

PARAMETRIZACJA SYGNAŁU MOWY. PERCEPTUALNE SKALE CZĘSTOTLIWOŚCI.

1

mgr inż. Kuba Łopatka

Katedra Systemów Multimedialnych

p. 628, tel. 348-6332

PLAN WYKŁADU

1. Potrzeba i istota parametryzacji
2. Klasyfikacja parametrów
3. Parametry czasowe
4. Parametry widmowe
5. Parametry formantowe
6. Parametry cepstralne
7. Parametry LPC
8. Perceptualne skale częstotliwości
9. Parametry w skalach perceptualnych



POTRZEBA I ISTOTA PARAMETRYZACJI

3

POTRZEBA PARAMETRYZACJI

- Co parametryzujemy?
- W jakim celu?
- W jaki sposób?
- Jak wykorzystamy te parametry?

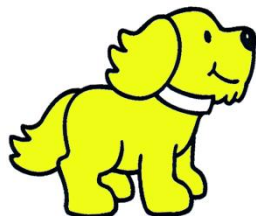
PARAMETR A CECHA OBIEKTU

obiekt

cecha

parametr

coś, co potrafimy
wyróżnić –
dotknąć, nazwać,
wskazać



Fazor

abstrakcyjna
jakość związana z
obiektem,
odróżniająca go
od innych
obiektów

żółty
krótkie łapy
duże uszy

liczbowe
wyrażenie
cechy
obiektu

kolor = 0xEFFD16
długość łap = 0.3
powierzchnia uszu = 2.5



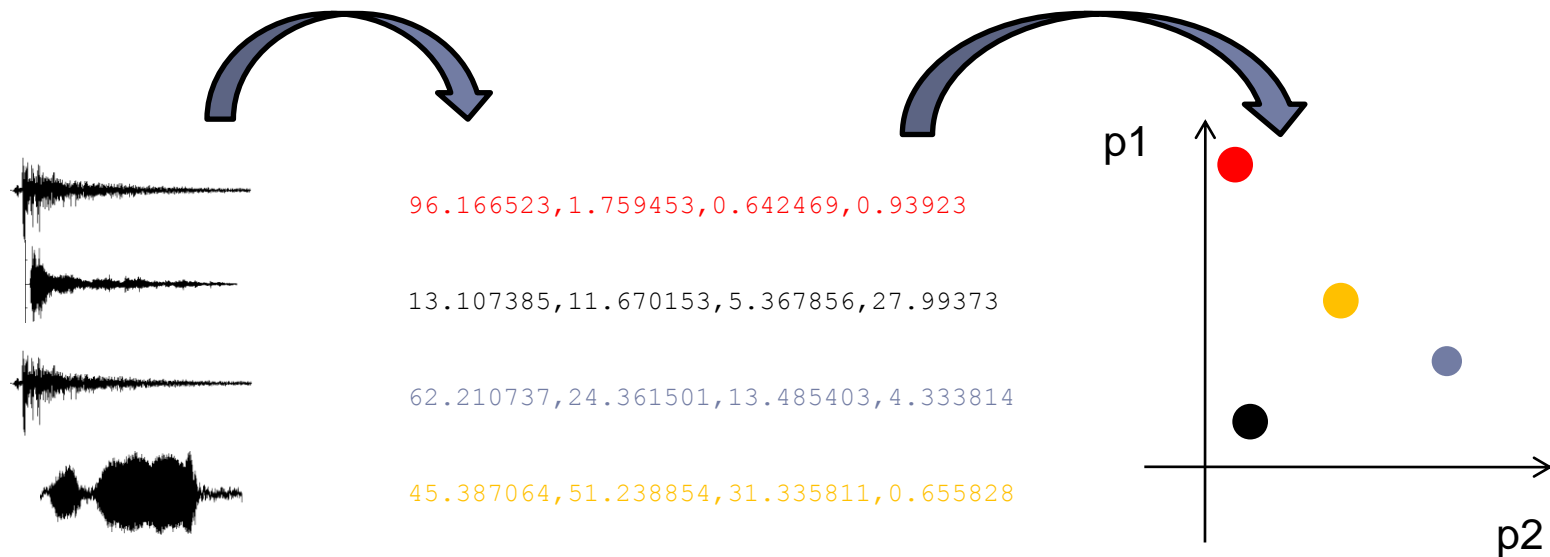
KLASA OBIEKTÓW

Obiekty, które mają wspólną jakość, tworzą **klasę** obiektów. Pewne cechy obiektów w ramach klasy powinny być zbliżone.

POTRZEBA PARAMETRYZACJI

- Komputery operują na liczbach, nie na abstrakcyjnych cechach.
- Abstrakcyjne cechy obiektu, oparte na subiektywnych wrażeniach, można mnożyć w nieskończoność. Parametryzacja uściśla i formalizuje opis obiektu.
- Wykorzystanie parametrów i ich analizy pozwala nam czasem zauważyć różnice, z których istnienia nie zdawaliśmy sobie sprawy.

ISTOTA PARAMETRYZACJI



Odwzorowanie dowolnych obiektów za pomocą wektorów liczb o skończonej długości - przejście z przestrzeni o nieskończonym wymiarze do przestrzeni K parametrów.

CEL PARAMETRYZACJI

- Odróżnienie od siebie obiektów różnych klas
- Rozpoznanie obiektu nieznanej klasy
- Weryfikacja przynależności obiektu do klasy



EKSTRAKCJA PARAMETRÓW

Proces obliczania parametrów nazywa się często **ekstrakcją cech obiektu** (lub *cech sygnału*), ang. *feature extraction*.

Ekstrakcja cech powinna być procesem powtarzalnym, deterministycznym i sformalizowanym matematycznie.

Wynikiem ekstrakcji parametrów jest **wektor cech** związany z obiektem.



PARAMETRIZACJA W MOWIE

Rozpoznawanie mowy

klasa – konkretna głoska

obiekty – nagrane sygnały zawierające głoskę

cechy – rozmieszczenie formantów, dźwięczność,
szumowość...

parametry - ...

Rozpoznawanie mówcy

klasa – konkretny mówca

obiekty – nagrane wypowiedzi mówcy

cecha – barwa głosu, wysokość głosu, rozmieszczenie
formantów...

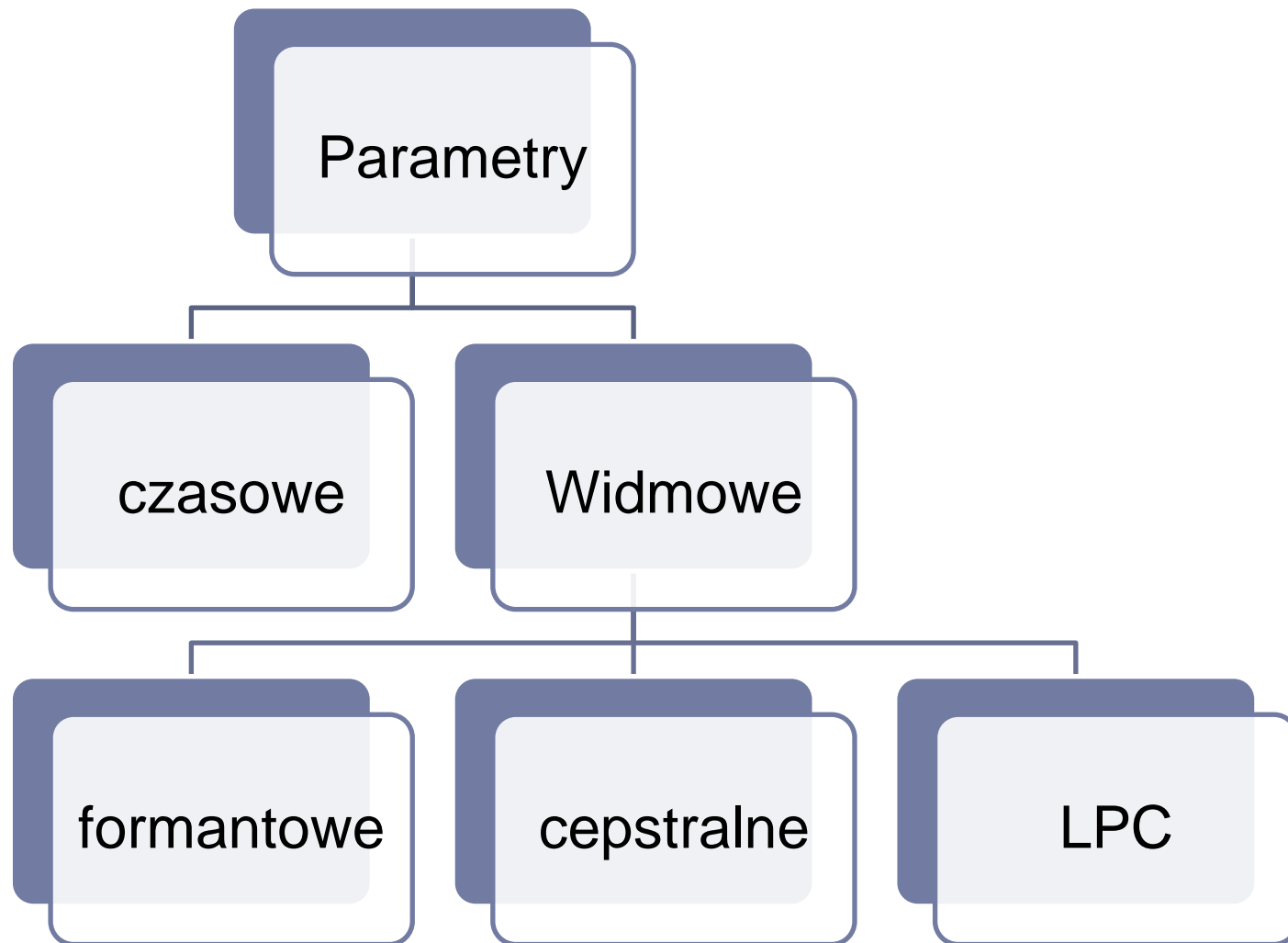
parametry - ...



KLASYFIKACJA PARAMETRÓW

12

KLASYFIKACJA PARAMETRÓW





PARAMETRY CZASOWE

14

PARAMETRY CZASOWE

Parametry **czasowe** są to parametry, które są wyznaczane wyłącznie na podstawie postaci czasowej sygnału.

Przykłady:

- Energia sygnału
- Środek ciężkości sygnału
- Obwiednia sygnału
- Gęstość przejść przez zero

ENERGIA SYGNAŁU

- Energię sygnału mowy obliczamy wg wzoru:

$$E = \sum_n x^2(n)$$

- Wartość skuteczna (RMS)

$$X_{sk} = \sqrt{\frac{1}{N} \sum_{n=1}^N x^2(n)}$$

OBWIEDNIA SYGNAŁU

- Obwiednia sygnału – wyznaczana w ramkach – obrazuje przebieg amplitudy

$$O(n) = \{o_1 \quad o_2 \quad \dots \quad o_N\}$$
$$o_n = \sqrt{\frac{1}{K} \sum_{k=1}^K x_n^2(k)}$$

Gdzie x_n – n -ta ramka sygnału

ŚRODEK CIĘŻKOŚCI SYGNAŁU

- Środek ciężkości – ang. *Temporal Centroid* – środek ciężkości obwiedni sygnału w dziedzinie czasu.

$$TC = \frac{\sum_{n=1}^N n \cdot O(n)}{\sum_{n=1}^N O(n)}$$

GĘSTOŚĆ PRZEJŚĆ PRZEZ ZERO

- Historycznie jeden z pierwszych parametrów obliczanych dla sygnału mowy. Ang. *zero crossing density* (ZCD). Wziął się stąd, że zbinaryzowana fala dźwięków mowy $\{-1;1\}$ jest dobrze rozpoznawana przez człowieka. Parametr może też być wyznaczony w ramkach w formie wektora.

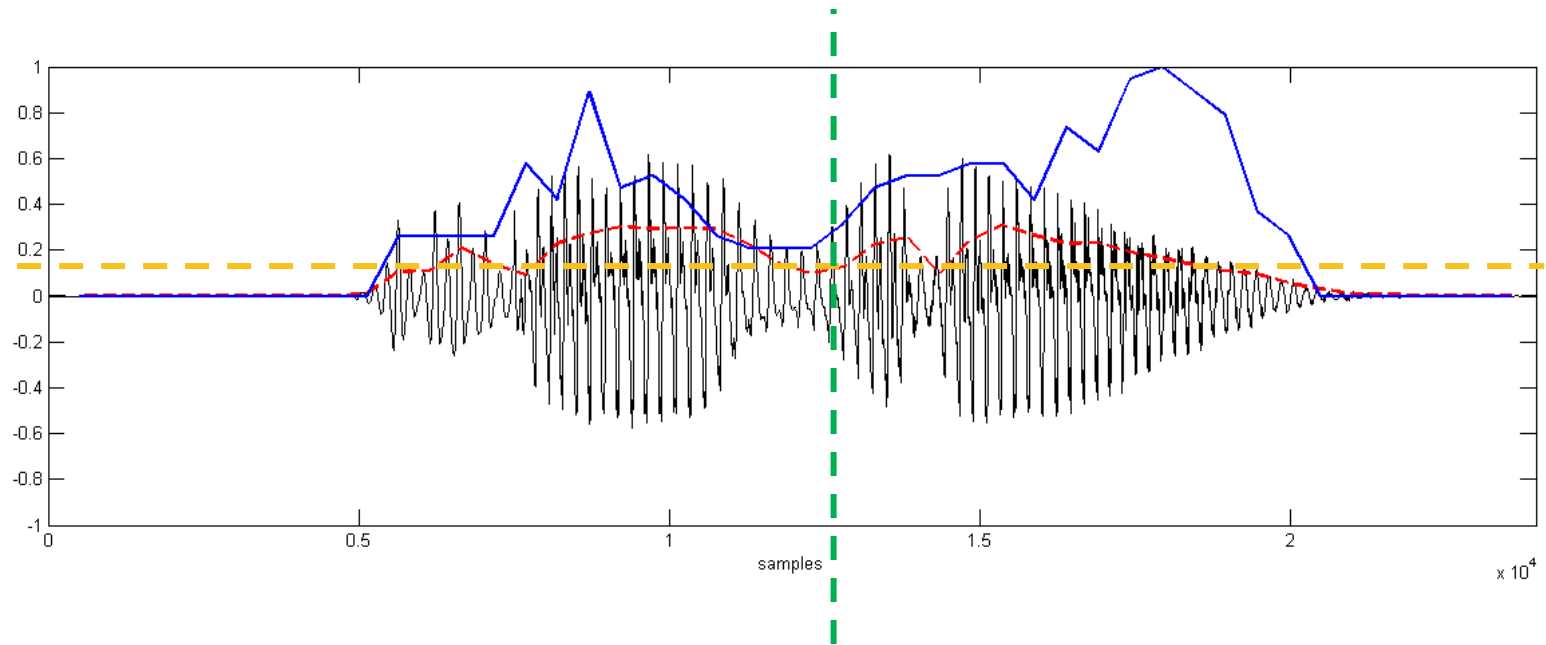
$$ZCD = \rho_0$$

$$\rho_0(n) = \{\rho_0(1) \quad \rho_0(2) \quad \dots \quad \rho_0(N)\}$$

Jeśli zastosowano preemfazę 6dB/oktawę, parametr oznacza się ρ_1

Jeśli zastosowano preemfazę 12dB/oktawę, parametr oznaczamy ρ_2

PARAMETRY CZASOWE



- Obwiednia
- Gęstość przejść przez zero
- $E=0.1627$
- $TC=12581$ [smpl]



PARAMETRY WIDMOWE

21

PARAMETRY WIDMOWE

- Parametry widmowe wyznaczone są na podstawie estymaty widma sygnału.

Estymacja widma sygnału:

- Funkcja widmowej gęstości mocy (*power spectral density* – *PSD*): periodogram, estymator Welcha, autokorelacja – **widmo mocy**
- Moduł DFT sygnału – **widmo amplitudowe**

MOMENTY WIDMOWE

- Momenty widmowy m -tego rzędu definiuje się następująco:

$$M(m) = \sum_{k=0}^{\infty} |G(k)| \cdot [f_k]^m$$

gdzie: $G(k)$ – wartość widma mocy dla k -tego pasma częstotliwości
 f_k – częstotliwość środkowa k -tego pasma

- Moment unormowany m -tego rzędu

$$M_u(m) = \frac{M(m)}{M(0)}$$

- Moment normalizujący zerowego rzędu ma sens mocy sygnału

$$M(0) = \sum_{k=0}^{\infty} |G(k)|$$

MOMENTY WIDMOWE

- Moment unormowany pierwszego rzędu ma sens środka ciężkości widma (*spectral centroid*):

$$M_u(1) = \sum_{k=0}^{\infty} \frac{|G(k)| \cdot f_k}{M(0)}$$

- Momenty unormowane centralne liczone są względem środka ciężkości widma:

$$M_{uc}(m) = \sum_{k=0}^{\infty} \frac{|G(k)| \cdot [f_k - M_u(1)]^m}{M(0)}$$

MOMENTY WIDMOWE

Inne momenty, które mają sensowną interpretację:

- Moment unormowany centralny drugiego rzędu – kwadrat szerokości widma

$$M_{uc}(2) = \sum_{k=0}^{\infty} \frac{|G(k)| \cdot [f_k - M_u(1)]^2}{M(0)}$$

- Moment unormowany centralny trzeciego rzędu oznacza skośność widma (ang. *skewness*)

$$M_{uc}(3) = \sum_{k=0}^{\infty} \frac{|G(k)| \cdot [f_k - M_u(1)]^3}{M(0)}$$

MOMENTY WIDMOWE

- Momenty unormowany centralne rzędu 2 i 4 wykorzystuje się do obliczenia kurtozy – miary płaskości widma sygnału:

$$kurtosis = \frac{M_{uc}(4)}{[M_{uc}(2)]^2}$$

inaczej:

$$kurtosis = \frac{1}{N} \sum_{j=1}^N \frac{(x_j - \bar{x})^4}{\sigma_x^4}$$

gdzie: x_j – j -ta obserwacja spośród N dostępnych obserwacji
 \bar{x} – średnia arytmetyczna dla wszystkich N obserwacji
 σ_x – odchylenie standardowe liczone na podstawie obserwacji

PŁASKOŚĆ WIDMOWA

- Płaskość widmowa (ang. *spectral flatness measure* – *SFM*) – stosunek średniej geometrycznej i arytmetycznej współczynników widma – miara harmoniczności sygnału

$$SFM = 10 \cdot \log \left\{ \frac{\left[\prod_{k=1}^{N/2} P \left(e^{j \frac{2\pi k}{N}} \right) \right]^{1/N/2}}{\frac{1}{N/2} \cdot \sum_{k=1}^{N/2} P \left(e^{j \frac{2\pi k}{N}} \right)} \right\}$$

$P \left(e^{j \frac{2\pi k}{N}} \right)$ - widmowa gęstość mocy

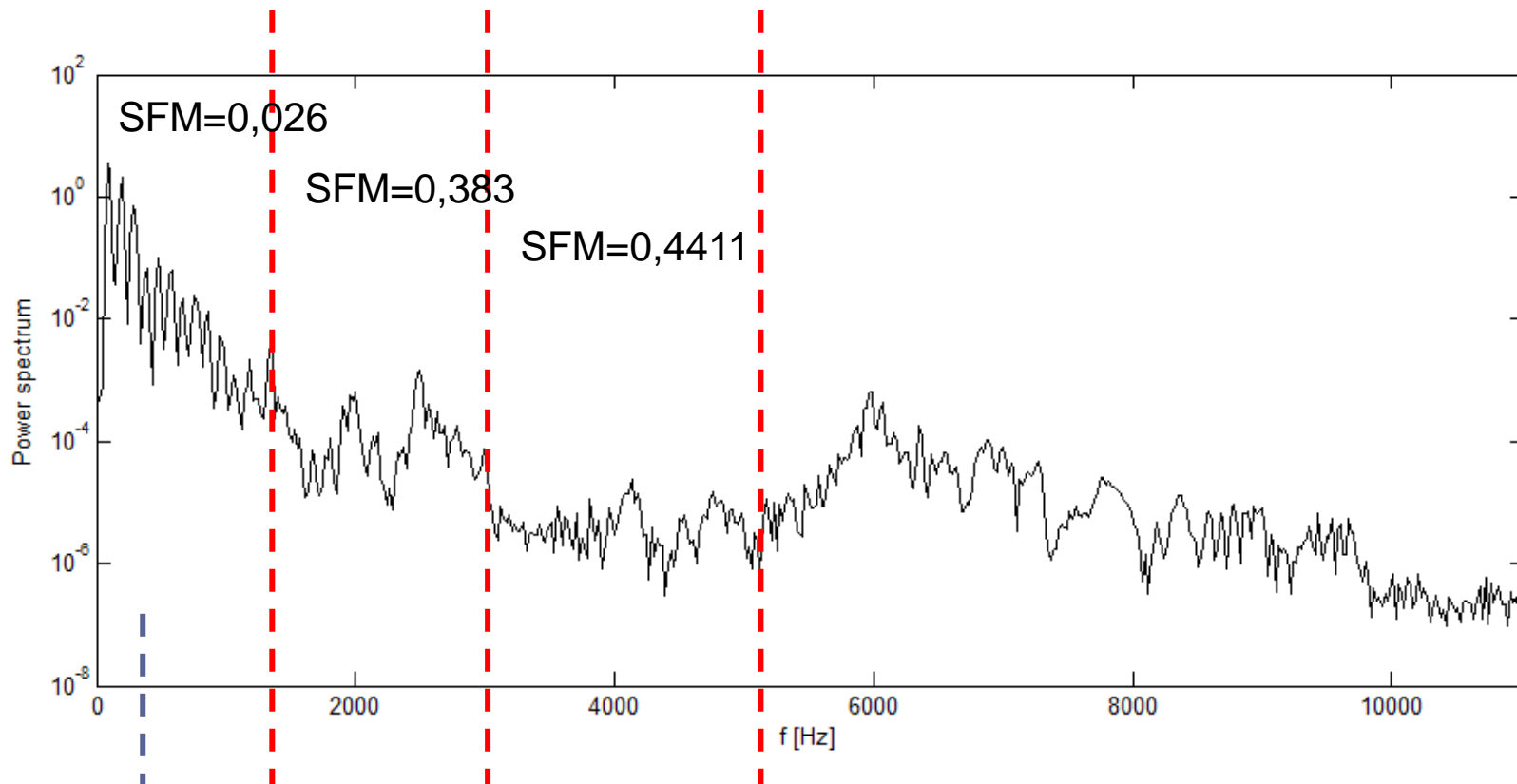
Parametr SFM może być również wyznaczany w pasmach

MPEG-7

- Ogrom parametrów widmowych (i nie tylko) zdefiniowanych jest w tzw. standardzie MPEG-7. Na przykład:
 - Audio Spectrum Envelope
 - Audio Spectrum Spread
 - Audio Spectrum Centroid
 - Harmonic Spectral Centroid
 - Harmonic Spectral Spread
 - Audio Spectrum Flatness
 - ...

Większość z nich jest jednak o wiele częściej stosowana dla sygnałów muzycznych.

PARAMETRY WIDMOWE



$\mu_1=178,34$ - środek ciężkości widma

$\mu_{c3} = 1,8 * 10^8$ – skośność

Kurtoza – 506.4653



PARAMETRY FORMANTOWE

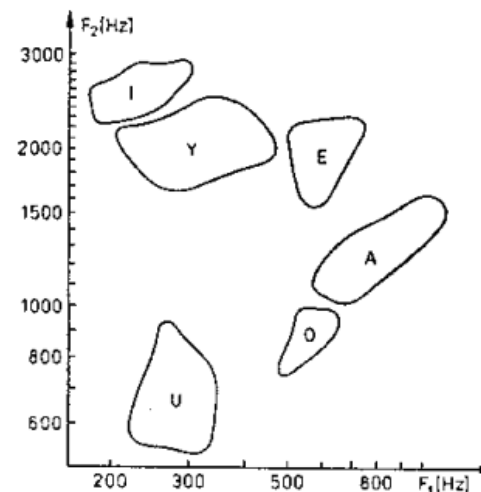
30

PARAMETRY FORMANTOWE

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



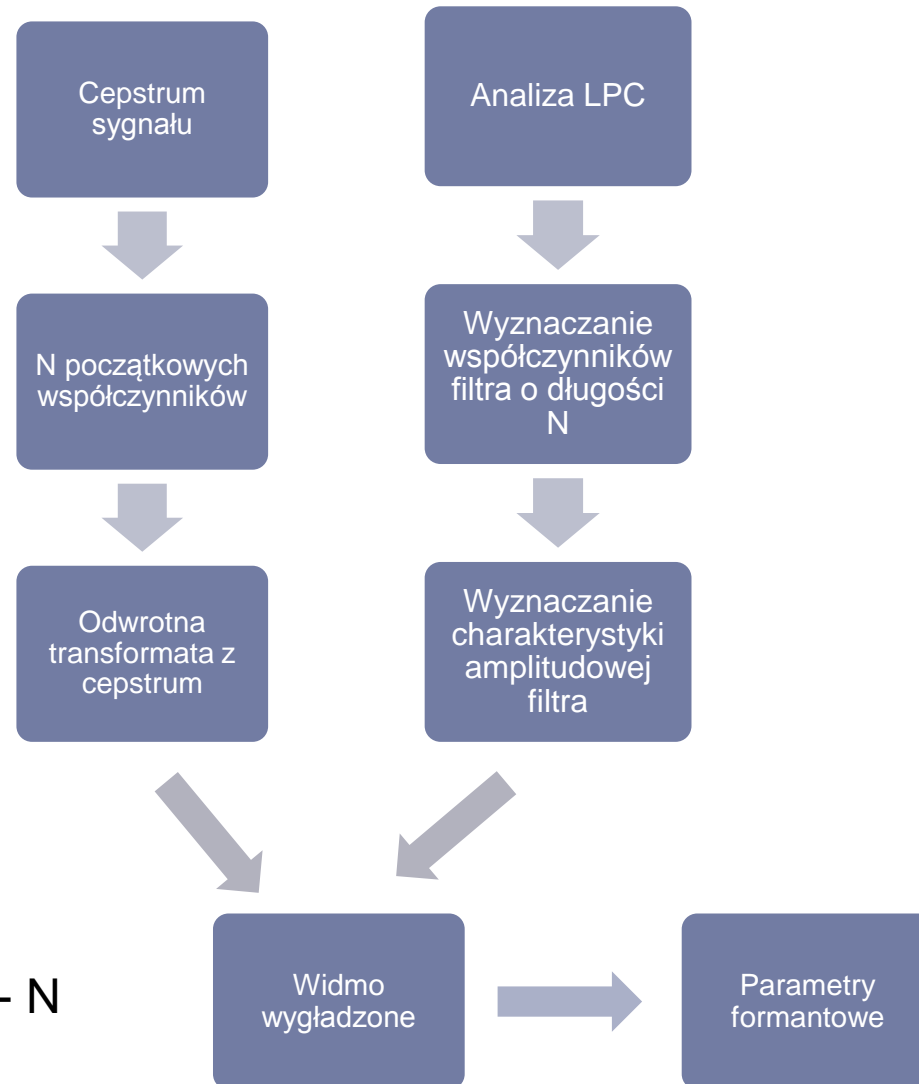
WYZNACZANIE PARAMETRÓW FORMANTOWYCH

- Do wyznaczania parametrów formantowych potrzebne jest tzw. wygładzone widmo sygnału (obwiednia widmowa – *spectral envelope*).

Metody wyznaczania widma wygładzonego:

- Cepstralna
- LPC

WYZNACZANIE PARAMETRÓW FORMANTOWYCH





PARAMETRY CEPSTRALNE

34

CEPSTRUM

Cepstrum to transformata Fouriera logarytmu widma.

Cepstrum zespolone:

$$\hat{X}(T) = F[\ln(X(f))]$$

Cepstrum mocy (logarytm widma amplitudowego)

$$\hat{X}(T) = F[\ln|X(f)|]$$

Cepstrum mocy (transformata kosinusowa)

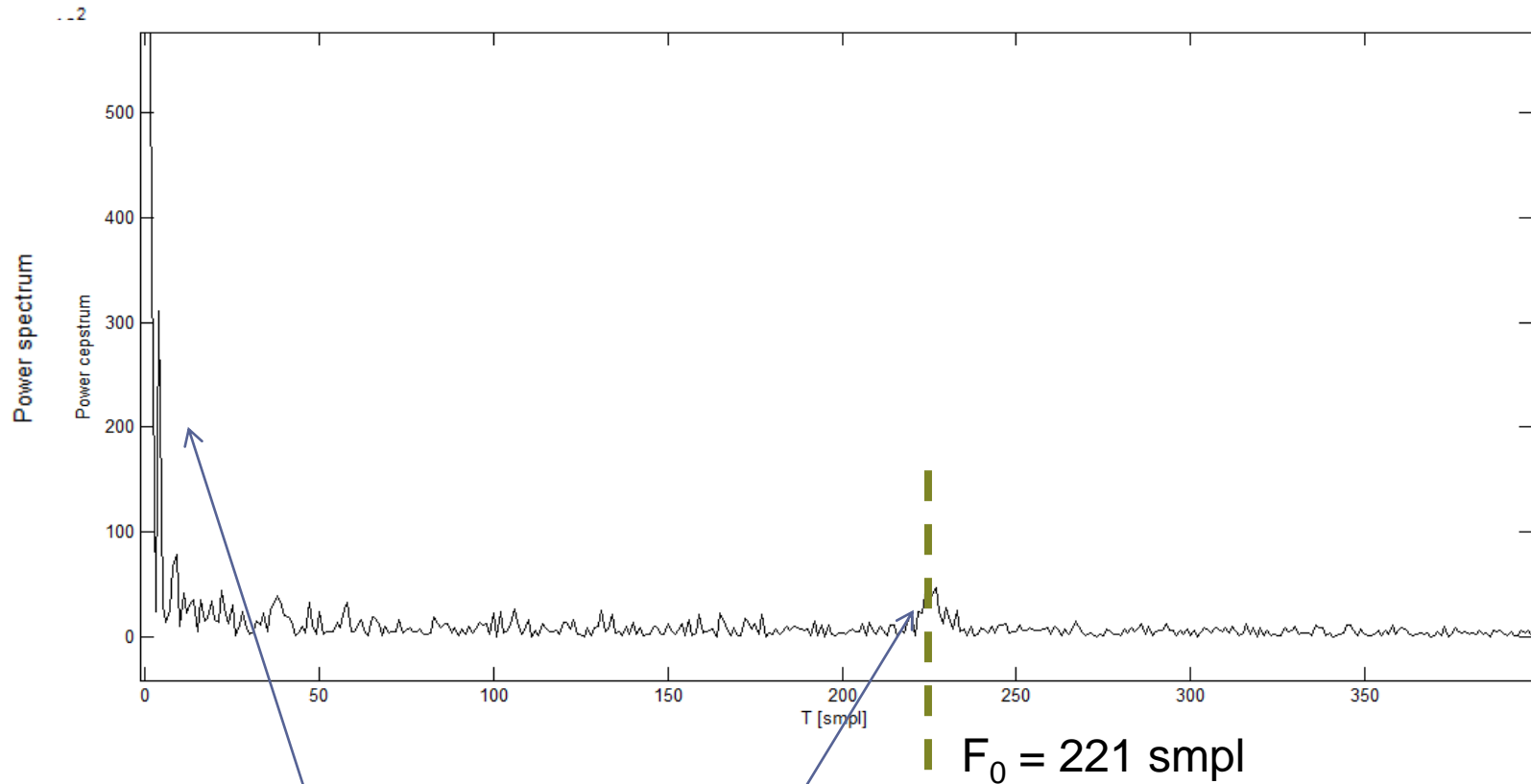
$$\hat{X}_c(k) = \sum_{n=0}^{N-1} [\ln|X(n)|] \cdot \cos\left(\frac{(n-0.5) \cdot k \cdot \pi}{N}\right)$$

PARAMETRY CEPSTRALNE

- Skala cepstrum odpowiada dziedzinie **czasu**
- Niskie współczynniki cepstralne niosą informacje o trakcie głosowym (rozpoznawanie mowy)
- Wysokie współczynniki cepstralne niosą informacje o tonie krtaniowym (ekstrakcja F_0)

Wektor parametrów cepstralnych to wektor wybranych współczynników cepstrum (lub parametrów wyznaczonych z tych współczynników).

PARAMETRY CEPSTRALNE



Charakterystyka traktu
głosowego

Charakterystyka tonu
krtaniowego

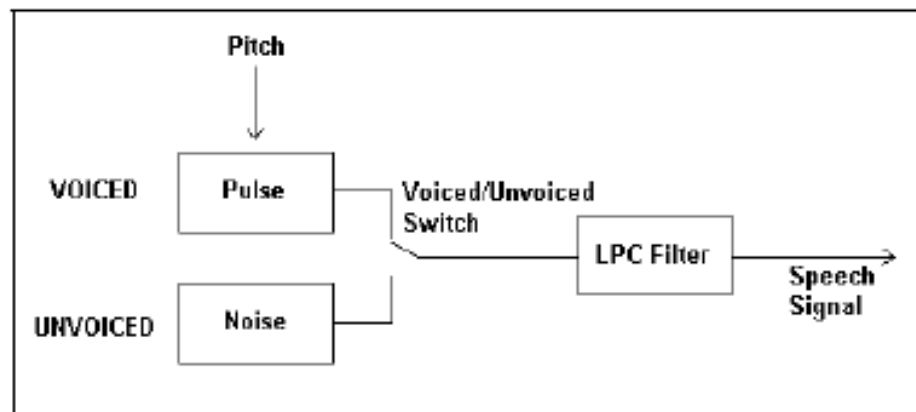


PARAMETRY LPC

38

LINIOWE KODOWANIE PREDYKCYJNE

Liniowe kodowanie predykcyjne (ang. *Linear Predictive Coding* – LPC) – technika analizy sygnału mowy polegająca na przedstawieniu sygnału mowy jako odpowiedzi filtru typu biegunowego (*all-pole filter*) na sygnał tonu krtaniowego.



LINIOWE KODOWANIE PREDYKCYJNE

Filtr biegunowy (AR – *autoregressive*)

- ma niezerowe współczynniki tylko w **mianowniku** transmitancji,
- Odzwierciedla rezonansową charakterystykę traktu głosowego.

$$H(z) = G \cdot \frac{1}{1 - \sum_{k=1}^p a_k \cdot z^{-k}}$$

LINIOWE KODOWANIE PREDYKCYJNE

Odpowiedź filtru biegunowego na pobudzenie:

$$v(n) = \sum_{k=1}^p a_k \cdot v(n - k)$$

Jest kombinacją liniową kolejnych próbek z wyjścia filtru. Oznacza to, że sygnał mowy można przewidzieć na podstawie jego poprzednich próbek. Stąd nazwa **liniowe kodowanie predykcyjne**.

Liczba próbek branych pod uwagę przy tej analizie jest zdeterminowana przez **rzęd** filtra (rzęd analizy LPC) – p .

LINIOWE KODOWANIE PREDYKCYJNE

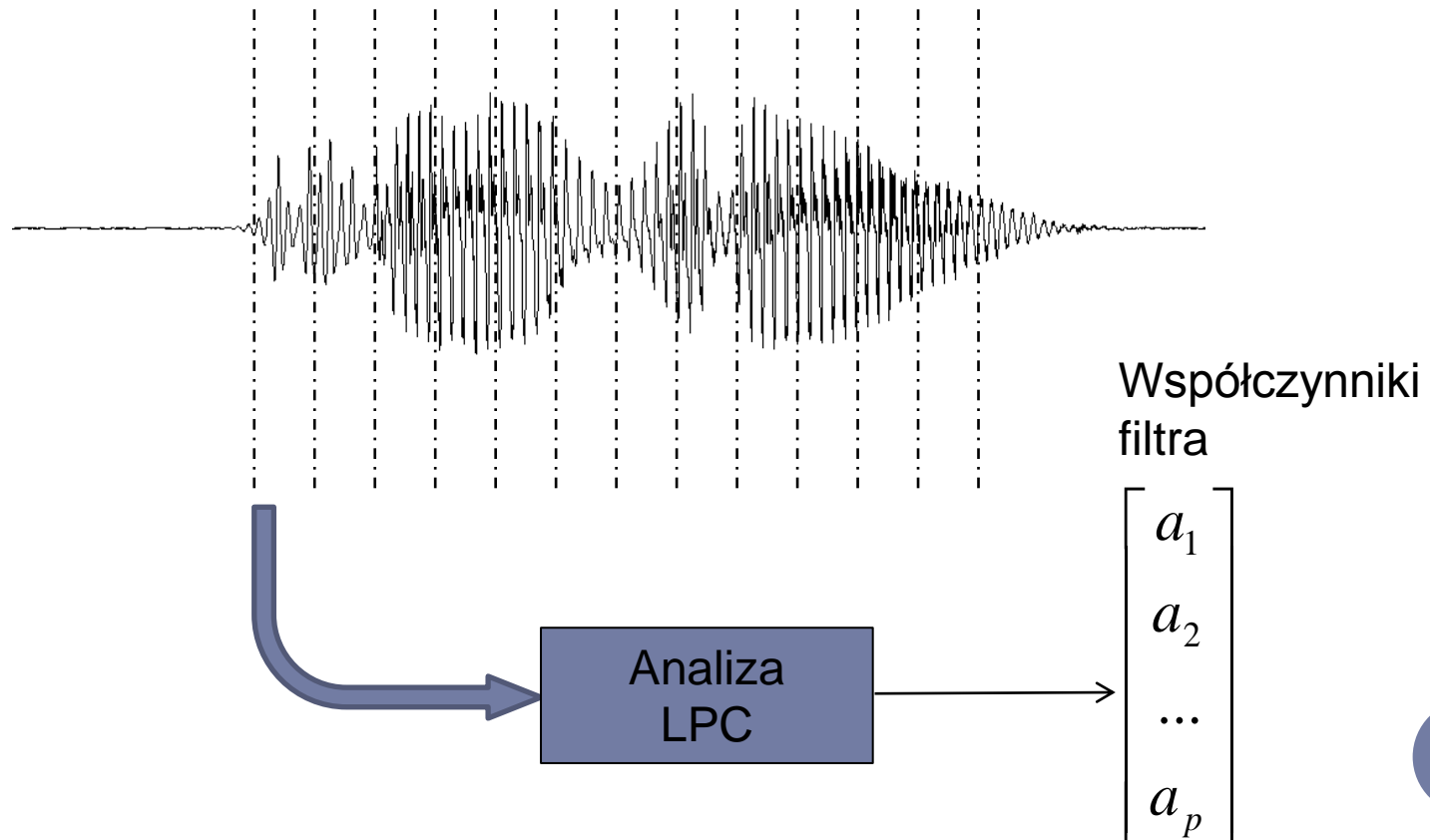
Wyznaczanie współczynników filtra LPC polega na zminimalizowaniu błędu między sygnałem a jego predykcją.

$$E = \sum_{n=1}^{N-1} e^2(n) = \sum_{n=1}^{N-1} \left[v(n) - \sum_{k=1}^p a_k \cdot v(n-k) \right]^2$$

Najczęściej rozwiązuje się ten problem metodą autokorelacyjną z zastosowaniem iteracyjnego odraczania macierzy (algorytmy Levinsona, Robinsona i Durbina).

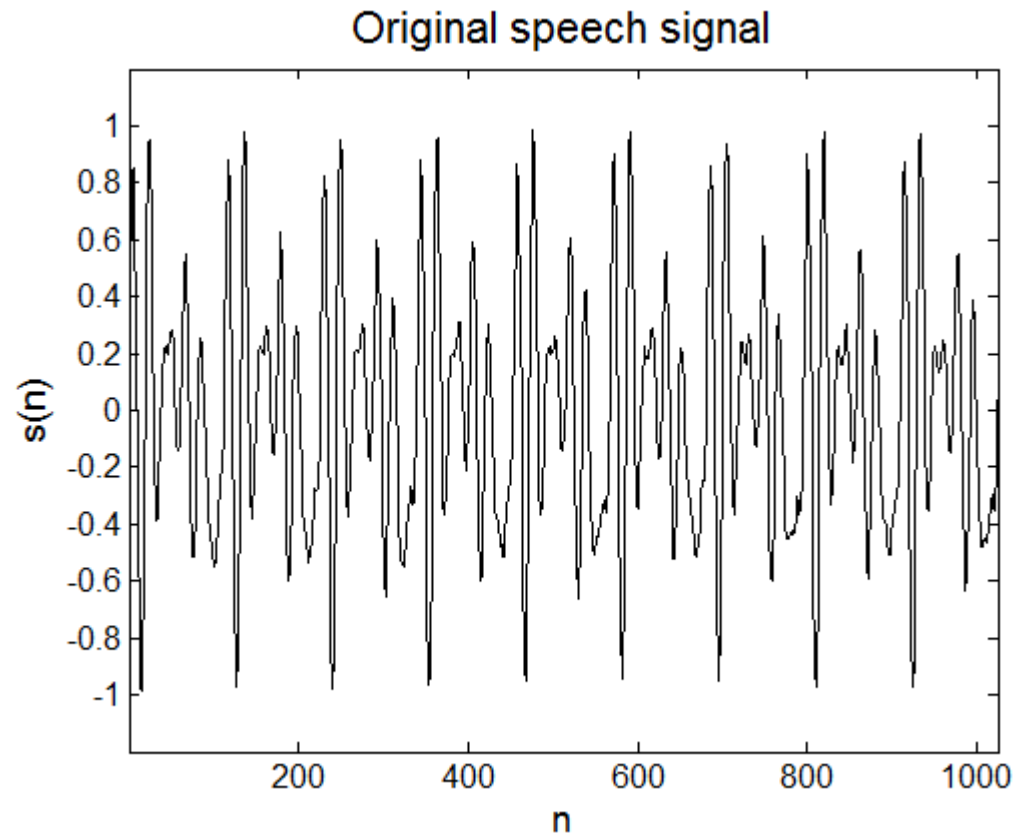
WYZNACZANIE PARAMETRÓW LPC

Parametry LPC to współczynniki filtra analizującego sygnał mowy. Wyznacza się je w ramkach, np. 25 ms.



MODELOWANIE SYGNAŁU MOWY FILTREM LPC

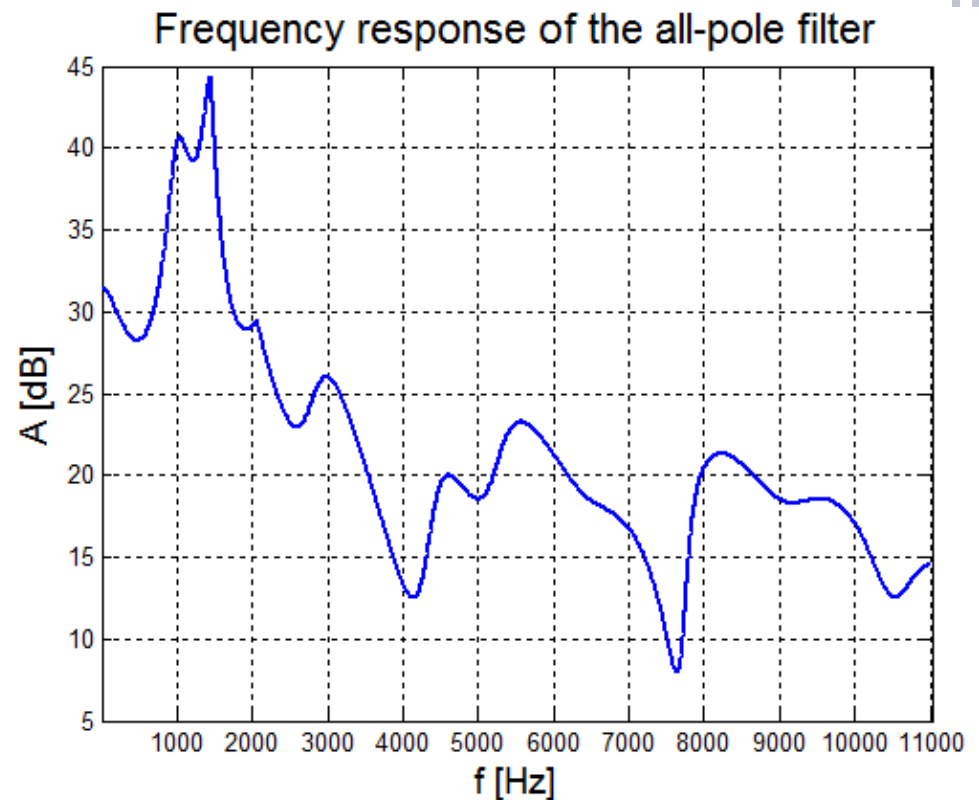
Analizujemy sygnał mowy (głoska *a*)



MODELOWANIE SYGNAŁU MOWY FILTREM LPC

- Wyznaczanie współczynników i charakterystyki filtra typu AR

```
lpc_ord = 20  
a = lpc(x, lpc_ord);  
[h, f] = freqz(1, a, 2048);
```



MODELOWANIE SYGNAŁU MOWY FILTREM AR

- Generowanie pobudzenia
pobudzenie sinusoidalne

```
t=1:L;  
for i=1:(fs/T)  
    comp=sin(t*2*i*pi/T);  
    y=y+comp;  
end
```

- pobudzenie szumem białym

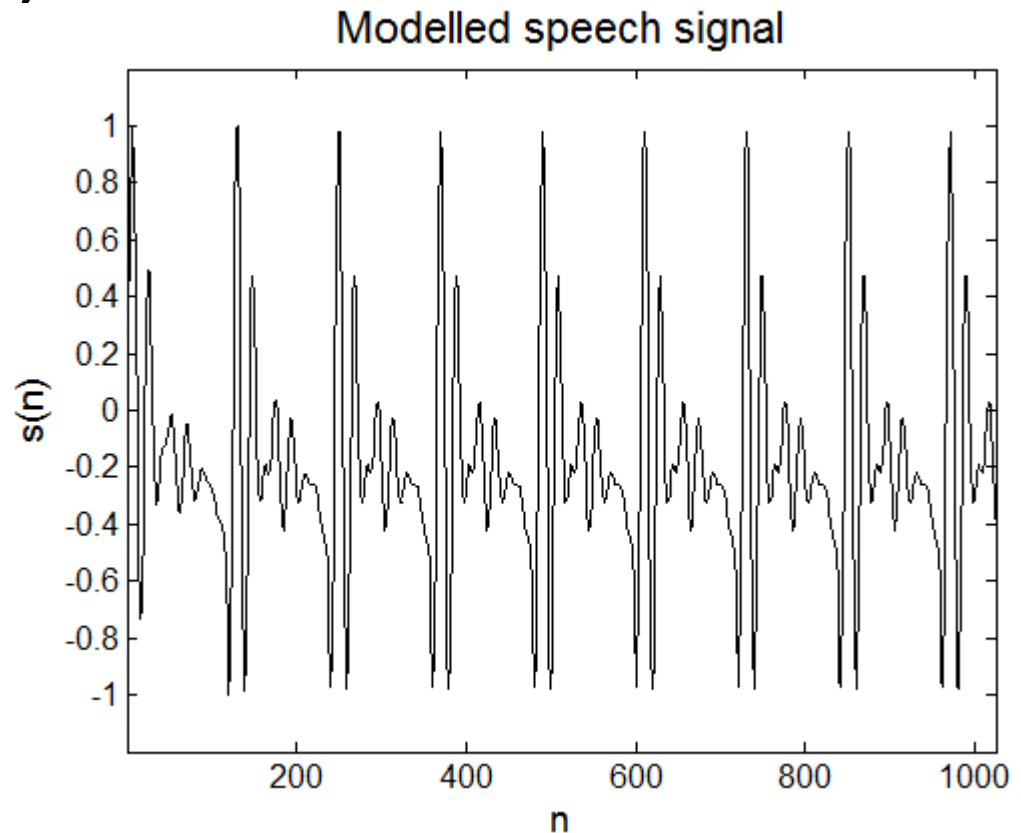
```
y=rand(1,L)
```



MODELOWANIE SYGNAŁU MOWY FILTREM AR

- Odtwarzanie sygnału na podstawie filtru i pobudzenia

`u=filter(1, a, y) :`





PERCEPTUALNE SKALE CZĘSTOTLIWOŚCI

48

PERCPETUALNE SKALE CZĘSTOTLIWOŚCI

Prawo Webbera-Fechnera głosi, że

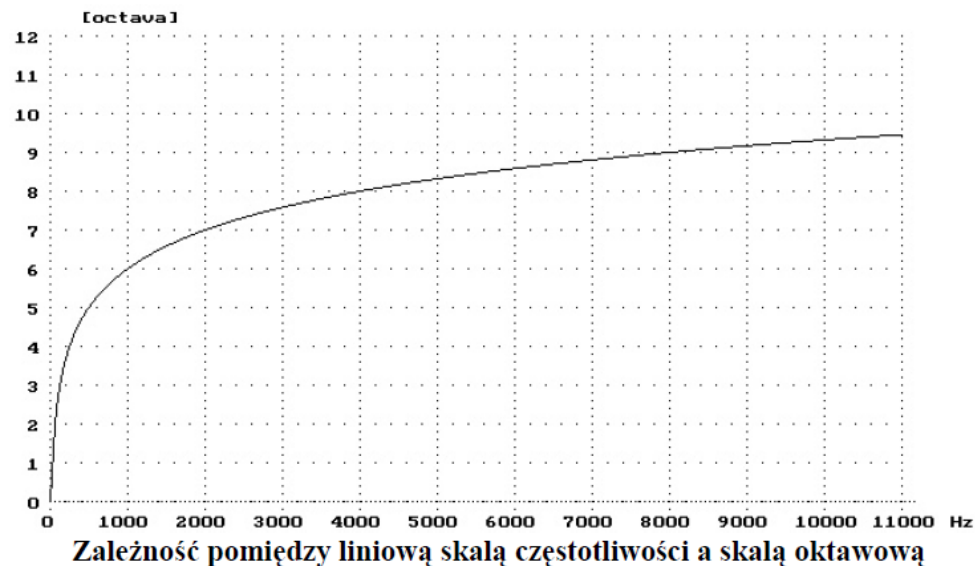
Reakcja układu biologicznego jest proporcjonalna do logarytmu pobudzającego go bodźca.

W ogólności można stwierdzić, że subiektywne **wrażenie** człowieka nie zależy w prosty sposób od obiektywnie mierzalnego pobudzenia. Oznacza to, że ludzkie ucho nie odpowiada liniowo na zwiększającą się częstotliwość.

SKALA OKTAWOWA

Najpowszechniejszą perceptualną skalą częstotliwości jest wykorzystywana w muzyce skala oktawa. Odpowiada ona strojowi równomiernie temperowanemu.

$$\text{oktawa} = \log_2(64 \cdot f)$$



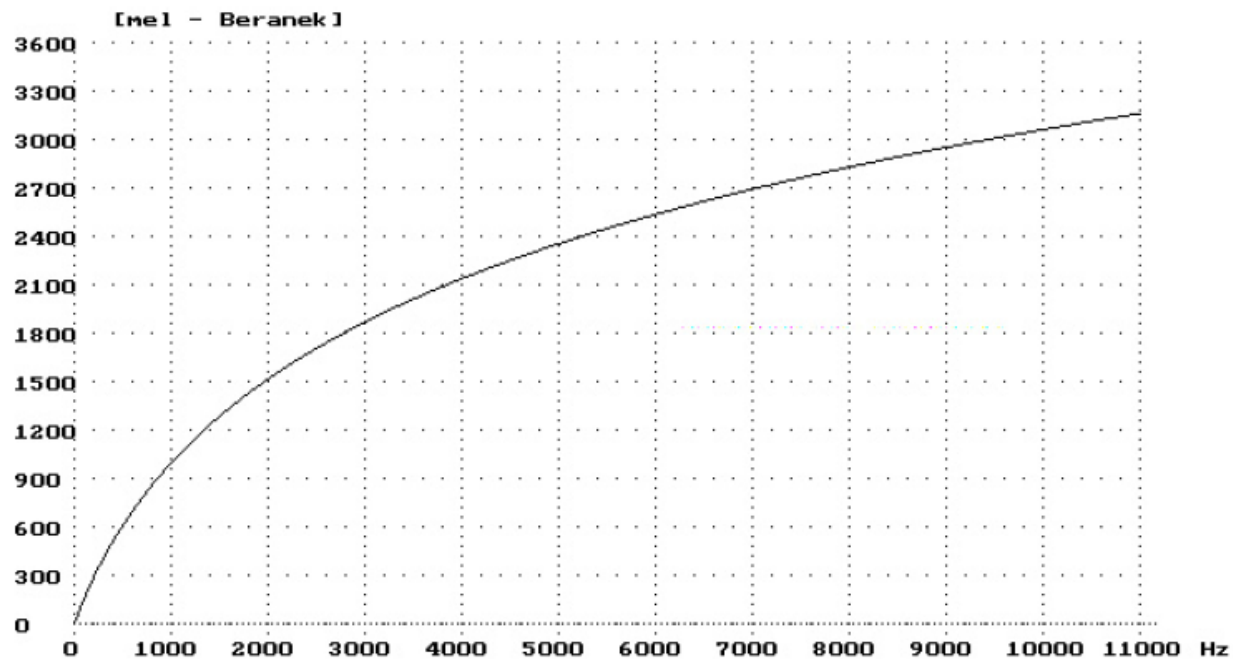
SKALA MELOWA

- Doświadczenie - zestroić dźwięki tak, by jeden był dwa razy wyższy od drugiego
- Wyznaczona w oparciu o tony proste
- Odpowiada ona subiektywnemu wrażeniu wysokości dźwięku
- W wyniku doświadczenia okazało się, że wrażenie wysokości zależy również od głośności dźwięku, stąd w definicji przyjęto poziom 40dB SPL (względem $20\mu\text{Pa}$).

SKALA MELOWA

Skala melowa wg Beranka (1000 mel = 1000 Hz)

$$M = 1127 \cdot \ln\left(1 + \frac{f}{0.7}\right)$$



Zależność pomiędzy liniową skalą częstotliwości a skalą melową Beranka

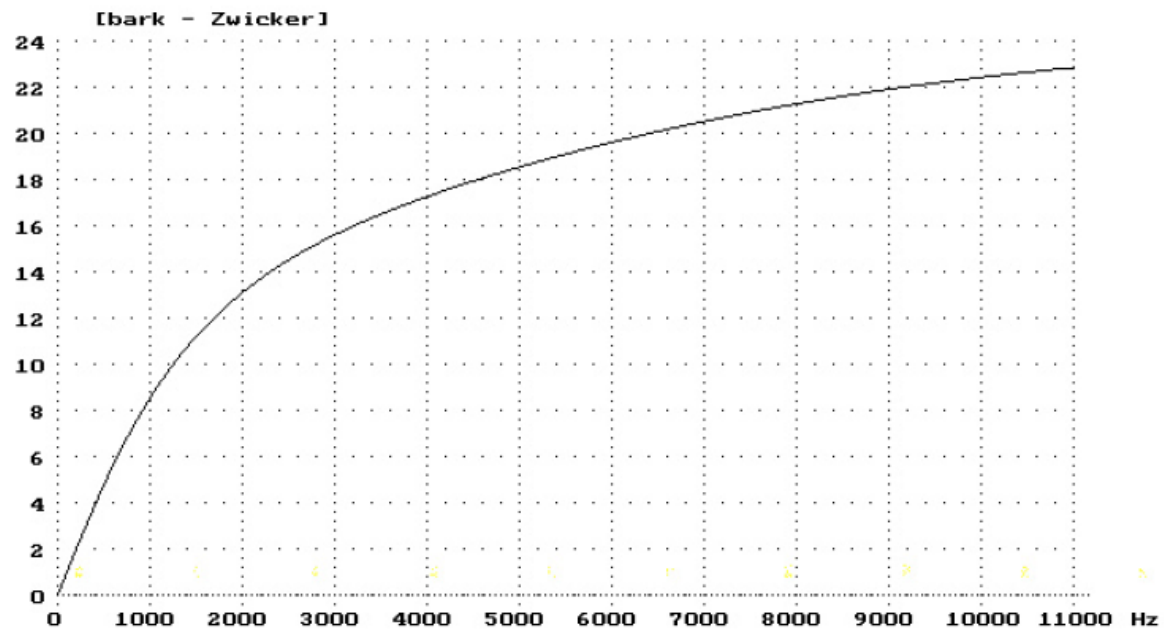
SKALA BARKOWA

- Odpowiada szerokości pasma krytycznego
- Pasma krytyczne (pojęcie podparte anatomią narządu Cortiego, teoriami słyszenia i doświadczeniami) to taki zakres częstotliwości, po którego przekroczeniu odczuwamy wyraźną zmianę głośności
- Wyróżnia się 24 pasma krytyczne
- Z pojęciem pasm krytycznych wiąże się również zjawisko maskowania

SKALA BARKOWA

Skala Barkowa wg Zwickera

$$b = 13 \cdot \arctan(0.76 \cdot f) + 3.5 \cdot \arctan\left(\left(\frac{f}{7.5}\right)^2\right)$$



Zależność pomiędzy liniową skalą częstotliwości a skalą barkową Zwickera

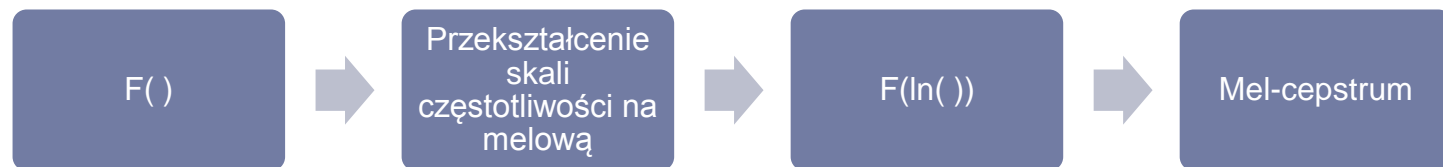


PARAMETRY W SKALACH PERCEPTUALNYCH

55

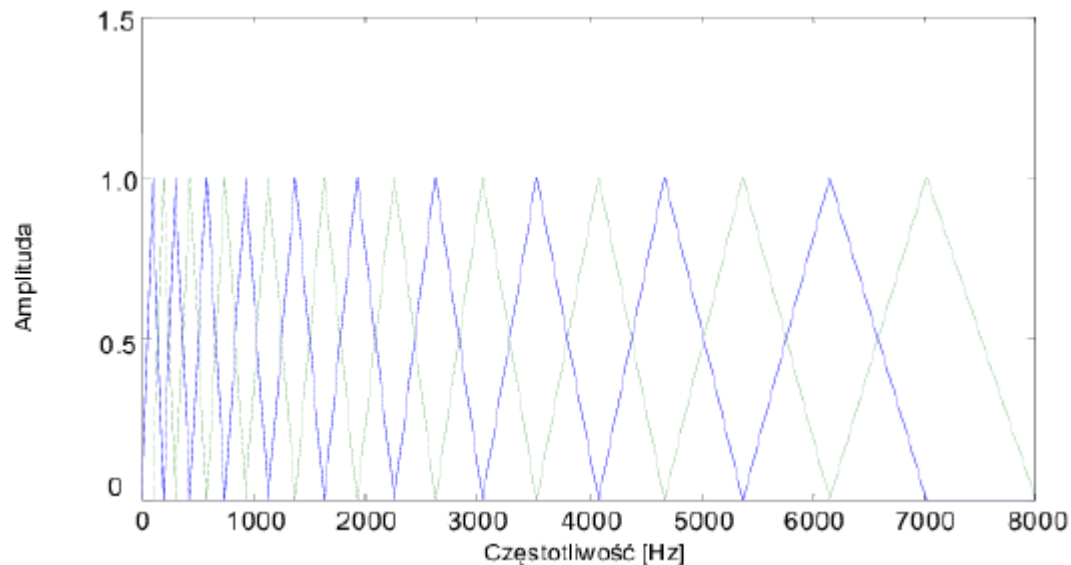
PARAMETRY MEL-CEPSTRALNE

Parametry mel-cepstralne (ang. MFCC – Mel-Frequency Cepstral Coefficients) to parametry szeroko stosowane w akustyce mowy oraz w kompresji sygnałów fonicznych. Powstają z cepstrum sygnału przedstawionego w skali melowej (mel-cepstrum).



PARAMETRY MEL-CEPSTRALNE

Skalę melową uzyskuje się poprzez filtrację sygnału bankiem filtrów o charakterystyce trójkątnej.



K- ty współczynnik mel-cepstralny odpowiada zawartości k-tego pasma. Zazwyczaj liczba pasm wynosi od 12 do 20.

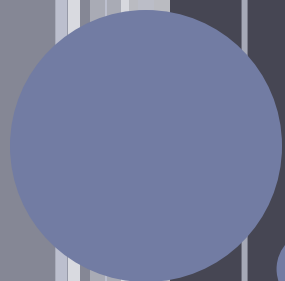
PARAMETRY MEL-CEPSTRALNE

- Wektor parametrów mel-cepstralnych to wektor współczynników cepstrum w odpowiednich pasmach melowych
- Mają za zadanie odzwierciedlać naturalną odpowiedź układu słuchowego na pobudzenie dźwiękami mowy
- Parametry mel-cepstralne cechuje mała wrażliwość na szum
- Są często wykorzystywane przy rozpoznawaniu mowy

$$\begin{bmatrix} MFCC_1 \\ MFCC_2 \\ MFCC_3 \\ \dots \\ MFCC_k \\ \dots \\ MFCC_K \end{bmatrix}$$

INNE PARAMETRY PERCEPTUALNE

- Parametry falkowe w skali barkowej
- Parametry z transformaty kosinusowej w skali barkowej
- Energia w pasmach melowych
- Energia w pasmach krytycznych
- ...



PODSUMOWANIE

OCENA PARAMETRÓW

Jak sprawdzić, czy parametr dobrze separuje klasy:

- Wykres na płaszczyźnie dwóch wybranych parametrów sprawdzający separowalność klas
- Próba klasyfikacji
- Narzędzia programowe do obróbki danych, np. WEKA
- Statystyczna ocena parametrów – np. statystyka Behrensa-Fishera

PODSUMOWANIE

- Parametryzacja to sposób **obiektywnego** opisu sygnału mowy
- Parametryzacja jest konieczna do rozpoznawania mowy lub mówcy
- Dobry parametr to taki, który dobrze odzwierciedla różnice między obiektami różnych klas
- Parametry powinny być dokładnie opisane matematycznie
- Warunki wyznaczania parametrów i ich wyniki powinny być powtarzalne
- Parametry mowy możemy wyznaczać w różnych dziedzinach (czasu, widma, cepstrum, LPC)
- Parametry wyznaczane w skalach perceptualnych odzwierciedlają naturalne wrażenia słuchowe