieciństwa opanowanie zdolności językowej dokonuje się równoległe do rozwoju aktywności intelektualnej, w wyniku czego nabywa się umiejętności przekazywania myśli w formie dogodnej do obróbki intelektualnej, to jest właśnie w postaci wyrazów i zdań.

U podstaw badania procesu tworzenia i percepcji sygnału mowy znajdują się trzy zasadnicze poziomy jego opisu: psycholingwistyczny, fizjologiczny oraz akustyczny (Rys. 1.1). Po stronie tzw. nadawczej w generacji sygnału mowy uczestniczą na wyższych piętrach dwa zasadnicze procesy - formowanie intelektualne wypowiedzi oraz artykulacja, zaś po stronie odbiorczej - słyszenie i rozumienie. Przy badaniu procesu generacji sygnału mowy na poziomie akustycznym, podstawowym narzędziem jest akustyka mowy, która traktuje narząd artykulacyjny człowieka jako specyficzne źródło akustyczne. Przebieg akustycznego sygnału mowy może być analizowany i interpretowany na każdym z wyżej wymienionych poziomów. Na najniższym, przedmiotem analizy jest fizyczna struktura poszczególnych dźwięków i zmiany w niej zachodzące podczas ich artykulacji. Natomiast na poziomie najwyższym (odpowiadającym wyższym czynnościom nerwowym), przedmiotem analizy jest psychologiczna realność struktury używanego języka komunikacji słownej przejawiająca się przede wszystkim w jego systemie fonologicznym. Podstawowym modelem do rozpatrywania zjawisk natury językowej może być przyjęty model teorii komunikacji operujący takimi pojęciami jak wejście, wyjście, źródło, nadawca, odbiorca, kanał, kodowanie, dekodowanie itp. Na każdym poziomie kodowania (lub dekodowania) operuje się odpowiednimi dla niego jednostkami. Na poziomie psycholingwistycznym niejako w sposób naturalny są nimi sylaba, wyraz i zdanie. Sam język jest ujmowany jako system relacji na poziomie fonologicznym, semantycznym i



syntaktycznym, czyli jako swoistego rodzaju struktura [SAUSSURE DE,1961]. Ujęcie stosunku myślenia do mowy jest jednym z zasadniczych problemów psychologicznych.



Rys. 1.1 Dziedziny wiedzy i procesy przetwarzania sygnału mowy będące u podstaw dwustronnej komunikacji słownej.

Istotne podstawy tych zagadnień zostały stworzone przez Vygotskiego [VYGOTSKI,1962]. W zasadzie jest to proces, w którym następuje stałe przechodzenie od słowa do myśli i od myśli do słowa. W procesie tym realna postać słowa jest pewną sekwencją dźwięków, która jako całość ma określone znaczenie, zaś myśl jest ustalaniem związków między rzeczami. Głównym narzędziem myśli są pojęcia, zaś pojęcia są znaczeniem słów. Tak więc w bardzo dużym uproszczeniu - znaczenie słowa jest łącznikiem między myślą i mową.

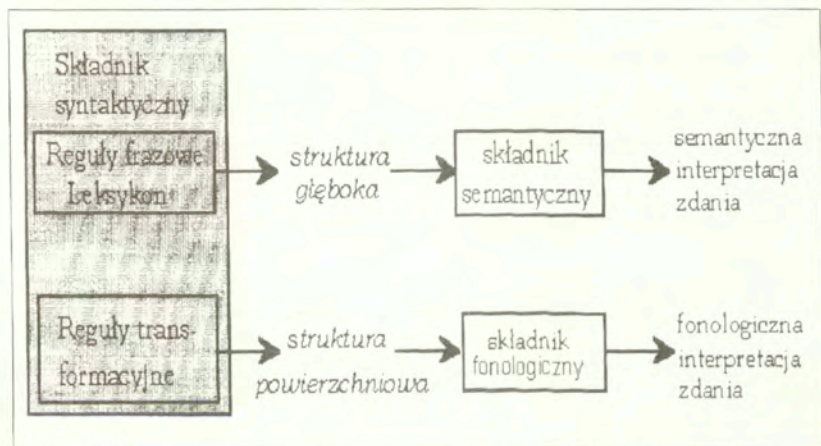
Znaczenie słowa przyjmuje się za podstawową jednostkę myślenia słownego. Jednakże znaczeniową formę słowa nie można wyłącznie rozpatrywać jako zjawiska należącego tylko do sfery myśli, bądź tylko mowy. Rozróżnienie pojęć język i mowa jest tu podstawowe. Język jest w naszym modelu traktowany jako system znaków dźwiękowych i określonych reguł fonologicznych, syntaktycznych i semantycznych rządzących kombinacją tych znaków w akty mowy. Przez mowę rozumie się konkretne akty użycia języka, czyli czynności werbalne człowieka. Język jest głównym narzędziem w procesie porozumiewania się ludzi, mowa - to konkretne realizacje tego procesu. Najprostszy model języka jest modelem sekwencyjnym, w którym zakłada się, że wypowiedź jest ciągiem elementów rozwijającym się w czasie i że istnieją powiązania między kolejnymi elementami językowymi (wyrazami, literami, fonemami). Oznacza to, że każdy element znajdujący się w sekwencji od lewej strony pociąga za sobą następujący po nim prawy, ostatnim zaś elementem jest znak końca ciągu, hamujący czynność jego generowania. Jednocześnie model ten ma charakter probabilistyczny, gdyż uwzględnia on prawdopodobieństwa, z jakim kolejne elementy ciągu mogą wystąpić na zadanych jego pozycjach, oraz wzajemne zależności między tymi prawdopodobieństwami. Ciągi elementów mogą być modelowane np. za pomocą dyskretnych modeli Markowa. Taki sposób modelowania jest często stosowany w systemach rozpoznawania mowy [YOUNG & BLOOTHOOFT, 1997]. Był on również zastosowany w systemie rozpoznawania wybranych klas dźwięków mowy polskiej [GUBRYNOWICZ & WRZOSKOWICZ, 1993], czy izolowanych wyrazów [GUBRYNOWICZ & WRZOSKOWICZ, 1996; WRZOSKOWICZ, 1996].

Prawdopodobieństwa te mogą być ustalone przez obliczenie częstotliwości występowania określonej jednostki w przyjętym zbiorze tekstów danego języka. Taką reprezentacją tych prawdopodobieństw może być, na przykład, słownik frekwencyjny słów, uwzględniający nie tylko częstotliwość występowania poszczególnych słów, ale również ich kombinacji, np. sekwencji określonych par słów, trzech słów itp. Zakłada się, że dla danego słowa jego związki z pewnymi słowami są silniejsze niż z innymi. Ustalenie prawdopodobieństwa występowania jakiegoś słowa w ciągu zdaniowym nie ma, oczywiście, nic wspólnego z ustalaniem przez lingwistykę strukturalistyczną możliwości pojawienia się jakiegoś słowa w danym miejscu zdania (relacja syntagmatyczna). W tym drugim wypadku chodzi o to, czy jedno słowo określa inne, a więc o związek semantyczny, zaś w pierwszym oblicza się, jakie jest prawdopodobieństwo tego faktu, a więc chodzi tu o związek statystyczny. Choć model statystyczny może być zastosowany do opisu mechanizmu generowania zdań, to ma on zasadnicze ograniczenia spowodowane głównie tym, że nie uwzględnia związków semantycznych między wyrazami. Na przykład, w modelu tym prawdopodobieństwo sekwencji "słyszę głos pterodaktyla" oraz "słyszę szybko palić" jest w obu przypadkach bliskie zeru, ale tylko to drugie jest dla języka polskiego niemożliwe syntagmatycznie. Inną bardzo istotną wadą modelu statystycznego jest jego liniowość, tzn. uniemożliwia on generowanie zdań podrzędnych wtrąconych.

Opisanie związków semantycznych między słowami umożliwia model Chomsky'ego oparty na teorii gramatyki generatywno-transformacyjnej. Gramatyka generatywna jakiegoś języka to jednak nie taka, której reguły pozwalają na generowanie, czyli tworzenie, produkowanie i odbieranie zdań, lecz jest modelem kompetencji mówiących [RUWET, 1982]. Umożliwia ona rozumienie lub opisywanie ściśle według reguł zdań danego języka. Rozróżnienie nadawcy i odbiorcy odbywa się dopiero na poziomie tzw. "struktury powierzchniowej" (patrz dalej). W zasadzie każda gramatyka jest generatywna (przynajmniej w zamierzeniu, bowiem nie operuje ściśle regułami) w podanym tu sensie tego pojęcia. Przymiotnik transformacyjna oznacza, że oprócz ścisłych reguł opisu zdań, gramatyka zawiera także reguły transformacji jednego zdania w drugie np. reguły przekształceń strony czynnej na bierną. W omawianym modelu wyróżnione są trzy podstawowe składniki:

1. syntaktyczny, który składa się z dwóch zasadniczych składników: reguł frazowych (opisu zdań) i słownika oraz ze składnika transformacyjnego zawierającego reguły transformacyjne;
2. semantyczny - będący interpretacją semantyczną danego zdania i
3. fonologiczny - będący interpretacją funkcji językowej dźwięków mowy.

Relacje między poszczególnymi składnikami modelu przedstawia Rys. 1.2 [KURCZ, 1976]. W modelu tym wyróżnia się dwie zasadnicze struktury, tzw. głęboką i powierzchniową.



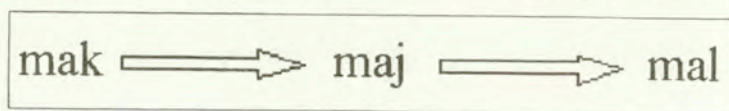
Rys.1. 2 Model Chomsky'ego generacji zdań (wg [KURCZ, 1976]).

Pierwsza prowadzi do interpretacji semantycznej zdania, przed przekształceniem jego w formę ostatecznej wypowiedzi, co następuje w strukturze powierzchniowej. W niej to zaś dokonywana jest złożona transformacja zdania na postać fonologiczną, która decyduje o brzmieniu danego zdania. Warto zwrócić uwagę, że relacje między tymi strukturami nie są zawsze jednoznaczne. Świadczą o tym przykłady zdań dwuznacznych, tzn. jednakowo brzmiących na poziomie fonologicznym, ale o zupełnie innym znaczeniu semantycznym (czyli o innej strukturze głębokiej).

Zadaniem języka jest połączenie znaczenia z dźwiękiem. Czyni to składnik syntaktyczny, zaś ocenę i ostateczną realizację tego celu zapewniają składniki interpretujące. Istota języka sprowadza się do relacji między strukturą głęboką a powierzchniową [KURCZ, 1976]. Przedstawiony skrótowo model języka oparty na gramatyce generatywno-transformacyjnej nie jest jedynym modelem, który jest stosowany w badaniach językoznawczych. Są również modele całkowicie odrzucające pojęcie struktury głębokiej. Tym nie mniej przedstawiony model Chomsky'ego pomaga sformalizować relacje istniejące między dźwiękiem i znaczeniem.

Językoznawcy wyróżniają pewne podstawowe dźwiękowe jednostki językowe, za pomocą których można opisać każdy język niezależnie od treści i form samego procesu porozumiewania się. Najmniejszą jednostką dźwiękową, którą można w danym języku wyodrębnić jest fonem. Składa się on z zespołu cech dystynktywnych wyróżniających go od innych fonemów. Sam fonem nie ma znaczenia, ale jest nosicielem jego zmiany w minimalnej jednostce lingwistycznej jaką jest słowo. Na przykład, zastąpienie w słowie jednego fonemu innym prowadzi bądź do zmiany znaczenia, czy wręcz do jego utraty.

Porównajmy jak zmienia się znaczenie w wyrazie jednosylabowym typu CVC przy zmianie pierwszej spółgłoski, tak jak to zostało przedstawione na Rys. 1.3:



Rys.1.3 Przykład zmiany znaczenia wyrazu przy wymianie w nim jednej głoski.

W podanym wyżej przykładzie zamiana głoski /k/ na /j/ powoduje zmianę znaczenia wyrazu, zaś w przypadku podstawienia na ostatnim miejscu głoski /l/ jego utratę (przynajmniej w języku polskim).

Każdy przekaz mówiony można przedstawić jako odpowiednią sekwencję dźwięków. Dziedzina zajmująca się sposobem wymawiania oraz ich klasyfikacją jest przedmiotem fonetyki, natomiast dziedziną, której przedmiotem jest badanie ich funkcji językowej jest przedmiotem fonologii. Częściowym dopełnieniem fonetyki jest

akustyka mowy, w której szczególnie nacisk położony jest na badanie fizycznej struktury dźwięków mowy, i jej zmian podczas ich artykulacji. Badania w zakresie akustyki mowy spowodowały burzliwy rozwój metod analizy przetwarzania sygnałów, nie tylko sygnału mowy i zostały wykorzystane w wielu innych dziedzinach.

Tak więc badając ogólne mechanizmy procesu formowania i odbierania przekazu językowego, na poziomie fonetycznym analizuje się również proces tworzenia jego struktury dźwiękowej, głównie zaś artykulację oraz percepcję, a także przetwarzanie przez umysł ludzki dźwięków mowy istotnych dla tego przekazu. Badania nad percepcją fonemów i rozpoznawaniem zdań są podstawowym narzędziem psycholingwistyki. W pracy jednak omawiane będą przede wszystkim akustyczne aspekty mowy i tylko z tego poziomu będą poruszane niektóre fonetyczne i fonologiczne zagadnienia bezpośrednio związane z formowaniem sygnału mowy. Głównym zaś punktem wyjścia do omawiania własności sygnału mowy będzie układ artykulacyjny człowieka i funkcjonowanie tego układu podczas generacji dźwięków mowy.

W zakresie badań fizjologicznych zasadniczym przedmiotem zainteresowania jest praca narządów artykulacyjnych podczas mówienia. Badania te obejmują przede wszystkim analizę ruchów poszczególnych elementów narządu artykulacyjnego (np. ruchów warg, języka itp.) podczas wymawiania określonych głosek, czy ich sekwencji. W badaniach tych niemal do ostatnich lat dominowała metoda kinorentgenograficzna umożliwiająca synchroniczną z zapisem dźwięku rejestrację ruchów artykulacyjnych, tj. zmian układów narządów przede wszystkim jamy ustnej, gardłowej i warg. Metoda kinorentgenograficzna jednak ma bardzo poważne ograniczenie, bowiem rejestrowany obraz układu narządów artykulacyjnych jest tylko dwuwymiarowy i daje tylko obraz podłużnego przekroju toru głosowego, od głośni do warg. Tak więc trzeci wymiar przekroju toru można tylko w przybliżeniu określić np. przez wykonanie odcisków gipsowych przedniej części toru, która znajduje się pod podniebieniem twardym, lub wykonując zdjęcia na wprost warg określając wielkość otworu wlotowego ust.

Dopiero stosunkowo zupełnie niedawno wprowadzono metodę trójwymiarowej wizualizacji toru gardłowo-ustnej, stosując metodę rezonansu magnetycznego (tzw. MRI [BAER i in.,1991; DEMOLIN i in.,1996; STORY i in.,1996; ALVAN i in.,1997;BADIN i in.,1998]). Jednakże metoda ta, wciąż jeszcze niedoskonała, narzuca osobie mówiącej wielokrotne powtarzanie w sposób ciągły artykułowania danego wyrazu, by móc zarejestrować w pełni przebieg zmian ukształtowania toru głosowego w przestrzeni 3-wymiarowej, a następnie post-synchroniczne nagranie dźwięku, gdyż aparatura rejestrująca obraz ma bardzo duży poziom zakłóceń akustycznych. Jednakże mimo wielu jeszcze ograniczeń metoda wizualizacji MRI jest niewątpliwie metodą przyszłościową, gdyż w przeciwieństwie do metody kinorentgenograficznej wykorzystującej naświetlanie promieniami X o stosunkowo dużym czasie ekspozycji i intensywności, jest w pełni bezpieczna dla lektora. Jednakże dla języka polskiego do

obecnej chwili nie ma innych danych, jak dane kinorentgenograficzne opublikowane przez Bolla i Földi [BOLLA&FÖLDI,1987], w których podano m.in. przekroje podłużne toru głosowego podczas artykulacji głosek polskich wymówionych pojedynczo wyrazach. W pracy tej, co prawda są niekiedy błędy w transkrypcji fonetycznej zarejestrowanych wyrazów, ale łatwe są one do skorygowania. Podane w tej książce dane dotyczące konfiguracji toru głosowego były podstawowym źródłem informacji przy komputerowym modelowaniu rezonantów języka polskiego oraz ich połączeń. Praca Bolla i Földi jest prawdopodobnie jedną z ostatnich, w której zastosowano w badaniach fonetycznych kinorentgenograficzną metodę dynamicznej wizualizacji toru głosowego.

Jak już wspomniano, po stronie odbiorczej przetwarzanie sygnału mowy odbywa się również na wielu poziomach. Na najniższym, to jest na poziomie błony bębenkowej tzw. ucha wewnętrznego, następuje przekodowanie fali akustycznej na ciąg impulsów nerwowych. Te są dalej przetwarzane na wyższych piętrach układu nerwowego, w wyniku czego następuje zdekodowanie poszczególnych dźwięków, ich sekwencji i wreszcie treści przekazywanej wiadomości. Badania percepcji i rozumienia mowy są bardzo złożone, bowiem choć struktura dźwięków czy ich sekwencji może być ściśle określona to dalszy mechanizm ich przetwarzania jest bardzo trudny do bezpośredniej obserwacji. Tym nie mniej badania nad percepcją mowy, a zwłaszcza percepcją poszczególnych dźwięków stanowiły przez wiele lat, i stanowią nadal jedną z zasadniczych metod poznania ich wewnętrznej struktury i roli spełnianej przy formowaniu zdania czy komunikatu. Jednakże, zagadnienia związane z samą percepcją mowy stanowią bardzo obszerne i całkiem odrębne zagadnienie i dlatego w tej pracy nie będą szerzej omawiane.

Podobnie jak w wielu dziedzinach nauki, symulacja komputerowa jest obecnie w zakresie poznawania własności sygnału mowy jedną z powszechniej stosowanych technik. Bowiem, jak w przypadku modelowania różnych systemów i procesów, tak i opracowanie modelu układu artykulacyjnego umożliwia odwzorowanie wielu cech i zależności istniejących w sygnale, a trudnych do bezpośredniego uchwycenia w nim przy stosowaniu standartowych metod pomiarowych. O popularności modelowania zadecydował fakt, że często jest jedyną możliwą metodą analizy i oceny działania układu rzeczywistego. Próby wykonania bezpośrednich pomiarów w celu dokonania obiektywnej oceny pracy niektórych skomplikowanych systemów i układów, prowadzą zazwyczaj do zakłóceń w działaniu badanego obiektu i wówczas konwencjonalne metody analityczne i pomiarowe okazują się mało skuteczne, a niekiedy wprowadzają dodatkowe źródła niekontrolowanych zakłóceń w jego pracy. W szczególności dotyczy to badania układów biologicznych w trakcie ich pracy, gdy wykonywane pomiary mają inwazyjny charakter. Modelowanie organu mowy oraz symulacja jego funkcji

głosotwórczych i artykulacyjnych jest tylko szczególnym przypadkiem metod analizy i oceny działania rzeczywistych układów biologicznych.

Poznanie mechanizmu powstawania mowy było od wielu lat przedmiotem badań szeregu ośrodków naukowych, analizujących proces generacji i percepcji mowy, zarówno od strony lingwistycznej, jak i fizycznej. Pełne podsumowanie wyników badań fizycznej struktury dźwięków mowy i jej powiązania z budową i działaniem narządów artykulacyjnych ostatnio zostało obszernie przedstawione w monografii Stevensa [STEVENS,1998]. Niemal od samego początku prowadzenia badań fonetycznych akustyczny sygnał mowy był przedmiotem bardzo wielu pomiarów wykonywanych dla głosek wielu języków. Choć w literaturze dominują dane dotyczące przede wszystkim języka angielskiego, to warto podkreślić, że w porównaniu z innymi językami, istniejąca dla języka polskiego baza danych zawierająca wyniki wszechstronnych pomiarów jest stosunkowo dość bogata. Jest ona rezultatem wieloletniej pracy polskich fonetyków ([JASSEM,1968-1980; JASSEM,1973; WIERZCHOWSKA,1980; DŁUSKA,1981; DUKIEWICZ,1995] i inni).

Kierunek badań nad modelowaniem toru głosowego, prowadzonych w wielu ośrodkach naukowych na świecie, szczególnie w USA (np. [ATAL i in.,1977;1989], [BAER i in.,1991], [LIN,1994], [SCHROETER i in., 1994], [STORY i in.,1996], [NARAYANAN i in.,1997], [ALVAN i in.,1997]) Japonii [SHIRAI,1993] , Szwecji ([LIN,1990], [FANT,1995], [LÓFQUIST,1997], [ENGWALL,1999], Francji ([MAEDA,1988], [CARRE i in.,1992], [BEAUTEMPS D., i in., 1995], [BADIN i in.,1998] jest podobny i oparty na tych samych klasycznych założeniach, stworzonych na początku lat pięćdziesiątych przez Dunna [DUNN,1950]. Przełomową pracą, która umożliwiła pełniejsze zrozumienie mechanizmu powstawania dźwięków mowy, była monografia szwedzkiego uczonego G. Fanta "*Acoustic theory of speech production*" [FANT,1960], w której przedstawił on teorię powstawania sygnału mowy oraz modelowania układu artykulacyjnego. Teoria ta została opracowana w oparciu o wyniki pomiarów techniką radiologiczną przekrojów toru głosowego podczas artykulacji samogłosek języka rosyjskiego, a także i innych głosek, w tym języka szwedzkiego i angielskiego. W pracy tej przedstawiono zasadę modelowania toru głosowego, traktowanego jako złożony układ akustyczny i polegającą na przedstawieniu jego w postaci rury o zmiennym przekroju, podzielonej na skończoną liczbę segmentów elementarnych. Każdy segment jest następnie modelowany za pomocą układu o stałych skupionych w postaci elektrycznego układu zastępczego, którego elementy są wyznaczone w oparciu o przyjęty system analogii akusto-elektrycznych.

Ogólnie biorąc obecnie można wyróżnić dwa sposoby podejścia do zagadnienia modelowania artykulacyjnego toru głosowego. Pierwszy (np. [MAEDA,1988; LIN, 1990; CARRE i in.,1992; BEAUTEMPS D., i in., 1995]), stosowany i udoskonalany przez wiele lat, opiera się na bezpośrednich pomiarach konfiguracji toru głosowego podczas artykulacji wybranych głosek, a następnie na wyznaczeniu funkcji przekroju i wyliczenia analitycznej funkcji transmitancji toru. W oparciu o tak określoną funkcję transmitancji dokonywana jest synteza dźwięku przy uwzględnieniu własności źródła pobudzenia krtaniowego i funkcji promieniowania ust. Dokładność modelowania jest dokonywana w płaszczyźnie akustycznej przez porównanie widm (głównie częstotliwości formantowych) sygnału syntetycznego i naturalnego zarejestrowanego podczas pomiarów konfiguracji .

Drugi sposób (np. [ATAL i in., 1989], [SCHROETER i in., 1994], [CIOCEA i in.,1998]) przebiega dokładnie w odwrotnym kierunku, to znaczy na podstawie pomierzonych charakterystyk widmowych wyznacza się funkcję powierzchni toru głosowego i dokładność modelowania jest określana w płaszczyźnie artykulacyjnej przez porównanie z funkcją powierzchni określoną dla pomierzonej konfiguracji geometrycznej toru głosowego. Ten sposób jest znacznie bardziej złożony niż pierwszy i niekiedy wyznaczone funkcje powierzchni przekroju znacznie odbiegają od oczekiwanych.

W przypadku badań sygnału mowy analiza niektórych procesów artykulacyjnych w płaszczyźnie fizycznej jest możliwa niemal wyłącznie poprzez modelowanie, jak na przykład nazalizacja, dynamiczne zmiany konfiguracji toru głosowego mające miejsce podczas wymawiania głosek o charakterze niestacjonarnym (spółgłoski płynne, drżące itp.). W oparciu o symulację komputerową procesów artykulacyjnych prowadzi się badania również, o charakterze poznawczym. Wnoszą one istotny wkład do rozwoju badań nad cechami fonetyczno-akustycznymi sygnału mowy, których wyniki znajdują również praktyczne zastosowania, jak na przykład w syntezie i automatycznym rozpoznawaniu mowy, a więc w dziedzinach mających znaczenie przy jak najszerzej pojętej dwustronnej komunikacji człowiek-maszyna. Warto jeszcze wspomnieć o jeszcze innym szczególnym zastosowaniu prac nad modelowaniem toru głosowego, a mianowicie do analizy w płaszczyźnie akustycznej wpływu zmian patologicznych w narządzie mowy, takich jak rozszczepy, guzki na fałdach głosowych, porażenie fałdów głosowych itp. Modelowanie tych zmian umożliwia bardziej dokładne uchwycenie związków istniejących między zmianami patologicznymi organu mowy, a wybranymi parametrami sygnału mowy [KACPROWSKI i in.,1976;KACPROWSKI,1980; GUBRYNOWICZ,1980].



Badania nad modelowaniem toru głosowego wykonane dla głosek języka polskiego są stosunkowo nieliczne i praktycznie biorąc były realizowane w jednym ośrodku [KACPROWSKI,1977b; KACPROWSKI,1980; NOWAKOWSKA i in.,1989;1993]. Przyjęta metoda modelowania była oparta na wyznaczeniu funkcji transmitancji toru głosowego, określonej na podstawie pomiarów kinorentgenograficznych jego konfiguracji geometrycznej. Ten sposób modelowania toru głosowego był również kontynuowany i rozwijany w obecnej pracy. Jak już wspomniano, modelowanie toru głosowego dla głosek stacjonarnych w zasadzie nie stanowi problemu, choć jak wskazują niektóre prace [BAER i in.,1991], nawet i w tych przypadkach występują niekiedy rozbieżności między wynikami pomiarowymi uzyskanymi dla mowy rzeczywistej i sygnałem syntetycznym otrzymanym z modelu. Natomiast nadal trudnym zagadnieniem jest modelowanie przebiegów tranzjentowym, zwłaszcza w przypadku głosek, których nie sposób wymówić bez sąsiedztwa samogłoski (jak np. spółgłoski "j l, ł"), i stosunkowo mało jest publikacji poruszających ten temat (można tu wymienić jedynie pracę [NARAYAN i in.,1997]). Warto jednak zaznaczyć, że względu na różnice w wymowie trudno bezpośrednio przenosić na grunt języka polskiego wyniki otrzymane dla innych języków.

W pracy przedstawione zostaną podstawowe aspekty akustyczne mowy, zaś zagadnienia fonetyczne, fonologiczne i lingwistyczne zostaną uwzględnione tylko w takim stopniu, na ile te cechy strukturalne mowy znajdują swoje odzwierciedlenie w fizycznej strukturze sygnału mowy. Wszystkie omawiane zagadnienia związane z akustyką mowy będą oparte na przykładzie dźwięków mowy języka polskiego wypowiedzianych w tekstach mowy ciągłej, zarejestrowanych w bazie danych *BABEL* [GUBRYNOWICZ,1998;1999].

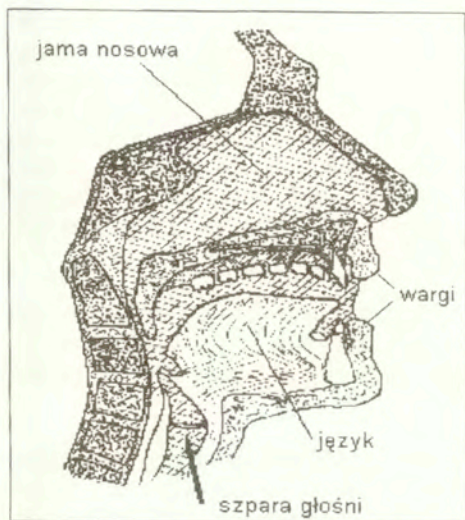
Omówiony zostanie układ artykulacyjny, jego budowa anatomiczna oraz funkcje jego poszczególnych elementów, a następnie będą podane zasadnicze informacje dotyczące zasobu dźwięków mowy polskiej, ich cech fonetyczno-akustycznych, klasyfikacji i podstawowych parametrów fizycznych stosowanych przy ich opisie. Istotną część pracy jest poświęcona teorii generacji akustycznego sygnału mowy, która stanowi podstawę przyjętej metody modelowania artykulacyjnego toru głosowego. W oparciu o powyższą teorię opisano zasady komputerowego modelowania układu artykulacyjnego mowy i wykorzystania modelowania artykulacyjnego do badań cech akustycznych określonych dźwięków mowy, bądź ich cech suprasegmentalnych, na przykład melodii mowy, akcentu itp. Przedstawione będzie opracowane w Pracowni Akustyki Mowy Zakładu Akustyki Cybernetycznej Instytutu Podstawowych Problemów Techniki PAN oryginalne oprogramowanie zastosowane w badaniach do modelowania funkcji transmitancji toru głosowego wyznaczonej dla dźwięków mowy zaliczanych do klasy rezonantów. Omówione będą otrzymane wyniki modelowania dla

tej klasy głosek o przebiegu ustalonym oraz niestacjonarnym, a także dla ich wybranych połączeń. W zakończeniu omówione będą aspekty praktyczne komputerowego modelowania toru głosowego, m.in. jako jednej z form poznawania cech fonetyczno-akustycznych dźwięków mowy.

## 2. Narząd artykulacyjny jako akustyczny układ generacji dźwięku.

### 2.1 Budowa anatomiczna organu mowy

Budowa anatomiczna organu mowy jest bardzo złożona zarówno pod względem strukturalnym, jak i czynnościowym, a jej szczegółowe opisy można znaleźć w licznych opracowaniach medycznych. Dlatego podany niżej opis anatomiczny zostanie ograniczony do najniezbędniejszego minimum, tak aby móc przedstawić funkcjonalne działanie układu artykulacyjnego jako swoistego rodzaju źródła akustycznego sygnału mowy.



Rys. 2. 1 Przekrój wzdłużny toru głosowego

części ustnej od przodu i z boków przez wargi i wewnętrzne ścianki policzków, od góry przez sklepienie jamy ustnej, na które składają się : szczcąka górna, podniebienie twarde i podniebienie miękkie, od dołu - przez okolice podjęzyczkową i powierzchnię języka. Ważny element w budowie toru ustno-gardłowego stanowią łuki zębowe. Krawędzie zębów przednich, siekaczy, biorą udział przy artykulacji tzw. głosek dentalizowanych. W dolnej, krtaniowej części toru ustno-gardłowego znajduje się zachełek krtaniowo-nagłośniowy, którego objętość zmienia się dla różnych dźwięków mowy. Na jej podstawie można wnioskować o pionowym położeniu krtani podczas fonacji. Jednakże jak stwierdzono, jego bezpośredni wpływ na kształtowanie artykułowanych dźwięków

Zasadniczą częścią układu artykulacyjnego bezpośrednio uczestniczącą w formowaniu dźwięków mowy jest tor gardłowo-ustno-nosowy rozciągający się od głośni aż do wylotu z ust na poziomie warg i/lub otworów nosowych. Na Rys.2.1 przedstawiono schematyczny przekrój wzdłużny toru głosowego, wraz z zaznaczeniem określonych wnek i niektórych struktur tworzących ich ściany. Wyodrębnia się w nim dwa oddzielne tory: tor gardłowo-ustny biegnący od głośni do wylotu z ust i obejmujący ponadgłośniową część krtani, jamę gardłową i ustną, oraz tor gardłowo-nosowy rozciągający się od głośni do otworów nosowych, obejmujący nadgłośniową część krtani, jamę gardłową i nosową. Tor gardłowo-ustny ograniczony jest w

mowy jest pomijalnie mały [BAER i in., 1991]. Znaczną część ścianek toru gardłowo-ustnego stanowią jednocześnie powierzchnie ruchomych członów narządu mowy, t.j. warg zmieniających wielkość otworu wylotowego ust, języka, podniebienia miękkiego, a także ruchomej zuchwy, toteż kształt i objętość toru artykulacyjnego ulegają podczas mówienia znacznym zmianom.

Oba odcinki torów mają wspólny segment początkowy i mogą przy wytwarzaniu dźwięku współpracować ze sobą, bądź funkcjonować osobno. Elementem sprzęgającym oba tory jest ruchomy języczek podniebienia miękkiego, będący fałdem mięśniowym, który podczas mówienia przywiera albo odchyła się od tylnej ścianki jamy gardłowej, sprzęgając obie części toru głosowego ze sobą, co następuje podczas artykulacji spółgłosek nosowych lub nazalizowanych. Kształt toru gardłowo-nosowego może ulegać tylko niewielkim zmianom, i wyłącznie w swej dolnej części, tam gdzie jedną z jego ścianek stanowi powierzchnia obsady języka. Dźwięki mowy powstają wówczas, gdy powietrze znajdujące się wewnątrz toru głosowego zostanie pobudzone do drgań przez jedno z dwóch (lub obu jednocześnie) źródeł fali akustycznej, które powstaje podczas przepływu powietrza z płuc. Energia potencjalna powietrza w płucach jest przetworzona na energię kinetyczną w płaszczyźnie drgających fałdów głosowych, lub w przedniej części toru głosowego, w miejscu w którym przepływ powietrza staje się turbulentny lub impulsowy. Powietrze to jest źródłem energii wytworzonego sygnału mowy.

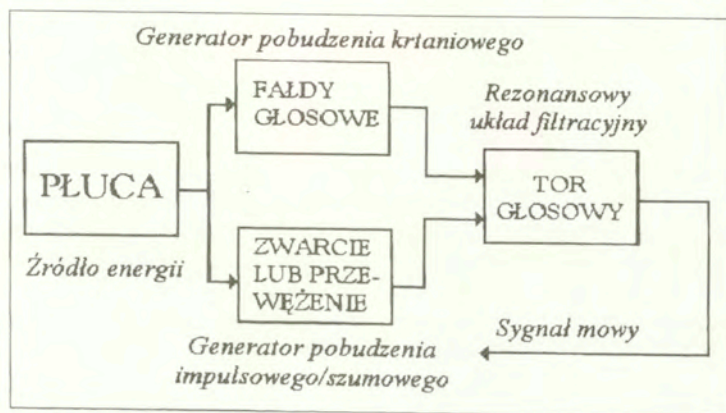
Struktury organu mowy, takie jak język, podniebienie miękkie oraz wargi, czy inne niekoniecznie ruchome, biorące udział w formowaniu dźwięków mowy, nazywane są artykulatorami. Ruchy takich artykulatorów jak na przykład języka, czy warg służą dwójkiemu celowi:

- a) do zmiany konfiguracji geometrycznej toru (przede wszystkim jego przekrojów poprzecznych i wzdłużnych, w mniejszym zaś stopniu jego długości), oraz
- b) do utworzenia ewentualnie innego źródła pobudzenia toru, znajdującego się powyżej fałdów głosowych.

Zmiany ukształtowania toru głosowego podczas mówienia pociągają za sobą przede wszystkim modyfikację charakterystyki jego transmitancji od źródła do wylotu ust i/lub nosa. Zmiany te odbywają się w stosunkowo dość płynny sposób, bez raptownych przejść od jednego stanu (konfiguracji toru) stacjonarnego do następnego. Jedynie podczas artykulacji głosek zwartych (jak /p,t,k/, czy zwarto-trących "c, cz" - /ts, tʃ/), które są wymawiane przy udziale przede wszystkim przedniej części języka odznaczającej się dużą ruchliwością. Otwarcie toru głosowego następuje z dużą prędkością, tak aby powstał impuls powietrza pobudzający do drgań powietrze znajdujące się w przedniej części toru, to znaczy między miejscem jego zamknięcia i wylotem z warg.

Jednakże, jak zobaczymy dalej, nie istnieje do końca jednoznaczna zależność między konfiguracjami przyjmowanymi przez tor głosowy oraz formowanym przez układ artykulacyjny przebiegami ciśnienia akustycznego, i różni mówcy mogą przekazywać ten sam komunikat za pomocą zróżnicowanych sekwencji kolejnych położeń artykulacyjnych. Fakt ten jest już od dość dawna dobrze znany z badań kinorentgenograficznych [por. LADEFOGED i in., 1972].

W celu przedstawienia organu mowy jako specyficznego źródła sygnału akustycznego, opis anatomiczny tego organu można znacznie uprościć, ograniczając się tylko do tych jego elementów, które są istotne z punktu widzenia pełnionych funkcji głosotwórczych i artykulacyjnych, tj. układu oddechowego, fonacyjnego i artykulacyjnego sensu stricto. Tego typu funkcjonalny schemat organu mowy został przedstawiony na Rys. 2.2.



Rys. 2.2 Funkcjonalny schemat organu mowy

W przypadku artykulacji głosek o pobudzeniu ciągłym, akustyczny sygnał mowy jest formowany poprzez pobudzenie do drgań powietrza, które znajduje się w torze głosowym. Powietrze jest pobudzane do drgań przez fałdy głosowe i/lub przez zawirowania jego strumienia powstające u wylotu przewężenia utworzonego ponad głośnią. W pierwszym przypadku, wypływające z płuc powietrze pobudza do drgań fałdy głosowe znajdujące się w krtani i wówczas pobudzenie ma charakter regularny, periodyczny. W drugim przypadku, w wyniku przepływu przez wąskie przewężenie tworzy się w części ustnej źródło szumowe o turbulentnym charakterze. Jak już wspomniano, odrębne źródło pobudzenia o charakterze impulsowym powstaje przy artykulacji głosek wybuchowych (np. /p,t,k/). Tak więc tor gardłowo-ustno-nosowy lub tylko jego odpowiednia część (powyżej miejsca przewężenia lub zwarcia) spełnia rolę rezonansowego układu filtracyjnego formującego obwiednię częstotliwościową sygnału pobudzającego.

## 2.2 Układ oddechowy.

Aparat oddechowy jest zbiornikiem mogącym u dorosłego człowieka pomieścić od 3.5 do 7 l powietrza. W czasie mówienia dorosły człowiek jednorazowo pobiera do płuc i wydycha z nich ok. 1.5 do 2 litrów powietrza [WIERZCHOWSKA,1980]. Wystarcza to przeciętnie na wymówienie ciągu głosek o łącznym czasie trwania od 1 do 3 sekund. W trakcie mówienia zadaniem układu oddechowego jest doprowadzenie do krtani i części ponadkrtaniowej układu artykulacyjnego strumienia powietrza o odpowiednio dużej energii kinetycznej, gdzie zostaje ona przekształcona na energię akustyczną formowanego sygnału mowy.

W przypadku artykulacji rezonantów (tj. samogłosek, spółgłosek płynnych, nosowych i drżących) fałdy głosowe o określonej masie, sprężystości i rezystancji strat są wprawiane w drgania wymuszone przez lokalne zmiany ciśnienia w głośni, wywołane przepływem powietrza do niej wypchanego z płuc przez kanały oskrzeli i tchawicę pod wpływem nacisku mięśni klatki piersiowej. Podczas artykulacji spółgłosek zwartych płuca wytwarzają nadciśnienia, tak aby w momencie szybkiego otwarcia toru głosowego powstał dostatecznie duży impuls powietrza.

Warto zaznaczyć, że praca płuc ma istotny wpływ na funkcjonowanie źródła krtaniowego, a tym samym na postać generowanego sygnału mowy. Często obserwuje się, że pod koniec wypowiedzania dłuższej sekwencji dźwięków z jednego wdechu, ciśnienie powietrza jest niewystarczające do podtrzymania regularnej pracy fałdów głosowych. Objawia się to krótkoterminową zmiennością okresu ich drgań, a także pojawieniem się dodatkowych impulsów. W wyniku tego głos jest odbierany jako bardziej chrypliwy i o niższej wysokości.

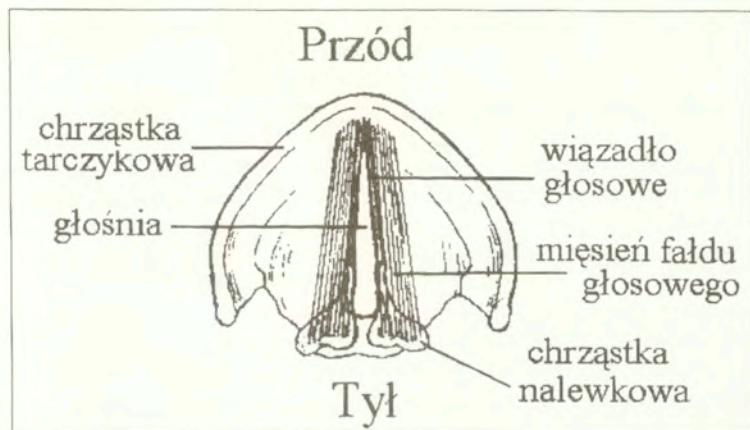
## 2.3 Układ fonacyjny

Dokładny opis budowy i działania źródła krtaniowego będzie przedstawiony dalej, w rozdziale poświęcony modelowaniu jego funkcji (patrz rozdz. 6). Obecnie zostaną podane jedynie podstawowe informacje umożliwiające podanie tylko ogólnych zasad formowania sygnału mowy w układzie artykulacyjnym człowieka. Jak już wcześniej wspomniano, dźwięczne pobudzenie krtaniowe jest najczęstszym typem fonacji, na przykład w przypadku mowy polskiej, ponad 3/5 głosek jest artykułowane przy pobudzeniu dźwięcznym toru głosowego. Dźwięczność sygnału lub jej lub brak ma znaczenie fonetyczne, a płynne zmiany częstotliwości i amplitudy pobudzenia wpływają na percepcję zmian wysokości i intensywności w sygnale mowy. Te właśnie zmiany w czasie decydują w dużej mierze o percepcji melodii (akcentu) wypowiedzi.

Obserwacja działania organów artykulacyjnych w sposób nie zakłócający wymawianie poszczególnych dźwięków mowy jest niezbędna dla poznania i

zrozumienia mechanizmów będących u podłoża samej artykulacji. Dzięki rozwojowi w ostatnich latach nieinwazyjnych metod obserwacji i rejestracji ruchów poszczególnych elementów układu artykulacyjnego (szczególnie metody stroboskopowe, szybkiej kinorentgenografii, ultradźwiękowe metody Dopplerowskie itp.) umożliwiły szczegółowe poznanie zarówno struktury, jak i mechanizmu działania poszczególnych organów. Szczególnie w przypadku źródła krtańowego, trudno dostępnego do bezpośredniej obserwacji metody te umożliwiły bardziej szczegółowe poznanie mechanizmu jego działania i opracowania odpowiedniego modelu. Szczególnie efektywna okazała się metoda elektrolaryngograficzna opracowana przez Fourcina i Abbertona [FOURCIN&ABBERTON,1971;1977]

Powietrze wypełniające kanał głosowy, jest podczas mówienia pobudzane do drgań przez wibrujące wiązadła głosowe, będące pasemkami tkanki łącznej, stanowiącymi krawędzie fałdów głosowych (Rys. 2.3). Fałdy te znajdują się na połowie wysokości krtani i składają się z mięśni i wiązań mających przyczepy przy chrząstkach, tworzących jej szkielet. Przebiegają one poziomo, od wewnętrznej powierzchni chrząstki tarczowej do wyrostków ruchomych chrząstek nalewkowych osadzonych na górnej krawędzi płytki chrząstki pierścieniowej. Przestrzeń utworzona między wiązadłami głosowymi nazywa się głośnia.



Rys. 2.3 Widok z góry na fałdy głosowe (wg [BORDEN&HARRIS,1984])

Podczas swobodnego oddychania głośnia ma kształt owalny, tak że powietrze przepływa przez nią w sposób niezakłócony, niezależnie od kierunku strumienia (do czy z płuc).. Natomiast w czasie mówienia fałdy głosowe są zbliżone do siebie na pewną odległość i znajdują się pod oddziaływaniem dwóch sił. Jedna powstaje w

wyniku zwiększonego oporu przepływu głośni dla strumienia powietrza idącego z płuc. Objawia się to przede wszystkim wzrostem ciśnienia podgłośniowego powyżej pewnej krytycznej wartości, przy której wiązadła zaczynają drgać. Dalsze zmiany ciśnienia podgłośniowego wpływają na intensywność pobudzenia dźwięcznego toru głosowego, w przybliżeniu proporcjonalnie do 3-4 potęgi tego ciśnienia [BROAD, 1973]. Oznacza to, że gdy ciśnienie podgłośniowe ulegnie 2-krotnemu wzrostowi, to intensywność pobudzenia dźwięcznego wzrośnie 8-16 razy, czyli o 9-12 dB. Tak więc już stosunkowo małe zmiany ciśnienia podgłośniowego wpływają w istotny sposób na jakość artykułowanego dźwięku. Zaś przy już bardzo dużym ciśnieniu podgłośniowym, jak to ma miejsce przy zwiększonym wysiłku głosowym, czy wręcz krzyku, następuje charakterystyczna zmiana barwy głosu. Jest to tzw. efekt Lombarda [LOMBARD, 1911]. Jednocześnie oprócz głośności, zmiany te mogą modyfikować, częstotliwość drgań fałdów głosowych.

Podczas procesu zwanego fonacją, odpowiednio napięte fałdy głosowe, rytmicznie otwierają się i zamykają, modulując w ten sposób strumień powietrza wypływającego z płuc. W momencie zamknięcia głośni, powietrze z płuc gromadzi się pod zamkniętymi fałdami powodując tworzenie się na nich coraz to większego nacisku. Gdy nacisk ten staje się większy od siły, z którą one zwierają się, następuje ich chwilowe rozchylenie, co umożliwia wpłynięciu do gardła niewielkiej porcji nagromadzonego pod głośnią powietrza. Nacisk powietrza na fałdy głosowe staje się wówczas mniejszy, toteż wiązadła głosowe mogą się ponownie zsunąć, głównie wskutek powstania podciśnienia w płaszczyźnie otworu głośni. Zamykanie się, lub otwieranie fałdów jest kontrolowane głównie przez mięśnie krtani. Quasi-periodyczne drgania fałdów głosowych powodują zmiany powierzchni, a więc i akustycznej impedancji otworu głośni, kształtując w ten sposób strumień przepływającego przez niego ośrodka, który przybiera postać ciągu dyskretnych impulsów. Schematycznie opisany proces powtarza się regularnie - w mowie od kilkudziesięciu do kilkuset, a w śpiewie nawet do tysiąca razy na sekundę. Szczegółowa analiza i modelowanie fizycznego działania źródła krtaniowego przedstawiono w Rozdziale 6.

Zasadniczym czynnikiem wpływającym na częstotliwość podstawową (inaczej mówiąc na wysokość głosu) są zmiany zachodzące na poziomie głośni, a dokładniej, w samych fałdach, ich rozmiarów, naprężenia, długości utworzonej przez nie szczeliny itp. Parametry fizyczne wiązadeł, które zmieniają się płynnie podczas mówienia wpływają na akustyczne charakterystyki głosu. Pierwszą z nich, która decyduje o intonacji, jest częstotliwość ich drgań. Zależy ona od czynnej długości fałdów, ich grubości, efektywnej masy drgającej i ich napięcia. Generalna zasada według której można przedstawić wpływ wyżej wymienionych czynników na częstotliwość podstawową wyraża się w taki sposób, że każda ich zmiana powodująca wzrost lub zmniejszenie szybkości ruchu wiązadeł powoduje odpowiednio wzrost lub zmniejszenie częstotliwości podstawowej.



Oprócz opisanych wyżej ruchów odbywających się na poziomie szparygłośni istnieje ruch samej krtani jako całości. Może być ona silnie podciągana ku górze w chwili przetykania, potem opuszczana. W czasie mówienia ruchy krtani są mniej energiczne lecz nieustannie zachodzą. W zależności od wymawianych głosek, od wysokości tonu krtaniowego oraz od wysiłku głosowego ruchy krtani odbywają się w różnych płaszczyznach, ku górze, ku dołowi, do przodu i do tyłu. Ruchy krtani w kilku płaszczyznach okazują się niezbędne podczas mówienia na równi z pracą fałdów głosowych. Niektórzy twierdzą, że ruchy krtani są najważniejsze w czasie fonacji i jako dowód podają szept, który odbywa się bez dźwięczności, ale bez ruchu krtani odbyć się nie może [DŁUSKA, 1981; WIERZCHOWSKA, 1980].

#### 2.4. Szumowe pobudzenie toru głosowego

W przypadku artykulacji głosek bezdźwięcznych, a także głosek o mieszanej strukturze dźwięczno-aperiodycznej, występuje odrębne źródło o szumowym lub impulsowym charakterze, i jest zlokalizowane na ogół w części ustnej toru głosowego. Źródło to powstaje bądź w pobliżu przewężenia utworzonego między językiem i podniebieniem (lub dziąsłami czy zębami), u wylotu którego przepływ powietrza płynącego z płuc staje się turbulentny. Powstałe w ten sposób zawirowania powietrza są źródłem pobudzenia szumowego odcinka toru głosowego zawartego między wylotem z przewężenia i wargami.

Warto jednak zaznaczyć, że mogą wystąpić przypadki, w których źródło krtaniowe może również funkcjonować jako źródło o charakterze szumowym, choć w mowie polskiej nie ma dźwięków, dla których tego rodzaju umiejscowienie źródła jest charakterystyczne. Wyróżnia się następujące rodzaje artykulacji, dla których źródło szumowe jest umiejscowione w krtani:

- a) Mowa szeptana – szum powstaje, gdy fałdy głosowe są zbliżone do siebie na tyle blisko, że w momencie osiągnięcia dostatecznie dużej prędkości przez strumień powietrza przepływającego między nimi przestaje być laminarny, a jednocześnie fałdy nie zostają pobudzone do drgań. Podobnego typu pobudzenie występuje często w różnego przypadkach schorzeń głosu, zwłaszcza porażen fałdów głosowych, które choć mogą zbliżyć się do siebie, nie są w stanie drgać. Warto zaznaczyć, że i w mowie naturalnej (zwłaszcza spontanicznej) może powstać chwilowe pobudzenie szumowe umiejscowione w głośni, przy przejściu między głoskami o mocno odmiennej artykulacji.
- b) Mowa sapaćca – szumowi powstającemu w głośni towarzyszy jednocześnie pobudzenie dźwięczne. Niedomykające się, drgające fałdy głosowe modulują przepływający między nimi strumień powietrza. Tego typu pobudzenie nazywane jest niekiedy dźwięcznością przydechową, która również może występować w głosie normalnym, przy przechodzeniu od pobudzenia dźwięcznego do szumowego. Wówczas w zakresie wyższych częstotliwości

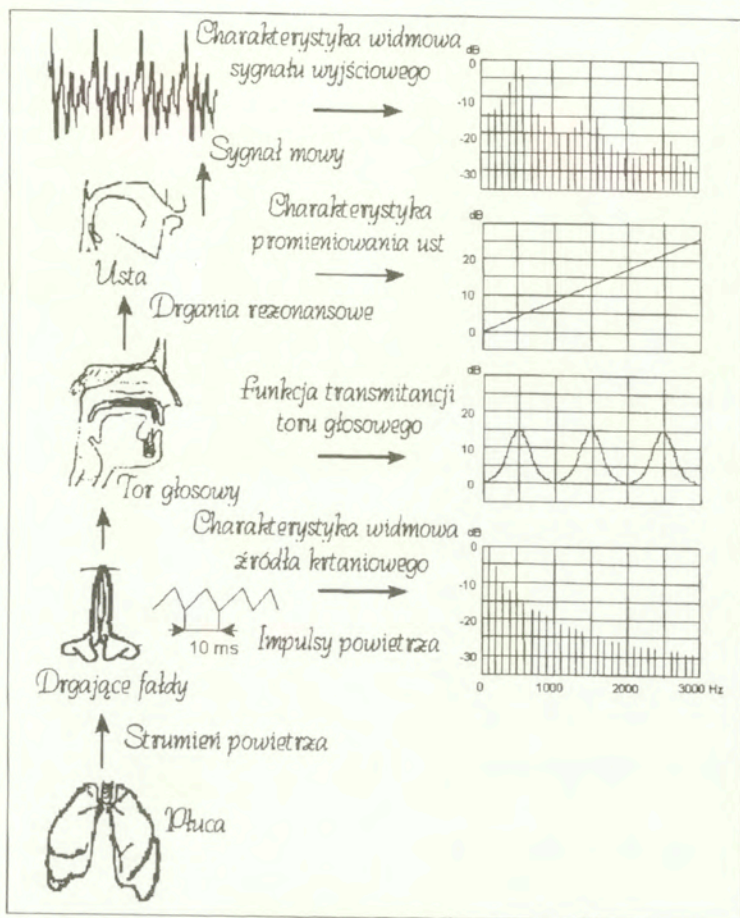
składowe harmoniczne są o znacznie mniejszej energii, niż przy normalnym pobudzeniu dźwięcznym.

## 2.5 Mechanizm generacji sygnału mowy

Ogólnie biorąc, podczas generacji sygnału mowy mogą powstać trojkiego rodzaju źródła pobudzenia toru gardłowo-ustnego, różnie w nim usytuowane i decydujące w zasadniczy sposób o fizycznej strukturze sygnału wyjściowego. Pierwszym z nich jest tzw. źródło krtaniowe położone na poziomie fałdów głosowych i uczestniczące przy artykulacji wszystkich głosek dźwięcznych. Drugi typ źródła powstaje w części ponadkrtaniowej, w wyniku powstania turbulentnego przepływu strumienia powietrza, tworzącego się za bardzo silnymi przewężeniami, formowanymi podczas artykulacji spółgłosek takich jak "s", "sz" itp. Źródło to ma charakter szumowy i jego położenie w torze gardłowo-ustnym zależy od typu artykułowanej spółgłoski. Dla głosek bezdźwięcznych języka polskiego, przepływ powietrza przez krtani odbywa się w sposób niezaburzony i źródło pobudzenia toru głosowego znajduje się na ogół w znacznej odległości od jego części krtaniowej. Trzeci typ źródła pobudzającego może powstać w miejscu chwilowego zamknięcia toru artykulacyjnego i ma charakter impulsowy. Tworzy się ono w wyniku jego nagłego otwarcia i wypuszczenia w bardzo krótkim czasie pewnej porcji powietrza. Te trzy źródła mogą występować oddzielnie przy artykulacji określonych głosek, równocześnie lub sekwencyjnie w parach. W drugim przypadku, przykładem jednoczesnego udziału, położonych w różnych miejscach toru głosowego źródeł krtaniowego i szumowego w artykulacji jest wymowa takich spółgłosek jak "ż", "ź". W ostatnim zaś przypadku, przykładem sekwencyjnego powstawania źródeł pobudzających tor głosowy może być artykulacja takich głosek jak "cz", "ć", w których na początku pojawia się krótki impuls powstający wskutek nagłego rozwarcia toru głosowego zamkniętego przez język w części ząbówowej, lecz na tyle małego, że po wygenerowaniu impulsu powstają zawirowania wypływającego powietrza za utworzoną szczeliną, w wyniku czego zostaje wygenerowany szum.

Na Rys. 2.4 przedstawiono dla głosek dźwięcznych poglądowy schemat przebiegu formowania w układzie artykulacyjnym akustycznego sygnału mowy. W lewej kolumnie idąc od dołu przedstawiono kolejno elementy układu uczestniczące w formowaniu sygnału, podając w prawej kolumnie ich charakterystykę akustyczną w płaszczyźnie częstotliwościowej. Schemat ten zakłada, że w zakresie rozpatrywanych częstotliwości układ artykulacyjny jest układem liniowym, zaś fala akustyczna rozchodząca się w torze głosowym jest falą płaską, czyli dla górnej częstotliwości przebiegu akustycznego wymiary poprzeczne toru są znacznie mniejsze od długości fali. Ten ostatni warunek jest spełniany dla częstotliwości nie większych niż 4000-4500 Hz.

W przypadku artykulacji pozostałych głosek ogólny schemat pozostaje ten sam, zmianie ulega charakterystyka źródła pobudzającego, jego miejsce położenia oraz odcinek toru głosowego uczestniczący bezpośrednio w formowaniu sygnału. Na przykład, przy artykulacji głosek bezdźwięcznych (szczelinowych) można przyjąć, że ze względu na dużą impedancję wewnętrzną źródła szumowego, odcinek toru głosowego zawarty między krtanią i miejscem zwarcia (szczeliną) nie uczestniczy w formowaniu sygnału mowy, a tylko przednia część toru zawarta między szczeliną i otworem wylotowym utworzonym przez wargi.



Rys. 2.4 Schemat formowania akustycznego sygnału mowy w układzie artykulacyjnym człowieka w przypadku głosek o pobudzeniu krtaniowym.

Na przedstawionym schemacie generacji sygnału mowy przyjęto w pewnym uproszczeniu piłokształtną postać przebiegu pobudzenia krtaniowego. Oznacza to, że czas podczas którego głośnia jest zamknięta jest pomijalnie mały, w stosunku do czasu rozwierania i zwierania fałdów. I choć odbiega to od rzeczywistego cyklu ruchu fałd, to widmo piłokształtnego przebiegu jest bardzo zbliżone do rzeczywistego widma tonu krtaniowego, które jest źródłem harmonicznym i ma opadającą obwiednię w stopniu  $-12$  dB/okt, tak jak to przedstawiono na rysunku. Sygnał ten podlega następnie filtracji w torze głosowym, który w przypadku modelowania go za pomocą rury cylindrycznej jest selektywnie wzmacniany wokół kolejnych częstotliwości rezonansowych (w pierwszym przybliżeniu, przy cylindrycznym przekroju toru głosowego - ćwierćfalowych), dla których powstają wewnątrz toru głosowego fale stojące odpowiednich rzędów. Stąd w widmie sygnału obserwuje się występowanie charakterystycznych maksimów nieco przesuniętych względem częstotliwości rezonansowych w kierunku wyższych częstotliwości, co jest wynikiem opadającej obwiedni tonu źródła tonu krtaniowego, nieco skorygowanej przez wznoszącą się charakterystykę promieniowania ust ( $+6$  dB/okt). Maksima te zwane formantami są jednymi z najpowszechniej stosowanymi w opisie dźwięków mowy parametrami fonetyczno-akustycznymi. Ten liniowy model generacji sygnału mowy jest u podstaw prowadzonych w dalszych rozdziałach rozważań. Warto zauważyć, że procesowi generacji fali dźwiękowej towarzyszy jednocześnie przepływ powietrza przez tor głosowy, jednakże o średniej prędkości znacznie mniejszej od prędkości propagacji fali akustycznej (wynosi ona w przybliżeniu ok.  $350$  m/s). Tak więc wpływ ruchu powietrza na propagację fali dźwiękowej jest pomijalnie mały.

### 3. Artykulacja dźwięków mowy

Zmienna podczas mówienia konfiguracja geometryczna części gardłowo-ustnej toru głosowego jest wyznaczana przede wszystkim przez układ ruchomych artykulatorów takich jak język, wargi, szczeka, języczek podniebienia miękkiego (velum), kość gnykowa itp. Część nosowa w zasadzie nie zmienia swego ukształtowania podczas artykulacji. Jedynie przy artykulacji spółgłosek nosowych i głosek nazalizowanych zmienia się kształt wielkość wlotu do jamy nosowej wyznaczonego przez velum.

Najbardziej dynamicznym elementem kształtującym konfigurację jamy ustnej podczas mówienia jest język. Zmiana jej konfiguracji pociąga za sobą w pierwszym rzędzie zmianę rozkładu częstotliwości rezonansowych (tj. biegunów) i zer funkcji transmitancji toru głosowego, a tym samym powoduje określone brzmienie generowanych dźwięków mowy. Klasyfikowanie dźwięków mowy, a przede wszystkim ich opis, jest oparte na analizie odpowiadających im przekrojów wzdłużnych toru głosowego oraz rodzaju jego pobudzenia. Na Rys. 3.1 przedstawiono zasadnicze typy ułożenia masy języka dla samogłosek, spółgłosek wybuchowych, trących i płynnych języka polskiego (przekroje opracowane na podstawie danych z [WIERZCHOWSKA, 1980]).

Zgodnie z tradycją fonetyczną, dźwięki mowy opisywane są najczęściej pośrednio, poprzez określanie niezbędnych do ich wytworzenia elementów lub ruchów narządu mowy. Tego typu opis, nazywany artykulacyjnym, jest trudnym do sformalizowania i przedstawienia w zwartej postaci. Koncentruje się on przede wszystkim na opisie ułożenia masy języka, a w przypadku spółgłosek określa się je najczęściej przez podanie ich tzw. miejsca artykulacji, to znaczy miejsca położenia zwarcia lub przewężenia utworzonego w części gardłowo-ustnej toru głosowego. W przypadku głosek bezdźwięcznych lub głosek o pobudzeniu mieszanym jest to miejsce, w którym powstaje w torze głosowym inne niż kraniowe (szumowe lub impulsowe) lub dodatkowe źródło pobudzenia. W przypadku samogłosek wyróżnia się trzy zasadnicze układy masy języka: położenie przednie, płaskie (centralne) oraz położenie tylne ([DŁUSKA, 1981]). Rozróżnienie to jest oparte na wyróżnieniu tej części masy języka, która jest wypiętrzona ku górze podczas artykulacji danej samogłoski. Jedynie w położeniu płaskim masa języka jest ułożona w miarę równomiernie wzdłuż całej jego długości.

Innym ważnym elementem opisu jest charakterystyka tzw. sposobu artykulacji, to jest całości kształtu warunków, w jakich dochodzi do generacji określonego dźwięku mowy, a zwłaszcza warunków przepływu powietrza przez przewężenie lub

przewycięzania zwarcia w części gardłowo-ustnej toru głosowego. Wyróżnia się dwójakiego rodzaju przewężenia:

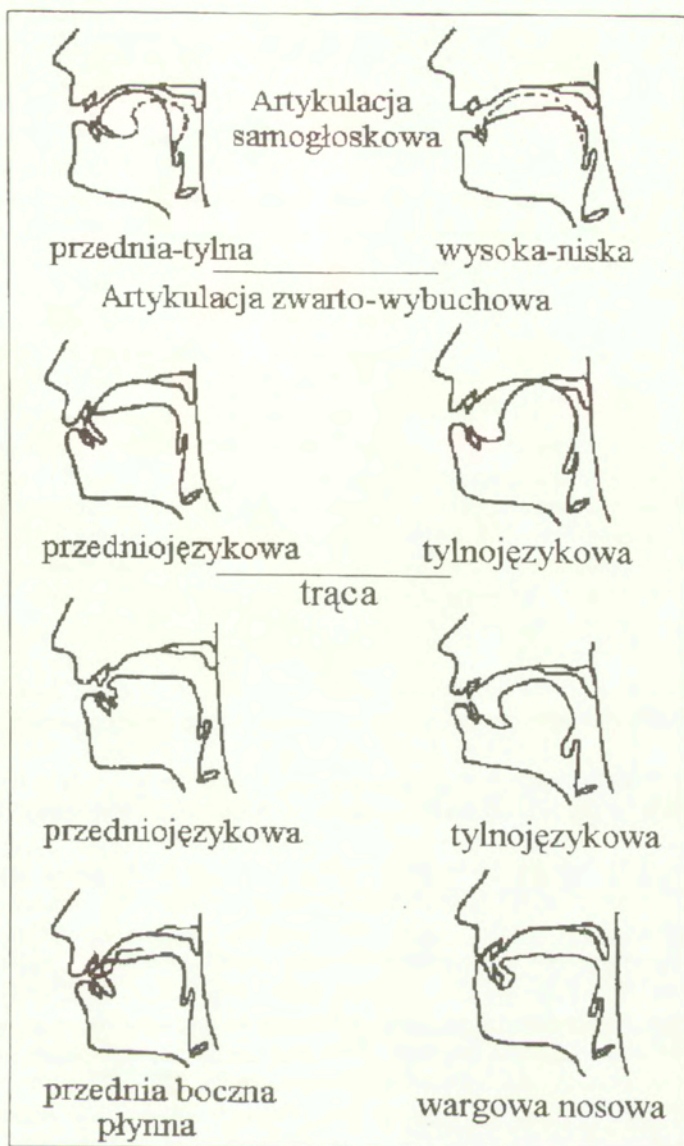
- a) otwór o dostatecznie dużej powierzchni przekroju, który pozwala na swobodny, niezaburzony (laminarny) przepływ powietrza, oraz
- b) szczelina, na brzegach której, u wylotu, następuje powstanie zawirowań (turbulencji) w przepływającym strumieniu powietrza, będącymi źródłami szumów.

W przypadku zwarcia następuje całkowite, chwilowe zamknięcie przepływu powietrza przez środkową płaszczyznę toru gardłowo-ustnego i kończy się ono zazwyczaj szybkim rozwarciem narządów mowy, w wyniku którego powstaje krótkotrwały impuls ciśnienia akustycznego (tzw. płożja), po którym następuje nieco od niego dłuższy segment szumowy (rzędu kilku-kilkunastu ms).

Ze względu na sposób artykulacji głoski języka polskiego dzielą się na: a) otwarte, b) otwarte nosowe (nazalizowane), c) szczelinowe, d) zwarto-wybuchowe, e) zwarto-szczelinowe, f) zwarte nosowe, g) drżące i h) boczne.

Głoski otwarte charakteryzują się tym, że wymawiane są bez zwarć i szczelin w środkowej części toru gardłowo-ustnego, nie powstają zatem przy ich artykulacji dodatkowe źródła drgań nieregularnych (szумы i płożje). Do tej grupy głosek zaliczane są przede wszystkim samogłoski oraz takie spółgłoski jak /j/, zwane często półotwartymi (lub płynnymi).

Klasa głosek otwartych nosowych obejmuje przede wszystkim samogłoski nosowe "e" i "a", których artykulacja odznacza się tym, że obok toru gardłowo-ustnego w ich formowaniu uczestniczy także tor gardłowo-nosowy. Głoski te w zasadzie są realizowane przy jednoczesnym, na ogół jednak asynchronicznym, udziale obu części toru głosowego, przy czym promieniowanie sygnału do otoczenia odbywa się zarówno przez usta jak i przez nozdrza. Na ogół, tor nosowy włącza się z pewnym opóźnieniem, w wyniku czego samogłoska nosowa jest realizowana w postaci dyftongu składającego się z dwóch części, pierwszej odpowiadającej danej samogłosce ustnej (/o/, /e/), po której następuje obniżenie się podniebienia miękkiego przy jednoczesnym zamykaniu się toru ustnego, w wyniku czego brzmienie pod koniec samogłoski nosowej staje się podobne do /n/.



Rys. 3.1 Niektóre wybrane konfiguracje toru głosowego dla głosek polskich (adaptowane z danych w: [WIERZCHOWSKA,1980])

Podczas artykulacji głosek szczelinowych, takich jak /s, ʃ, c/ (trące) powstaje w torze głosowym przewężenie, u wylotu którego powstają turbulencje powietrza. Przewężenie to musi być o dostatecznie małym przekroju, tak by prędkość wypływającego ze szczeliny powietrza osiągnęła odpowiednią prędkość. Z akustycznego punktu widzenia wylot szczeliny staje się źródłem pobudzenia szumowego i jego położenie w torze głosowym decyduje o tym jaka część toru ustnego uczestniczy w formowaniu dźwięku mowy. W pierwszym przybliżeniu można bowiem przyjąć, że część toru między głośnią i szczeliną nie uczestniczy w formowaniu dźwięku, wskutek bardzo dużej oporności wewnętrznej wylotu szczeliny w stosunku do oporności wejściowej przedniej części toru.

Zwarcie właściwe głoskom zwarto-trącym ( $\widehat{ts}$ ,  $\widehat{tʃ}$ ,  $\widehat{tc}$ ) kończy się zazwyczaj słabą plozją, po której bezpośrednio w środkowej płaszczyźnie narządów mowy powstaje szczelina analogiczna do szczeliny typowej dla odpowiedniej głoski trącej (np. szczelina w przypadku głoski  $\widehat{ts}$  ("c") jest taka sama jak dla /s/. W głosce zwarto-trącej o pobudzeniu mieszanym (takiej jak np. /dz/) faza początkowa artykulacji zaczyna się od momentu, w którym rozbrzmiewa stłumiony ton krtoniowy, a następnie podczas następnego fazy jest on modulowany w kanale gardłowo-ustnym. Głoski zwarto-trące, podobnie jak zwarto-wybuchowe, są silnie niestacjonarne. Przy ich przedłużonej artykulacji przekształcają się w odpowiednie głoski trące (np. "c"  $\widehat{ts}$  w "s" /s/, "cz"  $\widehat{tʃ}$  w "sz" /ʃ/, itp.)

Szczególnego rodzaju artykulacja ma miejsce w przypadku głosek nosowych, które są formowane przez część gardłowo-nosową o niemal stałej konfiguracji, oraz równolegle dołączonej do niej zamkniętej części ustnej o odpowiednio dobieranej długości. Udział części ustnej objawia się odpowiednią modyfikacją funkcji transmitancji, o czym szerzej będzie mowa w rozdziale 5. Wyjściowy sygnał jest promieniowany tylko przez nozdrza. Ponieważ nie ma w nim znacznie większych przewężeń, powoduje to, że głoski te są bardziej zbliżone do głosek otwartych, z punktu widzenia swobody przepływu powietrza od krtani do otoczenia, niż do spółgłosek zwartych. Jak dalej to zobaczymy fakt ten ma istotne znaczenie w przypadku analizy cech akustycznych tej klasy głosek, które raczej przemawiają za tym by spółgłoski nosowe zaliczyć do głosek otwartych, a nie zwartych, mimo zamkniętego podczas ich artykulacji toru gardłowo-ustnego. Ponieważ artykulacja spółgłosek nosowych jest zupełnie odmienna od artykulacji głosek otwartych, niekiedy wyodrębnia się klasę głosek spółotwartych, obejmującą również głoski boczne i drżące [WIERZCHOWSKA, 1980]. Głoski te są zawsze artykułowane przy udziale źródła krtoniowego.

Specyficzną głoską w języku polskim jest głoska drżąca /r/. Podczas jej artykulacji zmiany konfiguracji toru głosowego mają dość złożony przebieg, w



wyniku którego powstaje krótkotrwałe drganie jednego z odpowiednio napiętych ruchomych członów narządu artykulacyjnego - np. czubka języka (apex), czy koniuszka podniebienia miękkiego - uvula (jak to ma miejsce w standartowym francuskim /r/ roulé). Drgający człon jednorazowo lub kilkakrotnie zwiera się z odpowiednim odcinkiem podniebienia (w przypadku /r/ roulé z tylną powierzchnią języka), lub niekiedy zbliża się do niego na taką odległość, że następuje znaczne zwiększenie oporów przepływu powietrza w jamie ustnej. W obu przypadkach powstają w kanale głosowym szybkie i krótkotrwałe zmiany ciśnienia, których efektem są następujące po sobie impulsy o charakterze wokalicznym rozdzielone odcinkami o bardzo małym poziomie ciśnienia.

Bardzo specyficzne ukształtowanie toru głosowego powstaje przy artykulacji głoski bocznej /l/. Odznacza się ona tym, że jest wymawiana przy częściowym zwarciu w środkowej płaszczyźnie toru głosowego, zaś po bokach jamy ustnej powstają otwory, umożliwiające swobodny przepływ przez nią powietrza. W czasie zwarcia rozbrzmiewa lekko stłumiony ton krtaniowy modulowany w torze gardłowo-ustnym.

Podana wyżej klasyfikacja głosek ze względu na sposób artykulacji nie jest jedyną z możliwych, i w ciągu szeregu lat fonetycy przedstawiali wiele wariantów powyższego schematu klasyfikacyjnego. Jednakże, wydaje się, że przedstawiony schemat uwzględniający trzy stopnie zbliżenia narządów mowy (otwór, szczelina i zwarcie [WIERZCHOWSKA, 1980]) i jednocześnie wyróżniający człony narządu mowy uczestniczące w formowaniu dźwięku (a więc tym samym określający sposób artykulacji) jest chyba najbardziej spójny i jednocześnie dający się, jak to dalej zobaczymy, dość jednoznacznie opisać na płaszczyźnie cech akustycznych, co jest niezbędnym warunkiem przy opracowywaniu systemu modelowania układu artykulacyjnego człowieka.

Miejsce artykulacji umożliwia dokonanie dalszej, bardziej szczegółowej klasyfikacji głosek. Ponieważ położenie jego w jamie ustnej w znacznym stopniu zależy od układu masy języka, można przyjąć dla wszystkich głosek, zarówno dla otwartych jak i pozostałych, wspólne kryterium oparte na ocenie rozkładu masy języka. Jak dalej przekonamy się również i to kryterium znajduje w miarę jednoznaczne odwzorowanie na płaszczyźnie cech akustycznych. Do szczegółowego opisu miejsca artykulacji fonetycy określają położenie najwyższego punktu masy języka (tzw. wysklepienia masy języka) względem górnej powierzchni jamy ustnej podając względem jakiego jej odcinka występuje maksymalne zbliżenie. Podaje się więc, że dana głoska jest wargowa, czy zębowa, podniebienna (przednia, tylna) itp.

W ciągu szeregu lat przyjmowano na ogół, że każda fraza wypowiedzi stanowi w swoisty sposób zorganizowaną całość, przejawiającą się zarówno w jej strukturze segmentalnej (głoskowej) i sylabicznej, jak też w jej rozczłonkowaniu (podziale) rytmicznym i melodycznym [WIERZCHOWSKA, 1980]. W każdej frazie wyróżnia się jej początek poprzedzony przez pauzę (nie należy mylić jej ze zwarcim bezdźwięcznym, stanowiącym element głoski) zwany nagłosem, część środkową zwaną śródgłosem i część końcową (po której następuje pauza) zwaną wygłosem. Nagłos obejmuje wstępną fazę, podczas której następuje gotowość narządu artykulacyjnego do mówienia, a następnie, przejście do układu właściwego głosce wymawianej w nagłosie wypowiedzi. Wygłos wypowiedzi kończy się powrotem narządów mowy do położenia oddechowego. Warto podkreślić, że szczególnie w wygłosie wypowiedzi ruchy narządów mowy są wykonywane znacznie mniej dokładnie, z mniejszym nakładem energii, a także wolniej niż w nagłosie i śródgłosie. Przejawia się to przede wszystkim w osłabianiu wygłosowych zwarć (czy wręcz ich nie realizowaniu), w redukcji (zacierania się różnic) wygłosowych głosek otwartych, w zmniejszaniu się częstotliwości drgań strun głosowych, a także we wcześniejszym wygazaniu ich wibracji.

Słabość wygłosu pociąga za sobą bezdźwięczność wymawianych przed pauzą dźwięcznych głosek zwarto-wybuchowych, zwarto-szczelinowych i szczelinowych. W śródgłosie występuje co najmniej jedna głoska, na którą przypada tzw. szczyt głoski, to jest taka faza artykulacji, w której narządy mowy utrzymują się przez pewien czas w pozycji niezbędnej do wytworzenia tej głoski. W wypowiedziach, w których można wyróżnić więcej niż jedną głoskę, między poszczególnymi dźwiękami obserwuje się na ogół tzw. stany przejściowe. W tych momentach narządy mowy zmieniają układ z właściwego głosce poprzedzającej na układ właściwy głosce następującej. Wówczas momenty szczytowe przypadają na centralne fazy kolejno wymawianych głosek.

Podczas wymawiania ciągu głosek, jednym ze zjawisk silnie wpływającym na zmienność układów narządów artykulacyjnych właściwych poszczególnym głoskom jest koartykulacja. Jest ona spowodowana nakładaniem się ruchów artykulacyjnych właściwych sąsiadującym ze sobą głoskom. Na ogół występuje wcześniejsze przygotowanie się ruchomych narządów mowy, przede wszystkim masy języka i warg, do wymówienia następnej głoski. Na przykład, zaokrąglenie wargowe charakteryzujące samogłoskę /u/ zaczyna się zazwyczaj już podczas wymawiania dźwięków poprzedzających tę głoskę. Jednakże warto tu podkreślić, że wyprzedzenie to nie jest dowolne w czasie, lecz ma określoną wartość i jest ściśle przyporządkowane pozostałym ruchom artykulacyjnym wykonywanym podczas wymawiania tej samogłoski. Podobnie, podczas wymawiania dźwięków poprzedzających samogłoskę /i/ występuje tendencja do przybierania przez język przedniego położenia charakteryzującego tę głoskę. Według tradycyjnego interpretowania koartykulacji

zjawisko to traktuje się jako wynik wzajemnego oddziaływania na siebie poszczególnych jednostek fonetyczno-akustycznych. Zgodnie z tym ujęciem, każdą wypowiedź można przedstawić na płaszczyźnie fonologicznej w postaci konkatencji segmentów kanonicznych o określonych położeniach docelowych układu artykulacyjnego.

Wynikiem koartykulacji jest wzajemne oddziaływanie tych segmentów na siebie poprzez mechanizmy, których natury nie znamy, lecz manifestujące się poprzez "rozkładanie się" cech (np. zaokrąglenie wargowe, nazalizacja, dźwięczność itp.), to znaczy cechy typowe dla jednego segmentu odnajduje się w sąsiadujących z nim, a i niekiedy w dalszych. Postać kanoniczna segmentu określona za pomocą zbioru cech artykulacyjnych (lub innych) może być traktowana jako jego forma idealna, czy wręcz docelowa. Ciąg takich idealnych układów artykulacyjnych może ulegać wzajemnemu "dopasowywaniu się" poprzez fragmenty o charakterze przejściowym. Z drugiej strony, segmenty wypowiedziane w ciągu mogą "dopasowywać się" jeden do drugiego, aby zminimalizować amplitudę przejść. Tak więc przy takim ujęciu koartykulacji, jej głównym celem jest jak najbardziej możliwe zmniejszenie wysiłku artykulacyjnego, dzięki płynniejszym przejściom między kolejnymi głoskami.

Szczegółowe badania nad koartykulacją spowodowały, że obecnie odchodzi się coraz bardziej od takiego ujmowania tego zjawiska, a w szczególności od przedstawiania wypowiedzi na poziomie fonetyczno-akustycznym, w postaci ciągu jednostek segmentalnych oddziaływujących na siebie na poziomie artykulacyjnym. Coraz powszechniej przyjmuje się, że koartykulacja jest procesem na poziomie układu nerwowego i neuromuskularnego [FOWLER,1979; MARCHAL,1983], gdyż zdecydowanie wyraźnej struktury segmentalnej sygnału mowy nie obserwuje się ani na poziomie artykulacyjnym, ani na poziomie akustycznym. Jedyne co można stwierdzić, to tylko płynnie zachodzące zmiany jakościowe, lecz nie można podać w sposób precyzyjny dokładnych granic czasowych segmentów, tym bardziej, że często obserwuje się jednoczesność ruchów poszczególnych narządów artykulacyjnych. Powoduje to niemożność wyodrębnienia w sposób obiektywny poszczególnych segmentów. Tak więc wspomniany proces "rozszerzania się" się cech zachodzi na poziomie fonologicznym (czyli sterowania), a nie fonetycznym. Pierwsze próby takiego ujmowania koartykulacji były podejmowane już w latach 60-tych [CZISTOWICZ& KOŻEWNİKOW,1965; ÖHMAN,1966]. Szczególnie Öhman przedstawił przekonujące dowody na to, że nie można traktować strumienia mowy jako ciągu oddzielnych, ściśle zdefiniowanych ruchów, czy położeń układu artykulacyjnego odpowiadających poszczególnym fonemom, czy głoskom. Sterowanie układem artykulacyjnym jest bardziej złożone i zhierarchizowane, i obejmuje nie tylko poszczególne głoski, ale rozciąga się na sylaby, czy wręcz na całe wyrazy. Przed cytowanymi wyżej autorami już pewne wątpliwości co do segmentalnego ujmowania

sygnału mowy wyrażał von Essen [VON ESSEN, 1962], który twierdził, że artykulacja na poziomie wyrazu nie jest ciągiem dyskretnych czynności ruchowych. U podstaw bowiem leży jeden całościowy program artykulacyjny, którego realizacja w wyrazie czy w frazie stanowi wprawdzie nader skomplikowany, niemniej całościowy proces neuromotoryczny.

Tak więc obecnie koartykulację przedstawia się jako przejaw koprodukcji segmentów. Jest to możliwe w momencie przyjęcia założenia, że artykulacja samogłosek i spółgłosek odbywa się przy udziale dwóch różnych, lecz o skoordynowanym działaniu układów neuromuskularnych. Stąd czynnik czasu jest również istotnym elementem i nie można przedstawiać dowolnej wypowiedzi jako konkatenacji odpowiednich segmentów określanych w układzie przestrzennym cech, niezależnym od wymiaru czasu. W istocie wiele przemawia za jednoczesnością ruchów artykulacyjnych samogłoski i sąsiadujących z nią spółgłosek. Głównym efektem koartykulacji jest zacieranie się różnic pomiędzy sąsiadującymi ze sobą dźwiękami i tym samym do upodobnień, to jest do zmian postaci dźwiękowej form wyrazowych.

Głoska powstająca podczas nakładania się lub wzajemnego przenikania różnych ruchów artykulacyjnych musi wypaść inaczej niż wówczas, gdy ruchy te są wykonywane kolejno po sobie i są od siebie oddzielone czasowo. Należy przy tym zwrócić uwagę, że przeciwstawne sobie ruchy artykulacyjne wykazują skłonność do wyrównań, oraz że dominujący pod względem intensywności ruch może służyć inny, mniej istotny, zaś z czasowo różnych ruchów wykonuje się tylko jeden. Zjawisko wszelkiego wyrównywania ruchów przyjęto nazywać upodobnieniem (asymilacją). W wyniku jej może nastąpić zmiana bądź miejsca artykulacji (włącznie ze zmianą elementu narządu artykulacyjnego) lub sposób artykulacji. Zjawiska upodobnienia mogą być dwojakiego rodzaju [WIERZCHOWSKA, 1980]:

a) jeżeli dwie sąsiadujące ze sobą głoski są wymawiane przy użyciu tego samego elementu narządu mowy, wówczas istnieje tendencja do ich wymówienia z podobnym miejscem artykulacji;

b) jeżeli w ich artykulacji użyte są różne elementy, to element swobodny przybiera położenie jakie ma dla głoski sąsiedniej. Na przykład, jest to widoczne przy artykulacji spółgłosek zwarto-wybuchowych dwuwargowych z samogłoskami. Wówczas już przy artykulacji tego typu spółgłoski, masa języka przyjmuje układ specyficzny dla danej samogłoski.

Natomiast grupy spółgłoskowe (złożone ze spółgłosek zwarto-wybuchowych, zwarto-szczelinowych i szczelinowych są w nagłosie i w śródgłosie form wyrazowych

albo całkowicie dźwięczne, albo całkowicie bezdźwięczne (np. /fskotʃitc/ ("wskoczyć"), /krufka/, /pravda/ itp.). Zaś grupy mieszane złożone z tych spółgłosek i z głosek spółotwartych lub otwartych niesylabicznych, są najczęściej bezdźwięczne i wymawiane tak słabo, że często ulegają całkowitej redukcji (np. wyraz "ziarnko"-> /zarko/ ).

Jednakże nie wszystkie przesunięcia cech mogą być wyjaśnione przez upodobnienia. Na przykład nazalizacja samogłosek poprzedzających spółgłoski nosowe, czy zniknięcie opozycji dźwięczność - bezdźwięczność w ciągu spółgłosek nie zawsze służą złagodzeniu przejść od głoski do głoski, choć cecha danego segmentu obejmuje i segmenty sąsiadujące z nim. Tego typu asymilacje są wyjaśniane przez "przesunięcie czasowe". W tym przypadku ruch elementu narządu artykulacyjnego nie ulega poważniejszej modyfikacji. W istocie, bądź element ten zbyt długo jest utrzymywany w danym położeniu, bądź ruch do tego położenia musi się rozpocząć dostatecznie wcześniej, zanim zakończy się artykulacja głoski poprzedzającej. Upodobnienia można również traktować w kategoriach przesunięć czasowych, lecz nie obejmują one zmian w opozycjach dźwięczność-bezdźwięczność oraz nosowa-ustna. W sumie szereg zjawisk przemawia za dynamicznym ujmowaniem koartykulacji. Na przykład Öhman (ÖHMAN,1966), na podstawie badań elektromiograficznych, pokazał że efekt antycypacyjnej artykulacji jest widoczny w przypadku początkowej samogłoski /u/ i jej postaci zależy od tego, czy za nią następuje spółgłoska /d/ czy /g/ (a więc o przednim lub tylnym miejscu artykulacji). W drugim przypadku zaokrąglenie warg jest większe, zaś masa języka jest bardziej przesunięta ku tyłowi. Wskazuje to, że układ artykulacyjny "przygotowuje się" podczas wypowiedzi całej samogłoski do artykulacji następującej po niej spółgłosce.

Zasięg koartykulacji w niektórych przypadkach może być stosunkowo duży. Na przykład [MOLL&DANILOFF, 1971] stwierdzili, że opuszczanie się języczka podniebienia miękkiego może rozpocząć się już na początku pierwszej samogłoski w ciągu głosek, w których jedna lub dwie samogłoski poprzedzają spółgłoskę nosową. Podają oni, że w ciągu "free Ontario" opuszczenie się języczka podniebienia miękkiego niezbędne dla artykulacji spółgłoski /n/ rozpoczyna się już w momencie rozpoczęcia wymawiania samogłoski /i/. Szczególny przypadek koartykulacji był badany przez ([AMERMAN&DANILOFF,1977]). W swej pracy przedstawili, jak położenie i ukształtowanie masy języka odzwierciedlają w danym momencie wypowiedzi sylaby VC lub CV wzajemne oddziaływanie samogłoski i spółgłoski. Zaobserwowali, że w ciągu VC<sub>1</sub>C<sub>2</sub> spółgłoska C<sub>2</sub> ma wpływ na położenie koniuszka języka już podczas artykulacji samogłoski, zaś w ciągu C<sub>1</sub>C<sub>2</sub>V samogłoska ma wpływ na ułożenie tylnej części języka podczas wymawiania spółgłoski C<sub>1</sub>.

Szczegółowe badania elektropalatograficzne [BUTCHER&WEIHER, 1976] wykazały, że samogłoski w sylabie typu VCV również oddziałują na siebie, przy czym przeważa antycypacja artykulacyjna. Interesujące jest, że wzajemne oddziaływanie samogłosek zależy od typu spółgłoski rozdzielającej je. Szczególnie hamująco działają spółgłoski welarne. W tym przypadku, podczas ich artykulacji uczestniczy większa masa języka, co blokuje antycypację ruchu samogłoskowego. Szczególnie ważne wyniki uzyskał Marchal [MARCHAL,1983], który stosując również technikę elektropalatografii badał ciągi dwóch spółgłosek zwartow-ybuchowych, poprzedzonych samogłoską, lub po których ona następowała. Otrzymane wyniki wykazują, że ruchy artykulacyjne spółgłosek pojawiają się podczas wymawiania samogłoski, również w przypadku, gdy spółgłoska poprzedza samogłoskę. Na tej podstawie Marchal wyciąga wniosek, że ruch spółgłoskowy ma charakter ciągły.

MacNeilage i DeClerk [MACNEILADGE&DECLERK,1969] badając sylaby CVC wykazali, że artykulacja samogłoski i końcowej spółgłoski zależy od głoski je poprzedzającej. Podkreślili przy tym, że w szeregu przypadkach, kształt kanału głosowego podczas wymawiania samogłoski zbliża się ku temu, jaki przybiera podczas artykulacji spółgłoski poprzedzającej. Najbardziej widocznym efektem jest fakt, że w przypadku gdy samogłoska jest poprzedzona przez spółgłoskę /g/ objętość jamy ustnej ulega zmniejszeniu. Ogólnie biorąc, dla wszystkich samogłosek następuje wówczas zmniejszenie średniej objętości jamy ustnej. Przyczyną tego jest wypiętrzenia masy języka ku górze. Natomiast antycypacja artykulacyjna jest mniej częsta i na ogół znacznie słabsza. W efekcie spółgłoska następująca (końcowa) oddziałuje w małym stopniu na ukształtowanie toru głosowego podczas wymawiania samogłoski poprzedzającej. Ponadto, wpływ spółgłoski na samogłoskę jest słabszy, niż w odwrotnym przypadku. Cytowani autorzy wyciągają wniosek, że sterowanie podczas wymawiania sylaby CVC nie jest sterowaniem fonemicznym niezmiennym, czyli nie jest wynikiem konkatenacji sterowań narządu głosowego dla poszczególnych fonemów). Pewne z obserwowanych zjawisk koartykulacyjnych mogą być wyjaśnione przez przyjęcie założenia działania szeregu mechanizmów motorycznych wpływających na skurcz mięśni nakładający się na niezmiennie sterowanie fonemiczne.

W podsumowaniu, należy stwierdzić, że tradycyjne ujęcie zjawiska koartykulacji umożliwia interpretację tylko części obserwowanych zjawisk. Uwzględnianie jedynie zmian konfiguracji przestrzennej narządu artykulacyjnego podczas wymawiania poszczególnych głosek w celu opisywania ich za pomocą odpowiednich zbiorów cech, prowadzi do statycznego charakteryzowania dźwięków wypowiedzianych w łańcuchu mowy. Tymczasem ruchy artykulacyjne są programowane nie tylko w sposób przestrzenny, lecz i dynamiczny, to znaczy są

koordynowane w płaszczyźnie czasowej. Tak więc choć w wielu przypadkach posługujemy się pojęciem segmentów mowy, to jednak należy pamiętać, że mają one charakter dynamiczny. I pomimo że są różne jakościowo, tym nie mniej nie można między nimi postawić precyzyjnych granic czasowych. Zależność cech poszczególnych segmentów od kontekstu nie jest wynikiem "dopasowywania się" segmentów, lecz następuje wskutek realizacji pewnego planu wypowiedzi przedstawionego na poziomie fonologicznym w sposób abstrakcyjny. Plan ten stanowi podstawę do zorganizowania odpowiedniego przebiegu ciągu zdarzeń w płaszczyźnie fonetyczno-akustycznej, obejmującego również określone schematy koartikulacyjne oraz odpowiednie wymiary czasowe. Ponadto, w płaszczyźnie artykulacyjno-percepcyjnej dokonywana jest selekcja cech dystynktywnych, szczególnie z punktu widzenia percepcyjnego, umożliwiające dokonywanie opozycji między poszczególnymi klasami głosek w płaszczyźnie artykulacyjnej, akustycznej i percepcyjnej, z których dwie pierwsze mają charakter ściśle fizyczny, zaś trzecia - abstrakcyjny (bowiem opiera się już na analizie fonologicznej). Tak więc koartikulacja odwołuje się intuicyjnie do pojęcia segmentu, który z drugiej strony staje się pojęciem bardzo płynnym w momencie, gdy usiłuje się ustalić relacje istniejące między jednostkami lingwistycznymi i ich odpowiednikami w sferze artykulacyjnej i akustycznej.

Przedstawiony powyżej opis zjawiska koartikulacji oraz schematu organizacji wypowiedzi pokazuje jak bardzo trudno jest zaprogramować modelowanie generacji wypowiedzi w płaszczyźnie artykulacyjnej. Z tego powodu naturalnym wydaje się dążenie do uproszczenia organizacji procesu generacji ciągu segmentów przypisując im określone znaczenia niezależne od otoczenia, a zjawisko koartikulacji odtwarzać tylko w odcinkach czasowych odpowiadających przejściom (tranzjentom) od segmentu poprzedzającego do następującego. Tego typu podejście jest uzasadnione tylko w określonych przypadkach. Po pierwsze, gdy wzajemne oddziaływanie na siebie poszczególnych ruchów artykulacyjnych jest stosunkowo słabe lub gdy segmenty sąsiadujące ze sobą są podobnej natury (różnią się stosunkowo niewiele tylko jedną cechą, np. miejscem artykulacji, tak jak to ma miejsce w przypadku spółgłosek /m/ i /n/). W takim przypadku można zasób fonemów danego języka podzielić na pewne grupy, dla których można przyjąć, że koartikulacja będzie miała charakter liniowy, to jest nakładanie się ruchów artykulacyjnych następuje tylko w odcinkach czasowych, w których następuje przejście od jednej do kolejnej konfiguracji artykulacyjnej toru głosowego. W innych przypadkach należy uwzględnić koartikulację na całym odcinku (co najmniej o długości sylaby, ale często i dłuższym) i odpowiednio modyfikować również i docelową konfigurację toru w zależności od jej otoczenia. Jest to proces bardzo złożony i właściwie do tej pory nie rozwiązany.

Ponieważ w niniejszej pracy skupiono się na modelowaniu procesu artykulacji dla ciągów złożonych z rezonantów, zjawisko koartykulacji jest w ich przypadku znacznie mniej złożone i można było przyjąć z pewnym przybliżeniem, że jest wynikiem własności układu artykulacyjnego będącego swoistego rodzaju układem mechanicznym, którego ruchy są w mniejszym stopniu wynikiem sterowania centralnego określającego z góry ruchy artykulacyjne i konfiguracje toru głosowego odpowiednio do zadanej całej sekwencji fonemów. Przyjęto więc, że koartykulacja jest ograniczona przede wszystkim do odcinka przejściowego między sąsiadującymi ze sobą głoskami. Umożliwiło to łatwiejszą obserwację wzajemnego oddziaływania wybranych parametrów artykulacyjnych sąsiadujących ze sobą głosek.



## 4. Dźwięki mowy języka polskiego i ich cechy fonetyczno-akustyczne

### 4.1 Opis artykulacyjny dźwięków mowy polskiej

W mowie polskiej występuje stosunkowo wiele dźwięków o charakterze spółgłoskowym, a zwłaszcza takich, w których istotnym składnikiem są przebiegi nieregularne, szumowe bądź przebiegi impulsowe generowane przez szczeliny i zwarcia powstające w obrębie toru ustnego. Szczególnie liczne są te dźwięki, przy których wymawianiu zwarcia i szczeliny tworzą się w przedniej części jamy ustnej. Możliwości artykulacyjne narządów mowy są w przypadku tych dźwięków wyzyskiwane w bardzo dużym stopniu. Natomiast w porównaniu z innymi językami w mowie polskiej jest stosunkowo mniej dźwięków o charakterze samogłoskowym (otwartych), w których strukturze akustycznej dominują przebiegi w zasadzie regularne.

Jednym z naczelných problemów w fonetyce każdego języka jest stworzenie dla niego zwartego i jednoznacznego opisu artykulacyjnego systemu głosek. Tradycyjnie wyodrębnia się samogłoski i spółgłoski jako dwie oddzielne grupy opisując je w sposób niezależny od siebie. W przypadku języka polskiego, przyjmuje się, że jego system fonologiczny opiera się na 37 fonemach [JASSEM&ŁOBACZ,1971], to jest składa się z 6 samogłosek ustnych, 2 samogłosek nosowych oraz z 29 spółgłosek. W niniejszym rozdziale podjęto próbę nieco odmienną od tradycyjnej klasyfikacji głosek, głosek uwzględniając w nim nie tylko ich artykulację, lecz również ich strukturę akustyczną. Klasyfikacja ta była stworzona pod kątem jej wykorzystania w systemach automatycznego rozpoznawania mowy [GUBRYNOWICZ& WRZOSKOWICZ, 1993]

Ogólnie biorąc klasyfikację głosek polskich można dokonać w oparciu o omówione wcześniej, trzy zasadnicze elementy opisu artykulacyjnego, to jest sposób, miejsce i wysokość artykulacji. Klasyfikacja dźwięków mowy oparta na sposobie artykulacji może być przeprowadzona w bardzo różny sposób i z reguły jest dostosowana do własności fonetycznych i fonologicznych danego języka. Oznacza to, że przyjęty dla jednego języka schemat klasyfikacyjny umożliwiający jednoznaczny opis poszczególnych dźwięków mowy, może okazać się zawodny w przypadku próby opisu dźwięków mowy innego języka i na przykład, uniemożliwić jednoznaczne wyodrębnienie cech dystyngtywnych niektórych jego głosek. Istotny wpływ na określenie schematu klasyfikacyjnego głosek ma też aspekt praktyczny, a zwłaszcza cel do jakiego został on stworzony. Przykładem takiego systemu klasyfikacji głosek dostosowanego do struktury dźwiękowej języka polskiego jest system klasyfikacji oparty na sposobie i miejscu artykulacji opracowany pod kątem zastosowania go w systemie automatycznego rozpoznawania ograniczonego zbioru izolowanych wyrazów (10 klas głosek plus segment ciszy) [GUBRYNOWICZ i in.,1990]. W oparciu o podobny, bardziej rozszerzony schemat klasyfikacji głosek języka polskiego

opracowano system rozpoznawania mowy ciągłej, z zastosowaniem niejawnych modelach Markowa (tzw. HMM [GUBRYNOWICZ&WRZOSKOWICZ,1993]). W obu przypadkach interpretacja sposobu artykulacji głoski, czy jej miejsca artykulacji była interpretowana w oparciu o analizę widmowo-czasową sygnału mowy. W systemie rozpoznawania posługiwano się więc klasami głosek, a nie ich fonetycznymi odpowiednikami. Tego typu opis można więc stosować, na jednym z początkowych etapów rozpoznawania mowy, zawiązując w ten sposób, na przykład, zbiór rozpoznawanych na dalszym etapie jednostek (sylab, difonów, czy wyrazów)

Na Rys. 4.1 przedstawiono stosowany w cytowanych pracach schemat klasyfikacji głosek. Klasyfikacja ta oparta jest na dwóch zasadniczych elementach, to jest na sposobie i miejscu artykulacji. W sposobie artykulacji dodatkowo zróżnicowano głoski według ich czasowej struktury segmentalnej. Większość głosek języka polskiego ma strukturę monosegmentalną, to znaczy ich struktura czasowa jest jednolita, a podczas ich wymawiania nie ma takich zmian w artykulacji, by można było wyodrębnić sekwencje odmiennych faz artykulacyjnych, tak jak to ma miejsce, na przykład, w przypadku, wszystkich spółgłosek zwartych bezdźwięcznych (niezależnie, czy zwarto-trących, czy zwarto-wybuchowych). Podczas ich wymawiania można bowiem wyróżnić dwie odrębne fazy artykulacyjne. Pierwsza, będąca zwarciem bezdźwięcznym, po którym następuje druga, podczas której realizowany jest segment trący (lub wybuchowy), również bezdźwięczny. Podobnie jest w przypadku głosek zwartych dźwięcznych, w których obie fazy są dźwięczne. Wyraźne zróżnicowanie tych faz w przypadku głosek zwarto-trących w płaszczyźnie akustycznej (jak i artykulacyjnej) skłania niektórych autorów traktować je nie jako realizacje odrębnych fonemów, lecz jako sekwencji dwóch odpowiednich fonemów, to jest głoski zwarto-wybuchowej (/t/ lub /d/) i głoski trącej (np. [BIEDRZYCKI,1975]). Pogląd ten jednak kwestionuje Wierzchowska, wykazując, że sekwencja tych fonemów jest realizowana w mowie naturalnej inaczej, niż rzeczywista głoska zwarto-trąca [WIERZCHOWSKA, 1980].

Zasadniczą i najliczniejszą grupą wyróżnioną w powyższym schemacie klasyfikacyjnym stanowią tzw. rezonanty grupujące głoski dźwięczne, wyróżniające się od pozostałych głosek wyraźną formantową strukturą widma. Wyodrębniono w tej klasie 3 odrębne grupy: rezonanty samogłoskowe (o stacjonarnym charakterze konfiguracji toru głosowego podczas ich artykulacji), rezonanty nosowe (o stałej konfiguracji toru głosowego oraz z jednoczesnym udziałem toru nosowego) i rezonanty sylabiczne (odznaczają się płynną zmianą konfiguracji toru ustnego podczas ich artykulacji). Rezonanty z ostatniej grupy nie można wymówić bez udziału samogłoski (poprzedzającej lub następującej).

Nazwa „rezonanty” nie jest zbyt często stosowana w literaturze fonetycznej [BIEDRZYCKI,1975; SAWICKA,1995], choć z drugiej strony, tym terminem można objąć w sposób jednoznaczny grupę głosek, przy których artykulacji, z akustycznego

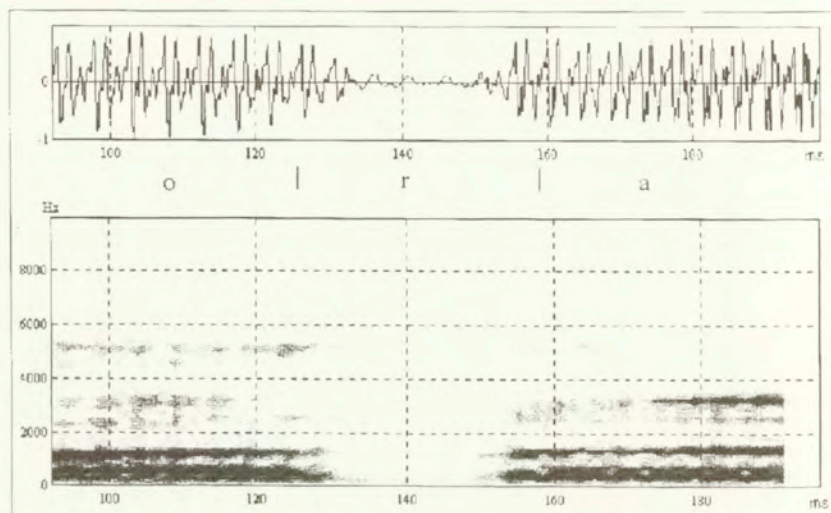
punktu widzenia, tor głosowy jest pobudzany na wysokości krtani, zaś propagowana w nim fala akustyczna rozchodzi się bez powstawania turbulencji i jest swobodnie promieniowana przez usta i/lub nos. Przy takiej definicji rezonantów, jak słusznie zauważa Sawicka [SAWICKA,1995] do tej klasy należy również zaliczyć samogłoski.

		MIEJSCE ARTYKULACJI					
		→					
SPOSÓB ARTYKULACJI	Polisegmentalne	Przednio-językowe /r/ - <b>AR</b>	r				
		Zwarty-trące dźwięczne <b>VA</b>	ɖʒ	ɖʒ̥	ɖʒ̥		
		Zwarty-trące bezdźwięczne <b>UA</b>	tʂ	tʂ̥	tʃ̥		
		Wybuchowe bezdźwięczne <b>UP</b>	p	t			k
		Wybuchowe dźwięczne <b>VP</b>	b	d			g
	Monosegmentalne	Szczelinowe bezdźwięczne <b>SS</b>	f	s	ʃ̥	ʃ̥	x
		Szczelinowe dźwięczne <b>VF</b>	v	z	ʒ̥	ʒ̥	
		REZONANTY nosowe <b>RN</b>	m	n		ɲ	ŋ
		sylabiczne <b>RS</b>		l		j	w
		samogłoski <b>RV</b>				i ɪ e	u o a

↑  
Wysokość  
artykulacji

Rys. 4.1 Schemat klasyfikacji głosek języka polskiego wg [GUBRYNOWICZ& WRZOSKOWICZ,1993]. Klasy głosek oznaczono dwuliterowymi kodami stosowanymi w systemie rozpoznawania izolowanych wyrazów.

Do klasy rezonantów należałoby również zaliczyć głoskę /r/ (tak jak to czyni Biedrzycki [BIEDRZYCKI,1978]), jednakże ze względu na jej polisegmentalną strukturę została ona wyodrębniona z tej grupy. Artykulacja tej głoski jest bardzo specyficzna i silnie zależy od jej miejsca w sekwencji głosek, a także od kontekstu. W zasadzie jest to głoska przednio-językowa dźwiękowa, drżąca. Jest ona artykułowana z płaskim układem masy języka poprzez co najmniej jednokrotne (lub niekiedy więcej) uderzenie koniuszka (apex) języka o górne dźiąsto, lub tylko zbliżenie się do niego na niewielką odległość. Jak silnie zależy realizacja tego fonemu od pozycji, pokazano to na Rys. 4.2, 4.3 i 4.4, przedstawiające oscylogramy i sonagramy tej głoski wymówionej kolejno w śródgłosie, w nagłosie oraz w wygłosie, odpowiednio w wyrazach "wczoraj", "rury" i "telewizor", w frazach wypowiedzianych przez jedną i tę samą osobę. Sonagramy są to trójwymiarowe spektrogramy, w których oś pozioma jest osią czasu, pionowa - częstotliwości, zaś stopień zaczernienia w określonym paśmie częstotliwości jest proporcjonalny do zwartej w nim energii. Wszystkie wyniki analiz widmowych przedstawione w tym rozdziale zostały wykonane dla fragmentów mowy ciągłej zarejestrowanej w językowej bazie danych BABEL [GUBRYNOWICZ,1998; 1999]. Analizy były wykonane za pomocą programu *Phonedit* opracowanego w Instytucie Fonetyki Uniwersytetu w Aix-en-Provence (Francja).



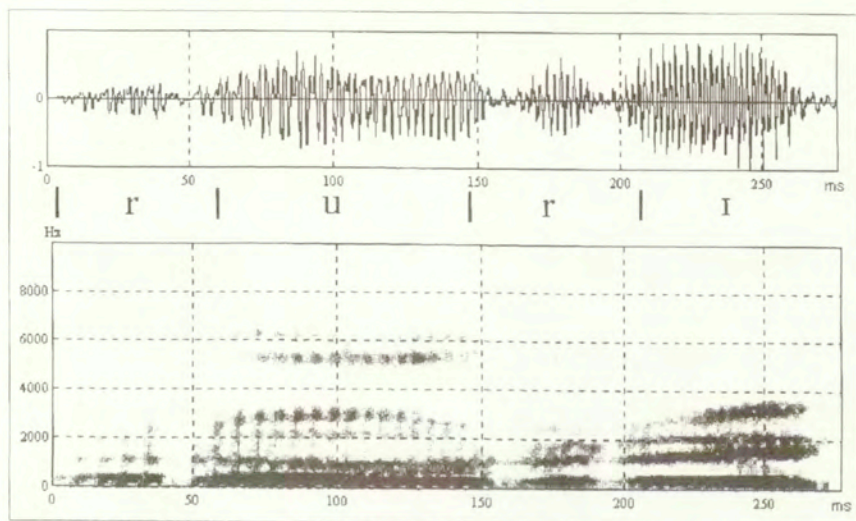
Rys. 4.2 Oscylogram i sonagram sekwencji /ora/ wypowiedzianej w wyrazie „wczoraj” /ftʃoraj/.

Warto porównać ze sobą przebiegi oscylograficzne i widma głoski /r/ wymówionej w śródgłosie w wyrazie "wczoraj" (Rys. 4.2) oraz w wyrazie "rury" (Rys. 4.3). W pierwszym przypadku widać, że nastąpiło tylko jedno uderzenie koniuszka języka o górne dźiąsto, zaś w drugim dwa, przedzielone między sobą krótkim

segmentem wokalicznym (o strukturze samogłoskopodobnej) zapewniającym pewną ciągłość w przejściu formantów samogłoskowych od /u/ do /i/.

Segment ten ma brzmienie zbliżone do tzw. „neutralnego e” (w transkrypcji fonetycznej IPA - /ə/), zaś jego częstotliwości formantowe są bliskie częstotliwościom rury rezonansowej o długości bliskiej długości toru głosowego ( $l = 17$  cm), to jest 500, 1500, 2500 Hz,... (częstotliwości te odpowiadają kolejnym falom stojącym o długości odpowiednio  $\frac{1}{4}l$ ,  $\frac{3}{4}l$ ,  $\frac{5}{4}l$ ,...). Warto podkreślić, że w percepcji głoski /r/ znaczenie ma nie tylko segment odpowiadający uderzeniu języczka, lecz również, stosunkowo krótkie, znajdujące się po obu stronach fragmenty samogłosek - poprzedzającej jak i następującej po nim. Granice tych części samogłosek wyznaczają tzw. ugięcia trzeciego i/lub czwartego formantu. Warto tu zaznaczyć, że w przypadku artykulacji /r/ w otoczeniu samogłoskowym, wspomniany element wokaliczny w mowie ciągłej występuje stosunkowo rzadko, raczej w przypadku, gdy akcentowana jest druga samogłoska.

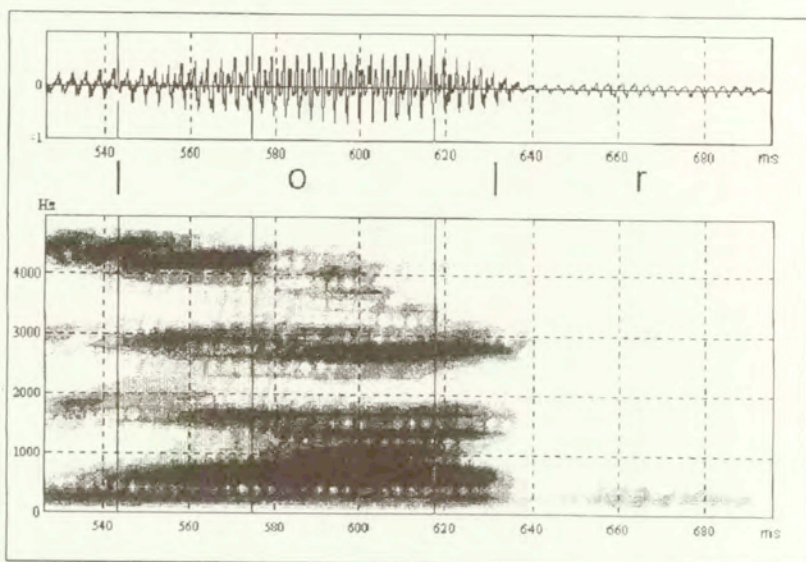
Podobny segment samogłoskopodobny („neutralne e”) występuje przy artykulacji głoski /r/ w nagłosie (por. wyraz "rury"), od którego zaczyna się jej artykulacja. Po nim następuje uderzenie języczka o znacznie krótszym czasie trwania, niż w przypadku artykulacji śródgłosowej i przejście do samogłoski /u/.



Rys.4.3 Oscylogram i spektrogram wyrazu „rury”.

Przypadek artykulacji /r/ w wygłosie został przedstawiony na Rys. 4.4 (wyraz „televizor”). Również i w tym przypadku głoska /r/ ma wyraźną polisegmentalną strukturę. Po samogłosce /o/ następuje krótki segment odpowiadający chwilowemu zbliżeniu się koniuszka języka do podniebienia twardego, po którym artykulację wyrazu „televizor” kończy segment samogłoskopodobny o stosunkowo małej amplitudzie.

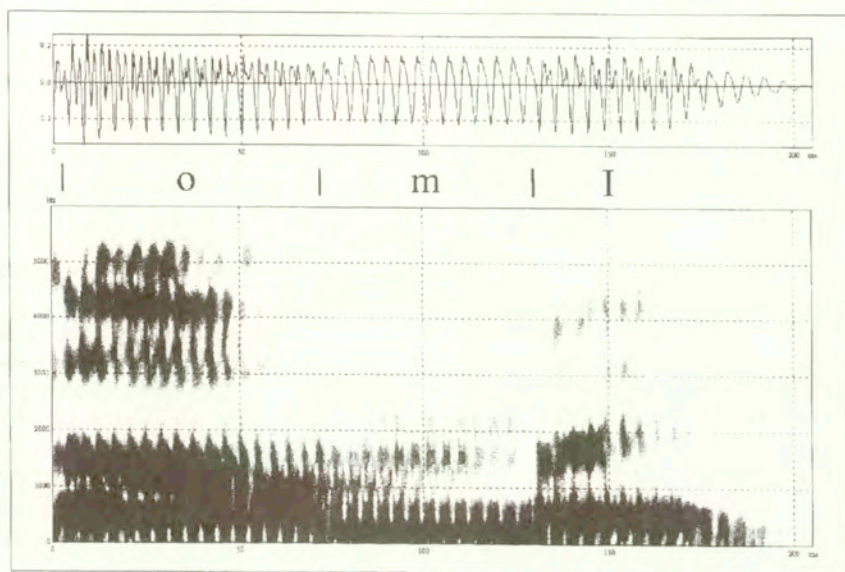
Na podstawie analizy widmowej jest trudno określić częstotliwości charakterystyczne dla segmentu uderzeniowego, natomiast z oscylogramu widać wyraźnie, że uderzenie języczka nie powoduje całkowitego zamknięcia toru głosowego. Utrudnia to wyznaczenie w sposób jednoznaczny granic segmentu odpowiadającego głosce /r/. Dopiero przy nieco zwolnionym wymawianiu głoski /r/ w sylabie typu  $V_1CV_2$ , gdzie  $V_1 = V_2$  można zaobserwować charakterystyczne ugięcia częstotliwości formantowych na skrajach segmentów samogłoskowych i spółgłoskowego. Kierunek ugięć formantów silnie zależy wówczas od miejsca artykulacji samogłosek. W przypadku, gdy samogłoski znacznie różnią się miejscem artykulacji, jak na przykład w wyrazie „rury” samogłoski /u/ - /i/, wówczas nad ugięciami formantów dominują ich przejścia od samogłoski poprzedzającej do następującej. Podobnie, efekt ten jest szczególnie widoczny przy artykulacji spółgłosek nosowych, zwłaszcza dwuwargowej /m/, dla której ułożenie języka podczas jej wymawiania może zmieniać się w znacznym stopniu.



Rys. 4.4 Oscylogram i spektrogram końcowego fragmentu sylaby /zor/ wymówionej w wyrazie „televizor”.

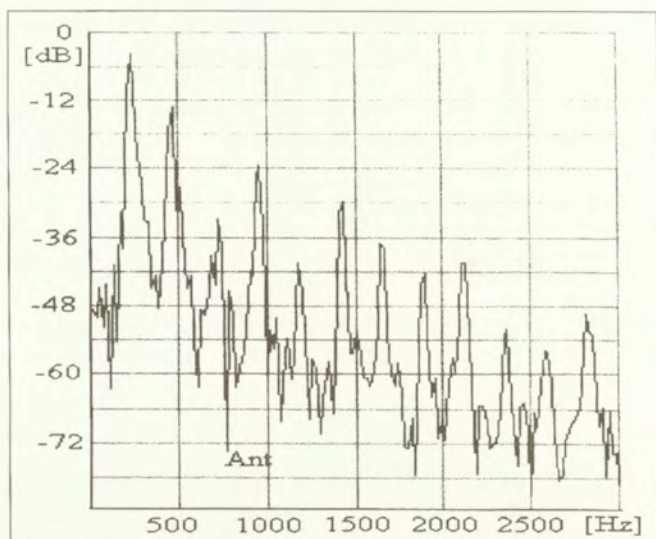
Specyficzną grupę w obrębie rezonantów stanowią dźwięki nosowe wytwarzane ze zwarciami w obrębie jamy ustnej, to jest spółgłoski nosowe /m, n, ŋ, ŋ/. Są one monofoniczne i mają dość stabilną strukturę czasową. Jest to jedyna klasa spółgłosek w języku polskim artykułowana przy udziale dwóch części: jamy ustnej i nosowej, stanowiących razem złożony układ rezonansowy, w którym struktura formantowa dźwięku mowy jest formowana przez część gardłowo-nosową, przy jednoczesnym bocznikującym oddziaływaniu jamy ustnej. Wynikiem jej oddziaływania jest pojawienie się w widmie spółgłosek nosowych charakterystycznego minimum ([FUJIMURA,1962; DUKIEWICZ,1967]).

Na Rys. 4.5 przedstawiono przykładowo widmo spółgłoski /m/ wypowiedzianej (przez głos kobiecy) w otoczeniu samogłoskowym, w wyrazie „domy”, wyciętym z fragmentu mowy ciągłej.



**Rys. 4.5** Oscylogram i spektrogram sekwencji /om/ wymówionej w wyrazie „domy”.

Na następnym rysunku (4.6) przedstawiony jest przekrój widmowy wyznaczony dla tej spółgłoski (w okolicy 100 ms). Widoczne na nim jest charakterystyczne, wąskie lokalne, minimum w pobliżu częstotliwości 760 Hz, które jest antyformantem (oznaczony na wykresie symbolem „Ant”).



Rys. 4.6 Przekrój widmowy spółgłoski /m/ wypowiedzianej w wyrazie domy”, wyznaczony w chwili 100 ms od początku przebiegu przedstawionego na Rys. 4.7.

Fujimura [FUJIMURA,1962] stwierdził, że częstotliwość antyformantu jest związana przede wszystkim z miejscem artykulacji spółgłoski nosowej, a więc z długością bocznikującej jamy ustnej. Przyjmując, że o częstotliwości antyformantu decyduje przede wszystkim rezonans półfalowy powstający w części gardłowo-ustnej (jak w przypadku rury obustronnie zamkniętej). Stąd można przyjąć za Fujimurą, że częstotliwość antyformantu głoski /m/ znajduje się w przedziale 750-1250 Hz, dla /n/ - w paśmie 1450-2200 Hz, zaś dla /ŋ/ - powyżej 3000 Hz. Należy zaznaczyć, że w mowie naturalnej jest bardzo często trudno określić dokładnie położenie antyformantu, który często powoduje deformację widma w zakresie niskich częstotliwości, ponieważ znajduje się stosunkowo blisko formantu drugiego (dla /m/), lub trzeciego (dla /n/), utrudniając jednocześnie dokładne określenie częstotliwości formantowych.

Analiza artykulacyjna spółgłosek nosowych sprowadza się do określenia miejsca artykulacji, czyli miejsca zwarcia, które uniemożliwia strumieniowi powietrza ujście na zewnątrz toru ustnego. Zwarcie takie dla /m/ powstaje między górną i dolną wargą, dla /n/ między końcem języka (apex) i tylną ścianą górnych zębów i dziąseł, dla /ŋ/ między częścią grzbietu języka i zadziąsłową częścią podniebienia twardego. Miejsce zwarcia ustnego wyznacza zatem, dla każdej z głosek, długość toru ustnego bocznikującego toru nosowego.



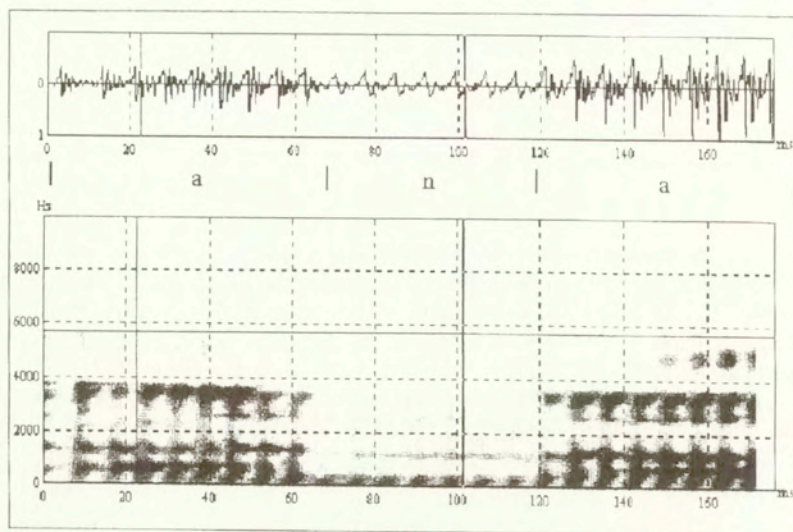
Dźwięk /m/ jest często wymawiany z płaskim układem powierzchni języka. W czasie jego artykulacji tor ustno-gardłowy jest prawie przez cały czas zamknięty na skutek zwarcia obu warg. Zwarcie to jest słabsze niż dla głosek zwarto-wybuchowych typu /p/ i /b/. Języczek podniebienia miękkiego otwiera dla powietrza przejście do jamy nosowej. Przekrój poprzeczny toru nosowego w czasie artykulacji nie ulega zmianom, jedynie wielkość wlotu do jamy nosowej (4 mm) decyduje o stopniu nosowości spółgłoski.

Dźwięk /n/ również wymawia się zwykle z płaskim układem masy języka, którego brzegi przylegają do sklepienia jamy ustnej na całej długości łuku zębowego. Pas kontaktu języka ze sklepieniem jamy ustnej ma kształt podkowy zwróconej ramionami ku tyłowi i jest nieco większy niż pas kontaktu właściwy dla spółgłosek przedniojęzykowych /t/ i /d/ (bardzo podobny układ artykulacyjny). Przy artykulacji głoski /n/, podniebienie miękkie otwiera przejście do toru nosowego (którego wysokość jest rzędu 7 mm), przez które przepływa strumień powietrza nieco silniejszy niż dla /m/. Podniebienie miękkie przybiera kształt zagięty, przy czym przednia jego część przedłuża podniebienie twarde. Wargi są lekko rozchylone, w odróżnieniu do /m/.

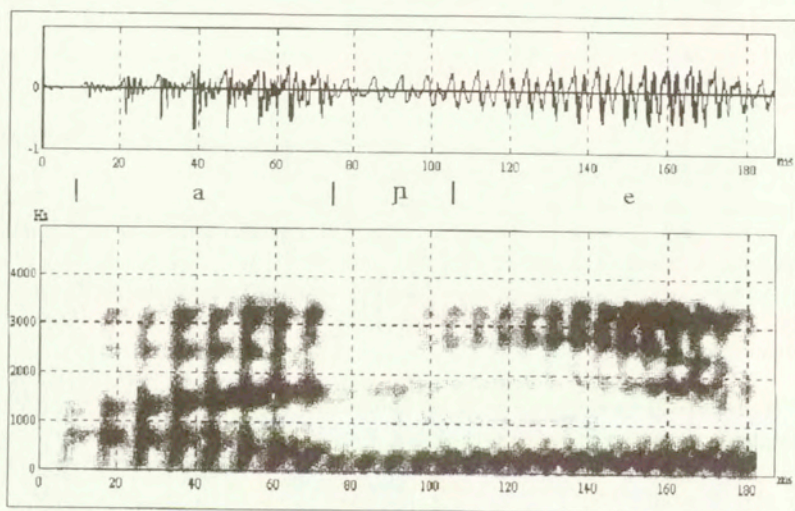
Obecność spółgłoski nosowej w ciągu głosek ma natomiast dość wyraźny wpływ koartylacyjny na sąsiadujące z nią głoskami, a zwłaszcza na bezpośrednio następujące po niej. Wpływ ten objawia się częściowo nazalizacją przylegającego do spółgłoski nosowej fragmentu następującej po niej głoski ustnej (najczęściej samogłoski). Na Rys. 4.7 i 4.8 przedstawiono odpowiednie przebiegi wyznaczone dla sekwencji /ana/ wypowiedzianej w ciągu wyrazów "a nadmiar" oraz sekwencji /aŋe/ w ciągu "a niemożliwe".

Spółgłoski nosowe wykazują bardzo silną tendencję do zachowania ustalonego przebiegu w połączeniu z innymi głoskami. Jest to spowodowane tym, że sygnał mowy podczas ich artykulacji jest formowany przede wszystkim przez gardłowo-nosową część toru głosowego. Ma ona wówczas stosunkowo stałą konfigurację geometryczną, a znaczna część energii sygnału mowy jest promieniowana przez otwory nosowe. Tak więc udział części ustnej ogranicza się w zasadzie do powstania w widmie antyformantów, zaś cała energia jest promieniowana niemal w całości przez otwory nosowe.

Częstotliwości formantowe polskich spółgłosek nosowych są na ogół zbliżone do następujących wartości: /m/ -  $F_1=300$  Hz,  $F_2=750$  Hz,  $F_3=1300$  Hz, /n/ -  $F_1=320$  Hz,  $F_2=1100$  Hz,  $F_3=1700$  Hz, /ɲ/ -  $F_1=350$  Hz,  $F_2=1250$  Hz,  $F_3=2100$  Hz, /ŋ/ -  $F_1=350$  Hz,  $F_2=1100$  Hz,  $F_3=1900$  Hz [JASSEM,1973]. Ta ostatnia spółgłoska występuje stosunkowo rzadko w języku polskim (częstotliwość występowania poniżej 0.1% - [JASSEM,1973]).



Rys. 4.7 Oscylogram i spektrogram sekwencji /ana/ wyciętej z sygnału mowy ciągłej, na granicy zbitki dwóch wyrazów „a nadmiar”.



Rys. 4.8 Oscylogram i spektrogram sekwencji /ape/ wymówionej w ciągu wyrazów „a niemożliwe”.

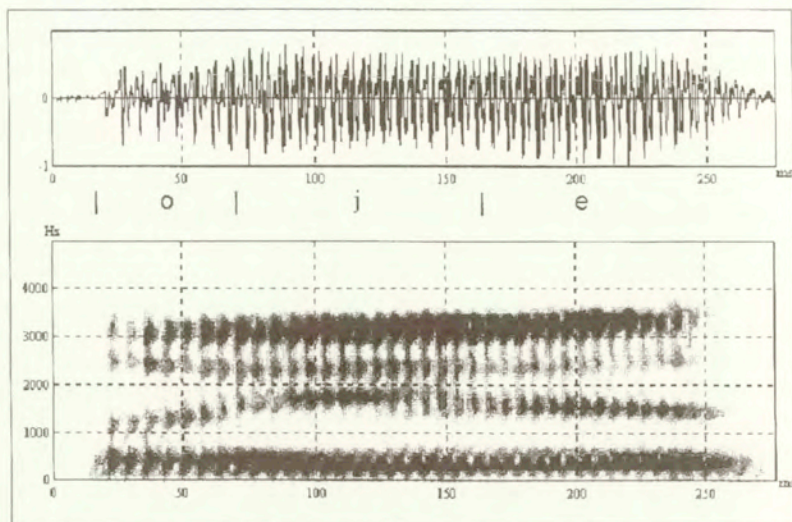
Warto w tym miejscu zaznaczyć, że podawane wyżej wartości liczbowe mają raczej orientacyjny charakter, gdyż budowa anatomiczna jamy nosowej nie jest jednakowa dla poszczególnych osób i jej zróżnicowanie powoduje znaczne odchylenia częstotliwości formantowych. Jak dalece zróżnicowania osobnicze w budowie jamy nosowej wykazują pomiary wykonane techniką MRI [DANG i in., 1994; 1996], z których wynika, że może ona znacznie różnić w sposób znaczny nie tylko objętością obu przejść (a także sinusów), ale także konfiguracją geometryczną. Powoduje to znaczne zróżnicowanie osobnicze funkcji transmitancji części gardłowo-nosowej toru artykulacyjnego.

Przy omawianiu spółgłosek nosowych należy jeszcze podkreślić pewien istotny fakt. Przez wiele lat w literaturze dominował pogląd, że przy percepcji spółgłosek nosowych dominującą rolę odgrywa część sygnału mowy generowana przy zamkniętej jamie ustnej (nasal murmur). Tymczasem badania percepcyjne wykazały [KUROWSKI i in., 1984; REPP, 1986], że sam segment nosowy ma tak bardzo mało zróżnicowaną charakterystykę widmową, która uniemożliwia w pewny sposób rozpoznawanie spółgłosek nosowych, natomiast najbardziej istotne informacje o ich miejscu artykulacji niosą segmenty tranzjentowe, zarówno początkowy, jak i końcowy. Segmenty te są widoczne na przytoczonych przykładach przebiegów widmowych spółgłosek nosowych (por. Rys. 4.5, 4.7, 4.8).

Następną grupą rezonantów tworzą spółgłoski sylabiczne, takie jak /l/, /j/ i /w/ o silnie tranzjentowym charakterze, i które nie sposób wymówić bez udziału samogłoski poprzedzającej lub następującej. Na Rys. 4.9, 4.10 i 4.11 przedstawiono przebiegi wyznaczone odpowiednio dla sekwencji /oje/, /ole/ i /owe/ "wycięte" z wyrazów "to jest", "koleżanki" i "wesołego". Porównując przedstawione dla tych sylab przebiegi widm widać, że najbardziej zmiennym jest widmo odpowiadające sylabie /oje/ i podane na rysunku granice dla spółgłoski /j/ należy traktować jako bardzo umowne. Warto zaznaczyć, że ruchy formantów odbywają się na znacznej części sygnału odpowiadającej samogłoskom, to znaczy, że zmiana konfiguracji toru głosowego rozpoczyna się jeszcze podczas artykulacji samogłoski /o/ i zmierza do konfiguracji odpowiadającej spółgłosce /j/, by następnie osiągnąć konfigurację typową dla samogłoski /e/.

Jednakże w podanym przykładzie wpływ następującej po tej spółgłosce samogłoski /e/ powoduje, że docelowa konfiguracja toru głosowego odpowiadająca głosce /j/ nie zostaje osiągnięta (świadczy o tym stosunkowo zbyt niska częstotliwość drugiego formantu - 1750 Hz zamiast 2200-2400 Hz, a także trzeciego), bowiem już w segmencie odpowiadającym tej spółgłosce rozpoczyna się ruch ku konfiguracji odpowiadającej samogłosce /e/ trwający w jej znacznej części. I choć gdy odtwarza się tylko segment odpowiadający głosce /j/, to słyszy się dźwięk zbliżony raczej do głoski "y", to odtwarzając go wraz z samogłoską poprzedzającą lub następującą, słyszy się głoskę /j/. Świadczy to o decydującej roli samego tranzjentu w percepcji

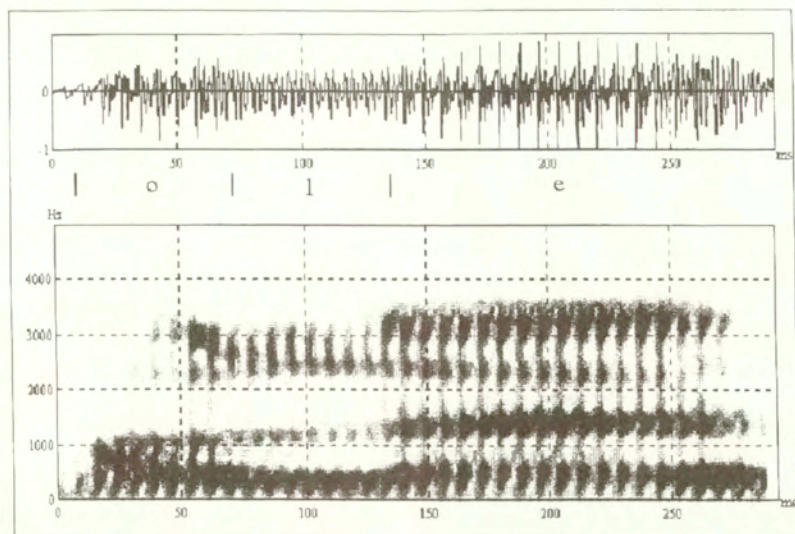
tego dźwięku, a nie fakt osiągnięcia konfiguracji docelowej toru głosowego, który może nigdy nie nastąpić.



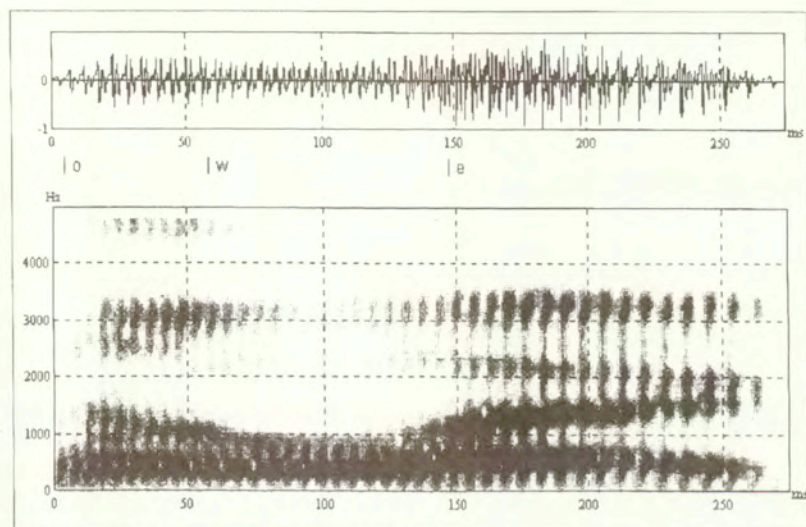
Rys. 4.9 Oscylogram i spektrogram przebiegu zarejestrowanego dla sekwencji głosek /oje/ wypowiedzianej w ciągu wyrazów „to jest”.

O ile dla głoski /j/ wypowiedzianej w kontekście samogłoskowym segment ustalony na ogół nie występuje (niekiedy jest bardzo krótki 10-30 ms), to w przypadku głosek /l/ i /w/ jest on zazwyczaj stosunkowo długi i tranzjenty formantów kończą się (lub zaczynają się) na wartościach typowych dla stanu ustalonego tych głosek.

Głoski typu /l/ noszą nazwę bocznych (niekiedy są zaliczane do płynnych). Jest to nazwa pochodząca od układu narządów artykulacyjnych podczas ich wymawiania. Masa języka wypełnia dno jamy ustnej, grzbiet języka bardzo często nawet zagłębia się łyżkowato. Z jednego boku jamy ustnej, między brzegiem jamy ustnej a pograniczem bocznych zębów i dziąseł, powstaje otwór. W przodzie jamy ustnej pas kontaktu języka zlokalizowany jest na wysokości wałka dziąsłowego. Utworzenie się szczelin bocznych różni w sposób zasadniczy artykulację /l/ od głosek szczelinowych. Powietrze bowiem, wydostające się przez kanały boczne, a dalej przez boczny rozstęp między zębami i boczne rozchylenie warg, prawie nie powoduje szmeru, a przednie zwarcie, osłabione bocznym odpływem powietrza też nie daje energicznego wybuchu, zatem głoska /l/ wymawiana jest z niewielkim zwarciem, kończącym się niekiedy bardzo słabą plozją.



Rys. 4.10 Oscylogram i spektrogram przebiegu zarejestrowanego dla sekwencji /ole/ wypowiedzianej w wyrazie „kolezanka”.



Rys. 4.11 Oscylogram i spektrogram przebiegu zarejestrowanego dla sekwencji /owe/ wypowiedzianej w wyrazie „wesolego”.

Formanty głoski /l/ są  $F_1 = 350-400$  Hz,  $F_2 = 1200 -1400$  Hz,  $F_3 = 2500-2700$  Hz [JASSEM,1973].

W przypadku głoski /w/ ruchy formantów są również znaczne, jednakże dla głoski /l/ są dużo szybsze (tzn. czas przejścia tranzjentowego jest znacznie krótszy). Można to wyjaśnić tym, że w przypadku artykulacji tej głoski zasadniczy ruch wykonuje koniusek języka, który ze względu na swoją małą masę, może odbywać się ze stosunkowo dużą prędkością. Tym nie mniej, mimo wolniejszego ruchu masy języka, ugięcia formantów samogłoskowych osiągają wartości docelowe głoski /w/, której częstotliwości formantowe wynoszą:  $F_1 = 300-500$  Hz,  $F_2 = 700-900$  Hz,  $F_3 = 2400-2700$  Hz. Konfiguracja toru artykulacyjnego podczas wymawiania tej głoski jest bardzo zbliżona do konfiguracji charakterystycznej dla samogłoski /u/. Powielając wycięty z sygnału segment ustalony głoski /w/ słyszy się wyraźnie tę samogłoskę.

#### 4.2 Samogłoski

Dla klasyfikacji artykulacyjnej samogłosek polskich, których zestaw jest ograniczony w porównaniu z innymi językami (np. 11 ustnych i 4 nosowe samogłoski w języku francuskim, 11 długich i krótkich samogłosek oraz 8 dyftongów w języku angielskim) wystarczy przyjąć zróżnicowanie konfiguracji masy języka wyznaczanej w dwóch prostopadłych względem siebie płaszczyznach, przebiegających przez linię środkową toru gardłowo-ustnego. Schemat klasyfikacyjny tego typu samogłosek ustnych podano niżej (wg [DŁUSKA, 1981]).

	przednie	centralne	tylne
wysokie	i	ɪ	
średnie	e		u
niskie		a	o

Rys. 4.12 Schemat klasyfikacji samogłosek polskich wg M. Dłuskiej [DŁUSKA,1981].

Powyższy schemat, nieco już przestarzały, jest oparty na określeniu ułożenia masy języka w płaszczyźnie poziomej (określenia "przednie", "centralne" i "tylne") oraz w płaszczyźnie pionowej, w której opisuje się względem stopnia wysklepienia masy języka ku górze (określenia "wysokie", "średnie" i "niskie"). Warto przy tym zwrócić uwagę, że różnica wysokości między średnią-przednią i średnią-tylną jest mniej istotna do jednoznacznej klasyfikacji głosek, niż różnica przesunięcia w płaszczyźnie poziomej

masy języka, tym nie mniej, jak później przekonamy się, nawet niewielkie stosunkowo zmiany ułożenia języka mogą mieć wymierny wpływ na mierzone cechy akustyczne głosek. W sumie, przedstawione powyżej granice między poszczególnymi głoskami są opisowe i mają raczej umowny charakter. Obecnie raczej stosuje się podział samogłosek według miejsca artykulacji na dwie tylko klasy – przednie i tylne [DUKIEWICZ, 1995].

Powyższy schemat klasyfikacyjny samogłosek znajduje swoje odbicie w ich strukturze widmowej, a przede wszystkim w częstotliwościach formantowych, zwłaszcza w częstotliwościach pierwszego i drugiego formantu. W Tabeli 4.1 podano średnie zakresy częstotliwości pierwszych czterech formantów F1, F2, F3 i F4 samogłosek polskich ([CALIŃSKI i in., 1970]). Wartości średnie otrzymano z 5 powtórzeń każdej samogłoski ustnej wypowiedzianej w izolacji, dla każdej z 10 osób oddzielnie (5 głosów męskich i 5 głosów żeńskich).

**Tab. 4.1** Średnie częstotliwości formantowe samogłosek polskich  
(na podstawie [CALIŃSKI i in., 1970])

Samogłoska	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]
/i/	188-275	2078-2836	2670-3432	3316-4144
/ɪ/	262-391	1689-2362	2424-3146	3124-4226
/e/	524-630	1580-2228	2468-3146	3064-4034
/a/	683-1021	1132-1566	2328-2860	3098-4088
/o/	493-679	788-1100	2410-3026	3194-3954
/u/	242-338	558-789	2266-3188	2942-4058

Z przytoczonych w Tab. 4.1 danych i schematu klasyfikacyjnego podanego na Rys. 4.12 widać wyraźnie, że im wyżej jest położona masa języka, tym niższa jest częstotliwość pierwszego formantu  $F_1$ , natomiast im bardziej jest ona przesunięta ku przodowi, tym wyższa jest częstotliwość drugiego formantu  $F_2$ . Warto zaznaczyć przy tym, że choć klasyfikacja ułożenia masy języka na trzy grupy, t.j. przednie, centralne i tylne, znajduje swoje odbicie na płaszczyźnie formantowej F1-F2 (zakresy wartości pozostałych formantów silnie zachodzą na siebie) to nie jest ono jednoznaczne, pomimo że częstotliwości drugiego formantu są silnie związane z miejscem artykulacji. Dotyczy to szczególnie samogłosek *i*, *e*, których zakresy wartości drugiego formantu  $F_2$  w znacznym stopniu pokrywają się. Klasyfikacja według wysokości artykulacji wydaje się być odwzorowywana bardziej jednoznacznie na płaszczyźnie formantowej (w częstotliwości  $F_1$ ).

Bardzo specyficzną grupą samogłosek w języku polskim są samogłoski nosowe / $\tilde{e}$ ,  $\tilde{o}$ /, które nie zostały zaznaczone na podanych schematach klasyfikacji głosek. Artykulacja samogłosek nosowych jest bardzo niestabilna i silnie zależy od nawyków osoby mówiącej, staranności wymowy, a także kontekstu, w jakim są wymawiane.

Dane w literaturze dotyczące artykulacji i cech widmowych samogłosek nosowych są dość niepełne i rozbieżne. W zasadzie przyjmuje się, że ich artykulacja w języku polskim jest asynchroniczna i ma charakter dyftongiczny (WIERZCHOWSKA, 1980). Na początku masa języka przybiera położenie takie jak podczas wymawiania odpowiedniego wariantu ustnego, a następnie języczek podniebienia miękkiego zaczyna się obniżać otwierając w ten sposób wlot do jamy nosowej. Artykulacja samogłoski nosowej kończy się na ogół spółgłoskowym segmentem nosowym. W przypadku następowania po nim spółgłoski zwarto-wybuchowej lub zwartej jego miejsce artykulacji jest zbieżne z jej miejscem artykulacji.

W przypadku artykulacji asynchronicznej, wskutek płynności przejścia między częścią ustną i nazalizowaną, trudno jest na podstawie odsłuchów i analizy widmowej określić między nimi wyraźną granicę. Płynność tego ruchu właściwie nie pozwala na interpretację bifonemiczną artykulacji samogłosek nosowych sugerowaną przez niektórych autorów (patrz np. [BIEDRZYCKI, 1978]). Interpretują oni samogłoski nosowe jako połączenie odpowiedniej samogłoski ustnej i nazalizowanej płynnej spółgłoski /w/ lub rzadziej /ɲ/. W zasadzie jest to raczej jedna z możliwych odmian artykulacji i trudno ustalić sztywne reguły uzależniające ich wymowę od kontekstu, bowiem silny wpływ na artykulację mają również indywidualne cechy mówcy. Należy podkreślić, że dyftongiczny charakter samogłosek nosowych powoduje ich wrażliwość na otoczenie. Przebieg ich artykulacji zależy zarówno od typu głoski poprzedzającej, jak i po nich następującej [DUKIEWICZ, 1995].

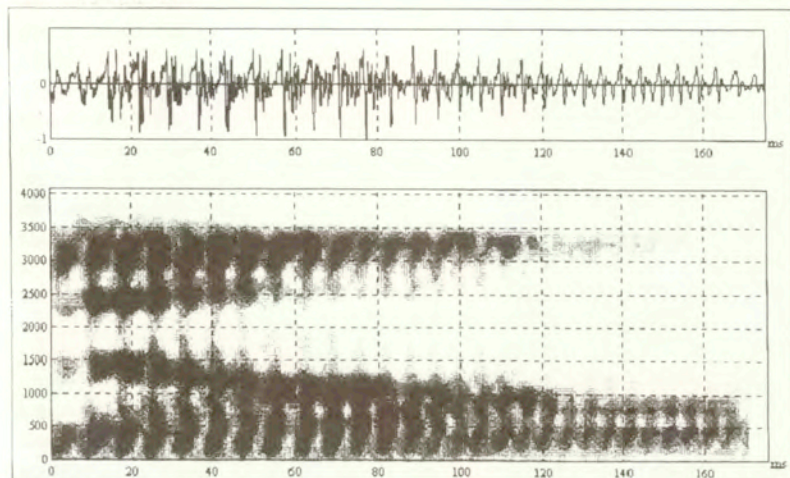
Często jednak występuje tendencja do całkowitej denazalizacji samogłosek nosowych i wymawiania jej w postaci sekwencji odpowiedniej samogłoski ustnej, po której następuje spółgłoska nosowa (najczęściej /m/, rzadziej /ɲ/). Ma to szczególnie miejsce, gdy po samogłosce nosowej następuje spółgłoska o przedniej artykulacji, ale nie jest to niezbędne. Na przykład przed spółgłoską welarną (np. przed /k/) inny jest typ spółgłoski nosowej następującej po samogłosce (wówczas jest /ɲ/, jak ma to miejsce w wyrazie „ręka” -> /rɛŋka/. Ogólna tendencja obserwowana obecnie w wymowie jest raczej denalizować całkowicie samogłoski nosowe, lub wymawiać je w postaci sekwencji samogłoski ustnej i spółgłoski nosowej, o zbliżonym miejscu artykulacji do spółgłoski po niej następującej. Na przykład, spółgłoskę /m/ w przypadku wyrazu „kapiel” (/kompjel/), /m/ w wyrazie „kęs” (/kɛns/), itp. [WIERZCHOWSKA, 1980]. Niestabilność artykulacji samogłosek nosowych powoduje, że trudno jest wyznaczyć ich niezmiennie cechy artykulacyjne (zwłaszcza stopień nazalizacji), a i tym samym ich cechy akustyczne. Dodatkowym utrudnieniem jest to, że samogłoska nosowa często jest inaczej artykułowana niż jej ustny wariant i w zasadzie na podstawie zdjęć kina rentgenograficznych można sądzić, że jej miejsce artykulacji jest mniej lub bardziej (w zależności od kontekstu) przesunięte ku tyłowi (por. przekroje artykulacyjne w [BOLLA&FÖLDI, 1987]) Warto przy tym wspomnieć, że nieregularny charakter polskich samogłosek nosowych



w przypadku języka polskiego jest przyczyną kwestionowania ich istnienia jako odrębnej klasy głosek [GUSSMAN,1980; RUBACH,1977a, b].

Obecność antyformantów w widmie samogłosek nosowych w mowie naturalnej jest często bardzo trudna do zaobserwowania, tym niemniej, jak dalej będzie można się przekonać, ich istnienie można stosunkowo łatwo wykazać na drodze analitycznej za pomocą modelu artykulacyjnego toru głosowego.

Na Rys. 4.13 przedstawiono końcowy fragment przebiegu zmian widma wraz z oscylogramem podczas artykulacji wyrazu "chwila?" wypowiedzianego na końcu zdania pytającego.



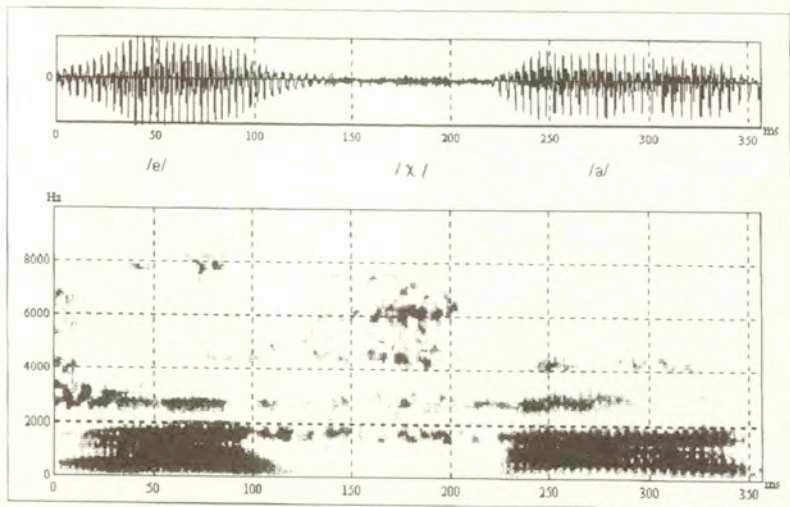
Rys. 4.13 Oscylogram i spektrogram przebiegu zarejestrowanego dla dyftongicznej wariantu samogłoski nosowej /o/ wypowiedzianej w wyrazie "chwila?" na końcu zdania pytającego.

Na rysunku, na którym przedstawiono przykład dyftongicznej realizacji samogłoski nosowej, można zaobserwować w obrębie samogłoski nosowej płynny ruch formantów od położenia odpowiadającego końcowemu fragmentowi głoski /l/ (pierwsze 15 ms na poniższym przebiegu) do położenia odpowiadające spółgłosce nosowej /m/ (począwszy od ok. 120 ms). Zmiana konfiguracji toru, a zwłaszcza stopniowe włączanie się części nosowej do formowania głoski powoduje obniżanie się częstotliwości drugiego formantu (rzędu 20%), a także pojawienie się tzw. antyformantów powstających wskutek wzajemnego bocznikowania się obu części toru głosowego.

### 4.3 Głoski szczelinowe (trące)

Spośród spółgłosek nie zaliczanych do klasy rezonantów, jedynie głoski trące (szczelinowe) są głoskami monosegmentalnymi i mogą być wymówione w sposób ciągły bez udziału samogłoski. Są one stacjonarne i charakteryzują się udziałem źródła szumowego w pobudzeniu toru głosowego. Źródło to powstaje u wylotu szczeliny utworzonej między językiem i podniebieniem (lub górnymi dżiąsłami) lub na krawędzi górnych zębów, na którą naprowadzany jest strumień powietrza. W obu przypadkach, w momencie gdy szybkość przepływu strumienia powietrza osiąga dostateczną wielkość u wylotu szczeliny przepływ powietrza staje się turbulentny i zawirowania jego stają się źródłem pobudzenia o charakterze szumowym. W zależności od położenia w torze głosowym tego źródła zależy charakterystyka widmowa artykułowanej głoski. Wiąże się to przede wszystkim z charakterystyką funkcji przenoszenia tej części toru głosowego, która jest pobudzana przez źródło szumowe, czyli przedniej części toru zawartej między utworzoną szczeliną i wylotem uformowanym przez wargi. Ponieważ szczelina ma bardzo małą powierzchnię przekroju w porównaniu z powierzchnią przekroju tylnej części toru głosowego, można więc w pierwszym przybliżeniu przyjąć, że wpływ tej części na charakterystykę artykułowanego dźwięku jest pomijalnie mały. Fakt ten potwierdzają badania modelowe [LEE&CHILDERS,1997], które dodatkowo wykazały, że z częstotliwości rezonansu można wyznaczyć efektywną długość przedniego odcinka toru głosowego formującego bezdźwięczną głoskę trącą. Przytoczone wyniki w cytowanej pracy potwierdziły, że wpływ tylnej części toru głosowego rozciągającej się między głośnia i szczeliną jest pomijalnie mały. Tak więc położenie miejsca artykulacji w torze głosowym, a tym samym usytuowanie źródła szumowego, ma decydujący wpływ na widmo generowanego dźwięku. Na kolejnych rysunkach przedstawiono przebiegi amplitudowe i widmowe spółgłosek trących o różnym miejscu artykulacji. Porównując ze sobą Rys. 4.14, 4.16 i 4.17 widać wyraźnie, jak wraz z artykulacją spółgłoski trącej o miejscu artykulacji coraz bardziej przesuniętym ku przodowi (od /x/, /ʃ/, /ç/, /s/ aż do /f/) obszar koncentracji energii w widmie przesuwa się w stronę coraz wyższych częstotliwości (spółgłoska /ç/, dla której nie podano przykładu, ma miejsce artykulacji bliskie spółgłosce /ʃ/).

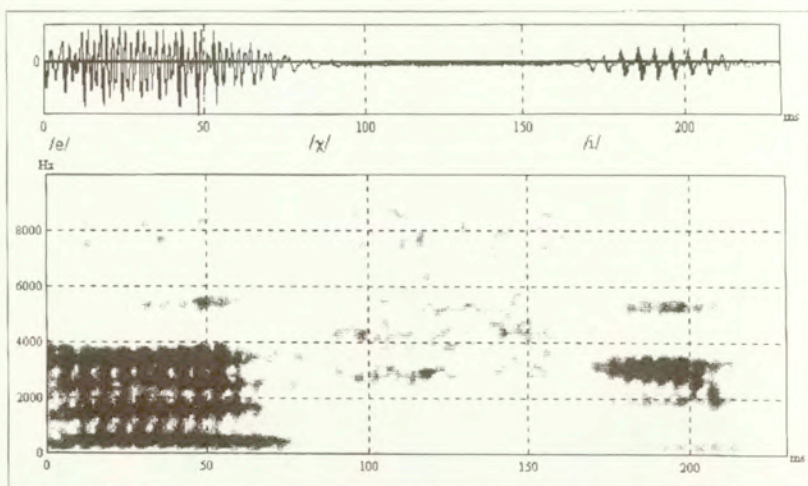
W przypadku artykulacji głoski /x/, jak w wyrazie "z orzechami" (Rys. 4.14), w formowaniu jej widma uczestniczy znaczna część długości toru głosowego (bowiem szczelina jest utworzona w jego tylnej części, między tylną częścią języka i podniebieniem miękkim). Z tego powodu w obrazie widmowym głoski /x/ wyraźnie widoczne są formanty (zwłaszcza  $F_2=1700$  Hz i  $F_3=2800$  Hz) ilustrujące przejście układu artykulacyjnego od konfiguracji samogłoskowej /e/ do konfiguracji odpowiadającej samogłosce /a/.



Rys. 4.14 Oscylogram i spektrogram przebiegu zarejestrowanego dla sekwencji głosek /exa/ wypowiedzianej w wyrazie "orzeciami".

Wpływ sąsiedztwa na artykulację głoski /x/ jest znaczny i objawia się on przede wszystkim zmianą położenia szczeliny, która przesuwa się w kierunku miejsca artykulacji głoski sąsiadującej (zarówno poprzedzającej, jak i następującej). W obrazie widmowym uwidacznia się ten wpływ zmianą przede wszystkim częstotliwości drugiego formantu. Dla porównania, jak kontekst samogłoskowy może wpływać na artykulację tej spółgłoski, na Rys. 4.15, przedstawiono oscylogram i sonagram przebiegu, zarejestrowanego dla sekwencji głosek /exi/ wypowiedzianej w wyrazie "Lechii". Gdy głoska ta jest wypowiedziana w kontekście /e-i/, obserwuje się przesunięcie się dolnego zakresu szumu powyżej 2000 Hz (-20 dB od wartości maksymalnej) i przesunięcie się całego zakresu szumów w stronę wyższych częstotliwości ( $F_2 = 2800$  Hz,  $F_3 = 4250$  Hz). Jest to wynikiem oddziaływania sąsiedztwa samogłoski /i/ na artykulację spółgłoski /x/. Ogólnie biorąc, dla każdej głoski trącej obserwuje się wpływ głoski po niej następującej na usytuowanie w torze głosowym szczeliny, zwłaszcza gdy jej miejsce artykulacji jest odległe w znacznym stopniu od miejsca położenia szczeliny. Jednakże wpływ ten jest najsilniejszy dla głoski /x/.

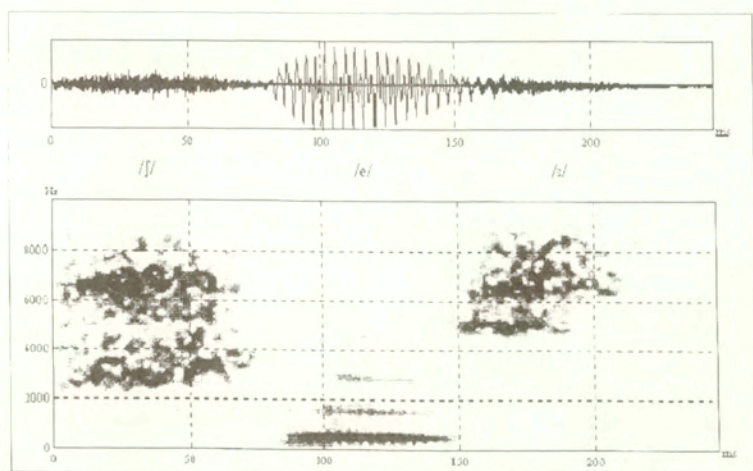
Dla głosek trących o bardziej przesuniętym do przodu miejscu położenia szczeliny, takich jak /s,ʃ/ koncentracja energii w widmie przesuwa się w kierunku wyższych częstotliwości (por. Rys. 4.16).



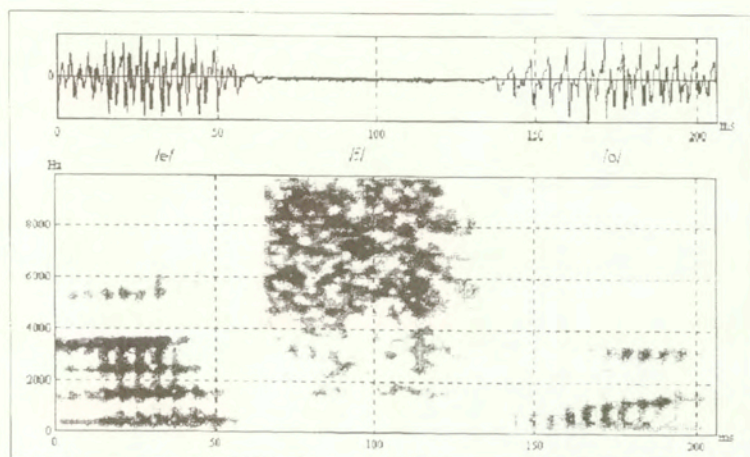
Rys. 4.15 Oscylogram i spektrogram przebiegu zarejestrowanego dla sekwencji głosek /exi/ wypowiedzianej w wyrazie „Lechii”.

Wynika to z wyższych częstotliwości rezonansowych krótszych odcinków toru głosowego formujących widmo tych spółgłosek. Dla spółgłoski /j/ obserwuje się dwa obszary koncentracji energii w widmie, w pobliżu 2700 Hz oraz 6600 Hz, natomiast dla spółgłoski /s/, o bardziej przedniej artykulacji - 5000 Hz oraz 7000 Hz. W drugim przypadku różnice między maksimami nie przekraczają 6 dB i można przyjąć, że obszar szumu głoski /s/ zawiera się w przedziale częstotliwości 4800-9000 Hz.

Jednakże w przypadku głoski /f/, efektu przesuwania się obszaru koncentracji energii widmie w kierunku wyższych częstotliwości wraz z przesuwaniem się miejsca artykulacji ku przodowi (głoska ta jest wargowo-zębowa) nie obserwuje się. Ponieważ źródło pobudzenia szumowego znajduje się niemal u wylotu toru głosowego, charakterystyka głoski jest formowana niemal wyłącznie przez charakterystykę promieniowania ust. Charakterystyka widmowa źródła szumu turbulentnego zależy od konfiguracji toru głosowego. Jeżeli źródło jest utworzone tylko przez szczelinę, tak jak to ma miejsce w przypadku głosek /s/ czy /f/, to obwiednia jego charakterystyki widmowej opada  $-6$  dB/okt, natomiast w przypadku gdy źródło szumowe powstaje w wyniku skierowania strumienia powietrza ze szczeliny na krawędź, tak jak to jest w przypadku głoski /f/, to obwiednia jego widma rośnie asymptotycznie  $+6$  dB/okt [LIN,1990]. Tak więc dla głoski /f/ jest charakterystyczne, że widmo jej jest narastające wraz ze wzrostem częstotliwości i praktycznie zakres widma jest ograniczony przez częstotliwość próbkowania sygnału (20 kHz), a nie przez charakterystyki rezonansowe toru głosowego.

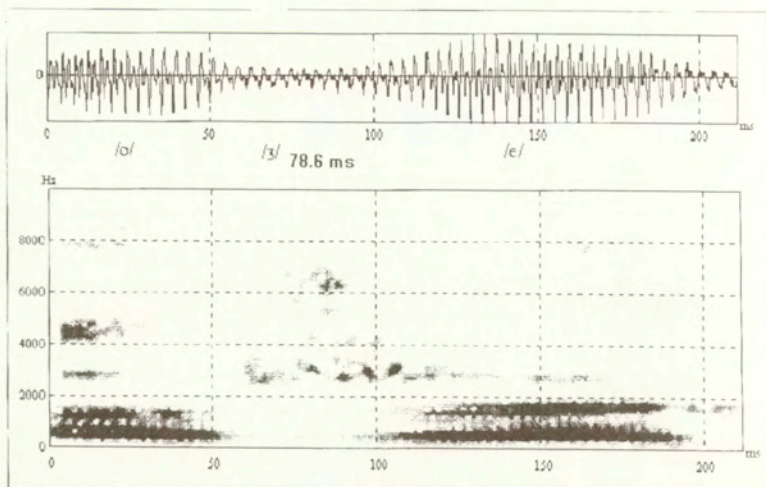


Rys. 4.16 Oscylogram i spektrogram przebiegu zarejestrowanego dla sylaby /jes/ wypowiedzianej w wyrazie "szesnaście".



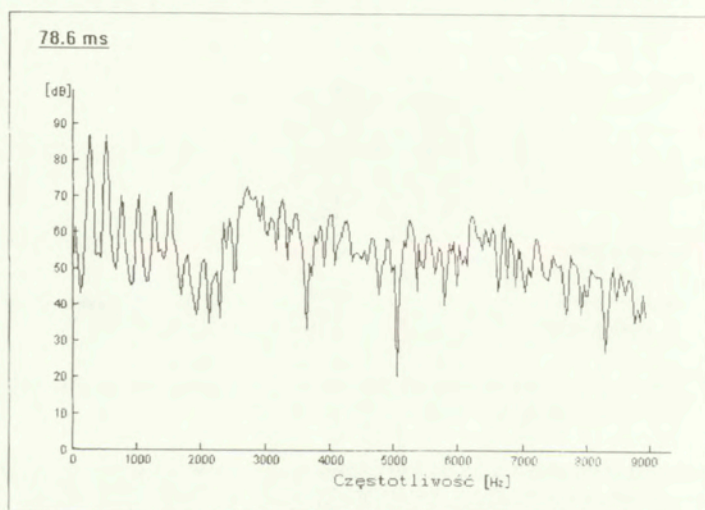
Rys. 4.17 Oscylogram i spektrogram przebiegu zarejestrowanego dla sylaby /efo/ wypowiedzianej w wyrazie "telefoniczne".

Głoski szczelinowe o mieszanym pobudzeniu, to jest, dźwięczno-szumowe, nie różnią się artykulacją od swoich bezdźwięcznych odpowiedników. Znamienne jest występowanie w widmie składowej harmoniczej o częstotliwości podstawowej o dość znacznej energii, natomiast składowa szumowa jest znacznie słabsza, niż w przypadku głosek bezdźwięcznych (por. głoski /f/ i /ʃ/ - Rys. 4.16 i 4.18).



Rys. 4.18 Oscylogram i spektrogram przebiegu zarejestrowanego dla sekwencji głosek /o3e/ wypowiedzianej w wyrazie "orzechami".

Na Rys. 4.19 przedstawiono przekrój widmowy pomierzony dla głoski /3/, w odległości 78.6 ms od początku sygnału. Na przekroju tym jest widoczna bardzo



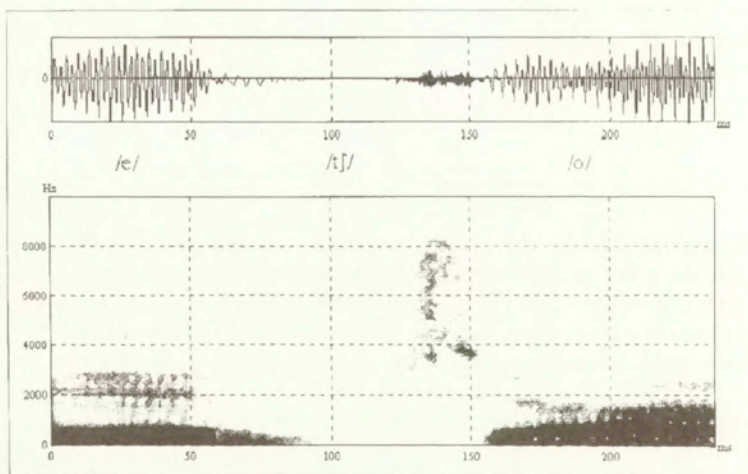
Rys. 4.19 Widmo głoski /3/ wypowiedzianej w wyrazie "orzechami" wyznaczone dla przekroju 78.6 ms odległym od początku sygnału (patrz Rys. 4.17).

wyraźna struktura harmoniczna ( $F_0 \approx 250$  Hz) widma w dolnym zakresie częstotliwości aż do ok. 2000 Hz, która jest w zakresie wyższych częstotliwości „przykryta” przez składową szumową. Obszary koncentracji energii są położone w podobnych pasmach częstotliwości, które obserwuje się dla głoski /f/.

#### 4.4 Głoski zwarto-szczelinowe

Specyficzną grupą dla języka polskiego są spółgłoski zwarto-szczelinowe (zwarto-trące), które często literaturze przedstawia się jako sekwencję odpowiedniej głoski wybuchowej (/t/ lub /d/, po której następuje właściwa głoska trąca. Jednakże badania przeprowadzone przez Stevensa [STEVENS, 1993], jednoznacznie wykazują na przykładzie głoski zwarto-trącej /tʃ/, że taki sposób interpretowania artykulacji głosek zwarto-trących jest niewłaściwy, ponieważ w ich początkowej fazie rozwarcia widmo generowanego w tym momencie szumu nie jest podobne do widma głoski zwarto-wybuchowej. Jest to spowodowane płynną zmianą położenia miejsca zwarcia do miejsca, w którym jest utworzona szczelina o dostatecznej wielkości. To przejście trwa stosunkowo długo, około 20-30 ms i generowany szum ma zmieniającą się charakterystykę widmową. Również i pod względem akustycznym spółgłoska zwarto-trąca różni się od odpowiadającej jej sekwencji głoski wybuchowej i trącej (można porównać tu na przykład, artykulację sekwencji wyrazów „koc spadł” oraz „kot spadł”). W artykulacji głosek zwarto-trących wyróżnia się dwie fazy - fazę zwarcia, po której następuje przejście do uformowania w torze ustnym szczeliny, w miejscu odpowiednim do artykulacji segmentu trącego.

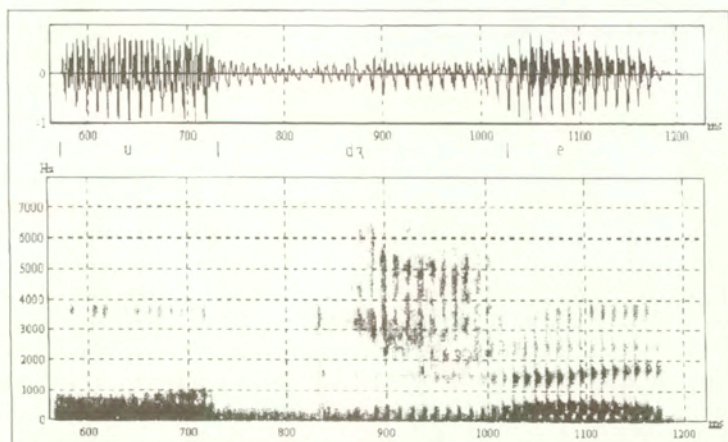
Ruch języka podczas wymawiania głoski zwarto-trącej jest złożony, zwłaszcza podczas ich wymawiania w otoczeniu rezonantów. Gdy głoska ta następuje po rezonancie, faza zwarcia często jest niekompletna, co objawia się brakiem wyraźnej ciszy w fazie zwarcia podczas artykulacji głoski bezdźwięcznej, lub nienaturalnie dużym poziomem składowej dźwięcznej podczas artykulacji takich głosek jak /dʒ/ czy /dʒ/, lub składa się z dwóch segmentów - segmentu dźwięcznego zwarcia następującego po głosce dźwięcznej i segmentu bezdźwięcznego (por. Rys. 4.20).



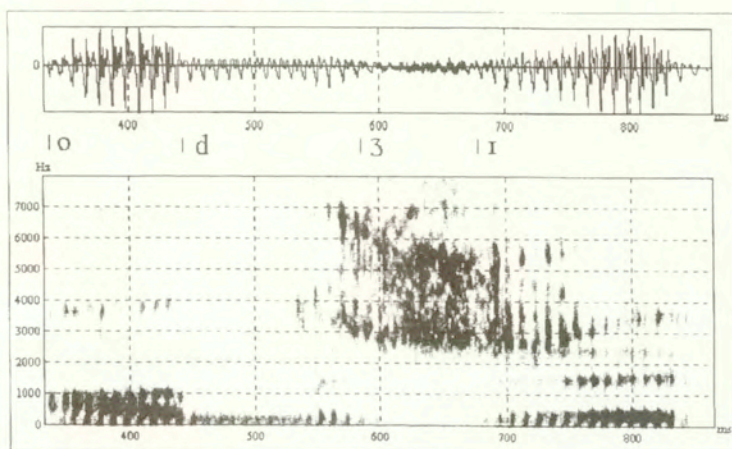
Rys. 4.20 Oscylogram i spektrogram przebiegu zarejestrowanego dla sekwencji głosek /eʃo/ wypowiedzianej w wyrazie "wieczorami".

Na Rys. 4.21 i 4.22 przedstawiono przebiegi dwusylabowe /u-dʒ-e/ oraz /o-dʒ-ɪ/ wymówionych w wyrazach "udrzeć" oraz "odzywka". Ilustrują one różnice w artykulacji głoski /dʒ/ i sekwencji /d-ʒ/. W przypadku artykulacji sekwencji /d-ʒ/ obserwuje się wyraźne zwiększenie amplitudy sygnału przed przejściem do głoski trącej oraz pojawienie się w widmie formantów o częstotliwościach zbliżonych do częstotliwości neutralnego /ə/. Występowanie segmentu samogłoskopodobnego, o stosunkowo krótkim czasie trwania (40-60 ms), w sekwencji dźwięcznej, między spółgłoską płozyjną i trącą jest dosyć częste. W przypadku artykulacji głoski zwarto-trącej, przejście między segmentem zwarcia zakończonym krótkim i o małej amplitudzie impulsem oraz segmentem trącym jest bardziej płynne.





Rys. 4.21 Oscylogram i spektrogram przebiegu zarejestrowanego dla dwusylabowej sekwencji /udze/ wypowiedzianej w wyrazie "udrzeć".

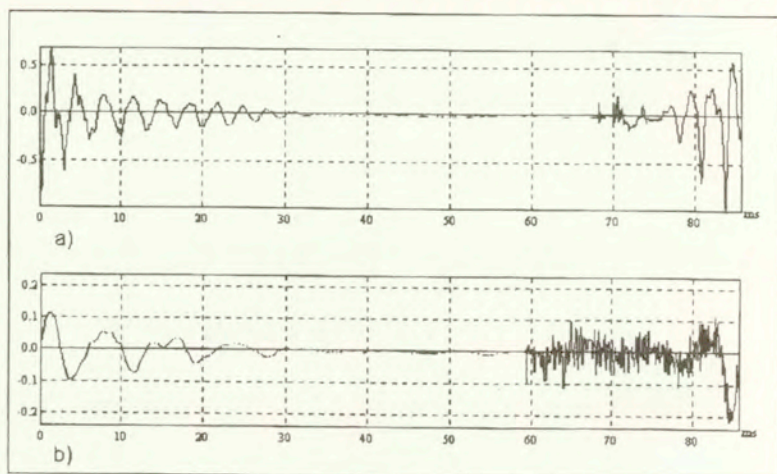


Rys. 4.22 Oscylogram i spektrogram przebiegu zarejestrowanego dla sekwencji dwóch sylab /odzi/ wypowiedzianej w wyrazie "odzywka".

#### 4.5 Głoski zwarto-wybuchowe

Podobnie jak w przypadku głosek zwarto-trących, artykulacja głosek zwarto-wybuchowych (plozyjnych) składa się z dwóch odrębnych faz - fazy zwarcia, po której następuje plozja mająca postać krótkiego impulsu >10 ms. Podobnie jak w przypadku innych głosek zwartych miejsce artykulacji może być położone między

wargami (głoska /p/) i tylną częścią języka (głoska /k/) i zależy w pewnym stopniu od kontekstu, zwłaszcza gdy miejsce artykulacji głoski następującej jest w znacznym stopniu odległe od miejsca zwarcia. Bardzo często w przypadku artykulacji głosek wybuchowych po impulsie występuje stosunkowo krótki segment szumowy, zwany przydechem (aspiracją), który w przeciwieństwie do niektórych spółgłosek innych języków, np. języka angielskiego, nie nosi informacji fonologicznych. Szum ten powstaje w trakcie rozwierania się miejsca zwarcia, tworzącego przez stosunkowo krótki okres szczelinę. Na Rys. 4.23 przedstawiono przebiegi oscylograficzne zarejestrowane dla spółgłoski /p/ wypowiedzianej w sekwencji "to pani" oraz w wyrazie "kupiłem". W pierwszym przypadku przy artykulacji spółgłoski /p/ po zwarciu następuje krótka plozja (~5 ms) i przejście do samogłoski otwartej, natomiast w wyrazie "kupiłem", plozja ta jest mniej silna i po niej następuje aspiracja o czasie trwania blisko 20 ms. W drugim przypadku, szybkość rozwierania jest mniejsza, bowiem następuje przejście do przedniej samogłoski przymkniętej /i/, wskutek czego przez pewien czas w torze głosowym istnieje szczelina powodująca turbulentny przepływ powietrza. W tym przypadku głoska składa się z 3 segmentów. Warto zwrócić uwagę, że zwarcie rozpoczyna się na końcowej (ok. 10 ms) części samogłoski /o/ (czy /u/), po której następuje segment ciszy, poprzedzający plozję.



**Rys. 4.22** Porównanie przebiegów czasowych spółgłoski /p/ wymówionej w sekwencji "to pani" (a) oraz w wyrazie "kupiłem" (b). W drugim przypadku widoczny jest występujący po plozji (ok. 60 ms na skali czasu) segment przydechowy.

W przypadku zwarto-wybuchowych dźwięcznych (/b,d,g/) podczas całego okresu zwarcia fałdy głosowe w zasadzie wibrują, choć niekiedy można zaobserwować w nim krótki (ok. 10-20 ms) segment ciszy.

Na Rys. 4.24 przedstawiono tradycyjny schemat klasyfikacyjny spółgłosek i oparty na sposobie i miejscu artykulacji, przy jednoczesnym uwzględnieniu rodzaju pobudzenia toru głosowego. Kolejność klas głosek została uporządkowana według stopnia otwarcia (tj. stopnia zbliżenia narządów mowy) malejącego w dół. U góry diagramu umieszczono najpierw te spółgłoski, które nie mają odpowiedników bezdźwięcznych. W klasie głosek otwartych wyróżniono tylko spółgłoski /j/ i /w/. Do klasy tej należą również samogłoski. Gdyby je nanieść na ten schemat, znalazłyby się one w klasie głosek palatalnych i tylny-językowych.



Rys. 4.23 Schemat klasyfikacji spółgłosek polskich w oparciu o sposób i miejsce artykulacji oraz typ pobudzenia toru głosowego.

Schemat umożliwia syntetyczne przedstawienie głosek według ich podobieństwa artykulacji, ale jednocześnie ich zróżnicowanie pod względem trzech zasadniczych parametrów artykulacyjnych, to jest sposobu, miejsca artykulacji oraz rodzaju pobudzenia.

#### 4.6 Koartykulacja w języku polskim

Jak już wspomniano w Rozdz. 3 każda wypowiedź stanowi w swoisty sposób zorganizowaną całość, co przejawia się zarówno w jej strukturze segmentalnej (głoskowej) i sylabicznej, jak też w jej rozczłonowaniu rytmicznym i melodycznym. W

każdej wypowiedzi można wyróżnić trzy zasadnicze fazy: *nagłos* - poprzedzony pauzą i stanowiący jej początkową część, *śródgłos* - część środkowa oraz *wyglós* - część końcowa, po której następuje pauza. Mowa jest realizowana jako ciągły proces, którego źródłem jest płynnie zmieniający swoją konfigurację układ artykulacyjny. Narzucony jakby z góry dyskretny opis sygnału mowy w postaci sekwencji wyżej opisanych jednostek fonologicznych, w rzeczywistości jest tylko pewnym uproszczonym przedstawieniem w płaszczyźnie językowej procesu mowy. Już z przedstawionego powyżej opisu dźwięków mowy polskiej można zorientować się, że na artykulację wielu głosek silnie mogą wpływać dźwięki sąsiadujące z nimi. Zjawisko koartykulacji definiuje się zazwyczaj jako nakładanie się ruchów artykulacyjnych właściwych sąsiadującym ze sobą głoskom i w zasadzie prowadzi do zacierania różnic między sąsiadującymi ze sobą dźwiękami i tym samym następuje niekiedy zmiana postaci dźwiękowej form wyższego rzędu, na przykład wyrazowych [WIERZCHOWSKA, 1980]. Wpływ ten może objawiać się na przykład modyfikacją miejsca, lub wysokości artykulacji, czy zmianą typu pobudzenia głoski. Zmiany te jednak nie zawsze są zazwyczaj tak silne, że następuje zmiana kategorii słyszanej głoski, choć i to niekiedy ma miejsce, zwłaszcza w przypadku zmiany typu pobudzenia. Podstawowe zjawiska występujące przy koartykulacji to upodobnienia, palatalizacja, zmiana typu pobudzenia oraz denazalizacja samogłosek nosowych o czym już była wcześniej mowa..

#### 4.6.1. Upodobnienia [WIERZCHOWSKA, 1980]

Ze względu na kierunek następujących upodobnień wyróżnia się upodobnienia regresywne i progresywne. W języku polskim najczęściej występuje upodobnienie pierwszego typu, które polega na wcześniejszym, częściowym przygotowaniu układu artykulacyjnego do wymówienia następnej głoski. Na przykład zaokrąglenie wargowe charakteryzujące samogłoskę /u/ zaczyna się zazwyczaj podczas wymawiania dźwięku poprzedzającego tę głoskę. Stosunkowo rzadziej występują upodobnienia drugiego typu, polegające na przedłużaniu na głoskę następną układu narządów mowy właściwego głosce poprzedzającej. Szczególnie charakterystyczne upodobnienia występują na stykach wyrazów wymówionych w sposób ciągły. Na przykład sekwencja wyrazów "jak duży", wymówiona w taki sposób jest najczęściej realizowana w postaci /jagduzi/, nie wykluczając jednak możliwości wymówienia jej również w postaci /jak#duzi/ z krótkim segmentem ciszy rozdzielającym oba wyrazy. Wskutek upodobnienia pod względem dźwięczności następuje zamiana głoski /k/ w /g/. Upodobnienia wewnątrz wyrazowe mają na ogół trwały charakter, ich wymowa jest ustabilizowana, i w transkrypcji fonetycznej tych wyrazów są niezmiennie (np. "babka" -> /bapka/, "stołówka" -> /stowufka/, "skrzat" -> /skʃat/, prośba -> /prozba/ itp.). Upodobnienia pod względem miejsca lub wysokości artykulacji nie muszą mieć tak drastycznego wpływu na zmianę formy dźwiękowej wyrazu jak upodobnienia pod względem dźwięczności i ujawniać się mogą, na przykład stosunkowo niewielką

zmianą częstotliwości formantowych wpływającą tylko barwę postrzeganej głoski, a nie jej kategorii.

Pozycja głoski w frazie również wpływa na jej artykulację. Każda wypowiedź może być podzielona na frazy przedzielone między sobą pauzami. W języku polskim istnieje silna tendencja do osłabiania wygłosu absolutnego (na końcu wypowiedzi) i jego ubezdźwięczniania. Tendencji tej nie ulegają rezonanty, choć gdy są wymawiane w grupie spółgłoskowej, także tracą dźwięczność. Natomiast w nagłosie i w śródgłosie grupy spółgłoskowe utworzone z 2 lub więcej spółgłosek zwarto-wybuchowych, zwarto-szczelinowych i szczelinowych są albo całkowicie dźwięczne, albo całkowicie bezdźwięczne, tak jak to przedstawiono w przykładach na upodobnienia pod względem dźwięczności. Grupy mieszane pod względem dźwięczności w języku polskim tworzyć mogą tylko rezonanty, jednakże wymawiane w środku utworzonych większych grup spółgłoskowych stają się najczęściej bezdźwięczne i są tak słabe, że często ulegają całkowitej redukcji, jak na przykład w wyrazach "jabłko" -> /japko/ zamiast /jabwko/, "rzemieślnik" -> /ɕemjɛɕnik/ zamiast /ɕemjɛɕɲik/ itp.

#### 4.6.2 Palatalizacja

Głoski o przednim układzie masy języka charakteryzują się wysklepieniem jego przedniej części ku górze, w stronę przedniej części podniebienia twardego. Stopień tego wysklepienia bywa różny - od niewielkiego, właściwego głoskom otwartym (np. /a/ do maksymalnego, właściwego, głoskom szczelinowym typu /i/, /ɛ/). Przy artykulacji głosek zwarto-wybuchowych, zwarto-szczelinowych, a także rezonantów spółgłoskowych wymawianych przed spółgłoską /j/ i samogłoską /i/ wyróżnia się dwa ruchy, jeden zasadniczy, polegający na przykład na tworzeniu zwarcia lub szczeliny o podobnym układzie masy języka jak dla głoski niezmiękczonej (twardej) oraz dodatkowy, polegający na jednoczesnym wzniesieniu środka języka do podniebienia twardego powodujący palatalizację, tj. zmiękczenie głoski i poszerzający zwarcie lub szczelinę w kierunku środka jamy ustnej. należą do kategorii głosek określanych jako miękkie. Informację o tym dodatkowym ruchu niesie zazwyczaj obserwacja przebiegu zmian częstotliwości formantowych, zwłaszcza pierwszego i drugiego formantu. Zmiękczenie objawia się obniżeniem F1 i wzrostem F2.

Przedstawiony opis artykulacyjny wraz z cechami akustycznymi dźwięków mowy polskiej ma istotne znaczenie praktyczne. W obecnym czasie wraz z rozwojem systemów dwustronnej komunikacji za pomocą mowy, opartych na zastosowaniu próbek mowy naturalnej (tzw. corpus-based systems) w procesie nauczania systemów rozpoznawania mowy, jak również w układach syntezy [YOUNG& BLOOTHOOFT,1997]. W celu wyznaczenia reprezentatywnych dla danego języka realizacji dźwiękowych poszczególnych próbek o rozciągłości przyjętej w danym systemie za podstawową (np. sylaba, difon, głoska itp.) należy dysponować

odpowiednio dużymi zbiorami nagrań tekstów wypowiedzianymi przez dużą liczbę osób. W celu zoptymalizowania pod względem wielkości tego typu zbiorów nagrań, należy dobrać teksty w taki sposób, by były one reprezentatywne dla danego języka zarówno pod względem strukturalnym, jak i dźwiękowym. Reprezentatywne dla danej populacji językowej tego typu zbiory nagrań, które zostały dokonane w uporządkowany sposób oraz zaopatrzone w szczegółowy opis, w którym każdemu fragmentowi zarejestrowanego sygnału jest przyporządkowana określona etykieta (np. nazwa głoski, sylaby itp.), przyjęto nazywać bazami danych mowy (tzw. speech database). Podstawą etykietyzacji jest właśnie jednoznaczny opis artykulacyjny głosek. Opis ten może być uproszczony i sprowadzony do postaci przedstawionej w niniejszym rozdziale, lub bardziej rozszerzony uwzględniający różne warianty artykulacyjne poszczególnych głosek, nie zmieniające ich znaczenia lecz tylko ich brzmienie. Opis ten nosi nazwę allofonicznego i stosowany jest przy bardziej szczegółowych badaniach fonetycznych. Jednakże w językowych bazach danych stosuje się na ogół opis uproszczony. W chwili obecnej dla szeregu języków istnieją już ogólnie dostępne bazy danych. Dominują bazy tworzone dla języka angielskiego pod kątem różnych zastosowań, przede wszystkim do systemów automatycznego rozpoznawania oraz syntezy mowy. Ostatnio podjęto pierwsze próby stworzenia tego typu baz danych dla mowy polskiej [GROCHOLEWSKI,1998; GUBRYNOWICZ,1998].

## 5. Zależność między konfiguracją geometryczną toru głosowego i sygnałem akustycznym mowy.

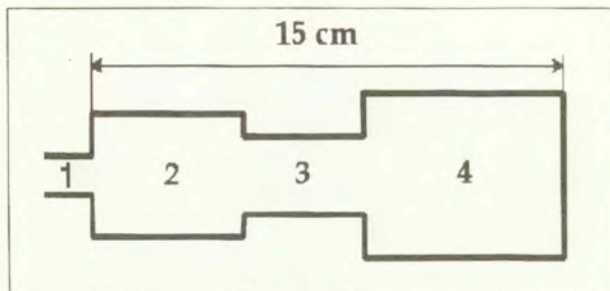
### 5.1 Wstęp

Podczas wymawiania określonego tekstu dokonywana jest realizacja zaprogramowanej sekwencji ruchów artykulacyjnych mających na celu uformowanie odpowiedniego sygnału akustycznego, o postaci umożliwiającej przesłanie danej informacji lingwistycznej. Sygnał ten jest tworzony przy udziale całego zespołu narządów artykulacyjnych, tworzących pewien złożony układ akustyczny. W pierwszym przybliżeniu można go odwzorować za pomocą zestawu szeregowo ze sobą połączonych odcinków cylindrycznych, pobudzanego na jednym końcu przez przebieg okresowy (tak jak to ma miejsce w przypadku artykulacji większości głosek dźwięcznych) i promieniującego falę akustyczną przez drugi koniec toru. Ilość i zarazem długość odcinków cylindrycznych zależy przede wszystkim od dokładności z jaką zamierza się odwzorować częstotliwościową funkcję transmitancji toru głosowego. Dla niektórych głosek (zwłaszcza samogłosek) może wystarczyć kilka segmentów o odpowiednio dobranej długości i wielkości powierzchni przekroju poprzecznego, dla innych zaś konieczne jest dokładniejsze odtworzenie geometrii toru głosowego i wówczas niezbędna liczba segmentów może wahać się między siedemnastoma a trzydziestoma kilkoma, zależnie od założonej długości toru (na ogół przyjmowana długość wynosi ok. 17,5 cm).

Skrajne uproszczenie geometrii toru głosowego, polega na modelowaniu jej za pomocą rury o stałym przekroju o powierzchni  $S=4.985 \text{ cm}^2$  i długości równej 17.8 cm. Funkcja częstotliwościowa rury o takiej geometrii ma maksima rozłożone w regularnych odstępach, w pobliżu częstotliwości 500, 1500, 2500, 3500 Hz i wyżej, i jest ona zbliżona do charakterystyki widmowej tzw. samogłosce neutralnej *ɪ* (tzw. *ie* neutralne). Jak już wspomniano, samogłoska ta nie ma znaczenia fonologicznego, jednakże występuje ona w rzeczywistym sygnale mowy stosunkowo dość często, na przykład jako segment wokaliczny spółgłoski *ri* wymawianej w nagłosie, czy w wygłosie, a także na przykład, między sekwencjami spółgłosek o kontrastowym pobudzeniu (bezdźwięczna/dźwięczna). Wyżej wymienione częstotliwości odpowiadają rezonansom kolejnych fal ćwierćfalowych, gdyż można przyjąć, że rura na poziomie głosi jest jednostronnie zamknięta.

Przykład bardziej złożonej reprezentacji geometrii toru głosowego przedstawiono na Rys. 5.1, na którym odwzorowano ją dla samogłoski */a/* za pomocą minimalnej liczby 4-ch odcinków cylindrycznych. Odcinek (1) odwzorowuje otwór wylotowy utworzony przez wargi, odcinek (2) - komorę utworzoną w przedniej części toru głosowego, odcinek (3) - zwężenie między masą języka i podniebieniem, zaś odcinek (4) - komorę w tylnej części, gardłowo-ustnej toru. Przy odpowiednim dobraniu

długości odcinków oraz ich przekrojów poprzecznych, możliwe jest uzyskanie w charakterystyce transmitancji maksimumów o częstotliwościach zbliżonych do częstotliwości formantowych samogłoski /a/.

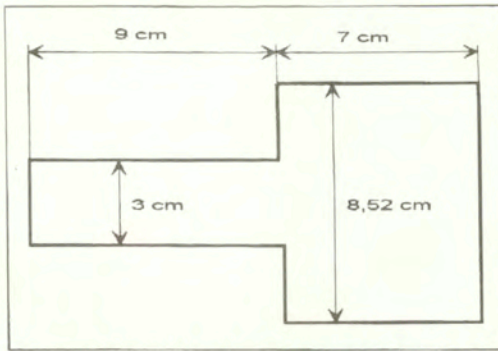


Rys. 5.1 Aproksymacja geometrii toru głosowego za pomocą 4-ch odcinków cylindrycznych.

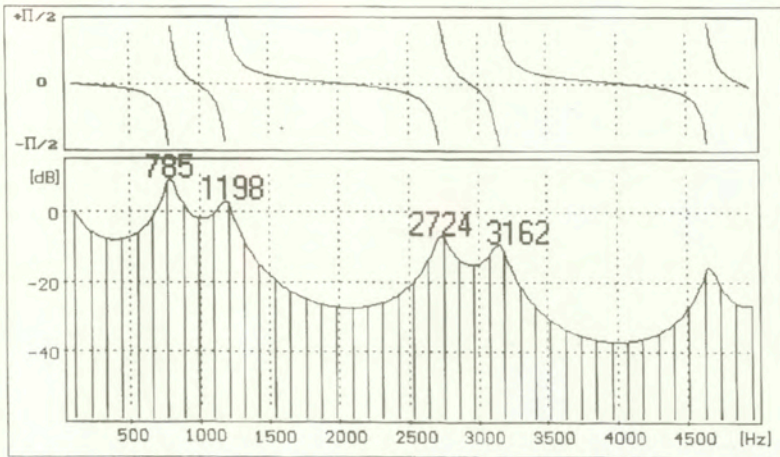
W przypadku trzech głównych z możliwych ułożeń masy języka, to jest skrajnie wypiętrzonej ku górze przedniej części języka, tylnej oraz układu płaskiego, możliwe jest przyjęcie w pierwszym przybliżeniu aproksymacji toru głosowego za pomocą jedynie dwóch odcinków rur cylindrycznych o odpowiednio dobranych przekrojach i wcale niekoniecznie odwzorowujących dokładnie rzeczywistą konfigurację geometryczną toru głosowego przybieraną podczas wymawiania określonej głoski, by uzyskać funkcję transmitancji o podobnych dla niej częstotliwościach formantowych. Na przykład, podany na Rys. 5.2 układ składający się z dwóch odcinków cylindrycznych o odpowiednio dobranych długościach i wielkościach przekrojów poprzecznych ma częstotliwości kolejnych maksimumów odpowiednio  $F_1 = 785$  Hz,  $F_2 = 1198$  Hz,  $F_3 = 2724$  Hz,  $F_4 = 3162$  Hz (Patrz Rys. 5.3). Częstotliwości te są zbliżone do częstotliwości formantowych samogłoski /a/, choć geometria toru głosowego podczas wymawiania tej głoski jest inna, niż przedstawiona na rysunku.

Na Rys. 5.3 przedstawiono charakterystykę widmową oraz fazową funkcji transmitancji układu akustycznego przedstawionego na Rys. 5.2. Warto zwrócić uwagę, że właśnie dla częstotliwości formantowych następuje skokowa zmiana znaku charakterystyki fazowej (z ujemnej na dodatnią). Ta własność charakterystyki fazowej może być wykorzystana do dokładnego wyznaczenia częstotliwości formantowych w układach za pomocą których modeluje się tor głosowy. Niestety tej możliwości nie ma przy analizie rzeczywistych sygnałów mowy, i częstotliwości formantowe są wyznaczone z widma amplitudowego, co może być obarczone dość znacznym błędem, zwłaszcza w przypadku nazalizacji, czy głosu o stosunkowo dużej częstotliwości podstawowej  $F_0$ .





Rys. 5.2 Układ akustyczny składający się z dwóch odcinków cylindrycznych, którego częstotliwościowa charakterystyka transmitancji jest zbliżona do charakterystyki toru głosowego podczas wymawiania samogłoski /a/.



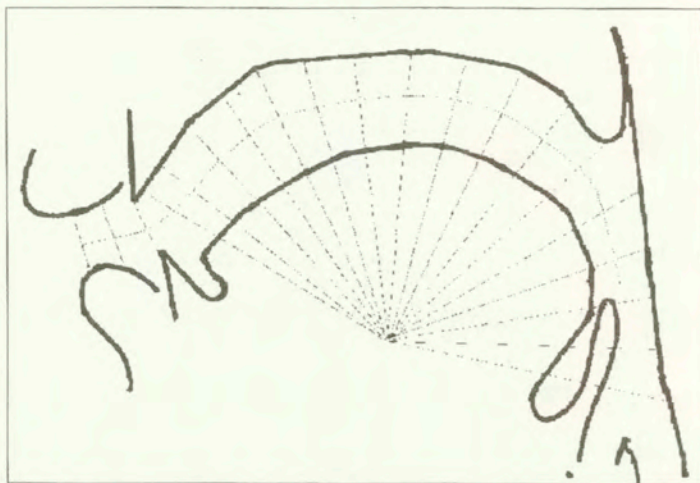
Rys. 5.3 Charakterystyka widmowa (u dołu) oraz fazowa (u góry) układu akustycznego składającego się z dwóch odcinków cylindrycznych (por. Rys. 5.2). Częstotliwości maksimum w widmie są zbliżone do częstotliwości formantowych samogłoski /a/. Widmo harmoniczne wyznaczono dla częstotliwości podstawowej  $F_0 = 125$  Hz, za pomocą programu Model (patrz rozdz.8).

Przedstawiony na Rys. 5.2 i 5.3 przypadek jest bardzo znamienny, bowiem wskazuje, że nawet dla modelu o konfiguracji znacznie odmiennej od rzeczywistej konfiguracji geometrycznej toru głosowego, otrzymana charakterystyka widmowa może być w dużym stopniu zbliżona do pomierzonej dla rzeczywistej głoski /a/. Świadczy to o tym, że nie zawsze istnieje jednoznaczna zależność między

charakterystyką widmową sygnału mowy a konfiguracją układu artykulacyjnego (zachodzi ona jedynie w odwrotną stronę) [FANT,1960].

## 5.2 Modelowanie artykulacji samogłoskowej.

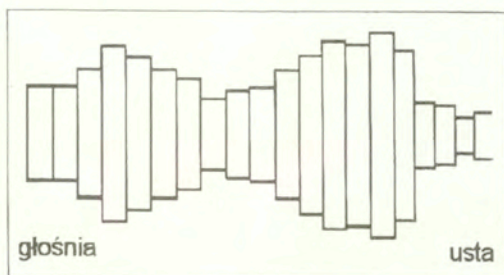
Już w latach 50-ych Fant (FANT,1960) porównując konfiguracje toru gardłowo-ustnego przy wymawianiu poszczególnych samogłosek i analizując ich charakterystyki widmowe wykazał, że w miarę dobre odwzorowanie zarówno od strony artykulacyjnej, jak i akustycznej uzyskuje się za pomocą modelu składającego się z kaskadowo połączonych ze sobą 16-20 odcinków rur cylindrycznych o długości ~1 cm. Na Rys. 5.4 przedstawiono przykładowy przekrój wzdłużny toru głosowego, wyznaczony w oparciu o jego zdjęcie kinorentgenograficzne [BOLLA&FÖLDI;1987], wykonane podczas artykulacji samogłoski /a/ w wyrazie /tak/. Na rysunku zaznaczono linie, wzdłuż których wyznaczane są kolejne przekroje poprzeczne (20). Linie te są w przybliżeniu prostopadłe do osi symetrii toru głosowego zaznaczonej na rysunku linią kropkowaną.



Rys. 5.4 Sposób wyznaczania przekrojów poprzecznych toru głosowego w oparciu o jego przekrój wzdłużny wyznaczony na podstawie zdjęcia kinorentgenograficznego.

Pomierzone odległości między powierzchnią języka i ścianami podniebienia, a także między ścianami gardła, są w pierwszym przybliżeniu traktowane jako średnice odpowiednich odcinków cylindrycznych, za pomocą których modeluje się geometrię

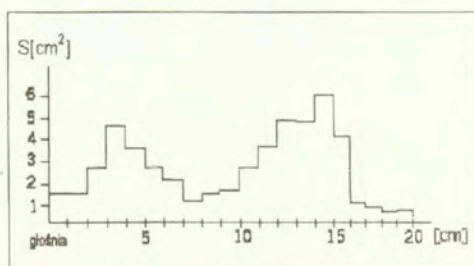
toru głosowego. Na Rys. 5.5 przedstawiono przykładowy wynik takiego modelowania dla samogłoski /u/.



Rys. 5.5 Cylindryczny model toru głosowego dla samogłoski /u/.

Konfigurację toru głosowego przy artykulacji samogłoski /u/ odwzorowano za pomocą 20 cylindrycznych odcinków o długości 1 cm oraz o średnicy zbliżonej do średnicy odwzorowywanych przekrojów poprzecznych. Warto zwrócić uwagę, że wskutek wysunięcia warg ku przodowi przy artykulacji samogłoski /u/ długość toru artykulacyjnego jest większa, niż na przykład przy artykulacji samogłoski /a/, dla której długość tę przyjmuje się równą 17 cm. Przedstawiony sposób wyznaczania przekrojów poprzecznych toru głosowego w oparciu o wykonane zdjęcia, jest podobny do metody przyjętej w pracy [LADEFOGED&BLADON,1982]. Jak dalej przekonamy się, funkcja transmitancji toru głosowego zależy przede wszystkim od kształtu funkcji przekroju (dokładniej od funkcji przekrojów poprzecznych), której przykład przedstawiono na Rys 5.6 dla samogłoski /u/. Została ona wyliczona na podstawie danych w [BOLLA &FÖLDI,1987].

Przedstawiony sposób modelowania toru głosowego za pomocą skończonej liczby odcinków cylindrycznych, daje wyniki w dziedzinie częstotliwości nie zawsze zbieżne z wynikami pomiarowymi. Przyczyną tego jest to, że wprawdzie funkcja kształtu toru głosowego stanowiąca podstawę opisu fonetycznego dźwięków mowy jest w miarę dobrze odwzorowana, to jednak funkcja transmitancji toru głosowego jest zależna, jak to zostanie dalej wykazane, od funkcji powierzchni przekrojów poprzecznych toru głosowego, a nie od jego funkcji kształtu przekroju wzdłużnego. Tym nie mniej w opisie fonetycznym głosek właśnie ta druga funkcja stanowi jego podstawę. Dlatego przy modelowaniu pożądanym jest odwzorowanie obu funkcji jednocześnie, co umożliwi ocenę spójności opisu fonetycznego z charakterystyką częstotliwościową dźwięków mowy.



**Rys. 5.6** Funkcja przekroju toru głosowego dla cylindrycznego modelu samogłoski /u/ przedstawionego na Rys. 5.5.  $S$  – powierzchnia przekroju poprzecznego w funkcji odległości jego od głośni, wyliczona na podstawie danych w [BOLLA&FÖLDI,1987]

Najobszerniejsze dane artykulacyjne dla mowy polskiej zostały zebrane przez zespół Laboratorium Fonetyki Instytutu Lingwistyki Węgierskiej Akademii Nauk i Katedry Fonetyki Wydziału Sztuk Uniwersytetu im. Eötvös Loránd w Budapeszcie [BOLLA&FÖLDI,1987]. W publikacji tej podane są przekroje wzdłużne toru głosowego wykreślone w oparciu o zdjęcia kinorentgenograficzne, wykonane podczas artykulacji głosek polskich wymówionych w różnych kontekstach, w obrębie izolowanych wyrazów. Załączono wymiary przekroju wzdłużnego toru głosowego wyznaczone w określonych jego punktach i dla różnych faz artykulacji (w zasadzie dla 5-ciu) opisywanej głoski. Ponadto, podano układ warg, zdjęcia podniebienia i języka po artykulacji obrazujące strefy ich kontaktu oraz spektrogramy z przekrojami widmowymi wraz z przebiegami czasowymi (te ostatnie dosyć niżej jakości). Jednocześnie zespół prof. K. Bolli udostępnił Pracowni Akustyki Mowy IPPT PAN zapisy kinorentgenograficzne na taśmie video, co umożliwiło w wątpliwych przypadkach powtórzenie niektórych analiz.

Jednym z powszechnie stosowanych przy modelowaniu zabiegów jest odtworzenie tylko w przybliżeniu układu masy języka w torze głosowym przy możliwie jak najdokładniejszym odwzorowaniu częstotliwościowej funkcji przenoszenia, co zapewnia uzyskanie również możliwie dobrej zrozumiałości dźwięku generowanego przez model. Jednakże skorygowana w ten sposób konfiguracja geometryczna modelu toru głosowego ma kształt niekiedy dość znacznie odbiegający od rzeczywistego, odpowiadającego modelowanej głosce. Jest to spowodowane tym, że konfiguracja geometryczna toru głosowego w wielu przypadkach, w istotny sposób odbiega od kształtu cylindrycznego (por. wyniki uzyskane techniką MRI np.[BAER i in., 1991]). W pracy podjęto próbę odwzorowania w modelu pomiarów wykonanych ze zdjęć kinorentgenograficznych, przy jednoczesnym uzyskaniu rozkładu częstotliwości formantowych zgodnego z pomierzonym w sygnale mowy zarejestrowanym dla lektora. W tym celu podjęto oryginalną próbę stworzenia modelu o niejednorodnym

kształcie, w którym różne części toru głosowego są odpowiednio modelowane odcinkami o przekrojach poprzecznych kołowym (część gardłowa), elipsoidalnym (tylna część jamy ustnej i otwór uformowany przez wargi) oraz trapezowym (część ustna znajdująca się pod podniebieniem twardym). W celu głębszego przedstawienia zasad modelowania toru głosowego zostaną omówione najpierw podstawy fizyczne jego modelowania jako złożonego układu akustycznego, modelowania źródeł pobudzających, zaś następnie podane będą wyniki modelowania artykulacji naturalnej w oparciu o opracowany w pracy niejednorodny model toru głosowego.

### 5.3 Podstawy fizyczne modelowania toru głosowego jako układu akustycznego

W cytowanej fundamentalnej pracy poświęconej badaniu i modelowaniu własności akustycznych układu artykulacyjnego, Fant [FANT,1960] porównując ze sobą konfiguracje toru gardłowo-ustnego przy wymawianiu poszczególnych samogłosek i analizując ich charakterystyki widmowe wykazał, że w miarę dobre, choć niekiedy mocno przybliżone odwzorowanie, zarówno od strony geometrii toru, jak i ich cech akustycznych, uzyskuje się za pomocą bardzo uproszczonego modelu składającego się z kaskadowo połączonych ze sobą czterech odcinków rur cylindrycznych o odpowiednio dobranych powierzchniach przekrojów poprzecznych do kierunku rozchodzenia się fali akustycznej, biegnącej od głośni do ust (taki model przedstawiono na Rys. 5.1). Model, którego łączna długość jest równa odległości otworu ustnego od głośni, odzwierciedla w przybliżeniu charakterystyki artykulacyjne dźwięków samogłoskowych (a przede wszystkim ułożenie masy języka), a także pozwala na odwzorowanie wpływu stopnia zaokrąglenia warg na charakterystykę widmową modelowanego dźwięku.

Przy założeniu, że w torze głosowym rozchodzi się wzdłuż jego osi podłużnej fala płaska, najbardziej właściwym od strony analitycznej jest modelowanie toru artykulacyjnego za pomocą układu transmisyjnego o stałych rozłożonych. Umożliwia on dokładniejsze niż model o stałych skupionych odwzorowanie zmieniającej się funkcji przekroju poprzecznego toru głosowego. W modelu tym można także uwzględnić wpływ znacznej części parametrów fizycznych oraz akustycznych toru artykulacyjnego związanych bezpośrednio z generacją sygnału mowy.

W modelu toru głosowego o stałych rozłożonych można zastosować aproksymację (znaną z teorii linii długich) za pomocą czwórników, z których każdy określa dla elementarnego odcinka toru zależności między parametrami wejściowymi i wyjściowymi fali akustycznej (w tym przypadku, są to ciśnienia i prędkości akustyczne) [KACPROWSKI,1962]. Warunek fali płaskiej rozchodzącej się wzdłuż osi podłużnej toru głosowego jest spełniony tylko dla tych częstotliwości, dla których największy wymiar poprzeczny toru jest mniejszy od  $1/4$  długości fali  $\lambda$ . Stąd więc, na przykład dla samogłoski /a/, dla której maksymalny wymiar poprzeczny toru jest rzędu

2-3 cm, wynika że, przy prędkości rozchodzenia się fali równej 352 m/s, aproksymacja jest poprawna, gdy maksymalna częstotliwość fali nie jest większa niż

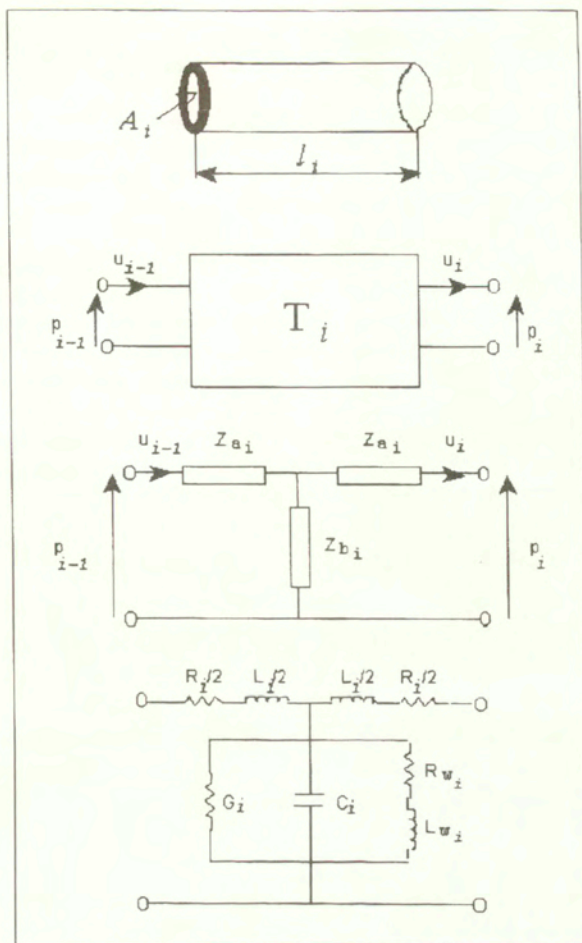
$$F_{\max} = c/4d_p = 3000 \sim 4400 \text{ Hz.}$$

Z przyjęcia dla modelu częstotliwości maksymalnej 4500 Hz, wynika ograniczenie na długość odcinka rury cylindrycznej. By dobrze odwzorowywał wpływ lokalnych zmian konfiguracji toru głosowego na charakterystykę częstotliwościową modelu, winien mieć długość  $l_1$  nie większą od 1/8 długości fali odpowiadającej częstotliwości  $F_{\max}$ , a więc

$$l_1 = 0.125 \sim 0.25 l_{4,5\text{kHz}} \gg 0.9 \text{ cm.}$$

Tak więc, aby uzyskać w miarę dokładne odwzorowanie kształtu toru głosowego względem fali akustycznej w nim rozchodzącej się, wystarczy aproksymować odcinkami cylindrycznymi o długości nie większej niż 1 cm.

Zgodnie z teorią układów akustycznych ([MORSE&INGARDEN,1968]) parametry elektryczne równoważnego czwórnika są wyznaczone przez fizyczne własności odcinka rury o długości  $l_1$  wypełnionej słupem powietrza. Uwzględnia się przy tym nie tylko własności ośrodka, w którym rozchodzi się fala akustyczna, lecz również straty wywołane drganiami ścianek toru głosowego. Poszczególne elementy czwórnika zastępczego mają ściśle określone znaczenie fizyczne, i dla przyjętego układu analogii akusto-elektrycznych, w którym indukcyjność  $L$  reprezentuje bezwładność (tzw. masę akustyczną) drgającego słupa powietrza,  $C$  - jego podatność,  $R$  - oporność akustyczną ośrodka (reprezentuje ona straty wywołane lepkością powietrza),  $G$  - przewodność akustyczną (reprezentuje ona straty wywołane przewodnością cieplną powietrza), schemat zastępczy jednorodnego odcinka rury cylindrycznej można przedstawić w postaci czwórnika typu T (może być również typu  $\Pi$ ), tak jak to przedstawiono na Rys. 5.7.



Rys. 5.7 Czwórnikowy model elementarnego odcinka toru głosowego.

Wielkości elektryczne podane obok odpowiednich elementów czwórnik równoważnego mają określone znaczenia fizyczne i są klasycznymi parametrami rury stratnej o długości  $l$ . Uwzględniają one zarówno własności akustyczne masy powietrza znajdującego się w danym odcinku rury, jak i jej ścianek, które również są pobudzone do drgań wymuszonych pod wpływem oddziaływania lokalnego ciśnienia akustycznego. Tak więc:

$L$  - reprezentuje bezwładność (masę akustyczną) powietrza zawartego w cylindrycznym odcinku toru,  
 $C$  - podatność akustyczną powietrza w odcinku toru,  
 $R$  - rezystancja akustyczna strat na tarcie wiskotyczne (lepkościowe),  
 $G$  - akustyczna konduktancja strat na przewodnictwo cieplne w powietrzu,  
 $R_w$  - rezystancja akustyczna współdrgających ścian odcinka cylindrycznego,  
 $L_w$  - masa akustyczna ścian odcinka cylindrycznego.

Na wejściu i na wyjściu elementarnego czwornika, liczone są ciśnienie akustyczne ( $p_i$ ) i prędkość objętościową fali akustycznej ( $U_i$ ) rozchodzącej się w odcinku modelowanego toru głosowego. Parametry te w przyjętym układzie analogii akusto-elektrycznych są reprezentowane odpowiednio przez napięcie na zaciskach czwornika i natężenie prądu wpływającego i wypływającego z czwornika.

Zależności między parametrami wejściowymi ( $p_{i-1}, U_{i-1}$ ) i wyjściowymi ( $p_i, U_i$ ) każdego elementarnego czwornika  $T$  w układzie analogii akusto-elektrycznych mogą być przedstawione w następującej postaci [ATAL i in., 1978]:

$$p_i = p_{i-1} \cosh(\gamma_i) + U_{i-1} Z_i \sinh(\gamma_i), \quad 5.1$$

$$U_i = \frac{p_{i-1}}{Z_i} \sinh(\gamma_i) + U_{i-1} \cosh(\gamma_i), \quad 5.2$$

gdzie  $p_i$  oraz  $p_{i-1}$  są odpowiednio ciśnieniami na wlocie i wylocie modelowanego segmentu, zaś  $U_i$  i  $U_{i-1}$  odpowiednimi prędkościami objętościowymi, zaś  $\gamma_i$  jest współczynnikiem propagacji określonym wzorem

$$\begin{aligned}
 \gamma_i &= \alpha_i + j\beta_i = \sqrt{Z_a Z_b} \\
 &= \sqrt{(R_i + j\omega L_i)(G_i + j\omega C_i + Y_w)}
 \end{aligned} \quad 5.3$$

W przypadku rury o sztywnych ścianach straty są reprezentowane przez współczynnik tłumienia  $\alpha$  określonego w postaci sumy

$$\alpha = \alpha_R + \alpha_G = \frac{R}{2 \cdot c \cdot L} + \frac{G}{2 \cdot c \cdot C} \quad 5.4$$

Powyższe zależności mogą być uporządkowane w postaci równania macierzowego:



$$\begin{vmatrix} p_i \\ U_i \end{vmatrix} = \begin{vmatrix} \cosh(\gamma_i) & Z_i \sinh(\gamma_i) \\ \sinh(\gamma_i)/Z_i & \cosh(\gamma_i) \end{vmatrix} \begin{vmatrix} p_{i-1} \\ U_{-1} \end{vmatrix}, \quad 5.5$$

gdzie  $Z_i$  jest impedancją charakterystyczną czwornika

$$Z_i = \sqrt{\frac{Z_{a_i}}{Z_{b_i}}}$$

Poszczególne elementy czwornika są określone przez poniższe wzory:

$$\begin{aligned} R_i &= \frac{S_i}{A_i^2} \sqrt{\frac{\omega \delta \mu}{2}}, \\ L_i &= \frac{\delta}{A_i}, \\ G_i &= \frac{S_i(\eta-1)}{\rho \cdot c^2} \sqrt{\frac{\lambda \cdot \omega}{2 \cdot c_p \rho}}, \\ C_i &= \frac{A_i}{\rho \cdot c^2}. \end{aligned} \quad 5.6$$

gdzie,

$A_i$  - powierzchnia przekroju poprzecznego i-tego odcinka rury,

$S_i$  - obwód przekroju i-tego odcinka rury,

zaś  $Y_m$  które jest admitancją reprezentującą straty powstałe wskutek drgań ścianek odcinka rury jest określone następująco

$$\begin{aligned} Y_m &= \frac{l}{R_m + j\omega L_m}, \text{ gdzie} \\ R_m &= \frac{1600}{S_i}, \\ L_m &= \frac{l \cdot 4}{S_i}, \end{aligned}$$

Przyjmuje się następujące wartości parametrów fizycznych ośrodka wypełniającego tor głosowy [KACPROWSKI,1977b]:

$c = 352 \text{ m/s}$  - prędkość rozchodzenia się fali akustycznej w torze głosowym,  
 $\rho = 1.14 \text{ kg} \times \text{m}^{-3}$  - gęstość ośrodka,  
 $C_p = 240 \text{ cal} \times \text{kg}^{-1} \times \text{deg}^{-1}$  - ciepło właściwe ośrodka przy stałym ciśnieniu  
 $\eta = 1.4$  - stała adiabatyczna ośrodka,  
 $\mu = 1.84 \times 10^{-5} \times \text{N} \times \text{s} \times \text{m}^{-2}$  - dynamiczny współczynnik lepkości ośrodka,  
 $\lambda = 5.5 \times 10^{-3} \times \text{cal} \times \text{m}^{-1} \times \text{deg}^{-1}$  - współczynnik przewodnictwa cieplnego.

Z przytoczonych wyżej wzorów (5.6) widać, że podczas artykulacji zmiennymi w czasie mogą być tylko te elementy czwornika, które są zależne od geometrii odwzorowywanego odcinka toru głosowego, a ściślej biorąc od wielkości powierzchni przekroju poprzecznego  $A_i$  oraz powierzchni zewnętrznej (obwodowej) odcinka  $S_i$  wyznaczającej wielkość strat, zarówno w słupie powietrza, jak i na wskutek drgań ścianek toru. Przy pominięciu strat (co można w pierwszym przybliżeniu uczynić), transmitancja segmentu jest wyznaczona przez wielkość powierzchni jego przekroju poprzecznego i jest nieliniową funkcją częstotliwości.

We wzorach na masę akustyczną ( $L$ ) i podatność akustyczną ( $C$ ) pominięto oddziaływanie między słuzówką jamy gardłowo-ustnej i rozchodzącą się wewnątrz falą akustyczną. Wskutek tarcia lepkościowego powietrza o ściany słuzówki wielkość skutecznej powierzchni  $S_i$  toru jest mniejsza, niż jej wartość nominalna wynikająca z jego geometrii. Natomiast straty wywołane przewodnictwem cieplnym słuzówki powodują, że powierzchnia ta uwzględniana przy obliczaniu podatności akustycznej ( $C$ ) ścianek toru jest większa. Tak więc zarówno wielkość  $L$ , jak i  $C$  powinny być większe, niż to wynika z podanych wyżej wzorów, jednakże poprawka ta zależy od przekroju powierzchni i częstotliwości, i na ogół jest ona pomijalnie mała. W przypadku, gdy  $A_i f > 74 \text{ cm}^2/\text{s}$  przyrost  $L$  jest mniejszy od 10% wartości nominalnej, zaś przyrost  $C$  mniejszy o 5% od wartości obliczonej ze wzoru [BENADE, 1968]. Tak więc wpływ ścianek słuzówki jest w większości przypadków do pominięcia.

Ze wzorów (5.5) i (5.6) wynika, że

$$\alpha_r = \frac{1}{2} \frac{S_i}{\rho c A_i} \sqrt{\mu \omega \frac{\rho}{2}} \quad 5.7$$

oraz

$$\alpha_G = \frac{S_i (\eta - 1)}{2 c A_i} \sqrt{\lambda c_p \rho \frac{\omega}{2}} \quad 5.8$$

Dla bezstratnego toru ( $R=G=0$ ) współczynnik tłumienia  $\alpha$  jest równy zeru. Wówczas współczynnik propagacji  $\gamma$  redukuje się do części urojonej  $j\beta$ , gdzie  $\beta=\omega/c$ . Natomiast dla pomijalnie małej wielkości strat, t.j.  $R \ll \omega L$  oraz  $G \ll \omega C$ , otrzymuje się

$$\beta = \frac{\omega}{c} \left( 1 - 0.5 \frac{R}{\omega L} \frac{G}{\omega C} \right), \quad 5.9$$

co również może być aproksymowane do wzoru  $\beta=\omega/c$ .

Podobnie jak  $\gamma_i$ , impedancja charakterystyczna czwornika elementarnego  $Z_i$  jest również wielkością zespoloną

$$Z_i = \frac{\rho C}{A_i} \left[ 1 - j(\alpha_R - \alpha_G) \frac{c}{\omega} \right], \quad 5.10$$

w której część urojona zanika, gdy straty wskutek tarcia i przewodnictwa cieplnego są sobie równe. Jednakże w rozpatrywanym układzie akustycznym nie ma to miejsca.

Z powyższych wzorów można wyprowadzić, że

$$\frac{\alpha_R}{\alpha_G} = (\eta - 1) \sqrt{\mu \frac{\lambda}{c_p}} = 0.444, \quad 5.11$$

czyli stosunek ten jest wielkością stałą, niezależną od geometrii toru głosowego.

Tak więc impedancja charakterystyczna  $i$ -tego segmentu może być aproksymowana z dość dobrym przybliżeniem do  $Z_i = \rho c/A_i$ , bowiem współczynniki  $\alpha_R$  i  $\alpha_G$  są zwykle pomijalnie małe.

Warto podkreślić, że przy wyprowadzaniu wzorów na elementy czwornika równoważnego odcinkowi rury cylindrycznej nie czyniono żadnych założeń co do długości tego odcinka. Aproksymacja toru głosowego była realizowana za pomocą zintegrowanych elementów linii długiej i była wyrażona za pomocą funkcji hiperbolicznych. Dopiero przy dokonywaniu dalszych uproszczeń wzorów przyjmowano założenie, że odwzorowywany odcinek jest na tyle krótki, że jest spełniony warunek  $\gamma l_1 \ll 1$ . Dopiero wówczas funkcje hiperboliczne mogą być uproszczone do pierwszych wyrażen ich rozwinięć szeregowych i impedancje  $Z_a$  i  $Z_b$  czwornika mogą być sprowadzone do następujących postaci:

$$Z_a = Z_i \tanh\left(\frac{\gamma_i l}{2}\right) = \frac{(R + j\omega L) l_i}{2}, \quad 5.12$$

$$Z_b = \frac{Z_i}{\sinh(\gamma_i l)} = \frac{l}{(G + j\omega C) l_i}, \quad 5.13$$

Całkowita funkcja przenoszenia toru głosowego może być wyznaczona poprzez konkatencję elementarnych czwórników, a dokładnie przez przemnożenie funkcji przenoszenia kolejnych odcinków toru głosowego o postaci

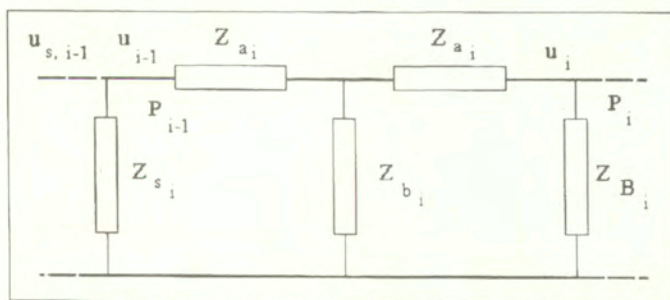
$$H_i(\omega) = \frac{U_i}{U_{i-1}}, \quad 5.14$$

W celu uniknięcia przy wyznaczaniu całkowitej funkcji transmitancji toru złożonego zapisu macierzowego, można wyznaczać w sposób "krokowy" impedancję widzianą na kolejnych zaciskach wejściowych czwórników widzianych z lewej strony, wyznaczanych poprzez kolejne ich prawostronne "zwijanie". Na Rys. 5.8 przedstawiono (i-1)-ty czwórnik obciążony prawostronnie impedancją  $Z_{B,i} = p_i/U_i$ . Wówczas stosunek prędkości objętościowych na wyjściu i na wejściu czwórnika elementarnego jest określony wzorem:

$$\frac{U_i}{U_{i-1}} = \cosh(\gamma_i l) + \frac{Z_{B,i-1} \sinh(\gamma_i l)}{Z_i}. \quad 5.15$$

Dokładnie taką samą zależność otrzyma się rozwiązując równanie macierzowe (5.5). Dwa specyficzne przypadki występują przy tak przyjętym zwijaniu i rozwijaniu macierzy łańcuchowej toru głosowego. Jeden z nich ma miejsce, gdy impedancja  $Z_{B,i-1} = 0$ , co odpowiada akustycznie promieniowaniu fali do otwartej przestrzeni. Wówczas impedancja promieniowania zamykająca najbardziej prawostronnie położony odcinek toru głosowego może być pominięta. Stąd,

$$Z_{B,i} = Z_{a,i} + \frac{Z_{a,i} Z_{B,i}}{Z_{a,i} + Z_{B,i}} = Z_i \tanh(\gamma_i l). \quad 5.16$$



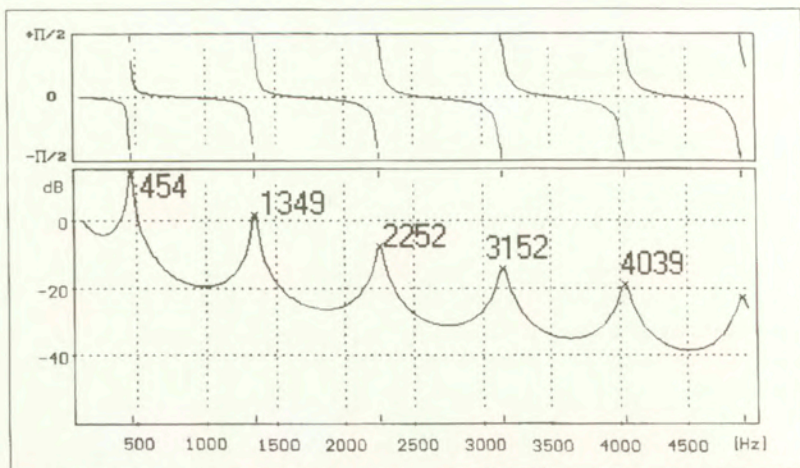
**Rys. 5.8** Czwórnik obciążony prawostronnie impedancją  $Z_B$  widzianą z jego zacisków wyjściowych i reprezentującą impedancję odcinka toru głosowego od bieżącego punktu do wylotu z ust.

Drugi przypadek, zachodzi gdy  $Z_{B,i-1} = A_b$ , to jest, gdy tor głosowy jest zamknięty sztywną ścianą. W przybliżeniu odpowiada to przypadkowi artykulacji, na przykład spółgłoski nosowej, gdy usta są zamknięte. Wówczas impedancja wejściowa  $i$ -tego czwornika  $Z_{B,i}$  jest określona wzorem

$$Z_{B,i} = Z_{a,i} + Z_{b,i} = Z_i \coth(\gamma_i). \quad 5.17$$

Impedancja  $Z_{s,i}$  jest dodatkową impedancją widzianą z zacisków wejściowych bieżącego odcinka od strony źródła krtaniowego.

Tak więc w przypadku, gdy tor głosowy przedstawiony jest w postaci jednorodnej rury stratnej, zamkniętej z jednej strony (od strony głośni) i otwartej na drugim końcu, to jej funkcja transmitancji będzie odbiegać nieco od funkcji wyznaczonej w układzie o stałych skupionych odwzorowujących bezstratny falowy rezonator Helmholtza. W przypadku rury bezstratnej kolejne maksima w jej funkcji transmitancji znajdą się dokładnie w punktach o częstotliwościach odpowiadających nieparzystym wielokrotnościom  $1/4$  długości rury. Jeżeli przyjmiemy jej długość równą 17,6 cm, to kolejne maksima funkcji transmitancji wypadną dla częstotliwości 500 Hz, 1500 Hz, 2500 Hz,.... Przy uwzględnione straty, tak jak to jest uwzględnione w układzie o stałych rozłożonych, wówczas częstotliwości maksimów ulegają obniżeniu, tak jak to pokazano na Rys. 5.9. Wielkość obniżenia zależy od wielkości przekroju poprzecznego rury.

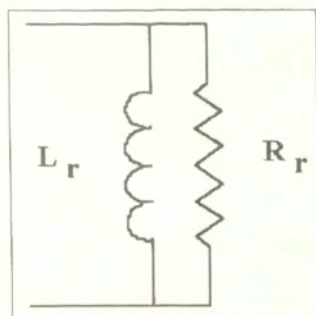


Rys. 5.9 Funkcja transmitancji (amplitudowa i fazowa) stratnej rury cylindrycznej zamkniętej na jednym końcu. Częstotliwości formantów w stosunku do funkcji rury bezstratnej uległy obniżeniu o ok. 10%. Charakterystyka wyliczona za pomocą programu *Model*.

Do pełnego przedstawienia i wyliczenia funkcji transmitancji toru głosowego, należy uwzględnić, że w przypadku artykulacji głosek dźwięcznych, powietrze w nim znajdujące się jest pobudzone do drgań przez źródło prądowe o skończonej impedancji, zaś na wyjściu tor głosowy jest zamknięty impedancją promieniowania ust. Najbardziej dokładny jej model dla głosek ustnych ma postać drgającego tłoka umieszczonego w sferycznej odgradzie. W bardziej uproszczonym modelu, odgroda jest nieskończenie płaska i impedancja promieniowania ust może być wyliczona w przybliżony sposób ze wzoru [FANT,1960]

$$Z_r \approx \frac{\rho c}{A_r} \left[ \frac{(\beta r)^2}{2} + j \frac{8}{3\pi} (\beta r) \right], \quad 5.18$$

Jednakże ten sposób modelowania funkcji promieniowania ust jest nadal dość złożony i jak wykazały badania eksperymentalne [FANT,1960], do pełnego zamodelowania toru artykulacyjnego wystarczają uproszczone modele funkcji promieniowania ust. Jeden z nich, został przedstawiony w układzie analogii akustoelektrycznych na Rys. 5.10.



Rys.5.10 Model impedancji promieniowania ust w układzie analogii akusto-elektrycznych [FANT,1960].

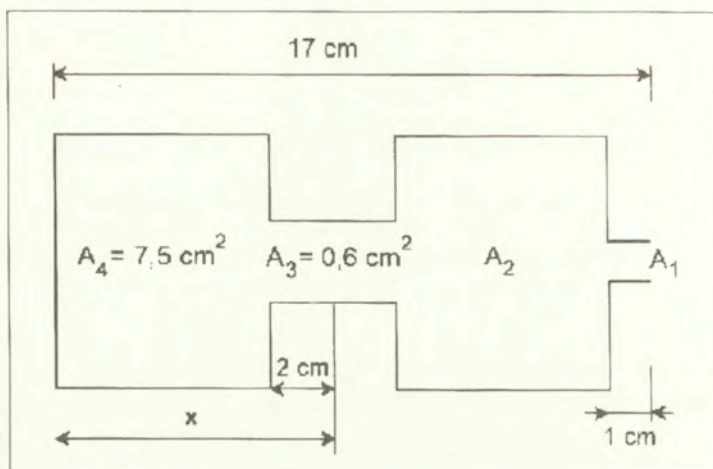
Impedancja promieniowania ust jest określona następującym wzorem uwzględniającym parametry akustyczne toru głosowego [WAKITA&FANT, 1978]:

$$Z_r = \frac{K_s \rho \omega^2}{4\pi c} + \frac{j\omega 8\rho}{3\pi\sqrt{\pi} A_0} = R_r + j\omega L_r, \quad 5.19$$

$$K_s = \begin{cases} 1 + (0.6/1600)f & 0 \leq f \leq 1600 \text{ Hz} \\ 1.6 & f \geq 1600 \text{ Hz} \end{cases} \quad 5.20$$

Przedstawiony na początku rozdziału, uproszczony, 4-ro segmentowy model toru głosowego (Rys. 5.1) może posłużyć do zbadania zasadniczych relacji jakie istnieją między rozmiarami i położeniem przewężenia, a także wpływem zaokrąglenia warg. Jak już wcześniej wspomniano, tego typu model, w którym dwie komory, przednia i tylna są połączone ze sobą silnie zwężonym odcinkiem jest stosunkowo bliski konfiguracji geometrycznej toru głosowego powstającej podczas wymawiania samogłoski /u/, czy /i/. Przednia część toru głosowego jest zakończona krótkim odcinkiem o stosunkowo małej powierzchni przekroju, aproksymującą część wylotową uformowaną przez wargi. Na Rys.5.11 [FANT,1960] przedstawiono ten model wraz z podanymi wymiarami, które są zbliżone w pewnym stopniu do rzeczywistych. Dla tak zamodelowanej konfiguracji toru głosowego wykonano obliczenia częstotliwości formantowych w funkcji położeniu przewężenia, dla różnych stopni zaokrąglenia warg, a także zmiennych wymiarów przewężenia. W obliczeniach uwzględniono funkcje strat, oporność źródła krztaniowego oraz oporność promieniowania ust. Przesuwanie

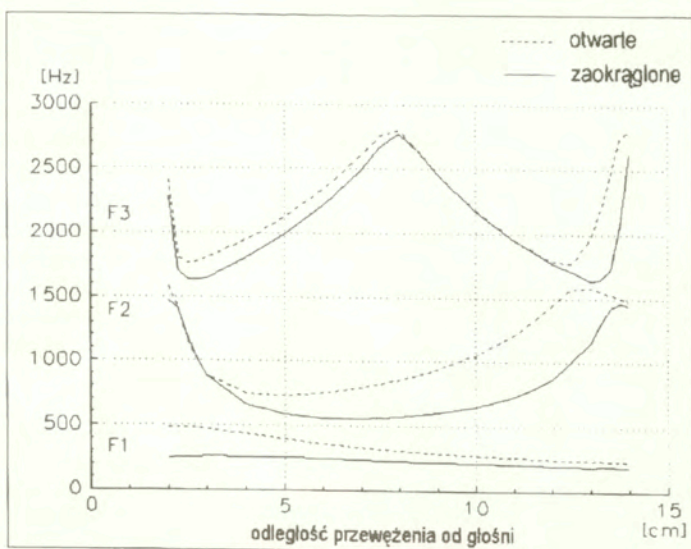
przewężenia dokonywano w taki sposób, by całkowita objętość modelu toru głosowego pozostawała niezmienna, dzięki czemu bezwzględna wielkość strat pozostawała stała.



**Rys. 5.11** Geometria uproszczonego modelu toru głosowego przyjęta w obliczeniach funkcji transmitancji dla różnych położeń przewężenia i jego wymiarów.  $A_1$  – powierzchnia przekroju otworu wylotowego ust,  $A_2$  – powierzchnia przekroju przedniej części toru gardłowo-ustnego,  $A_3$  – powierzchnia przekroju szczeliny,  $A_4$  – powierzchnia przekroju tylnej części toru gardłowo-ustnego. Wszystkie powierzchnie są powierzchniami przekrojów poprzecznych do kierunku rozchodzenia się fali akustycznej [FANT,1960].

Pomimo tak dalece uproszczonej aproksymacji geometrii toru głosowego, można zaobserwować pewne podstawowe zależności między jego modelowanym kształtem i odpowiadającymi mu charakterystykami widmowymi generowanego sygnału wyjściowego. Na Rys. 5.12 przedstawiono zależność częstotliwości formantowych od odległości od głośni utworzonego w torze głosowym przewężenia. Na rysunku tym widać wyraźną zależność częstotliwości drugiego formantu od miejsca położenia przewężenia. Ogólny trend jest taki, że im przewężenie jest bardziej przesunięte ku przodowi, tym wyższa jest częstotliwość drugiego formantu, co jest zgodne z obserwowaną tendencją w naturalnych dźwiękach mowy. Ponadto, widoczny jest wpływ wielkości przekroju otworu wylotowego na częstotliwości maksimum funkcji transmitancji toru.



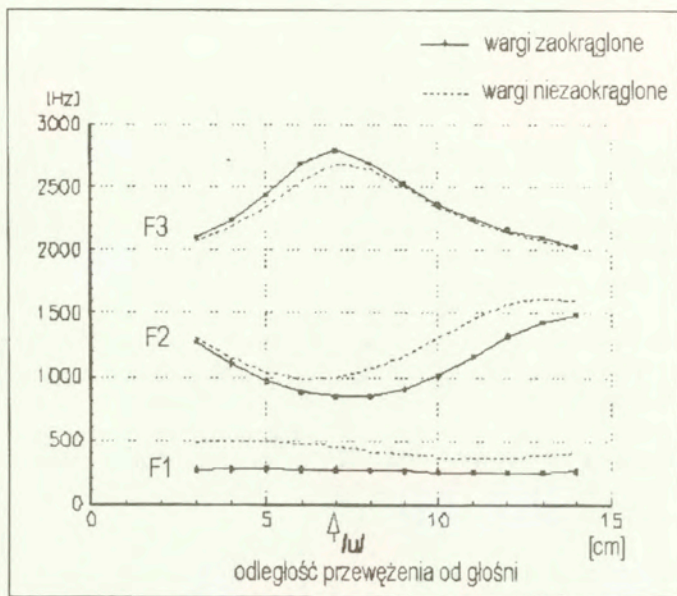


Rys. 5.12 Przebieg zmian częstotliwości formantowych w zależności od położenia środka przewężenia w 4-ro odcinkowym modelu cylindrycznym (z uwzględnieniem strat) oraz wielkości otworu wylotowego (dla powierzchni  $A_1=0.25 \text{ cm}^2$  zbliżonej do wartości rzeczywistej dla /u/ oraz  $A_1=2.6 \text{ cm}^2$  podobnej jak dla /e/). Przebiegi wyliczone przy zastosowaniu programu *Model*.

Ogólnie biorąc, zwiększenie zaokrąglenia ust powodujące zmniejszenie powierzchni przekroju otworu wylotowego, w wyniku czego następuje obniżenie częstotliwości formantowych. Przykładowo, dla częstotliwości drugiego formantu obniżenie to może wynieść ok. 30-40%, przy nie uwzględnieniu przy tym faktu, że zaokrąglanie głosek w języku polskim powoduje jednocześnie wydłużenie toru głosowego (do 18-19 cm). Zaś zwiększenie długości toru głosowego powoduje dodatkowe obniżenie częstotliwości formantowych.

W rzeczywistym torze głosowym nie jest możliwe osiągnięcie tak dużego zakresu zmian położenia przewężenia bez jednoczesnej zmiany kształtu części toru położonych przed i za miejscem artykulacji, zwłaszcza dla przypadków odpowiadających skrajnym położeniom przewężenia, w bezpośrednim sąsiedztwie głośni, lub warg. Ograniczenie to dotyczy zwłaszcza przewężeń o bardzo małej powierzchni przekroju poprzecznego. Choć więc teoretycznie zmiany położenia środka przewężenia w torze głosowym względem głośni mogą w rozważanym modelu zachodzić w zakresie od 2 do 14 cm, to w praktyce, z powodów anatomicznych, przyjmowana minimalna odległość środka przewężenia od głośni nie może być mniejsza od 4 cm.

Na Rys.5.13 przedstawiono analogiczny do Rys. 5.12 przebieg zmian częstotliwości formantowych dla 20-odcinkowego modelu, w którym odtwarzana konfiguracja toru głosowego jest bardziej zbliżona do rzeczywistej konfiguracji samogłoski /u/, niż w przypadku modelu 4-ro segmentowego. Na wykresie zaznaczono położenie środka przewężenia odpowiadające konfiguracji dla samogłoski /u/ (por. Rys. 5.5 i 5.6). Przesuwanie przewężenia odbywa się (w lewo lub w prawo) w sposób symetryczny, przy sekwencyjnym przesuwaniu segmentów przeciwnie do ruchu przewężenia, tak aby sumaryczna objętość toru głosowego nie ulegała zmianie. Trendy zmian częstotliwości formantowych w zależności od zmian konfiguracji toru głosowego są podobne jak w przypadku modelu 4-ro odcinkowego, aczkolwiek nie są już aż tak duże (szczególnie jest to widoczne dla trzeciego formantu). Wpływ zaokrąglenia jest szczególnie widoczny w przebiegu zmian częstotliwości F1 i F2 zmniejszając nieco wpływ położenia przewężenia na ich wartości.

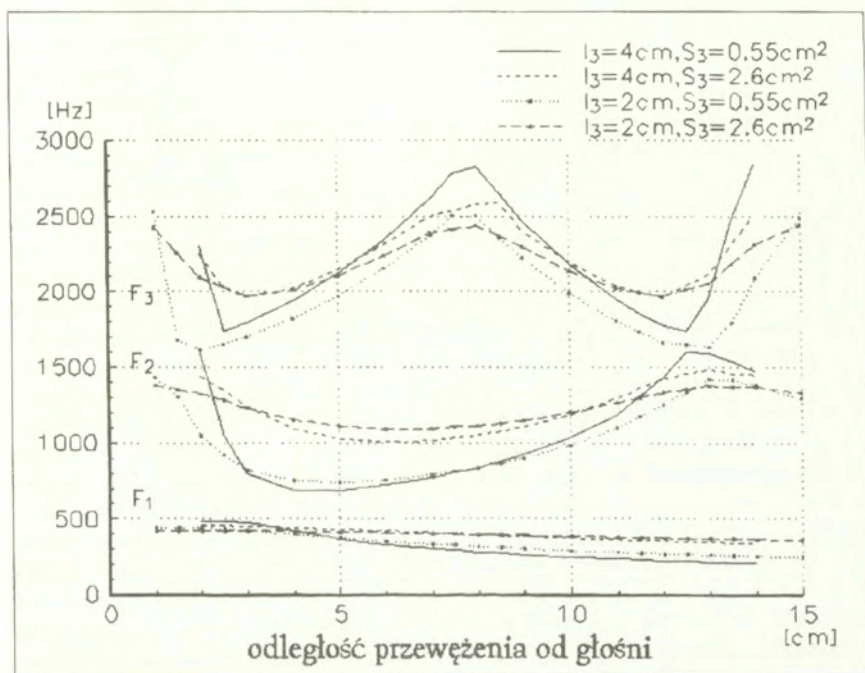


**Rys. 5.13** Przebieg zależności częstotliwości formantowych od położenia środka przewężenia dla 20-odcinkowego modelu cylindrycznego. Na rysunku zaznaczono położenie przewężenia odpowiadające samogłosce /u/. Przebiegi wyliczone przy zastosowaniu programu *Model*.

Tak więc, w oparciu o przedstawiony model toru głosowego można określać w wybiórzy sposób wpływ poszczególnych parametrów artykulacyjnych (to jest

położenia i wielkości przekroju poprzecznego przewężenia, zaokrąglenia warg) na charakterystykę widmową artykułowanych głosek, a zwłaszcza na ich częstotliwości formantowe. Pomijając odcinek znajdujący się w pobliżu głośni (jest on w zasadzie jeszcze w obrębie krtani) można generalnie stwierdzić, że im dalej przewężenie jest oddalone od głośni tym większa jest częstotliwość drugiego formantu. Znajduje to potwierdzenie w wynikach pomiarów częstotliwości formantowych samogłosek polskich (por. Tab. 4.1). Na przykład, w przypadku samogłosek, im bardziej artykulacja danej samogłoski jest przednia, tym częstotliwość formantu F2 jest większa. Tak więc dla samogłoski /i/ częstotliwość ta jest największa.

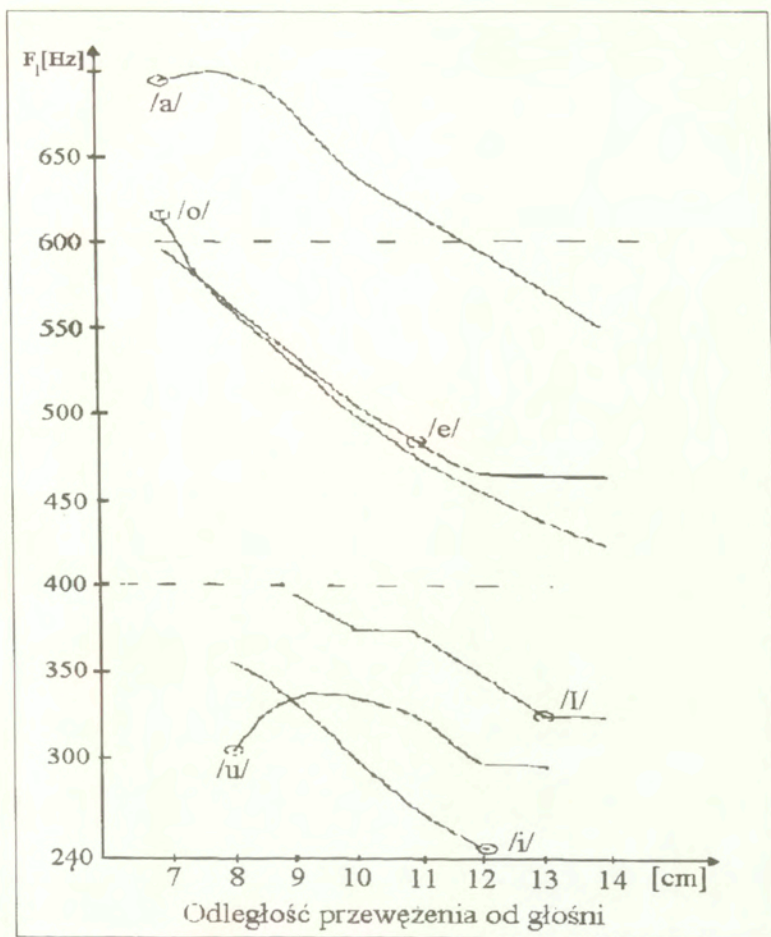
Na Rys. 5.14 przedstawiono wpływ wymiarów przewężenia, długości ( $l_3$ ) i powierzchni przekroju poprzecznego ( $S_3$ ) na przebieg zmian częstotliwości formantowych od położenia miejsca artykulacji.



**Rys. 5.14** Zależność częstotliwości formantowych od położenia środka przewężenia w torze głosowym dla różnych długości  $l_3$  i wielkości przekroju poprzecznego  $S_3$  przewężenia. Przebiegi wyliczone przy zastosowaniu programu *Model*.

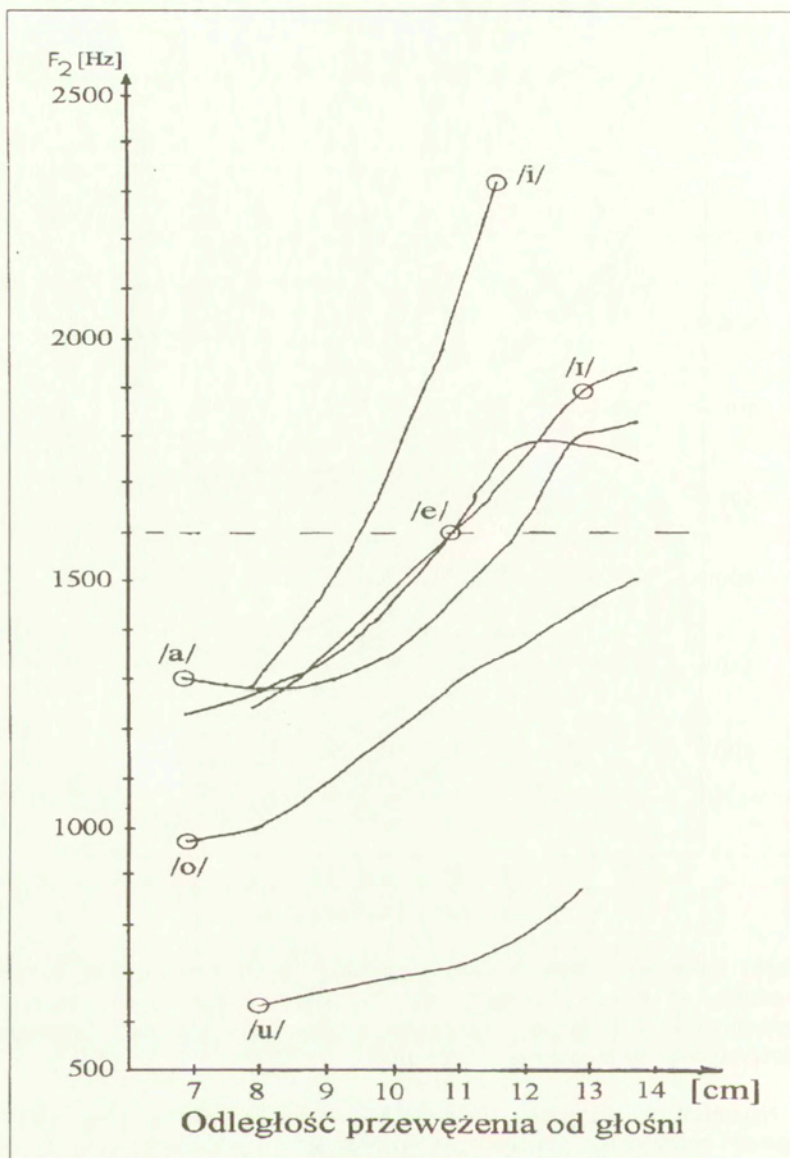
Z powyższego rysunku widać, że zwiększenie wielkości przekroju poprzecznego przewężenia powoduje zmniejszenie zakresu zmian częstotliwości formantowych, zwłaszcza  $F_1$  i  $F_2$ , przy zmianie jego położenia. Charakterystyczne jest obniżanie się częstotliwości trzeciego formantu w środkowym zakresie zmian położenia przewężenia przy zmniejszaniu jego długości

Na następnych wykresach (Rys. 5.15, 5.16 i 5.17) przedstawiono jak zmienia się dla poszczególnych modelowanych konfiguracji samogłoskowych częstotliwość pierwszego i drugiego formantu, wraz ze zmianą położenia przewężenia (definiowanego jako odległość środka przewężenia od głośni), przy zachowaniu stałej objętości toru głosowego. Oczywiście, w praktyce, warunek stałości objętości toru głosowego nie jest zachowany, ale wykresy te otrzymane w badaniach modelowych pokazują, jak nawet stosunkowo małe zmiany położenia przewężenia mogą wpływać na częstotliwości pierwszych dwóch formantów poszczególnych samogłosek. Długość przewężenia dla różnych konfiguracji samogłoskowych toru głosowego jest tu czynnikiem ograniczającym zakres zmian położenia przewężenia. Dlatego dla samogłosek otwartych (/a/, /o/, /e/) może być on większy, niż w przypadku samogłosek przymkniętych (/i/, /I/, /u/). Z Rys. 5.15 widać, że dla wszystkich konfiguracji samogłoskowych, z wyjątkiem konfiguracji odpowiadającej samogłosce /u/, przesunięcie przewężenia toru głosowego ku przodowi, powoduje obniżenie częstotliwości formantu  $F_1$  (ku tyłowi jego wzrost), tak jak to obserwuje się w mowie naturalnej. Również uporządkowanie krzywych przebiegu zmian  $F_1$  dla poszczególnych konfiguracji samogłoskowych jest zgodne z klasyfikacją samogłosek według wysokości artykulacji, tzn. najniżej przebiega krzywa odpowiadająca konfiguracji samogłoski /i/, najwyżej dla konfiguracji odpowiadającej samogłosce /a/. Dość specyficzny, nieregularny w stosunku do pozostałych krzywych przebieg zmian częstotliwości pierwszego formantu dla samogłoski /u/ jest spowodowany wpływem zaokrąglenia ust, oddziałującym na własności przedniej części toru głosowego. Również charakterystyczne jest występowanie dla niektórych konfiguracji, w pobliżu ich typowego miejsca artykulacji, zakresu zmian położenia przewężenia, w obrębie którego zmiany częstotliwości pierwszego formantu są stosunkowo niewielkie (por. konfiguracje /a/, /e/, czy /I/). W sumie, z przedstawionych wyników modelowania widać, że maksymalna zmiana częstotliwości pierwszego formantu wywołana zmianą położenia przewężenia nie przekracza 20-30% i jego częstotliwość jest wyznaczona przede wszystkim z wielkości przekroju poprzecznego przewężenia (tj. wysokości artykulacji) wynikającego z przyjętej danej konfiguracji samogłoskowej.



**Rys.5.15** Zależność częstotliwości pierwszego formantu  $F_1$  od położenia przewężenia w torze głosowym, dla różnych modelowanych konfiguracji samogłoskowych. Kółkami oznaczono położenie środka przewężenia charakterystyczne dla modelowanej samogłoski.

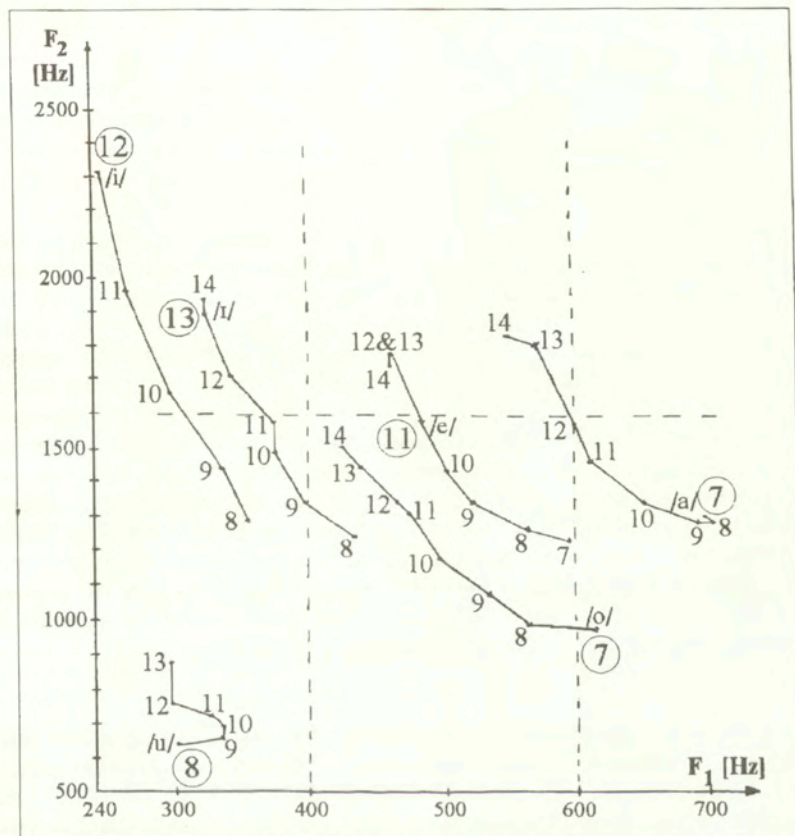
Natomiast w przypadku drugiego formantu (Rys. 5.16), dla wszystkich konfiguracji bez wyjątku, ten sam ruch przewężenia ku przodowi powoduje wzrost częstotliwości  $F_2$  (z wyjątkiem konfiguracji /e/, dla której pod koniec następuje niewielki spadek). W tym przypadku, widać że miejsce położenia przewężenia jest czynnikiem bardziej różnicującym artykulację samogłoskową, niż konfiguracja toru



**Rys. 5.16** Zależność częstotliwości drugiego formantu  $F_2$  od położenia przewężenia w torze głosowym, dla różnych modelowanych konfiguracji samogłoskowych. Kółkami oznaczono położenie przewężenia charakterystyczne dla modelowanej samogłoski.

głosowego (przynajmniej w przypadku samogłosek /a/, /e/, /i/, dla których krzywe przebiegają stosunkowo blisko siebie). Znamienne jest równoległy przebieg krzywych zmian dla niemal wszystkich konfiguracji samogłoskowych (z wyjątkiem /u/). Wskazuje on, że każda konfiguracja ma jakby swoją „linię” zmian, nie krzyżującą się, czy zachodzącą na pięć pozostałych.

Na Rys. 5.17 przedstawiono w płaszczyźnie F1-F2, w sposób zintegrowany zmiany częstotliwości pierwszych dwóch formantów wraz ze zmianą odległości przewężenia od głośni.



Rys. 5.17 Przebieg zmian częstotliwości formantowych w płaszczyźnie F1-F2 w zależności od położenia przewężenia w torze głosowym, wyznaczone dla sześciu konfiguracji samogłoskowych. Przy zaznaczonych punktach podano odległości przewężenia od głośni. Kółkami oznaczono położenie przewężenia charakterystyczne dla modelowanej samogłoski.

Podobnie jak dla poprzednich wykresów widać, że charakter zmian częstotliwości dla wszystkich konfiguracji z wyjątkiem /u/ jest podobny i nachylenie krzywych jest bardzo zbliżone do siebie. Tak znacznie odmienny przebieg dla samogłoski /u/ pokazuje jak istotny jest wpływ zaokrąglenia ust na częstotliwości pierwszych dwóch formantów.

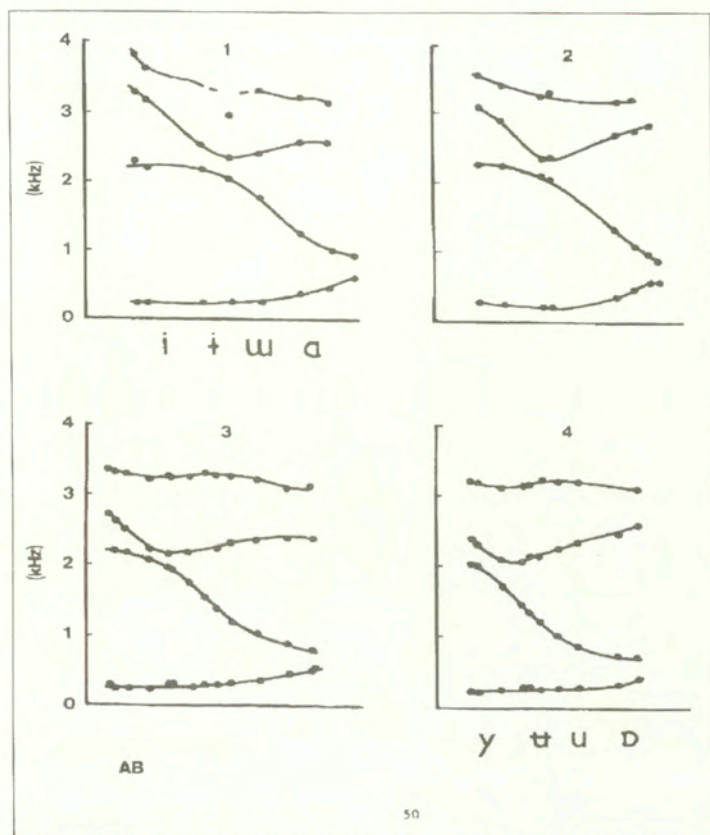
Wracając do nomogramów przedstawionych na Rys. 5.12 i 5.14 (po raz pierwszy, w postaci bardziej rozbudowanej opublikowano w [FANT,1960]) należy zasygnalizować, że były przez wiele lat traktowane jako nieco abstrakcyjne, nie mające zbyt wiele wspólnego z rzeczywistymi konfiguracjami toru głosowego i będącymi nie do odtworzenia przez osobę mówiącą. Dopiero wiele lat później Ladefoged i Bladon [LADEFOGED&BLADON,1982] wykazali, że wytrawni fonetycy są w stanie tak artykułować dźwięki samogłoskopodobne, by zmieniając w kontrolowany i w miarę płynny sposób położenie miejsce artykulacji w torze głosowym otrzymywać charakterystyki zbliżone do tych, które zostały przedstawione na Rys. 5.12, w zakresie odległości przewężenia od głośni powyżej 7 cm.

Eksperyment był prowadzony przez dwóch doświadczonych fonetyków, którzy starali się bardzo starannie wymawiać dźwięk samogłoskopodobny zmieniając jednocześnie w możliwie kontrolowany sposób, bądź miejsce tworzonego w torze głosowym przewężenia, bądź dla zadanego położenia jego wielkość (czyli wysokość artykulacji). Kontrolowano następujące parametry artykulacji: 1) położenie przewężenia w torze głosowym, 2) stopień otwarcia wylotu ust. Osoba mówiąca starała się tak artykułować, by następowało stopniowe przesuwanie się przewężenia od tyłu ku przodowi. W żadnym momencie, eksperymentatorzy nie byli w stanie dokładnie określić miejsce przewężenia w torze głosowym, lecz nie mieli wątpliwości, że przesunięcie jego następowało od skrajnego położenia znajdującego się nieco bliżej głośni niż położenie odpowiadające samogłosce tylnej /u/ (otwartej, niezaokrąglonej) jak w wyrazie angielskim "car", do położenia nieco za położeniem odpowiadającym samogłosce /i/ (samogłoska przednia, przymknięta jak w wyrazie angielskim "see", lub polskim "pić"). Jedyne, co było pewne, że nastąpiło jego przesunięcie ku przodowi w stosunku do poprzedniej artykulacji.

W badaniach, starano się by przewężenia miały stały przekrój odpowiadający samogłosce /i/, co kontrolowano częściowo za pomocą odpowiedniej aparatury ultradźwiękowej. Stopień zaokrąglenia warg był wyznaczany przez dokładny pomiar powierzchni wylotu ust. Każda osoba zarejestrowała odpowiednie ciągi dźwięków samogłoskopodobnych dla 4 stopni zaokrąglenia warg, przesuując miejsce artykulacji stopniowo od strony głośni w kierunku ust. Pozioma współrzędna na wykresach nie jest wyskalowana, ponieważ nie było możliwe dokładne wyznaczenie położenia przewężenia w torze głosowym, z tego powodu punktami zaznaczono umownie tylko te miejsca dla których wyznaczono częstotliwości pierwszych



czterech formantów. W ten sposób uzyskano wyniki zbliżone do podanych przez Fanta dla powierzchni przewężenia równej  $A=0.65 \text{ cm}^2$ . Przykładowy diagram wyznaczony dla jednego z fonetyków biorących udział w opisanym eksperymencie przedstawiono na Rys. 5.18.



**Rys. 5.18** Przykład zmiany częstotliwości formantowych przy przesuwaniu w sposób kontrolowany miejsca artykulacji w torze głosowym przez wytrawnego fonetyka [LADEFOGED&BLADON,1982], dla czterech stopni zaokrąglenia warg. Dla niektórych konfiguracji toru głosowego można przypisać określoną głoskę (znaczenie fonetyczne). Kierunek zmian położenia przewężenia jest od strony ust ku głośni.

Uzyskane przebiegi częstotliwości formantowych są podobne w swym ogólnym charakterze do już wcześniej omówionych. Dla tych konfiguracji toru głosowego, którym można było przypisać określone znaczenie fonetyczne, na wykresach oznaczono te miejsca położenia przewężenia odpowiednim symbolem

fonetycznym. Choć są pewne różnice między przebiegami otrzymanymi dla modelu i tymi które otrzymali w swoich badaniach wspomniani fonetycy, jednak ogólny charakter zmian jest w obu przypadkach podobny.

Przedstawione wyniki modelowania toru głosowego, oparte na zastosowaniu analogii akusto-elektrycznych, oraz wyniki próby weryfikacji tego modelu w rzeczywistym organie artykulacyjnym, potwierdzają, że wyprowadzone dla modelu zależności uwzględniają, w znacznej mierze fizyczne własności narządu artykulacyjnego.

## 6. Modelowanie pobudzenia krtaniowego

### 6.1 Podstawy modelowania własności źródła krtaniowego.

Jednym z ważnych problemów dotyczących zarówno syntezy, jak i oceny stanu źródła krtaniowego jest analiza kształtu i parametrów fali pobudzenia krtaniowego, która w mowie naturalnej ulega podczas artykulacji znacznym zmianom. W przypadku syntezy mowy, jednym z istotniejszych zadań jest określenie postaci oraz zakresu tych zmian, które trzeba odtworzyć, aby uzyskać możliwie naturalne brzmienie syntetycznego sygnału. W celu odtworzenia z akustycznego sygnału realizowana jest trzystopniowa analiza sygnału - odwrotna filtracja sygnału wypowiedzi, resynteza po zastąpieniu naturalnej fali pobudzenia krtaniowego wydzielonym, oraz subiektywna ocena otrzymanego sygnału. W procesie odwrotnej filtracji, funkcja transmitancji toru głosowego jest reprezentowana dla danej konfiguracji przez odpowiedni czwórnik. Przelfiltrowanie sygnału przez odwrotną funkcję transmitancji czwórnika daje na wyjściu funkcję pobudzenia toru głosowego. W czwórnikowym modelowaniu toru głosowego jest ona funkcją prędkości objętościowej na poziomie fałdów głosowych.

W celu wyjaśnienia zasady działania źródła krtaniowego, którego zasadniczym elementem są drgające fałdy głosowe, konieczne jest poznanie przebiegu zmian czasowych dwóch podstawowych parametrów, to jest powierzchni otworu głośni ( $A_g$ ) oraz prędkości objętościowej ( $U_g$ ) powietrza przepływającego przez głośnię. Jednakże niezbędne jest przedtem poznanie budowy, zasad funkcjonowania oraz podstaw modelowania źródła krtaniowego uczestniczącego w formowaniu wszystkich dźwięków mowy o strukturze harmoniczej lub harmoniczo-szumowej.

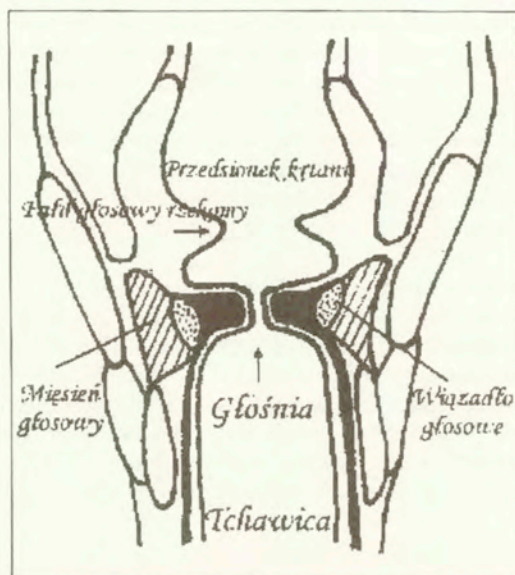
### 6.2 Budowa i funkcjonowanie krtani.

Dostęp do krtani jest utrudniony i najczęściej, aby dokonać wizualizacji fałdów głosowych podczas fonacji w sposób najmniej zakłócający artykulację, stosuje się coraz powszechniej światłowód (tzw. fiberoskop) wprowadzany przez jeden z otworów nosowych. Należy jednak zaznaczyć, że nie wszystkie osoby znoszą dobrze obecność obcego ciała w części nosowej, a tym bardziej w części ponadkrtaniowej, poniżej języczka podniebienia miękkiego. Powszechnie stosowana obecnie stroboskopia umożliwia obserwację działania źródła krtaniowego tylko podczas artykulacji izolowanych głosek otwartych wymawianych w stanie ustalonym (w praktyce jest to niemal wyłącznie samogłoska /a/).

Na Rys. 6.1 przedstawiono widok od przodu pionowego przekroju krtani, przedstawiający wewnętrzną konfigurację jej poszczególnych elementów

uczestniczących w fonacji. Zasadniczymi elementem źródła krtaniowego jest zespół fałdy głosowe - fałdy kieszonki krtaniowej (fałdy rzekome) - kieszonka krtaniowa (Morganiego) i może on przybierać trzy zasadnicze położenia:

- położenie oddechowe, w którym fałdy kieszonki krtaniowej i fałdy głosowe są rozchylone, umożliwiając w ten sposób swobodny przepływ powietrza przez głośnie,
- położenie prefonacyjne, w którym fałdy kieszonki krtaniowej i fałdy głosowe są zbliżone do siebie,
- położenie fonacyjne, w którym fałdy kieszonki są rozchylone, zaś tkanki łączne fałdów głosowych wykonują cykliczne ruchy zamykając i rozwierając kolejno głośnie.



Rys. 6.1 Widok przekroju pionowego krtani

Podczas fonacji, przy niskiej częstotliwości drgań, fałdy głosowe są stosunkowo słabo napięte i cykliczne zamykanie i rozwieranie głośni następuje wskutek oddziaływania ciśnienia podgłośniowego na błonę śluzówkową fałdów głosowych. Przy zamkniętej głośni, najpierw dolna część fałdów głosowych rozchyla się, a następnie ruch ten przenosi się ku górnej ich części. Podobnie, zamykanie się głośni następuje w wyniku zbliżenia się do siebie najpierw ich części dolnej, a następnie górnej. To przesunięcie fazy w ruchach dolnej i górnej części fałdy głosowego, zaobserwowane już przez Farnswortha (FARNSWORTH, 1940), jest rzędu  $80^{\circ}$ - $90^{\circ}$ .

### 6.3 Podstawy działania źródła krtaniowego

Podczas fonacji chrząstki nalewkowe przybliżają się do siebie w celu zwężenia szpary głośni. Wskutek wytworzenia ciśnienia podgłośniowego przez powietrze wychodzące z płuc wiązadła głosowe są wprowadzane w ruch drgający, poruszając się do i od linii symetrii głośni, cyklicznie zamykając i otwierając głośnię modulując w ten sposób strumień przepływającego powietrza w ciąg quasi-periodycznych impulsów. Każdy z tych impulsów pobudza do drgań powietrze znajdujące się ponad głośnią w torze gardłowo-ustno-nosowym. Ruch wiązadeł głosowych, jest złożony, bowiem nie poruszają się one jako jedna sztywna całość. Różne partie każdego z wiązadeł głosowych drgają w z pewnym przesunięciem w czasie, zarówno w płaszczyźnie pionowej, jak i poziomej.

Drgania wiązadeł głosowych, które są drganiami typu relaksacyjnego, są wynikiem działających na nie przemiennych sił, otwierających i zamykających głośnię. Siły te są trojkiego pochodzenia - aerodynamicznego, mięśniowego i elastycznego, i oddziałują w złożony sposób w krtani na proces fonacji. Teoria opisująca działanie tych sił nazywana jest aerodynamiczno-mioelastyczną teorią fonacji (VAN DEN BERG, 1958) Teoria jest dość skomplikowana, ale można skupić się na analizie trzech zasadniczych sił oddziałujących na fałdy głosowe. Niezależnie od przyjmowanego modelu, ruch fałdów głosowych jest wynikiem oddziaływania na nie dwóch zasadniczych sił: typu mechanicznego (napinanie fałdów przez mięśnie) oraz typu aerodynamicznego, wywołanego przez różnicę ciśnień powietrza, powstałą po obu stronach fałdów głosowych.

Jednym z zasadniczych czynników regulujących cykliczne zamykanie się i otwieranie się głośni jest napięcie wiązadeł. Siła zamykająca powstaje przede wszystkim w wyniku naciągania ich między chrząstkami nalewkowymi i chrząstką tarczową, wskutek czego fałdy są pod oddziaływaniem napięcia wzdłużnego. Podczas artykulacji dźwięcznych głosek wiązadła głosowe są w zasadzie rozwarte (w położeniu fonacyjnym) i wprowadzane są w ruch drgający przez strumień powietrza przepływający przez głośnię. Im większe jest ciśnienie podgłośniowe, tym większa jest siła dośrodkowa i tym samym większa jest częstotliwość drgań fałdów głosowych. Prędkości przepływu przez szparę głośni powietrza, są dość znaczne i mogą sięgać nawet 0.2-0.5 Macha.

Trzecia siła, która może wystąpić podczas fonacji, jest przyłożona poprzecznie do wiązadeł głosowych. Im ona jest większa, tym wyższa jest częstotliwość drgań fałdów głosowych. Wzrost częstotliwości następuje w wyniku skrócenia części drgającej fałdów, które jednak nie pociąga za sobą zmiany napięcia wzdłużnego, a jedynie zmniejszenie masy drgającej. Jednakże wyżej wymienione czynniki wpływające na częstotliwość podstawową nie oddziałują w sposób niezależny jeden od drugiego.

Sposób w jaki zmieniają się parametry fizyczne źródła krtaniowego zależy przede wszystkim od długości, masy i napięcia fałdów głosowych. Zmiana wymiarów fałdów głosowych następuje głównie wskutek ruchu chrząstek nalewkowych i tarczowej. Szerokość szpary głośni zmienia się zgodnie z ruchami zwierającymi lub rozwierającymi chrząstek nalewkowych. Siła dociskająca fałdy głosowe do siebie jest określona przez zwieranie się chrząstek nalewkowych, które może być spowodowane przez ich ruchy wahadłowe i przesuwanie się w stawach chrząstkowych. Długość fałdów głosowych z drugiej strony, ulega zmianie na wskutek obrotu w stawie chrząstki tarczowej. Obrót ten umożliwia przesunięcie się chrząstek nalewkowych do tyłu od chrząstki tarczowej napinając w ten sposób wiązadła głosowe, co zwiększa jednocześnie ich napięcie i grubość. Ruchy chrząstek są oczywiście wywołane przez napinanie odpowiednich mięśni krtani.

Siły aerodynamiczne oddziaływujące na wiązadła głosowe są wynikiem oddziaływania wewnątrzgłośniowego ciśnienia powietrza. Składowe tego ciśnienia wyznaczone są w przybliżeniu z prawa Bernoulliego, które wynika z uproszczonego równania zachowania energii.

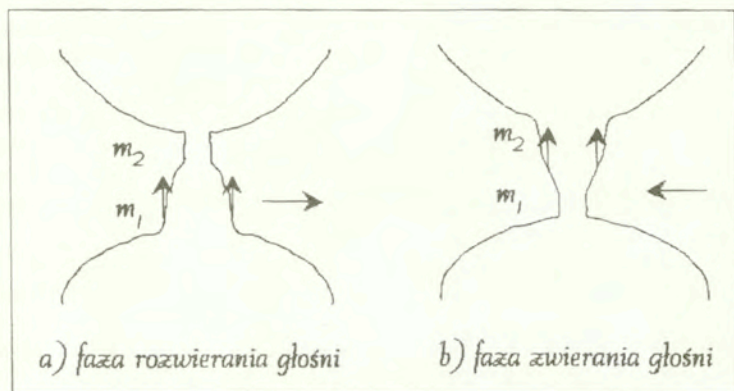
$$P + 1/2 rU^2 = \text{constant},$$

gdzie P jest ciśnieniem powietrza, r - gęstością i U- prędkością objętościową powietrza. Tak więc, gdy prędkość powietrza wzrasta, ciśnienie wewnątrz głośni maleje i na odwrót. Ciśnienie tuż pod wiązadłami jest w przybliżeniu równe ciśnieniu podgłośniowemu, zaś u wylotu głośni - nieco mniejsze od ciśnienia atmosferycznego. Jest to wynikiem tego, że w tym punkcie toru głosowego suma energii kinetycznej i potencjalnej przepływającego powietrza musi być równa statycznej energii potencjalnej wynikającej z ciśnienia atmosferycznego. Stąd, powietrze przepływając przez głośnię maleje w sposób ciągły, od ciśnienia podgłośniowego, do nieco mniejszego od atmosferycznego.

Na Rys. 6.2 przedstawiono dwa podstawowe stadia drgań fałdów głosowych (wg [TITZE, 1988]), przy założeniu, że ich ruch jest symetryczny względem osi symetrii głośni (co nie zawsze ma miejsce, zwłaszcza w przypadku patologii krtani). Na rysunku schematycznie pokazano obserwowane przesunięcie fazowe ruchu górnej i dolnej części fałdu o masach oznaczonych odpowiednio  $m_1$  i  $m_2$ .

Przy określonych średnich prędkościach cząsteczek w strumieniu powietrza przepływającego przez górne i dolne odcinki głośni o odpowiednich powierzchniach przekrojów  $A_1$  i  $A_2$ . W wyniku napięcia wzdłużnego, na każdą z mas  $m_1$  i  $m_2$  oddziałuje siła sprężysta skierowana w stronę położenia spoczynkowego. Ponadto, wskutek przesunięcia w ruchu górnej i dolnej części fałdu oddziałuje między nimi dodatkowa siła sprężysta.

Powstające w głośni siły aerodynamiczne oddziałują na fałdy głosowe w taki sposób, by początkowe zmiany położenia masy  $m_1$  były kontynuowane, w kierunku do linii symetrii głośni, lub na zewnątrz od niej (zależnie od fazy zwierania lub rozwierania fałdów). Z tego powodu początkowe położenie równowagi jest niestabilne jednakże ruch dolnej masy jest z jednej strony ograniczony jest zwieraniem się głośni, z drugiej nieliniowym narastaniem sił sprężystych oddziałujących na fałdy przy większych zmianach ich położenia. Podtrzymujące się drgania więzadeł głosowych są wynikiem dynamicznej wzajemnej wymiany sił pochodzenia mechanicznego i aerodynamicznego oddziałujących na więzadła głosowe i na momenty mas  $m_1$  i  $m_2$ . Pełne wyjaśnienie procesu drgań więzadeł wymaga nie tylko określenia wielkości sił oddziałujących na nie, lecz również momentu ich występowania, tzn. względnych faz sił, przesunięć i prędkości mas  $m_1$  i  $m_2$ . Ponieważ siły pochodzenia aerodynamicznego oddziałujące na masę  $m_2$  są małe w stosunku do oddziałujących na masę  $m_1$ , ruch masy  $m_2$  odbywa się głównie pod wpływem mechanicznej siły sprężystości sprzęgającej ze sobą obie masy. Tak więc masa  $m_2$  jest "ciągniona za sobą" przez masę  $m_1$ , co wyjaśnia opóźnienie ruchu w cyklu drganiowym górnej części więzadeł głosowych względem dolnej.



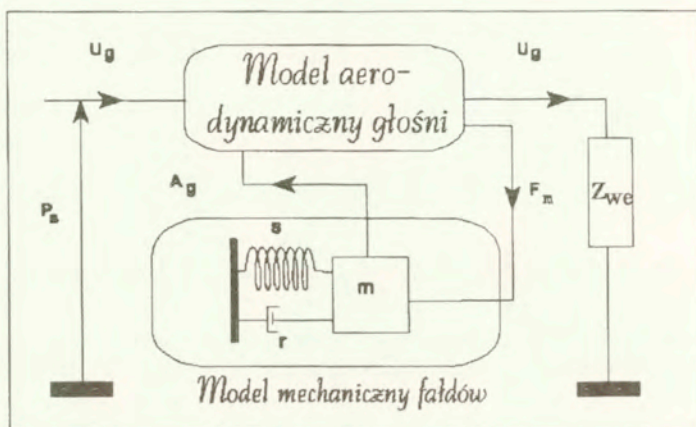
**Rys. 6.2** Stadia drgań fałdów głosowych wraz z oznaczeniem kierunków oddziałujących na nie sił pochodzenia aerodynamicznego (wg [TITZE,1988]).

Istotne w wyjaśnieniu mechanizmu fonacji powstającej w wyniku drgań fałdów głosowych jest to że ruch ich nie jest bezpośrednio przekazywany do otaczającego powietrza, lecz odbywa się poprzez modulację objętości przepływającego przez głośnię strumienia powietrza. Formowane w ten sposób impulsy powietrza są rzeczywistym źródłem dźwięczności. Natomiast bezpośredni udział drgań więzadeł w widmie generowanego dźwięku jest akustycznie pomijalnie mały. Dlatego w pierwszym przybliżeniu, przy analizie podstaw powstawania impulsów powietrza pobudzających

do drgań powietrze znajdujące się w części ponadkrtaniowej, można stosować model jedno-masowy, w którym masa zastępcza fałdu  $m$  jest równa sumie mas  $m_1$  i  $m_2$ .

Z dotychczasowych rozważań wynika więc, że w modelu źródła krtaniowego należy wyodrębnić dwa jego zasadnicze zespoły – mechaniczny oraz aerodynamiczny, tak jak to przedstawiono na Rys.6.3. W schemacie tym uwzględniono również wpływ impedancji wejściowej toru głosowego  $Z_{we}$  obciążającej je. W modelu jedna część odwzorowuje powstawanie w układzie aerodynamicznym siły ( $F_m$ ) oddziaływującej na ścianki fałdów głosowych, zaś druga opisuje oddziaływanie części mechanicznej głośni na warunki przepływu powietrza przez nią, poprzez modyfikację powierzchni przekroju głośni ( $A_g$ ).

Podstawową zasadą jest, że ruchy drgające ścianek fałdów są wywołane oddziaływaniem na nie strumienia powietrza przepływającego przez głośnię. Fałdy tworzą więc z jednej strony układ mechaniczny, który w pewnych warunkach przepływu może znaleźć się w stanie niestabilnym. Następuje wówczas ciągły przepływ energii od strumienia powietrza płynącego przez głośnię do ścianek fałdów, co niweluje straty wywołane przez tarcie wewnętrzne w fałdach głosowych. Bezwładnościowe i sprężyste własności drgającej ścianek fałdów (masa oraz sztywność) w połączeniu z geometrią głośni wyznaczają zakres drgań, zarówno ich amplitudę, jak i częstotliwość.



Rys. 6.3 Ogólny schemat funkcjonalny modelu źródła krtaniowego. Opis oznaczeń w tekście pracy.

Dodatni przepływ energii od strumienia powietrza do drgających ścianek fałdów następuje tylko wówczas, gdy siła pobudzająca jest wyłącznie pochodzenia aerodynamicznego i ma składową będącą w fazie z prędkością ich poruszania się.



Podstawowe równanie ruchu dla układu drgającego składającego się z masy i sprężyny ma postać:

$$m \frac{d^2 x}{dt^2} + r \frac{dx}{dt} + Sx = F_m(x, x, t), \quad 6.1$$

gdzie  $m$ ,  $r$  i  $S$  są odpowiednio masą, tłumieniem i sztywnością układu, zaś  $x$ ,  $dx$  - przemieszczeniem i prędkością,  $F_m$  - siłą pobudzającą oraz  $t$  - czas. Współczynnik  $r$  reprezentuje oporność wynikającą z istnienia sił rozproszonych - takich jak np. siły wiskoelastyczne w tkankach, które zmniejszają w istocie szybkość przemieszczania się masy fałdu  $W$  przypadku, gdy siła  $F_m$  jest niezmienna w czasie, równanie różniczkowe drugiego rzędu jest autonomiczne (to znaczy, że rozwiązanie jest funkcją niezależną od czasu) i wówczas opisuje układ będący w stanie drgań samowzbudnych, a nie wymuszonych (z udziałem siły zewnętrznej). Zależność siły pobudzającej od prędkości przepływu jest tu zasadnicza dla powstawania drgań. W momentach, gdy jest ona zgodna z kierunkiem wektora prędkości, energia strumienia powietrza jest przekazywana do układu masa-sprężyna-tłumienie, w innych momentach jest z niego odbierana. Ponieważ siła ta powstaje wskutek zjawiska Bernoulliego wewnątrz układu, wykonuje on więc drgania swobodne. Gdy siła ta ma znak ujemny, następuje zbliżenie do siebie ścianek fałdów, a w krańcowym przypadku ich zetknięcie. Ruchowi ścianek fałdów przeciwdziała ich sztywność  $S$  powodująca powstanie odpowiedniej siły zwrotnej. W momencie zwarcia się fałdów siła Bernoulliego jest równa zeru i siła zwrotna powoduje otwieranie się głośni. To przemienne powstawanie i zanikanie wymienionych sił uzyskiwane jest przynajmniej na dwa różne sposoby:

- 1) wskutek wytwarzania się przeciwnie skierowanych ciśnień ponad- i podgłośniowego oraz
- 2) w wyniku zmieniającej się geometrii szpary głośni, co powoduje powstanie wewnątrz niej zmiennego rozkładu ciśnień.

W pierwszym przypadku, oscylacje są ułatwione przez bezwładnościowe własności słupa powietrza w części ponadkraniowej toru głosowego, zaś w drugim przez niejednorodną deformację fałdów. Oba przypadki nie wykluczają się wzajemnie i podczas normalnej fonacji, oba wymienione mechanizmy mogą występować jednocześnie.

Rozpatrując pierwszy przypadek, należy uwzględnić, że częstotliwość podstawowa  $F_0$  jest z reguły poniżej częstotliwości pierwszego formantu (przynajmniej dla głosów męskich), stąd słup powietrza w torze głosowym zachowuje się jak masa powietrza, która jest jako całość tylko przyspieszana lub opóźniana [ROTHENBERG,

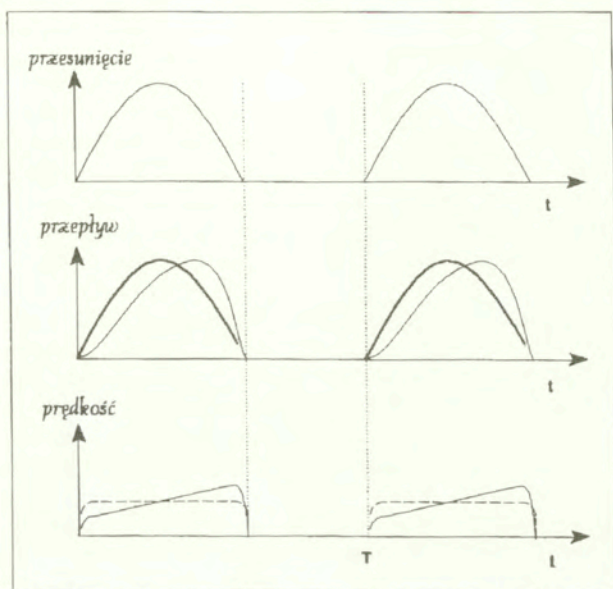
1980]. W momencie otwierania się głośni, wzrasta przepływ powietrza przez głośnię, w wyniku czego następuje przyśpieszenie słupa powietrza znajdującego się w torze gardłowo-ustnym, zgodnie z kierunkiem przepływu. Do nadania tego przyśpieszenia potrzebne jest wytworzenie u wylotu głośni ciśnienia wyższego od atmosferycznego.

Natomiast ciśnienie wewnątrz głośni wzrasta odpowiednio powodując rozchylenie się fałdów. Siła oddziaływująca na fałdy jest więc zgodna z kierunkiem prędkości poruszania się fałdów i energia jest do nich przekazywana. Natomiast w okresie zamykania się głośni wspomniany słup powietrza dąży do zachowania poprzedniego momentu pędu skierowanego ku wylotowi ust. Pociąga to za sobą zmniejszenie ciśnienia ponadgłośniowego, to jest ciśnienia powstającego u podstawy słupa powietrza i jednocześnie zmniejszanie się średniego ciśnienia wewnątrz głośni i tym samym malenie siły oddziaływującej w kierunku przeciwnym do ruchu powrotnego fałdów głosowych. Tak więc w pewnym momencie, tuż przed ich zwarciem, średnie ciśnienie wewnątrz głośni może stać się ujemne, i może wystąpić siła oddziaływująca, tym razem zgodnie z kierunkiem ruchu fałdów. Nie jest to zresztą niezbędne dla podtrzymania drgań. Wystarczy tylko, aby była dostatecznie mała podczas zamykania, mniejsza niż podczas otwierania głośni oraz, by siły oddziaływujące na ścianki fałdów były wystarczająco duże do pokonania sił tarcia powietrza o ścianki.

Bezwładność słupa powietrza nie wpływa na ruchy boczne ścianek fałdów głosowych i ich prędkość. Natomiast w sposób zasadniczy wpływa na kształt fali pobudzenia krtaniowego, to jest na przebieg czasowy wielkości i prędkości przepływu powietrza, jak to przedstawiono na Rys. 6.4 ([ROTHENBERG,1980]).

Kształt fali pobudzenia krtaniowego (przebiegu czasowego zmian wielkości wydatku strumienia powietrza) był szczegółowo omawiany i modelowany (np. [ROTHENBERG,1980; 1981; FANT, 1982]). Jej cechą znaną zarówno dla fali natężenia, jak i prędkości przepływu jest jej stosunkowo wolne narastanie w okresie rozwierania się fałdów, jak i gwałtowne opadanie w końcowej fazie ich zwięcia. Prawoskośność fali pobudzenia krtaniowego wynika z bezwładności masy powietrza znajdującego się w torze gardłowo-ustnym. Nadanie przyśpieszenia tej masie wymaga pewnego czasu, stąd osiągnięcie odpowiedniej wartości następuje z pewnym opóźnieniem w stosunku do ruchu ścianek fałdów. Tak więc szczyt przepływu powietrza jest opóźniony względem momentu maksymalnego rozwarcia fałdów. W przypadku nie uwzględniania wspomnianej bezwładności masy powietrza (odpowiednie przebiegi zostały oznaczone na Rys. 6.4 linią przerywaną) zachowana jest pełna synchronizacja między przebiegiem zmian wielkości strumienia przepływającego powietrza i ruchami ścianek fałdów głosowych. Hamowanie (podobnie jak i przyśpieszanie) masy powietrza również dokonywane jest z pewnym opóźnieniem, jednakże w momencie szybkiego zamykania się głośni, gwałtownie maleje jednocześnie strumień powietrza. Asymetria przebiegów jest głównym

czynnikiem w powstawaniu siły oddziaływującej na fałdy, która zależy przede wszystkim od prędkości przepływu powietrza. Przebieg czasowych zmian prędkości przepływu cząsteczek powietrza w obrębie głośni można wyznaczyć dzieląc funkcję przepływu przez funkcję przesunięcia (to jest dzieląc górny wykres przez środkowy), bowiem ta ostatnia jest proporcjonalna do przekroju poprzecznego głośni. Również i w tym przypadku funkcja prędkości jest asymetryczna względem maksimum, wzrastając stopniowo w fazie rozwierania. Przy pominięciu wpływu bezwładności masy powietrza funkcja prędkości jest stała.



Rys. 6.4 Wyidealizowane przebiegi czasowe parametrów źródła krtaniowego: przesunięcia fałdów, przepływu powietrza przez głośnię oraz prędkości. Linia kreskowaną oznaczono przebiegi, przy nie uwzględnieniu bezwładności masy powietrza znajdującego się w części ponadkrtaniowej [ROTHENBERG,1980].

Zmiany ciśnienia wewnątrzgłośniowego  $P_w$ , a więc tym samym siły oddziaływującej na fałdy głosowe można wyznaczyć w oparciu o prawo Bernoulliego. Przy pomijalnie małej składowej kinetycznej ciśnienia w obszarze ponadgłośniowym

$$P_w = P_s - \frac{1}{2} \rho u^2, \quad 6.2$$

gdzie  $P_s$  jest ciśnieniem podgłośniowym,  $\rho$  - gęstością powietrza, zaś  $u$  - prędkością przepływu. Z powyższego wzoru widać silną zależność ciśnienia wewnątrzgłośniowego

od prędkości przepływu. Różniczkując powyższe równanie względem prędkości, przy niezmiennym ciśnieniu podgłośniowym,

$$\Delta p_w = -\rho u \Delta u$$

6.3

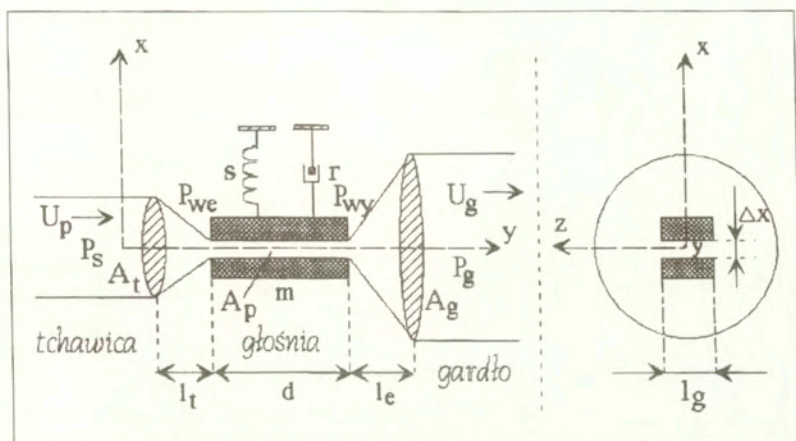
uzyskuje się, że przy wzroście liniowej prędkości przepływu  $u$  następuje malenie ciśnienia oddziaływującego na fałdy głosowe, co jest kluczem do wyjaśnienia niestabilności układu przez nie uformowanego, a więc tym samym wyjaśnia mechanizm wprowadzania ich w ruch drgający. W momencie, gdy znak ciśnienia staje się dodatni, siła oddziaływująca na ścianki głośni utworzone przez fałdy rośnie ze wzrostem prędkości przepływu, zaś prędkość ruchu fałdów maleje. Siła ta w efekcie tłumi ruch fałdów w tej fazie drgań. Należy zwrócić uwagę, że istnienie w pewnej fazie cyklu ujemnego ciśnienia Bernoulliego nie jest wystarczające by powstało tłumienie o ujemnym znaku. W przypadku, gdy nie istnieje bezwładność masy powietrza, wszystko co zyskuje się podczas fazy rozwierania, jest stracone podczas zamykania głośni i nie ma przemiany energii z jednej postaci na drugą, co ma miejsce przy asymetrii przebiegów. Więcej energii jest przekazywane ze strumienia przepływającego powietrza, niż tracone, głównie wskutek tarcia w fałdach głosowych.

Inny ważny czynnik, o zupełnie odmiennym charakterze niż wyżej omówiony, powodujący powstawanie w głośni sił również zależnych od prędkości przepływu, jest związany z falą poprzeczną rozchodzącą się w fałdzie głosowym. Powoduje ona chwilowe zmiany przekroju głośni i tym samym zmiany rozkładu ciśnień wewnątrz niej. Zmiana kształtu głośni pociąga za sobą, przy założeniu, że prawo zachowania energii Bernoulliego może być stosowane (nawet w przybliżeniu) to ciśnienie rośnie od góry ku dołowi w przypadku zbieżnej konfiguracji głośni (rozwieranie), natomiast w przeciwnej fazie (zwieranie), maleje. Wynika to z faktu, że zgodnie z prawem Bernoulliego i zasadą zachowania ciągłości przepływu, wzrostowi (lub maleniu) powierzchni przekroju kanału głośni towarzyszy wzrost (lub odpowiednio - malenie) w nim ciśnienia. Dzięki temu powstaje niezerównoważone ciśnienie oddziaływujące na fałdy. Ponieważ średnie ciśnienie w głośni jest większe w przypadku jej zbieżnego kształtu, niż rozbieżnego, może wytworzyć się wewnątrz niej siła zależna od prędkości strumienia powietrza, od którego następuje przepływ energii do fałdów głosowych.

#### 6.4 Model jednomasowy

Jednym z najprostszych do zamodelowania układu drgającego fałdów głosowych jest model jednomasowy (taki zresztą założono w Rys. 6.3). Zakłada się wówczas symetrię ruchów prostoliniowych obu fałdów głosowych, zgodną w fazie ich dolnych i górnych części, co jest słuszne przy założeniu istnienia tylko podstawowego modu drgań. Jest to spełnione przy pełnej sztywności ścianek obu fałdów.

Głośnię można wówczas przedstawić za pomocą odcinka cylindrycznego o powierzchni przekroju zmieniającej się podczas przepływu przez nią powietrza. Natomiast propagacja dźwięku poprzez głośnię może być rozpatrywana jako jednowymiarowa. To założenie jest słuszne dla przypadków, gdy długość fali jest znacznie większa od wymiarów poprzecznych głośni. Uproszczony model jednomasowy jest przedstawiony na Rys. 6.5, przy założeniu prostopadłościennego kształtu szczeliny głośni.



Rys. 6.5 Jednomasowy model źródła kraniowego z wyróżnionymi parametrami mechanicznymi i aerodynamicznymi (oznaczenia objaśnione w tekście).

Chwilowa wielkość powierzchni przekroju głośni wynosi

$$A_g = A_{g_0} + 2l_g \cdot \Delta x. \quad 6.4$$

Przy założeniu ciągłości i quasistacjonarności (równomierności) przepływu, co jest spełnione dla małych rozmiarów głośni, oraz jednocześnie dla dużej prędkości przepływu przez nią powietrza, z aerodynamicznych własności krtani wynika zależność między prędkością objętościową  $U_g$  powietrza i różnicą ciśnień pod-  $P_s$  i ponadgłośniowego  $P_g$ . Stosując system analogii ciśnienie - napięcie elektryczne oraz prędkość objętościowa - natężenie prądu, można zdefiniować oporność głośni  $R_g$  jako stosunek różnicy powyższych ciśnień do prędkości objętościowej

$$R_g = \frac{P_s - P_g}{U_g}. \quad 6.5$$

Jednocześnie oporność ta jest sumą dwóch składników: a) oporności kinetycznej  $R_k$  reprezentującej straty kinetyczne wywołane wahaniami ciśnienia w obszarze wlotu

głośni o długości  $l_1$  oraz w odcinku  $l_e$ , gdzie jej przekrój się rozszerza oraz b) oporności  $R_v$  reprezentującej straty wskutek tarcia lepkościowego powietrza o ścianki głośni ([FLANAGAN, 1972]). Tak więc

$$R_g = R_t + R_e \quad 6.6$$

Różnice ciśnień  $P_s - P_{we}$  oraz  $P_{wy} - P_g$  mogą być wyliczone z równania zachowania ciągłości przepływu (stałości wydatku powietrza wpływającego i wypływającego z głośni) oraz z równania Bernoulliego. W strefie zwężania się głośni ( $l_1$ ) mamy więc

$$P_s + \frac{1}{2} \rho V_s^2 = P_{we} + \frac{1}{2} \rho V_{we}^2 = \text{const} \quad (r - e \text{ Bernoulliego}) \quad 6.7$$

oraz

$$V_s \cdot A_t = V_{we} \cdot A_p = U_g \quad (\text{zachowanie wydatku powietrza}), \quad 6.8$$

gdzie gęstość powietrza  $\rho = 1.29 \text{ kg/m}^3$ , zaś  $V_s$  i  $V_{we}$  są prędkościami cząsteczkowymi odpowiednio na powierzchniach  $A_t$  i  $A_p$ .  $A_t$  jest powierzchnią przekroju poprzecznego tchawicy.

Z dwóch poprzednich równań (6.7 i 6.8) wynika, że

$$P_s - P_{we} = \frac{\rho U_g^2}{2} \left[ \frac{1}{A_p^2} - \frac{1}{A_t^2} \right]. \quad 6.9$$

Podobnie w strefie nadgłośniowej o wzrastającym przekroju

$$P_{wy} - P_g = \frac{\rho U_g^2}{2} \left[ \frac{1}{A_g^2} - \frac{1}{A_p^2} \right], \quad 6.10$$

gdzie  $A_g$  jest powierzchnią przekroju toru głosowego u jego wejścia. Podstawiając ciśnienie Bernoulliego

$$P_{Be} = \frac{\rho U_g^2}{2 A_p^2} \quad 6.11$$

otrzymuje się

$$P_s - P_{we} = P_{Be} \left[ 1 - \left( \frac{A_p}{A_t} \right)^2 \right],$$

$$P_{wy} - P_g = - P_{Be} \left[ 1 - \left( \frac{A_p}{A_g} \right)^2 \right] = - P_{Be} (1 - N^2),$$
**6.12**

gdzie  $N = A_p/A_g$ .

W rzeczywistości, zmiana powierzchni przekroju otworu głośni jest na tyle gwałtowne, że u jej wylotu przepływ staje się turbulentny, w wyniku czego powstają dodatkowe straty. Współczynnik ten wynosi 0.37 [VAN DEN BERG i in., 1957], stąd

$$P_s - P_{we} = P_{Be} (1 + 0.37) = 1.37 P_{Be}.$$
**6.13**

Podobnie jest w strefie, w obrębie której głośnia się rozszerza do wlotu toru głosowego.

Istnienie u wylotu głośni pierścieniowej warstwy przyściennej powietrza oraz strefy przepływu turbulentnego powoduje konieczność wprowadzenia dodatkowego czynnika strat ciśnienia  $\Delta p$ :

$$\Delta p = \frac{\rho U_g^2}{2} \left[ \frac{1}{A_p} - \frac{1}{A_g} \right].$$
**6.14**

Stąd

$$P_{wy} - P_g = \frac{\rho U_g^2}{2 A_p^2} (1 - N^2) - \frac{\rho U_g^2}{2 A_p^2} (1 - N)^2 =$$

$$= - P_{Be} \cdot 2N(1 - N).$$
**6.15**

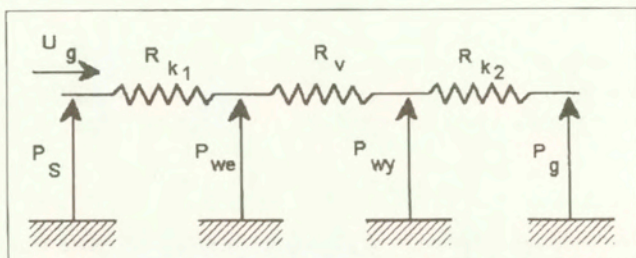
Wyrażenie (6.15) jest zgodne z tym, które zastosowali Ishizaka i Flanagan ([ISHIZAKA & FLANAGAN, 1972]) w modelu dwumasowym (patrz dalej). Wartość wyrażenia  $2N(1-N)$ , zależnie od fazy ruchu fałdów głosowych jest rzędu 0.05-0.4. Dla małych otworów głośni, można przyjąć, że

$$P_{wy} - P_g = -0.4 P_{Be}.$$
**6.16**

Przy dostatecznie małej powierzchni przekroju głośni można przyjąć, że przepływ powietrza w jej obrębie jest laminarny i rezystancja  $R_v$  reprezentuje straty wywołane tarciami lepkościowym powietrza o ścianki głośni. Zgodnie z równaniem ruchu dla przepływów laminarnych, spadek ciśnienia jest proporcjonalny do przebytej

drogi  $d$ , prędkości objętościowej  $U_g$  i odwrotnie proporcjonalnej do trzeciej potęgi powierzchni przekroju przepływu. Tak więc

$$P_{we} - P_{wy} = 12\mu \cdot d \cdot l_g^2 \frac{U_g}{A_p^3}, \quad 6.17$$



Rys. 6.6 Elektryczny model równoważny źródła krtaniowego

Wartości poszczególnych oporności określone są następująco (Rys. 6.6):

$$R_{k1} = 1.37 \frac{\rho}{2} \frac{|U_g|}{A_p^2},$$

$$R_{k2} = -\frac{\rho}{2} \frac{|U_g|}{A_p^2} 2N(1-N), \quad 6.18$$

$$R_v = 12\mu \cdot d \frac{l_g^2}{A_p^3}.$$

Wówczas pełna oporność akustyczna głośni w stanie statycznym wynosi

$$R_g = R_v + R_{k1} + R_{k2}. \quad 6.19$$

W celu opisanego działania głośni w stanie dynamicznym należy uwzględnić szereg czynników, zwłaszcza te, które opisują zachowanie się toru głosowego, w którym rozchodzi się fala akustyczna, a także wpływ głośni na jego własności transmisyjne. W pierwszym przybliżeniu, można zamodelować go w postaci bezstratnej rury o stałym przekroju  $A_1 = 5 \text{ cm}^2$  i długości  $l = 17 \text{ cm}$ , oraz założyć, że rozchodzi się w nim fala płaska. Z jednej strony jest on obciążony impedancją  $Z_r$  promieniowania otworu ustnego, zamodelowanego do postaci tła kołowego o powierzchni  $A_2$ ,



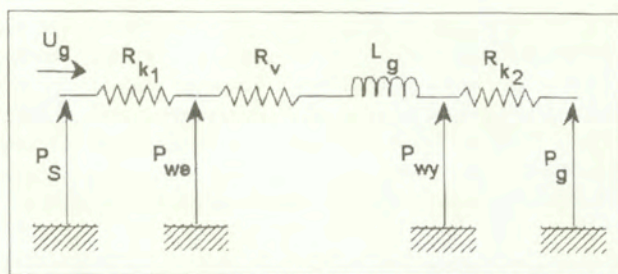
drgającego w nieskończonej płaskiej odgradzie. Natomiast impedancja wejściowa toru obciążająca źródło kraniowe może być wyliczona ze wzoru

$$Z_t = Z_0 \frac{Z_r \cdot \cos \beta l + j Z_0 \cdot \sin \beta l}{Z_0 \cdot \cos \beta l + j Z_r \cdot \sin \beta l} \quad 6.20$$

w którym  $Z_0 = \rho c / A_t$  jest impedancją falową rury,  $\beta = \omega / c = 2\pi / \lambda$  -przesuwnością falową, zaś  $Z_r$  - impedancją promieniowania tłoka kołowego o promieniu  $r$  i powierzchni  $A_t = \pi r^2$  (patrz Rys. 5.10 oraz wzór 5.18)

Analogicznie można przyjąć, że głośnia widziana od strony toru głosowego zachowuje się podobnie jak otwór wylotowy uformowany przez usta, to znaczy również można ją zamodelować za pomocą tłoka o powierzchni kołowej, drgającego w nieskończonej odgradzie (powierzchnia przekroju krtani jest znacznie mniejsza od powierzchni wejściowej toru głosowego). Impedancja promieniowania głośni składa się również z dwóch równoległe połączonych ze sobą elementów  $R_g$  oraz  $L_g$ , i równoległe dołączonych do oporności wejściowej toru głosowego. Jednakże z uwagi na duży stosunek powierzchni wejściowej toru głosowego do powierzchni głośni (ta ostatnia przeciętnie wynosi ok.  $5 \text{ mm}^2$ )  $Z_t \ll Z_g$  i źródło kraniowe można w pierwszym przybliżeniu traktować jako źródło prądowe o stałej wydajności prędkości objętościowej, niezależnej od chwilowej konfiguracji kanału głosowego.

Uwzględniając masę współdrgającego powietrza w głośni jako element o charakterze bezwładnościowym, ostateczny jej model ma postać przedstawioną na Rys. 6.7.



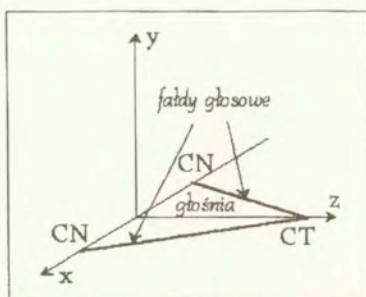
Rys. 6.7 Elektryczny model zastępczy głośni po uwzględnieniu drgającej w niej masy powietrza  $L_g = \rho \cdot d / A_p$ .

### 6.5 Opis ruchu fałdów głosowych.

Na Rys.6.8 przedstawiony jest schematycznie układ fałdów głosowych w przestrzeni trójwymiarowej. Zaznaczono na nim punkty zamocowania fałdów

głosowych w krtani. Dowolny punkt znajdujący się na wewnętrznej stronie fałdu głosowego wykonuje (o czym była już wcześniej mowa) złożony ruch wskutek oddziaływania wielu sił względem poszczególnych wyróżnionych osi:

- a) ruch o charakterze periodycznym wzdłuż osi  $x$ , który jest zasadniczy dla modelowania źródła krtaniowego,
- b) ruch względem osi  $y$  w wyniku oddziaływania różnicy ciśnień istniejącej między obszarem pod- i nad-głośniaowym (ruch ten jest również o charakterze periodycznym), lecz jego amplituda jest stosunkowo niewielka.
- c) ruch względem osi  $z$ , który jest opisywany w modelach bardziej złożonych, tzw. wielomasowych, z których najbardziej popularny jest model dwumasowy (patrz punkt 6.6).



**Rys. 6.8** Schemat zamocowania fałdów głosowych w krtani. CT – chrząstka tarczowa, CN – chrząstki nalewkowe. Kierunek przepływu powietrza jest wzdłuż osi  $y$ .

W pierwszym przybliżeniu, masa fałdów głosowych może być traktowana jako jedna całość poruszająca się wzdłuż osi  $x$ . W stanie dynamicznym ruch ich może być opisany za pomocą równania różniczkowego drugiego rzędu (wzór 6.1). Zastępcza masa drgających fałdów głosowych  $m$  była w literaturze różnie oceniana, od 0.15 g [ISHIZAKA, 1966] do 0.24 g [FLANAGAN&LANDGRAF, 1968]. Na ogół przyjmuje się obecnie, że wynosi ona ok. 0.18 g [KOIZUMI i in., 1987]. Współczynnik  $r$  w równaniu (6.1) reprezentuje tłumienność systemu, która pochodzi od :

- bezwładności związanej z masą błony śluzówkowej fałdów głosowych,
- ich sztywności ( $1/k$ ),
- tarcia lepkościowego błony śluzówkowej, tkanki łącznej i mięśniowej fałdów głosowych, oraz
- zjawiska przyczepności (lepkości) wzajemnej fałdów głosowych cyklicznie stykających się ze sobą przez pewien czas.

Współczynnik ten określony jest wzorem [FLANAGAN, 1972]:

$$r = \frac{\xi}{\sqrt{mk}}, \quad 6.21$$

gdzie  $\xi$  jest lepkościowym współczynnikiem tłumienia drgań.

Sprężystość  $s(x)$  reprezentuje nieliniarną sztywność przesunięciową, modelowaną przez sztywność zawieszenia masy  $m$ , przeciwstawiającej się chwilowym przemieszczeniom się fałdu z położenia spoczynkowego  $x_0$  oraz odkształceniom lepkosprężystym w fazie zwarcia obu fałdów, gdy stykają się ze sobą. Ruch w tej fazie nie ustaje momentalnie, bowiem masy fałdów przesuwały się nieco ku sobie, ulegając chwilowej, odwracalnej deformacji.

Sprężysta siła powrotna  $F_s$  jest zatem sumą dwóch składników [ISHIZAKA & KANEKO; 1968]:

a) wywołanego sprężystością przesunięciową

$$f_s = kx + \eta_k kx^3, \quad 6.22$$

w którym  $\eta_k$  jest współczynnikiem nielinierności sprężystości przesunięciowej  $k$ ;

oraz

b) wywołanego sprężystością odkształceniową, określonego następująco

$$f_h = h(x - x_0) + \eta_h h(x - x_0)^3, \quad 6.23$$

w którym  $\eta_h$  jest współczynnikiem nielinierności sprężystości odkształceniowej  $h$ , zaś  $x_0 = -A_0/2l_g$  oznacza odległość między płaszczyzną neutralną (spoczynkową) krawędzi fałdów głosowych i płaszczyzną ich styku.  $A_0$  jest powierzchnią spoczynkową otworu głośni na długości odcinka  $d$ , odpowiadającego zastępczej masie  $m$ . Efekt sprężystości odkształceniowej występuje oczywiście tylko w zakresie dużych przesunięć masy  $m$ , większych od wartości granicznej  $x_0$ , gdy jest spełniony warunek graniczny

$$x \leq x_0 = -\frac{A_0}{2l_g}. \quad 6.24$$

W celu określenia zależności siły  $F_m$  (pochodzenia aerodynamicznego) od prędkości objętościowej przepływu i powierzchni szpary głośni należy przyjąć następujące założenia:

- a) na drodze  $l_1$  zwężenia tchawicy u wlotu do głośni ciśnienie wejściowe  $P_{we}$  równa się ciśnieniu Bernoulliego, zwiększonemu o współczynnik ustalony na drodze empirycznej [VAN DEN BERG i inni, 1957]  $\approx 0.37$  (jest to wynik zwężenia się przekroju na drodze przepływu powietrza, o czym już wcześniej była mowa);
- b) spadek ciśnienia na długości głośni  $d$  pochodzi od laminarnej rezystancji przepływu określonej ogólnym wzorem na rezystancję akustyczną głośni  $R_g$  (wzór 6.19),
- c) różnica ciśnień między głośnią i wlotem do toru głosowego, na jego rozszerzającym się odcinku ma wartość ujemną i jest określona wzorem 6.16.

Wówczas składową statyczną  $F_m$  można obliczyć z zależności

$$F_m = P_m \cdot l_g \cdot d, \quad 6.25$$

przy założeniu, że ciśnienie w środku głośni  $p_m$  wynosi

$$P_m = \frac{P_{we} + P_{wy}}{2}. \quad 6.26$$

Ciśnienie u wlotu głośni  $p_{we}$  wynosi

$$P_{we} = P_s - 1.37 \frac{\rho U_g^2}{2 A_p^2}, \quad 6.27$$

zaś ciśnienie u wylotu  $p_{wy}$

$$P_{wy} = P_{we} - 12\mu \frac{dl_g^2 U_g}{A_p^3}. \quad 6.28$$

Stąd średnie ciśnienie w głośni

$$P_m = P_{we} - 6\mu \frac{dl_g^2 U_g}{A_p^3}. \quad 6.29$$

W celu obliczenia prędkości objętościowej  $U_g$  w funkcji powierzchni przekroju  $A_p$  głośni należy rozwiązać poniższe równanie

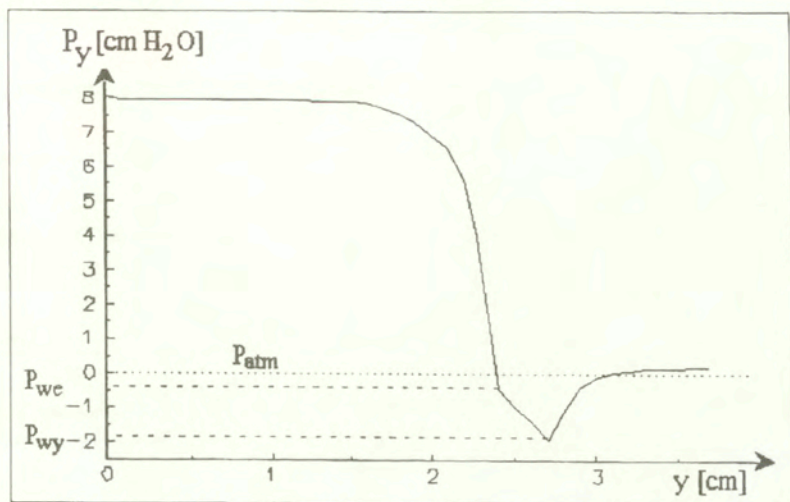
$$\Delta P = P_s - P_g = 1.37 \frac{\rho U_g^2}{2 A_p^2} + 12\mu \frac{dl_g^2 U_g}{A_p^3} - 2N(1-N) \frac{\rho U_g^2}{2 A_p^2}, \quad 6.30$$

lub o postaci

$$[1.37 - 2N(1 - N)] \frac{\rho U_g^2}{2 A_p^2} + 12\mu \frac{dU_g}{A_p} - \Delta P = 0 .$$

6.31

Zakładając, że prędkość liniowa przepływu powietrza w głośni jest znacznie mniejsza od prędkości ruchu fałdów oraz uwzględniając fakt, że wszystkie wymiary głośni są mniejsze od długości fali rozchodzącej się w torze artykulacyjnym, można w oparciu o wzory 6.27, 6.28 i 6.29 wyznaczyć rozkłady spadku ciśnienia na drodze przepływu powietrza od tchawicy aż do wlotu toru głosowego (Rys. 6.9).



Rys. 6.9 Przebieg spadku ciśnienia powietrza między tchawicą i gardłem.

W modelu funkcjonowania źródła krtaniowego należy uwzględnić dodatkowo wzajemne oddziaływanie w płaszczyźnie akustycznej źródło-tor głosowy. Jak już wspomniano w poprzednim rozdziale, w najprostszym ujęciu tor głosowy może być aproksymowany za pomocą kaskadowo ze sobą połączonych 17-u odcinków cylindrycznych o jednakowej długości  $l$  i zmiennej funkcji powierzchni przekroju  $A(x)$ . Układ ten może być opisany za pomocą macierzy

$$\|Z\| \times \|\bar{U}\| = \|\bar{P}\| ,$$

gdzie  $\bar{U}$  jest modułem funkcji prędkości objętościowej w torze głosowym

$$U(x, t) = \bar{U}(x, t) e^{j\omega t} ,$$

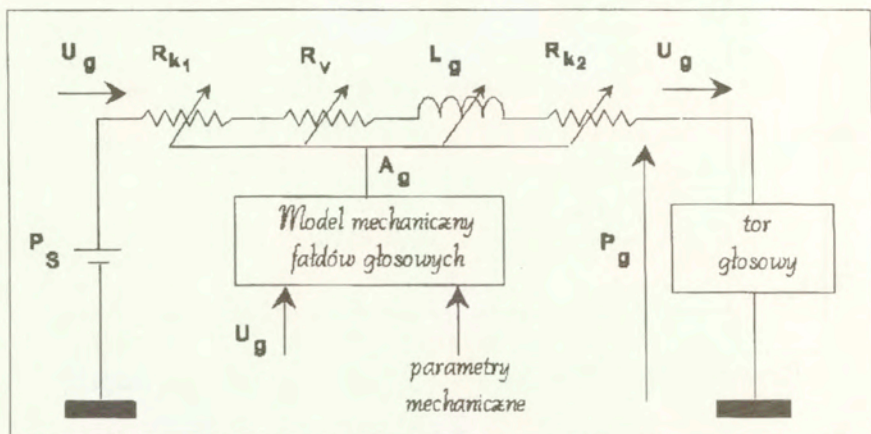
zaś  $\bar{P}$  jest modułem funkcji ciśnienia

$$P(x, t) = \bar{P}(x, t) e^{j\omega t}.$$

Dla każdej określonej konfiguracji toru głosowego odpowiadającej na przykład samogłoskom polskim można wyliczyć impedancję wejściową toru głosowego widzianą od strony źródła krztaniowego, która może być przedstawiona w uogólnionej postaci

$$Z_{we} = R_{we} + jX_{we}(\omega).$$

Tak więc model źródła krztaniowego obciążonego impedancją wejściową toru głosowego może mieć postać przedstawioną na Rys. 6.10.



Rys. 6.10 Pełny model jednomasowy źródła krztaniowego.

### 6.6 Dwumasowy model źródła krztaniowego

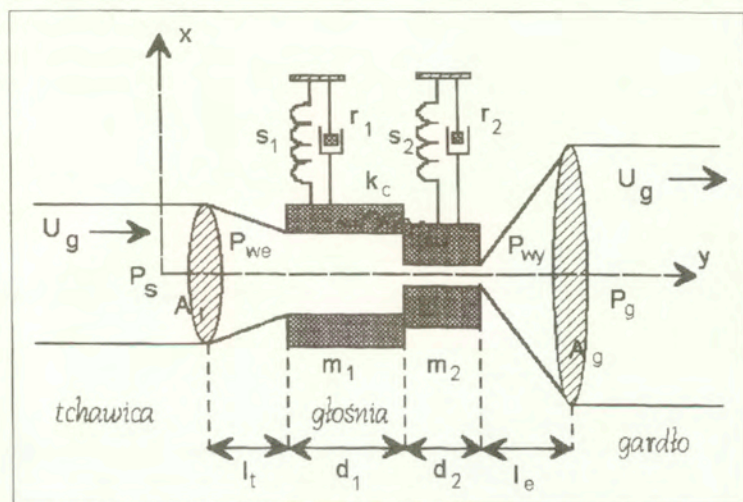
W istocie swej model jednomasowy jest dość grubym przybliżeniem układu rzeczywistego i choć umożliwia wyprowadzenie pewnych zasadniczych zależności aeroakustycznych, to jednak nie uwzględnia bardzo istotnych cech strukturalnych fałdów głosowych dotyczących zarówno ich kinetyki, jak ich odkształcalności postaciowej, o której była mowa na wstępie rozdziału. Przy dokładniejszej analizie drgań fałdów, można wyróżnić w nich przynajmniej dwie części, dolną i górną, wykonujące fazowo przesunięte względem siebie ruchy, przy czym dla głosu normalnego przyjmuje się obustronną zgodność faz ruchów odpowiednich części

fałdów (co nie ma często miejsca przy patologicznych zmianach głośni, takich jak na przykład przy częściowych jedno- lub obustronnych porażeniach fałdów). Równocześnie zakłada się dopuszczalność ruchów poszczególnych części fałdów, zarówno wzdłuż osi poprzecznej ( $x$ ) głośni, jak i prostopadłej, względem której odbywa się cykliczne zamykanie oraz otwieranie głośni. Ogólny schemat modelu dwumasowego, w którym każdy fałd jest przedstawiony za pomocą dwóch mas sprężyste ze sobą połączonych przedstawiono na Rys. 6.11. Głębokość szczeliny głośni jest podzielona na dwa odcinki  $d_1$  i  $d_2$  odpowiadające masom  $m_1$  i  $m_2$ , reprezentującym dolną i górną część fałdu głosowego, przy czym powierzchnie otworu głośni na długościach odcinków  $d_1$  i  $d_2$  są określone wzorami:

$$A_{g1} = A_{g,1} + 2 x_1(t) l_g$$

$$A_{g2} = A_{g,2} + 2 x_2(t) l_g$$
6.32

We wzorach tych  $A_{g01}$  i  $A_{g02}$  oznaczają powierzchnie średnie dla położenia, wokół których odpowiednie części fałdów wykonują boczne, w pełni symetryczne ruchy drgające.



Rys. 6.11 Dwumasowy model źródła krztaniowego.

W modelu obie masy fałdów połączone są sprężystością sprzęgającą  $K_c$ , która reprezentuje sztywność fałdów głosowych na zginanie w płaszczyźnie prostopadłej do kierunku ich drgań. Sprężystość ta jest parametrem silnie nieliniowym i zmienia się odpowiednio do zmian grubości i sztywności samych fałdów w wyniku oddziaływania

na nie mięśnia tarczowo-nalewkowego, który napina je podczas fonacji. Siła zwracająca  $f_c$  pochodząca od tej sprężystości jest określona wzorem

$$f_c = 2s_c(x_1 - x_2)^3(d_1 + d_2)^2, \quad 6.33$$

Przy założeniu liniowości odkształceń (czyli przy małych amplitudach drgań),

$$f_c \approx K_c(x_1 - x_2). \quad 6.34$$

Dla normalnych warunków fizjologicznych sprężystość sprzęgająca  $K_c$  zmienia się w szerokich granicach, wokół średniej rzędu 0.25 N/m (ISHIZAKA & FLANAGAN, 1972). Jednak w miarę wzrostu wartości sprężystości sprzęgającej ( $K_c \rightarrow A$ ) model dwumasowy stopniowo upodabnia się do modelu jednomasowego. Ma to miejsce, na przykład, w przypadku nadmiernego wzrostu napięcia fałdów głosowych.

Przedstawione na Rys. 6.11 elementy  $s_1$  i  $s_2$  reprezentują nieliniarne sztywności obu części fałdów głosowych wynikające z oddziaływania na nie przede wszystkim mięśni głosowych (znajdujących się wewnątrz fałdów). Przeciwstawiają się one chwilowym przesunięciom  $x_1(t)$  oraz  $x_2(t)$ , a jednocześnie odkształceniom lepkosprężystym w fazie zwarcia, gdy fałdy częściowo stykają się wzajemnie, ulegając chwilowej i odwracalnej deformacji. Tak więc sprężystą siłą powrotną można przedstawić za pomocą sumy dwóch składników

$$f_j = f_{h_j} + f_{h_j}, \quad 6.35$$

gdzie  $j=1,2$ .

Składnik  $f_{h_j}$  oznacza siłę powrotną, wywołaną sprężystością przesunięciową  $k_j$ , i jest określony wzorem

$$f_{h_j} = k_j x_j (1 + \eta_{k_j} x_j^2), \quad 6.36$$

gdzie  $j=1,2$ , zaś  $h_{kj}$  jest współczynnikiem nieliniowości sprężystości przesunięciowej  $k_j$ . Składnik  $f_{h_j}$  oznacza siłę pochodzącą od sprężystości odkształceniowej  $h_j$

$$f_{h_j} = h_j (x_j - x_{o_j}) [1 + \eta_{h_j} (x_j - x_{o_j})^2], \quad 6.37$$

w którym  $h_{hj}$  jest współczynnikiem nieliniowości sprężystości odkształceniowej  $h_j$ , a  $x_{o_j} = -A_{goj}/2l_g$  oznacza odległość między płaszczyzną równowagi w "położeniu fonacyjnym" fałdów głosowych i płaszczyzną ich styku. Tylko w momencie styku fałdów głosowych następuje efekt sprężystości odkształceniowej. Zależność sił



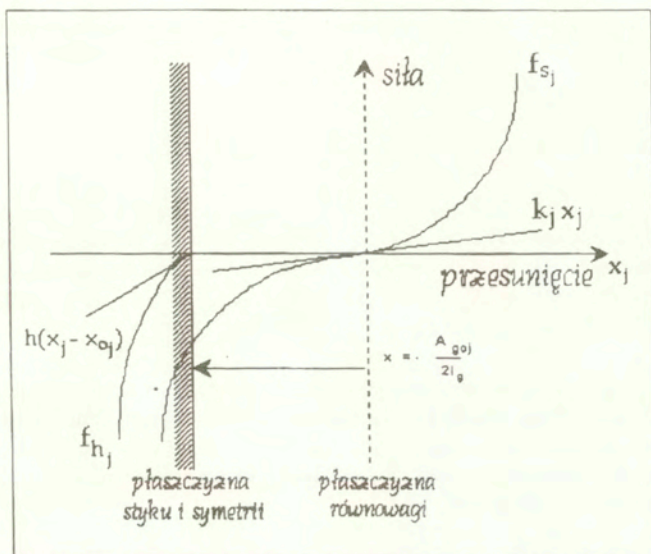
powrotnych, wywołanych sprężystością przesunięciową i odkształceniową, od chwilowych przesunięć fałdów głosowych z położenia równowagi przedstawiono na Rys. 6.12.

Przeciętne wartości parametrów sprężystościowych układu drgającego fałdów głosowych w normalnych warunkach fizjologicznych są następujące [ISHIZAKA&FLANAGAN,1972]:

$$k_1 = 50-100 \text{ N/m}, h_1 \gg 3k_1,$$

$$k_2 = 5-50 \text{ N/m}, h_2 = 3k_2,$$

$$h_{k1}, h_{k2} = 50-100, h_{h1}, h_{h2} = 250-500.$$



Rys. 6.12 Zależność sił powrotnych, wywołanych sprężystością przesunięciową  $k_j$  i odkształceniową  $h_j$ , od chwilowych przesunięć fałdów z położenia spoczynkowego (wg [ISHIZAKA&FLANAGAN,1972]).

Równania ruchu górnej i dolnej części fałdu mają postać podobną jak we wzorze (6.1), przy czym uwzględniona jest dodatkowa siła powstała wskutek wzajemnego przemieszczenia się górnej masy względem dolnej:

$$m_1 \frac{d^2 x_1}{dt^2} + r_1 \frac{dx_1}{dt} + s_1 x_1 + s_c (x_1 - x_2) = F_{m1}$$

6.38

$$m_2 \frac{d^2 x_2}{dt^2} + r_2 \frac{dx_2}{dt} + s_2 x_2 + s_c (x_2 - x_1) = F_{m2}$$

Rezystancja strat  $r_1$  i  $r_2$  pochodzi od:

1) tarcia lepkościowego tkanki śluzówkowej, łącznej i mięśniowej fałdów głosowych,

2) zjawiska przyczepności wzajemnej (lepkości) ścianek fałdów głosowych stykających się ze sobą w fazie zwarcia.

Wpływ oporności strat wygodnie jest przedstawić przy pomocy odpowiednich współczynników tłumienia drgań  $x_1$  i  $x_2$ , dla górnej i dolnej części fałdów określonych, zgodnie z poniższymi wzorami

$$r_1 = 2\xi_1 \sqrt{m_1 k_1}, \quad r_2 = 2\xi_2 \sqrt{m_2 k_2},$$

6.39

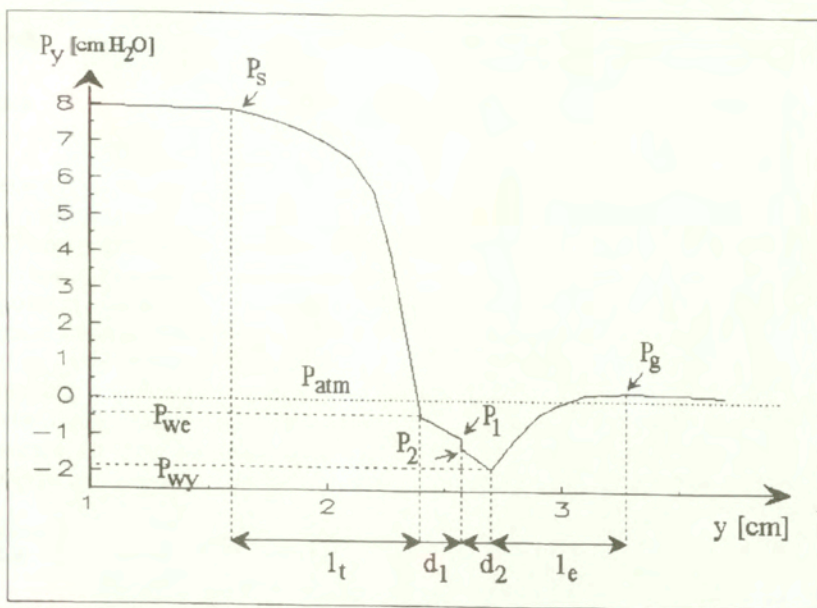
w których  $k_1$  i  $k_2$  są liniowymi współczynnikami sprężystości przesunięciowej występującymi we wzorze (6.36). Dla fazy rozwarcia fałdów współczynniki tłumienia wynoszą dla normalnych warunków fonacji odpowiednio  $x_1=0.1$  i  $x_2=0.6$ , natomiast w fazie zwarcia, zakładając tłumienie krytyczne  $x_1=1.1$  i  $x_2=1.6$  [ISHIZAKA & FLANAGAN, 1972].

Podobnie jak w przypadku modelu jednomasowego, spadek ciśnienia wzdłuż głośni (zgodnie z kierunkiem przepływu powietrza) jest również liniowy, z tą jednak różnicą, że na granicy dolnej i górnej części fałdu, tam gdzie występuje połączenie obu mas  $m_1$  i  $m_2$ , następuje stosunkowo niewielkie, dodatkowe, skokowe zmniejszenie ciśnienia o wartość określoną wzorem

$$\Delta p = \frac{1}{2} \rho (u_{g1} - u_{g2}) = \frac{1}{2} \rho U_g^2 (A_{g2}^2 - A_{g1}^2)$$

6.40

Spadek ciśnienia wywołany skokową zmianą przekroju głośni jest równoważny różnicy gęstości energii kinetycznej przepływu ośrodka, wywołanej zmianą prędkości liniowej od  $u_{g1}$  do  $u_{g2}$  wskutek zmiany przekroju przepływu z  $A_{g1}$  na  $A_{g2}$ , przy założeniu ciągłości strugi  $U_g = \text{const}$ . Stąd, rozkład ciśnienia od wlotu aż do wylotu głośni ma przebieg nieco odmienny od przedstawionego na Rys. 6.9 dla modelu jednomasowego. Na rozkładzie ciśnienia wyznaczonym dla modelu 2-masowego widoczny jest wspomniany skok ciśnienia (Rys. 6.13).



Rys. 6.13 Rozkład ciśnienia w głośni między tchawicą i gardłem wyznaczony w oparciu o model dwumasowy źródła krtańowego (wg [ISHIZAKA&FLANAGAN, 1972]).

Podobnie jak poprzednio, uwzględnia się fakt, że rozmiary głośni, tj. szczeliny między fałdami głosowymi są znacznie mniejsze od długości fali  $l$ , a prędkość liniowa  $u_g$  powietrza w głośni jest znacznie mniejsza od prędkości drgań wiązadeł głosowych  $v = dx/dt$ , zatem rozkład spadku ciśnienia w głośni można wyznaczyć w pierwszym przybliżeniu na podstawie praw przepływu quasi-stacjonarnego i jednowymiarowego, to jest w kierunku osi  $y$  (Rys. 6.11).

Kolejne spadki ciśnienia na drodze przepływu powietrza są określone wzorami

$$P_s - P_{we} = 0.69 \rho (U_s^2 A_{g1}^2) + \int_0^{l_t} \frac{\rho}{A_t(x)} dx \frac{d}{dt} U_s, \quad 6.41$$

$$P_{we} - P_1 = 12 \mu d_1 l_g^2 A_{g1}^3 U_s + \frac{\rho d_1}{A_{g1}} \frac{d}{dt} U_s, \quad 6.42$$

$$P_1 - P_2 = \frac{1}{2} \rho U_s^2 (A_{g2}^2 - A_{g1}^2), \quad 6.43$$

$$P_2 - P_{wy} = 12\mu d_2 l_g^2 A_{g2}^3 U_g + \frac{\rho d_2}{A_{g2}} \frac{d}{dt} U_g, \quad 6.44$$

$$P_{wy} - P_g = -\rho U_g^2 A_{g2}^2 \frac{A_{g2}}{A_g} \left(1 - \frac{A_{g2}}{A_g}\right). \quad 6.45$$

Podobnie jak poprzednio, przyjęto że w odcinku zężenia się tchawicy  $l$ , spadek ciśnienia u wlotu głośni równa się ciśnieniu Bernoulliego, zwiększonemu o wspomniany przy modelu jednomasowym współczynnik empirycznym  $\omega \approx 0.37$ . Spadki ciśnień na odcinkach  $d_1$  i  $d_2$  pochodzą od laminarnych rezystancji przepływu określonych wzorem 6.19, natomiast spadek ciśnienia na skokowej zmianie przekroju głośni, tj. w miejscu styku mas  $m_1$  i  $m_2$ , jest równoważny różnicy gęstości energii kinetycznej przepływu powietrza, wywołanej zmianą prędkości liniowej od  $u_{g1}$  do  $u_{g2}$  wskutek zmiany powierzchni przekroju od  $A_{g1}$  do  $A_{g2}$  przy jednoczesnym założeniu ciągłości prędkości objętościowej  $U_g$ . Spadek ciśnienia w miejscu wylotu głośni, gdzie następuje szybkie powiększenie powierzchni przekroju, ciśnienie dąży do wartości ciśnienia atmosferycznego. Spadek ten jest obliczony na podstawie praw przepływu Newtona

$$f = \frac{d}{dt} mv. \quad 6.46$$

Zgodnie z założeniem ciągłości strugi  $U_g$

$$\rho U_g (u_g - u_{wy}) = A_g (P_{wy} - P_g) \quad 6.47$$

lub

$$(P_g - P_{wy}) = \frac{1}{2} \rho u_{g2}^2 [2N(1-N)] = \frac{1}{2} \rho - \frac{U_g}{A_g} 2 - [2N(1-N)] = P_{B2} [2N(1-N)], \quad 6.48$$

gdzie podobnie jak dla modelu jednomasowego  $N = A_{g2}/A_g$ ,  $P_{B2}$  – jest ciśnieniem Bernoulliego przy wlocie do toru głosowego oraz  $A_g$  powierzchnią przekroju wlotu do toru głosowego. Wartość składnika  $2N(1-N)$  zmienia się w zakresie od 0.05 do 0.40 i wpływa na wzajemne oddziaływanie między źródłem krztaniowym i torem głosowym.

We wzorach uwzględniono nie tylko statyczne składniki przepływu powietrza, ale również dynamiczne, związane z bezwładnością masy akustycznej powietrza drgającego w odpowiednich przekrojach głośni. Efekt ten wyraża się dodatkowymi składnikami typu  $(pd/A)(dU_g/dt)$  we wzorach 6.41, 6.42 i 6.44.

Na podstawie powyższych równań i układu analogii elektro-akustycznych można utworzyć zastępczy schemat elektryczny głośni, w którym prędkość objętościowa  $U_g$  jest reprezentowana przez natężenie prądu, zaś spadki ciśnień przez odpowiednie spadki

napięć. Równoważny układ głośni jest przedstawiony na Rys. 6.14, zaś poszczególne jego elementy określone są poniższymi wzorami

$$\begin{aligned}
 R_l &= 1.37 \frac{\rho}{2} \frac{|U_g|}{A_{g1}^2}, & L_l &= \int_0^l \frac{dx}{A_l(x)}, \\
 R_{v1} &= 12 \frac{\mu l_g^2 d_1}{A_{g1}^3}, & L_{g1} &= \frac{\rho d_1}{A_{g1}}, \\
 R_{l2} &= \frac{1}{2} \rho (A_{g2}^2 - A_{g1}^2) |U_g|, \\
 R_{v2} &= 12 \frac{\mu l_g^2 d_2}{A_{g2}^3}, & L_{g2} &= \frac{\rho d_2}{A_{g2}}, \\
 R_g &= - \frac{\rho}{2} \frac{2}{A_{g2} A_g} \left[ 1 - \frac{A_{g2}}{A_g} \right] |U_g|.
 \end{aligned}
 \tag{6.49}$$

Wówczas impedancja akustyczna głośni w modelu dwumasowym jest określona wzorem

$$Z_g = (R_{k1} + R_{k2}) |U_g| + (R_{v1} + R_{v2}) + j\omega(L_{g1} + L_{g2} + L_l),
 \tag{6.50}$$

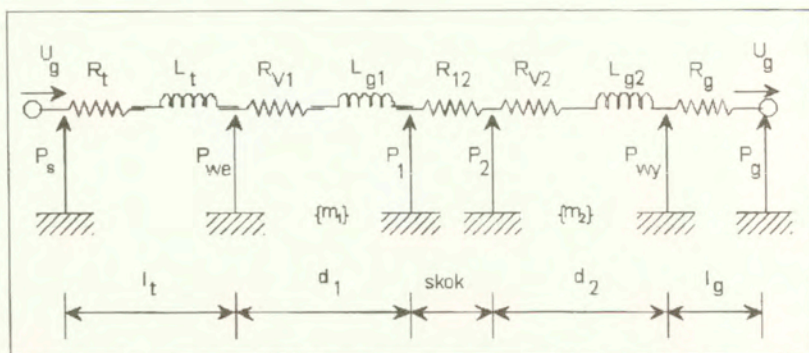
gdzie

$$R_{k1} = \frac{0.19\rho}{A_{g1}^2}, \quad R_{k2} = \frac{\rho \left[ 0.5 - \frac{A_{g2}}{A_g} \left( 1 - \frac{A_{g2}}{A_g} \right) \right]}{A_{g2}^2}.$$

Na ogół  $L_l$  jest pomijalnie małe w stosunku do sumy  $L_{g1} + L_{g2}$  i powyższy wzór na impedancję akustyczną głośni  $Z_g$  daje zgodny wynik dla modelu jednomasowego (patrz wzór 6.18), gdy podstawimy  $A_{g1} = A_{g2} = A_p$

Wszystkie parametry układu są funkcjami czasu i są zależne od powierzchni przekrojów  $A_1$  i  $A_2$  utworzonymi między odpowiednimi masami  $m_1$  i  $m_2$  obu fałdów. Powierzchnie te wyznaczone z ruchu drgającego obu par mas fałdów, który z kolei jest funkcją przede wszystkim ciśnienia podgłośniowego  $P_s$ , współczynnika napięcia fałdów głosowych  $Q$ , powierzchni spoczynkowej przekroju głośni  $A_{g0}$  oraz prędkości objętościowej w wylotu głośni  $U_g$ .

Na Rys. 6.14 przedstawiono układ równoważny głośni wraz z wszystkimi elementami opisanymi powyższymi wzorami.



**Rys. 6.14** Układ równoważny głośni w systemie analogii akusto-elektrycznych w dwumasowym modelu źródła krztaniowego (wg [ISHIZAKA&FLANAGAN,1972]).

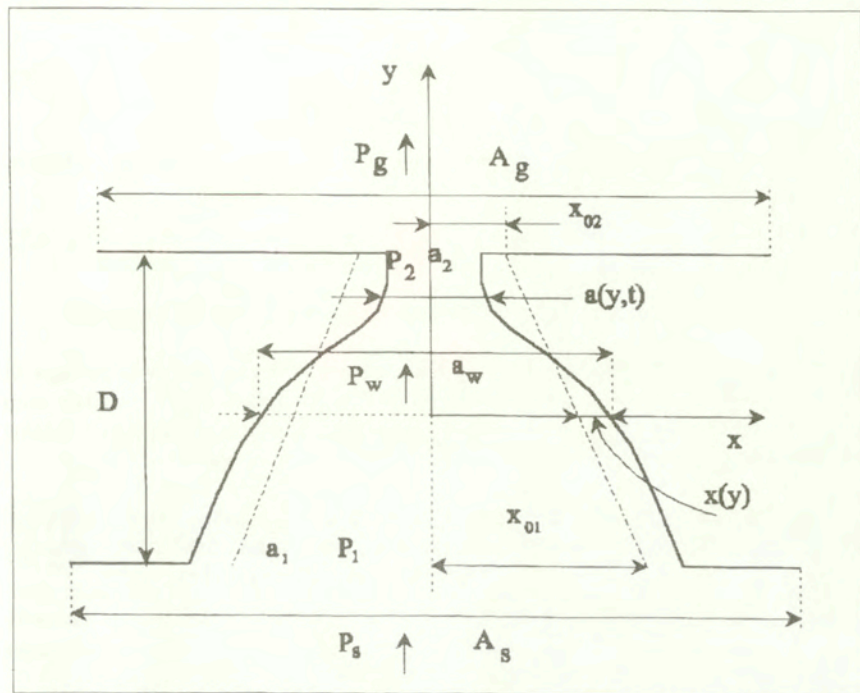
Impedancje szeregowo usytuowane w górnym ramieniu czwornika są określone przez akustyczne inertancje oraz nieliniowe stratności w otworze głośni między fałdami głosowymi, zależnymi od parametrów przepływu. Przebieg spadku ciśnienia w głośni wyliczony dla modelu dwumasowego jest bardzo podobny do spadku wyliczonego dla modelu jednomasowego. Różnica polega na pojawieniu się na granicy między dolną i górną częścią fałdu dodatkowego skokowego spadku ciśnienia zależnego od strat o charakterze lepkościowym. Choć oba modele wyjaśniają wiele zjawisk zachodzących w głośni, tym nie mniej nie uwzględniają wiele zjawisk nieliniowych związanych przede wszystkim z własnościami mechanicznymi fałdów głosowych modulujących przepływ strumienia powietrza.

### 6.7 Nieliniowy model źródła krztaniowego dla małej amplitudy drgań

Z badań eksperymentalnych [ITTZE,1988] wynika, że w fałdach głosowych, a dokładniej w tkance łącznej i nabłonkowej rozchodzi się ku górze fala powierzchniowa, w wyniku czego zmienia się kształt i masa górnej części fałdu. Bezpośrednie uwzględnienie faktu rozchodzenia się fali powierzchniowej w fałdzie prowadzi do dużej złożoności obliczeń. Wprowadzenie dodatkowych stopni swobody w modelu źródła krztaniowego umożliwia zarówno uwzględnienie rozchodzenia się fali powierzchniowej, jak i przesunięcie fazowe ruchu dolnej i górnej części fałdu głosowego.

6.7.1 Model uwzględniający rozchodzenie się w fałdach głosowych fal powierzchniowych

Na Rys.6.15 [TITZE,1988] przedstawiono schemat głośni uwzględniający nieliniową geometrię fałdów głosowych widzianych w przekroju pionowym. W oparciu o ten schemat będą wyprowadzone zależności opisujące ruch fałdów i zmianę prędkości cząsteczkowej strugi przepływającego przez głośnię powietrza. Na rysunku wyróżniono odpowiednio powierzchnie przekrojów części pod- ( $A_s$ ) i ponad-głośniowej ( $A_g$ ), przekrojów części wlotowej ( $a_1$ ), środkowej (śródgłośni -  $a_w$ ) i wylotowej ( $a_2$ ) głośni.



Rys. 6.15 Schematyczne przedstawienie przekroju głośni wraz z określeniem ciśnień, przekrojów i wychyleń w wybranych miejscach ([TITZE,1988]).

Powierzchnie  $A_s$  i  $A_g$  są stałe, zaś pozostałe powierzchnie przekrojów są funkcjami czasu i położenia w głośni. Tak więc powierzchnia przekroju w obrębie głośni jest funkcją dwóch zmiennych

$$a(y,t) = 2l_g(x_0(y) + x_1(y,t)), \quad 6.51$$

gdzie

$y$  - jest odległością od środkowego punktu głośni wzdłuż linii poprowadzonej zgodnie z kierunkiem przepływu powietrza,  
 $x_0$  - jest odległością krawędzi wlotu głośni od jej osi symetrii  $y$ , zaś  
 $x_1(y,t)$  - jest funkcją czasową wychylenia się masy słuzówki pokrywającej fałd głosowy w zależności od jej położenia względem osi symetrii.

Hirano (HIRANO,1977) wykazał, że fale powierzchniowe rozchodzące się w tkance słuzówkowej fałdów są zgodne z kierunkiem przepływu powietrza. Tak więc w najprostszym przypadku (jednowymiarowym), fale te mogą być opisane za pomocą jednowymiarowego równania ruchu falowego o postaci

$$\frac{\partial^2 x_1}{\partial t^2} = c^2 \frac{\partial^2 x_1}{\partial y^2}, \quad 6.52$$

gdzie  $c$  jest prędkością propagacji fali. Ogólne rozwiązanie ma postać równania d'Alemberta

$$x_1(y,t) = x_1\left(t - \frac{y}{c}\right), \quad 6.53$$

co można łatwo zweryfikować przez podstawienie jego do poprzedniej zależności.

Z ostatniego równania wynika, że propagacja fali w tkance słuzówki pociąga za sobą opóźnienie ruchu górnej części fałdów względem ich dolnej. Dzięki właśnie temu opóźnieniu powstaje niezbędna dla wywołania ruchu drgającego fałdów niezrównoważenie układu.

Dla uproszczenia, załóżmy, że dla momentu bezpośrednio poprzedzającego proces fonacji, boki głośni są liniowo zbieżne, a sam jej przekrój ma kształt trapezowy (na Rys. 6.15 zaznaczono linią przerywaną). Tak więc funkcja  $x_0(y)$  ma postać

$$x_0(y) = \frac{1}{2}(x_{01} + x_{02}) - \frac{y}{D}(x_{01} - x_{02}), \quad 6.54$$

gdzie  $x_{01}, x_{02}$  są odległościami od osi symetrii  $y$ , odpowiednio dolnej i górnej krawędzi głośni, zaś  $D$  jest grubością fałdów głosowych.

W celu wyznaczenia przemieszczenia wywołanego rozchodzącą się falą powierzchniową ogólną postać rozwiązania d'Alemberta ruchu falowego można rozwinąć na szereg Taylora względem punktu  $y$ , znajdującego się w połowie drogi ruchu fali powierzchniowej, zachowując tylko składniki 0-ego i 1-ego rzędu



$$x_1(t - \frac{y}{c}) = x_1(t, 0) + \frac{y}{c} \frac{\partial x_1}{\partial t} \Big|_{t,0} + \dots \approx x - (\frac{y}{c} \frac{\partial x}{\partial t}), \quad 6.55$$

gdzie  $x$  oraz  $\delta x/\delta t$  są odpowiednio przemieszczeniem i prędkością fałdu głosowego określonymi w punkcie  $y$ .

Rozwinięcie to umożliwia uzyskanie bezpośredniej zależności przekroju poprzecznego w głośni  $a(y)$  od prędkości przepływu powietrza. W celu określenia zakresu ważności rozwinięcia przyjmijmy w wyrażeniu na  $x_1$  rozwiązanie o postaci harmoniczej  $\sin(t - y/c)$ . Rozwijając funkcję sinus różnicy do postaci

$$\sin(\omega t - \frac{\omega y}{c}) = (\sin \omega t) \cos \frac{\omega y}{c} - (\cos \omega t) \sin \frac{\omega y}{c} \quad 6.56$$

i przyjmując, że dla małych kątów

$$\sin \frac{\omega y}{c} \approx \frac{\omega y}{c}, \quad \cos \frac{\omega y}{c} \approx 1. \quad 6.57$$

Rozwiązanie zależności (6.55)

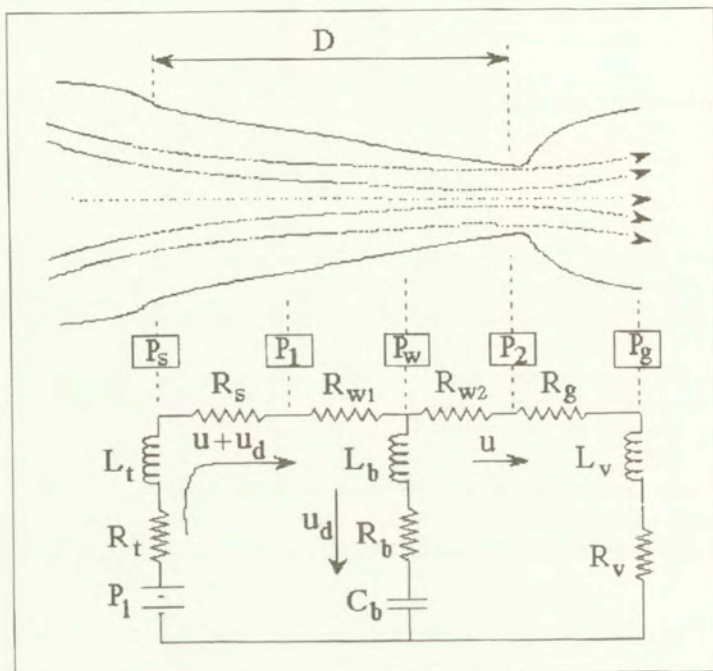
$$\sin(\omega t) - \frac{\omega y}{c} \cos \omega t. \quad 6.58$$

Aproksymacja dla małych kątów nie jest zbyt dobra dla celów numerycznych, ponieważ błędy przybliżeń dla  $(\omega y)/c > 0.75$  przekraczają 10%. Dla częstotliwości 100 Hz i prędkości fali powierzchniowej 1 m/s, aproksymacja ta będzie zbyt niedokładna dla grubości fałdu przekraczającej 2 mm (TITZE, 1988). Z drugiej strony, rozwinięcie tego typu od strony koncepcyjnej jest bardzo pouczające, a jednocześnie dogodne również od strony analitycznej. Żaden z dalej wyciąganych wniosków nie będzie obciążony w/w 10% błędem obliczeń numerycznych.

Dla tej postaci przybliżenia przybliżenia, przesunięcie fazy w ruchu dolnej i górnej krawędzi fałdów, dla  $D=2$  mm wynosi  $80^\circ$ . Powierzchnie przekroju głośni u jej wlotu i wylotu są określone dla  $y = \pm D/2$ . Wprowadzając opóźnienie  $t = D/2c$  otrzymuje się

$$\begin{aligned} a_1 &= 2l_g(x_{01} + x + \tau \frac{\partial x}{\partial t}) \\ a_2 &= 2l_g(x_{02} + x - \tau \frac{\partial x}{\partial t}) \end{aligned} \quad 6.59$$

W celu określenia prędkości objętościowej powietrza, średniego ciśnienia wewnątrz głośni, oraz wyznaczenia równania ruchu należy rozpatrzeć w nieco zmodyfikowanej postaci układ o stałych skupionych opisujący warunki przepływu przez głośnię. Na Rys. 6.16 przedstawiono schemat zastępczy głośni, który umożliwia określenie warunków ruchu błony słuzówkowej pokrywającej fałdy. Przedstawione na nim nieliniowe oporności przepływu w częściach zbieżnych i rozbieżnych głośni, są reprezentowane w układzie równoważnym odpowiednio przez elementy  $R_c$  i  $R_g$ , zaś elementy opornościowe wewnątrz głośni przez elementy  $R_{w1}$  i  $R_{w2}$ , które również są nieliniowe. Prędkość przepływu  $u$  wylotu głośni jest równa  $u$ , zaś  $u$  jej wejścia  $u+u_d$ , gdzie  $u_d$  jest prędkością objętościową powstałą wskutek ruchu oscylacyjnego ścian fałdów. Ponieważ w równaniach (6.59) wszystkie przemieszczenia są określone dla punktu znajdującego się w połowie grubości fałdów, mechaniczne  $w'$  zności fałdów mogą być również przedstawione w postaci skupionej za pomocą trzech elementów ( $C_f$  - podatności,  $L_b$  - bezwładności oraz oporności strat  $R_b$ ) i przypisane dla tego punktu. Elementy te jak wiadomo są bezpośrednio związane ze sztywnością, masą i tłumiennością drgającej słuzówki fałdów głosowych.



Rys. 6.16 Schemat zastępczy głośni uwzględniający nieliniowe warunki przepływu strugi powietrza przez głośnię [TITZE, 1988].

Ponieważ przemieszczenie względem osi  $y$  zostało przyjęte jako liniowa funkcja tej zmiennej (patrz równanie 6.55), średnia prędkość objętościowa przepływu może być przedstawiona dla punktu leżącego na osi  $x$  w postaci

$$u_c = 2l_g D \frac{\partial x}{\partial t}. \quad 6.60$$

Jednakże zmiany ciśnienia w obrębie głośni nie są liniowe. Średnie ciśnienie wewnątrz głośni  $P_w$  jest jedyną siłą pobudzającą tkankę śluzówkową okrywającą fałdy do ruchu i może być ono wyznaczone, jeżeli rozkład ciśnienia  $P(y)$  wewnątrz głośni jest znany. Jednak możliwe jest wyznaczenie jego z równania ruchu śluzówki

$$\begin{aligned} P_w &= \frac{I}{D} \int_{-\frac{D}{2}}^{+\frac{D}{2}} P(y) dy = 2l_{gd} L_b \left( \frac{\partial^2 x}{\partial t^2} + R_b \frac{\partial x}{\partial t} \right) \\ &= M \frac{\partial^2 x}{\partial t^2} + B \frac{\partial x}{\partial t} + Kx \end{aligned} \quad 6.61$$

gdzie  $M$ ,  $B$  i  $K$  są masą skuteczną, tłumieniem i podatnością wyrażonymi na jednostkę powierzchni drgającej tkanki śluzówkowej, wyznaczonymi dla układu o stałych skupionych, w punkcie położonym na początku osi  $x$  ( $y=0$ ).

Ponieważ funkcja  $P(y)$  nie jest liniowo zależna od współrzędnej  $y$ , nie można jej przyjąć więc jako równą średniej arytmetycznej z ciśnienia wejściowego i wyjściowego głośni ( $P_1$  i  $P_2$ ), jak również dzielić oporność całkowitą głośni na dwie równe części. Choć rysunek 6.16 nie przedstawia metody rozwiązywania oczkowego układu równań umożliwia przedstawienie pewnej koncepcji problemu analizy przepływu. Zależności wyznaczające spadek ciśnienia na głośni oraz ciśnienie pobudzające, określone w pracach [VAN DEN BERG i in., 1957; ISHIZAKA & MATSUDAIRA, 1972], w których przyjęto stacjonarny przepływ powietrza przez szparę głośni, również można zastosować przy analizie przepływu modulowanego [TITZE, 1988]. Z wyników uzyskanych w powyższych pracach wynika, że można przyjąć istnienie dwóch różnych stref przepływu powietrza. Jedna, od wlotu przez całą długość głośni, dla której przepływ jest laminarny i może być opisany zmodyfikowanym równaniem Bernoulliego. Natomiast druga, u wylotu głośni ( $y=D/2$ ), dla której przepływ staje się turbulentny i straty energii są wówczas znaczne. Mniej niż 20% składowej kinetycznej ciśnienia u wylotu głośni jest odzyskane przy przejściu z przekroju o powierzchni  $a_2$  do przekroju o powierzchni  $A_g$ . Stąd spadek ciśnienia na głośni wynosi

$$P_1 - P_2 = (k_c - k_e) \frac{\rho}{2} |u| \frac{u}{a_2} = k_t P_{k2}, \quad 6.62$$

gdzie

$k_c$  - jest współczynnikiem strat składowej kinetycznej ciśnienia w obrębie przepływu laminarnego (wlot i cała głośnia),

$k_e$  - współczynnik odtworzenia składowej kinetycznej ciśnienia (dla obszaru przepływu turbulentnego),

$\rho$  - gęstość powietrza,

$u$  - prędkość objętościowa przepływu powietrza u wylotu głośni,

$k_t = k_c - k_e$  - jest współczynnikiem spadku ciśnienia wzdłuż głośni,

$P_{k2}$  - składowa kinetyczna ciśnienia u wylotu głośni.

Oporność strat wywołanych lepkością jest uwzględniona we współczynniku  $k_e$  wyznaczanego doświadczalnie. Cytowani wcześniej autorzy podają następujące wartości współczynników  $1.0 < k_c < 1.4$  oraz  $K_e = 2(a_2/A_g(1-a_2/A_g)) < 0.2$ .

Dolna granica współczynnika  $k_c$  ma zastosowanie dla bardzo zbieżnych ścianek głośni, zaś górna w przypadku, gdy głośnia przybiera kształt prostokątny. Przy współczynniku  $k_c = 1.0$ , całkowity spadek ciśnienia wzdłuż głośni aż do jej wylotu jest pochodzenia kinetycznego (nie ma strat energii). Ponieważ zakres zmian współczynnika  $k_c$  jest względnie mały, wynika stąd że i straty energii na tym odcinku przepływu nie są zbyt duże. Zależność określająca wielkość współczynnika  $k_c$  została wyprowadzona z zasady zachowania ciągłości energii kinetycznej. Podczas niemal całego okresu drgań fałdów głosowych współczynnik  $k_c$  jest mniejszy od 0.1, co sugeruje że u wylotu głośni nie następuje odtworzenie składowej kinetycznej ciśnienia.

Zakładając, że straty energii rosną zgodnie z kierunkiem przepływu, ciśnienie wewnątrz głośniowe (aż do składowej 1-ego rzędu) może być wyznaczone z równania Bernoulliego

$$P(y) = P_2 + P_{k2} \left( 1 - \frac{a_2^2}{a^2(y)} \right) \quad 6.63$$

gdzie  $P_2$  jest ciśnieniem u wylotu z głośni. Całkując lewą stronę równania (6.61) otrzymuje się wyrażenie na średnie ciśnienie wewnątrzgłośniowe

$$P_w = P_2 + P_{k2} \left\{ 1 - \left[ \frac{a_2^2}{D} (-a^{-1}) \left( \frac{\partial a}{\partial y} \right)^{-1} \right]^{+\frac{D}{2}} \right\} \quad 6.64$$

By otrzymać taką postać rozwiązania, należało przyjąć założenie, że gradient  $\delta a / \delta y = (a_2 - a_1) / D$  jest niezależny od  $y$ , co jest spełnione w przypadku założenia trapezowego kształtu przekroju pionowego głośni. Podstawiając ten gradient do powyższego równania otrzymuje się

$$P_w = P_2 + P_{k2} \left(1 - \frac{a_2}{a_1}\right). \quad 6.65$$

Ponieważ odtworzenie składowej kinetycznej ciśnienia wyjściowego wynosi

$$P_2 - P_g = k_e P_{k2}, \quad 6.66$$

to  $P_2$  z ostatniego równania można podstawić do poprzedniego, zaś wyjściową składową kinetyczną  $P_{k2}$  zastąpić wyrażeniem  $(P_s - P_g)/k_i$  (patrz zależność 6.62). W wyniku, otrzymuje się zależność między wartością średniego ciśnienia pobudzającego i różnicą ciśnień ponad- i pod-głośniowego:

$$P_w = P_g + \frac{1}{k_i} (P_s - P_g) \left(1 - k_e - \frac{a_2}{a_1}\right). \quad 6.67$$

W poprzednio rozpatrywanych modelach źródła krtaniowego, przyjmowano, że ciśnienie pobudzające fałdy głosowe do drgań jest równe różnicy ciśnień pod- i ponad-głośniowego. To jest prawdziwie jedynie wówczas, gdy ciśnienie na wejściu toru głosowego ( $P_g$ ) jest równe zeru. Tak jest, gdy przekrój głośni jest bardzo zbieżny i na jej wyjściu nie ma odtworzenia ciśnienia kinetycznego. Dla tak raczej trywialnego przypadku wspomniana różnica ciśnień staje się po prostu równa ciśnieniu podgłośniowemu, co w rzeczywistości nie ma miejsca.

Jak już wcześniej wspomniano w sposób tylko opisowy, skuteczne ciśnienie pobudzające jest zależne od geometrii przekroju głośni. Podczas jej rozwierania się stosunek  $a_2/a_1$  jest mniejszy, niż podczas jej zamykania, co powoduje asymetrię siły pobudzającej. W momencie zbliżania się przekroju głośni do kształtu prostokątnego siła pobudzająca maleje, jeżeli  $P_g$  maleje. To może mieć miejsce, gdy wspomniana różnica ciśnień jest duża. Ujemne ciśnienie pobudzające może również wystąpić, gdy  $P_g$  staje się ujemne, co może wystąpić tuż przed zamknięciem głośni, gdy bezwładność masy powietrza znajdującej się nad szparą głośni powoduje powstanie siły ssącej ponad fałdami, lub gdy przekrój głośni jest rozbieżny w kierunku ku górze.

#### 6.7.2 Analiza ruchu fałdów przy braku sprzężenia głośni z torem głosowym.

W przypadku, gdy ciśnienie podgłośniowe jest równe ciśnieniu powietrza w płucach  $P_L$ , zaś ciśnienie ponadkrtaniowe jest równe atmosferycznemu ( $P_g = 0$ ), wówczas ciśnienie pobudzające jest

$$P_w = \frac{P_L}{k_l} (1 - k_e - \frac{a_2}{a_1}). \quad 6.68$$

Ten przypadek jest stosunkowo łatwy do zasymulowania w laboratorium na wyciętych fałdach głosowych. Wówczas tylko ciśnienie w płucach jest zmienne, bowiem oba ciśnienia, pod- i ponad-głośniowe są ustalone. Rozpatrując typowy przypadek, w którym ponadgłośniowa powierzchnia jest duża (to jest  $a_2 \ll A_g$ , lub  $k_e = 0$ ). Podstawiając zależności (6.59) do wzoru (6.68), a następnie otrzymany wynik na ciśnienie  $P_g$  do lewej strony wzoru (6.61) (odpowiedź tkanki śluzówkowej) otrzymuje się równanie ruchu o postaci

$$\left( \frac{P_L}{k_l} \right) \frac{\left( x_{o1} - x_{o2} + 2\tau \frac{\partial x}{\partial t} \right)}{\left( x_{o1} + x + \tau \frac{\partial x}{\partial t} \right)} = M \frac{\partial^2 x}{\partial t^2} + B \frac{\partial x}{\partial t} + Kx. \quad 6.69$$

Jest to autonomiczne, nieliniowe równanie różniczkowe o ogólnej postaci identycznej do przedstawionej wzorem (6.1), w którym nie występuje jawna zależność od czasu  $t$ . Jednakże należy zwrócić uwagę, że w ciśnieniu pobudzającym występuje czynnik, który zmienia się bezpośrednio z prędkością ruchu tkanki śluzówki. I choć czynnik ten występuje zarówno w liczniku, jak i w mianowniku lewostronnego wyrażenia, to jednak w liczniku jest on dodatkowo pomnożony przez współczynnik 2. Jest oczywistym, że ciśnienie pobudzające jest większe dla dodatniej pochodnej przemieszczenia, niż dla ujemnej, co jest zgodne z rozważaniami w poprzedniej części.

W celu uzyskania uproszczenia obliczeń przyjmijmy, że amplituda drgań jest stosunkowo mała i jej funkcja ma postać  $x = x_g + \Delta x$ , gdzie  $x_g$  jest średnim (statycznym) przemieszczeniem fałdu, zaś  $\Delta x$  jego składową oscylacyjną o małej amplitudzie. Składowa statyczna spełnia równanie kwadratowe

$$\overline{P_w} = \left( \frac{P_L}{k_l} \right) \frac{x_{o1} - x_{o2}}{x_{o1} + x_{tr}} = Kx_{tr}, \quad 6.70$$

kótre otrzymuje się przy podstawieniu w równaniu (6.67)

$$\frac{\partial^2 x}{\partial t^2} = \frac{\partial x}{\partial t} = 0.$$

Rozwiązanie określone jest wówczas wzorem

$$x_{tr} = -\frac{x_{01}}{2} \pm \sqrt{\left(\frac{x_{01}}{2}\right)^2 + \frac{P_L x_{01} - x_{02}}{k_t K}} \quad 6.71$$

Ponieważ dodatnie ciśnienie w płucach może powodować przesunięcie tylko w dodatnią stronę, a więc powyższe wyrażenie ma sens fizyczny tylko, gdy występuje "plus" przed pierwiastkiem. Aby móc dalej prowadzić analizę, zakłada się, że połączone, prefonacyjne i statyczne przemieszczenie ścianki fałdu, razem wzięte są znacznie większe niż składowa dynamiczna (oscylacyjna) przesunięcia. Rozkładając mianownik we wzorze (6.69) wokół dwumianu  $x_{01} + x_{tr}$ , przy jednoczesnym założeniu, że

$$x_{01} + x_{tr} \phi \phi \left( \Delta x + \tau \frac{\partial x}{\partial t} \right)^2 \quad 6.72$$

i mnożąc go przez licznik, a następnie zachowując tylko liniowe składniki, otrzymuje się stałe współczynniki przy zmiennej  $\Delta x$  i jej pochodnych. Wartości skutecznego tłumienia  $B'$  dla małej amplitudy oscylacji oraz sztywności  $K'$  wyrażone na jednostkę powierzchni mogą być określone poprzez odpowiednie zsumowanie współczynników występujących przy przesunięciach i prędkościach w rozwiązaniu zlinearyzowanego równania różniczkowego drugiego rzędu. Współczynniki te są określone następującymi zależnościami

$$B' = B - \frac{P_L \tau (x_{01} + x_{02} + 2x_{tr})}{k_t (x_{01} + x_{tr})^2} = B - B^* \quad 6.73$$

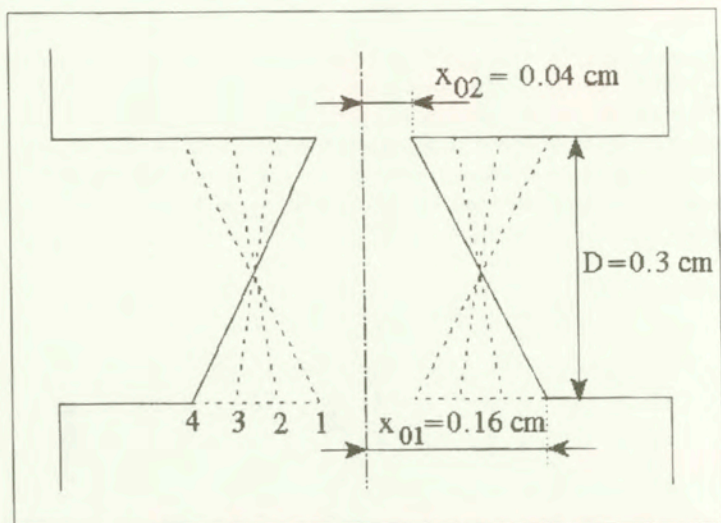
$$K' = K + \frac{P_L x_{01} - x_{02}}{k_t (x_{01} + x_{tr})^2} = K + K^* \quad 6.74$$

gdzie  $B^*$  i  $K^*$  są odpowiednio aerodynamicznym tłumieniem i sztywnością. Natomiast masa drgająca tkanki śluzówkowej przypadająca na jednostkę powierzchni fałdów jest ta sama, tzn.  $M^* = M$ . Progowe ciśnienie pobudzające  $P_L$ , to jest ciśnienie, przy którym tłumienie jest równe 0, można wyznaczyć ze wzoru (6.73) podstawiając  $B=0$  oraz  $t = D/2c$

$$P_L = \frac{2k_t (x_{01} + x_{tr})^2}{D x_{01} + x_{02} + 2x_{tr}} Bc \quad 6.75$$

Z powyższej zależności widać, że progowe ciśnienie pobudzające maleje, gdy maleją tłumienie w błonie śluzówkowej, lub prędkość rozchodzenia się w niej fali powierzchniowej, lub wielkość otworu wlotowego głośni, oraz gdy wzrasta grubość

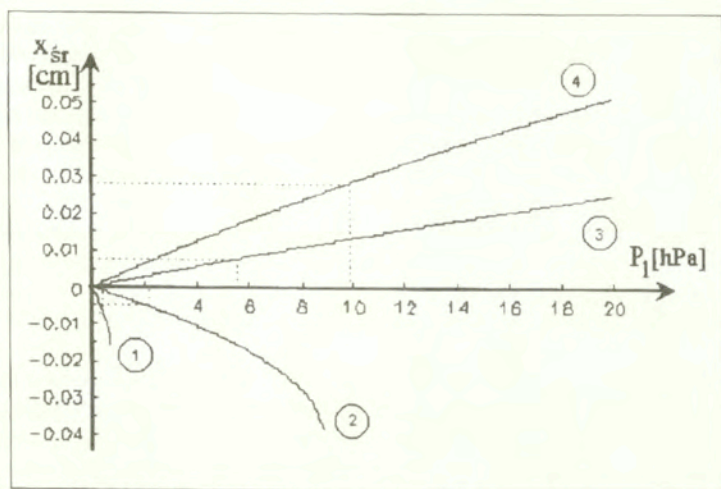
fałdów. Iloczyn  $Bc$  w powyższym wzorze będzie określany jako skuteczne ciśnienie tłumienia. Wzrost tłumienia lepkościowego w tkance fałdu głosowego może być zniwelowany wzrostem opóźnienia  $t$  (wskutek mniejszej prędkości rozchodzenia się fali powierzchniowej) w błonie słuzówkowej. Aby wyznaczyć ciśnienie  $P_L$  ze wzoru (6.71) należy wyznaczyć  $x_{sr}$  w funkcji  $P_L$ , a następnie wstawić je do wzoru (6.75). Obecność w tym ostatnim czynników  $x_{01}$  i  $x_{02}$  sygnalizuje, że progowe ciśnienie pobudzające, przy którym fałdy zostają wprowadzone w stan oscylacji zależy w istotny sposób od konfiguracji szpary głośni. Na Rys. 6.17 [TITZE,1988] przedstawiono cztery zasadnicze, odpowiednio ponumerowane (od 1 do 4) konfiguracje głośni - dwie zbieżne (o dużej i małej zbieżności) oraz dwie analogiczne wymiarowo, o przekroju rozbieżnym. Punkt środkowy, znajdujący się na ścianie fałdu w połowie jej wysokości jest odległy od osi symetrii  $y$  o 0.10 cm i nie zmienia swojego położenia. Natomiast, przy zmianie konfiguracji głośni  $x_{01}$  i  $x_{02}$  zmieniają się symetrycznie w przeciwstawnych sobie kierunkach, w odstępach co 0.04 cm, od 0.04 do 0.16 cm.



Rys. 6.17 Cztery podstawowe konfiguracje przekroju głośni rozpatrywane przy modelowaniu źródła krztaniowego [TITZE,1988].

Na Rys. 6.18 przedstawiono przebieg wzajemnej zależności statycznego przesunięcia  $x_{sr}$  od progowej wartości ciśnienia pobudzającego  $P_L$ . W celu jej wyznaczenia konieczne jest rozwiązanie równania kwadratowego dla 4-ch wybranych konfiguracji (patrz wzór 6.70).





Rys. 6.18 Zależność statycznego wychylenia ścianek fałdów  $x_{sr}$  od ciśnienia podłożniowego  $P_L$  dla poszczególnych konfiguracji szpary głośni. Oznaczenia konfiguracji zgodne z oznaczeniami na Rys. 6.17. Liniami kropkowanymi zaznaczono położenia równowagi, dla których drgania fałdów mają charakter stacjonarny.

W celu wyznaczenia zależności  $P_L$  od ciśnienia tłumieniowego ( $B_c$ ) należy rozwiązać zależność (6.75) dla przemieszczenia statycznego. Warto zwrócić uwagę, że przemieszczenie to jest dodatnie dla przekrojów zbieżnych, zaś ujemne dla rozbieżnych. Jest to bezpośrednią konsekwencją faktu, że ciśnienie Bernoulliego bądź "wypycha" na zewnątrz, lub "zasysa" środkową część ścianek fałdów, tak aby powstało pewne położenie równowagi, wokół którego następują oscylacje, przy czym, przy zwiększającej się zbieżności przekroju głośni, maleje zakres oscylacji. Na przykład dla konfiguracji oznaczonej liczbą 2 (lekką rozbieżną), oscylacje mogą być podtrzymywane przy ciśnieniu  $P_L = 2.12$  hPa, przy skutecznym ciśnieniu tłumieniowym rzędu 10 hPa. Natomiast dla konfiguracji 3 (lekką zbieżną), przy tym samym ciśnieniu tłumieniowym, dla podtrzymania w układzie oscylacji niezbędne jest ciśnienie  $P_L$  rzędu 5.45 hPa. Zaś w przypadku skrajnych konfiguracji (1 oraz 4) odpowiednie ciśnienia  $P_L$  wynoszą 0.42 oraz 9.95 hPa.

Na wykresach przedstawionych na Rys. 6.18 zaznaczono (liniami kropkowanymi) położenia równowagi wyznaczone dla odpowiednich pobudzeń progowych, dla których charakter oscylacji jest stacjonarny (tzn. amplituda drgań jest stała). Powyżej danego położenia amplituda stopniowo rośnie, poniżej maleje (wskutek przewag tłumienia).

W obliczeniach przyjęto wartości współczynników takie jak w modelu dwumasowym. Współczynnik spadku ciśnienia na szparze głośni  $k_1$  przyjęto równy 1.1. Dla tych wartości i geometrii szpary głośni, zakres przesunięcia statycznego nie przekracza  $\pm 0.05$  cm dla całego zakresu zmian ciśnienia płucnego 0-20 hPa. Dla różnych konfiguracji głośni statyczne ciśnienie Bernoulliego  $P_{we}$  (równaniu 6.70) jest ujemne i może w swej wielkości przewyższać oddziaływanie liniowej sztywności sprężyny  $Kx_{sr}$ , co powoduje wyjście układu ze stanu równowagi. Może wówczas nastąpić zwarcie fałdów, co matematycznie jest reprezentowane przez zespoloną postać rozwiązania zależności (6.71). Objawia się to pojawieniem się punktów o składowej urojonej w dół od krzywych 1 i 2 przedstawionych na Rys. 6.18 (krzywe te urywają się w momencie pojawienia się liczby ujemnej pod pierwiastkiem we wzorze 6.71).

Jednakże założony model drgań o małej amplitudzie uniemożliwia analityczne rozwiązanie tego problemu w przypadku braku równowagi układu, to jest braku stałego wychylenia, wokół którego odbywają się oscylacje. Ponadto, przy dużej amplitudzie drgań, pojawiają się istotne czynniki ograniczające ich zakres wielkości zmian, zwłaszcza takie jak nieliniowy charakter zmian sztywności fałdów, czy wręcz zderzanie się fałdów ze sobą. Inny czynnik upraszczający w przyjętym przybliżeniu, to skokowa zmiana konfiguracji przy przejściu od tchawicy do głośni, tak jak to przedstawiono na Rys. 6.15, co w rzeczywistości nie ma miejsca i zmiana ta odbywa się w sposób płynny, ograniczając w ten sposób prefonacyjne kształty przybierane przez szparę głośni.

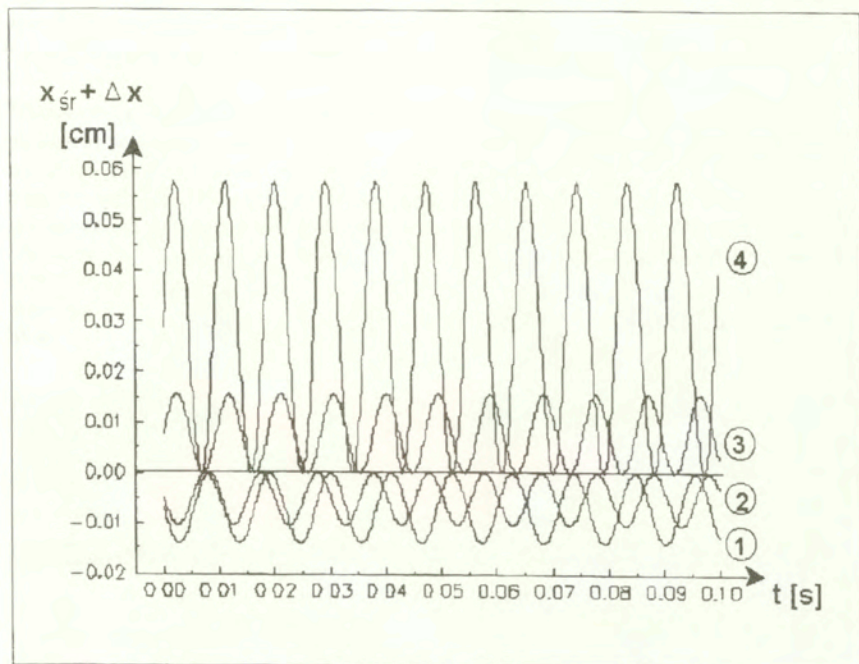
Najmniejsza wartość  $x_{o1}$  jest równa 0.04 cm, która jest mniej więcej tego samego rzędu co największa wartość  $x_{sr}$ . Tak dla większości konfiguracji głośni można więc przyjąć, że  $x_{sr}$  jest pomijalnie małe względem  $x_{o1}$ . To prowadzi do znacznego uproszczenia równania (6.75) i przybliżona zależność progowego ciśnienia pobudzającego od ciśnienia tłumieniowego ma wówczas liniową postać

$$P_L = 2 \frac{k_1}{D} \frac{x_{o1}^2}{x_{o1} + x_{o2}} Bc. \quad 6.76$$

Istotnym sposobem weryfikacji założeń dla układu o małej amplitudzie drgań jest rozwiązanie numeryczne nieliniowego równania różniczkowego i porównanie go z poprzednimi wynikami. Można to łatwo uzyskać stosując konwencjonalne metody numeryczne. Wyniki rozwiązania równania dla wszystkich rozpatrywanych uprzednio czterech konfiguracji głośni przedstawiono na Rys. 6.19 (zastosowano pakiet MATHEMATICA) przedstawiając przebieg zmian czasowych  $x = x_{sr} + \Delta x$  dla wszystkich czterech rozpatrywanych konfiguracji szpary głośni (podobny wykres jest w pracy [TITZE, 1988]).

Na Rys. 6.19 statyczne przesunięcie sinusoid dokładnie pokrywa się z przesunięciem statycznym wynikającym z przebiegu krzywych  $x_{sr}$ , wyznaczonych dla poszczególnych przekrojów prefonacyjnych głośni. Ponieważ amplituda sinusoidalnych

przemieszczeń  $\Delta x$  jest mniejsza od 0.01 cm, a najmniejsze przemieszczenie wokół położenia równowagi  $x_{01}+x_{sr}$  wynosi  $0.08\pm 0.01$ , tak więc rozwinięcie dwumianowe wokół  $x_{01}+x_{sr}$ , umożliwiające linearyzację procesu opisanego zależnością (6.69) jest w zasadzie słuszne, bowiem kryterium prawdziwości tego rozwinięcia  $(x_{01}+x_{sr})^2 \gg (\Delta x)^2$  jest spełnione.



**Rys. 6.19** Wykres czasowy przebiegu średniego przemieszczenia środka ścianki fałdu głosowego przy ciśnieniu pobudzającym równym wartości progowej. Przebiegi wyznaczono dla 4 konfiguracji szpary głośni.

Pozostaje rozpatrzyć najprostszy przypadek konfiguracji głośni, to jest, gdy szpara głośni jest prostokątna, co już było przedmiotem rozważań dla modelu jednomasowego. Z równania (6.71), a także z rysunku 6.18, wynika że przy  $x_{01}=x_{02}$  nie nastąpi statyczne przesunięcie ( $x_{sr}=0$ ). Wówczas progowe ciśnienie, przy którym powstaną oscylacje jest określone wzorem o uproszczonej postaci

$$P_L = \frac{k_t}{D} x_{01} B c. \quad 6.77$$

Ze wzoru tego widoczna jest bezpośrednia zależność wielkości progowego ciśnienia pobudzającego od szerokości szpary głośni, przy którym następuje ruch drgający ścianek fałdów. Im bliżej są one siebie, tym łatwiej wprowadzić je w ruch drgający o małej amplitudzie. Na przykład, jeżeli dla wszystkich szerokości szpary ciśnienie strat  $Bc$  jest rzędu 10 hPa, to poziom ciśnienia progowego zmienia się od 1.5 hPa dla  $x_{01}=x_{02}=0.04$ , aż do 5.9 hPa dla  $x_{01}=x_{02}=0.16$ . Warto wreszcie zwrócić uwagę, że dla tego zakresu zmian  $x_{01}$  zakres zmian ciśnienia progowego, choć jest stosunkowo duży, jest jednak mniejszy od zakresu zmian wywołanego samą zmianą kształtu głośni.

Ostatni przykład umożliwia porównanie uzyskanych wyników z wynikami otrzymanymi dla modelu dwumasowego. Jako punkt wyjścia można przyjąć równanie ruchu dla dolnej masy  $m_1$ , dla współczynnika tłumienia ruchu  $r_1$ , współczynnika sztywności  $s_1$  (6.43). Przy proporcjonalności siły oddziaływania między obiema masami do przemieszczenia jednej względem drugiej oraz mechanicznej sztywności  $s_c$  sprzęgającej górną i dolną masę równej zero, równanie ruchu ma postać

$$m_1 \frac{\partial^2 x_1}{\partial t^2} + r_1 \frac{\partial x_1}{\partial t} + s_1 x_1 = 2\phi(x_1 - x_2). \quad 6.78$$

Ponieważ w modelu ciągłym mechaniczne sprzężenie istniejące między górną i dolną ścianką fałdów jest określone w obliczeniach poprzez uwzględnienie rozchodzenia się w błonie śluzówki fałdu z prędkością  $c$  fali powierzchniowej, stąd  $s_c$  nie występuje w ostatniej zależności.

Aerodynamiczna w swoim charakterze sztywność sprzęgająca  $\Phi$  obie masy, dla przypadku  $m_1=m_2$  jest określona wzorem

$$\begin{aligned} \phi &= l_g \frac{D}{2} \frac{\rho v^2}{2} \frac{1}{2x_0} = l_g \frac{D}{2} \frac{P_{k_1}}{2\rho_0} \\ &= l_g \frac{D}{2} \frac{P_L}{k_1 x_0}, \end{aligned} \quad 6.79$$

gdzie  $v$  jest średnią liniową prędkością wychylenia ścianek, gdy szpara głośni ma przekrój prostokątny. Ostatnie przekształcenie jest zgodne z przyjętym wcześniej przybliżeniem, że  $k_c \approx 0$ ,  $P_g = P_L$  oraz  $P_g = 0$  (patrz wzór 6.62). Podstawiając  $x_2 = x_1 - 2\tau \delta x_1 / \delta t$  podobnie jak w rozwinięciu na szereg Taylora, które zostało zastosowane dla określenia przemieszczenia punktu środkowego ścianki fałdu (wzór 6.55), z tą różnicą, że w tym przypadku rozwinięcie to zostało wykonane w równaniu ruchu dolnej masy  $m_1$ . Tak więc równanie (6.78) przybiera postać

$$m_1 \frac{\partial^2 x}{\partial t^2} + (r_1 - 4\tau\phi) \frac{\partial x}{\partial t} + s_1 x_1 = 0. \quad 6.80$$

W celu porównania progowej wartości tłumienia ( $r_1-4\tau\Phi$ ), z tą która wynika z zależności (6.77), należy uwzględnić, że tłumienie B jest wyrażone na jednostkę powierzchni fałdu równej  $l_g \times D$ . Ponieważ powierzchnia jego dolnej części o masie  $m_1$  jest równa połowie powierzchni fałdu, stąd

$$B = \frac{2r_1}{l_g D} = \frac{8\tau\phi}{l_g D} \quad 6.81$$

Podstawiając wzór na  $\Phi$  (6.79) do powyższej zależności oraz podstawiając  $\tau=D/2c$ , wzór (6.81) przybiera identyczną jak wzór (6.77).

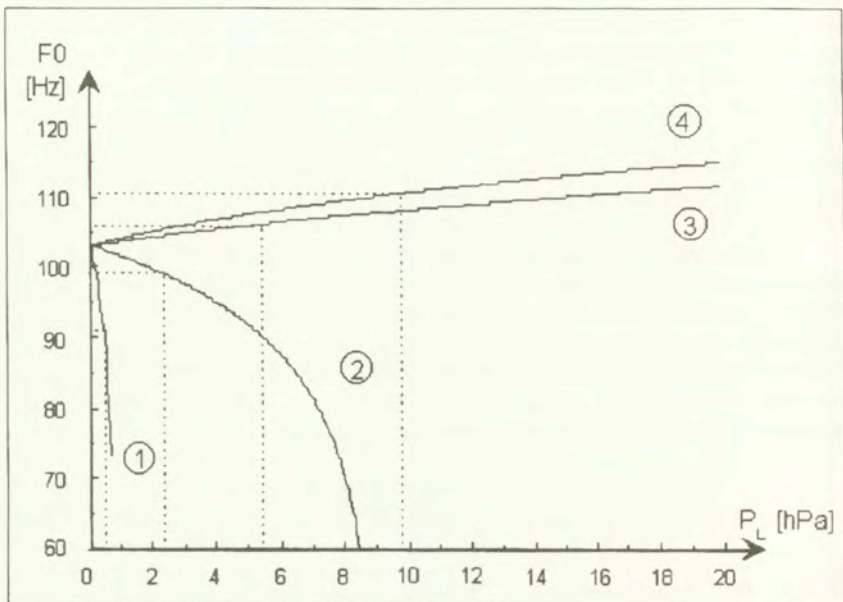
W modelu dwumasowym jednym z podstawowych warunków powstania oscylacji jest to, że sztywność mechaniczna sprężająca obie masy musi być mniejsza od składowej aerodynamicznej sztywności  $\Phi$ . Jest to równoważne stwierdzeniu, że prędkość fali powierzchniowej rozchodzącej się w zewnętrznej tkance śluzówkowej fałdu musi być nie mniejsza od  $2D\Phi/r_1$ . Inaczej opóźnienie  $\tau$  w równaniu (6.80) jest zbyt małe, by oporność  $4\tau\Phi$  przewyższyła oporność strat  $r_1$ .

Z rozwiązania równania ruchu (6.69) wynika że częstotliwość podstawowa drgań ścianek fałdów jest wprost proporcjonalna do sztywności sumarycznej  $K'$  (wzór 6.74) i odwrotnie proporcjonalna do ich masy. Ze wzoru (6.79) wynika, że dla zbieżnej konfiguracji szpary głośni (3 i 4) składowa aerodynamiczna sztywności jest dodatnia i powoduje wzrost częstotliwości podstawowej  $F_0$ . Jednocześnie wielkość tego przyrostu jest wprost proporcjonalna do ciśnienia  $P_L$ , co było już stwierdzone dla mowy naturalnej (LADEFOGED,1963). Zależność ta ma postać

$$F_0 = \frac{1}{2\Pi} \sqrt{\frac{(K + K')}{M}} \quad 6.82$$

Na Rys. 6.20 przedstawiono dla czterech konfiguracji szpary głośni zależność częstotliwości podstawowej od ciśnienia  $P_L$ , przy założeniu jednostkowej masy drgającej fałdów  $M=0.476\text{g/cm}^2$  i ich sztywności jednostkowej  $K=0.02 \text{ N/m}^3$  ([TITZE,1988]). Początkowa częstotliwość własna fałdów przy takich wartościach parametrów wynosi prawie 103 Hz. Dla konfiguracji zbieżnych wzrost ciśnienia  $P_L$  pociąga za sobą wzrost częstotliwości drgań fałdów w stosunku 0.5 Hz/hPa, natomiast dla konfiguracji rozbieżnych częstotliwość  $F_0$  maleje w znacznie większym stopniu 5-10 Hz/hPa. Ujemny znak składowej aerodynamicznej sztywności jest oczywisty ze wzoru (6.74), bowiem dla konfiguracji 1 i 2  $x_{01} < x_{02}$ . Ponadto,  $x_{sr}$  jest również ujemne, gdy  $x_{01}$  jest małe, bowiem następuje szybkie zbliżanie się do punktu osobliwego we wzorze na  $K'$  (wykresy dla konfiguracji rozbieżnych 1 i 2 urywają się w pobliżu punktu osobliwego).

Jednakże tak znaczne zmiany w kierunku ujemnym muszą być przyjmowane z dużą ostrożnością. Przy oscylacjach o dużej amplitudzie, zderzanie się ścianek fałdów ze sobą i nieliniowe zmiany ich sztywności mogą przeważać nad zmianami sztywności wyprowadzonymi dla małych amplitud oscylacji. I tak w przypadku modelu dwumasowego rozpatrywanego wcześniej, tego typu zależność między częstotliwością podstawową i ciśnieniem podgłośniowym (to jest występowanie ujemnej aerodynamicznej sztywności) nie obserwuje się. Stanowi to ostrzeżenie przed zbyt daleko idącym ekstrapolowaniem wniosków wynikających z modelu słusznego jedynie dla drgań o małej amplitudzie.



**Rys. 6.20** Zależność częstotliwości drgań fałdów głosowych  $F_0$  od ciśnienia podgłośniowego. Na krzywych zaznaczono liniami kropkowanymi punkty odpowiadające wartościom progowym ciśnienia pobudzającego dla poszczególnych konfiguracji głośni (wg [ITTZE,1988]).

Również z Rys. 6.19 przedstawiającego przebiegi zmiany wychylenia  $\Delta x[t]$  dla czterech konfiguracji szpary głośni (przy ciśnieniach progowych zgodnych z zależnością (6.75)), widoczna jest zależność częstotliwości podstawowej od ciśnienia  $P_L$  i rozmiaru zbieżności (lub rozbieżności) szpary głośni. Częstotliwości drgań dla konfiguracji głośni (1-4) wynoszą odpowiednio 92,99,106 i 111 Hz. Tak więc częstotliwość podstawowa  $F_0$  zależy głównie od ciśnienia podgłośniowego  $P_L$  i stopnia zbieżności (lub rozbieżności) szpary głośni. Jak podaje Hirano [HIRANO,1983] wzrost

napięcia fałdów głosowych powoduje zbliżenie do siebie dolnych części fałdów, wskutek czego przekrój szpary głośni przybiera kształt prostopadłościenny (lub nawet lekko rozbieżny). Wiadomo [HIRANO,1977], że mięśnie głosowe są dla rejestru piersiowego znacznie aktywniejsze niż w przypadku rejestrów wyższych (np. falsetu). Przyjmując, że konfiguracja (4) szpary głośni odpowiada prefonacyjnemu ustawieniu dla rejestru falsetowego (zwanemu również głowowym), zaś konfiguracja (2) - rejestrowi piersiowemu, oczywistym staje się że znacznie większe zmiany częstotliwości podstawowej mogą być uzyskane łatwiej poprzez zmianę rejestru, niż poprzez zmianę ciśnienia podgłośniowego. Dla typowego ciśnienia w płucach rzędu 8 hPa [TITZE,1988] częstotliwość  $F_0$ , przy zmianie konfiguracji głośni ze zbieżnej (4) na rozbieżną (2), może zmaleć się w sposób skokowy ze 107 Hz do 71 Hz. Podobnie gwałtowne zmiany mogą wystąpić w amplitudzie pobudzenia krtaniowego, co jest widoczne na przebiegach przedstawionych na Rys.6.19.

Ponieważ średnie przesunięcie fałdu w kierunku osi symetrii szpary głośni (przy zmianie konfiguracji z (4) na (2) przy ciśnieniu podgłośniowym  $P_g=P_L=8$  hPa jest rzędu 0.05 cm (Rys. 6.18), zaś średnia odległość prefonacyjna - 0.1 cm, tak więc podczas oscylacji fałdów następuje zredukowanie szerokości szpary głośni rzędu 50%. Pociąga to za sobą dłuższe i bardziej gwałtowne zamykanie się szpary w przypadku oscylacji o dużej amplitudzie. Gdy ciśnienie pobudzające przekroczyłoby nieco tę wartość ( $B'$  staje się ujemne), ścianki głośni zetknęłyby się ze sobą w momencie przejścia do konfiguracji (2), nawet przy zachowaniu warunku drgań o małej amplitudzie. Jednakże w praktyce, należy uwzględnić fakt, że charakterystyka sztywności fałdów w zależności od wychylenia nie jest liniowa (patrz wzór 6.41) i przy zwiększonym ich napięciu skuteczna sztywność  $K$  rośnie znacznie szybciej, przeciwdziałając w ten sposób pełnemu zwarciu. Nieliniowe zmiany charakterystyki sztywności powodują znacznie większe zmiany częstotliwości  $F_0$ , niż w przypadku modelu liniowego.

6.7.3 Warunki oscylacji przy uwzględnieniu bezwładnościowego sprzężenia źródła z torem głosowym.

Jak wykazał Rothenberg [ROTHENBERG,1981], przy częstotliwości  $F_0$  poniżej pierwszego formantu, reaktancyjna część impedancji wejściowej toru głosowego ma charakter głównie bezwładnościowy. Można zatem impedancję tę przedstawić w układzie o stałych skupionych, za pomocą dwóch szeregowo ze sobą połączonych elementów  $I_1$  oraz  $R_1$ . Podgłośniowe obciążenie jest podobnego charakteru i może być traktowane z obciążeniem nadgłośniowym w sposób łączny.

Niech ciśnienie pobudzające tor głosowy u jego wejścia jest określone wzorem

$$P_g = R_1 u + I_1 \frac{\partial u}{\partial t} \quad 6.83$$

W celu określenia wpływu obciążenia źródła przez tor głosowy, założmy że ścianki fałdów poruszają się synchronicznie na całej swej powierzchni, wskutek czego nie ma przesunięcia fazowego między dolną i górną krawędzią fałdu, czyli  $\tau = 0$ , zaś  $a_1 = a_2 = a$ . Prowadzi to bezpośrednio do przypadku modelu jednomasowego, dla którego warunki pobudzenia są maksymalnie zależne od wielkości sprzężenia z torem głosowym.

Dla uproszczenia można przyjąć jak poprzednio, że  $K_c = 0$  ( $a_2 \ll A_2$ ), w wyniku czego warunek powstania w układzie oscylacji w pełni zależy od ciśnienia wejściowego, to jest

$$P_w = P_g,$$

zgodnie ze wzorem (6.67).

Łącząc ze sobą wzór (6.83) z równaniem ruchu (6.61) otrzymuje się równanie o postaci

$$R_1 u + I_1 \frac{\partial u}{\partial t} = M \frac{\partial^2 x}{\partial t^2} + B$$

Po podstawieniu  $u(t) = a \cdot v(t)$ , gdzie  $v(t)$  jest prędkością cząsteczkową powietrza u wylotu głośni, zaś zgodnie ze wzorami (6.59)  $a(t) = 2 \cdot l_g [x_0 + x(t)]$ , równanie ruchu przybiera zatem postać

$$\begin{aligned} M \frac{\partial^2 x}{\partial t^2} + (B - 2 \cdot l_g \cdot v) \frac{\partial x}{\partial t} + (K - 2 l_g \cdot R_1 \cdot v - 2 l_g \cdot I_1 \frac{\partial v}{\partial t}) x \\ = 2 l_g \cdot x_0 (R_1 \cdot v + I_1 \frac{\partial v}{\partial t}). \end{aligned} \quad 6.84$$

Powyższe równanie ma dwie niewiadome -  $x(t)$  i  $v(t)$ , tak więc niezbędne jest druga zależność w celu otrzymania rozwiązania. Przyjmując jak poprzednio, że spadek ciśnienia między głośnią i wylotem z płuc jest pomijalnie mały, to jest  $P_w = P_L$ , to równanie przepływu (6.661) przybiera postać

$$P_L - R_w(t) - I_1 \frac{\partial u(t)}{\partial t} = K_1 \frac{\rho}{2} v^2. \quad 6.85$$

Korzystając z zasady zachowania ciągłości strugi, to jest  $u = av$ , oraz podstawiając do lewej strony powyższego równania otrzymuje się

$$\frac{\rho}{2} K_1 v^2 + 2 l_g (R_1 (x_0 + x) + I_1 \frac{\partial x}{\partial t}) v + 2 l_g I_1 (x_0 + x) \frac{\partial v}{\partial t} - P_L = 0. \quad 6.86$$



Powyższe nieliniowe równanie różniczkowe wraz z równaniem (6.84) wyznacza charakterystykę drgań układu jednomasowego obciążonego po stronie akustycznej.

Dokładne rozwiązanie analityczne tej pary równań jest złożone. Jednakże możliwe jest otrzymanie rozwiązania przybliżonego, przy zbliżonych do rzeczywistych warunkach. Po pierwsze należy wyodrębnić składowe stałe w zmiennych  $x(t)$  i  $v(t)$ , które można przedstawić w postaci

$$x(t) = x_{sr} + \Delta x(t), \quad v(t) = v_{sr} + \Delta v(t).$$

Następne uproszczenie polega na założeniu, że składowa oscylacyjna prędkości cząsteczkowej zmienia się przede wszystkim wraz ze strumieniem przemieszczenia i tylko w niewielkim stosunkowo stopniu ze zmianą przepływu pobudzającego (strumienia przepływającego przez szparę głośni). Przepływ przemieszczenia u wylotu z głośni równy  $2I_g D \cdot \partial x / \partial t$  jest ujemny, to znaczy, że powietrze jest z powrotem wciskane do środka wskutek bocznych ruchów słuzówki. Zakładając, że tylko ten wsteczny przepływ wpływa na oscylacyjne zmiany  $v$ , można więc napisać

$$\Delta v = -b \frac{\partial x}{\partial t}, \quad 6.87$$

gdzie  $b$  jest współczynnikiem stałym.

Podstawiając w równaniu (6.84) wyrażenia na zmienne  $v$  i  $x$ , oraz łącząc wszystkie współczynniki przy  $\Delta x$  i jej pochodnych, oraz eliminując składniki nieliniowe wyższego rzędu (iloczyn zmiennej  $\Delta x$  i jej pochodnych są pomijalnie małe przy małej amplitudzie oscylacji) otrzymuje się

$$\begin{aligned} M' &= M + 2I_g \cdot I_1 \cdot b(x_0 + x_{1r}), \\ B' &= B - 2I_g \cdot v_{1r} + 2I_g \cdot R_1 \cdot b(x_0 + x_{1r}), \\ K' &= K - 2I_g \cdot R_1 \cdot v_{1r}. \end{aligned} \quad 6.88$$

Powyższe wzory określają odpowiednio efektywną masę, tłumienność i sztywność układu oscylacyjnego. Dodatkowo składowe statyczne spełniają poniższe warunki:

a) z równania (6.84)

$$(K - 2I_g \cdot R_1 \cdot v_{1r}) = 2I_g \cdot R_1 \cdot v_{1r} r x_0, \quad 6.89$$

b) z równania (6.86)

$$\frac{\rho}{2} k_1 v_{sr}^2 + 2l_g R_l (x_0 + x_{sr}) v_{sr} - P_L = 0. \quad 6.90$$

W ostatniej zależności  $v_{sr}^2$  dominuje nad składnikiem liniowym  $v_{sr}$ . Bez tej dominacji, układ byłby za mało nieliniowy, by mógł wejść w stan oscylacji. W pierwszym przybliżeniu składowa stacjonarna prędkości cząsteczkowej dla bezstratnego toru głosowego ( $R_l=0$ ) wynosi

$$v_{sr} \cong \sqrt{\frac{2P_L}{k_1 \rho}} \quad 6.91$$

Stąd z równania (6.89) wynika, że  $x_{sr} \approx 0$ . Natomiast z zależności (6.89 i 6.90) poprzez kolejne aproksymacje uzyskuje się coraz dokładniejsze wartości  $v_{sr}$  i  $x_{sr}$  przy  $R_l \neq 0$ . Zaczynając od wartości  $v_{sr}$  wyznaczonej z zależności (6.91) i stanowiącej początkową estymatę, można następnie z równania (6.89) obliczyć  $x_{sr}$  i wstawić ją do równania (6.90). Z rozwiązania równania kwadratowego uzyskuje się lepszą estymatę  $v_{sr}$ , która następnie może być zastosowana do dokładniejszej aproksymacji  $x_{sr}$ , i tak ponownie od początku w kolejnych krokach aproksymacji.

Dla przykładu rozpatrzmy, jakie parametry akustyczne źródła otrzymuje się przy przyjęciu następujących parametrów słuchówki i powietrza:  $K=200\ 000$  dyn/cm<sup>3</sup>,  $M=0.476$  g/cm<sup>3</sup>,  $k_t=1.1$ ,  $x_0=0.04$  cm,  $l_g=1.4$  cm,  $I_t=0.01$  dyn s<sup>2</sup>/cm<sup>5</sup>,  $R_t=3.1$  dyna s/cm<sup>5</sup>. Inercja  $I_t$  jest oszacowana dla rury cylindrycznej  $\rho l/A$ , gdzie  $l$  jest długością rury (17.5 cm), zaś  $A$  jest polem powierzchni jej przekroju poprzecznego (2 cm<sup>2</sup>).  $R_t$  jest oszacowane bardzo zgrubnie z dobroci  $Q$  rezonansu toru głosowego i szerokości pasma pierwszego formantu. Dla  $F_1=500$  Hz i  $B_{F1}=50$  Hz

$$\frac{\omega_1 I_t}{R_t} = Q = 10.$$

Stąd

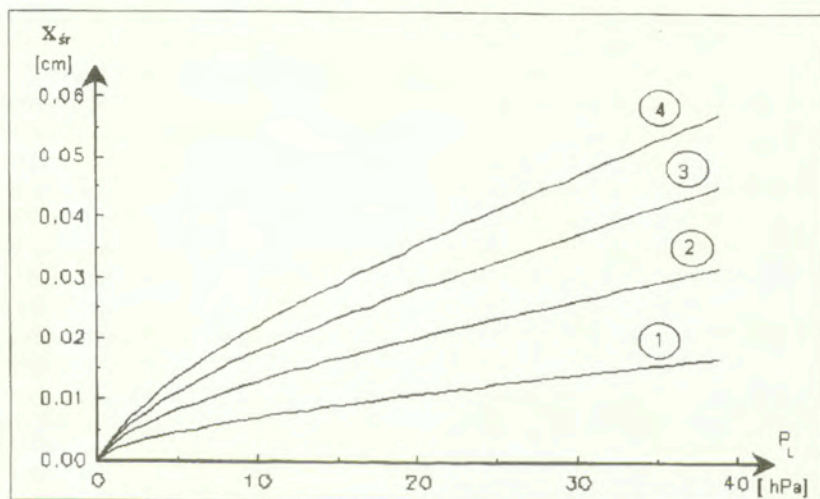
$$R_t = \frac{\omega_1 I_t}{Q}.$$

W obliczeniu tym zakłada się układ kaskadowo połączonych obwodów rezonansowych o stałej oporności (patrz Rozdz. 5). Uzyska się w najlepszym razie przybliżoną zależność  $R_t$  od częstotliwości. Wpływ na nią ma wiele czynników, wśród których na pierwszym miejscu należy wymienić oporność promieniowania u wylotu warg, oporność strat wywołaną drganiami ścian toru głosowego i oporność strat spowodowaną ich przewodnictwem cieplnym. Choć wpływ wyżej wymienionych czynników nie jest zbyt wielki, tym niemniej należy być ostrożnym w stosowaniu przybliżonych zależności przy poszukiwaniu dokładnych wartości równoważnej inercji i oporności toru głosowego.

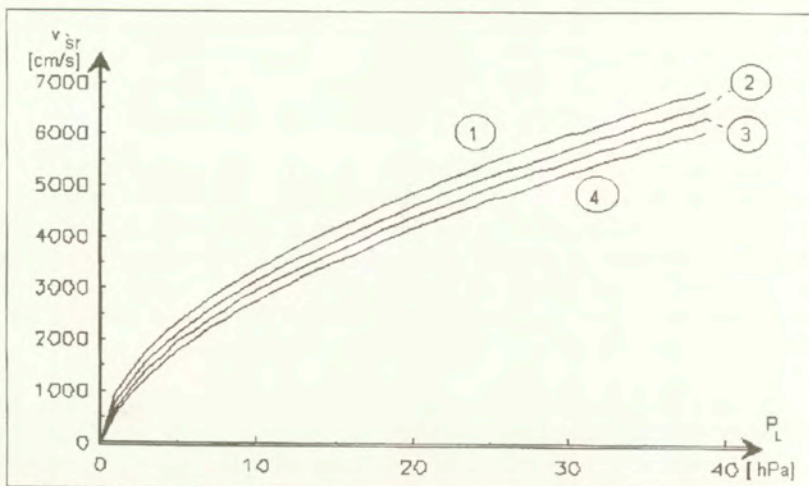
Dla wyżej podanych wartości parametrów i ciśnienia progowego 18 hPa, początkowa wartość  $v_{sr}$  (z r-a 6.91) wynosi 5304 cm/s. Stąd  $x_{sr}=0.012$  cm. Po wykonaniu drugiego kroku aproksymacji otrzymuje się  $v_{sr}=4956$  cm/s (z r-a 6.90). Po następnym kroku iteracji:  $v_{sr}=4963$  cm/s,  $x_{sr}=0.0114$  cm. Na ogół po 4-5 kroku iteracji uzyskuje się dokładność na piątym miejscu po przecinku.

Rodzina krzywych zależności między  $v_{sr}$  i  $x_{sr}$  od ciśnienia podgłośniowego  $P_L$  przedstawiono na Rys. 6.21 i 6.22. Numerami oznaczono krzywe odpowiadające kolejnym szerokościom prefonacyjnym szpary głośni  $x_0$  (od 0.04 do 0.16 cm), której przekrój jest prostokątny. Z podanych wykresów widać  $x_{sr}$  rośnie wraz ze wzrostem połowy szerokości szpary głośni  $x_0$ , zaś  $v_{sr}$  w tym samym czasie nieco maleje. Natomiast dla szpary głośni prostopadłościennej nie ma ujemnych wartości  $x_{sr}$ , tak jak to miało miejsce w przypadku konfiguracji rozbieżnych głośni.

Występowanie tylko dodatnich wartości  $x_{sr}$  dla wszystkich wielkości szpar głośni jest wynikiem oddziaływania składowej statycznej ciśnienia ponadgłośniowego  $R_t \cdot u_{sr}$ , gdzie  $u_{sr}$  jest prędkością objętościową odpowiadającą prędkości liniowej  $v_{sr}$ . Składowa ta powoduje odpychanie ścian fałdów od linii symetrii głośni dla wszystkich jej konfiguracji, co utrudnia wejście fałdów w stan oscylacji. Stąd impedancja ponadgłośniowa (a dokładniej oporność  $R_t$ ) podwyższa wielkość progowego ciśnienia pobudzającego i jednocześnie pociąga za sobą wzrost skutecznego tłumienia drgań.



Rys. 6.21 Zależność statycznego przemieszczenia ścianki fałdu  $x_{sr}$  od ciśnienia podgłośniowego  $P_L$ . (wg zależności w [TITZE,1988]).



Rys. 6.22 Zależność średniej prędkości cząsteczkowej  $v_{sr}$  od ciśnienia podgłośniowego  $P_L$  (wg zależności w [TITZE,1988]).

W celu przeanalizowania wpływu obciążenia toru głosowego na tłumienie i wielkość progowego ciśnienia pobudzającego należy wrócić do wzorów na skuteczną masę, tłumienie i sztywność układu oscylacyjnego (wzory 6.88) wynikające z ruchu fałdów o amplitudzie  $\Delta x$ . Szczególnie istotny jest wzór na  $B'$ , z którego wynika, że obciążenie głośni przez tor głosowy powoduje pojawienie się dwóch składników o przeciwnych sobie znakach. Pojawienie się składnika o ujemnym znaku umożliwia wejście fałdów w stan oscylacji. Składnik ten reprezentuje inertancyjne obciążenie głośni przez tor głosowy. Natomiast dodatni składnik powoduje wzrost tłumienia  $B'$  w układzie i utrudnia powstanie oscylacji. Porównując ze sobą zależności na  $B'$  (z 6.88 i 6.73) można zauważyć, że inertancja toru głosowego spełnia podobną rolę jaką odgrywało w poprzednim modelu opóźnienie  $\tau$  w propagacji fali powierzchniowej w tkance śluzówkowej, między dolnym i górnym brzegiem fałdu głosowego. Inertancja powietrza znajdującego się w części ponadkrtaniowej powoduje powstanie opóźnienia w odpowiedzi systemu, zgodnie z kierunkiem przepływu. Tak więc zależność między  $I_t$  i  $\tau$  jest nie tylko intuicyjna. W istocie kolumna powietrza ponad fałdami oddziałuje jak górna masa w modelu dwumasowym, opóźniając zmianę ciśnienia ponad masą dolną.

Warunek  $B'=0$ , aby w układzie mogły powstać oscylacje prowadzi wprost do następującej zależności

$$v_{sr} = \frac{B}{2I_s I_t} + \Delta \omega_t b(x_0 + x_{st}), \quad 6.92$$

gdzie  $\Delta\omega_1=R_1/I_1$  jest szerokością pierwszego formantu toru głosowego. Przy jego zadanej konfiguracji i określonych parametrach źródła krztaniowego prędkość  $v_{sr}$  jest minimalną prędkością cząsteczkową (liniową) niezbędną do powstania w układzie oscylacji. Progowe ciśnienie podgłośniowe  $P_L$  i prędkość cząsteczkowa są bezpośrednio ze sobą związane. Warto przy tym zwrócić uwagę, że funkcje prędkości (Rys. 6.22) mają bardzo mały rozrzut, co sugeruje, że  $v_{sr}$  jest prawie niezależne od prefonacyjnej szerokości szpary głośni, a więc tym samym od wielkości przepływu. Z zależności (6.90) wynika, że  $v_{sr}$  jest zależne od  $x_0$  w sposób uchwytty tylko dla większych wymiarów szpary głośni. W innych przypadkach, obowiązuje prosta zależność kwadratowa

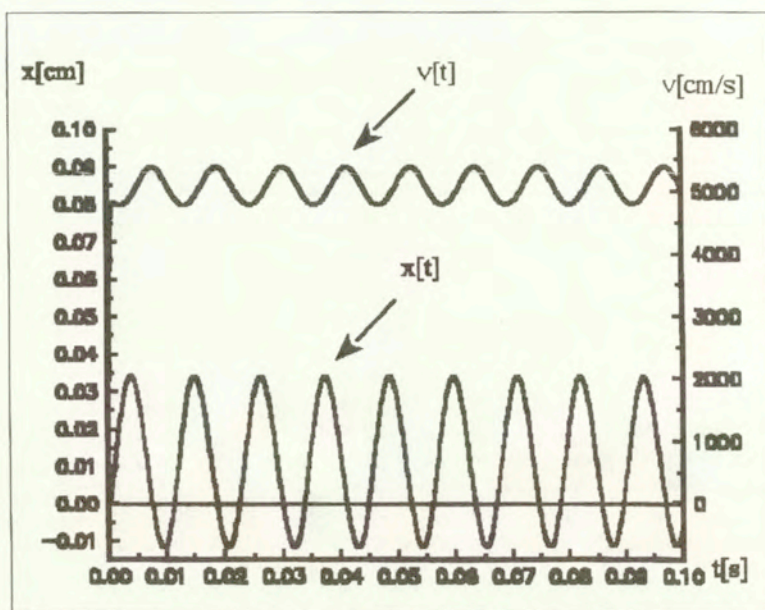
$$P_L = \frac{\rho}{2} k_1 v_{sr}^2.$$

Z równania (6.92) wynika, że zwiększenie długości fałdu  $l_g$  może pociągnąć za sobą obniżenie progowej prędkości liniowej i w podobny sposób, wzrost grubości fałdu powoduje obniżenie progowego ciśnienia pobudzającego, tak jak to miało miejsce w przypadku asynchronicznego ruchu brzegów ścian fałdów (wzór 6.75). Dodatkowo, wzrost składowej inertancji wejściowej toru głosowego obniża progową wartość prędkości cząsteczkowej, chociaż wzrost szerokości pasma formantowego zwiększa ją.

Współczynnik  $b$  występujący we wzorze (6.92), a także we wcześniejszych zależnościach (6.87 i 6.88) wiąże ze sobą składowe oscylacyjne prędkości cząsteczkowej z prędkością ruchu ścianek fałdów (patrz wzór 6.92). Metoda prób i błędów została zastosowana do wyznaczenia wielkości tego współczynnika. Metodę Runge-Kutta zastosowano do numerycznego rozwiązania równań (6.84) i (6.86), w celu wyznaczenia przebiegów czasowych przemieszczenia i prędkości cząsteczkowej.

Na Rys. 6.23 przedstawiono wyniki rozwiązań, przy takich samych parametrach toru głosowego i słuzówki. Podobnie jak w ostatnim przykładzie, ciśnienie  $P_L$  zostało tak dobrane, aby nie wystąpiło malenie lub wzrost amplitudy oscylacji. Ciśnienie to wyniosło 18 hPa. Warto zwrócić uwagę, że składowa oscylacyjna prędkości cząsteczkowej  $\Delta v$  jest bardzo mała w porównaniu z jej składową statyczną  $v_{sr}$  ( $\Delta v=3.3$  m/s, zaś  $v_{sr}=50$  m/s). Wartości te są zgodne z uprzednio otrzymanymi w sposób iteracyjny wynikami -  $v_{sr}=49.63$  m/s i  $x_{sr}=0.011$  cm.

Natomiast założenie, że pochodna  $\Delta x$  jest przesunięta o  $180^\circ$  jest przesunięta względem  $\Delta v$  (zależność 6.87) nie sprawdziło się. Obserwowane na Rys. 6.23 przesunięcie wynosi ok.  $150^\circ$ . Pociąga to za sobą bez wątpienia błąd w oszacowaniu progowych wartości ciśnień pobudzających, jednakże, jak dalej będzie to wykazane, błędy te nie są zbyt duże. Estymata współczynnika  $b$  z Rys. 6.23 wynosi  $\omega\Delta x_a/\Delta v_a=30$ , gdzie  $\omega$  jest pulsacją równą  $2\pi\times 88$  Hz,  $\Delta x_a$  - amplitudą przemieszczenia równą 0.02 cm, zaś  $\Delta v_a$  - amplitudą prędkości cząsteczkowej (330 cm/s).



Rys. 6.23 Przebiegi czasowe przemieszczenia  $x(t)$  i prędkości cząsteczkowej  $v(t)$  wyznaczone z zależności 6.84 i 6.86.

Dokładniejsze oszacowanie wartości współczynnika  $b$  podał Titze [TITZE,1988]. Dla różnych wartości  $B$  i  $x_0$  wyznaczał progowe wartości ciśnienia pobudzającego poprzez symulację. Następnie obliczył  $P_L$  jako funkcje zmiennej czynnika  $B/(2l_g I_t)$ . Zmienna ta ma wymiar prędkości i może być traktowana jako skuteczna prędkość tłumieniowa dla wartości progowej ciśnienia, przy obciążeniu bezstratnym torem głosowym (wzór 6.92 przy  $\Delta\omega_1 b=0$ ). Powyższa skuteczna prędkość tłumieniowa odpowiada skutecznemu ciśnieniu tłumieniowemu ( $B_c$ ). Zastosowana procedura polega na wyznaczaniu wartości  $P_L$  poprzez obliczanie  $x_{sr}$  i  $v_{sr}$  z zależności (6.89 i 6.90) i znajdowaniu odpowiadającej mu wartości  $B/(2l_g I_t)$ .

Warto wspomnieć, że choć teoretycznie sprzężenie inertancyjne źródła z torem głosowym stanowi pewną alternatywę dla mechanizmu fali powierzchniowej rozchodzącej się w słuzówce fałdu, by źródło krtaniowe weszło w stan oscylacji, to w praktyce ma ono ograniczony zasięg. Tym nie mniej, w rehabilitacji osób po usunięciu fałdów głosowych (u tzw. bezkrtaniowców) mechanizm sprzężenia inertancyjnego wydaje się być jedyną alternatywą przy generowaniu dźwięcznego pobudzenia toru głosowego. Jest to wykorzystywane przy nauczaniu bezkrtaniowców poprzez w pewnym okresie rehabilitacji tzw. dźwięcznego odbicia (ructus).

Oczywiście sprzężenie bezwładnościowe, którego wielkość zależy głównie od masy akustycznej toru głosowego, przez który przepływa powietrze, silnie zależy od jego konfiguracji. Zwłaszcza obecność przewężeń w części gardłowej pociąga za sobą duże wahania obciążenia inertancyjnego. Jednakże głównym problemem jest wyznaczenie składowej opornościowej. Dla przypadku wyidealizowanego, to jest, gdy szerokość pasma formantowego jest równa zeru ( $R_1=0$ ), wyrażenie na progową wartość ciśnienia pobudzającego otrzymuje się z podstawienia wzoru (6.92) do (6.90)

$$P_L = \frac{\rho}{2} k_1 \left( \frac{B}{2l_n I_1} \right)^2, \quad (R_1 = 0) \quad 6.93$$

dla wszystkich wartości  $x_0$ .

Dla typowych wartości parametrów źródła, wcześniej już podanych, otrzymuje się że  $P_L=12.3$  hPa. Jednakże straty w torze głosowym nie są pomijalnie małe. W sumie obciążenie bezwładnościowe może co najwyżej być elementem towarzyszącym dominującemu mechanizmowi pobudzania fałdów, wywołanemu rozchodzeniem się fali powierzchniowej w śluzówce okrywającej fałdy, i co najwyżej powodować ewentualnie obniżenie wartości progowej ciśnienia pobudzającego. Natomiast w przypadku normalnej fonacji nie może stać się dominującym mechanizmem.

Z zależności (6.88) dotyczącej sztywności, wynika że oporność akustyczna toru głosowego przekształca się na składową o ujemnym znaku mechanicznej sztywności fałdów głosowych. Wcześniej pokazano, że akustyczna inertancja toru głosowego transformuje się po stronie mechanicznej w ujemną oporność mechaniczną fałdów głosowych (w tłumieniu). Jest to wynikiem istnienia różnicy w ruchu powietrza i śluzówki rzędu pochodnej (dokładnie o jeden rząd). W przypadku braku ruchu śluzówki fałdu mamy do czynienia ze stałym przepływem powietrza przez głośnię, natomiast stały ruch śluzówki powoduje przyspieszanie ruchu powietrza.

Ilościowo, ujemna składowa sztywności może być znaczna. Dla powyższych wartości parametrów źródła, składowa wynikająca z oddziaływania toru głosowego w przeliczeniu na jednostkę powierzchni ścianki fałdu wynosi  $43\ 100$  dyn/cm<sup>3</sup>, co jest znaczne w porównaniu z ich sztywnością mechaniczną ( $200\ 000$  dyn/cm<sup>3</sup>). Dla małej amplitudy oscylacji częstotliwość ich drgań jest

$$F_0 = \frac{1}{2\pi} \sqrt{\frac{K'}{M'}}, \quad 6.94$$

gdzie  $M'$  jest określone wzorem (6.88). W przypadku istnienia sprzężenia opornościowo-bezwładnościowego częstotliwość podstawowa może tylko ulec zmniejszeniu, ponieważ  $K'$  jest zawsze mniejsze od  $K$ , zaś  $M' > M$ . Dla przyjętych wartości parametrów źródła krtaniowego częstotliwość podstawowa maleje ze 103 Hz

(bez obciążenia  $R_c=I_c=0$ ) do 86.6 Hz. Wartość ta pokrywa się z częstotliwością przebiegu przedstawionego na Rys. 6.23. To obniżanie ("ściągnięcie w dół") częstotliwości podstawowej było już wcześniej wykazane w pracy (ISHIZAKA&FLANAGAN,1972), dla rur różnej długości przyłączanych do modeli jedno- i dwu-masowych. Model jednomasowy był bardziej wrażliwy na "ściągnięcie w dół" częstotliwości  $F_0$  i tylko przy obciążeniu inertancyjnym mógł wejść w stan oscylacji.

Przedstawiona metoda opisanie zasadniczych własności oscylacyjnych fałdów głosowych, oparta na koncepcji rozchodzenia się w słuzówce okrywającej fałdy fali powierzchniowej, jest w znacznej mierze oparta na badaniach Titzego [TITZE,1988;1994;TITZE&STRONG,1975], który pierwszy podjął próbę uwzględnienia zjawisk nieliniowych związanych z ruchem drgającym fałdów głosowych. Przyjęcie założenia, że fala powierzchniowa rozchodzi się w kierunku zgodnym z kierunkiem przepływu strumienia powietrza przez głośnię, umożliwia wyprowadzenie równań różniczkowych ruchu fałdów o dość prostej postaci. Po zlinearyzowaniu tych równań dla przypadku oscylacji o małej amplitudzie, uzyskuje się rozwiązania o przybliżonej postaci dla układu o skutecznej masie, tłumieniu i sztywności.

Otrzymane wyniki nie różnią się zasadniczo od tych otrzymanych dla modelu dwumasowego. Wydaje się, że dominujący mechanizm uzyskania sił wymuszających drgania zależnych od prędkości przepływu powstaje w wyniku przesunięcia czasowego między oddziaływaniem ciśnienia na dolną i górną krawędź ściany fałdu, a dokładniej ujmując, opóźnionego oddziaływania ciśnienia na dolną krawędź ścianki fałdu, co jest wywołane opóźnionym ruchem górnej krawędzi względem dolnej. W przypadku modelu o małej amplitudzie drgań, opóźnienie to wynika ze skończonej prędkości propagacji fali powierzchniowej w słuzówce, natomiast w modelu 2-masowym, mamy do czynienia z dwoma sprzężonymi ze sobą oscylatorami sterowanymi przepływem strumienia powietrza, który utrzymuje górną i dolną masę powietrza w pewnym przesunięciu fazowym względem siebie.

Propagacja fali powierzchniowej musi iść z dołu do góry. Dla przeciwnego kierunku propagacji źródło krztaniowe nie wejdzie w stan oscylacji, tym bardziej, że kierunek przechodzenia energii byłby wówczas od słuzówki do strumienia powietrza. Stwierdzenie to, pociąga za sobą, że w modelu 2-masowym dolna masa musi zawsze w ruchu wyprzedzać górną. Tylko wówczas powstaje zjawisko ujemnego tłumienia, które jak wykazano w modelu o małej amplitudzie oscylacji jest odwrotnie proporcjonalne do prędkości propagacji fali powierzchniowej (lub wprost do opóźnienia między ruchami górnej i dolnej krawędzi ściany fałdów.

Istnienie opóźnienia w oddziaływaniu na krawędzie fałdów było również wykorzystane przy sformalizowaniu inertancyjnego oddziaływania toru głosowego na

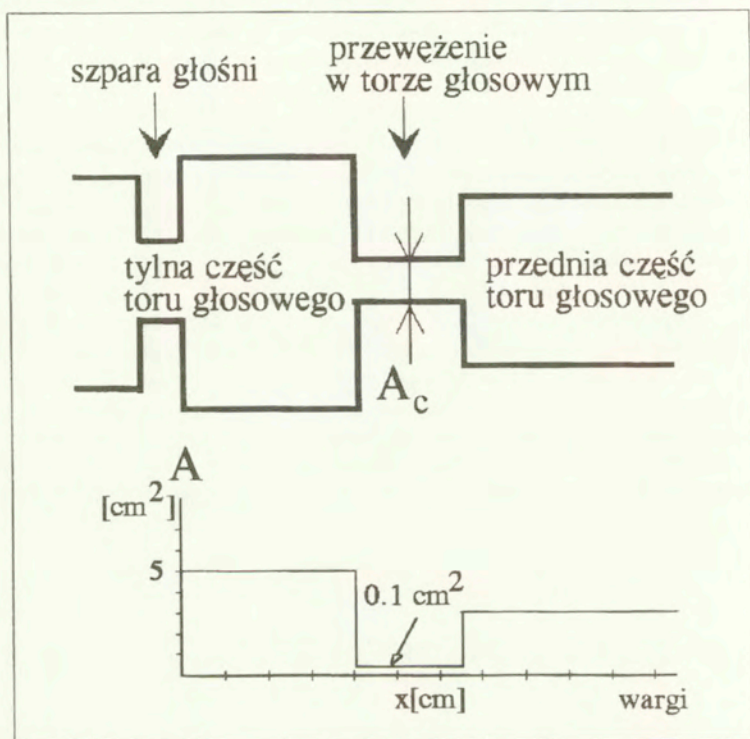


źródło krtaniowe. Ponieważ narastanie i gwałtowne zanikanie strumienia powietrza przepływającego przez głośnię i tor głosowy następują z pewnym opóźnieniem względem ruchu śluzówki, a więc ciśnienie wymuszające drgania jest większe podczas rozwierania się głośni, niż podczas jej zamykania. Ta zależna od prędkości strumienia powietrza siła wymuszająca przekazuje energię do fałdów głosowych i pomaga podtrzymywać stan ich oscylacji. Jak wykazano, tłumienie skuteczne maleje wraz ze wzrostem inertancji toru głosowego i rośnie, gdy oporność toru również rośnie. Wyłącznie podatnościowe obciążenie źródła przez tor głosowy uniemożliwia całkowicie powstanie oscylacji.

Na zakończenie rozdziału, warto podkreślić, że modelowanie źródła krtaniowego, a zwłaszcza ruchu fałdów głosowych oraz badanie kształtu fali pobudzenia krtaniowego (przebiegu czasowego zmian wielkości wydatku strumienia powietrza) jest ważne nie tylko ze względu na wyjaśnienie mechanizmu powstawania drgań fałdów głosowych przy artykulacji głosek dźwięcznych. Jak wykazują badania (m.in. [MARASEK,1997; STEVENS,1998]) kształt fali pobudzenia krtaniowego wpływa w decydujący sposób na jakość brzmienia głosu, przede wszystkim na jego barwę. Również wszelkiego rodzaju patologie fałdów głosowych (np. porażenia jedno- lub dwustronne, częściowe niedomykanie się fałdów, guzki itp.) powodują istotną zmianę kształtu fali pobudzenia krtaniowego. Jednakże w tym przypadku, do jednoznacznej oceny samej patologii źródła krtaniowego konieczna jest bardziej dokładna analiza ruchu fałdów głosowych, a tę najlepiej jest realizować przy zastosowaniu elektrogłotograficznej metody (metoda EGG) oceny zmian wielkości otworu głośni, przebiegu zmian w czasie. Tylko bowiem ta metoda umożliwi bezpośrednią obserwację przebiegu cyklu zwierania się i rozwierania się fałdów głosowych podczas mówienia, nie wprowadzając przy tym dodatkowych utrudnień w procesie artykulacji [FOURCIN&ABBERTON,1971;1977; CHILDERS i in.,1986; MARASEK,1997].

## 7. Źródło pobudzenia bezdźwięcznego

Blisko 40% dźwięków mowy polskiej jest wymawiane przy współdziałaniu turbulentnego źródła szumowego powstającego w przewężeniu (lub u jego wylotu) utworzonym w torze głosowym (tak jak to ma miejsce przy artykulacji niektórych głosek trących), czy wskutek opływania przeszkody w nim powstałej podczas artykulacji tego typu głosek. Dokładne, analityczne ustalenie położenia źródła szumowego w torze głosowym oraz wyznaczenie jego parametrów fizycznych nie jest łatwe. Na Rys. 7.1 przedstawiony jest model toru głosowego odwzorowujący jego przybliżoną geometrię podczas artykulacji spółgłosek trących bezdźwięcznych.



Rys. 7.1 Uproszczony model toru głosowego i jego funkcja przekroju dla spółgłosek trących.

W przypadku artykulacji głosek trących można przyjąć, że ich charakter jest statyczny i w pierwszym przybliżeniu powstają w wyniku pobudzenia przedniej części toru głosowego przez źródło szumu turbulentnego umiejscowione w

przewężeniu toru głosowego lub w pobliżu wylotu powietrza z przewężenia [FANT,1960]. Źródło to powstaje, gdy podczas przepływu powietrza z tylnej części toru głosowego przez przewężenie następuje tak znaczne zwiększenie prędkości cząsteczkowej powietrza, wskutek czego struga powietrza staje się nieciągła i tworzą się w niej zawirowania i turbulencje. Miejsce ich powstania zależy zarówno od wielkości strumienia powietrza, geometrii przewężenia oraz lepkości powierzchni jego ścianek. Turbulencje mogą również powstać w przypadku uderzenia strumienia powietrza z odpowiednią prędkością o przeszkodę, np. o zęby, jak to ma miejsce w przypadku artykulacji spółgłoski /f/.

Wymiary przewężenia znacznie zależą od artykułowanej głoski, zarówno jego długość jak i wielkość jego przekroju poprzecznego. W przypadku, gdy szczelina jest utworzona między językiem i podniebieniem długość jej może sięgać kilku centymetrów. Gdy powietrze przepływa przez przewężenie, to spadki ciśnienia powstają przede wszystkim wskutek strat u samego wlotu oraz wylotu z przewężenia, niż w nim samym na jego długości. Źródłem strat są zawirowania powstające właśnie w pobliżu krawędzi wlotu i wylotu przewężenia. Na drodze eksperymentalnej stwierdzono [STEVENS,1971], że spadek ciśnienia powstały u zwężenia i rozszerzenia na krańcach przewężenia jest proporcjonalny ciśnienia dynamicznego  $\frac{1}{2}\rho V^2$  w przewężeniu, gdzie  $\rho$  jest gęstością powietrza, zaś  $V$  prędkością liniową powietrza w przewężeniu. W dalszych rozważaniach wygodniej będzie posługiwać się prędkością objętościową  $U = V \cdot A$ , gdzie  $A$  jest powierzchnią przekroju poprzecznego przewężenia. Stąd spadek ciśnienia wywołany stratami u wlotu i wylotu przewężenia można określić wzorem

$$\Delta p_L = k_L (\rho U^2 / 2A^2), \quad 7.1$$

gdzie  $k_L$  jest pewną stałą, której wielkość zależy od kształtu wlotu i wylotu przewężenia. Na podstawie przeprowadzonych badań modelowych stwierdzono [STEVENS,1971], że w przypadku łagodnego przejścia od szerokiego przekroju poprzecznego toru głosowego do utworzonego w nim wąskiego przewężenia wartość tego współczynnika może być bardzo mała i nie przekraczać 0.05, natomiast przy skokowej zmianie przekrojów poprzecznych współczynnik ten zawiera się w zakresie 1,2-1,7. Jednakże badania porównawcze wykonane dla modeli krtani [BERG VAN DEN i in., 1957] o wlocie do krtani zmieniającym się w sposób płynny i skokowym wylocie wykazały, że współczynnik ten wynosi około 0.87. Należy więc oczekiwać, że przy łagodnie zmieniającym się wlocie i wylocie szczeliny współczynnik ten będzie jeszcze mniejszy. Wydaje się więc, że dla wszystkich przewężeń jakie mogą być utworzone w torze głosowym podczas artykulacji spółgłosek trących, w tym również przez wargi i język, można przyjąć jako wartość średnią tego współczynnika 0.9 [STEVENS,1971].

Dodatkowo, oprócz strat na wlocie i wylocie z przewężenia powodujących spadek na nim ciśnienia, występuje dodatkowy spadek wywołany opornością  $R$  przepływu laminarnego przez przewężenie. Dla prostokątnego przekroju szczeliny opór ten został określony wzorem (6.18) przy omawianiu modelu jednomasowego źródła krtaniowego. W przypadku, gdy przekrój przewężenia jest kołowy o średnicy  $d$  wówczas oporność ta jest określona wzorem

$$R = 12\mu l / \pi d^4. \quad 7.2$$

Drugi składnik spadku ciśnienia w przewężeniu jest pochodzenia kinetycznego i jest wprost proporcjonalny do prędkości objętościowej przepływu powietrza  $U$  i wynosi  $RU$ . Jednakże w większości przypadków (to jest, gdy powierzchnia przekroju szczeliny jest większa od  $0.05 \text{ cm}^2$ , lub dla niezbyt małych prędkości przepływu) jest on pomijalnie mały w stosunku do spadku ciśnienia wywołanego oporem lepkościowym powietrza. Wówczas można przyjąć, że spadek ciśnienia na przewężeniu utworzonym w torze głosowym jest określony wzorem

$$\Delta p = 0.9 \frac{\rho U^2}{2A^2}. \quad 7.3$$

Ogólnie biorąc podczas artykulacji większości spółgłosek trących, warunki przepływu powietrza przez tor głosowy są wyznaczone przez dwa w nim istniejące przewężenia – przez szparę głośni oraz przez przewężenie utworzone w części ponadkrtaniowej, między językiem i podniebieniem. W przypadku, gdy przewężenie utworzone w części gardłowo-ustnej toru głosowego jest znacznie mniejsze od przekroju szpary głośni wówczas spadek ciśnienia  $\Delta p$  w nim jest praktycznie równe ciśnieniu podgłośniowemu  $p_g$ , to jest ok. 8 cm  $\text{H}_2\text{O}$ . Również dla małych prędkości przepływu, to jest poniżej 200-300  $\text{cm}^3/\text{s}$ , ponadkrtaniowe ciśnienie jest równe ciśnieniu podgłośniowemu, ponieważ spadek ciśnienia na drodze od płuc do krtani jest pomijalnie mały. Dla dużych prędkości przepływów spadek ciśnienia na szparze głośni wzrasta, ale ciśnienie ponadkrtaniowe jest porównywalne z ciśnieniem podgłośniowym. W przypadku, gdy powierzchnia szpary głośni jest bardzo mała w stosunku do przewężenia, warunki przepływu powietrza są określone przez szparę głośni i cały spadek ciśnienia przepływającego powietrza jest w jej obrębie.

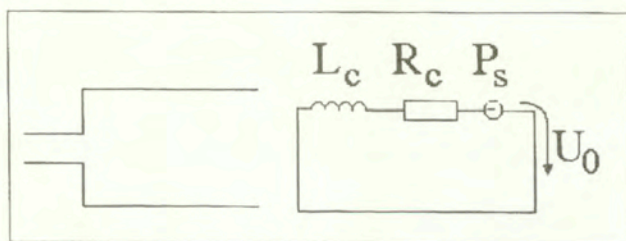
Gdy przepływ powietrza przez przewężenie odbywa się ze znaczną prędkością, to powstają w nim turbulencje, powodujące fluktuacje prędkości strumienia powietrza. Turbulencje te występują przy artykulacji głošek trących. Nieco inny charakter mają turbulencje powstające, gdy strumień powietrza uderza o krawędź powierzchni ciała stałego. Oddziaływujące wówczas siły powodują powstanie szeregu źródeł o postaci dipolowej.

Badania eksperymentalne nad turbulencjami powstającymi podczas przepływu powietrza przez przeszkodę umieszczoną w przewodzie cylindrycznym wykazały [STEVENS, 1971], że moc akustyczna promieniowana z takiego źródła modelowana za pomocą dipola jest proporcjonalna do czwartej potęgi prędkości objętościowej przepływu  $U^4$ . W przypadku umieszczenia przeszkody w rurze nieskończenie długiej z jednej strony, zaś z drugiej zamkniętej opornością promieniowania, wówczas promieniowana moc staje się proporcjonalna do szóstej potęgi prędkości przepływu  $U^6$ . Ponieważ spadek ciśnienia na przeszkodzie (czy przewężeniu)  $\Delta p \sim \frac{1}{2}\rho V^2$ , więc promieniowana moc może być określona wzorem

$$P = \frac{k\Delta p^3 d^2}{\rho^2 c^3}, \quad 7.4$$

gdzie  $d$  jest średnicą rury,  $c$  prędkością rozchodzenia się dźwięku.

W przypadku artykulacji głosek trących źródło szumowe może znajdować się bądź u wylotu przewężenia, bądź na przeszkodzie, na którą pada strumień powietrza. Na Rys. 7.2 przedstawiono prosty model odpowiadający w przybliżeniu artykulacji głoski /x/.



Rys. 7-2 Model artykulacji głoski trącej o przewężeniu utworzonym w tylnej części toru głosowego ([LEE, 1990]).

Na Rys. 7.3 przedstawiono model równoważny artykulacji głoski, dla której uproszczoną konfigurację toru głosowego przedstawiono na Rys. 7-1. W obu przypadkach przedstawionych na Rys. 7-2 i 7-3 własności przewężenia są modelowane poprzez szeregowe połączenie indukcyjności  $L_c$  oraz oporności  $R_c$ . W pierwszym przypadku funkcja transmitancji układu, definiowana jako stosunek prędkości objętościowej u wylotu ust do ciśnienia generowanego przez źródło, ma postać

$$\frac{U_0}{P_s} = \frac{1}{sL_c + R_c}, \quad 7.5$$

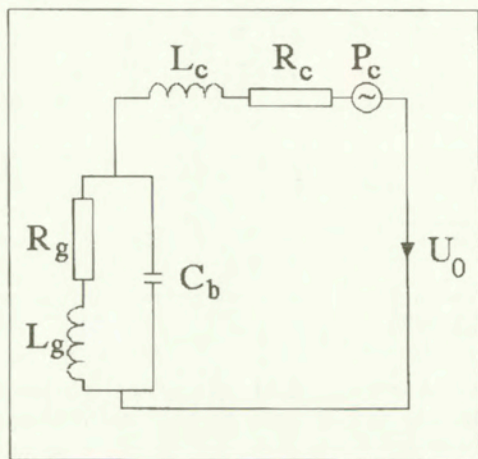
gdzie  $L_c$  oraz  $R_c$  są odpowiednio indukcyjnością i opornością przewężenia. Są one określone poniższymi wzorami:

$$L_c = \rho \frac{l_c}{A_c} \quad 7.5a$$

oraz

$$R_c = \frac{k_c \cdot \rho \cdot v_c}{A_c} + \frac{32 \cdot \mu \cdot l_c}{A_c \cdot d_c^2}, \quad 7.5b$$

gdzie  $A_c$  jest powierzchnią przekroju poprzecznego przewężenia ( $A_c = 0.25 \cdot \pi \cdot d_c^2$ ,  $l_c$  jego długością,  $k_c$  współczynnikiem kształtu, zaś  $v_c$  prędkością cząsteczkową strumienia powietrza w przewężeniu. Drugi składnik we wzorze (7.5b) reprezentuje straty wskutek tarcia strumienia powietrza o ścianki przewężenia, które dla wymiarów przewężenia spotykanych dla większości głosek trących są pomijalnie małe w stosunku do pierwszego składnika reprezentującego straty o charakterze kinetycznym. Funkcja transmitancji opisana wzorem (7.5) ma tylko jeden biegun rzeczywisty dla  $s = -R_c/L_c$  odpowiadający górnej częstotliwości przenoszenia przewężenia.



Rys. 7-3 Model równoważny konfiguracji toru głosowego przedstawionej na Rys. 7-1 ([LEE,1990]).

Gdy przewężenie znajduje się pośrodku toru głosowego, tak jak to przedstawiono na Rys. 7.1, wówczas funkcja transmitancji układu równoważnego (Rys. 7.3) ma postać

$$\frac{U_0}{P_s} = \frac{s^2 + s \cdot (R_g/L_g) + 1/(L_g \cdot C_b)}{Z}, \quad 7.6$$

gdzie

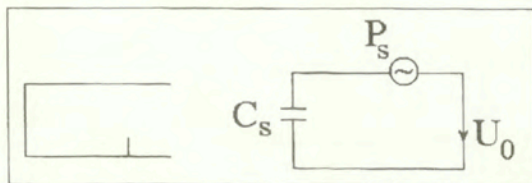
$$Z = L_c \cdot \left[ s^3 + s^2 \frac{(R_g \cdot L_c + R_c \cdot L_g)}{L_g \cdot L_c} + s \cdot \frac{L_c + L_g + R_c \cdot R_g \cdot C_b}{L_g \cdot L_c \cdot C_b} + \frac{R_c + R_g}{L_g \cdot L_c \cdot C_b} \right]$$

Dla modelu bezstratnego, to jest gdy  $R_c = 0$  oraz  $R_g = \infty$  funkcja transmitancji redukuje się do postaci

$$\frac{U_0}{P_s} = \frac{1}{L_c} \cdot \frac{s}{s^2 + \frac{1}{L_c \cdot C_b}}. \quad 7.7$$

Funkcja ta ma jeden biegun dla  $s=0$  i dwa bieguny odpowiadające rezonansowi Helmholtza układu utworzonego z przewężenia i komory utworzonej za nim, w tylnej części toru głosowego.

Przy artykulacji większości spółgłosek trących źródło pobudzenia szumowego powstaje również na przeszkodzie utworzonej na drodze strumienia powietrza płynącego w torze głosowym z prędkością odpowiednio dużą, by przepływ stał się turbulentny. Dla najprostszego przypadku, to jest rury z krótką przeszkodą wewnątrz niej dla przepływającego powietrza (Rys. 7.4).



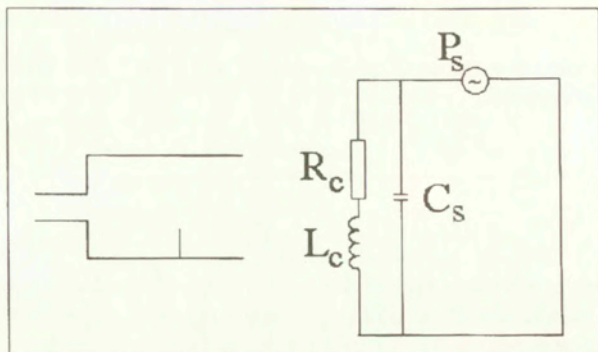
Rys. 7.4 Model rury ze źródłem turbulencji w pobliżu przeszkody

W układzie tym impedancja przewężenia jest nieskończenie duża. W zakresie niskich częstotliwości wszystkie oporności i indukcyjności są bliskie zeru, a pojemność tej części toru głosowego, która znajduje się od strony warg przed przeszkodą jest zwarta. Funkcja transmitancji układu równoważnego ma więc postać

$$\frac{U_0}{P_s} = s \cdot C_s, \quad 7.8$$

gdzie  $C_s$  jest podatnością części rury znajdującej się między tylną ścianą i źródłem turbulentnym. Funkcja transmitancji ma jedno zero dla  $s=0$ .

W przypadku gdy impedancja przewężenia ma pewną skończoną wartość, wówczas model pobudzenia bezdźwięcznego toru głosowego ma postać przedstawioną na Rys. 7.5.



Rys. 7.5 Model układu z przewężeniem i przeszkodą ([LEE,1990]).

Funkcja transmitancji tego układu ma postać

$$\frac{U_0}{P_s} = C_s \cdot \frac{s^2 + s \cdot \frac{R_c}{L_c} + \frac{1}{L_c \cdot C_s}}{s + \frac{R_c}{L_c}}, \quad 7.9$$

gdzie  $L_c$  jest takie same jak we wzorze (7.5a), natomiast oporność strat  $R_c$  określona wzorem (7.5b) jest powiększona o dodatkowy składnik uwzględniający straty turbulენტne wewnątrz przewężenia i ma postać

$$R_c = \frac{k_c \cdot \rho \cdot v_c}{A_c} + \frac{32 \cdot \mu \cdot l_c}{A_c \cdot d_c^2} + \frac{\xi \cdot \rho \cdot l_c \cdot v_c}{2 \cdot A_c \cdot d_c}, \quad 7.9a$$

gdzie

$$\xi = 0.3164 \cdot R_e^{-0.25},$$



$A_c$  - jest powierzchnią przekroju poprzecznego przewężenia i jest równe  $A_c = 0.25 \cdot \pi \cdot d_c^2$

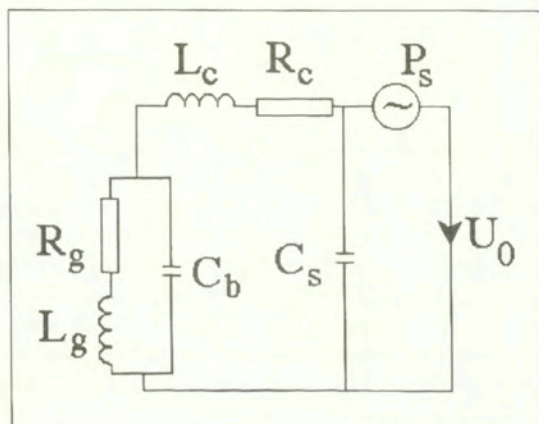
$L_c$  - długość szczeliny

$R_e$  - liczba Reynoldsa równa [FANT,1960]

$$R_e = v_c \cdot \frac{d_c}{\nu} \quad 7.9b$$

gdzie  $\nu = \mu/D$  jest kinetycznym współczynnikiem lepkościowym i  $v_c$  średnią prędkością cząsteczkową powietrza przepływającego przez szczelinę. Dla przeciętnych rozmiarów szczeliny tworzonej podczas artykulacji spółgłosek trących suma strat wskutek tarcia i turbulencji strumienia powietrza jest znacznie mniejsza od strat o charakterze kinetycznym i w większości przypadków można je pominąć.

Zależność (7.9) ma dwa zera odpowiadające rezonansowi Helmholtza (układ szczelina – jama przednia toru głosowego) oraz jeden biegun rzeczywisty dla  $s = -R_e/L_c$ , który określa szerokość krzywej rezonansowej. Gdy w modelu uwzględnić wpływ tylnej komory w torze głosowym oraz impedancję głośni, tak jak to przedstawiono na Rys. 7.3, to model równoważny stanowi kombinację modelu z Rys. 7.3 i 7.5 (patrz Rys. 7.6).



Rys. 7.6 Model równoważny konfiguracji toru z Rys. 7.1 z uwzględnieniem przeszkody usytuowanej tak jak to przedstawiono na Rys. 7.5 ([LEE,1990]).

Na ogół można założyć, że  $L_g$  jest pomijalnie mała, wówczas zależność (7.9) dla modelu równoważnego przedstawionego na Rys. 7.6 przybiera postać

$$\frac{U_0}{P_c} = C_s \frac{A}{B}, \quad 7.10$$

gdzie

$$A = s^3 + s^2 \frac{L_c + R_c \cdot R_g \cdot C_b}{L_c \cdot R_g \cdot C_b} + s \frac{C_s \cdot R_g + C_s \cdot R_c + C_b \cdot R_g}{C_s \cdot L_c \cdot C_b \cdot R_g} + \frac{1}{C_s \cdot L_c \cdot C_b \cdot R_g}$$

oraz

$$B = s^2 + s \frac{L_c + R_c \cdot R_g \cdot C_b}{L_c \cdot R_g \cdot C_b} + \frac{R_c + R_g}{L_c \cdot C_b \cdot R_g}.$$

Powyższa zależność ma trzy bieguny i dwa zera. W celu określenia częstotliwości charakterystycznych, najłatwiej będzie pominąć w pierwszym przybliżeniu straty, czyli gdy podstawi się  $R_c=0$  i  $R_g=\infty$ , wówczas

$$\frac{U_0}{P_c} = C_s \frac{s^2 + s \cdot \frac{C_s + C_b}{L_c \cdot C_b \cdot C_s}}{s^2 + \frac{1}{L_c \cdot C_b}}. \quad 7.11$$

Z zależności (7.11) widać wyraźnie, że dwa bieguny powyższego równania wyznaczają częstotliwość rezonatora Helmholtza utworzonego przez szczelinę i tylną komorę (jej podatność  $C_b$ ). Ponadto, w powyższej zależności występuje pojedyncze zero (dla  $s=0$ ) oraz para zer związana z układem rezonatorów utworzonego przez tylną i przednią część, przy czym ta druga znajduje się między źródłem szumowym i szczeliną (podatności  $C_b$  i  $C_s$ ). Przedstawiony model pobudzenia szumowego jest zbliżony do rzeczywistych warunków przy artykulacji spółgłoski /s/. Podczas jej wymawiania struga powietrza wyphywająca ze szczeliny utworzonej w przedniej części toru głosowego między językiem i podniebieniem twardym uderza o krawędź górnych zębów przednich.

Analizując charakterystyki częstotliwościowe przedstawionych modeli powstawania pobudzenia szumowego, można stwierdzić, że w przypadku gdy:

- a) źródło pobudzenia szumowego znajduje się u wylotu szczeliny lub pośrodku niej, funkcja źródła ma o jeden biegun więcej niż zer, w wyniku czego obwiednia widma pobudzenia maleje ze wzrostem częstotliwości w stosunku  $-6$  dB/okt;

- b) źródło pobudzenia szumowego jest usytuowane w pobliżu przeszkody, funkcja źródła ma o jedno zero więcej niż biegunów, w wyniku czego obwiednia jego widma wzrasta asymptotycznie z nachyleniem  $+6$  dB/okt.

W obu przypadkach na osi rzeczywistej funkcji transmitancji źródła znajduje się co najmniej jedno zero lub biegun.

Wyżej wymienione własności modeli zastępczych źródeł pobudzenia szumowego wynikają z charakteru jego impedancji. Modele równoważne dla źródła powstającego na przeszkodzie (Rys. 7.4, 7.5 i 7.6), wskazują, że dominująca składowa impedancji źródła ma charakter pojemnościowy, natomiast w przypadku powstania źródła pobudzenia w przewężeniu lub u jego wylotu, składowa ta ma charakter indukcyjny.

## 8. Komputerowy model układu artykulacyjnego.

### 8.1 Wstęp

Przy modelowaniu własności źródła pobudzającego oraz układu artykulacyjnego dąży się przede wszystkim do odtworzenia konfiguracji geometrycznej toru głosowego, aby tak odwzorować jego funkcję transmitancji, że otrzymany w efekcie końcowym wyjściowy sygnał akustyczny był możliwie zbliżony do brzmienia modelowanej głoski. W początkowo prowadzonych badaniach, koncentrowano się głównie na możliwie dokładnym odwzorowaniu funkcji transmitancji za pomocą prostych układów akustycznych, składających się zazwyczaj z krótkich odcinków rur cylindrycznych, w wyniku czego otrzymana funkcja odwzorowująca konfigurację toru głosowego często w znaczny sposób różniła się od rzeczywistej. Powodowało to znaczne utrudnienie przy modelowaniu stanów przejściowych przy artykulacji głosek niestacjonarnych, a także modelowanie płynnych zmian tej konfiguracji przy artykulacji ciągu głosek.

Należy podkreślić, że przez wiele lat podstawą do modelowania konfiguracji geometrycznej toru głosowego były dwuwymiarowe zdjęcia rentgenowskie [FANT, 1960], a później kinorentgenograficzne (jak np. w [BOLDA&FOLDI,1987]), również dające tylko przebieg zmian przekroju wzdłużnego toru głosowego. Ostatnio odnotowuje się coraz szybszy postęp w wizualizacji geometrii toru głosowego, m.in. zastosowanie techniki MRI (patrz [BAER i in., 1991]) zaczyna umożliwiać coraz dokładniejszą w oparciu o trójwymiarowy obraz toru, choć jednak wciąż w dość ograniczonym zakresie, weryfikację tworzonych modeli artykulacyjnych poszczególnych głosek. Jednym z poważniejszych ograniczeń jest przede wszystkim rozgraniczenie rejestracji obrazu geometrii toru głosowego i wymawianego dźwięku mowy. Jest to spowodowane dużym poziomem zakłóceń wywoływanych przez aparaturę wizualizacyjną. Ponadto, wskutek stosunkowo długiego czasu skanowania (1 przekrój na 1 sekundę) konieczne jest wymawianie głosek w sposób ciągły w nienaturalnie długi sposób co wymaga od lektora idealnego utrzymywania konfiguracji toru głosowego podczas artykulacji danej głoski, w nienaturalnej pozycji leżącej z głową skierowaną ku górze. Następnie, konieczne jest osobne, ponowne nagranie dźwięku, przez naśladowanie oryginalnego sygnału zarejestrowanego wraz obrazem i podawanego przez słuchawki osobie, na której dokonywano eksperymentu wizualizacji toru głosowego. Jest oczywiste, że w eksperymentach tego typu mogą uczestniczyć tylko bardzo wprawne osoby. Pozycja leżąca osoby mówiącej powoduje drobne przesunięcia słuzówki toru głosowego wskutek efektu grawitacji, stosując technikę MRI występują trudności w dokładnym wyznaczeniu położenia linii zębów. Wszystko to powoduje, że jak na razie dokładność wyznaczania przekrojów poprzecznych toru głosowego techniką MRI nie jest jeszcze wystarczająco dokładna i wymaga udoskonalenia [DEMOLIN i in.,1995]. Tych wad nie ma technika zdjęć kinorentgenograficznych umożliwiającą bieżącą rejestrację zmian w czasie przekroju podłużnego toru głosowego podczas wymawiania nie tylko pojedynczych głosek, lecz

również dłuższych sekwencji wyrazowych wraz z sygnałem mikrofonowym. Ale niestety ta technika nosi poważne zagrożenie dla zdrowia osoby poddawanej tego typu eksperymentom i z tego powodu wykonywanie zdjęć kino rentgenograficznych zostało obecnie całkowicie zaniechane.

Pomimo bardzo złożonych warunków, w jakich dokonywana jest trójwymiarowa wizualizacja toru głosowego, otrzymane metodą MRI obrazy wyraźnie wskazują, że konfiguracja geometryczna toru głosowego w znaczny sposób odbiega od kształtu cylindrycznego, najczęściej przyjmowanego przy modelowaniu toru [np. FANT,1960; MERMELSTEIN, 1967; MRAYATI M. in., 1988; LIN, 1994; SCHOENTGEN & CIOCEA,1995]. Również prostsze metody trójwymiarowej wizualizacji (choć dające tylko częściowy obraz toru, jak to ma miejsce w przypadku metod ultradźwiękowych) wskazują, że geometria toru głosowego jest bardzo złożona. Z tego powodu, w wielu przypadkach, przy modelowaniu geometrii toru głosowego, nie modeluje się bezpośrednio kształtu przekrojów poprzecznych toru głosowego, lecz ich powierzchnie. Pociąga to za sobą konieczność przy modelowaniu sekwencji głosek lub głosek niestacjonarnych określenia czasowej funkcji zmian powierzchni poszczególnych przekrojów, co jest jak na razie ustalane w dość arbitralny sposób, zakłada się bądź zmiany o charakterze liniowym nie dające zbyt dobrych wyników, lub nieliniową funkcję zmian typu S, która w niektórych przypadkach sprawdza się nieco lepiej, niż funkcja liniowa. Ale funkcja strat toru głosowego, która zależy właśnie od kształtu przekroju jest wyznaczana wówczas w sposób dość arbitralny.

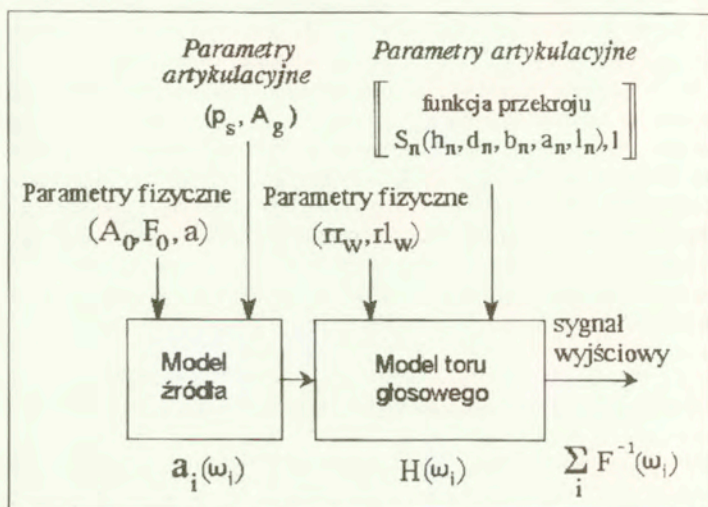
W niniejszym rozdziale pracy opisano wyniki modelowania w oparciu o aproksymację poszczególnych przekrojów poprzecznych toru za pomocą figur geometrycznych o różnym od kołowego kształtu, takich jak na przykład elipsa, czy trapez. W celu uproszczenia obliczeń sygnału wyjściowego zastosowano mniej skomplikowany niż opisany w Rozdz. 6 model kraniowego źródła pobudzającego. Opisany wcześniej model jest bardzo złożony i raczej ma zadanie wyjaśnienie samej istoty funkcjonowania źródła kraniowego. W dalszych rozważaniach zastosowano uproszczony model pobudzający odwzorowujący falę pobudzającą tor głosowy o pólkształtnym przebiegu.

## 8.2 Ogólny schemat realizacji modelu artykulacyjnego

W rozdziale 5-ym przedstawiono ogólne podstawy modelowania artykulacyjnego. Stanowiły one wyjściową bazę do opracowania oryginalnego komputerowego modelu artykulacyjnego, którego pierwsza wersja została opisana w pracy [NOWAKOWSKA&ŻARNECKI,1989]. Model ten działał poprawnie dla stacjonarnych przebiegów samogłoskowych i spółgłosek nosowych, a także przy jego pomocy przeprowadzono próbę ujęcia zjawiska nazalizacji ujmowanej w sposób dynamiczny, na przykładzie modelowania sylaby /ma/, tak by móc odzwierciedlić

przebieg zmian charakterystyki widmowej funkcji transmitancji toru głosowego podczas artykulacji tego typu sylaby. W obliczeniach należało podać tylko konfigurację toru głosowego odpowiadające spółgłosce /m/ oraz samogłosce /a/ oraz czas przejścia od jednej konfiguracji do następnej. Ponieważ podczas artykulacji sylaby *spółgłoska dwuwargowa nosowa – samogłoska* masa języka nie zmienia swojego ułożenia w znaczny sposób, modelowanie konfiguracji toru gardłowo-ustno-nosowego za pomocą odcinków rury cylindrycznej o liniowo zmiennych w czasie przekrojach było wystarczające i nie wprowadzało dodatkowych artefaktów. Jednakże podejmowane później próby modelowania w tym układzie połączeń samogłosek z innymi spółgłoskami nosowymi dawały w wyniku przebiegi o nienaturalnym, a niekiedy zniekształconym brzmieniu. Z tego powodu przyjęto inny sposób modelowania konfiguracji geometrycznej toru głosowego o czym będzie mowa w dalszej części.

Na Rys. 8.1 przedstawiono ogólny schemat blokowy dynamicznego sterowania artykulacyjnym modelem toru głosowego. Wyróżniono w nim dwa rodzaje parametrów zastosowanych do sterowania modelem: parametry fizyczne oraz artykulacyjne. Pierwsza grupa parametrów określa fizyczne własności ośrodka wypełniającego tor głosowy, mechaniczne własności ścianek toru  $r_{rw}$  i  $rl_w$ , a także fizyczne własności źródła pobudzającego (częstotliwość i amplitudę drgań oraz obwiednię widma).



**Rys. 8.1** Schemat blokowy sterowania artykulacyjnym modelem komputerowym (objaśnienia stosowanych oznaczeń parametrów w tekście pracy).

Parametry ośrodka przyjmuje się jako niezienne w czasie, a ich wartości są zadawane przed modelowaniem określonej konfiguracji geometrycznej toru. Natomiast

parametry źródła krtaniowego mogą być zmienne, a przebieg zmian określa się podając na przykład ich wartości początkowe i końcowe dla określonego segmentu czasowego (niekoniecznie pokrywającego się z głóską, czy sylabą).

Parametry artykulacyjne określają przede wszystkim konfigurację geometryczną toru głosowego podczas wymawiania danej głoski (lub jej fragmentu). Do tej klasy parametrów zaliczono również parametry określające własności źródła utworzonego przez fałdy głosowe. Tak więc średnia powierzchnia szpary głóśni ( $A_g$ ) wyznacza jego średnią impedancję akustyczną, zaś ciśnienie podgłóśniowe, czyli ciśnienie strugi powietrza ( $p_s$ ) przepływającego z płuc do krtani i wyznaczające energię fali akustycznej rozchodzącej się w torze głosowym. Zespół parametrów fizycznych źródła krtaniowego jest jak wykazano bardzo złożony i związany jest przede wszystkim z mechaniką drgań fałdów głosowych (patrz Rozdz. 6). Na Rys 8.1 w sposób umowy podano przede wszystkim jego specyficzne parametry akustyczne ( $A_0$  i  $F_0$ ), to jest amplitudę i częstotliwość pobudzenia krtaniowego. Parametry określające bezpośrednio geometrię toru głosowego opisują funkcję przekroju  $\Psi_l(S_n)$  odpowiadającej danej jego konfiguracji podczas wymawiania modelowanej głoski lub segmentu, gdzie  $l$  jest długością toru głosowego, zaś  $S_n$  jest powierzchnią przekroju  $n$ -tego odcinka toru. Powierzchnia ta zależy od przyjętego kształtu przekroju poprzecznego toru głosowego i dla jednakowej jego wysokości będzie różna. Stąd przykładowo model toru głosowego o eliptycznym kształcie przekroju poprzecznego będzie miał inną funkcję przekroju  $\Psi_l(S_n)$  (a tym samym i funkcję transmitancji), niż model o kołowym kształcie przekroju, choć oba modele mogą mieć identyczny kształt przekroju podłużnego.

Podstawowym problemem przy modelowaniu jest określenie funkcji przekroju poprzecznego toru głosowego dla określonej jego konfiguracji geometrycznej odpowiadającej artykulacji zadanej głoski. Stosunkowo jeszcze do niedawna, jedyną płaszczyzną, wzdłuż której można było określić ułożenie narządów artykulacyjnych (z wyjątkiem warg) była boczna płaszczyzna pionowa, względem której można było wykonać kinorentgenograficzne zdjęcie toru głosowego. Wyznaczony w ten sposób obraz jego przekroju podłużnego stanowił przez wiele lat podstawę do aproksymacji toru głosowego za pomocą szeregu (17-34) odcinków cylindrycznych. Podstawowym problemem było wyznaczenie ich średnic, tak by pole powierzchni przekroju danego odcinka cylindrycznego była równa powierzchni przekroju odwzorowywanego odcinka toru głosowego. Brak trójwymiarowych danych toru głosowego powodował heurystyczne dobieranie średnic poszczególnych odcinków modelu, w taki sposób, aby w końcowym wyniku otrzymać rozkład częstotliwościowy formantów możliwie bliski rzeczywistemu rozkładowi, wyznaczonemu dla dźwięku mowy zarejestrowanego przy odpowiadającej mu konfiguracji toru artykulacyjnego.

Tego typu aproksymacja funkcji transmitancji toru głosowego w płaszczyźnie częstotliwościowej, może jednak prowadzić do dość znacznych różnic między funkcją przekroju modelu i rzeczywistą, bez właściwie praktycznej możliwości ich

weryfikacji. Jak pokazuje praca [CARRE i in.,1992], przy zadanej konfiguracji geometrycznej można wyróżnić w modelu odcinki, których zmiany powierzchni przekroju poprzecznego mają mniejszy wpływ na rozkład częstotliwości formantowych, niż pozostałe, wskutek czego można otrzymać w modelu dla dwóch dość znacznie różniących się funkcji przekroju zbliżone do siebie rozkłady formantów.

Do problemu wyznaczania funkcji przekroju toru głosowego można więc podchodzić w dwojaki sposób. Pierwszy polega na wyznaczaniu jej postaci na podstawie mierzonych parametrów sygnału akustycznego (niekoniecznie częstotliwości formantowych), drugi zaś pośrednio, z pomierzonego przekroju podłużnego toru głosowego, tworząc uprzednio pewien zbiór zależności wymiarów poprzecznych toru głosowego od wymiarów pionowych wyznaczonych ze zdjęcia rentgenograficznego oraz położenia miejsc, w których wyznaczane są powierzchnie przekrojów poprzecznych. Wyznaczone zależności uwzględniają jednocześnie pewne naturalne ograniczenia o charakterze anatomicznym, takie jak na przykład te, które uniemożliwiają tworzenie zbyt gwałtownych zmian w przekroju poprzecznym na krótkim odcinku modelowanego toru głosowego (np. w środkowej czy tylnej części masy języka).

Problem wyznaczania funkcji przekroju toru głosowego w oparciu o mierzony zespół parametrów akustycznego sygnału mowy był badany od wielu lat i szereg metod zostało zaproponowanych (przede wszystkim dla dźwięków mowy o pobudzeniu dźwięcznym), z których należy wymienić metodę w której zastosowaną przekształcenie Fouriera do wyznaczania logarytmu funkcji przekroju (np. [MERMELSTEIN,1967;SCHROEDER,1967;WAKITA,1973]). Metoda ta była w ciągu szeregu lat udoskonalana i powstały różne jej odmiany (np. [NAKAJIMA T. i in.,1975; SHIRAI&KOBAYASHI,1986; SCHROETER&SONDHL,1994]), które umożliwiają już modelowanie dynamicznych zmian konfiguracji toru głosowego na podstawie analizy akustycznego sygnału mowy. Należy podkreślić, że bezpośrednie wyznaczanie konfiguracji toru głosowego z sygnału akustycznego mowy jest utrudnione tym, że jak już wspomniano wcześniej (patrz Rozdz. 5), parametry fizyczne sygnału nie tylko zależą od konfiguracji geometrycznej toru, ale również od parametrów fizycznych źródła pobudzającego oraz charakterystyki promieniowania ust. Oddzielenie tych wpływów przy wyliczaniu funkcji przekroju toru głosowego poprzez zastosowanie odwrotnej filtracji zaproponował stosunkowo dawno Wakita [WAKITA,1973]. Jednak metoda ta okazała się nie wystarczająca, bowiem zbyt wiele parametrów układu artykulacyjnego było nieznanymi, a przyjęta charakterystyka filtracji zbyt uproszczona. Mimo znacznych niejednoznaczności powstających przy wyznaczaniu funkcji przekroju toru głosowego, szczególnie popularne były i są próby jej określenia w oparciu o pomierzone częstotliwości formantowe (np. [LADEFOGED i in.,1978; SHIRAI&KOBAYASHI,1986; LIN, 1990; SCHOENTGEN&CIOCEA,1995]). Metody te zakładają pewne wyjściowe



konfigurację toru głosowego, a następnie w oparciu o wyznaczone parametry artykulacyjne, takie jak długość toru artykulacyjnego, położenie miejsca artykulacji, długość przewężenia toru głosowego, powierzchnia pierwszego od strony głośni przekroju początkowego itp., wyznaczana jest funkcja powierzchni przekroju. Wprawdzie dla otrzymanej w ten sposób konfiguracji toru głosowego otrzymuje się zbliżony do rzeczywistego rozkład formantów, jednakże sama kształt toru wyznaczony w ten sposób niekiedy, znacznie odbiega od wyznaczonych metodą rentgenograficzną, czy MRI.

Jak już wspomniano inne podejście wyznaczania funkcji przekroju opiera się na analitycznym opracowaniu zależności między nią i funkcją przekroju podłużnego toru głosowego określaną ze zdjęć rentgenograficznych. Zależności te można określić dla poszczególnych konfiguracji toru w sposób heurystyczny za pomocą tzw. „ $\alpha\beta$  modelu” [HEINZ&STEVENS,1965], udoskonalonego ostatnio przez Beautemps i in. [BEAUTEMPS i in.,1995]. Zgodnie ze wcześniejszą wersją tego modelu powierzchnia przekroju  $A(x)$  wyznaczana jest za pomocą funkcji odległości  $x$  od głośni danego przekroju o wymiarze pionowym  $d$ , zgodnie ze wzorem:

$$A(x) = \alpha(x)d(x)^{\beta(x)}, \quad 8.1$$

gdzie wartości współczynników  $\alpha$  i  $\beta$  są odpowiednio dobierane dla modelowanego, rzeczywistego toru głosowego i położenia w nim danego przekroju. Model Heinza i Stevensa był potem wielokrotnie udoskonalany (np.[SUNDBERG i in.,1987; BAER i in.,1991;PERRIER i in.,1992]). W wielu rozwiązaniach przyjmuje się dla współczynnika  $\beta$  stałą wartość i wynosi ona na ogół 1.5-2. W ostatniej wersji [BEAUTEMPS i in.,1995], uniknięto dość arbitralnego wyznaczania odcinków toru głosowego, dla których wyznaczane były przedziały zmian wartości współczynników  $\alpha_{\min}$  i  $\alpha_{\max}$ . Wyznaczane są one bowiem za pomocą rozkładu Fouriera w przedziale  $0 < x < l_{\text{tot}}$ , gdzie  $l_{\text{tot}}$  jest długością toru głosowego. Wówczas, odpowiednio do wysokości przekroju podłużnego  $d$  i jego odległości od głośni  $x$  współczynnik  $\alpha$  jest wyznaczany w przedziale  $[\alpha_{\min}(x), \alpha_{\max}(x)]$  na podstawie liniowej zależności [BEAUTEMPS i in.,1995]:

$$\alpha(d, x) = \alpha_{\min}(x) \text{ jeśli } d < d_{\min}$$

$$\alpha(d, x) = \alpha_{\min}(x) + \frac{\alpha_{\max}(x) - \alpha_{\min}(x)}{d_{\max} - d_{\min}} \cdot (d - d_{\min}), \text{ dla } d_{\min} < d < d_{\max}$$

$$\alpha(d, x) = \alpha_{\max}(x) \text{ jeśli } d > d_{\max}$$

W pracy [PERIER i in.,1992] stwierdzono, że dobre wyniki otrzymuje się, gdy  $d_{\min}=1$  oraz  $d_{\max}=2$  i dla takich wartości były wyliczane wartości współczynnika  $\alpha(d,x)$ . Warto zaznaczyć, że w obszarze krtani ten model wyliczania powierzchni przekroju nie jest stosowany ze względu na jej zbliżony do ostrosłupa kształt. W tej części toru głosowego model oparty na wyznaczaniu współczynników  $\alpha$  i  $\beta$  daje zupełnie fałszywe wyniki. Z tego powodu początkowy odcinek toru głosowego ponad fałdami głosowymi o długości 2 cm się, modeluje się po prostu za pomocą odcinka cylindrycznego o stałej powierzchni przekroju równej  $1.8 \text{ cm}^2$ . Podane w pracy Beautemps i in. [BEAUTEMPS i in.,1995] wyniki wyznaczania funkcji przekroju dla samogłosek /i,a,u/ oraz bezdźwięcznych spółgłosek trących zostały zweryfikowane poprzez porównanie częstotliwości formantów wyliczonych dla modelu i pomierzonych dla sygnału mowy generowanego przez osobę, dla której wykonano zdjęcie przekroju podłużnego toru głosowego techniką MRI. Rozbieżności między wyliczonymi i pomierzonymi częstotliwościami formantów są na ogół mniejsze, niż te podawane we wcześniejszych pracach, jednakże w przypadku samogłosek znaczne rozbieżności są zwłaszcza w przypadku pierwszego formantu i różnice wynoszą ponad 10% (a w przypadku /i/ ponad 17%). W przypadku spółgłosek trących błędy te są pomijalnie małe, co było tym łatwiejsze do osiągnięcia z uwagi na to, że odcinek toru głosowego efektywnie uczestniczący w formowaniu dźwięku mowy jest znacznie krótszy, niż w przypadku samogłosek, bowiem, jak już wspomniano w Rozdz. 4, tylko rezonans przedniej części znajdującej się między wargami i utworzoną szczeliną kształtuje widmo głosek o pobudzeniu szumowym. Wpływ tylnej części toru głosowego jest pomijalnie mały.

W przypadku modelowania głosek polskich na podstawie obrazu przekroju podłużnego toru głosowego powstały zasadnicze trudności, by móc zastosować metodę modelowania jego funkcji przekroju w oparciu o wyznaczenie współczynników  $\alpha$  i  $\beta$ . Podstawowym i jedynym źródłem informacji o konfiguracji toru głosowego podczas artykulacji głosek polskich jest wspomniana praca Bolla i Földi [BOLLA&FÖLDI,1987], w której podano dla poszczególnych głosek jedynie ograniczony zbiór danych, i niestety nie zawsze wolny od błędów (zwłaszcza w interpretacji fonetycznej zarejestrowanych dźwięków mowy). W cytowanej pracy brak jest danych o przekroju poprzecznym toru głosowego, co uniemożliwiło jego modelowanie za pomocą odcinków cylindrycznych, których przekroje poprzeczne byłyby wyznaczane zgodnie ze wzorem (8.1), stąd nie jest możliwe zweryfikowanie poprawności doboru współczynników  $\alpha$  i  $\beta$ .

Z tego powodu przyjęto inną, oryginalną metodę odwzorowywania trójwymiarowej konfiguracji toru głosowego, dzieląc go na trzy odcinki, w których kształt przekroju poprzecznego jest aproksymowany do jednej z figur płaskich, a mianowicie trapezu, elipsy lub koła [NOWAKOWSKA, GUBRYNOWICZ, ŻARNECKI,1993]. W przyjętym rozwiązaniu geometria toru głosowego jest określana w sposób skokowy co 1 cm, w prostopadłych do osi toru płaszczyznach. Jako

wyściową przyjęto konfigurację toru przybieraną podczas artykulacji samogłosek i aproksymowano ją za pomocą odcinków o przekroju eliptycznym w części gardłowej i ustnej (w części pod podniebieniem miękkim) oraz wargowej, o przekroju trapezowym - w części ustnej znajdującej się pod podniebieniem twardym oraz o przekroju kołowym - w części nosowej, której konfiguracja nie ulega zmianie (z wyjątkiem części wlotowej), a więc funkcja przekroju jest stała i może być odwzorowywana za pomocą odcinków cylindrycznych). Istotne jest tylko w tym przypadku właściwe odwzorowanie ruchów języzka podniebienia miękkiego, od których zależy stopień sprzężenia części nosowej z torem gardłowo-ustnym. Warto podkreślić, że już dawniej analizowano przydatność modeli o innym (zwłaszcza eliptycznym), niż powszechnie stosowanym, kołowym kształcie przekroju, ale nie stwierdzano istotnych różnic między nimi [WAKITA, 1973]. Jednakże stosunkowo niedawno zaproponowano modelowanie toru głosowego za pomocą rury o przekroju eliptycznym uzyskując bardzo dobrą zgodność wyliczonej funkcji transmitancji z wyznaczoną dla rzeczywistego toru [MIKI i in., 1996].

W zastosowanym modelu aproksymację geometrii części ustnej znajdującej się pod podniebieniem twardym dokonano w oparciu o własne badania opisane w pracy [NOWAKOWSKA i in., 1993]. Według opisanego tam sposobu aproksymacji geometrii toru głosowego poszczególne segmenty w zadanych odcinkach toru  $n$  są opisywane przez podanie ich długości ( $l_n$ ) i w zależności od przyjętego kształtu ich powierzchni przekroju ( $S_n$ ):

- średnicy dla kształtu kołowego ( $d_n$ ), lub
- średnic elipsy dla kształtu eliptycznego ( $a_n$ ,  $b_n$ ), bądź
- długości podstaw i wysokości trapezu ( $a_n$ ,  $b_n$  i  $h_n$ ).

Liczba segmentów w modelu, na które jest dzielony tor głosowy jest stała i wynosi 20. Wydłużanie, lub skracanie jego długości dla poszczególnych artykulacji realizuje się poprzez zmiany długości segmentów modelujących przednią część toru głosowego, to jest w obrębie warg. Mają one zmienną długość w zakresie od 0.2 (dla /i/) do 1 cm (dla /u/). W ten sposób modeluje się zmianę długości toru głosowego wywołaną wysunięciem się warg do przodu przy artykulacji głosek zaokrąglonych, tak jak to ma miejsce na przykład dla samogłoski /u/. Zakres zmian długości toru głosowego zawiera się więc w przedziale 15-20 cm .

Istotnym problemem przy każdym modelowaniu toru głosowego jest jego złożoność, a tym samym czasochłonność obliczeń niezbędnych do wyliczenia funkcji transmitancji toru głosowego dla zadanej konfiguracji. Funkcja ta stanowi podstawę do wyliczenia sygnału wyjściowego na impedancji promieniowania ust (i/lub ewentualnie nosa) przez jej przemnożenie przez funkcję źródła, przy uwzględnieniu jego admitancji i impedancji wejściowej toru głosowego. Jak już wspomniano w Rozdziale 5 metoda wyliczania zespolonej funkcji transmitancji toru głosowego przedstawianego w postaci szeregowo połączonych ze sobą elementarnych czwórników typu T jest bardzo

czasochłonne, jeżeli wyznaczana jest ona poprzez bezpośrednie wyznaczanie łańcucha macierzy typu A. Tak więc bezpośrednie wyliczanie sygnału wyjściowego, a ściślej biorąc mnożenie przez siebie funkcji transmitancji szeregowo ze sobą połączonych czwórników o elementach zespolonych w celu wyznaczenia zespolonej funkcji transmitancji toru głosowego jest bardzo czasochłonne. Dlatego w modelu zastosowano metodę tzw. równań Kirchoffa [KACPROWSKI,1980], polegającą na obliczaniu kolejnych, poczynając od impedancji promieniowania, impedancji zastępczych poszczególnych czwórników obciążonych impedancją widzianą na ich zaciskach wyjściowych. Wyniki pośrednie są następnie wykorzystywane do wyliczania rozptyłu prądów  $i_i$  w kolejnych czwórnikach poczynając od źródła, aż do otrzymania wartości prądu w impedancji promieniowania (por. Rys.5.8).

Metoda ta szczególnie ułatwia obliczenia w przypadku dołączenia kanału nosowego do kanału ustno-gardłowego. Impedancja obciążająca odcinek kanału gardłowo-ustnego w miejscu dołączenia kanału nosowego traktowana jest jako impedancja równoległa do impedancji wejściowej tego odcinka i ich impedancja wypadkowa występuje w kolejnych krokach obliczania impedancji pośrednich. Podobnie rozptyw prądów w tym punkcie jest na drugim etapie obliczeń wyznaczany bezpośrednio z prawa Kirchoffa. Sygnał akustyczny u wylotu nozdrzy jest obliczany identycznie jak sygnał u wylotu ust, przy czym impedancję promieniowania oblicza się na podstawie średniej powierzchni otworów nozdrzy. W przypadku modelowania spółgłosek nosowych i nazalizowanych przyjęto, że wyjściowy sygnał akustyczny jest rejestrowany w polu dalekim i jest algebraiczną sumą dwóch oddzielnie wyliczonych sygnałów, to jest promieniowanego z ust oraz z nozdrzy.

W obliczeniach sumarycznego sygnału wyjściowego przyjęto, że pobudzenie krtaniowe jest w dziedzinie czasu funkcją periodyczną o kształcie trójkątnym (lub piłozębatym), której charakterystyka widmowa jest funkcją dyskretną, o stałym nachyleniu obwiedni -12 dB/okt. Zakładając, że w zakresie częstotliwości akustycznych układ elektryczny będący analogiem toru głosowego jest układem liniowym, można zastosować zasadę superpozycji. W myśl tej zasady przyjmuje się, że jeżeli pobudzenie układu jest w dziedzinie częstotliwości harmoniczne, to odpowiedź jego jest odpowiednią sumą odpowiedzi na poszczególne składowe pobudzenia. Tak więc przyjmując, że dla prąka podstawowego (czyli dla  $f = F_0$ ) amplituda pobudzenia ma ustaloną wartość  $A_0$ , to dla każdej częstotliwości dyskretniej będącej wielokrotnością składowej podstawowej  $f = k \cdot F_0$  można obliczyć wartość amplitudy odpowiedniej składowej harmonicznej odpowiedzi. Sumując je w każdym obwodzie zastępczym otrzymuje się wartości prądów przepływających przez poszczególne elementy układu zastępczego.

W ten sposób można wyznaczyć składowe częstotliwościowe sygnału wyjściowego, a zwłaszcza można określić charakterystykę częstotliwościową natężenia prądu płynącego w impedancji promieniowania. W dziedzinie częstotliwości prąd ten

jest funkcją zespoloną. Ponieważ otrzymany w wyniku modelowania sygnał ma być odsłuchiwany, obliczane jest więc widmo napięcia wyjściowego na impedancji promieniowania, które też jest funkcją zespoloną. Sygnał ten w postaci ciągłej liczony jest według odwrotnej dyskretnej transformaty Fouriera:

$$x(t) = \sum_i \operatorname{Re}(H_i u_i) \sin 2\pi F_0 t + \operatorname{Im}(H_i u_i) \cos 2\pi F_0 t \quad 8.2$$

gdzie  $u_i$  jest  $i$ -tą składową zespoloną widma. Jest to więc sygnał periodyczny o częstotliwości podstawowej  $F_0$  odpowiadającej częstotliwości pobudzenia krtaniowego. W celu wyprowadzenia sygnału poprzez przetwornik cyfrowo-analogowy, sygnał ten musi być przedstawiony w postaci cyfrowej dla zadanej częstotliwości próbkowania  $F_s$ , na ogół równej 10-20 kHz.

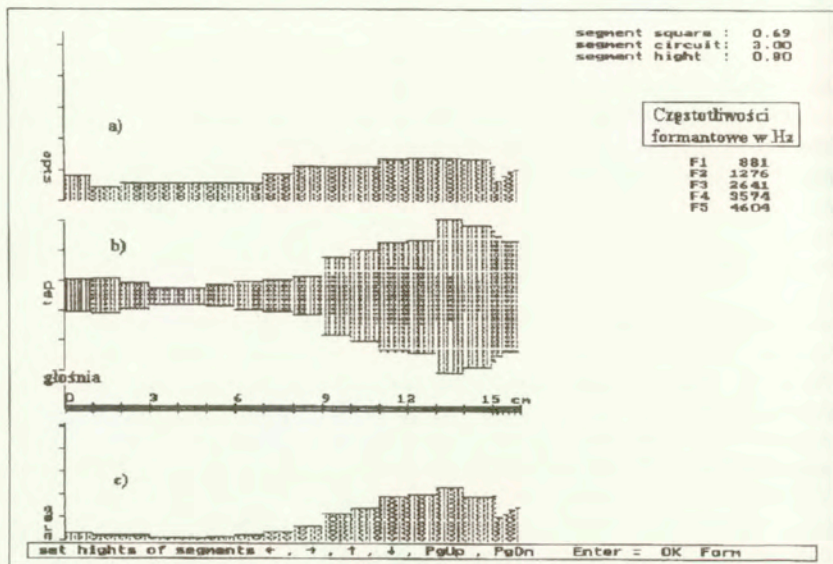
Przesyłanie kolejnych wartości obliczonego sygnału na wejście przetwornika cyfrowo-analogowego połączonego ze wzmacniaczem akustycznym i głośnikiem daje w efekcie sygnał akustyczny syntetycznego sygnału mowy. Otrzymywany w ten sposób sygnał zależy bezpośrednio jedynie od zadanej konfiguracji artykulacyjnej toru oraz wartości wyżej wymienionych parametrów toru głosowego.

### 8.3 Opis programu „MODEL”.

Opracowany program komputerowy ma bardzo rozbudowaną część edycyjną. Umożliwia ona bezpośrednio wprowadzanie z klawiatury danych określających konfigurację geometryczną modelowanego toru głosowego, jej wizualizację, wyznaczanie częstotliwości formantowych dla zadanych konfiguracji toru głosowego oraz system automatycznego programowania przebiegu zmian konfiguracji od kształtu początkowego do kształtu końcowego wraz z wyświetlaniem funkcji przekrojów w stanach pośrednich i odpowiadającym im częstotliwościom formantowym. Punktem wyjścia w obliczeniach było wyznaczanie funkcji transmitancji toru głosowego przy jednoczesnym założeniu, że charakterystyka częstotliwościowa źródła krtaniowego jest funkcją o stałej fazie [FANT, 1960].

Częstotliwości formantowe, czyli częstotliwości, dla których funkcja przenoszenia kanału artykulacyjnego ma bieguny są obliczane na podstawie analizy przebiegu fazy widma wyliczonego sygnału wyjściowego. W tym celu znajduje się częstotliwości, dla których kąt wyznaczany przez wektor wodzący wykresu tej charakterystyki na płaszczyźnie zmiennej zespolonej przekracza wartość  $[\pi/2 + k\pi]$ , gdzie  $k=0,1,2,\dots$ . Po znalezieniu przedziału, gdzie takie przekroczenie następuje, właściwą wartość częstotliwości formantowych oblicza się metodą iteracyjną z dokładnością do 0.1 Hz. Ponieważ istnieje ryzyko dwukrotnego przekroczenia wartości krytycznych między poszczególnymi harmonicznymi (zwłaszcza dla dużych wartości  $F_0$ ), tak więc przy obliczaniu formantów charakterystykę amplitudowo-fazową sygnału

wyznacza się z krokiem  $0.1 \cdot F_0$ . Dzięki temu uzyskano możliwość dokładnego wyznaczenia częstotliwości formantowych nawet dla takich sygnałów, dla których kolejne formanty są położone blisko siebie lub w przypadku wystąpienia nazalizacji, gdy formanty mogą być zamaskowane przez antyformant. Metody analizy sygnału akustycznego oraz algorytmy analizy przebiegów czasowych stosowane przy analizie rzeczywistego sygnału mowy nie pozwoliłyby na uzyskanie ani takiej pewności w detekcji kolejnych formantów, ani takiej dokładności przy wyznaczaniu ich częstotliwości. Dodatkową zaletą przedstawionej metody analizy sygnału wyjściowego na płaszczyźnie zespolonej jest możliwość określenia miejsc występowania antyformantów nawet wówczas, gdy na charakterystyce widma amplitudowego nie są one zauważalne.

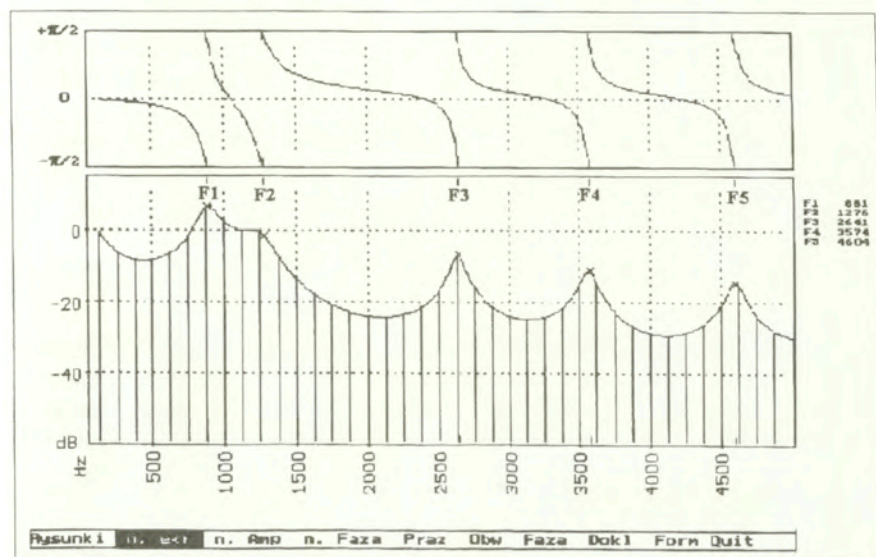


Rys. 8.2. Widok modelowanej konfiguracji toru głosowego z boku (a), z góry (b) oraz jego funkcji przekroju (c) wyznaczonej dla modelowanej samogłoski /a/. U góry z prawej strony podane są częstotliwości formantowe otrzymane dla tej konfiguracji.

Na Rys. 8.2 przykładowo podano widok konfiguracji geometrycznej toru głosowego przedstawiany przez program *Model* na monitorze komputera dla modelowanej samogłoski /a/. Pokazano na nim idąc od góry przekrój wzdłużny toru głosowego (a), widok z góry (b) zaznaczając w obrębie podniebienia twardego (od 9 do 15 cm) wymiary dolnej i górnej podstawy trapezu aproksymującego przekrój poprzeczny poszczególnych segmentów toru głosowego, oraz (c) przebieg funkcji przekrojów  $S_n$  dla modelowanej konfiguracji. Z prawej strony, u góry diagramu, podane

są wymiary geometryczne wskazanego kursorem segmentu (w tym przypadku  $n=1$ ) i poniżej wyliczone dla analizowanej konfiguracji częstotliwości pierwszych pięciu formantów. Dla zadanej konfiguracji podane są wyliczone częstotliwości formantowe wyznaczone w oparciu o charakterystykę fazową funkcji transmitancji toru głosowego (patrz Rys. 8.3.) wyliczoną automatycznie przez program.

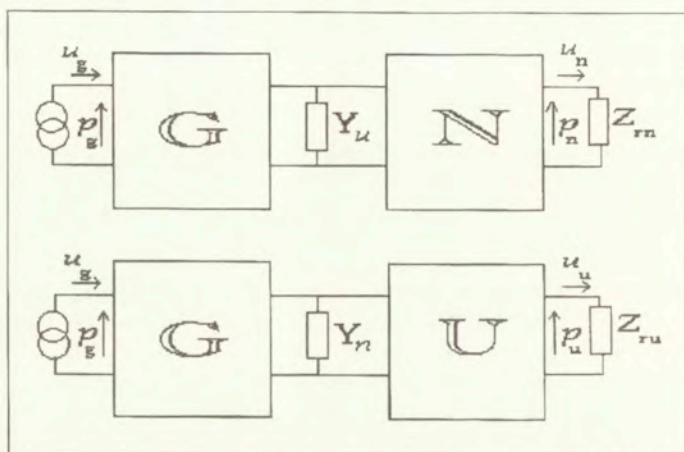
Na Rys. 8.3 przedstawiono amplitudowo-fazową funkcję transmitancji wyliczoną dla przedstawionej na Rys. 8.2 konfiguracji toru głosowego modelu samogłoski /a/. Widoczne są bardzo wyraźne skoki charakterystyki fazowej (skokowa zmiana znaku funkcji transmitancji z ujemnego na dodatni) na częstotliwościach formantowych, co zgadza się zresztą z teorią modelowania toru głosowego za pomocą czwórników i fakt ten został wykorzystany do precyzyjnego wyznaczania częstotliwości formantowych. W podanym przykładzie, maksimum w charakterystyce amplitudowej odpowiadające drugiemu formantowi F2 jest mało wyraźne, natomiast skokowa zmiana znaku fazy jest jednoznaczna.



Rys. 8.3 Widmo amplitudowo-fazowe modelowanej samogłoski /a/ o konfiguracji toru głosowego przedstawionej na Rys. 8.2. Widmo prążkowe obliczono dla częstotliwości podstawowej  $F_0=125$  Hz. Obok wykresu podano częstotliwości pierwszych pięciu formantów (w Hz)

Innym trudnym przypadkiem w mowie naturalnej jest, jak już wspomniano, wyznaczanie częstotliwości formantowych dla samogłosek nazalizowanych. Dołączenie kanału nosowego powoduje pojawienie się w widmie dodatkowych

maksimów nie związanych bezpośrednio z konfiguracją geometryczną gardłowo-ustnej części toru, ale wynikające przede wszystkim z bocznikującego oddziaływania części nosowej. Bezpośrednie obliczanie sygnału wyjściowego toru artykulacyjnego w przypadku włączenia części nosowej jest złożone i w celu uproszczenia obliczeń w opracowanym modelu wyróżnia się dwa oddzielne tory, to jest gardłowo-ustny formujący sygnał promieniowany przez usta oraz gardłowo-nosowy formujący sygnał promieniowany przez nos. W obu torach uwzględnia się odpowiednie admitancje wejściowe bocznikujące w miejscu rozgałęzienia się torów, to jest admitancję widzianą od strony wlotu do części nosowej  $Y_n$ , i bocznikującą w tym miejscu tor gardłowo-ustny, oraz admitancję wejściową części ustnej  $Y_u$  bocznikującą w tym samym miejscu tor gardłowo-nosowy, tak jak to przedstawiono na Rys. 8.4.



**Rys. 8.4** Schematyczne przedstawienie torów gardłowo-nosowego (u góry) i gardłowo-ustnego z uwzględnieniem wzajemnego obciążenia w miejscu rozgałęzienia się części nosowej i ustnej, odpowiednio admitancjami  $Y_u$  oraz  $Y_n$ . Wyjścia torów są zamknięte odpowiednimi impedancjami promieniowania, to jest  $Z_m$  (część nosowa) oraz  $Z_u$  (część ustna). Sumaryczne ciśnienie akustyczne fali promieniowanej do ośrodka w polu dalekim jest  $p_u + p_n$  [NOWAKOWSKA&ŻARNECKI, 1989].

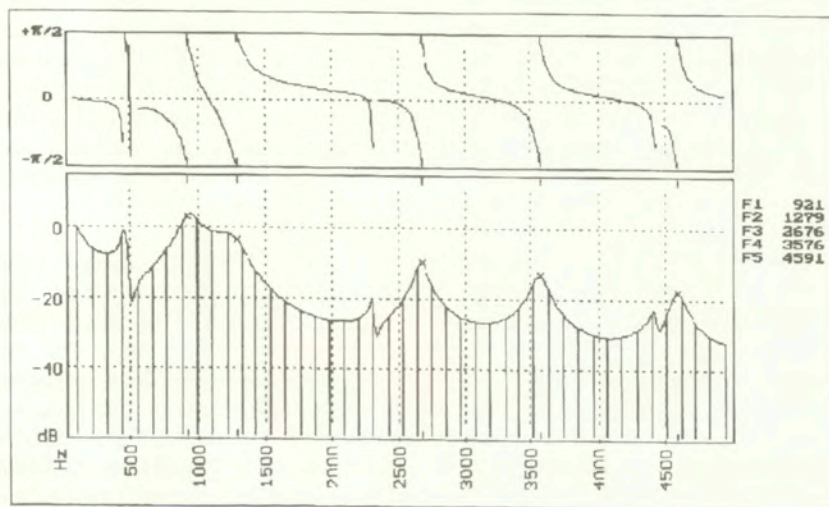
Na Rys. 8.5 przedstawiono przykładowo zamodelowany wpływ podłączenia części nosowej dla konfiguracji toru gardłowo-ustnego odpowiadającej samogłosce polskiej /a/. Wpływ ten objawia się pojawieniem ostrych minimów w charakterystyce widmowej (np. 490 Hz), wywołanych wystąpieniem w funkcji transmitancji toru głosowego zer powstających w wyniku równoległego podłączenia admitancji wejściowej toru nosowego. W podanym przykładzie, dla częstotliwości w których występują dodatkowe maksima nie będące formantami (490 Hz, 2350 Hz i 4420 Hz), można łatwo zauważyć, że w charakterystyce fazowej skoki fazy są zupełnie o innym przebiegu, niż w przypadku formantów, co pozwala uniknąć w tym przypadku



błędno zaliczenia ich do formantów, jak to ma miejsce w wielu systemach modelowania toru głosowego, w których formanty są definiowane jako lokalne maksima funkcji transmitancji i w przypadku modelowania nosowania wymagają ręcznej korekty (por. np. model Maeda dostępny w Internecie [MAEDA,1993]) Warto zaznaczyć, że w przypadku mowy naturalnej antyformanty spowodowane podłączeniem części nosowej w charakterystyce widmowej sygnału mowy ulegają często zatarciu, m. in., wskutek stosunkowo drobnej niestabilności częstotliwości podstawowej F0.

Zjawisko nazalizacji wywołane sąsiedztwem spółgłoski nosowej (wskutek koartykulacji), a także nosowania spowodowanego zmianami patologicznymi w torze artykulacyjnym (np. niedomykanie się podniebienia miękkiego), może spowodować błędne określenie częstotliwości formantowych wymawianych głosek ustnych [GUBRYNOWICZ,1990].

W opracowanym programie „Model” zadawane mogą być konfiguracje toru artykulacyjnego stacjonarne, jak również i zmieniające się w czasie. W drugim przypadku przyjmuje się założenie istnienia pewnej ciągłości w zmianie konfiguracji toru głosowego podczas wymawiania takich głosek, jak na przykład /j,w/.



Rys. 8.5 Charakterystyka amplitudowo-fazowa funkcji transmitancji nazalizowanej samogłoski /a/, obliczonej za pomocą modelu .

Jeżeli w chwili t1 trójwymiarowa geometria toru głosowego {g1} jest opisana przez wymiary geometryczne kolejnych segmentów toru, zaś w chwili docelowej t2

konfiguracja ta jest opisana za pomocą innej geometrii  $\{g_2\}$ , to ścianki poszczególnych segmentów muszą się przemieszczać przestrzennie w pewnym skończonym, różnym od zera czasie  $t_w = t_2 - t_1$ . Można przyjąć więc, że w czasie  $t_w$  mamy kolejno  $N$  różnych konfiguracji toru artykulacyjnego. W poprzednich badaniach [NOWAKOWSKA i in., 1989; 1993; GUBRYNOWICZ i in., 1995], przyjęto minimalny krok czasowy, podczas którego konfiguracja toru głosowego jest stała, równy okresowi pobudzenia krztaniowego, czyli  $N = t_w \cdot F_0$ . Umożliwiało to znaczne przyspieszenia obliczania sygnału mowy.

Jednakże założenie takie okazało się bardzo niewygodne przy modelowaniu dynamicznych zmian konfiguracji toru głosowego, bowiem nakładało konieczność spełnienia przy wyliczaniu sygnału spełnienia warunku wielokrotności czasu trwania tych zmian  $t_w$  względem  $T_0$  ( $1/F_0$ ), co wykluczało na przykład możliwość programowania ciągłych przebiegów tranzjentowych ze zmienną częstotliwością podstawową  $F_0$ . Ponadto, przy zadanym czasie  $t_w$  przejścia od jednej konfiguracji do następnej warunek ten mógł być spełniony tylko dla określonego zbioru wartości  $F_0$ . Niespełnienie tego warunku, nawet w minimalnym stopniu, powodował znaczne zniekształcenia w sygnale syntetycznym odbierane jako "chrypienie" wywołane nieciągłymi zmianami w sygnale. Dlatego w obecnej wersji programu „Model” przy opisie zmian geometrii toru głosowego przyjęto krok czasowy równy okresowi próbkowania sygnału (tj. 0.1 ms), czyli  $N = t_w \cdot F_s$ , co pociągnęło za sobą znaczne zwiększenie obliczania konfiguracji pośrednich (blisko 100 razy), lecz dzięki temu uzyskano sygnał syntetyczny o wysokiej jakości. W obliczeniach przyjęto, że przestrzenne przemieszczanie się ścianek toru głosowego jest liniowe w funkcji czasu. Dla każdej kolejnej konfiguracji obliczano odpowiadające jej widmo sygnału wyjściowego. Następnie z odwrotnego przekształcenia Fouriera wyznaczono wartości amplitudy sygnału dla kolejnych dyskretnych chwil czasowych  $t_k$ .

Powstający podczas dynamicznej zmiany konfiguracji toru głosowego sygnał, jest wyliczany przy przejściu od zadanej konfiguracji początkowej do końcowej (przy zachowaniu stałego kroku dyskretyzacji w dziedzinie czasu) z sekwencji widm odpowiadających kolejnym pośrednim geometriom toru artykulacyjnego. Tak więc modelując przejście tranzjentowe, wystarczy w programie *Model* określić początkową i końcową konfigurację geometryczną układu artykulacyjnego oraz czas przejścia między tymi konfiguracjami. W podobny sposób, można płynnie zmieniać niektóre parametry modelowania (np. obwiednię amplitudy, częstotliwość podstawową) istotne dla brzmienia syntezywanego sygnału, choć nie wpływają one w sposób znaczący na jego zrozumiałość. Na przykład, nawet niewielkie zmiany częstotliwości podstawowej  $F_0$  w sygnale syntetycznym mają duży wpływ na jego subiektywną ocenę naturalności brzmienia, zwłaszcza, gdy uwzględnia się zmiany mikromelodyczne oraz w szczególnych przypadkach, na odróżnienie znaczenia twierdzącego lub pytającego generowanej frazy. W zasadzie wszystkie parametry modelu mogą być zmieniane w sposób ciągły, lub dyskretny, synchronicznie z okresem tonu podstawowego.

Na Rys. 8.6 przedstawiono przykładowy wydruk funkcji sterującej wyznaczonej do generacji przebiegu tranzjentowego połączenia sylabicznego typu VCV (/ajo/). W pierwszej kolumnie wydruku (GEOU) podane są docelowe konfiguracje toru gardłowo-ustnego odpowiadające poszczególnym głoskom i nie ulegające przez pewien krótki czas zmianie. W kolumnie CZAS podano w [ms] odpowiednie czasy trwania poszczególnych konfiguracji. "Gwiazdkami" oznaczono segmenty przejściowe o zmieniającej się w czasie konfiguracji toru. Powtórzenie nazwy konfiguracji (np. „A” odpowiadającej głosce /a/), sygnalizuje, że sama konfiguracja toru głosowego nie ulega zmianie, lecz zmieniają się inne parametry (w tym przypadku amplituda sygnału - "ampl", czy częstotliwość podstawowa  $F_0$ ) generowanego sygnału. Poza amplitudą sygnału i częstotliwością  $F_0$  można niezależnie dobrać wartości takich parametrów źródła pobudzenia jak wielkość powierzchni szpary głośni ( $A_g$ ), nachylenie obwiedni amplitudy charakterystyki widmowej źródła w pełnym zakresie częstotliwości (dB\_oct), a także w zakresie powyżej 800 Hz (dB\_oct8) oraz ciśnienie podgłośniowe ( $P_g$ ). Kolumna GEON podaje ukształtowanie części wlotowej toru nosowego. W podanym przykładzie wlot do nosa jest praktycznie zamknięty przez język podniebienia miękkiego (średnica wlotu wynosi bowiem 0.05 mm).

AJO									
GEOU	GEON	CZAS	F0	dB_oct	A_g	P_g	ampl	dB_oct8	
A	Nos05	32	125	-12	0.2	10	1.8	-12	
	***	24							
A	Nos05	32	128	-12	0.2	10	1.2	-12	
	***	16							
A	Nos05	16	132	-12	0.2	10	1.5	-12	
	***	70							
J	Nos05	24	128	-12	0.2	10	1.3	-12	
	***	112							
O	Nos05	32	135	-12	0.2	10	1.3	-12	
	***	32							
O	Nos05	64	138	-12	0.2	10	1.5	-12	

Rys. 8.6 Wydruk wartości parametrów funkcji sterujących modelowaniem przebiegu tranzjentowego dla sekwencji VCV /ajo/.

W obliczeniach zastosowano uproszczony model źródła krtaniowego. Założono, że nie ma wzajemnego oddziaływania między źródłem i torem głosowym. Źródło pobudzenia krtaniowego może więc być określone jako źródło prędkości objętościowej ( $u_g$ ) powietrza przepływającego przez głośnię. W układzie zastępczym traktowane jest ono jako źródło o stałej wydajności prądowej, niezależnej od zadanej

konfiguracji geometrycznej toru głosowego. Jego impedancja może być przedstawiona w postaci szeregowego połączenia oporności oraz indukcyjności i jest określona wzorem [KACPROWSKI,1977]:

$$Z_g = R_g + j\omega L_g \quad 8.3$$

$$= 12\mu d l^2 A_g^{-3} + 0.44\rho A_g^{-2} |u_g| + j\omega\rho d A_g^{-1}$$

gdzie

- $A_g$ - powierzchnia skuteczna głośni,
- $l$  - długość szpary głośni,
- $d$  - głębokość tej szpary,
- $u_g$ - prędkość objętościowa powietrza przepływającego przez głośnię.

Badania wykonane przez Badin i Fanta [BADIN&FANT,1984] wykazały, że wzór ten jest ważny dla szerokiego zakresu wartości parametrów przepływu:

- ciśnienia podgłośniowego -  $P_s = 64$  cm H<sub>2</sub>O,
- szerokości szpary głośni -  $0.1$  w(t)= $A_g(t)/l = 2$  mm i
- średniej prędkości objętościowej -  $u_g = 2000$  cm<sup>3</sup>/s.

Zakładając, że wymiary geometryczne źródła krtaniowego są stałe i wynoszą odpowiednio  $d=0.3$  cm i  $l=1.8$  cm, można przyjąć następujący wzór określający impedancję głośni:

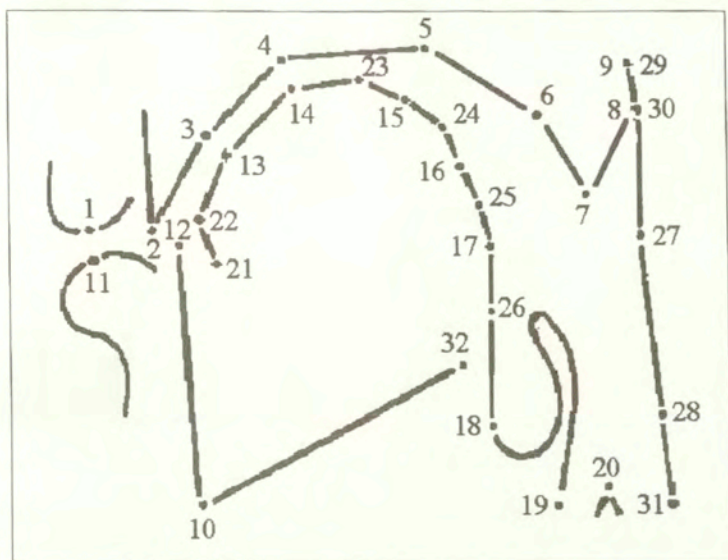
$$Z_g = 3.72 \cdot 10^{-4} \frac{l}{w^3} + 2.3 \cdot 10^{-2} \frac{l}{w} \sqrt{P_s} + j \cdot 1.95 \cdot 10^{-3} \frac{f}{w} \quad 8.4$$

W obliczeniach impedancji zakłada się na ogół średni wysięk głosowy, któremu odpowiadają wartości  $P_s \gg 10$  cm H<sub>2</sub>O i  $w=0.12$  cm, zaś przyjmowany zakres analizowanych częstotliwości jest na ogół poniżej 4.5 kHz.

8.4 Przyjęta metoda modelowania rezonantów języka polskiego i ich połączeń w oparciu o dane artykulacyjne głosek polskich.

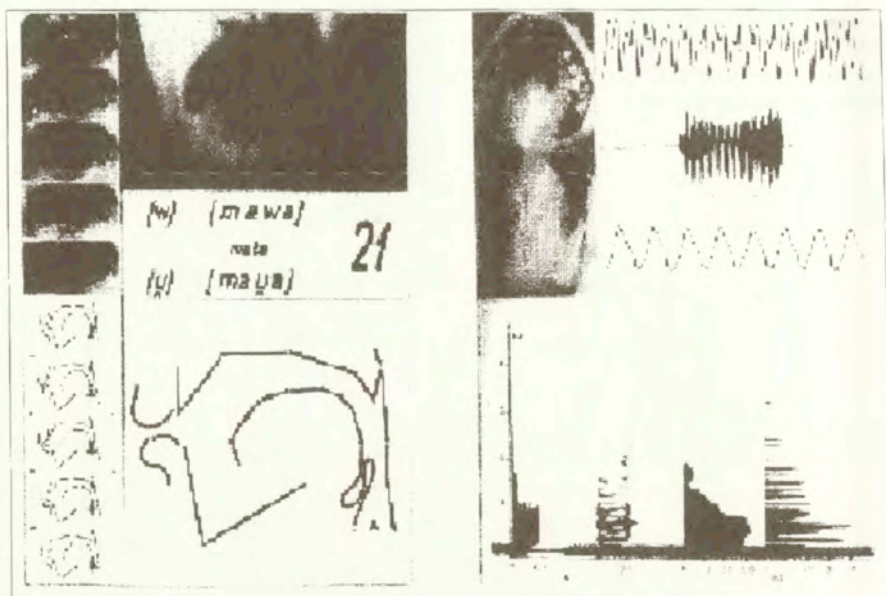
W badaniach nad modelowaniem artykulacyjnym rezonantów języka polskiego za pomocą programu „Model” zastosowano wyniki unikalnych pomiarów wykonanych dla głosek języka polskiego przez naukowców węgierskich [BOLLA&FÖLDI,1987]. Na podstawie zdjęć kinorentgenograficznych przedstawiających zmiany przekroju wzdłużnego toru głosowego, zarejestrowanych podczas wymawiania przez polskiego lektora izolowanych wyrazów, wybrali oni określone punkty pomiarowe w torze gardłowo-ustnym, wyznaczając odległości między nimi dla określonych faz artykulacji

analizowanej głoski. Schematyczne rozmieszczenie punktów pomiarowych na zdjęciach przekroju podłużnego toru głosowego przedstawiono na Rys. 8.7. Dane ze zdjęć kinorentgenograficznych uzupełniono pomiarami układu warg oraz powierzchni styku języka z podniebieniem wyznaczonymi z fotolingogramów i fotopalatogramów.



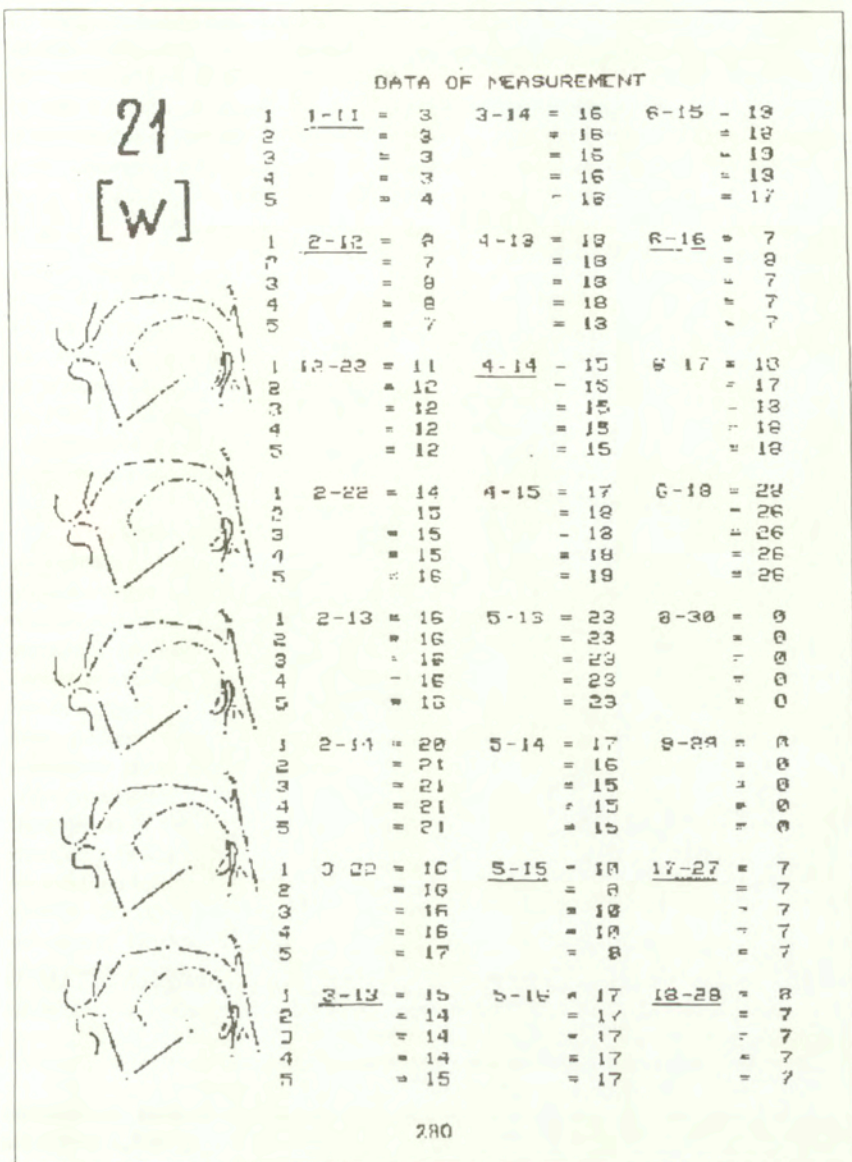
**Rys. 8.7** Schematyczne rozmieszczenie punktów pomiarowych na zdjęciu kinorentgenograficznym toru głosowego [BOLLA&FÖLDI, 1987].

Na Rys. 8.8 przedstawiono przykładowo zestaw danych graficznych dotyczący głoski /w/ wymówionej w wyrazie „mała” /mawa/. Po lewej stronie rysunku przedstawiono zmiany przekroju podłużnego toru głosowego oraz układu warg w pięciu równoodległych momentach czasowych, podczas jej wymawiania. Po prawej stronie rysunku, przedstawiono palatogram (u góry) i lingogram (poniżej) pokazujące zaznaczone czarnym kolorem powierzchnie styku podniebienia i języka wyznaczone dla tej głoski. Z boku, u góry podano przebieg akustyczny sygnału jej odpowiadający, w skali czasowej rozciągniętej i normalnej, poniżej, na wysokości lingogramu – glotogram przedstawiający impulsy pobudzenia krtaniowego zaś u samego dołu – charakterystykę widmowo-czasową głoski /w/ oraz przekrój widma wyznaczony w środkowej fazie jej artykulacji.



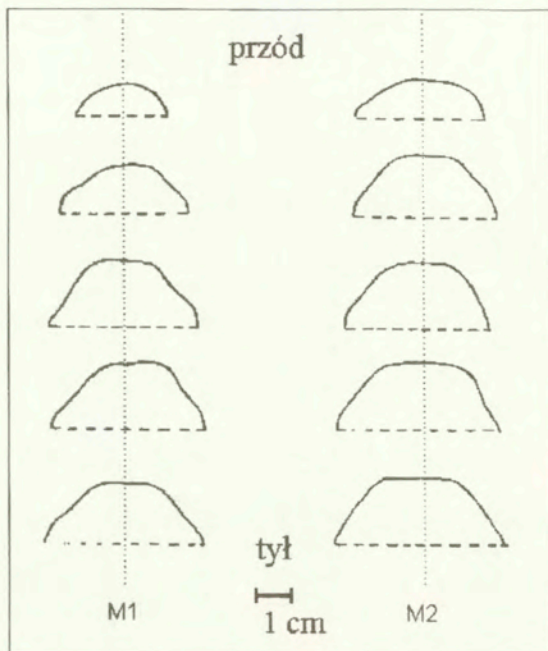
**Rys. 8.8** Graficzna forma przedstawienia danych artykulacyjnych i akustycznych wyznaczonych dla spółgłoski /w/ wypowiedzianej w wyrazie /mawa/ (“mała”) [BOLLA&FÓLDI, 1987].

Liczbową formę przedstawiania wyników pomiarów przekroju podłużnego toru głosowego przedstawiono na Rys. 8.9. Dla każdej głoski wyróżniano pięć faz artykulacyjnych o jednakowym czasie trwania. Autorzy wybrali 24 pary punktów pomiarowych (zgodnie z oznaczeniami podanymi na Rys. 8.7) i dla każdej pary wyznaczyli odległości między nimi (w mm), podczas każdej z faz artykulacyjnych.. Najistotniejszą z punktu widzenia modelowania głosek stacjonarnych jest faza środkowa, odpowiadająca tak zwanej “czystej” fazie artykulacji danej głoski. Wyznaczona dla tej fazy konfiguracja toru głosowego może być również traktowana jako „docelowa” przy modelowaniu głosek o tranzjencyjnym charakterze (tak jak to właśnie ma miejsce w przypadku głoski /w/). Na wydruku podkreślono te odległości między punktami pomiarowymi, które mogły być wykorzystane przy modelowaniu konfiguracji toru głosowego, a więc są to odległości między punktami leżącymi naprzeciw siebie. W sumie danych do modelowania konfiguracji geometrycznej toru głosowego nie ma zbyt wiele, szczególnie w obrębie części kraniowej (tylko 2). Pozostałe pomiary miały charakter pomocniczy i służyły tylko do dokładniejszej rekonstrukcji graficznej przekroju wzdłużnego toru głosowego.



Rys. 8.9 Forma przedstawienia w pracy [BOLLA&FÖLDI,1987] wyników pomiarów konfiguracji geometrycznej toru głosowego dla 5-ciu faz artykulacji spółgłoski /w/ wypowiedzianej w wyrazie "mała" (/mawa/) (patrz objaśnienia w tekście).

W celu uzupełnienia tych w sumie dość niepełnych danych i określenia górnego kształtu przedniej części toru głosowego, wykonano dla 10 osób odciski stomatologiczne podniebienia twardego, a następnie zrobiono z nich odlewy. W odstępach co 1 cm wykonano na nich prostopadłe do powierzchni odwzorowującej kształt podniebienia twardego przekroje. Przekroje te umożliwiły wyznaczenie przez aproksymację niektórych wymiarów poprzecznych toru głosowego (w cytowanej pracy wymiary poprzeczne są tylko w obrębie warg). Na Rys. 8.10 przedstawiono w ten sposób wyznaczone dla dwóch osób kształty przekrojów poprzecznych w obrębie podniebienia twardego [NOWAKOWSKA i in., 1993].



**Rys. 8.10** Przykładowe przekroje poprzeczne toru głosowego (w odstępach co 1 cm) w obrębie podniebienia twardego wyznaczone dla dwóch osób z odcisków stomatologicznych ([NOWAKOWSKA i in., 1993]).

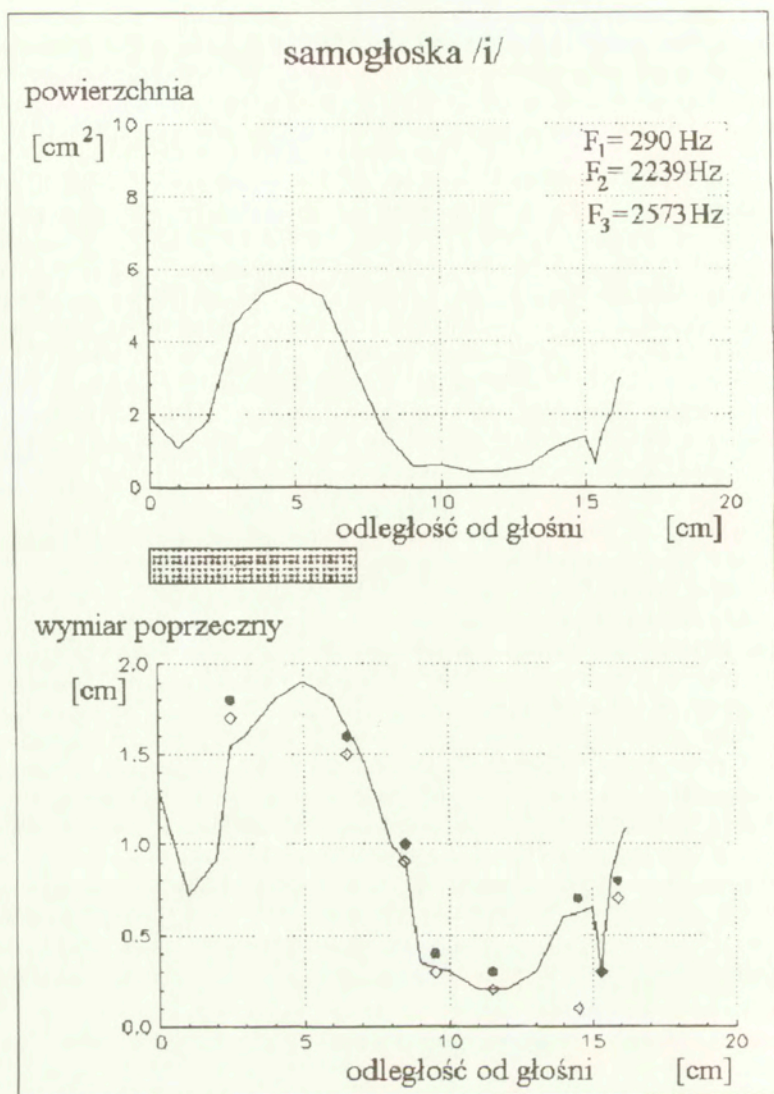
Z przedstawionych rysunków widać, że przy założeniu płaskiej powierzchni języka przy artykulacji (co jest do przyjęcia przy wymawianiu samogłosek, a zwłaszcza samogłoski /a/), to z pewnym przybliżeniem można przyjąć, że przekroje poprzeczne toru głosowego w obrębie podniebienia twardego (z wyjątkiem bezpośredniego sąsiedztwa przednich zębów) można aproksymować do kształtu trapezowego. Badania



robione za pomocą rezonansu magnetycznego (MRI) umożliwiające trójwymiarową wizualizację toru głosowego (patrz zwłaszcza [STORY i in., 1996]) potwierdzają częściowo słuszność tej aproksymacji dla samogłosek. Należy jednak tutaj podkreślić, że istotnym jest odtworzenie przede wszystkim przebiegu zmian funkcji przekroju toru głosowego, a nie tylko samego jego kształtu.

U uzupełnione w ten sposób dane znajdujące się w cytowanej pracy dokonywano próby rekonstrukcji funkcji powierzchni przekrojów rzeczywistego toru głosowego, zachowując w modelu te same odległości między odpowiednimi punktami pomiarowymi (tj. te same wymiary poprzeczne). Jednocześnie pozostałe wymiary modelu były dobierane w taki sposób, by uzyskać częstotliwości formantów możliwie bliskie rzeczywistym częstotliwościom modelowanej głoski, przy jednoczesnym zachowaniu odpowiednich proporcji wymiarów podłużnych przekrojów do poprzecznych, tak by były zbliżone do danych podawanych w literaturze [BAER i in., 1991]. Warto zaznaczyć, że węgierscy autorzy udostępnili wykonywane jednocześnie z filmem kinorentgenograficznym nagrania dźwiękowe, co umożliwiło pomiary częstotliwości formantowych modelowanych głosek.

Przykład rekonstrukcji funkcji powierzchni przekroju poprzecznego dla samogłoski /i/ wypowiedzianej w wyrazie „bigos” przedstawiono na Rys. 8.11. U dołu przedstawiono widok boczny przekroju toru głosowego wraz z zaznaczonymi wymiarami podanymi w cytowanej pracy (kropki). U góry podano funkcję powierzchni przekroju modelowanej konfiguracji toru głosowego wraz z odpowiadającymi jej częstotliwościami formantowymi. Wartości te pokrywają się z dużą dokładnością z rzeczywistymi, z błędem nie większym niż 1.4% (patrz Tab. 8.1). Dla porównania, u dołu podano również odległości (jasne romby) między punktami pomiarowymi podanymi przez Bolla i Földi(1987) dla środkowej fazy artykulacji spółgłoski /j/ wypowiedzianej w wyrazie „jama”. Jak widać różnice w tej fazie artykulacji /j/ i samogłoski /i/ nie są zbyt wielkie i z powodzeniem można zastępować przy realizacji tranzjentu /j/ konfiguracją docelową toru głosowego o wymiarach i kształcie zbliżonym do konfiguracji geometrycznej samogłoski /i/. Widoczna na rysunku (u dołu) znaczna różnica dla tych głosek w wysokości segmentu toru głosowego położonego w odległości 14 cm od szpary głośni nie ma jak to sprawdzono na modelu istotnego wpływu na częstotliwości formantowe  $F_1$ ,  $F_2$ ,  $F_3$ .



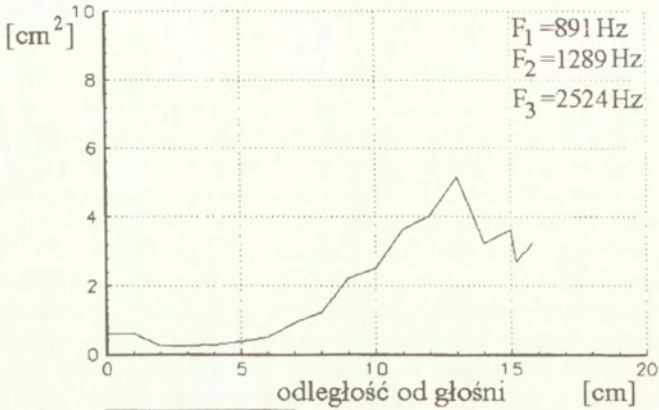
**Rys. 8.11** Funkcja powierzchni przekroju samogłoski /i/ (u góry) otrzymana w wyniku modelowania konfiguracji geometrycznej toru głosowego. U dołu podano odpowiadający jej przekrój wzdłużny toru. Kropkami zaznaczono wymiary poprzeczne toru głosowego podane w pracy [BOLLA&FÖLDI,1987]. Dla porównania jasnymi rombami zaznaczono odległości wyznaczone przez tych samych autorów dla spółgłoski /j/ (środkowa faza artykulacji).

Na następnych rysunkach (8.12 i 8.13) przedstawiono analogiczne wykresy dla samogłoski /a/ wypowiedzianej w wyrazie „tak” oraz samogłoski /u/ wymówionej w wyrazie „góra”. Wyznaczona funkcja transmitancji dla aproksymowanej funkcji powierzchni samogłoski /a/ ma częstotliwości formantowe bardzo bliskie pomierzonym dla rzeczywistego sygnału mowy i ich wartości są zawarte przedziałach typowymi dla samogłoski /a/ języka polskiego (patrz [CALIŃSKI i in., 1970]).

Nieco odmienny wynik modelowania otrzymano w przypadku samogłoski /u/, dla której funkcja powierzchni przekroju (u góry, linia ciągła) ma niezbyt typowy dla tej samogłoski przebieg. Jak już wspomniano w Rozdziale 5 uproszczona konfiguracja toru głosowego podczas artykulacji samogłoski /u/ może być przedstawiona w postaci dwóch komór połączonych ze sobą krótkim przewężeniem. Dla takiej konfiguracji funkcja powierzchni charakteryzuje się na ogół dwoma dość rozległymi wierzchołkami (tak jak to zaznaczono na rysunku linią kropkowaną). Jednakże osoba, dla której był wykonywany zapis kinorentgenograficzny wymówiła tę samogłoskę z nieco ściśniętym gardłem, prawdopodobnie pod wpływem kontekstu spółgłoski /g/ (co spowodowało znaczne zmniejszenie objętości tylnej komory w torze głosowym i tym samym znaczne podniesienie częstotliwości pierwszego formantu. W literaturze podaje się dla tej samogłoski wymówionej w izolacji częstotliwości pierwszego formantu rzędu 280-330 Hz, choć dla niektórych osób niekiedy spotyka się i wartości bliskie 400 Hz [CALIŃSKI i in., 1970]. Pomimo znacznej różnicy między konfiguracjami toru głosowego, w obu przypadkach częstotliwości drugiego i trzeciego formantu są zbliżone do siebie.

samogłoska /a/

powierzchnia



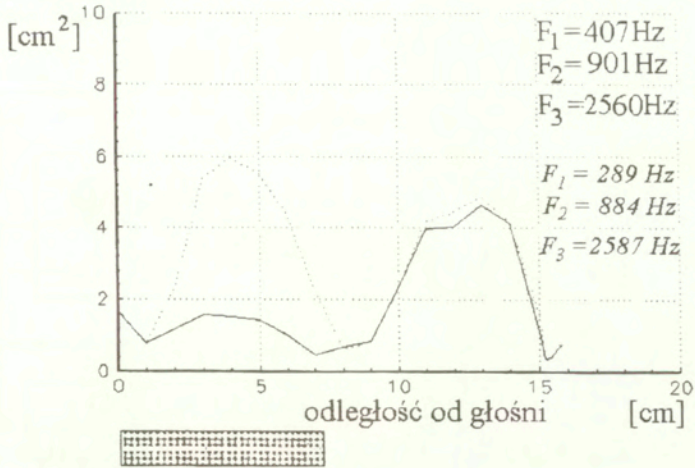
wymiar poprzeczny



Rys. 8.12 Wynik modelowania funkcji powierzchni przekroju (u góry) dla samogłoski /a/. U dołu podano odpowiadający jej przekrój wzdłużny toru. Kropkami zaznaczono wymiary poprzeczne toru głosowego podane w pracy [BOLLA&FÖLDI,1987].

## samogłoska /u/

powierzchnia

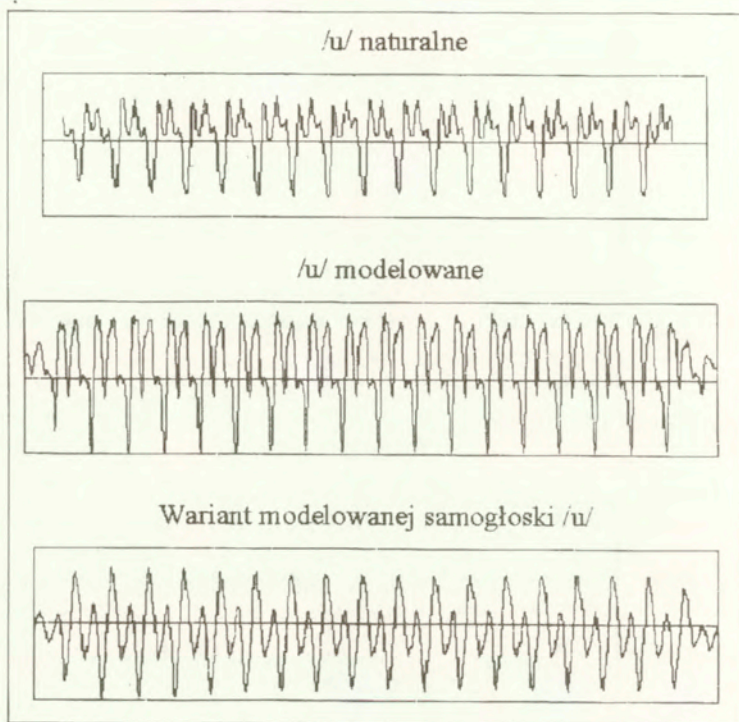


wymiar poprzeczny



Rys. 8.13 Wynik modelowania funkcji przekroju dla samogłoski /u/ (linia ciągła). W górnej części rysunku podano również typowy przebieg funkcji powierzchni przekroju dla tej samogłoski (linia kropkowana) wraz z odpowiadającymi tej konfiguracji częstotliwościami formantowymi (poniżej, kursywą).

Na Rys. 8.14 podano przebiegi czasowe dla samogłoski /u/: u góry - wypowiedzianej przez mówcę w wyrazie „góra”, w środku przebieg modelowany w oparciu o zarejestrowaną dla niej konfigurację geometryczną toru głosowego lektora, zaś u dołu - wyliczony sygnał dla funkcji powierzchni przekroju o typowym dla tej samogłoski, tzn. dwuwierzchołkowym kształcie. Oba syntetyczne warianty są bardzo dobrze identyfikowane, choć, wskutek dużej różnicy w częstotliwości pierwszego formantu, różnią się znacznie między sobą brzmieniem.



**Rys. 8.14** Przebiegi czasowe samogłoski /u/ naturalnej, modelowanej i wariantu modelowanego /u/. Przebiegi te różnią się przede wszystkim częstotliwością pierwszego formantu, która wynosi odpowiednio  $F1=414$ , 401 oraz 289 Hz.

W poniższej Tabeli 8.1, podano dla skrajnych w sensie artykulacyjnym samogłosek /i,a,u/, częstotliwości pierwszych trzech formantów pomierzonych dla modelowanych samogłosek naturalnych oraz syntetycznych. Różnice w pomierzonych częstotliwościach formantowych samogłosek naturalnych i ich odpowiedników

syntetycznych nie są duże (<4%). Stosunkowo są one największe dla samogłoski /a/ ( $\Delta F_1=1\%$ ,  $\Delta F_2=3.9\%$ ,  $\Delta F_3=2.6\%$ ), co może być spowodowane nie do końca dokładną przy modelowaniu rekonstrukcją konfiguracji geometrycznej toru głosowego, spowodowaną niepełnym jego opisem.

**Tab. 8.1** Porównanie częstotliwości formantowych samogłosek naturalnych i syntetycznych.

samogłoski	naturalne			syntetyczne		
	F <sub>1</sub> [Hz]	F <sub>2</sub> [Hz]	F <sub>3</sub> [Hz]	F <sub>1</sub> [Hz]	F <sub>2</sub> [Hz]	F <sub>3</sub> [Hz]
/i/	286	2246	2562	290	2239	2573
/a/	901	1250	2592	891	1289	2524
/u/	414	894	2552	407 (289)	901 (884)	2560 (2587)

Jak już wspomniano, nie było możliwości zweryfikowania odwzorowywanej dla danej głoski geometrii toru głosowego w przestrzeni trójwymiarowej z jego rzeczywistą konfiguracją. Z tego powodu starano się, by linia odwzorowująca przekrój podłużny toru przechodziła możliwie dokładnie przez punkty pomiarowe podane przez węgierskich autorów, zaś poza nimi dobierano ją tak, by uzyskać dokładnie te same częstotliwości formantowe, co w modelowanej samogłosce. Ponadto, dodatkowym kryterium poprawności doboru odpowiednich konfiguracji toru głosowego były przebiegi formantowe modelowanych przebiegów tranzjentowych w wyrazach wypowiedzianych przez lektora, dla którego wykonano zapisy kinorentgenograficzne. Warto podkreślić, że dokładność odwzorowania rozkładu częstotliwości formantowych i ich przebiegu zmian czasowych jest przez wielu autorów nadal podstawowym kryterium dokładności odtworzenia konfiguracji rzeczywistego toru głosowego [BEAUTEMPS i in., 1995; BADIN i in., 1998]

W programie *Model* nie ma ograniczeń przy modelowaniu dynamicznych zmian konfiguracji toru głosowego i w zasadzie maksymalna liczba konfiguracji geometrycznych toru odpowiadająca kolejnym punktom czasowym jest określona wielkością przydzielonej pamięci. W najprostszym przypadku, można dla każdego tranzjentu określić tylko początkową i końcową konfigurację geometryczną toru, natomiast pośrednie konfiguracje (tj. przekroje wzdłużne toru głosowego) są wyliczane poprzez aproksymację liniową dla zadanego czasu trwania przejścia. Oczywiście wyznaczone pola powierzchni przekrojów poprzecznych toru głosowego, a tym samym odpowiadające tym konfiguracjom chwilowe funkcje transmitancji toru głosowego są nieliniowymi funkcjami czasu. W przedstawionym na Rys. 8.6 sekwencji wartości

parametrów sterujących generowaniem sygnału dla sylaby /ajo/, pokazano że możliwe jest zadanie płynnej zmiany częstotliwości podstawowej F0, obwiedni amplitudy, a także obwiedni widma źródła kraniowego i charakterystyki promieniowania ust (w podanym przykładzie są one stałe i wynoszą 12 dB/okt).

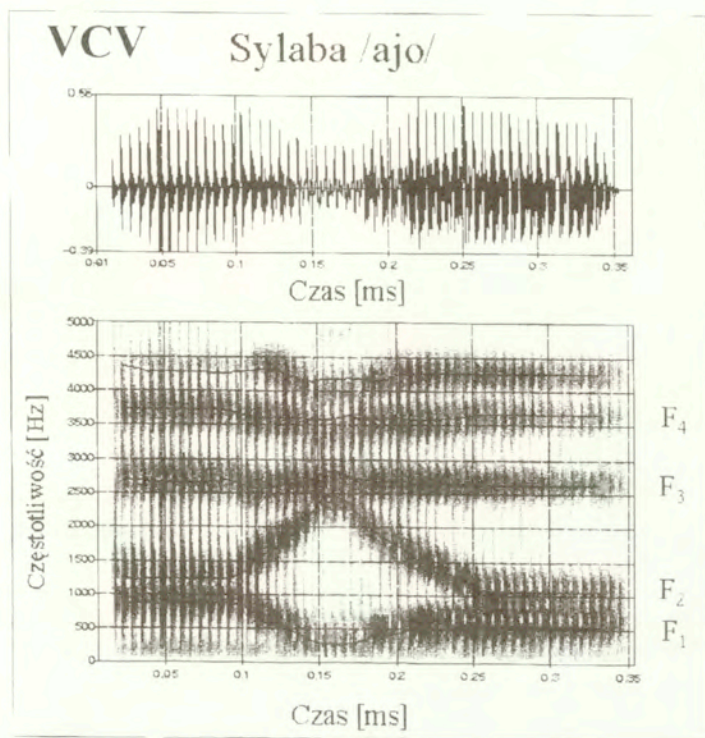
W tym przypadku choć konfiguracja nie zmienia się w przypadku artykulacji samogłoski /a/ i /o/, to w celu zaprogramowania przebiegu zmian konturu melodycznego, podane są wartości docelowe F0 wraz z ich położeniem czasowym i określeniem przy jakiej konfiguracji toru należy obliczać sygnał czasowy. W zaprogramowanym połączeniu sylabicznym /ajo/ podano 3 konfiguracje geometryczne odpowiadające samogłoskom /a/, /o/ i spółgłosce /j/ oraz 6 punktów czasowych z określonymi wartościami F0, wraz z czasami przejścia między nimi. Przykładowo, czas trwania samogłoski /a/ w obliczanym przebiegu został zaprogramowany na 120 ms i czas trwania tranzjentu do spółgłoski /j/ wynosi tu 70 ms, natomiast początkowa częstotliwość F0 dla tej samogłoski przyjęto 125 Hz i czas trwania tego segmentu wynosi 30 ms, następnie zaś częstotliwość wzrasta do 128 Hz po upływie 25 ms.

Przy modelowaniu opracowany program wyznacza najpierw konfiguracje pośrednie toru głosowego opisując w ten sposób przejście od konfiguracji początkowej odpowiadającej głosce /a/ do konfiguracji docelowej /j/, a następnie końcowej odpowiadającej samogłosce /o/. Dla każdej konfiguracji pośredniej określoną dla bieżącego punktu czasowego, wyliczana jest funkcja transmitancji toru, a następnie poprzez jej odwrotną transformatę Fouriera i jej splot z funkcją źródła wyznaczana jest funkcja czasową sygnału mowy. Konfiguracje pośrednie są wyznaczane z krokiem odpowiadającym przyjętej częstotliwości próbkowania, w tym przypadku równej 10 kHz. W obliczeniach zakłada się liniową zmianę funkcji przekroju podłużnego toru głosowego, co nie oznacza jednak, że funkcja powierzchni przekroju poprzecznego (a tym samym funkcja transmitancji toru głosowego) zmienia się w sposób liniowy.

Na Rys. 8.15 przedstawiony jest obliczony przebieg czasowy połączenia sylabicznego /ajo/ wraz z jego charakterystyką widmową. Na widmie tym zaznaczono liniami ciągłymi wyznaczone dla modelowanej sekwencji częstotliwości formantowe F1, F2, F3 i F4. Należy podkreślić, że wygenerowany przebieg nie modeluje rzeczywistego połączenia sylabicznego /ajo/, bowiem przyjęte konfiguracje samogłosek i spółgłoski /j/ zostały wyznaczone z ich danych artykulacyjnych podanych przez autorów węgierskich dla różnych wyrazów (/a/ z wypowiedzi w wyrazie „tak”, /o/ - w wyrazie „kot” i spółgłoska /j/ - „jama”). W tym przypadku konfiguracje geometryczne toru głosowego nie zostały wyznaczone z wyrazu zawierającego sekwencję /ajo/ i raczej chodziło o sprawdzenie, jak program *Model* wylicza przebiegi o tranzjentowym charakterze i czy sygnał w taki sposób wygenerowany jest percepcyjnie zbliżony do rzeczywistego. Nieformalna ocena odsłuchowa nie pozostawiała wątpliwości co do rozpoznania połączenia sylabicznego /ajo/, zaś jego naturalność brzmienia zależy przede wszystkim od właściwego odwzorowania konturu melodycznego. Ponieważ w



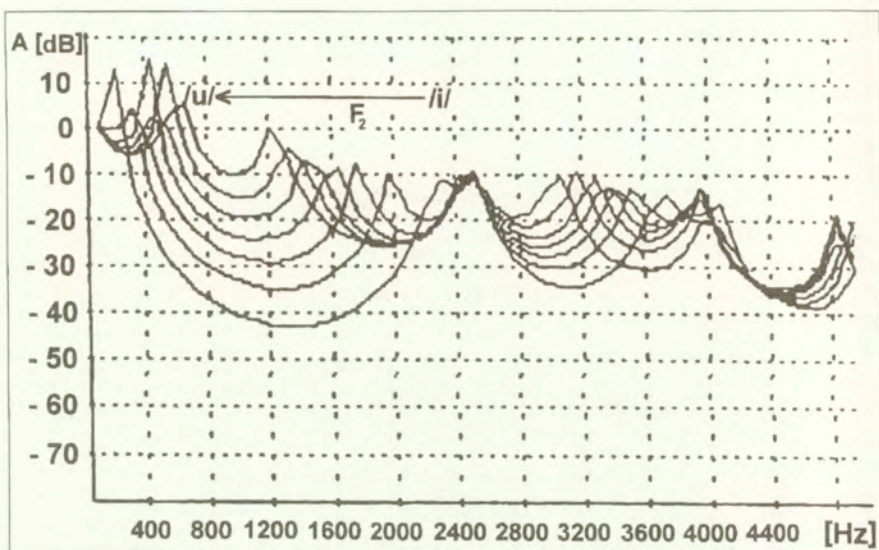
zarejestrowanym przez autorów węgierskich materiale nie występuje to połączenie sylabiczne, więc nie było możliwe zweryfikowanie przebiegu zmian częstotliwości formantowych w przebiegu syntetycznym z analogicznym przebiegiem /ajo/ wypowiedzianym przez lektora, dla którego wykonano aproksymacje konfiguracji geometrycznych toru głosowego. W przypadkach innych połączeń porównanie takie było z reguły wykonywane (patrz dalej) i podstawowym kryterium podobieństwa sygnału syntetycznego i naturalnego było podobieństwo przebiegów widmowych, a zwłaszcza przebiegów formantowych.



**Rys. 8.15** Przebieg czasowy oraz sonogram połączenia sylabicznego /ajo/ wygenerowanego w modelu przy wartościach funkcji sterujących podanych na Rys. 8.6. Liniami ciągłymi zaznaczono przebiegi formantowe wyznaczone z wygenerowanego sygnału.

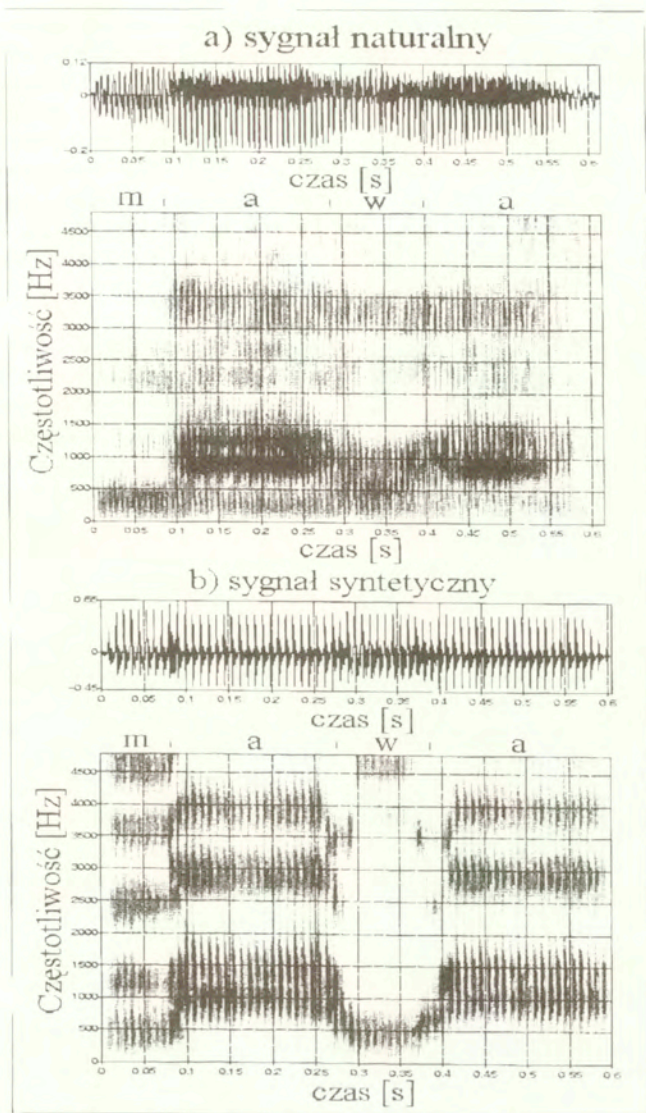
Na Rys. 8.16 przedstawiono, przykładowo przebieg zmian charakterystyki funkcji transmitancji wyliczony dla wybranych (7) pośrednich konfiguracji geometrycznych toru głosowego, wyznaczonych przez program *Model* dla fazy

przejścia między konfiguracjami samogłosek skrajnych /i/ oraz /u/. Konfiguracje toru głosowego dla artykulacji samogłoskowej dobrano tak, aby uzyskać w obu przypadkach częstotliwości formantowe zbliżone do wartości średnich podawanych dla mowy polskiej. Na rysunku tym obserwuje się, że wraz z przesuwaniem się ku tyłowi miejsca artykulacji samogłoskowej następuje charakterystyczne, bardzo wyraźne przemieszczanie się drugiego formantu w stronę niższych częstotliwości, od ok. 2200 Hz (dla /i/) do częstotliwości poniżej 800 Hz (dla /u/). Zaś dla pierwszego formantu ruch jego odbywa się w przeciwnym kierunku od ok. 280 Hz do ok. 450 Hz.



Rys. 8.16 Schematyczny przebieg zmian funkcji transmitancji toru głosowego przy zmianie jego konfiguracji odpowiadającej samogłosce /i/ w kierunku do konfiguracji odpowiadającej samogłosce /u/.

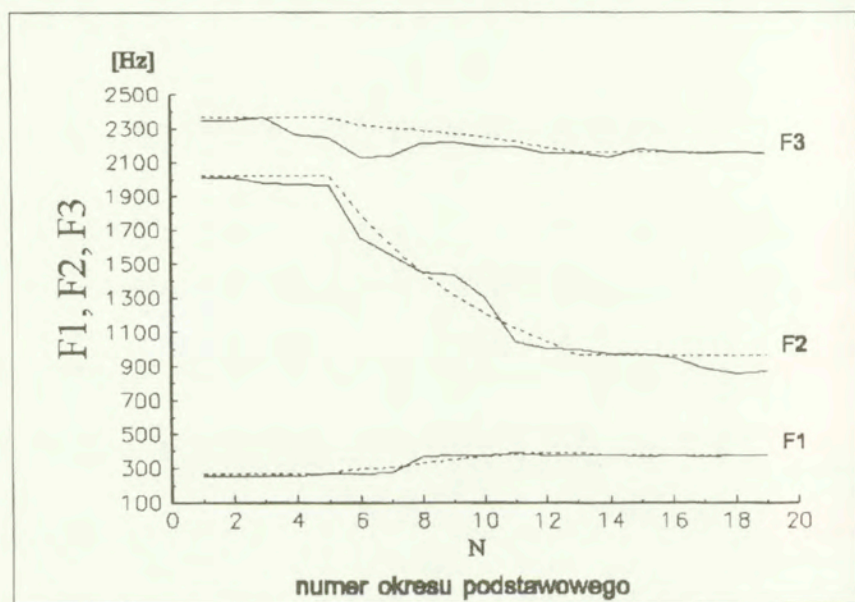
Przykładowy wynik zaprogramowania bardziej złożonego przebiegu, będącego sekwencją rezonantów /mawa/ w wyrazie „mała” przedstawiono na Rys. 8.17. U góry, wraz z przebiegiem czasowym podano sonagram tego wyrazu wypowiedzianego przez lektora, zaś u dołu sygnał wraz z tą samą sekwencją rezonantów wyliczony w oparciu o zaprogramowaną, zmienną w czasie konfigurację geometryczną toru głosowego. Jako dane wejściowe dla procedury *Model* podawane są konfiguracje geometryczne odpowiadające kolejnym głoskom wraz z zadaniem ich czasem trwania oraz czasami przejść od jednej do następnej głoski.



Rys. 8.17 Przebieg czasowy i widmowy sygnału mowy dla wyrazu “mała” /mawa/ wypowiedzianym przez lektora (a) oraz dla sygnału syntetycznego utworzonego z tej samej sekwencji głosek, wyznaczonego w oparciu o docelowe konfiguracje geometryczne toru głosowego odpowiadające poszczególnym głoskom (b).

Przedstawione obrazy widmowe sygnału naturalnego i syntetycznego są zbliżone do siebie. Jednakże są widoczne pewne różnice między nimi, jak na przykład, nieco wyższa częstotliwość drugiego formantu dla samogłoski /a/, znacznie wyższa częstotliwość czwartego formantu dla tej samej głoski, oraz stosunkowo niski poziom trzeciego formantu dla głoski /w/. Różnice te wynikają m.in. z tego, że dla głoski /a/ podana w cytowanej pracy konfiguracja toru głosowego była wyznaczona, jak już wspomniano, dla wyrazu „tak”, a nie „mała”. Bardziej dokładne porównywanie przebiegów czasowych i widmowych jest jednak utrudnione z powodu stosunkowo małego poziomu sygnału mowy do zakłóceń, zarejestrowanego podczas wykonywania zdjęć kino rentgenograficznych.

W tej sytuacji wydaje się, że jedynie dokładna analiza formantowa, mimo znacznych swoich ograniczeń, pozwala na bardziej zobiektywizowany sposób porównywania modelowanych przebiegów z ich naturalnymi odpowiednikami, tak jak to przedstawiono na Rys. 8.18. Pokazano na nim przykładowo realizację tranzjentu /j-u/ na płaszczyźnie formantowej, wymówionego w wyrazie "tiul" (linie ciągłe) przez lektora, dla którego były wykonywane kinorentgenogramy.



Rys. 8.18 Przebiegi częstotliwości formantowych w diadzie /j-u/ wyznaczonych synchronicznie z okresem tonu krztaniowego w mowie naturalnej (w wyrazie „tiul”) i syntetycznej (linia kropkowana).

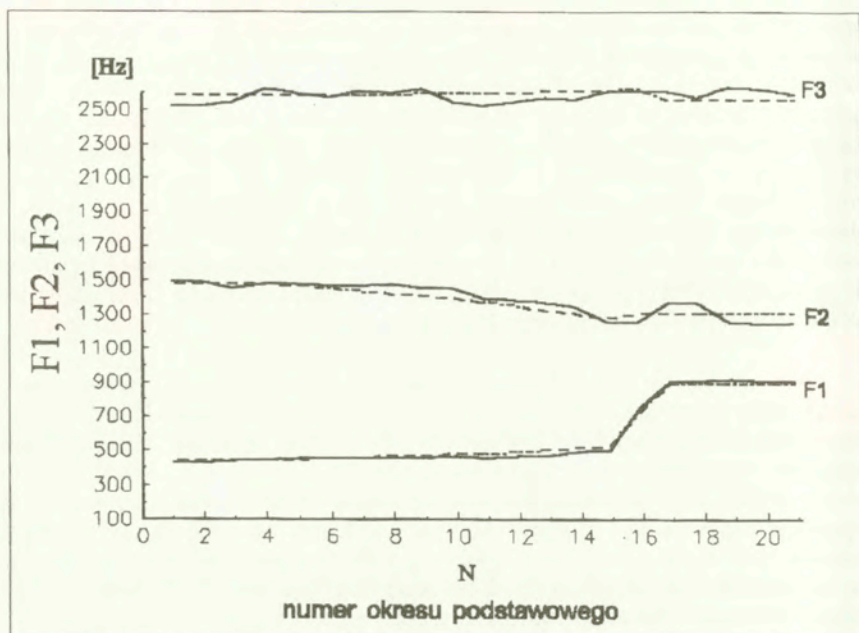
Na tej samej płaszczyźnie naniesiono przebieg tego samego tranzjentu (linie kreskowane), uzyskany w wyniku dynamicznego modelowania zmian konfiguracji geometrycznej toru głosowego od konfiguracji odpowiadającej głosce /j/ do konfiguracji odpowiadającej głosce /u/ (patrz Rys. 8.11 i 8.13). Jak widać z podanych przebiegów artykulacja samogłoski /u/ przez lektora w wyrazie „tiał” jest bardzo zbliżona do tej co w wyrazie „góra”.

Oba przebiegi wyznaczono synchronicznie z okresem tonu krtaniowego, wprowadzając tym samym swoistego rodzaju normalizację czasową, umożliwiającą ich bezpośrednie porównanie ze sobą. Z porównania przebiegu zmian częstotliwości formantów wraz z kolejnymi okresami tonu krtaniowego, widać że w przyjętym systemie modelowania, zmiana częstotliwości podstawowej ma wpływ na przebieg tranzjentowych zmian częstotliwości formantowych, pomimo że wartości sygnału wyjściowego w programie *Model* są wyznaczane asynchronicznie do F0, zgodnie z przyjętą częstotliwością próbkowania sygnału. W podanym przykładzie, ze względu na niestacjonarny charakter spółgłoski /j/, trudno postawić wyraźną granicę między nią i samogłoską /u/ po niej następującą, pomimo bardzo szczegółowej analizy formantowej tej sekwencji. Można co najwyżej stwierdzić, że przejście od głoski /j/ do samogłoski /u/ następuje w okolicach 9-12 okresu tonu podstawowego.

Nie mniej, w przypadku gdy dąży się do dokładnego odwzorowania rzeczywistego przebiegu formantów nie zawsze wystarczy podanie dla każdej z głosek tworzących sylabę jednej tylko konfiguracji toru. Na Rys. 8.19 przedstawiony został przypadek modelowania sylaby /la/ wypowiedzianej w wyrazie „lato”. Przejście między głoską /l/ i /a/ następuje między 15 a 17 okresem tonu krtaniowego. W segmencie odpowiadającym głosce /l/ wyraźnie można wyodrębnić dwa człony – pierwszy (od 1 do 15 okresu tonu krtaniowego), który odpowiada wolno zmieniającej się konfiguracji toru głosowego podczas artykulacji głoski /l/, oraz drugi, podczas którego następuje szybkie przejście do głoski /a/. Warto jednak zaznaczyć, że w praktyce pierwszy człon głoski /l/ można by zamodelować jako stacjonarny (o długości ok. 100 ms), i nie wpłynęłoby to w istotny sposób na zrozumiałość tej sylaby. Pociągnęłoby to tylko zwiększenie nieco różnic w ogólnym obrazie przebiegów formantowych wyznaczonych dla mowy naturalnej oraz syntetycznej.

Porównując z przebiegami w mowie naturalnej w dziedzinie widmowej i formantowej, tak jak to przykładowo zostały przedstawione na Rys. 8.17 i 8.18 posłużyły do programowania iloczynów poszczególnych faz artykulacji modelowanych sylab, czy wyrazów. W tym celu, dokonano szeregu analiz realizacji połączeń rezonantów typu VC, CV i VCV dla mówcy, którego głoski były modelowane. Jednocześnie dokładność odwzorowania przebiegów widmowych, a zwłaszcza formantowych stanowiło podstawowe kryterium w ocenie jakości modelowania. Niestety posiadane dane dotyczące konfiguracji geometrycznych toru głosowego wyznaczone w oparciu o zapisy kinorentgenograficzne są niepełne, i z tego powodu nie

było możliwe w pełni dokładne odwzorowanie w modelu artykulacyjnym wszystkich możliwych połączeń rezonantów istniejących w języku polskim i zweryfikowanie ich z realizacjami w mowie naturalnej. A więc, gdy realizowano połączenie nie występujące w nagraniach autorów węgierskich, starano się wówczas wykorzystując dane dotyczące konfiguracji głosek wypowiedzianych w innych kontekstach, zrealizować przebieg podobny (w płaszczyźnie widmowej i formantowej) do rzeczywistego zarejestrowanego przez inną osobę.



Rys 8.18 Przebieg częstotliwości formantowych w diadzie /la/ w wyrażeniu „lato” wyznaczonych synchronicznie z okresem tonu krztaniowego w mowie naturalnej i syntetycznej (linia kropkowana).

Dla stosunkowo małej liczby połączeń przebiegi formantowe były porównywane w nieliniowej skali czasu, której jednostką czasu stanowił jeden okres pobudzenia krztaniowego, tak jak to przedstawiono na Rys. 8.18 i 8.19. Jednakże pomiar tego typu był wykonany w sposób ręczny i nie można było tego zrealizować dla większego zbioru danych. Dlatego w większości przypadków wyznaczano częstotliwości formantowe w sposób zautomatyzowany z krokiem 5 ms, ręcznie korygując ewentualne błędy, występujące zwłaszcza dla spółgłosek nosowych oraz głosek nazalizowanych (np. "na styku" samogłoski ustnej i spółgłoski nosowej), a także w przypadkach "bliskiego"

położenia sąsiadujących ze sobą formantów, wskutek czego były interpretowane jako jedno maksimum (np. miało to niekiedy miejsce dla samogłoski /a/ - F1 i F2, czy spółgłosek /i,j/ - F2 i F3). Jak już wcześniej wspomniano (punkt 8.3), występowanie w przypadku głosek nazalizowanych lub spółgłosek nosowych tzw. antyformantu deformuje strukturę formantową charakterystyki widmowej, powodując często tzw. "rozszczepienie" pierwszego lub drugiego formantu i pojawienie się co najmniej jednego dodatkowego lokalnego maksimum tzw. fałszywego formantu, który w mowie naturalnej niekiedy trudno odróżnić od rzeczywistego, bowiem w naturalnym sygnale nie jest możliwe wyznaczenie charakterystyki fazowej.

#### 8.4. Wyniki modelowania połączeń polskich rezonantów

W języku polskim nie istnieje ostra granica między samogłoskami i spółgłoskami. Oprócz samogłosek właściwych, ustnych czy nazalizowanych istnieją głoski posiadające w swojej artykulacyjnej, a także i akustycznej naturze cechy zarówno samogłosek, jak i spółgłosek [WIERZCHOWSKA,1981]. Do tej grupy zaliczają się przede wszystkim tzw. głoski przejściowe /j,w/, czasami w literaturze nazywana półspółgłoskami (semi-vowels) lub głoskami tranzjentowymi. Głoski te są w języku polskim reprezentowane w transkrypcji ortograficznej postaci jako "i", jak w wyrazie „tiul”, lub "j" („kij”) oraz jako "u" („auto”), lub "ł" (cały”). Głoska /j/ ma zdecydowane pokrewieństwo artykulacyjne z samogłoską /i/. Artykulacja /j/ jest miękka, średnio-językowo, średnio-podniebienna. Porównując palatogramy samogłoski /i/ i spółgłoski /j/ nie obserwuje się znaczących różnic w kształcie i wielkości powierzchni styku języka i podniebienia. Głoska /j/ przy miejscu artykulacji i stopniu otwarcia niewiele różniącym się od /i/ odznacza się krótszym czasem trwania, w stosunku 1:1.85 [ŁOBACZ,1970], zwłaszcza części ustalonej. Jak można było przekonać się z Rys. 8.11, konfiguracje toru głosowego odpowiadające tym głoskom różnią się przede wszystkim w przedniej, ząbówkowej części toru.

Z akustycznego punktu widzenia głoska /j/ jest przebiegiem niestacjonarnym, o formantach docelowych niemal takich samych jak dla samogłoski /i/. Dźwięk /j/ wymawia się w języku polskim najczęściej na pograniczu sylab. Jest głoską, która nie może istnieć bez co najmniej jednostronnego sąsiedztwa samogłoskowego i nigdy nie staje się ośrodkiem sylaby.

Głoskę /j/ modelowano w typowych pozycjach - nagłosu, wygłosu i śródgłosu, w otoczeniu samogłoskowym. Głoska /j/ wymawiana w nagłosie charakteryzuje się bardzo krótkim czasem trwania fragmentu ustalonego (często segment początkowy sylaby tego typu nie ma wyraźnej części ustalonej). Przy modelowaniu sylaby, czas trwania początkowej części ustalonej wynosił nie więcej niż 20-25 ms (wydłużenie tego czasu powodowało percepcję samogłoski /i/, po której słyszana była głoska /j/). Następujący długi tranzjent do samogłoski trwa ok. 120-130 ms, po którym następuje

samogłoska o czasie trwania nie mniejszym niż 100-120 ms, tak by brzmienie sylaby było zbliżone do naturalnego (by nie było wrażenia "przycięcia" samogłoski). Modelowane sylaby brzmiały bardzo zrozumiale i słuchacze uczestniczący w nieformalnych odsłuchach często nie byli w stanie odróżnić je od izolowanych sylab nagranych dla głosu o zbliżonej częstotliwości podstawowej.

Modele sylab, w których głoska /j/ jest w wygłosie, badano na przykładach połączeń -ej, -aj, -oj. Samogłoski we wszystkich modelowanych sylabach trwały średnio 120-140 ms, po czym następował tranzjent długości 70-80 ms do konfiguracji docelowej odpowiadającej głosce /j/. Czas trwania tej konfiguracji wynosił średnio 60-70 ms. Przy krótszym czasie trwania odnosi się wrażenie "przyciętej" od końca sylaby, natomiast wydłużanie jej powodowało percepcję dodatkowej, końcowej samogłoski /i/, to jest na przykład, zamiast /aj/ było postrzegana sekwencja /aji/.

Głoskę /j/ w śródgłosie modelowano w połączeniach typu: ajo, oja, aja czy eja. Przejście od samogłoski poprzedzającej do /j/ trwało we wszystkich przypadkach średnio 70 ms. Czas trwania konfiguracji docelowej /j/ nie przekraczał 25 ms, po której następował kolejny tranzjent do samogłoski kończącej sylabę i wynoszący ok. 90-120 ms. Podobnie jak w poprzednich przypadkach iloczasy samogłoski następującej po /j/ wynosił ok. 120 ms.

Zbliżone do wyznaczonych dla /j/ czasy przejścia otrzymano dla spółgłoski "ł" (/w/). Warto podkreślić, że /w/ w tej pozycji często występuje w wyrazach zapożyczonych, w których po /a/ i po /e/, jest realizowana przez drugą część samogłoskowego dyftongu /au/, pod względem funkcjonalnym równającym się spółgłosce. W wyrazach typu *auto, skaut, autor, automat* czy *Europa* występuje właśnie tego typu realizacja (np. *auto* --> /awto/).

Realizacja głoska /w/ w nagłosie była projektowana dla sylab typu /wa, we, wo, wi/. We wszystkich modelowanych sylabach, z wyjątkiem /wi/ głoska ta trwa krótko, w granicach 45-60 ms. Optymalny czas przejścia do samogłoski wynosi ok. 80 ms. Czas trwania samogłoski nie wpływa na zrozumiałość sylaby jedynie na jej akcentowanie. Oddzielnego projektu wymagała diada /wi/. Połączenie tego typu w języku polskim nie występuje w wyrazach, a jedynie w mowie ciągłej na ich granicy. W wyniku eksperymentów przeprowadzonych na modelu oraz odpowiednich analiz rzeczywistego sygnału okazało się, że czas ustalony głoski /w/ trwa maksimum 10 - 15 ms, po czym następuje długi tranzjent o czasie trwania 100-120 ms do samogłoski /i/, która w tym przypadku trwa stosunkowo krótko bo, 30 -50 ms. Wydłużanie czasu ustalonego dla /w/ powoduje odbiór tej sylaby jako /ui/, a nie /wi/.

Głoskę /w/ w wygłosie projektowano w połączeniach typu /iw, aw, ow, ew/. Optymalny czas trwania segmentu przejściowego od samogłosek do /w/ wynosił



ok.100-120 ms. Dopuszczalny jest krótszy tranzjent, aż do 80 ms, ale nie należy go bardziej skracać, ponieważ w sposób wyraźny spada zrozumiałość tej głoski. Optymalny czas trwania segmentu ustalonego /w/ wynosił w granicach 30-45 ms.

W śródgłosie głoska /w/ została zrealizowana w sylabach typu /awa, ewa, awo, awe/. Przejście czasowe od samogłoski trwającej około 70-90 ms do głoski /w/ wynosi we wszystkich przypadkach 50 ms. Konfiguracja ustalonego segmentu /w/ trwa ok.75 ms. Iloczas tranzjentu do samogłoski następującej po /w/ jest nieco większy od początkowego i wynosi ok. 70 ms. Następująca po /w/ samogłoska może mieć czas trwania 80-140 ms i o wyborze odpowiedniej wartości decyduje jedynie ocena naturalności brzmienia.

Odrębną ze względu na artykulację grupę modelowanych głosek stanowią spółgłoski ustne /r, l/ oraz nosowe /m,n,ɲ/. Wyróżnia je występowanie dodatkowego segmentu wokalicznego w przypadku wymawiania ich w izolacji. Głoski te wymawiane w izolacji, nie zaczynają się od swej spółgłoskowej artykulacji, ani się na niej nie kończą, ale rozwijają przed nią (lub po niej) wstępny element wokaliczny. Podobnie jest niekiedy, gdy są wymawiane w sąsiedztwie spółgłosek, zwartych czy szczelinowych. Oddziela je od nich wówczas krótki (20-40 ms) element wokaliczny. Tylko obustronne sąsiedztwo samogłosek właściwych wypiera niejako, lub zastępuje go. W wyniku badań przy zastosowaniu opisanego modelu artykulacyjnego otrzymano, że element wokaliczny w odbiorze percepcyjnym przypomina głoskę pośrednią między /a/ i /e/, bliską tzw. neutralnemu "e" (w transkrypcji /ə/).

Głoska /l/ ze względu na swoją specyficzną artykulację, podczas której następuje częściowe tylko przymknięcie toru głosowego, w wyniku czego powstają dwa boczne kanały wokół koniuszka języka, nie mogła być odwzorowana w modelu w sposób podobny, jak to przyjęto dla pozostałych głosek. Ponadto, brak w pracy [BOLLA&FÖLDI,1987] danych dotyczących rozmiarów bocznych kanałów w przedniej części toru głosowego, spowodował, że w modelowaniu tej głoski przednia, ząbowa część toru była odwzorowywana przez przewężenie, którego powierzchnia przekroju była równoważna powierzchni tych otworów, w sensie częstotliwości formantowych. Takie uproszczenie może spowodować, że przy modelowaniu przebiegów tranzjentowych, uzyskane reguły połączeń głoski /l/ z samogłoskami są słuszne dla sygnału syntetycznego, ale mogą stanowić niekiedy, jedynie pewne przybliżenie zjawisk zachodzących w mowie naturalnej.

W języku polskim głoska /l/ nie jest wymawiana przed samogłoską /l/. Wygenerowane połączenia tranzjentowe /la, lo, le/ wykazały, że konfiguracja ustalona głoski /l/ może trwać ok. 80-100 ms, zaś przejście do samogłoski musi zawierać się w przedziale 15-20 ms. Iloczas samogłoski może wynosić ok.100 ms. Artykulacja w wygłosie głoski /l/ była symulowana w sylabach typu /al,e,l,o/. Czas trwania samogłoski początkowej wynosił 100-120 ms. Czas przejścia do /l/ jest nieco większy,

niż przy modelowaniu tranzjentu w nagłosie i musi zawierać się w granicach 40-60 ms, tak by osiągnąć możliwe naturalne brzmienie tej głoski. Segment ustalony /l/ w syntetycznych sylabach trwa ok. 100 ms. W sygnale rzeczywistym izolowanej sylaby czas artykulacji /l/ wynosił średnio 120-160 ms. W wyrazach zakończonych np. na -al /np. kowal/ końcówka /al/ trwa jednak znacznie krócej, niż w przypadku wymawiania jej w izolacji i całkowity czas jej artykulacji wynosi ok. 160 ms.

W śródgłosie modelowano sylaby typu /ala, ale, ali, elo, ole, olo, ule, uli/. W sylabach przejście od samogłoski /u,a,e,i/ w kierunku /l/ może zawierać się w przedziale 30-65 ms, zaś ustalona konfiguracja /l/ może trwać od 45 do 65 ms, a więc znacznie krócej, niż gdy jest wymawiana w nagłosie lub wygłosie. Natomiast przejście od samogłoski tylnej /o/ do /l/ wymaga czasu trwania rzędu 65-80 ms. W tym przypadku ustalona konfiguracja /l/ jest krótsza i trwa 30-40 ms. Przejście od /l/ w kierunku następującej po nim samogłoski nie wykazuje znaczących różnic pod względem typu samogłoski i wynosi w granicach 20-35 ms.

Przytoczone wyżej reguły programowania przebiegów tranzjentowych w sylabach zawierających spółgłoskę /l/ w śródgłosie dają w wyniku przebiegi widmowe odbiegające w sposób zasadniczy od przebiegów wyznaczonych w mowie naturalnej dla podobnych sylab wypowiedzianych w izolowanych wyrazach. Pewne różnice w iloczasach mogą wynikać z powodu percepcyjnej optymalizacji modelowanej sylaby, która wymówiona w izolacji często trwa nieco dłużej, niż te same połączenie głosek wypowiedziane w wyrazie. Jednak rozrzut iloczasów podany w regułach modelowania głoski /l/ jest na ogół zbliżony do iloczasów wyznaczonych dla mowy naturalnej.

W przypadku spółgłosek nosowych, ze względu na częstość występowania w mowie naturalnej zajęto się przede wszystkim modelowaniem artykulacji i połączeń kontekstowych dla /m/ i /n/.

Częstotliwość pierwszego antyformantu nie zawsze zresztą widocznego w widmie spółgłosek nosowych wypowiedzianych w mowie ciągłej, jest ściśle związana z odległością między miejscem zwarcia ustnego i miejscem sprzężenia torów artykulacyjnych. Niesie ona pośrednią informację o efektywnej długości toru gardłowo-ustnego biorącego udział w artykulacji spółgłoski nosowej.

Jak już wspomniano wcześniej, przy artykulacji głosek /m/ i /n/ niekiedy występuje, podobnie jak dla /t/ czy /l/, element wokaliczny, lecz zanika on o wiele częściej, niż dla głosek ustnych tej samej grupy. Jedną z przyczyn jest brak ograniczeń, co do czasu trwania fragmentu ustalonego głosek nosowych. W przypadku, gdy element wokaliczny zaniknie lub zostanie zastąpiony inną samogłoską, to i tak fragment

ten może trwać dość długo, na przykład, ok. 100-130 ms nie tracąc przy tym nic na zrozumiałości.

Głoskę /m/ modelowano i analizowano w różnych położeniach kontekstowych. Utworzono sylaby, w których /m/ występuje w nagłosie typu: /ma, me, mo, mi/. Czas trwania ustalonej konfiguracji /m/ w nagłosie wynosił 120 ms. Przejście do samogłoski następującej po /m/ jest stosunkowo krótkie i wynosi 20-40 ms. Granica ta jest dość wyraźna i w zasadniczy sposób wpływa na zrozumiałość sylaby. Przedłużanie tranzjentu do samogłoski powoduje, że głoska /m/ traci na wyrazistości i zaczyna często być odbierana jako coś pośredniego między /m/ i /n/. Ponadto obserwuje się bardzo wyraźny wpływ głoski nosowej na samogłoskę sąsiednią, objawiający się obecnością cech nosowości na widmie samogłoski. W efekcie samogłoska ta zostaje wyraźnie podzielona na część nazalizowaną i ustną. Efekt ten jest wyraźny na widmie modelowanego sygnału oraz sygnału rzeczywistego oraz znajduje wyrazne potwierdzenie w odbiorze percepcyjnym. Czas trwania części ustnej samogłoski nie wymaga ścisłych reguł i decyduje on tylko o zaakcentowaniu jej w sylabie.

Głoska /m/ w wygłosie modelowano była w sylabach typu /am, em, om, im/. Podobnie jak w przypadku nagłosu /m/ silnie oddziałuje na samogłoskę, w tym przypadku poprzedzającą. Po części zdecydowanie ustnej samogłoski następuje część nazalizowana, której iloczas wynosi ok. 25-40 ms. Sygnał odpowiadający konfiguracji ustalonej /m/ ma amplitudę mniejszą niż sąsiadująca samogłoska i trwa w granicach 100-130 ms. Przedłużanie tej części sylaby powoduje przesadne jej przeciąganie, przez co brzmienie całej sylaby jest mało naturalne.

Do badania głoski /m/ w sródgłosie posłużyły modelowane połączenia typu /ama, ema, ime, emo, amo, ima/. W przypadku, gdy /m/ występowało w sródgłosie zostały potwierdzone reguły modelowania faz tranzjentowych sylaby ustalone dla pozycji nagłosu i wygłosu. Analizowane sylaby trwały średnio 400 ms, podobnie jak zmierzone na sygnale rzeczywistym z nagrań dla izolowanych sylab. Przejście od nazalizowanego fragmentu samogłoskowego do /m/ wynosiło 15 ms, zaś stan ustalony głoski /m/ trwał 120 ms. Czas ten może być krótszy, lecz nie powinien wynosić mniej niż 70 ms. Przejście od /m/ do nazalizowanej części samogłoski wynosi 15-20 ms. Ustalenie jak długo słyszy się wpływ głoski nosowej na samogłoskę po niej występującą jest wynikiem szczegółowych badań na sygnale modelowanym i rzeczywistym.

Długość toru ustnego biorącego udział w artykulacji głoski /n/ wynosi ok. 14-15 cm, tym samym częstotliwość pierwszego antyformantu jest wyższa niż w przypadku /m/ i wynosi ok. 1200-1400 Hz (dla /m/ wynosi ona ok. 800-900 Hz). Głoska /n/ nie występuje w mowie polskiej bezpośrednio przed /i/ oraz /j/. Mamy wtedy do czynienia

z oddzielną głoską nosową w transkrypcji fonetycznej oznaczaną /n/, dla której częstotliwość antyformantu jest bliska 3000 Hz.

Głoskę /n/ w nagłosie testowano w sylabach typu /ne, na, no/. Część ustalona /n/ w sylabach wymienionego typu może trwać od 50 do 70 ms. Percepcja /n/ jest odbierana faktycznie przez znacznie dłuższy czas, niż by to wynikało z iloczasu ustalonego /n/, ponieważ po nim następuje dłuższy niż dla /m/ tranzjent do samogłoski, który percepcyjnie jest też odbierany jako część tej spółgłoski. Przejście to wynosi 40-65 ms, po czym następuje nazalizowana część samogłoski trwająca ok. 40 ms. Dalej (po upływie ok. 40 ms) zaczyna się jej ustny wariant. Sylaby syntezowane z bezpośrednim przejściem od /n/ do samogłoski ustnej, przy pominięciu ustalonej części nazalizowanego fragmentu samogłoskowego wykazywały znacznie mniejszą zrozumiałość i rozpoznawanie ich niekiedy było wątpliwe. Świadczyłyby to, że przejście do konfiguracji samogłoskowej jest szybsze niż zamknięcie wlotu do toru nosowego.

Głoska nosowa /n/ w pozycji wygłosu była modelowana dla sylab typu /an, en, on, in/. We wszystkich sylabach tranzjent od nazalizowanej części samogłoski do /n/ wynosił 25-30 ms. Czas przejścia od części ustnej do nazalizowanej nie wpływa w stopniu istotnym na zrozumiałość sylaby. W modelowanych sylabach wynosił on ok. 30 ms. Część ustalona głoski /n/ może trwać w tym przypadku wyjątkowo długo, bowiem nawet 140 ms. Nie zmienia to zrozumiałości syntetycznej sylaby, jedynie podkreśla jej wybrzmiewanie. Zrozumiałość wszystkich sylab uznawana była jako bardzo dobra.

W pozycji śródgłosu /n/ modelowano w połączeniach: /ana, ane, ina, ine, ona/. Wpływ spółgłoski nosowej na sąsiednie samogłoski jest w obu kierunkach podobny, jak w omówionych wyżej przypadkach (jest on symetryczny). Nazalizowany segment samogłoskowy trwa ok. 30-40 ms. Czas trwania segmentu ustalonego /n/ może ulegać zmianom w granicach 60-120 ms. Czas trwania samogłosek dobierano pod względem naturalności brzmienia połączenia. Przy zachowaniu tych reguł, a w szczególności przestrzeganiu odpowiednich iloczasów (ok. 30 ms) tranzjentu w kierunku /n/ i z /n/ do nazalizowanej części samogłoski, brzmienie modelowanych sylab oceniano jako bardzo dobre.

Jak już wspomniano specyficzną klasę głosek w języku polskim stanowi głoska drżąca /r/. Na ogół jest ona dźwięczna, często wymawiana wraz z krótkim elementem wokalicznym o charakterze samogłoskowym (por. [DŁUSKA,1981]) i przy jego artykulacji konfiguracja toru głosowego przyjmuje kształt pośredni między konfiguracją odpowiadającą głosce /a/ a konfiguracją odpowiadającą głosce /e/. Czas trwania segmentu wokalicznego poprzedzającego krótki segment ustalony /r/ wynosi ok. 10-15 ms. Natomiast czas trwania elementu wokalicznego kończącego głoskę /r/ może być nieco dłuższy (15-25 ms). Iloczas głoski /r/ jest więc zazwyczaj krótki,

rzadko przekracza 30-40 ms. Przedłużanie artykulacji pojedynczego elementu spółgłoskowego nie może odbyć się bez zmiany brzmienia sylaby. W przesadnej wymowie /r/ rzadko przekracza się trzy drgnięcia przedniej części języka, nawet w wygłosie. Przy wymowie spokojnej /r/ wykonywane jest tylko jedno krótkotrwałe zwarcie. Wynika stąd, że powtarzanie się po sobie kolejnych zwarć nie jest dla tej głoski nieodzownym elementem sposobu artykulacji, a jedynie służy jej podkreśleniu i przedłużeniu.

Gdy głoska /r/ jest wymawiana w nagłosie lub wygłosie zawsze można wyodrębnić trzy fazy artykulacyjne, których kolejność zależy od położenia, a także od kontekstu – to jest artykulacja segmentu wokalicznego, zwarcie pełne (lub częściowe) oraz przejście do/od samogłoski. Konieczność wykonywania przez organ artykulacyjny za każdym razem wszystkich faz składających się na artykulację /r/ jest najważniejszą cechą artykulacyjną tej głoski. Należy jednak zaznaczyć, że gdy jest ona wymawiana w śródgłosie następuje zlanie się elementu wokalicznego z sąsiadującym segmentem samogłoskowym

W wygłosie badano spółgłoskę /r/ w modelowanych sylabach typu /mar, mer, mor, mir/. Przyczyną wybrania tego typu sylab były zdecydowanie złe rezultaty badań percepcyjnych sylab modelowanych bez udziału spółgłoski początkowej. Podobnie źle były identyfikowane tego typu diady "wycięte" z mowy naturalnej. Przyczyny tego zjawiska nie zostały wyjaśnione. Drżące /r/ w wygłosie trwa dłużej niż w pozostałych pozycjach i niekiedy występują w niej dwa zwarcia i dwa krótkie segmenty wokaliczne. Użykuje się w ten sposób podkreślenie drżenia. Syntetyczna głoska /r/ w wygłosie posiada krótki, ok. 10 ms element wokaliczny, po którym następuje tranzjent ok. 10-15 ms do /r/. Konfiguracja głoski /r/ trwa 10-15 ms, po czym następuje kolejne przejście ok. 15 ms do elementu wokalicznego. Kończący element wokaliczny trwa ok. 15 ms. W przypadku modelowania dwukrotnego zwarcia, następuje przejście (ok. 10-15 ms) do kolejnego elementu wokalicznego, po którym występuje ponowny tranzjent do segmentu /r/ ustalonego (10 ms), o co najmniej dwukrotnie zmniejszonej amplitudzie w stosunku do pierwszego elementu wokalicznego.

Analizy kolejnych symulacji głoski /r/ wykazały, że:

- element wokaliczny poprzedzający /r/ musi być krótki i jego czas trwania nie może przekraczać 15 ms,
- czas przejścia od elementu wokalicznego do /r/ wygłosowego nie przekracza 30 ms, w przeciwnym razie może zostać odebrany jako oddzielny dźwięk, zbliżony do /e/; najlepsze brzmienie uzyskano przy czasie przejścia ok. 10-15 ms,
- ustalony fragment głoski /r/ nie powinien trwać zbyt długo (nie więcej niż 40 ms), ponieważ może sylaba może zostać odebrana ze słyszalnym /l/ w wygłosie,

- czas tranzjentu do kończącego elementu wokalicznego nie powinno przekraczać 25 ms, zaś sam element wokaliczny nie może trwać dłużej niż 20-25 ms, aby uniknąć w fazie wybrzmiewania percepcji lekko słyszalnego /e/.

Analizy połączeń typu /mar, mer, mor, mir/ wykazały zależność optymalnej, ze względu na brzmienie /r/ wygłosowego, długości tranzjentu od rodzaju samogłoski. Dla przejścia od samogłoski środkowej /a/ do /r/ przyjęto długość tranzjentu rzędu 60-80 ms, od samogłosek przednich /e/ oraz /i/ - 40-60 ms, a od samogłosek tylnych /o/, /u/ - 20-25 ms.

Głoskę /r/ w nagłosie symulowano według projektu zawierającego element wokaliczny o nieco dłuższym, niż w poprzednich przypadkach iloczynie - ok. 20-25 ms, po którym następuje krótkie przejście do /r/, ok. 10-15 ms. Konfiguracja głoski /r/ w nagłosie utrzymywana była przez 15 ms. Długość tranzjentu do końcowego elementu wokalicznego o czasie trwania 15 ms mieści się w granicach 10-25 ms. Łączny czas trwania głoski /r/ wraz elementami wokalicznymi wynosił 70-85 ms. Zaprojektowaną w ten sposób głoskę /r/ łączono z samogłoskami tworząc diady typu /ra, re, ro, ri/. Tranzjent od /r/ do samogłoski trwał na ogół ok. 40 ms (dla /i,o/), 60 - 80 ms (dla /a,e/).

Głoskę /r/ w śródgłosie modelowano w połączeniach typu /ara, era, ura, ore, ira, ari, iro, ori/. Wybrano dla niej model zawierający element wokaliczny o długości 10 ms, przejście do konfiguracji /r/ trwające w granicach 10-15 ms, potem część ustaloną /r/ ok. 10 -15 ms i przejście do końcowego elementu wokalicznego o długości 15 ms, które trwało ok. 10-25 ms. Iloczas tranzjentu od samogłoski do /r/ (czyli do niemal pełnego zamknięcia toru głosowego) może zawierać się w granicach 30-60 ms, natomiast od /r/ do samogłoski końcowej ok. 40 ms.

W podsumowaniu można stwierdzić, że przy modelowaniu sylab tworzonych z głosek należących do klasy rezonantów czasy trwania spółgłosek o charakterze ustalonym, a także o typowo tranzjentowej naturze mają decydujący wpływ nie tylko na jakość brzmienia sylaby, lecz często na jej identyfikację. Czas przejścia między głoskami o ustalonym przebiegu jest na ogół krótki, ale może być niekiedy zmieniany w dość dużych granicach, zależnie od pozycji artykulacji. W wygłosie czas przejścia do spółgłoski jest na ogół dłuższy, niż w przypadku nagłosu. Artykulacja spółgłosek rezonantowych w różnym stopniu zależy od kontekstu samogłoskowego. Najmniej zależna od niego jest artykulacja dwuwargowej spółgłoski nosowej /m/, najnocniej zaś spółgłoski drżącej /r/. Przeprowadzone badania umożliwiły doświadczalną weryfikację opracowanego modelu toru artykulacyjnego, na przykładzie nie tylko przebiegów stacjonarnych takich jak samogłoski, czy spółgłoski nosowe, ale również przebiegów o szybko zmieniającej się konfiguracji toru podczas artykulacji takich głosek, jak /j/ czy /w/.

## 9. Zakończenie

Przedstawione w pracy akustyczno-artykulacyjne podstawy generacji sygnału mowy stanowiły bazę do opracowania metody artykulacyjnego modelowania procesu generacji sygnału mowy polskiej o pobudzeniu krtniowym. Na przykładzie trójwymiarowego modelowania konfiguracji geometrycznej toru głosowego, które zrealizowano w oparciu o unikalne dla mowy polskiej pomiary wykonane na podstawie zdjęć kinorentgenograficznych (BOLLA&FÖLDI,1987), dokonano próby zastosowania oryginalnej aproksymacji konfiguracji toru dla poszczególnych polskich rezonantów stosując w określonych jego odcinkach przekroje o kształtach elipsoidalnych, kołowym, bądź trapezowym. Dzięki temu sterowanie zmianą w czasie przekroju podłużnego toru głosowego pociąga za sobą odpowiednią zmianę funkcji przekroju poprzecznego i tym samym odpowiednie ukształtowanie funkcji transmitancji toru głosowego dla zadanego momentu artykulacji. Istotnym elementem opracowanej metody modelowania przebiegów o tranzjentowym charakterze jest to, że dzięki zastosowaniu aproksymacji nieregularnego kształtu toru głosowego do szeregowo połączonych ze sobą odcinków o przekroju kołowym, eliptycznym i trapezowym stało się możliwe zautomatyzowane programowanie większości przejść typu CV i VC przez zadanie początkowej i końcowej konfiguracji toru, bez konieczności określania szczegółowego, nieliniowego w czasie, przebiegu zmian funkcji przekroju. Takie uproszczenie było niemożliwe w przypadku modelowania toru głosowego do postaci rury o stałym kształcie przekroju, kołowym, czy eliptycznym, ponieważ liniowa zmiana w czasie jednego wymiaru, to jest wysokości przekroju nie oddaje właściwego przebiegu częstotliwości formantowych.

Opracowana komputerowa metoda modelowania toru głosowego może być bardzo efektywnym narzędziem badawczym do analizy własności akustycznych dźwięków mowy o pobudzeniu dźwięcznym. W rozdziale 5 pracy, przedstawiono, jak można wykorzystać badania modelowe do bardziej szczegółowej analizy związków istniejących między częstotliwościami formantowymi i wybranymi elementami konfiguracji samogłoskowych (por. Rys. 5.15, 5.16 i 5.17). Główny jednak nacisk położono na określeniu relacji zachodzących przy artykulacji ich w połączeniach z innymi głoskami, a zwłaszcza modelowania takich zjawisk zachodzących w mowie naturalnej jak koartykulacja, czy nazalizacja. Na przykładzie połączeń polskich rezonantów spółgłoskowych z samogłoskami wykazano przydatność modelowania artykulacyjnego do badania w płaszczyźnie akustycznej przebiegów tranzjentowych (Rozdz. 8.5).

Przeprowadzone badania były pierwszymi zrealizowanymi dla języka polskiego, w których podjęto próbę wykorzystania do modelowania unikalnych pomiarów przekrojów toru głosowego wykonanych ze zdjęć kinorentgenograficznych.

Choć dane te były ograniczone tylko przekrojów wzdłużnych toru głosowego tylko dla otworu wylotowego utworzonego przez wargi autorzy cytowanej pracy podają również jego wymiary poprzeczne), stanowiły one punkt wyjścia do dobierania kształtu i wymiarów poprzecznych toru głosowego w taki sposób, by kształt podużny toru głosowego dla określonej głoski był zbliżony do wyznaczonego ze zdjęć oaz by jednocześnie częstotliwości formantowe wyznaczone z modelu były zgodne z pomierzonymi dla tej głoski wypowiedzianej przez lektora w kontekście wyrazowym.

Zastosowany w badaniach oryginalny program *Model* (Rozdz. 8.3) umożliwia modelowanie toru głosowego dla przebiegów o charakterze nie tylko stacjonarym, jak na przykład dla samogłosek, lecz również dla przebiegów silnie tranzjentowych, jak na przykład spółgłosek płynnych. Symulowanie ciągłości zmian w konfiguracji toru głosowego podczas przejścia od jednej konfiguracji do drugiej realizowano ze stałym krokiem czasowym, przy założeniu, że zmiany przekroju podłużnego toru głosowego dokonują się w sposób liniowy w funkcji czasu.

Za pomocą programu *Model* przebadano również modelowanie toru głosowego za pomocą rury o stałym kształcie cylindrycznym i eliptycznym. Wykazano we wcześniejszych publikacjach ([NOWAKOWSKA, i in., 1989]), że przy modelowaniu głosek stacjonarnych (samogłoski, spółgłoski nosowe) uzyskuje się dobre wyniki modelowania (zarówno w płaszczyźnie formantowej, jak i czasowej).

Założenie liniowości zmian w czasie konfiguracji geometrycznej toru głosowego jest dość dużym przybliżeniem, zwłaszcza przy przejściach między konfiguracjami znacznie różniącymi się między sobą. Szczególnie zawodne jest w przypadku, gdy tor głosowy jest odwzorowywany w postaci rury cylindrycznej o zmiennej średnicy. Znacznie lepsze wyniki modelowania przebiegów tranzjentowych uzyskano, gdy w torze głosowym wyróżniono trzy odcinki o odmiennym kształcie. Część gardłową modelowano za pomocą odcinka cylindrycznego, część toru znajdującą się pod podniebieniem miękkim za pomocą rury o przekroju eliptycznym, zaś kształt przekroju przedniego odcinka toru głosowego ograniczonego od góry przez podniebienie twarde aproksymowano jako trapezowy. Wówczas liniowe zmiany wymiarów przekroju podłużnego powodują bardziej adekwatne zmiany funkcji przekroju poprzecznego i tym samym generują bardziej zbliżone do rzeczywistych przebiegi zmian częstotliwości formantowych. Zgodność przebiegów formantowych w wyliczonych przebiegach widmowych modelowanych połączeń rezonansów z analogicznymi pomierzonymi dla mowy naturalnej były podstawą do obiektywnej oceny jakości modelowania.

Przy modelowaniu wyników modelowania czasowych przebiegów zmian funkcji przekroju toru głosowego stwierdzono, że dla stosunkowo krótkich tranzjentów ten sposób generacji przekrojów pośrednich między konfiguracją



początkową i końcową dla danego tranzjentu jest wystarczający. Jednakże przy dłuższych tranzjentach (rzędu 80-150 ms) konieczne jest niekiedy określenie dodatkowej, pośredniej konfiguracji docelowej, tak by móc dobrze odwzorować dynamiczne zmiany zachodzące podczas rzeczywistej artykulacji modelowanego przebiegu (tak jak to miało miejsce, na przykład, w przypadku spółgłoski płynnej /l/ wymówionej w nagłosie (patrz Rys. 8.18), dla której wyodrębniono dodatkową konfigurację docelową. W ten sposób wyodrębniono zmiany konfiguracji toru głosowego zachodzące podczas artykulacji samej spółgłoski o niestacjonarnym charakterze.

Wykonane badania modelowe na syntetycznych głoskach oraz diadach i triadach typowych dla języka polskiego potwierdzają, a niekiedy i uzupełniają wiedzę dotyczącą fonetyki akustycznej rezonantów. W Rozdz. 8.4 i 8.5 przedstawiono sposób modelowania artykulacyjnego polskich rezonantów oraz ich połączeń typu VC, CV oraz VCV wypowiedzianych w wyrazach, ukształtowanie toru artykulacyjnego w czasie wymawiania ich faz ustalonych, charakterystykę częstotliwościową odpowiadającą danej konfiguracji oraz wygenerowano sygnały akustyczne im odpowiadające. Ostateczna weryfikacja poprawności modelowania artykulacji polegała na ocenie przede wszystkim dokładności odwzorowania przebiegów formantowych, a także zrozumiałości sygnałów syntetycznych.

Modelowanie płynnych zmian geometrii toru głosowego zachodzących podczas artykulacji dźwięków mowy o niestacjonarnym charakterze stwarza możliwość badania mechanizmu koartykulacji oraz dokładnych reguł nim rządzących. Zaprezentowano to w Rozdz. 8.5, zwłaszcza na przykładzie głoski drżącej /r/ i jej różnych połączeń. Modelowanie jej jest bardzo trudne do ze względu na jej polisegmentalny charakter, zmieniający się zależnie od kontekstu, w jakim jest ona wypowiedziana. Zasadniczym segmentem tej głoski jest krótki segment uderzeniowy, podczas którego następuje przymknięcie toru głosowego. Jednakże, by efekt ten był słyszalny segment ten musi wystąpić w otoczeniu samogłoskowym lub samogłoskopodobnym. Z tego powodu przy modelowaniu należy uwzględnić położenie /r/ w wyrazie (frazie). Uzyskane reguły modelowania /r/ poszerzają wiedzę dotyczącą artykulacji tej głoski, jej iloczasu i roli w sylabie z uwzględnieniem położenia w śródgłosie, nagłosie i wygłosie. Potwierdzono postulowane w literaturze występowanie segmentu wokalicznego w sąsiedztwie właściwej konfiguracji /r/ ([DŁUSKA, 1981; DUKIEWICZ, 1995]), gdy jest ono wymawiane w nagłosie lub wygłosie. Ustalono czas trwania segmentu wokalicznego oraz jego konfigurację artykulacyjną bliską konfiguracji tzw. neutralnego „e”.

W modelowaniu połączeń rezonantów, w sposób niemal naturalny uwzględnione jest zjawisko koartykulacji, które szczególnie jest widoczne w

połączeniach typu: samogłoska - spółgłoska nosowa, samogłoska-spółgłoska nosowa-samogłoska, dla których zakres wpływu nazalizacji na samogłoski sąsiednie można stosunkowo łatwo określić. Występowanie tego zjawiska przeanalizowano w mowie naturalnej i reguły dotyczące wpływu nazalizacji (czasu trwania segmentu nazalizowanego samogłoski) potwierdziły zasadność uwzględnienia przy modelowaniu częściowej nazalizacji segmentu samogłoskowego, gdy jest wymawiany w sąsiedztwie spółgłoski nosowej.

Reguły przejść tranzjentowych badane we wszystkich prezentowanych połączeniach rezonantów okazują się być dość rygorystyczne. Zmiany przejść czasowych w większych granicach, niż opisane w poprzednim rozdziale, w sposób zasadniczy wpływają na percepcję generowanych sylab, obniżając ich naturalność brzmienia, a przede wszystkim ich zrozumiałość. Przy modelowaniu diady (bądź triady) ustalenie właściwego czasu trwania przejścia od jednej głoski do drugiej jest najważniejszym elementem w tworzeniu danego połączenia głosek i gwarantuje poprawnie identyfikowany percepcyjnie sygnał syntetyczny. Pozostałe elementy takie jak: początkowe narastanie i opadanie amplitudy sygnału, niewielkie zmiany częstotliwości podstawowej oraz sterowanie zmianą amplitudy w obrębie głosek, czy tranzjentu, mogą wpłynąć w sposób zauważalny na naturalność brzmienia sygnału, nie pogarszając jednak w sposób istotny jego zrozumiałości.

W literaturze istnieją dla języka polskiego pewne wyrwykowe wyniki eksperymentów dotyczących modelowania samogłosek i ich połączeń z niektórymi spółgłoskami, które wykonano za pomocą syntezy formantowego SYNFOR I i II [KACPROWSKI&MIKIEL, 1965a;1965b; KUDELA,1970]. Jednakże trzeba zaznaczyć, że syntezy formantowy modeluje w istocie widmowe przebiegi sygnałów naturalnych za pomocą kaskadowo lub równoległe połączonego ze sobą zestawu filtrów górnoprzepustowych o odpowiednio dobranych częstotliwościach rezonansowych zmieniających się w czasie. Tak więc jest to raczej odtwarzanie sygnału mowy w płaszczyźnie widmowej, a nie klasyczne modelowanie toru głosowego jako złożonego układu akustycznego.

Na zakończenie, warto podkreślić, że wprowadzenie ostatnio nowych technik wizualizacji konfiguracji toru głosowego, zwłaszcza MRI, mimo wszelkich swych ograniczeń umożliwia już trójwymiarową wizualizację jego geometrii. Już te dane dostępne w literaturze (por. [BAER i in.,1991; STORY i in.,1996]) wskazują, że przekroje poprzeczne toru głosowego są niekiedy swym kształtem dość odległe od geometrycznych figur regularnych. Jednakże brak jakichkolwiek danych o trójwymiarowej konfiguracji geometrycznej toru głosowego dla głosek języka polskiego, spowodował że nadal podstawowym źródłem o konfiguracji toru głosowego jest, pomimo swoich niedociągnięć (np. wrażliwa niekiedy transkrypcja

fonetyczna dźwięków mowy), wspomniana monografia [BOLLA&FÖLDI,1989], w której podano przekroje podłużne toru głosowego dla większości dźwięków mowy polskiej, w różnych fazach ich artykulacji. Dane te były uzupełnione ponadto nagraniami video oraz zapisami dźwiękowymi udostępnionymi autorowi niniejszej pracy.

Należy oczekiwać, że zagadnienie modelowania narządu artykulacyjnego jako złożonego, trójwymiarowego układu akustycznego, będzie dla mowy polskiej kontynuowane, już w oparciu o zdjęcia wykonywane techniką MRI. W wielu przypadkach, modelowanie układu artykulacyjnego jest niekiedy jedyną drogą do pełnego wyjaśnienia mechanizmu niektórych zjawisk występujących przy generacji sygnału mowy, czy działania niektórych elementów np. źródła krtaniowego. Szczególnie trudnym do określenia są związki zachodzące między zmianami w sygnale akustycznym i zmianami patologicznymi w obrębie narządów artykulacyjnych (np. częściowe porażenia fałdów głosowych, guzki, rozszczep podniebienia itp.). Wydaje się, że w niektórych przypadkach, jedynie poprzez modelowanie będzie można w sposób jednoznaczny ustalić zależności między dysfunkcją narządu artykulacyjnego i fizyczną postacią akustycznego sygnału mowy. Ale należy podkreślić, że dla wielu zmian chorobowych jest to bardzo złożone zagadnienie, i zapewne z tego powodu, tak bardzo mało jest prac podejmujących tego typu badania.

### Podziękowanie

Prof. Kalman Bolla i dr Eva Földi z Instytutu Lingwistyki Węgierskiej Akademii Nauk, w Budapeszcie za udostępnienie nagrań dźwiękowych i video, zarejestrowanych podczas wykonywania zdjęć kinorentgenograficznych oraz za udostępnienie dodatkowych materiałów pomiarowych. Wandzie Nowakowskiej za wieloletnią współpracę w badaniach nad modelowaniem artykulacyjnym głosek języka polskiego oraz Piotrowi Żarnekiemu za współpracę w realizacji i rozwijaniu komputerowego modelu toru głosowego.

## LITERATURA

- Alvan A., Narayanan S., Haker K. (1997), Toward articulatory-acoustic models for liquid approximations based on MRI and EPG data. Part II. The rhotics. , *J. Acoustic. Soc. Am.* **101**, 1078-1089
- Atal B.S., Chang J.J., Mathews M.V., Tukey J.W. (1978), Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique, *J. Acoustic. Soc. Am.* **63**,5, 1535-1555.
- Atal B.S., Rioul O. (1989), Neural networks for estimating articulatory positions from speech, *J. Acoustic. Soc. Am.* **86**,123-131.
- Badin P., Fant G. (1984), Notes on vocal tract computation, *STL-QPSR* 2-3/1984, 53-108.
- Badin P., Bailly G., Rayaudi M., Segebarth C. (1998), A three-dimensional linear articulatory model based on MRI data, *Proc. 3<sup>rd</sup> ESCA Int. Workshop on Speech Synthesis*, 249-254.
- Baer T., Gore J.C., Gracco L. C., Nye P.V., (1991) Analysis of vocal tract shape and dimension using magnetic resonance imaging: vowels, *J. Acoust. Soc. Am.*, **90**, 2, (Pt 1), 799-828.
- Benade A.H., (1968) On the propagation of sound waves in a cylindrical conduit, *J. Acoust. Soc. Am.* **44**, 616-623.
- Beautemps D., Badin P., Laboissiere R. (1995), Deriving vocal tract area functions from midsagittal profiles and formant frequencies: A new model for vowels and fricatives consonants based on experimental data, *Speech Communication*, **16**, 1, 27-47.
- Berg Van Den J.W., Zantema J.T., Doornenbal P. jr, (1957), On the air resistance and the Bernoulli effect of the human larynx, *J. Acoust. Soc. Am.* **29**, 626-631.
- Berg Van Den J.W., (1958) Myoelastic-aerodynamic theory of voice production, *J. of Speech and Hearing Res.*, 1,227-244.
- Biedrzycki L. (1978), *Fonologia angielskich i polskich rezonantów*, PWN, Warszawa.
- Bolla K., Földi E. (1987), *A phonetic conspectus of Polish. The articulatory and acoustic features of Polish speech sounds.*, Hungarian Papers in Phonetics 18, Budapest.
- Borden G.J., Harris K.S. (1984), *Speech Science Primer: Physiology, Acoustics and Perception of Speech*. 2<sup>nd</sup> ed., Williams&Wilkins, Baltimore.
- Broad D. J. (1973), *Phonation w: Normal aspects of speech, hearing and language*, F.D. Minifie, T.J. Hixon, F. Williams, Prentice-Hall, Englewood Cliffs, 127-167.
- Butcher A., Weiher E. (1976), An electropalatographic investigation of coarticulation in VCV sequences, *J. of Phonetics*, **3**, 39-47.
- Californski T., Jassem W., Kaczmarek Z. (1970), Investigation of vowel formant frequencies as personal voice characteristics by means of multivariate analysis of variance, *Speech Analysis and Synthesis*, vol. 2, (ed. W. Jassem), Warsaw, 7-26.

- Carré R., Mrayati M. (1992), Distinctive regions in acoustic tubes speech production modelling, *J. Acoustique*, **5**, 141-159.
- Chan D., Fourcin A., Gybbon D., Grandstrom B., Huckvale M., Kokkinakis G., Kvale K., Lamel L., Lindberg B., Moreno A., Mouropoulos J., Senia F., Transcoso L., Velt C. 't, Zeilinger J. (1995) EUROM - A spoken Language Resource for the EU, *Proc. EUROSPEECH'95*, Madrid, vol. 1, 867-870.
- Childers D.G., Hicks D.M., Moore G.P., Alsaka Y.A. (1986), A model for vocal vibration motion, contact area, and the glotogram, *J. Acoust. Soc. Am.* **80**(5), 1309-1320.
- Ciocea S., Schoentgen J. (1998), Semi-analytic formant-to-area mapping, *Etudes&Travaux*, n° 2, ULB, Bruxelles.
- Cole R., Mariani J., Uszkoreit H., Zaenen A., Zue V. (1996), *Survey of the state of the art in human language technology*, Center for Spoken Language Understanding, Oregon Graduate Institute, <http://www.cse.ogi.edu/CSLU/HLTSurvey/>
- Czistowicz L.A., Kożewnikow V.A. i in. (1965), *Riecz: Artikulacja i wospriatje.*, Nauka, Moskwa-Leningrad.
- Dang J., Honda K., Suzuki H. (1994), Morphological and acoustical analysis of the nasal and the paranasal cavities, *J. Acoust. Soc. Am.*, **96**, 2088-2100.
- Dang J., Honda K. (1996), Acoustical modeling of the vocal tract based on morphological reality: Incorporation of the paranasal sinuses and the piriform fossa, *Proc. 1<sup>st</sup> ESCA Tutorial and Research Workshop on Speech Production Modeling: From Control Strategies to Acoustics*, Autrans, 49-52.
- Demolin D., Hombert J.-M., Lecuit V., Segebarth Ch., Soquet A. (1995), An MRI study of French vowels, *Proc. EUROSPEECH'95*, Madrid, 2235-2238.
- Dłuska M. (1981), *Fonetyka polska. Artykulacje głosek polskich*, PWN, Warszawa-Kraków (wznowienie z 1947 r).
- Dukiewicz L. (1967), *Polskie głoski nosowe*, PWN, Warszawa.
- Dukiewicz L. (1995), *Fonetyka [w:] Fonetyka i Fonologia* (L. Dukiewicz i I. Sawicka), Wyd. Inst. Jęz. Polskiego PAN, Kraków.
- DUNN H.K. (1950), The calculation of vowel resonances and an electrical vocal tract, *J. Acoust. Soc. Am.*, **22**, 740-755.
- Engwall O. (1999), Vocal tract modelling in 3D, *STL-QPSR 1-2/1999*, 31-38.
- von Essen O. (1962), *Allgemeine und angewandte Phonetik*, 3-e Auf., Akademie-Verlag, Berlin, (*Fonetyka ogólna i stosowana*, PWN, Warszawa, 1967)
- Fant G. (1960), *Acoustic theory of speech productions*, Mouton and Co.,s' Gravenhage.
- Fant G. (1982), Preliminaries to analysis of the human voice source, *STL-QPSR 4/1982*, 1-27.
- Fant G. (1995), The LF-model revisited . Transformations and frequency domain analysis, *STL-QPSR 2-3/1995*, 119-156.
- Farnsworth D.W. (1940), High-speed motion pictures of the human vocal cords, *Bell Lab. Rec.*, **18**, 203-208.

- Schroeder M. (1967), Determination of the geometry of the human vocal tract by acoustic measurements, *J. Acoust. Soc. Am.*, **41**, 4 (Part 2), 1002–1010.
- Schroeter J., Sondhi M. (1994), Techniques for estimating vocal-tract shapes from the speech signal, *IEEE Trans. On Speech and Audio Processing*, **2**, 1, 133–150.
- Shirai K., Kobayashi T. (1986), Estimation and generation of the articulatory motion from speech wave, *Speech Comm.*, **5**, 159–170.
- Shirai K. (1993), Estimation and generation of the articulatory motion using neural networks, *Speech Comm.*, **13**, 1, 45–51.
- Stevens K.N. (1971), Airflow and Turbulence Noise for Fricative and Stop Consonants: Static Considerations, *J. Acoust. Soc. of Am.*, **50**, nr 4 (part 2), 1180–1192.
- Stevens K.N. (1993), Modelling affricate consonants, *Speech Comm.*, **13**, 1, 35–43.
- Stevens K.N. (1998), *Acoustic Phonetics*, MIT Press, Cambridge.
- Story B., Titze I., Hoffman E. (1996), Vocal tract area functions from magnetic resonance imaging, *J. Acoust. Soc. of Am.*, **100**, 537–554.
- Sundberg J., Johansson C., Wilbrand H., Ytterbergh C. (1987), From sagittal distance to area. A study of transverse, vocal tract cross-sectional area, *Phonetica*, **44**, 1, 76–90.
- Titze I.R. (1988), The physics of small-amplitude oscillation of the vocal folds, *J. Acoust. Soc. Am.*, **83**, 4, 1536–1552.
- Titze I.R. (1994), *Principle of Voice Production*, Prentice Hall, Englewood Cliffs.
- Titze I.R., Strong W. (1975), Normal modes in vocal cord tissues, *J. Acoust. Soc. Am.*, **57**, 736–744.
- Vygotski L.S. (1962) *Thought and Language*, M.I.T. Press, Cambridge.
- Wakita H. (1973), Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveform, *IEEE Trans. Audio Electroacoust.* **AU-21**, 417–427.
- Wakita H., Fant G. (1978), Toward a better vocal tract model, *STL-QPSR* 11978, 9–29.
- Wierzchowska B. (1980), *Fonetyka i fonologia języka polskiego*, Ossolineum, Warszawa-Wrocław.
- Wrzosek A. (1996), Analiza efektywności niejawnych modeli Makowa w rzeczywistych systemach automatycznego rozpoznawania mowy polskiej, Rozpr. dokt., IPPT PAN.
- Young S., Bloothoof G. eds., (1997) *Corpus-based methods in language and speech processing*, Kluwer Ac Pub. Dordrecht.



56520