

TEORIA WYTWARZANIA DŹWIĘKÓW MOWY, FORMANTY, MODELOWANIE WYTWARZANIA DŹWIĘKÓW MOWY.

1

dr inż. Kuba Łopatka / dr hab. inż. Józef Kotus
Katedra Systemów Multimedialnych

PLAN WYKŁADU

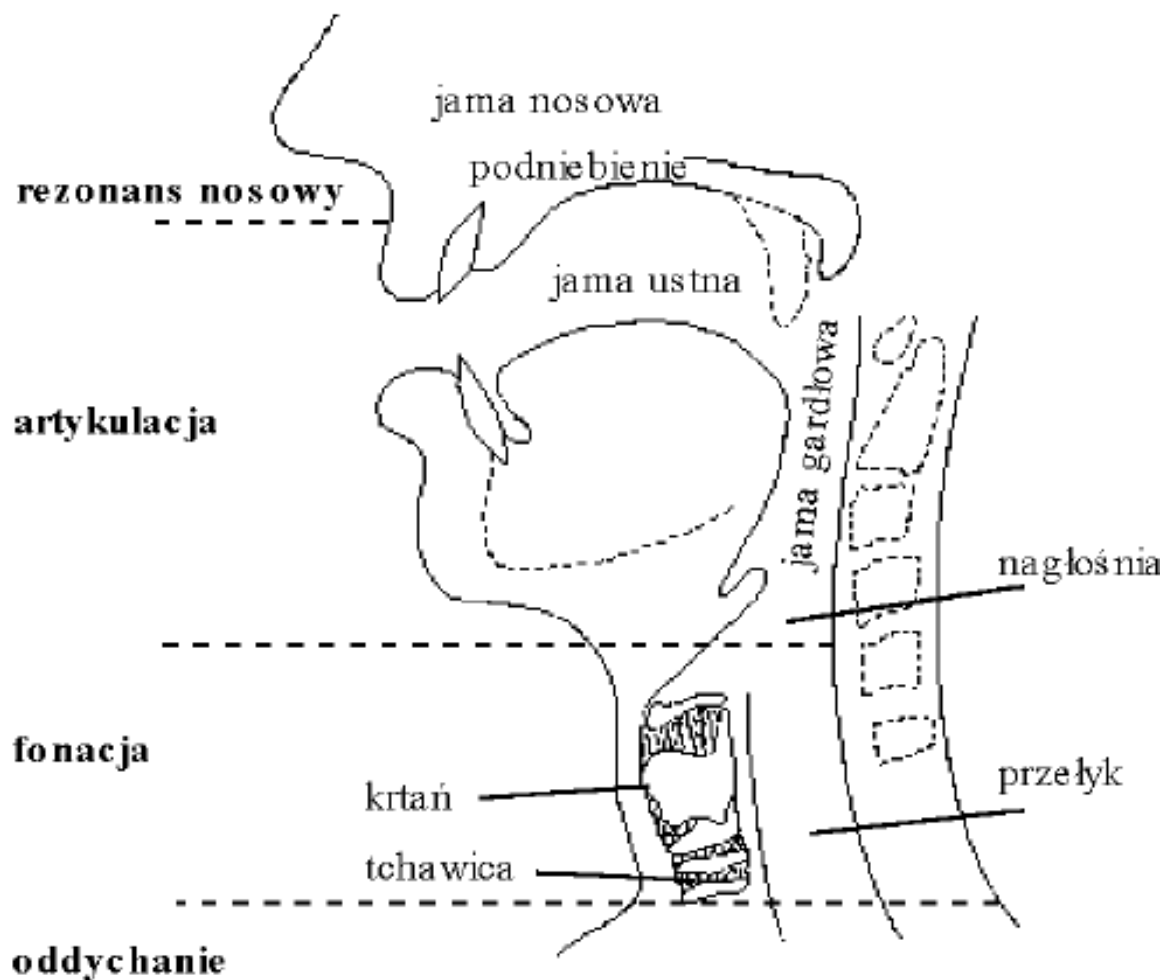
1. **Teoria wytwarzania dźwięków mowy**
 - Ogólna teoria wytwarzania dźwięków mowy
 - Ton krtaniowy
 - Pobudzenie szumowe
 - Transmitancja traktu głosowego
2. **Formanty**
 - Pojęcie formantu
 - Parametry formantowe
 - Formanty F1 i F2, trójkąt samogłosek
3. **Modelowanie wytwarzania dźwięków mowy**
 - Modelowanie fizyczne
 - Modelowanie elektryczne i komputerowe
 - Model źródło-filtr

The left side of the slide features a series of vertical stripes in various shades of green and grey. Overlaid on these stripes are several green circles of different sizes, arranged in a vertical line. The largest circle is at the top, with smaller ones below it, creating a decorative border.

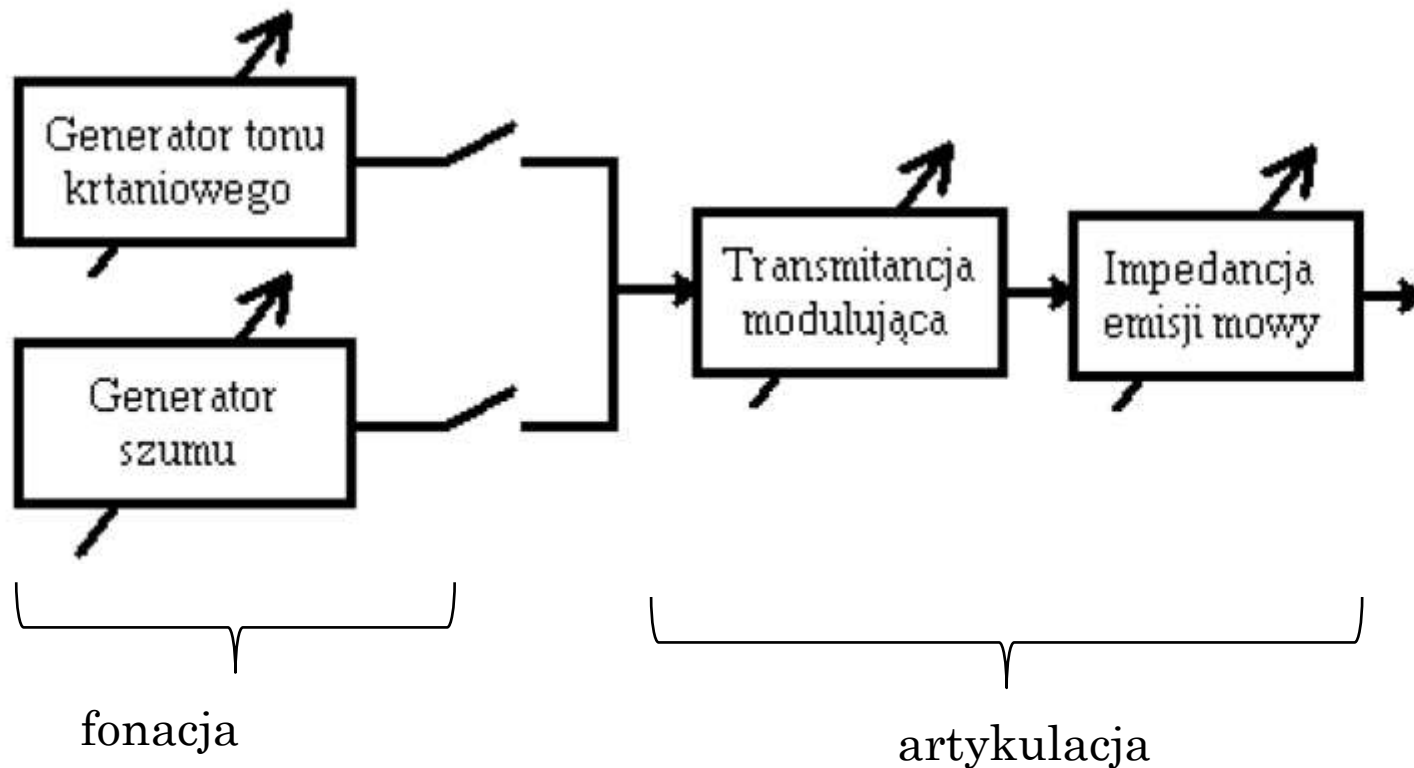
TEORIA WYTWARZANIA DŹWIĘKÓW MOWY

3

PRZYPOMNIENIE – ANATOMIA



OGÓLNY SCHEMAT ZASTĘPCZY WYTWARZANIA MOWY



Podstawowym założeniem tej teorii jest **niezależność** charakterystyk tonu krtaniowego i artykulatorów.

TON KRTANIOWY

- Jest to sygnał akustyczny wytworzony przez drgające struny głosowe.
- Stanowi pobudzenie dla głosek dźwięcznych i zwartych.
- Powstaje w procesie **fonacji**.
- Trudno go zbadać, bo zawsze jest w postaci zmodulowanej.
- Sam w sobie jest słaby i bezbarwny, ulega wzmocnieniu i nabiera brzmienia dopiero w wyższych partiach traktu głosowego.
- Podstawowy parametr – częstotliwość podstawowa – f_0 (lub F_0)- zgodna z częstotliwością podstawową mowy.

TON KRTANIOWY

Zakres częstotliwości podstawowej tonu krtaniowego:
w mowie (typowe):

- mężczyźni: 85-180Hz
- kobiety: 165-255Hz
- dzieci: 250-300Hz

W śpiewie w zależności od rodzaju głosu:

- bas, baryton: 65Hz-349Hz
- tenor: 130-523Hz
- alt: 195-698Hz
- mezzosopran, sopran: 220-1280Hz

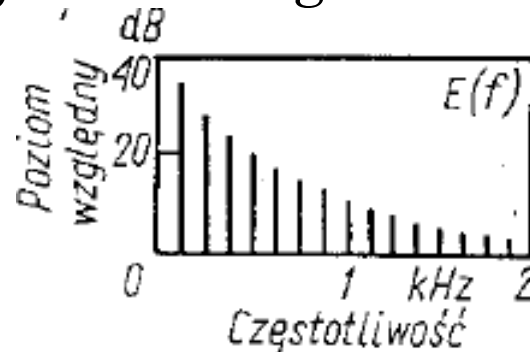
TON KRTANIOWY

Postać czasowa:



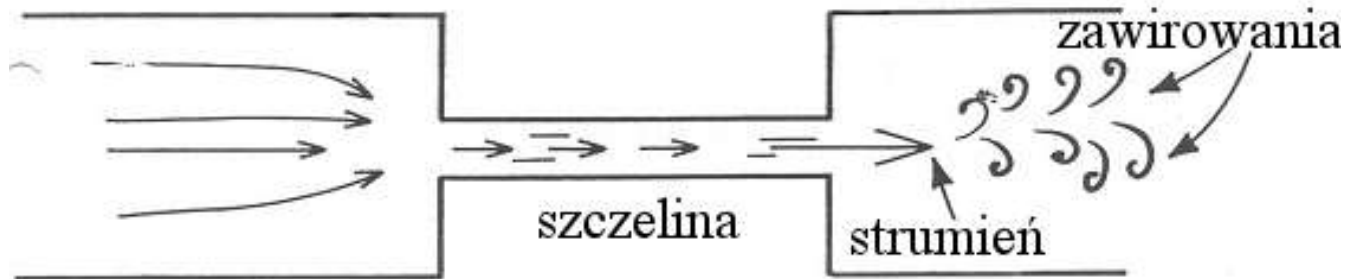
Widmo:

- Obecne są wszystkie harmoniczne częstotliwości podstawowej opadające ok. 6-12dB/okt. (w praktyce nie jest tak regularnie)



POBUDZENIE SZUMOWE

Przy artykulacji głosek szumowych strumień wydychanego powietrza przechodzi z przepływu **laminarnego** do przepływu **turbulentnego** (zjawisko znane w hydrodynamice).

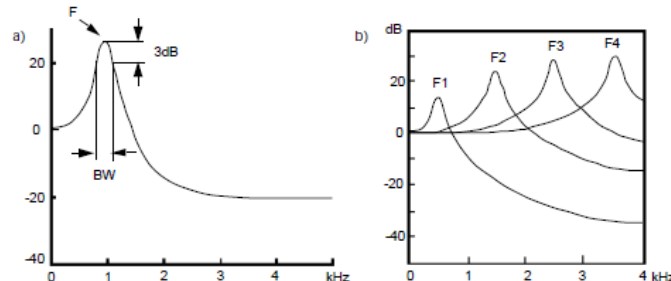


Przy przejściu powietrza przez szczelinę dokonuje się konwersja energii kinetycznej na akustyczną.

Przy nagłym otworzeniu drogi przepływu powietrza podczas artykulacji spółgłosek zwartych powietrze tworzy **falę udarową**, która jest nośnikiem energii akustycznej.

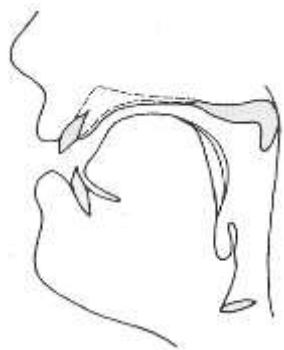
TRANSMITANCJA TRAKTU GŁOSOWEGO

- Z punktu widzenia akustycznego trakt głosowy to układ rezonatorów.
- Kształt i połączenia między tymi rezonatorami zmieniają się w zależności od ułożenia narządów artykulacyjnych – zuchwy, języka, warg, podniebienia.
- Charakterystyka częstotliwościowa traktu głosowego jest połączeniem krzywych rezonansowych.

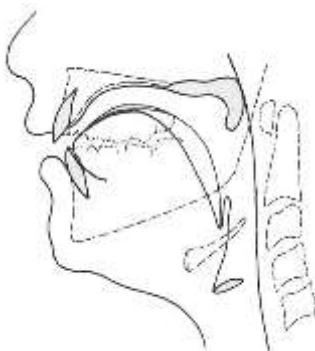


TRANSMITANCJA TRAKTU GŁOSOWEGO

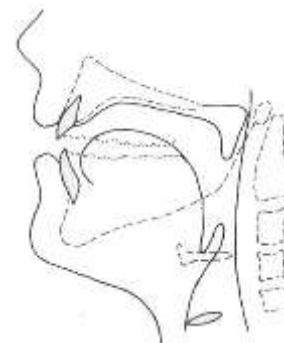
Różne konfiguracje narządów mowy:



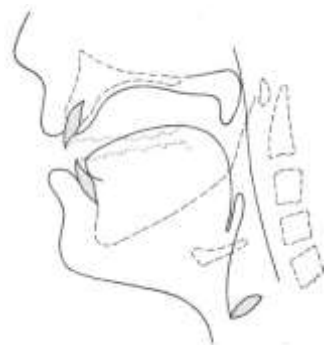
i



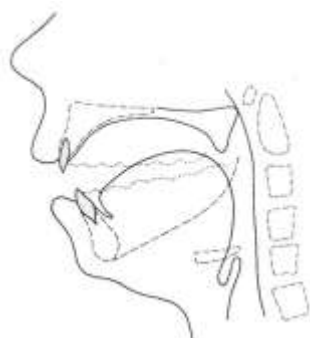
y



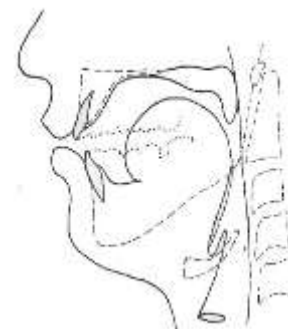
e



a



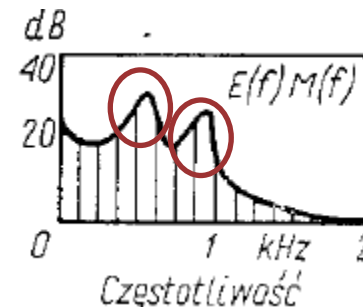
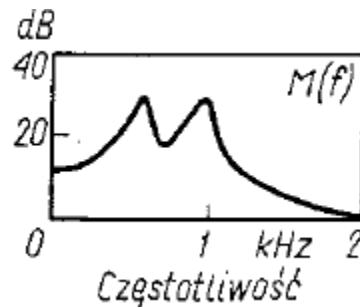
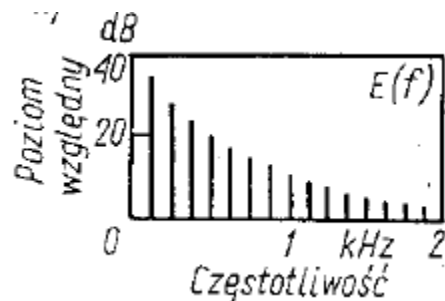
o



u

TRANSMITANCJA TRAKTU GŁOSOWEGO

- Rezonans akustyczny narządów mowy powoduje podbicie pewnych częstotliwości (pewnych obszarów widma).
- To, które obszary widma zostaną podbite, zależy od ułożenia narządów mowy – przez zmianę ułożenia narządów mowy zmieniają się wymiary rezonatorów.
- W efekcie w widmie sygnału mowy obserwuje się charakterystyczne skupiska energii.



Widmo tonu krtaniowego
sygnału mowy

charakterystyka traktu głosowego

widmo



FORMANTY

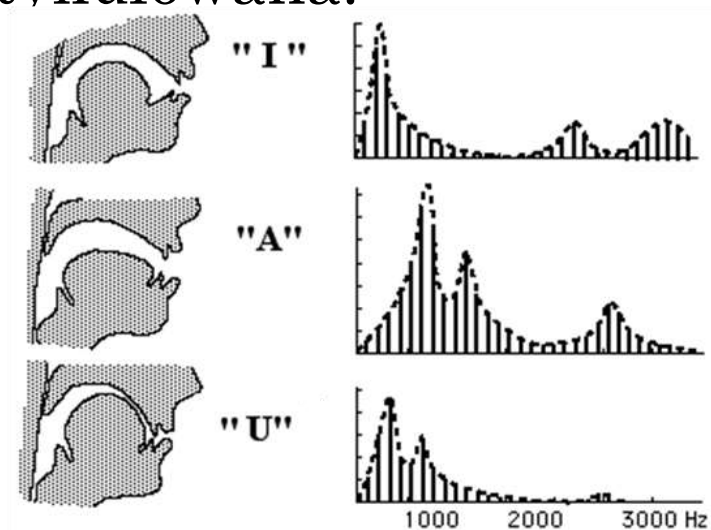


13



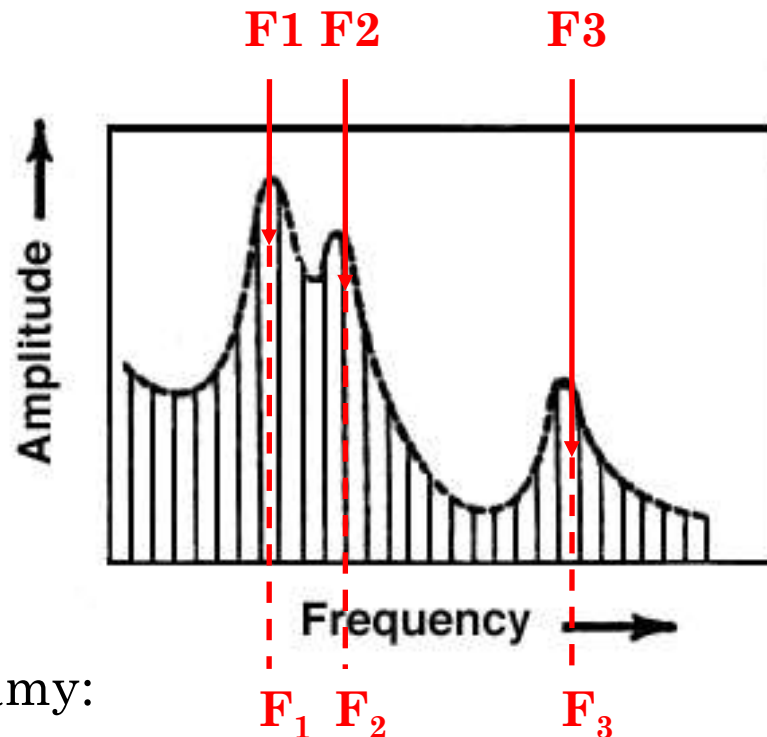
FORMANTY

- Skupisko energii w widmie sygnału nazywamy **formantem**.
- Formant to jedno z najważniejszych pojęć w akustyce mowy.
- Rozmieszczenie formantów zależy od tego, jaka głoska jest artykulowana.



FORMANTY

Formanty oznaczamy:



Częstotliwości
formantowe oznaczamy:

Czasem formantem F_0 nazywa się ton krtaniowy (o częstotliwości podstawowej F_0)

FORMANTY

- Ucho ludzkie jest czułe na formanty.
- Teoria głosi, że formanty leżą u podstaw rozpoznawania mowy przez człowieka. W związku z tym wykorzystuje się je przy automatycznym rozpoznawaniu mowy.

FORMANTY

Parametry formantowe to:

- Częstotliwość formantu F_N
- Poziom formantu A_N (wyrażany w dB, względny unormowany do najsilniejszego formantu albo bezwzględny)

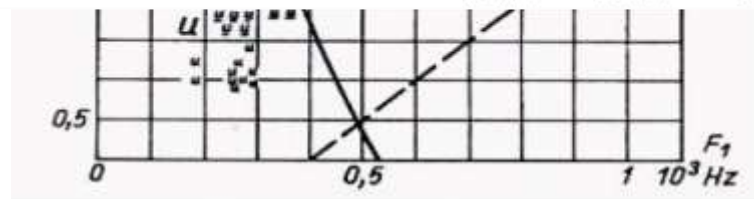
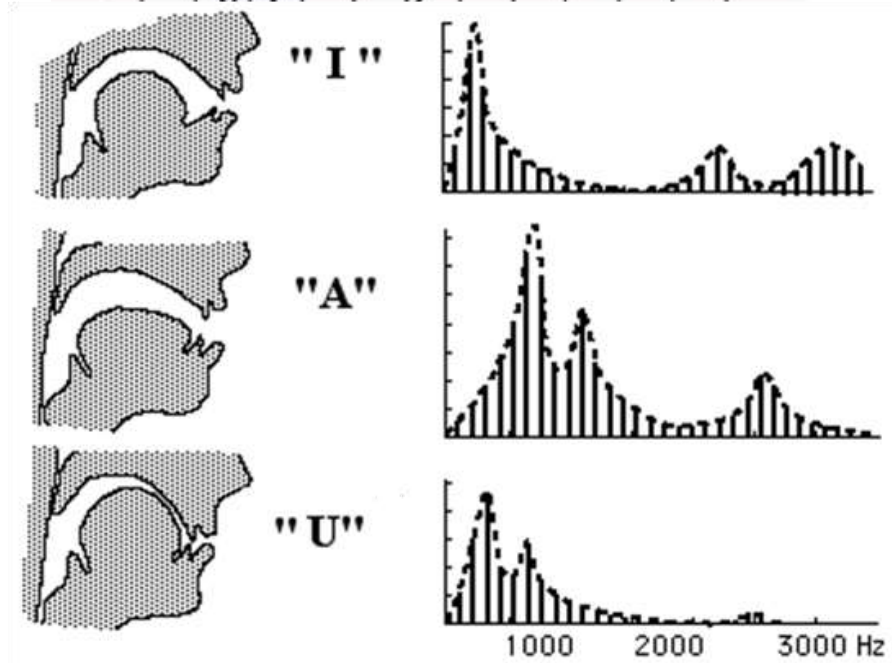
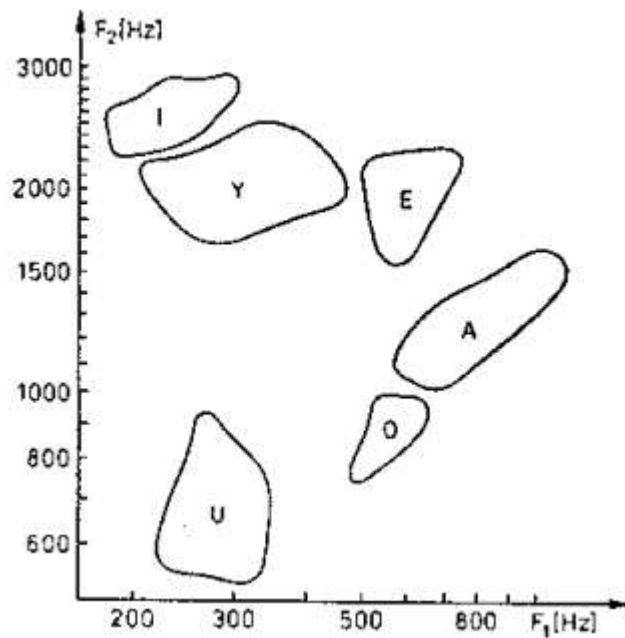
Fonem	częstotliwości [Hz]				poziomy względne [dB]			
i	210	2750	3500	4200	0	-15	-15	-27
e	380	2640	3000	3600	0	-12	-16	-20
a	780	1150	2700	3500	0	-7	-25	-25
y	240	1550	2400	3300	0	-12	-20	-30
o	400	730	2300	3200	0	-3	-30	-35
u	270	615	2200	3150	0	-13	-40	-50
w	600	1700	2900	4100	-9	0	-2	-10
sz	-	2300	2900	3600	-	-9	-8	0
h	500	1700	2500	4200	-12	0	-10	-17
z	-	1750	2950	4300	-	-6	-10	0

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

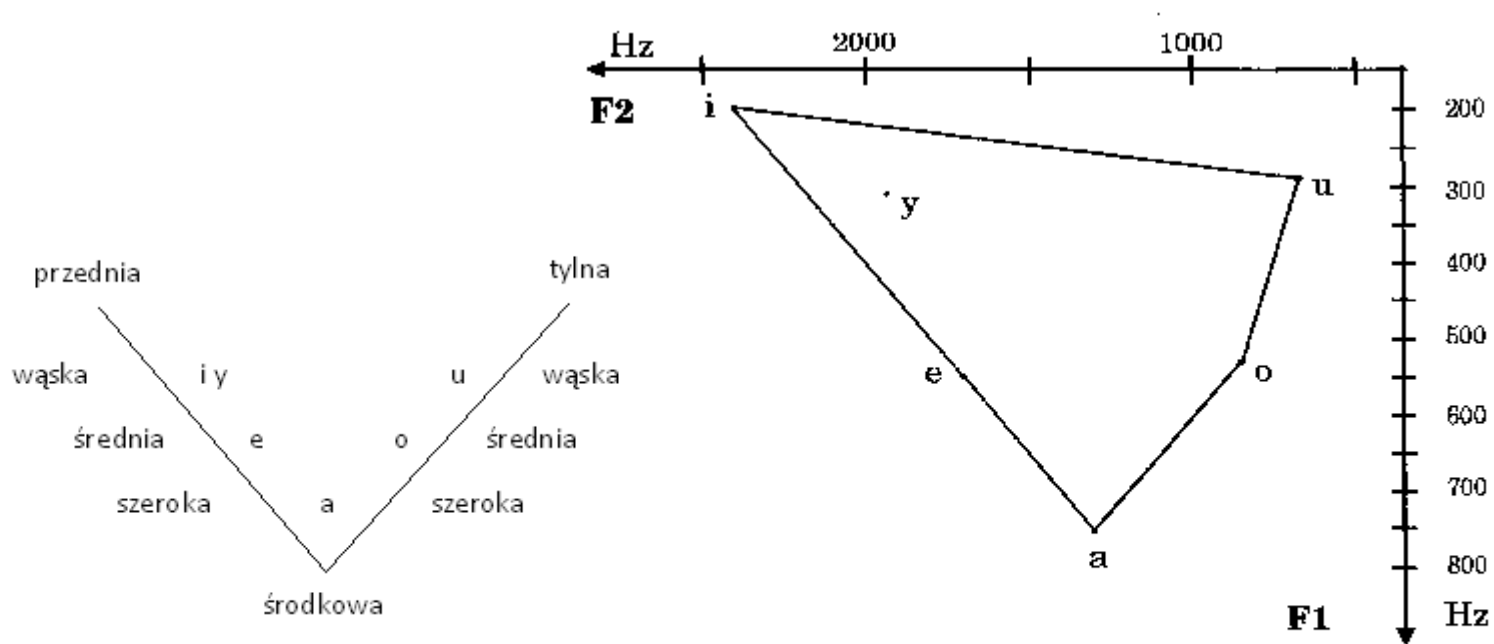
DWA PIERWSZE FORMANTY

- Dwa pierwsze formanty (F1 i F2) zależą od ułożenia języka (dzielącego jamę ustną na dwie wnęki).
- Niektóre spośród głosek (zwłaszcza samogłoski) można odróżnić na podstawie tylko tych dwóch formantów.
- W akustyce mowy rozpowszechniony jest wykres na płaszczyźnie częstotliwości F_1 i F_2 – tzw. „trójkąt samogłosek”.

TRÓJKĄT SAMOGŁOSEK



TRÓJKĄT SAMOGŁOSEK



The slide features a dark green background with a vertical decorative element on the left consisting of several thin, light-colored stripes. To the right of these stripes are several green circles of varying sizes, arranged in a descending pattern from top to bottom. The largest circle is at the top, followed by a smaller one, then a medium-sized one containing the number 22, and finally two more smaller circles at the bottom.

MODELOWANIE MECHANIZMÓW WYTWARZANIA MOWY

22

MODELOWANIE WYTWARZANIA MOWY

Modelowanie wytwarzania mowy wymaga 3 składników:

- Modelowanie pobudzenia
- Modelowanie traktu głosowego
- Modelowanie emisji mowy

W praktyce najważniejsze są dwa pierwsze elementy.

MODELOWANIE POBUDZENIA

- Głoski dźwięczne – modelowanie tonu krtaniowego

Ton krtaniowy modeluje się najczęściej przebiegiem piłokształtnym, który ma zbliżone właściwości widmowe.



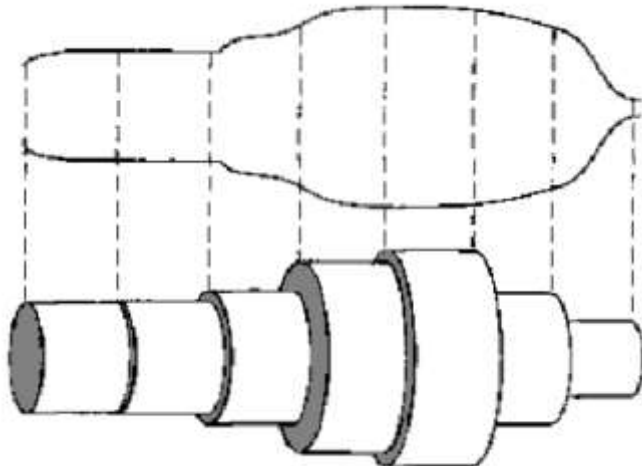
- Głoski bezdźwięczne – pobudzenie szumowe.

Modeluje się szumem o widmie opadającym 6dB/okt.

MODELOWANIE TRAKTU GŁOSOWEGO

- Model fizyczny

Fizycznie trakt głosowy można zamodelować jako połączenie ściętych stożków lub walców. W pierwszym przypadku powstaje model **tubowy**, zachowujący ciągłość przekroju, w drugim - model **cylindryczny**.

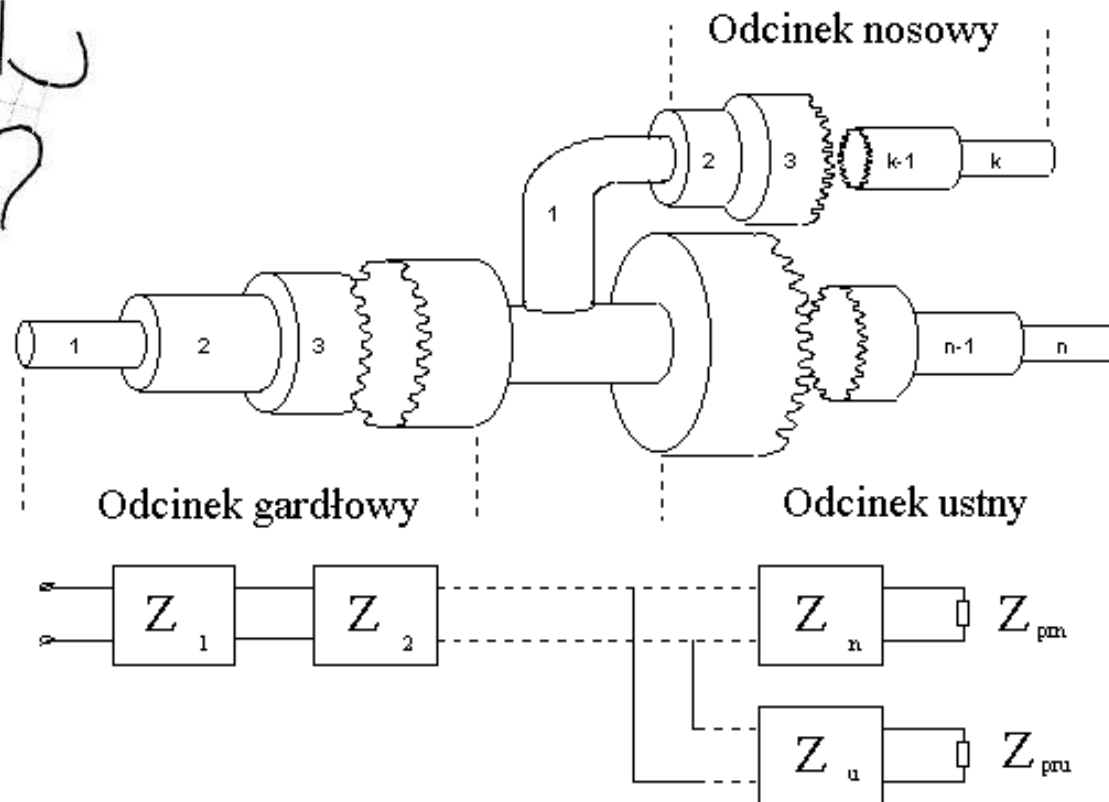
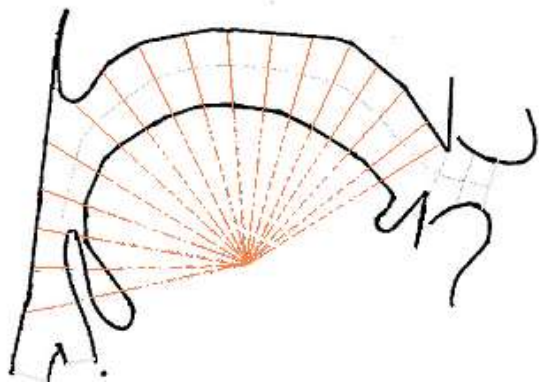


model tubowy

model
cylindryczny

MODELOWANIE TRAKTU GŁOSOWEGO

- Model fizyczny



MODELOWANIE TRAKTU GŁOSOWEGO

- rezonator Helmholtza (umożliwia modelowanie pojedynczego formantu)
- podwójny rezonator Helmholtza (umożliwia modelowanie dwóch formantów)
- modele złożone z kilku rur zakończonych płaską tarczą kołową (odgroda) imitującą
- charakterystykę promieniowania ust jako nadajnika dźwięku
- trójparametrowy model Fanta, uwzględniający rozkład biegunów i zer na płaszczyźnie zespolonej i podstawowe trzy parametry: miejsce artykulacji (miejsce największego przewężenia kanału), stopień tego przewężenia (powierzchnia przekroju) oraz kształt otworu wylotowego ust
- model Markela-Graya

MODELOWANIE TRAKTU GŁOSOWEGO

- Model elektryczny

Odwzorowanie charakterystyki częstotliwościowej za pomocą układu filtrów elektrycznych o charakterystyce rezonansowej (LC).

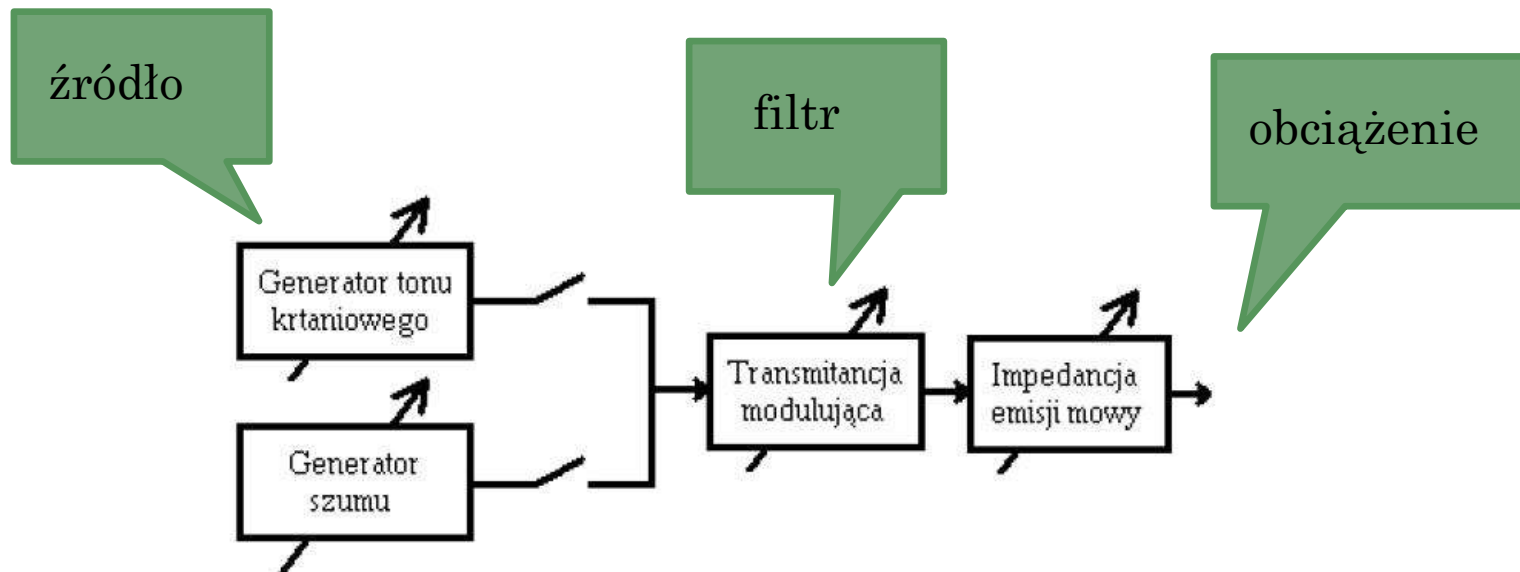
- Modele komputerowe:

- Model LPC – odwzorowanie charakterystyki traktu głosowego za pomocą filtra biegunowego.
- Model artykulacyjny – matematyczne odwzorowanie mechanizmów generowania mowy.

Więcej na wykładzie o **syntezie mowy**.

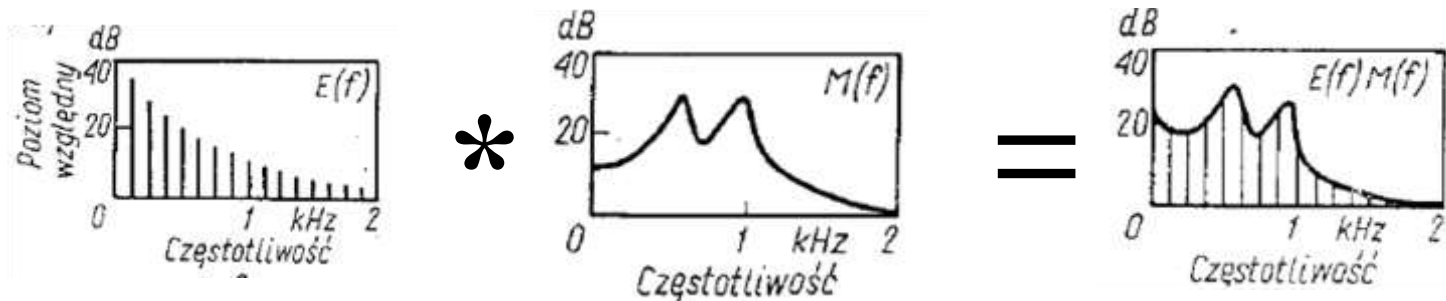
MODEL ŹRÓDŁO-FILTR

W większości modeli wytwarzania dźwięków mowy powtarza się idea, by schemat zastępczy generowania mowy przedstawiać w układzie **źródło-filtr**.



MODEL ŹRÓDŁO-FILTR

Jak w teorii obwodów, odpowiedź układu powstaje z wymnożenia charakterystyk pobudzenia i filtra.



The slide features a decorative left margin with several vertical bars of varying shades of green and a cluster of five green circles of different sizes. The largest circle is at the top left, with four smaller circles arranged below and to its right.

DZIĘKUJĘ ZA UWAGĘ!