Akustyka Mowy
Parametryzacja sygnału mowy. Perceptualne skale częstotliwości.

Przygotował: mgr inż. Kuba Łopatka
Uzupełnił i prezentuje: dr hab. inż. Józef Kotus
Katedra Systemów Multimedialnych

Plan wykładu

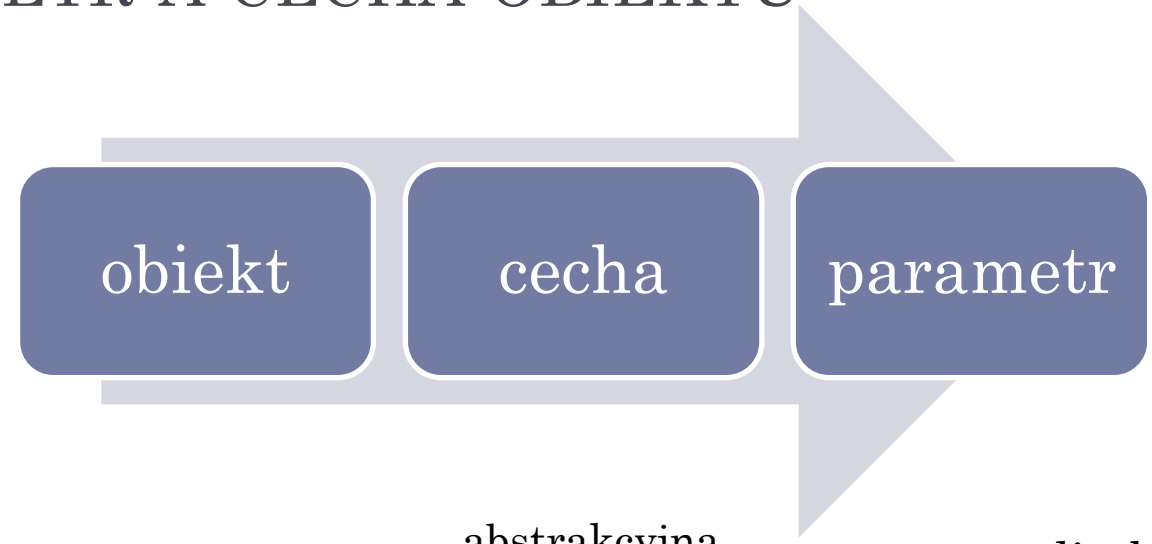
1. Potrzeba i istota parametryzacji
2. Klasyfikacja parametrów
3. Parametry czasowe
4. Parametry widmowe
5. Parametry formantowe
6. Parametry cepstralne
7. Parametry LPC
8. Perceptualne skale częstotliwości
9. Parametry w skalach perceptualnych

Potrzeba i istota parametryzacji

Potrzeba parametryzacji

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

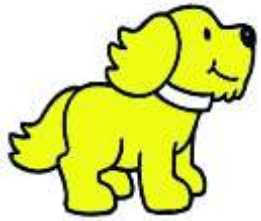
Parametr a cecha obiektu



coś, co potrafimy wyróżnić – dotknąć, nazwać, wskazać

abstrakcyjna jakość związana z obiektem, odróżniająca go od innych obiektów

liczbowe wyrażenie cechy obiektu



Fazor

żółty
krótkie łapy
duże uszy

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

Klasa obiektów

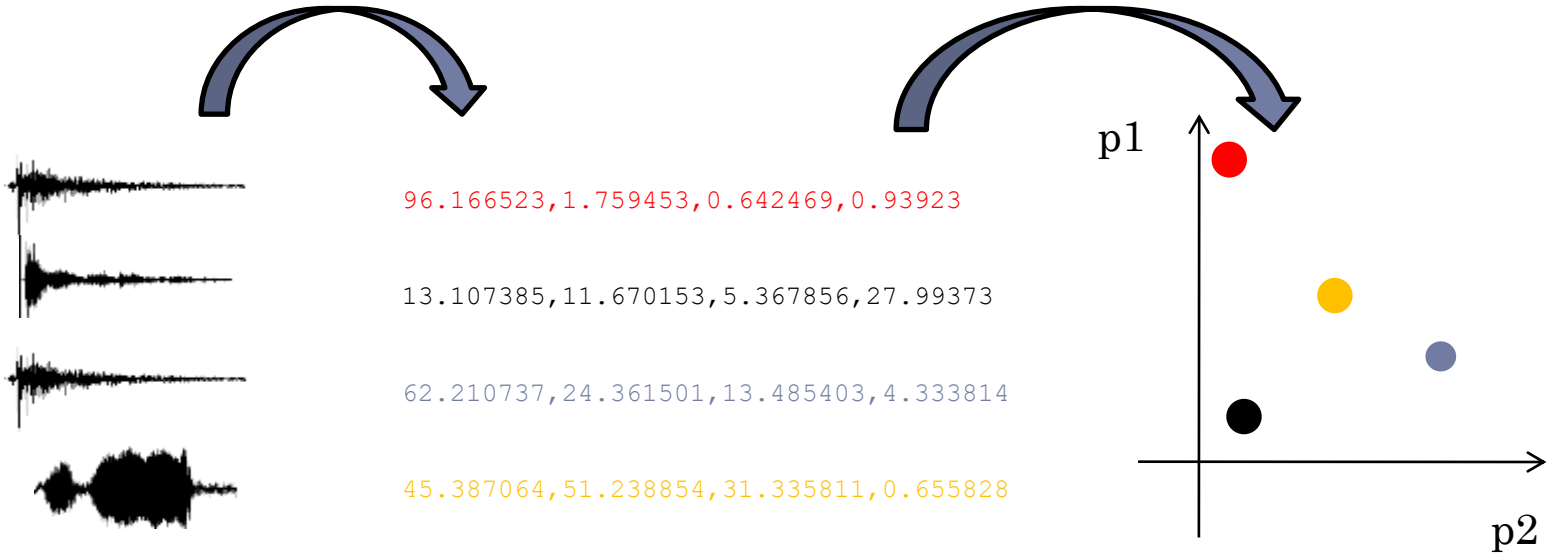
Obiekty, które mają wspólną jakość, tworzą **klasę** obiektów.

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.
- **Analiza parametryczna jest najbardziej obiektywną formą analizy sygnału mowy.**

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*.

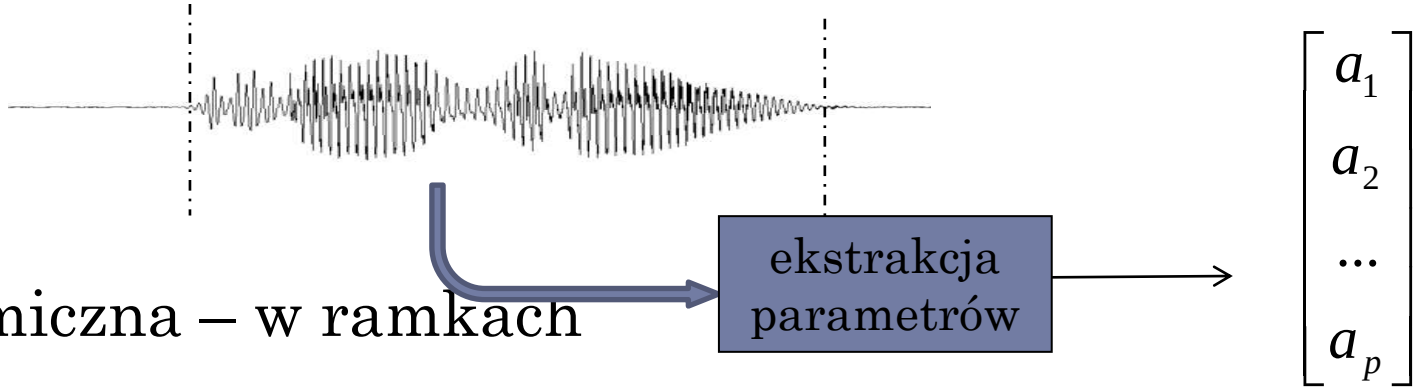
Ekstrakcja cech powinna być procesem powtarzalnym, deterministycznym i sformalizowanym matematycznie.

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

Ekstrakcja parametrów

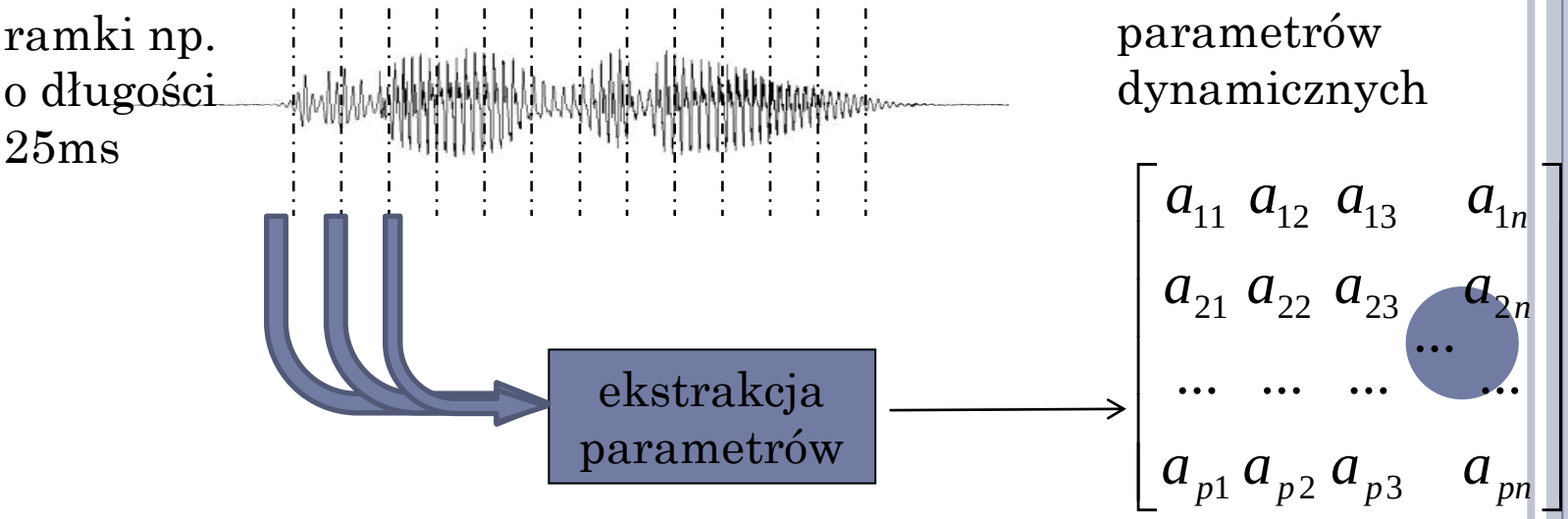
- statyczna – na całym sygnale

wektor parametrów



- dynamiczna – w ramkach

macierz parametrów dynamicznych



Parametryzacja w mowie

Rozpoznawanie mowy

klasa – konkretna **głoska**

obiekty – nagrane sygnały zawierające głoskę

cechy – rozmieszczenie formantów, dźwięczność, szumowość…

parametry - …

Rozpoznawanie mówcy

klasa – konkretny **mówca**

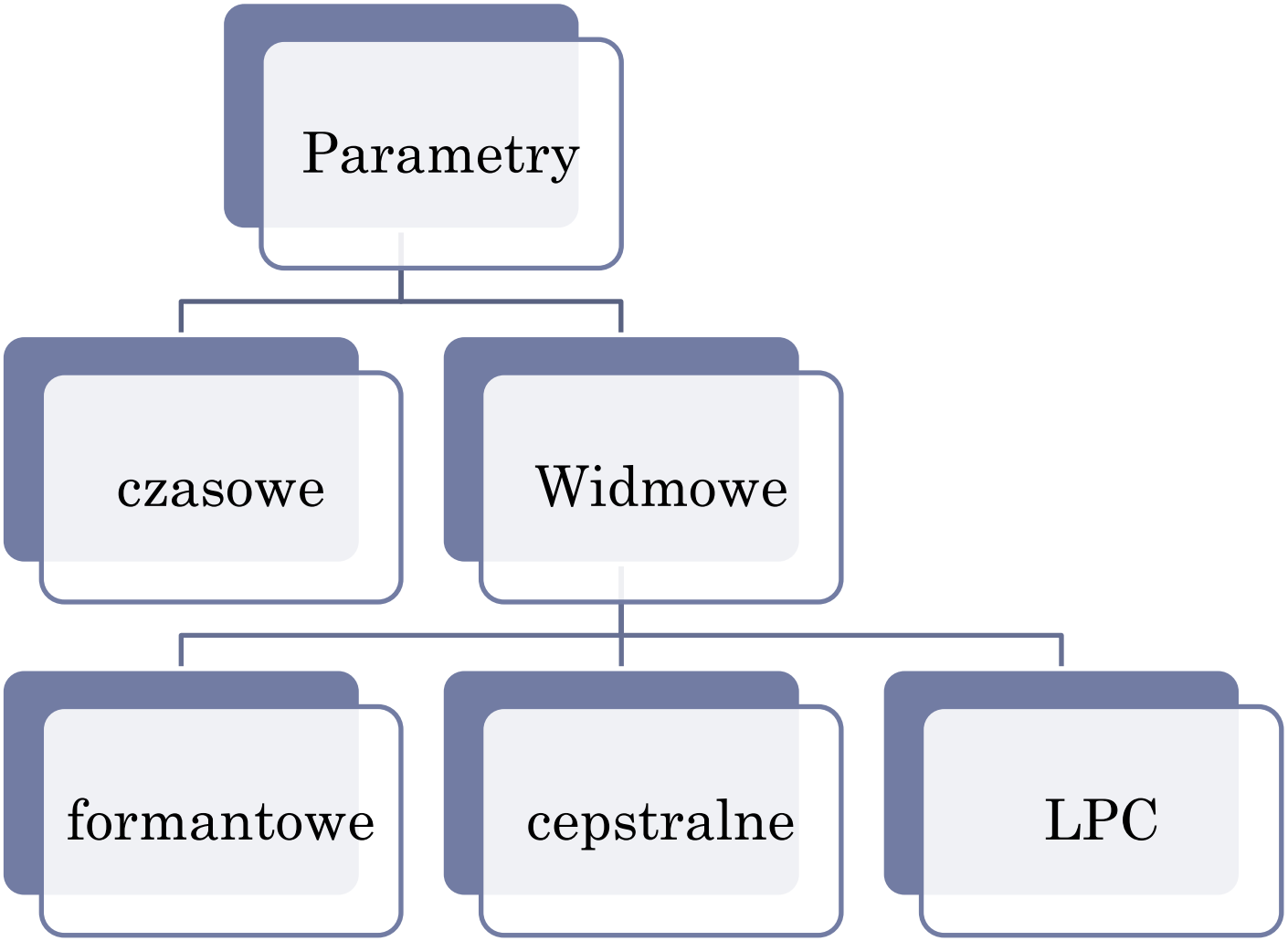
obiekty – nagrane wypowiedzi mówcy

cecha – barwa głosu, wysokość głosu, rozmieszczenie formantów…

parametry - …

Klasyfikacja parametrów

Klasyfikacja parametrów



Parametry czasowe

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
- Jitter – średnia, względna **różnica długości** sąsiednich **okresów podstawowych**
- Shimmer – średnia, względna **różnica amplitudy** sąsiednich **okresów podstawowych**

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.

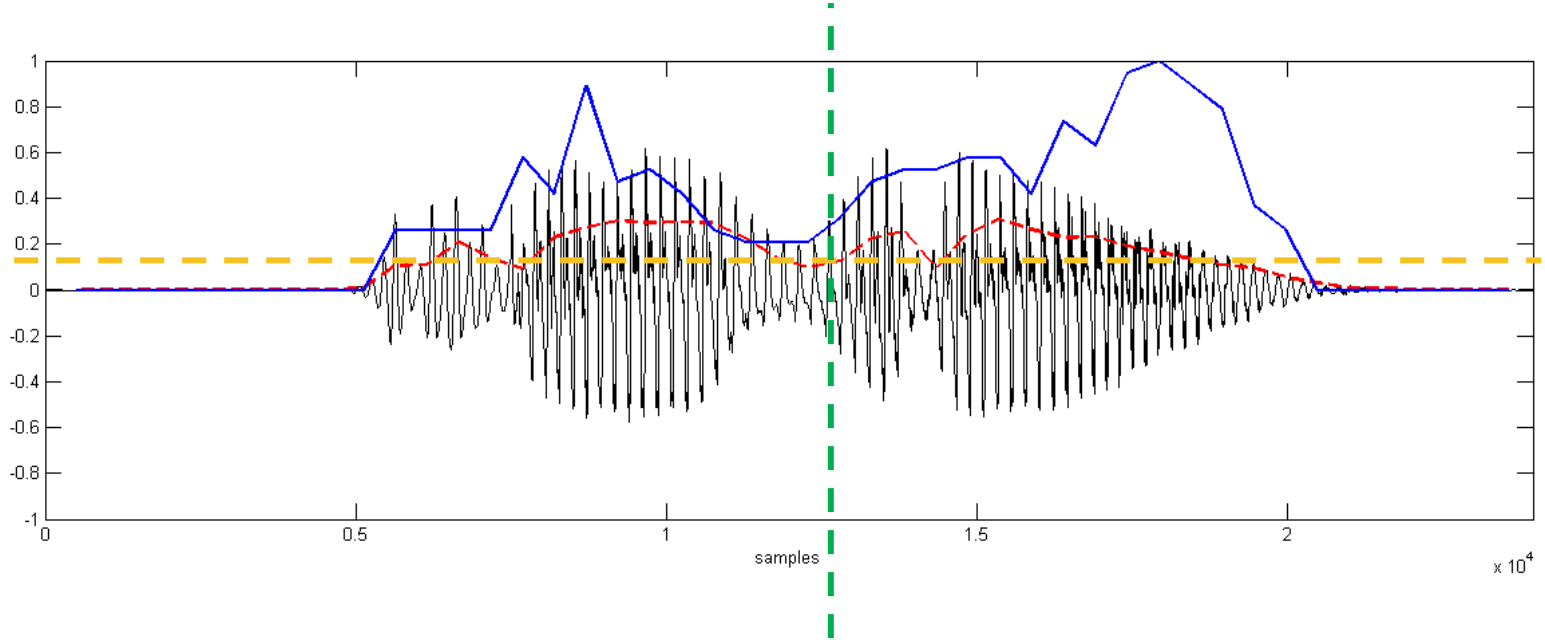
$$ZCD = \rho_0$$

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

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

Parametry widmowe

- Parametry widmowe wyznaczane 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, wykorzystywany do normalizacji momentów wyższych rzędów

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

Momenty widmowe

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

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

- *Unormowany moment pierwszego rzędu odgrywa w analizie mowy znaczącą rolę i można wykazać jego przydatność w zadaniach rozpoznawania – szczególnie głosek szumowych.*
- *Mniej przydatne są momenty-nawet unormowane-wyższych rzędów, gdyż są one w oczywisty sposób skorelowane ze sobą, a ponadto, w odróżnieniu od momentu pierwszego rzędu, nie mają przekonywającej interpretacji.*

Momenty widmowe

- Momenty unormowane **centralne** liczone są względem środka ciężkości widma:

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

- *Przykładowo, użyteczny w analizach jest centralny unormowany moment drugiego rzędu Mcu(2), reprezentujący kwadrat "szerokości" widma.*
- *Użyteczne bywają - obok wspomnianego Mcu(2) także momenty rzędów: 3, 4 i 5, pozwalające opisywać różne typowe deformacje struktury widma.*

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 unormowane 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
x – średnia arytmetyczna dla wszystkich N obserwacji
σ_x – odchylenie standardowe liczone na podstawie obserwacji

Momenty widmowe

- Parametry wywodzące się z momentów widmowych znajdują zastosowanie w analizie i rozpoznawaniu mowy.
- Parametry te mogą być wyznaczane dla całego widma, mogą również dotyczyć wydzielonych jego fragmentów.
- Przykładowo parametr *M(0) wyznaczany w odpowiednich pasmach częstotliwości może* być **użyteczny**, na przykład, **do wykrywania różnicy pomiędzy głoskami dźwięcznymi i bezdźwięcznymi** (w głoskach dźwięcznych i tylko w nich występuje obszar koncentracji energii w zakresie niskich częstotliwości (około l00 Hz), co jest związane z występowaniem podstawowej harmonicznej tonu krtaniowego).
- Wiele innych głosek można rozróżniać, badając stosunki zawartości energii w wybranych pasmach widma - a do tego nadaje się doskonale parametr M(0).

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

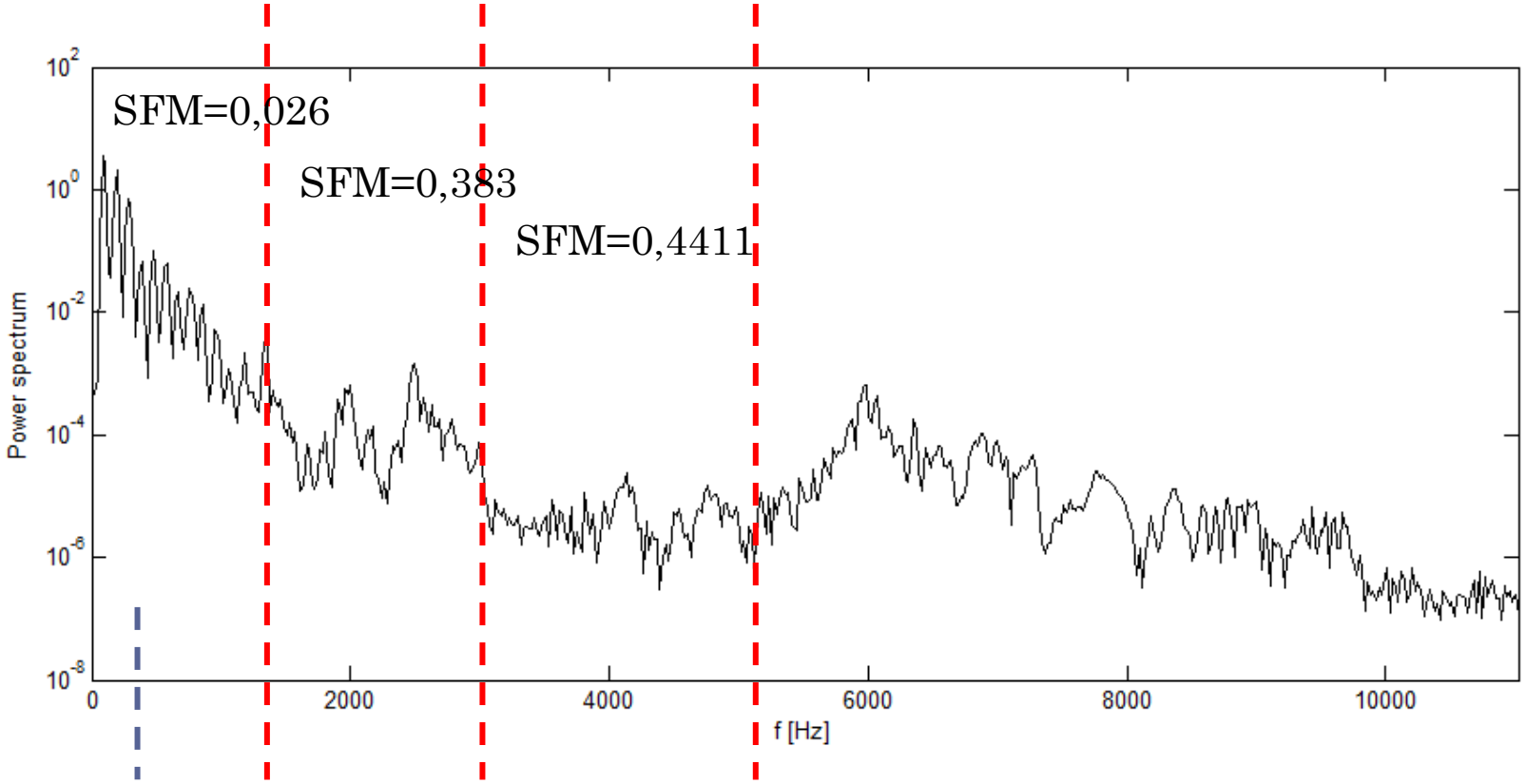
Parametr SFM może być również wyznaczany w pasmach

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



Mu1=178,34 - środek ciężkości widma

Muc3 = 1,8 * 10^8 – skośność

Kurtoza widma – 506.4653

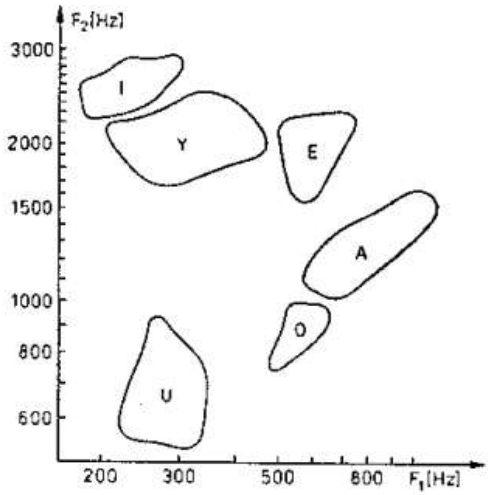
Parametry formantowe

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

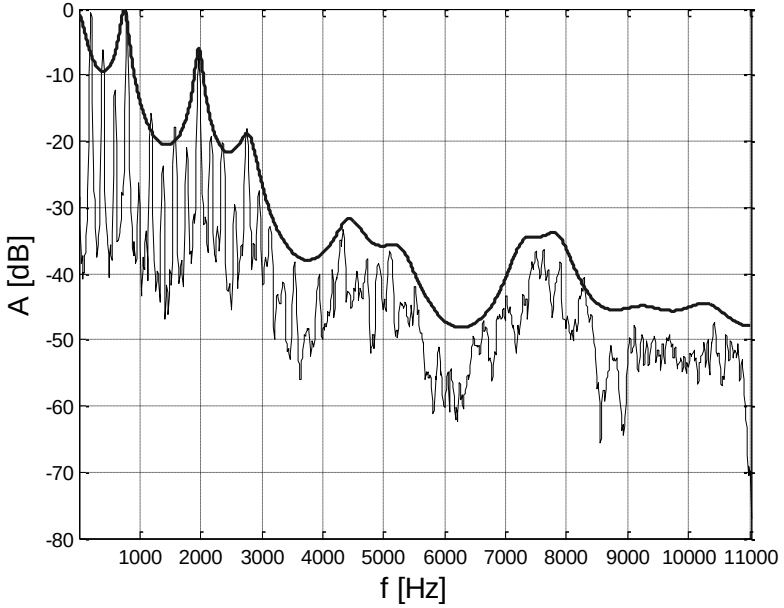


Wyznaczanie parametrów formantowych

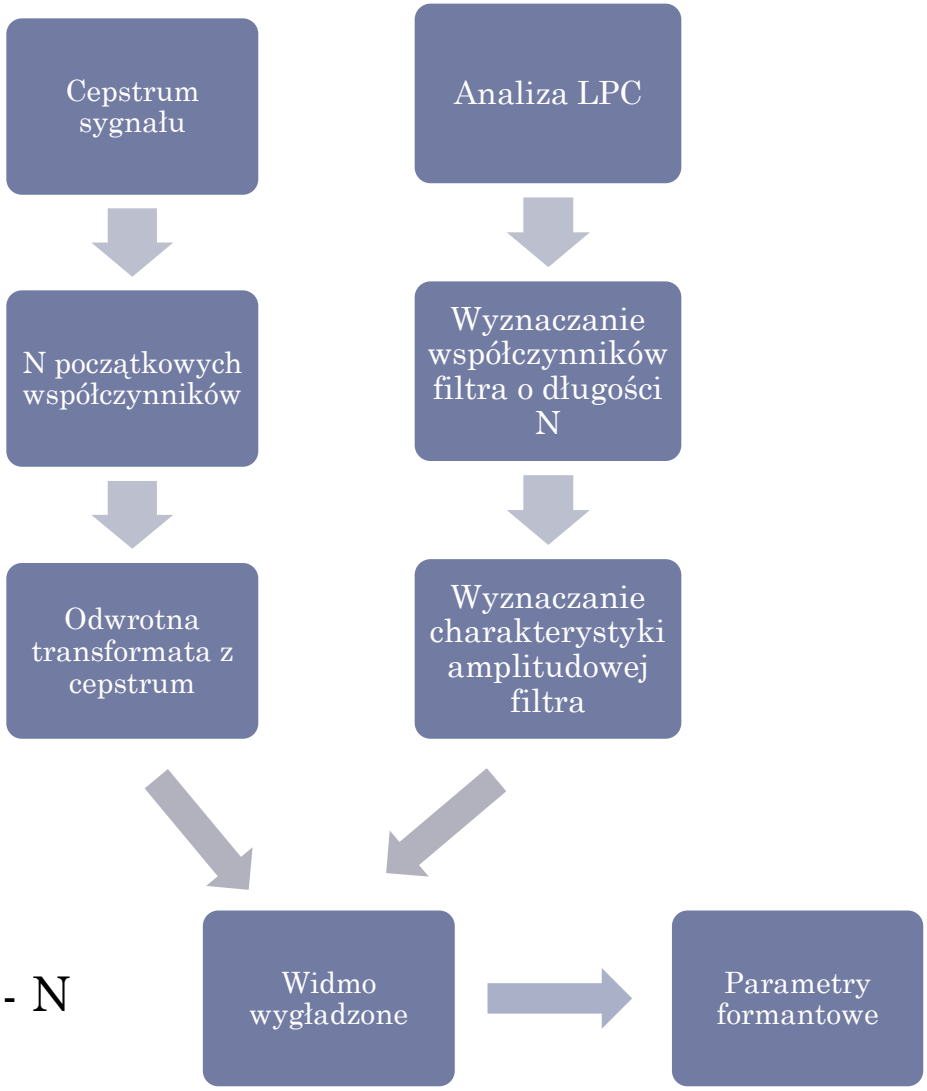
- Do wyznaczania parametrów formantowych potrzebne jest tzw. wygładzone widmo sygnału (obwiednia widmowa – *spectral envelope*).

  Metody wyznaczania widma wygładzonego:

- Cepstralna
- LPC



Wyznaczanie parametrów formantowych



Rząd wygładzania - N

Parametry cepstralne

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)

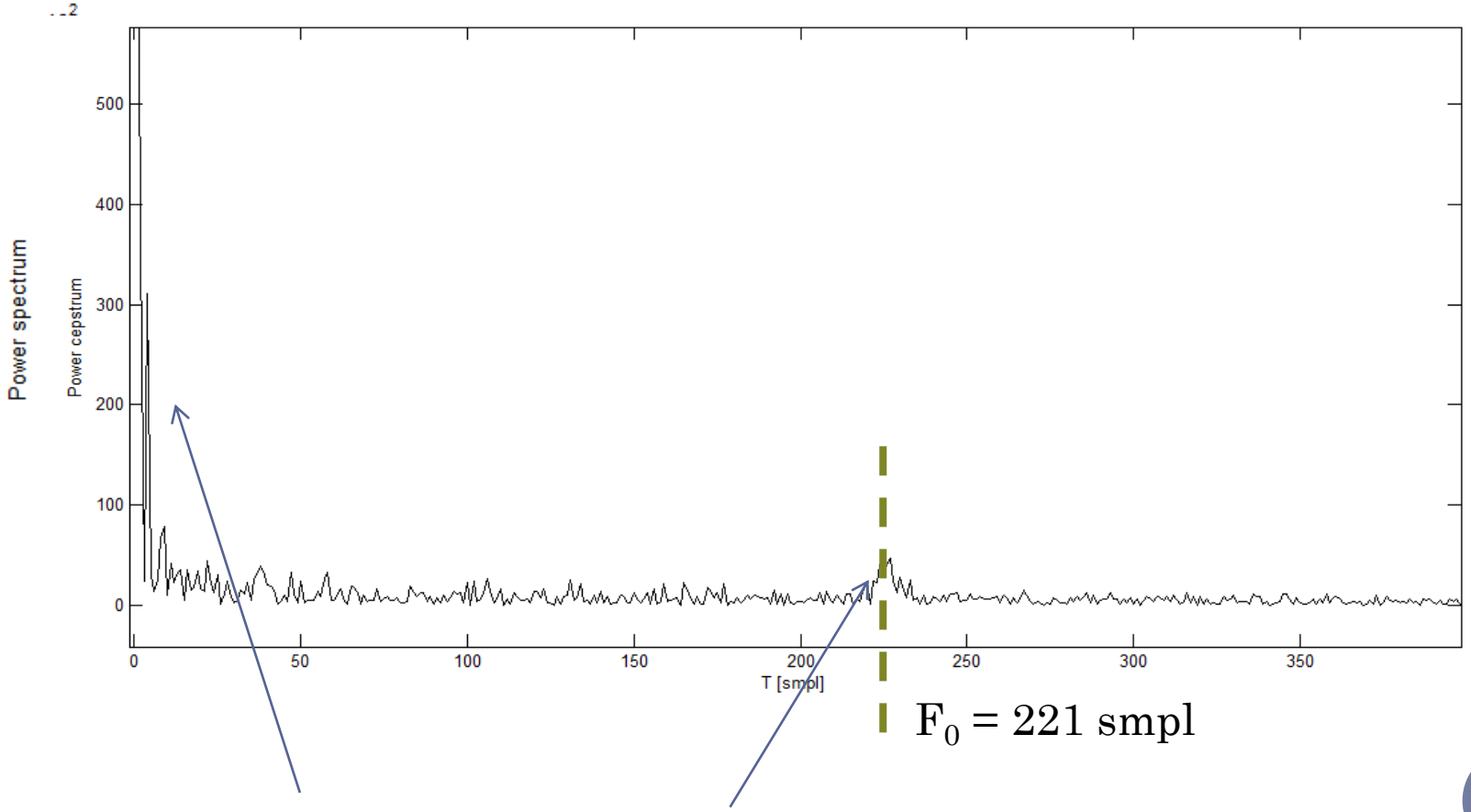
$$\hat{X}_c(k) = \sum_{n=0}^{N-1} [\ln|X(n)|] \cdot \cos\left(\frac{(n-0.5)\cdot k \cdot \pi}{N}\right)$$

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 cepstralne

- Po dokonaniu operacji odwrotnych do używanych przy tworzeniu cepstrum otrzymuje się sygnał o gładkiej obwiedni widma, odpowiadającej ruchom artykulacyjnym narządów mowy.
- Taka czynność, nazywana **wygładzaniem cepstralnym**, jest nieocenioną pomocą przy wszelkich dalszych analizach i badaniach sygnału mowy.
- Możliwe jest także wykorzystanie analizy cepstratnej do wydzielenia tonu krtaniowego oraz do tworzenia mowy syntetycznej.
- Często stosuje się ją do usuwania z sygnału różnych form zakłóceń, echa, pogłosu, zniekształceń wnoszonych przez proces rejestracji sygnału (tą metodą "czyści" się archiwalne nagrania o dużej wartości historycznej łub artystycznej).

Parametry LPC

Liniowe kodowanie predykcyjne

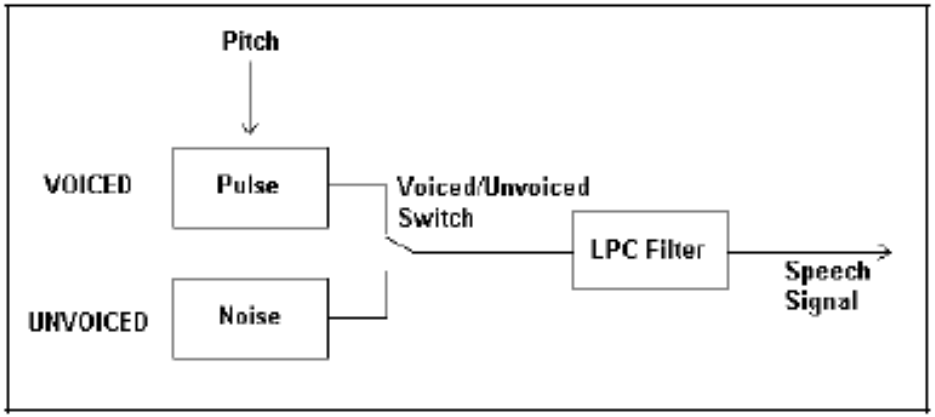
Liniowe kodowanie predykcyjne (ang. *Linear Predictive Coding* – LPC)

W kontekście analizy mowy uzyskuje się za jej pomocą bardzo zróżnicowane wyniki:

- *od opisu sygnału w formie ułatwiającej jego rozpoznawanie,*
- *przez oszczędny opis, wykorzystywany do kompandowanego przesyłania sygnału przez łącza,*
- *aż do celów badawczych, gdzie za pomocą predykcji liniowej wyznacza się geometryczne parametry tonu głosowego w trakcie procesu artykulacji określonych głosek.*

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 **rząd** filtra (rząd analizy LPC) – *p*.

Liniowe kodowanie predykcyjne

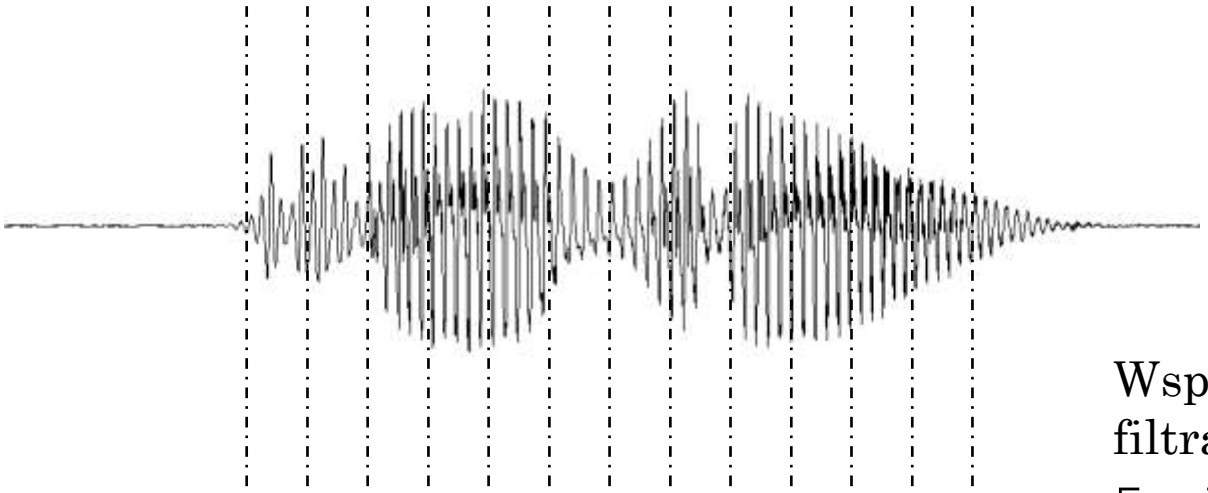
Wyznaczanie współczynników filtra LPC polega na zminimalizowaniu błędu między sygnałem a jego predykcją.

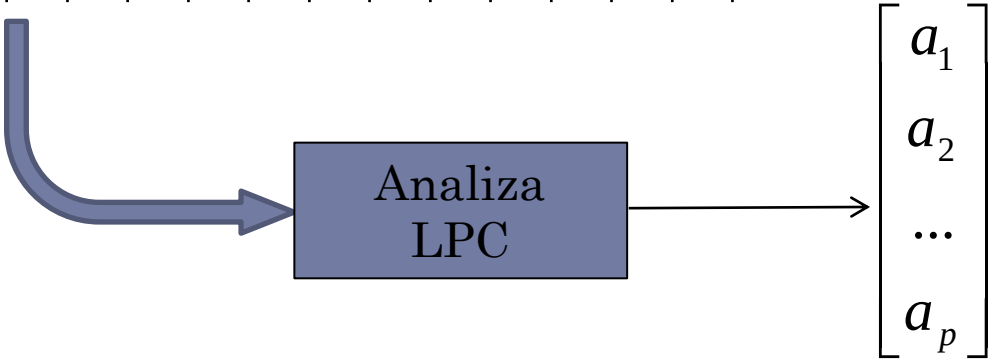
$$E = \sum_{n=1}^{N-1} e^2(n) = \sum_{n=1}^{N-1} \left[v(n) - \sum_{k=1}^{p} a_k \cdot v(n-k) \right]^2$$

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

Wyznaczanie parametrów LPC

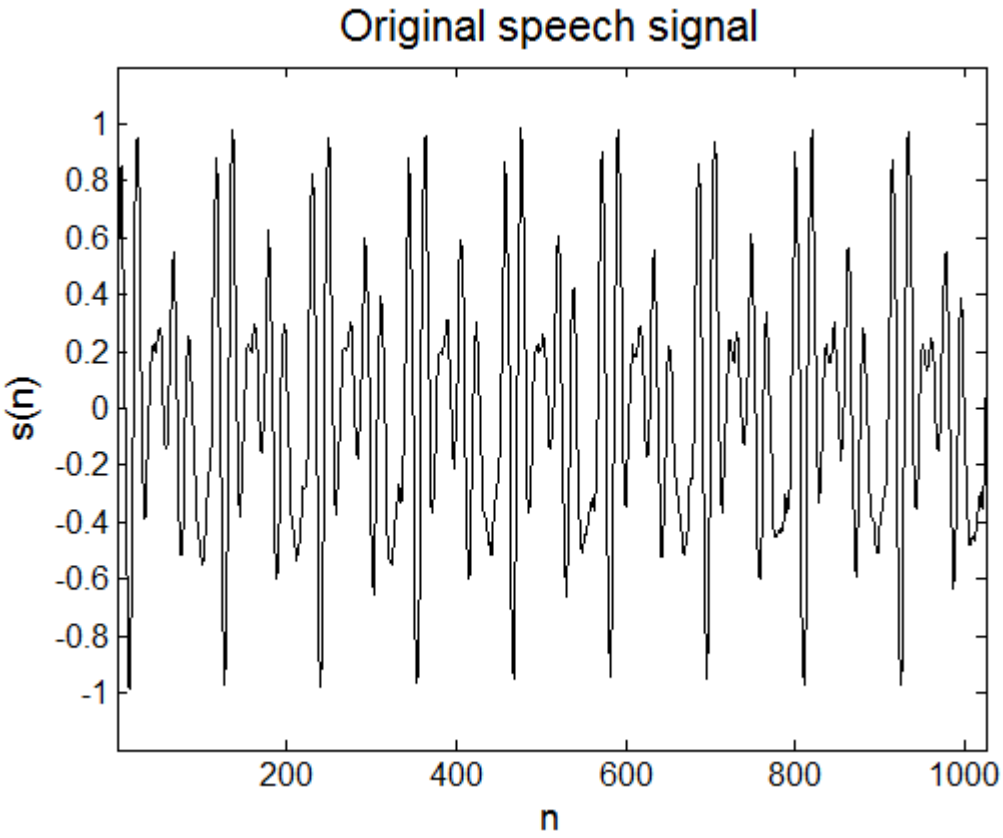
Parametry LPC to współczynniki filtra analizującego sygnał mowy. Wyznacza się je w ramkach, np. 25 ms.





Modelowanie sygnału mowy filtrem LPC

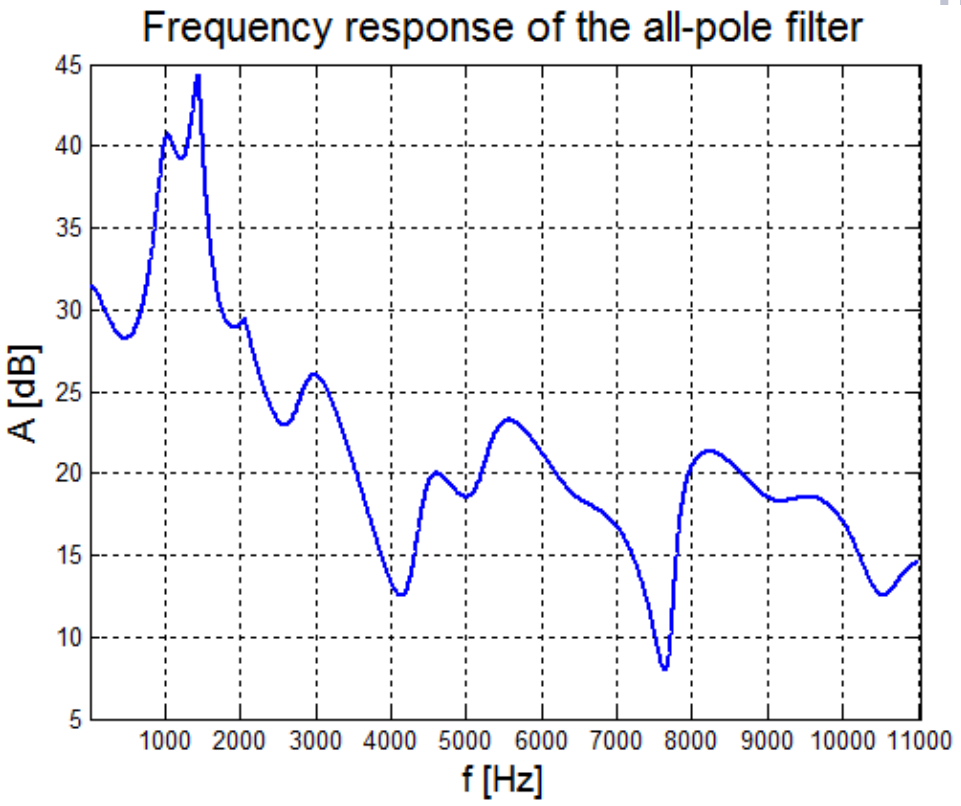
Analizujemy sygnał mowy (głoska *a*)



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
```

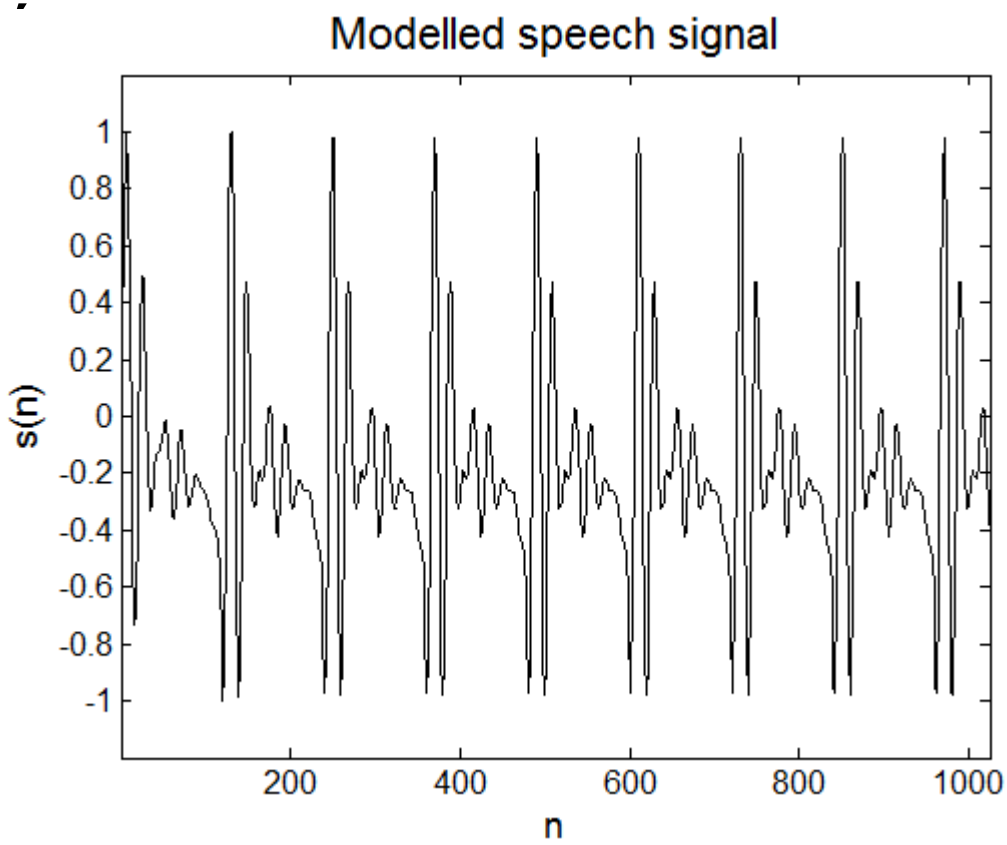
pobudzenie szumem białym

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

Modelowanie sygnału mowy filtrem AR

- Odtwarzanie sygnału na podstawie filtru i pobudzenia

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



Technika LPC - podsumowanie

- W sumie technika predykcji liniowej może w dużym stopniu zastępować wszystkie wcześniej omówione techniki analizy sygnału mowy, gdyż może służyć do analizy widmowej sygnału,
- pozwala wykrywać szczegóły jego obwiedni (na przykład formanty),
- dostarcza parametrów umożliwiających efektywne rozpoznawanie sygnału i jego oszczędne przesyłanie,
- wreszcie stanowi efektywne narzędzie w diagnostyce medycznej narządów mowy.

Perceptualne skale częstotliwości

Percpetualne skale częstotliwości

Prawo Webbera-Fechnera głosi, że

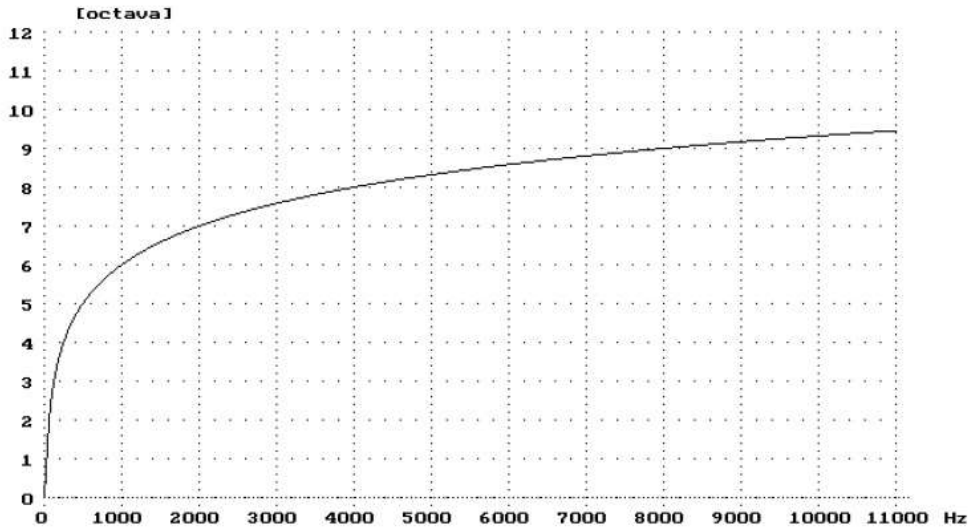
Reakcja układu biologicznego jest proporcjonalna do logarytmu pobudzającego go bodźca.

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 oktawowa. Odpowiada ona strojowi równomiernie temperowanemu.

$$oktawa = \log_2(64 \cdot f)$$



Zależność pomiędzy liniową skalą częstotliwości a skalą oktawową

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 20μPa).

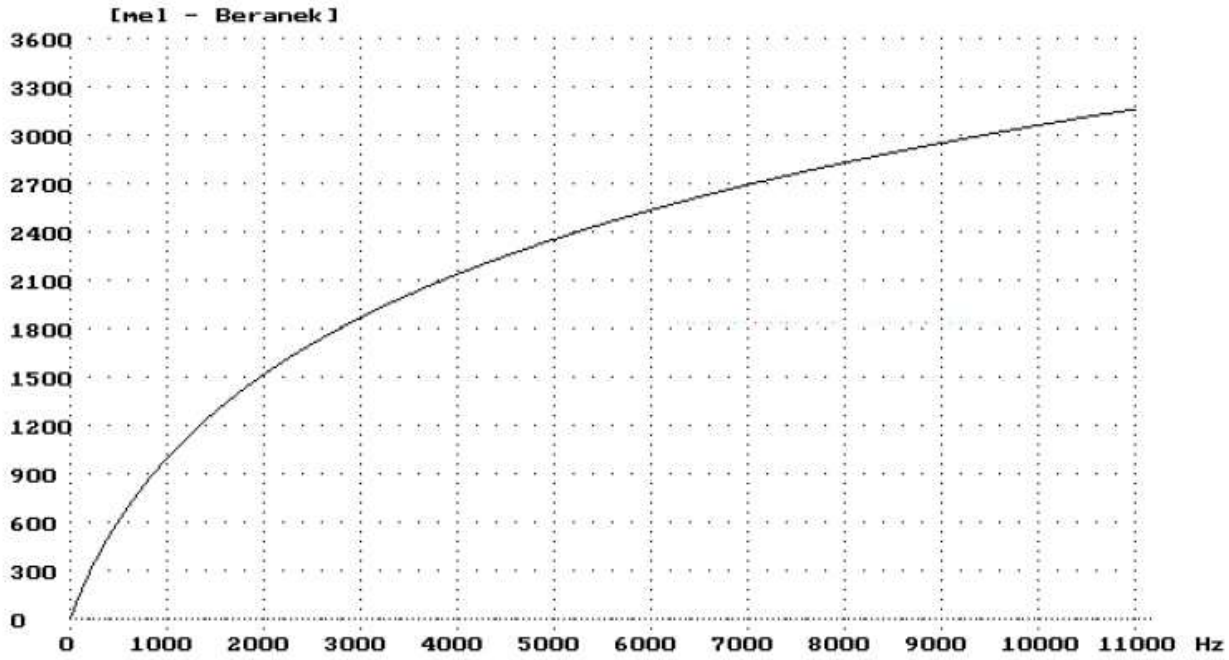
Własności skali mel

- Punktem odniesienia jest ton 1000 Hz o poziomie 40 dB – 1000 meli = wysokość tonu o częstotliwości 1000 Hz
- Dla każdego tonu dobiera się drugi ton o częstotliwości odbieranej subiektywnie jako o dwukrotnie niższej (lub wyższej) wysokości, lub dokonuje się podziału danego zakresu częstotliwości na 4 percepcyjnie jednakowe interwały
- Do 500 Hz skala meli pokrywa się ze skalą częstotliwościową. Powyżej – zależność jest logarytmiczna
- 100 mel = 1 Bark

Skala melowa

Skala melowa wg Beranka (1000 mel = 1000 Hz)

$$M = 1127 \cdot \ln\left(1 + {}^{f}\!/_{0.7}\right)$$



Zależność pomiędzy liniową skalą częstotliwości a skalą melową Beranka

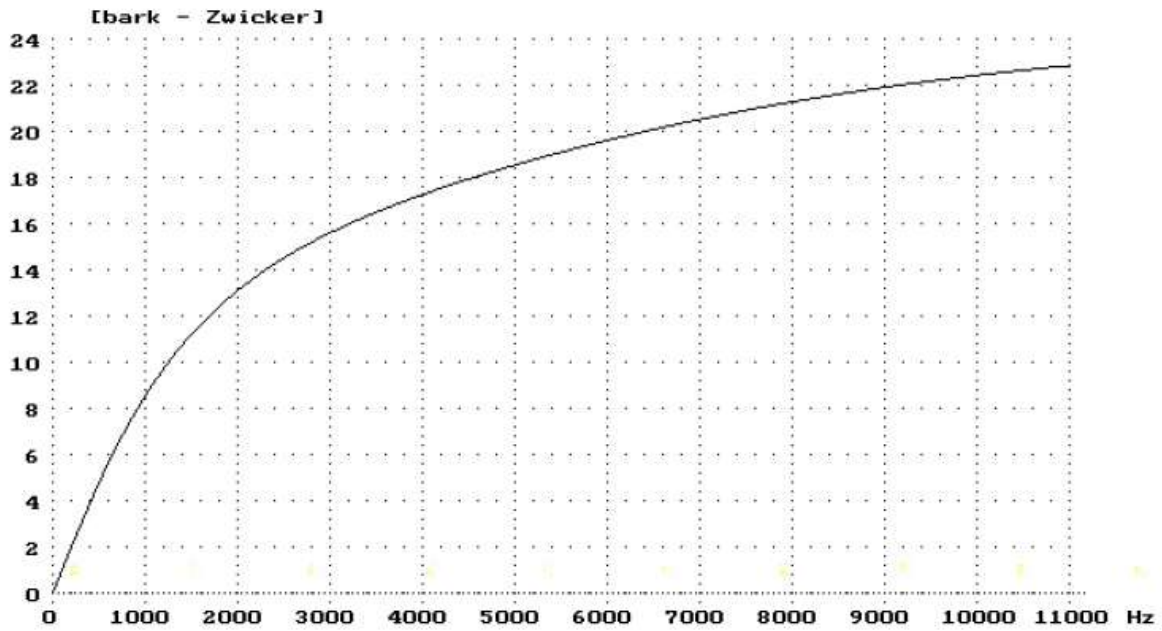
Skala barkowa

- Odpowiada szerokości pasma krytycznego
- Pasmo 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

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



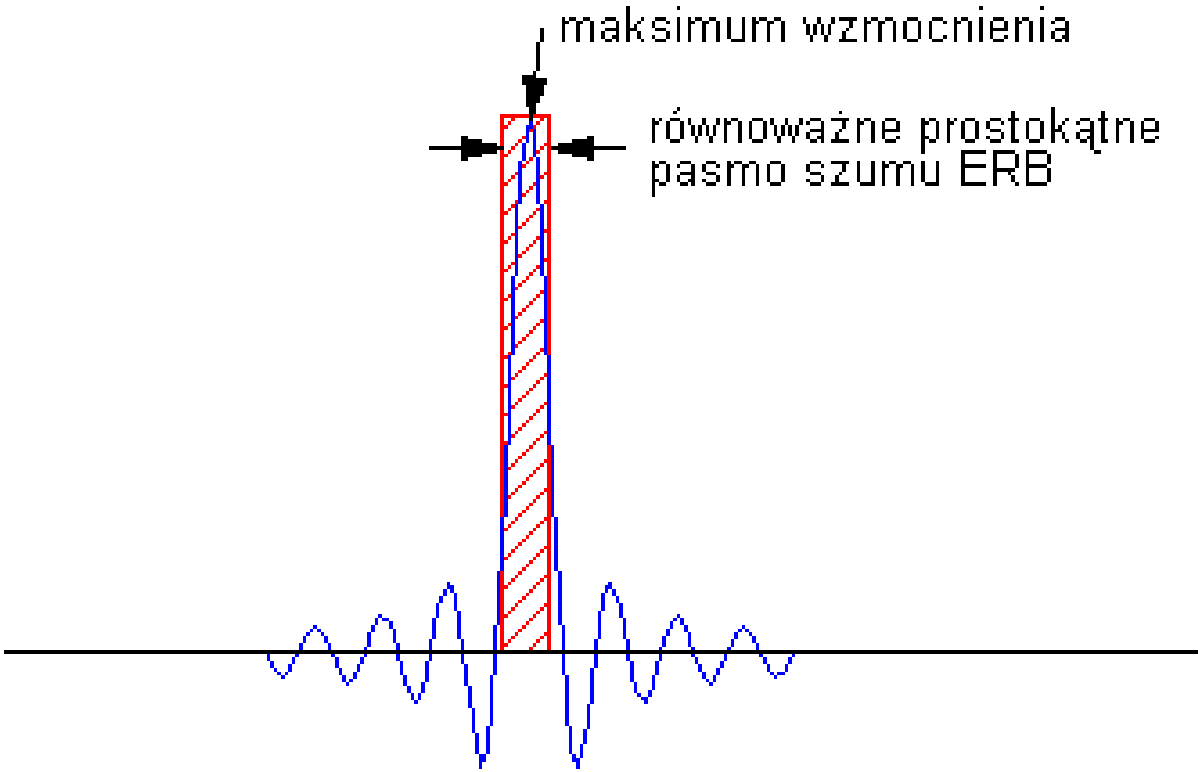
Zależność pomiędzy liniową skalą częstotliwości a skalą barkową Zwickera

Dane filtrów w skali Bark

Nr pasma [bark]	Dolna częstotliwość [Hz]	Górna częstotliwość [Hz]	Szerokość pasma [Hz]
1	0	100	100
2	100	200	100
3	200	300	100
4	300	400	100
5	400	510	110
6	510	630	120
7	630	770	140
8	770	920	150
9	920	1080	160
10	1080	1270	190
11	1270	1480	210
12	1480	1720	240
13	1720	2000	280
14	2000	2320	320
15	2320	2700	380
16	2700	3150	450
17	3150	3700	550
18	3700	4400	700
19	4400	5300	900
20	5300	6400	1100
21	6400	7700	1300
22	7700	9500	1800
23	9500	12000	2500
24	12000	15500	3500

Definicja pasma ERB

ERB – equivalent rectangular bandwith
jest szerokością filtru prostokątnego przepuszczającego szum o tej samej mocy i tej samej mocy szczytowej, co filtr modelowany



Własności skali ERB

- Skala ERB jest wyrażana w Hz
- Zakres 16 000 Hz dzieli się na 40 pasm
- Szerokość pasma również zależy od częstotliwości środkowej

Trzy percepcyjne skale częstotliwości

- Skala Bark:

$$Bark(f) = \begin{cases} .01f, & 0 \le f < 500 \\ .007f + 1.5, & 500 \le f < 1220 \\ 6\ln(f) - 32.6, & 1220 \le f \end{cases}$$

- Skala Mel :

$$Mel(f) = 2595 \log_{10}(1 + \frac{f}{700})$$

- Skala ERB :

$$ERB(f) = 24.7(4.37f + 1)$$

Wrażenie słuchowe wysokości dźwięku

- Na wrażenie wysokości dźwięku wpływa:
  - Częstotliwość dźwięku prostego – ma wpływ zasadniczy
  - Poziom ciśnienia akustycznego - ma wpływ nieznaczny, widoczny tylko dla skrajnych częstotliwości.
  - Czas trwania - ma wpływ na pojawianie się wrażenia wysokości i utrwalenie się tego wrażenia.

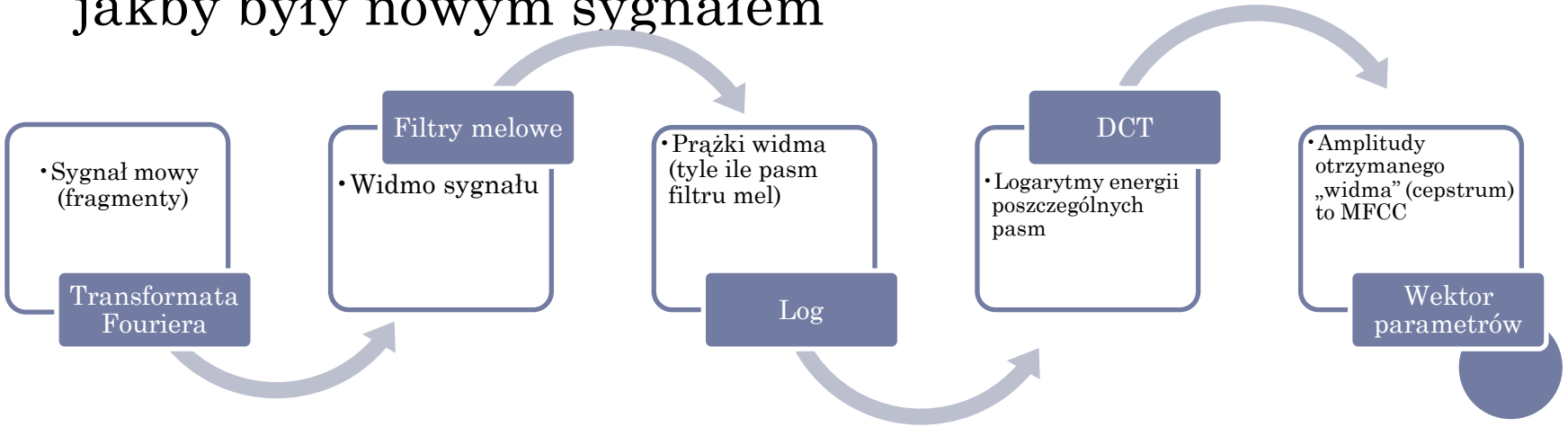
Parametry w skalach perceptualnych

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).

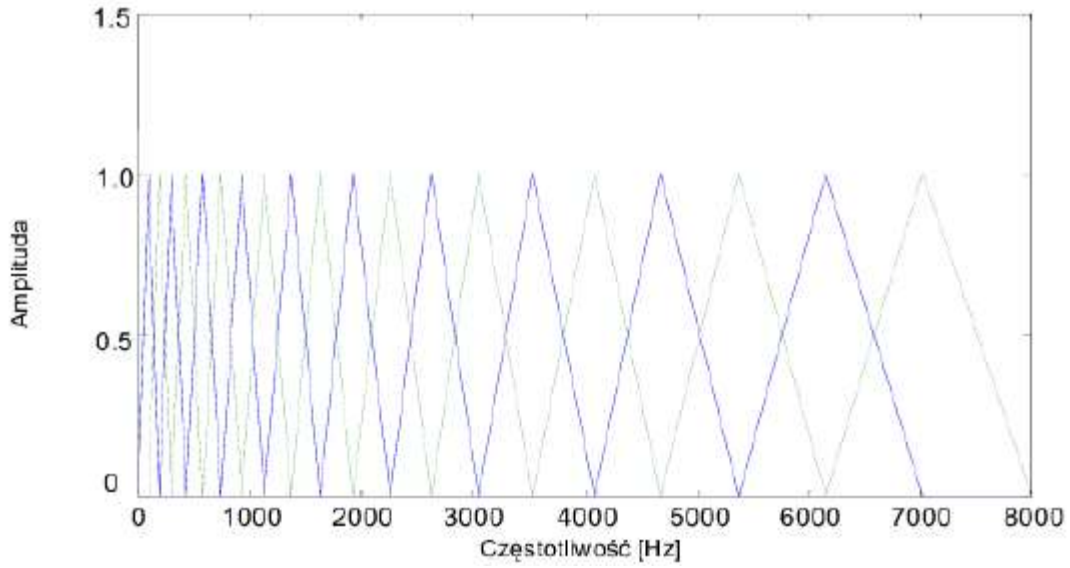
Parametry mel-cepstralne

- Sygnał dzielony na okna wg zadanego algorytmu (fonemy, głoski, wg energii itd.)
- Jeden filtr melowy – jeden prążek – jeden współczynnik – jeden parametr mel-cepstralny
- Transformata kosinusowa logarytmów współczynników uzyskanych z filtracji sygnału, tak jakby były nowym sygnałem



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

Dziękuję za uwagę!