

Akustyka mowy

Analiza widmowa

Plan wykładu

- Wykład obejmuje przedstawienie podstawowych zagadnień związanych z analizą widmową sygnału mowy.
- W pierwszej części wykładu zostaną przedstawione **techniki dekompozycji sygnału mowy** na komponenty widmowe. Zostaną przedstawione parametry analizy widmowej szczególnie użyteczne w analizie sygnału mowy.
- W drugiej części wykładu zostaną przedstawione **techniki obrazowania sygnału mowy** za pomocą uprzednio przedstawionych technik.
- Trzecia część wykładu będzie miała charakter warsztatów. Zostaną przedstawione **narzędzia do analizy widmowej sygnału mowy** oraz możliwości ich zastosowania na wybranych przykładach.

Wprowadzenie

- Opis sygnału w dziedzinie częstotliwości jest podstawową, rutynowo stosowaną i w istocie najbardziej przydatną formą jego opisu
- Użyteczność widmowej prezentacji sygnału – w szczególności w odniesieniu do sygnału mowy – wynika z kilku faktów
 - w procesie artykulacji mowy kształtowana jest głównie obwiednia amplitudowo-częstotliwościowa sygnału (poprzez odpowiednio formowaną strukturę rezonansową traktu głosowego)
 - W procesie percepcji następuje etap wydzielenia składowych o poszczególnych częstotliwościach
- Analiza widmowa może być realizowana wieloma metodami i może służyć do różnych celów

Wprowadzenie

- Teoria Fouriera mówi nam, że każdy sygnał opisany w dziedzinie czasu może zostać opisany za pomocą jednego lub więcej sygnałów sinusoidalnych z odpowiednią amplitudą, fazą i częstotliwością.
- Oznacza to, że każdy sygnał przedstawiony w dziedzinie czasu możemy przedstawić również w dziedzinie częstotliwości.
- W kontekście naszych rozważań **widmo sygnału** to zbiór sinusów które odpowiednio połączone dadzą ten sygnał w dziedzinie czasu.
- Teoretycznie aby przetransformować sygnał z dziedziny czasu do dziedziny częstotliwości przebieg sygnału musi być znany w całym czasie czyli nieskończoności. Jednak w praktyce do analizy wykorzystujemy skończony fragment sygnału.

Co nam daje analiza widmowa

- W odniesieniu do wielu sygnałów, w tym do sygnału mowy, prawdziwe jest twierdzenie, że **świadomie kształtowane składowe sygnału mieszczą się głównie w jego amplitudowo-częstotliwościowej charakterystyce**, podczas gdy **wpływ czynników losowych determinuje w pierwszym rzędzie strukturę charakterystyk fazowo-częstotliwościowych**
- Po dokonaniu analizy widmowej rozdzielenie wymienionych składowników staje się bardzo proste
- Podobnie (choć przy użyciu nieco bardziej złożonych metod) **analiza częstotliwościowa pozwala rozróżnić te własności sygnału, za które odpowiedzialne jest źródło tonu, od tych, które są wynikiem procesu modulacji sygnału w narządach mowy**

Metody dekompozycji sygnału

- Analiza Fouriera
- Analiza za pomocą filtrów o procentowo stałej szerokości
- Analiza Falkowa
-

Transformata Fouriera:

$F(\omega)$ jest innym sposobem patrzenia na funkcje lub fale:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-i\omega t) dt$$

$\omega = \frac{2\pi}{T} = 2\pi f$ jest częstością kołową



Jean Baptiste Joseph Fourier
(1768 - 1830),
francuski matematyk

$F(\omega)$ jest transformatą fourierowską danej funkcji $f(t)$. **Zawiera ona tę samą informację, co funkcja $f(t)$.**

Mówimy, że $f(t)$ żyje w **d dziedzinie czasu** (przestrzeni), a $F(\omega)$ żyje w **d dziedzinie częstości**.

Transformata Fouriera zamienia charakterystyki czasowe lub przestrzenne zjawisk, na ich charakterystyki częstotliwościowe.

Odwrotna transformata Fouriera

Transformata Fouriera pozwala nam przejść od $f(t)$ do $F(\omega)$:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-i\omega t) dt$$

$\omega = \frac{2\pi}{T} = 2\pi f$ jest częstością kołową

A co z przejściem w drugą stronę?

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \exp(i\omega t) d\omega$$

Dyskretna transformata Fouriera

Przypuśćmy, że mamy N kolejnych wartości zmierzonych w odstępach czasu Δ , tak że

$$h_k \equiv h(t_k), \quad t_k = k\Delta, \quad k = 0, 1, 2, \dots, N - 1$$

Zamiast próbować znaleźć transformatę dla wszystkich wartości f oszacujmy ją jedynie w konkretnych punktach, danych przez:

$$f_n \equiv \frac{n}{N\Delta}, \quad n = -\frac{N}{2}, \dots, \frac{N}{2}$$

Po przybliżeniu całki otrzymujemy $H_n \equiv \sum_{k=0}^{N-1} h_k e^{2\pi i k n / N}$

Zastosowane powyżej przekształcenie nosi nazwę **dyskretnej transformaty Fouriera**

Algorytm FFT

Idea

Sama idea algorytmu opiera się na tzw. lemacie **Danielsona-Lanczosa**. Odkryli oni, że pojedyncza DFT o długości N , jest równoważna sumie dwóch transformacji o długości $N/2$, jedna z nich jest złożona z nieparzystych punktów spośród oryginalnych N , a druga z parzystych.

$$\begin{aligned} H_n &= \sum_{j=0}^{N/2-1} e^{2\pi i n j / (N/2)} f_{2j} + W^n \sum_{j=0}^{N/2-1} e^{2\pi i n j / (N/2)} f_{2j+1} = \\ &= H_n^e + W^n H_n^o \end{aligned}$$

H_n^e oznacza n -ty składnik transformacji o długości $N/2$, stworzony z parzystych (even) punktów, a H_n^o odpowiednio z nieparzystych (odd).

Algorytm FFT

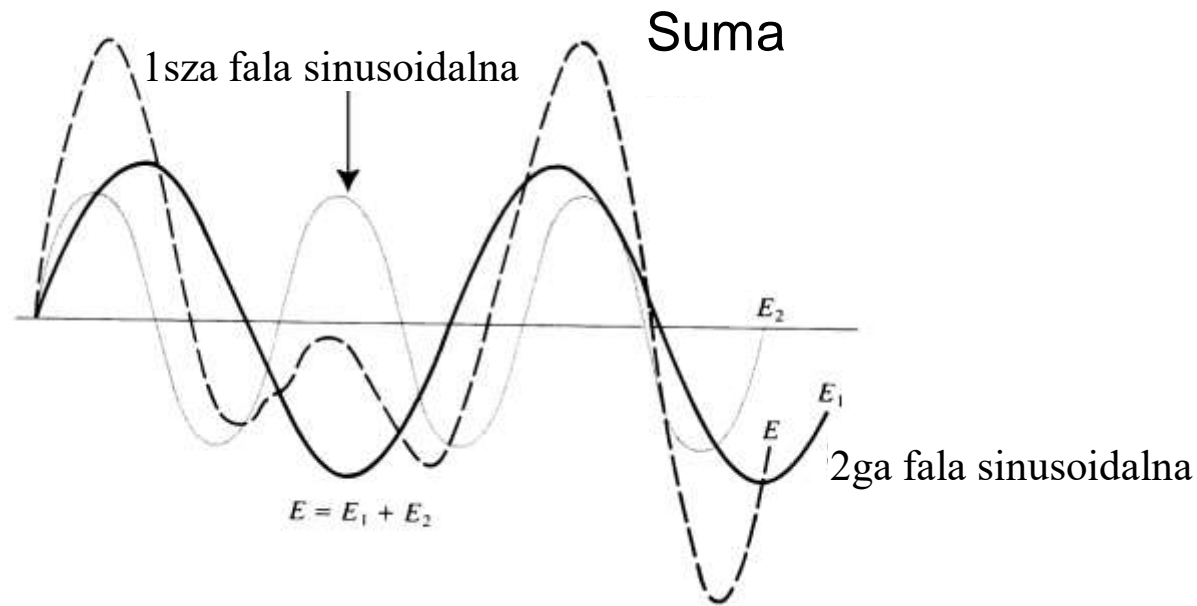
Algorytm Cooley'a-Tukey'a

Przykład wyznaczania transformaty dla $N = 8$ punktów							
$a_0 \ a_1 \ a_2 \ a_3 \ a_4 \ a_5 \ a_6 \ a_7$							
$a_0 \ a_2 \ a_4 \ a_6$				$a_1 \ a_3 \ a_5 \ a_7$			
$a_0 \ a_4$		$a_2 \ a_6$		$a_1 \ a_5$		$a_3 \ a_7$	
000 = 0	100 = 4	010 = 2	110 = 6	001 = 1	101 = 5	011 = 3	111 = 7
a_0	a_4	a_2	a_6	a_1	a_5	a_3	a_7
000 = 0	001 = 1	010 = 2	011 = 3	100 = 4	101 = 5	110 = 6	111 = 7
$b_0 = a_0 + a_4$		$b_2 = a_2 + a_6$		$b_4 = a_1 + a_5$		$b_6 = a_3 + a_7$	
$b_1 = a_0 - a_4$		$b_3 = a_2 - a_6$		$b_5 = a_1 - a_5$		$b_7 = a_3 - a_7$	
$c_0 = b_0 + b_2$		$c_2 = b_0 - b_2$		$c_4 = b_4 + b_6$		$c_6 = b_4 - b_6$	
$c_1 = b_1 + \omega_4 b_3$		$c_3 = b_1 - \omega_4 b_3$		$c_5 = b_5 + \omega_4 b_7$		$c_7 = b_5 - \omega_4 b_7$	
$d_0 = c_0 + c_4$				$d_4 = c_0 - c_4$			
$d_1 = c_1 + \omega_8 c_5$				$d_5 = c_1 - \omega_8 c_5$			
$d_2 = c_2 + \omega_8^2 c_6$				$d_6 = c_2 - \omega_8^2 c_6$			
$d_3 = c_3 + \omega_8^3 c_7$				$d_7 = c_3 - \omega_8^3 c_7$			

$$W_N^{k - \frac{N}{2}} = e^{i2\pi(k - \frac{N}{2})/N} = e^{i2\pi k/N} e^{-i\pi} = -e^{i2\pi k/N} = -W_N^k$$

Przykład: Fale anharmoniczne będące sumami oscylacji sinusoidalnych:

Rozważmy sumę 2 fal sinusoidalnych (to jest harmonicznnych) o różnych częstościach:

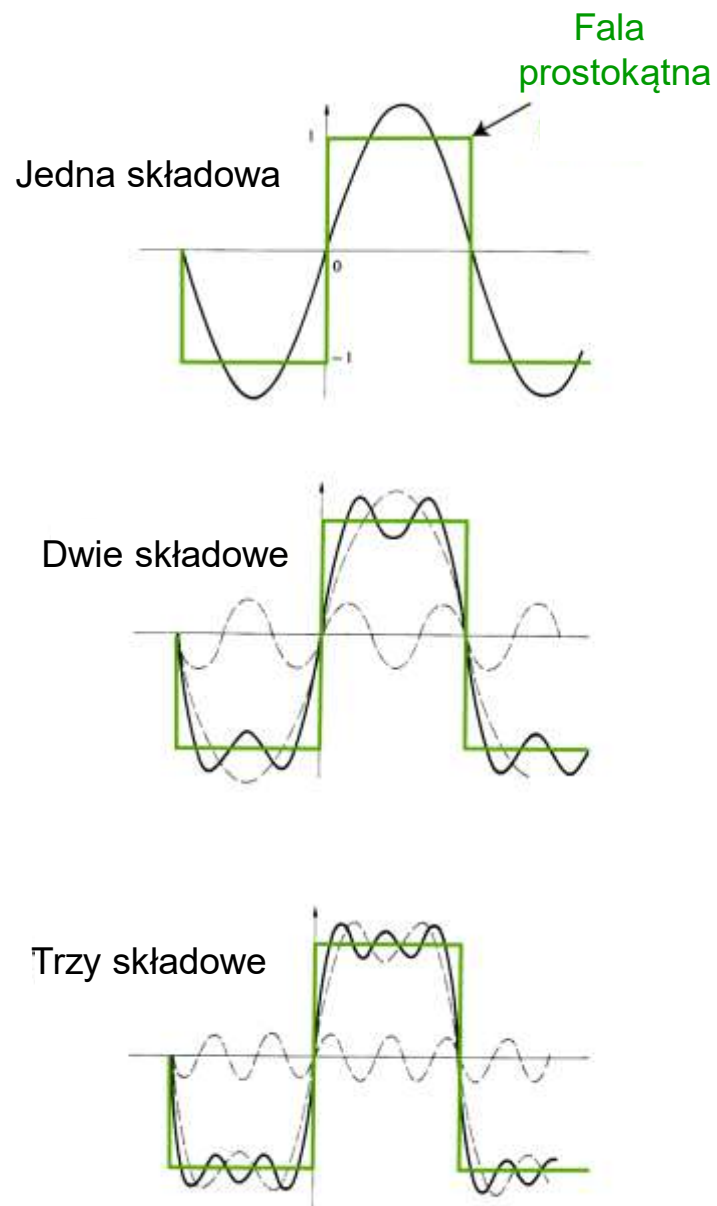


Fala będąca ich sumą jest okresowa, ale nie harmoniczna.
Większość fal to fale nieharmoniczne.

Fourierowska reprezentacja fali prostokątnej:

Falę prostokątną zapisać można jako sumę funkcji harmonicznycych.

Im więcej składowych harmonicznycych jest sumowanych, tym lepsze jest przybliżenie przebiegu prostokątnego.

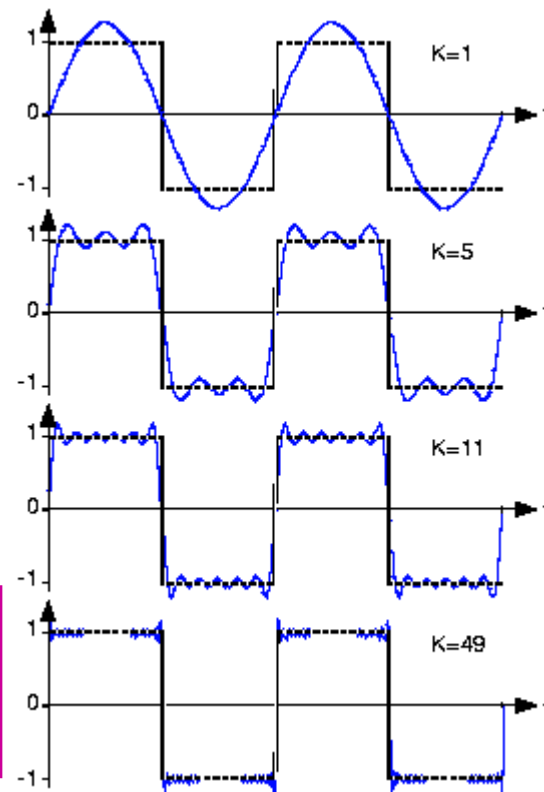


Fourierowska reprezentacja fali prostokątnej:

Przybliżenie funkcji $\text{rect}(t)$ szeregiem
Fouriera.

K – liczba członów uwzględnionych w sumie

$$f(t) = \frac{1}{\pi} \sum_{m=0}^{\infty} F_m \cos(mt) + \frac{1}{\pi} \sum_{m=0}^{\infty} F'_m \sin(mt)$$

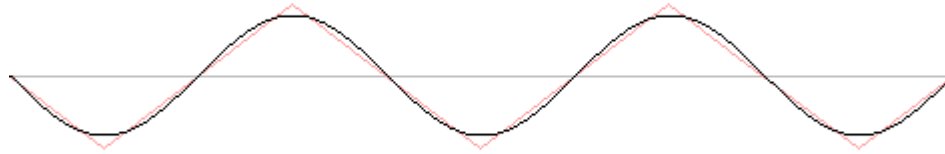


Wniosek:

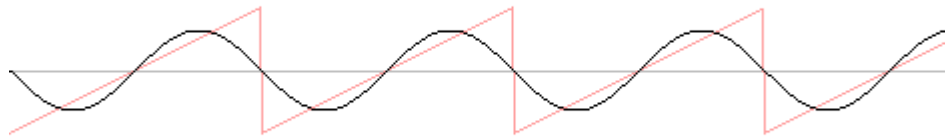
Sygnaly (w tym sygnał mowy) można otrzymać jako superpozycję nieskończonej liczby funkcji \sin i \cos . Współczynniki rozwinięcia zależą od charakteru funkcji, którą chcemy przedstawić. Ich znajomość jednoznacznie określa funkcję, jako taką.

Fourierowska reprezentacja: fali prostokątnej:

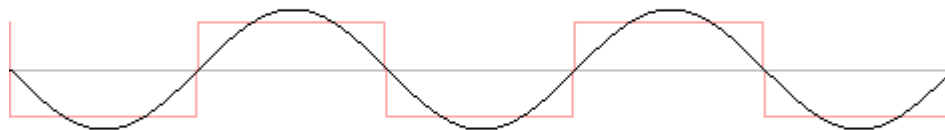
harmonics: 1



harmonics: 1



harmonics: 1



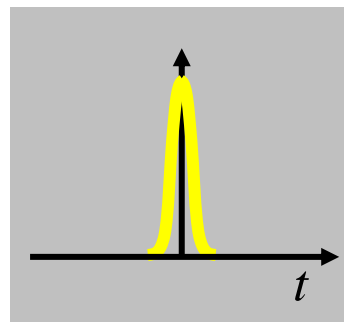
Transformata Fouriera funkcji skalującej: $f(at)$

$$F\{f(at)\} = F(\omega/a) / |a|$$

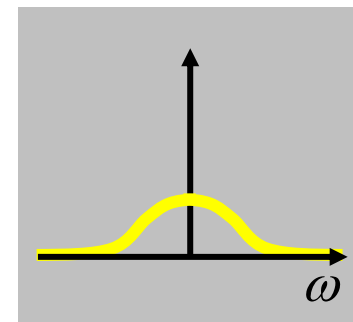
Im krótszy impuls,
tym szersze spektrum!

- jest to w istocie zasada nieoznaczoności

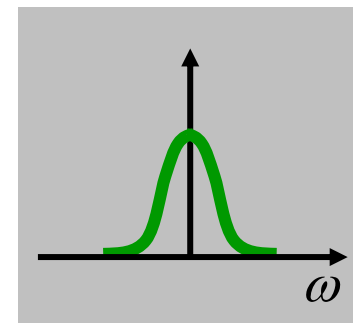
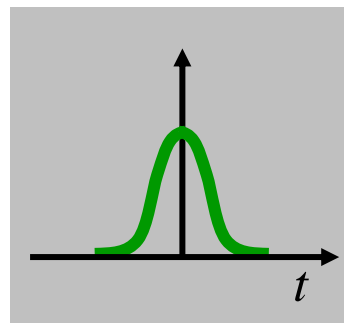
Krótki impuls



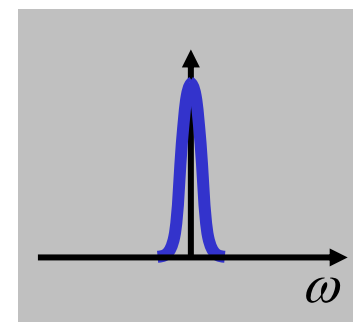
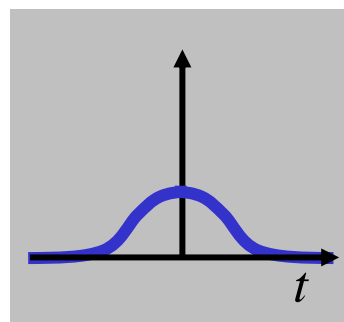
$F(\omega)$



Impuls średniej długości



Długi impuls



Impulsy długie a krótkie:

Relacja nieoznaczoności: iloczyn szerokości czasowej i spektralnej impulsu:

$$\Delta\omega \Delta t \geq 2\pi$$

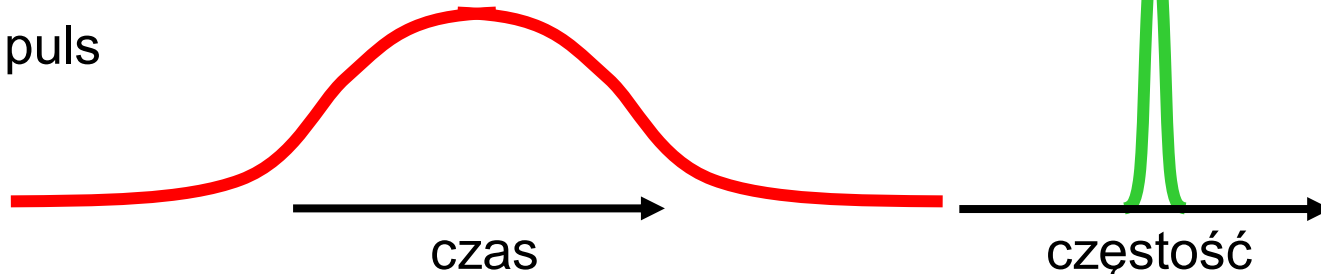
lub:

$$\Delta\nu \Delta t \geq 1$$

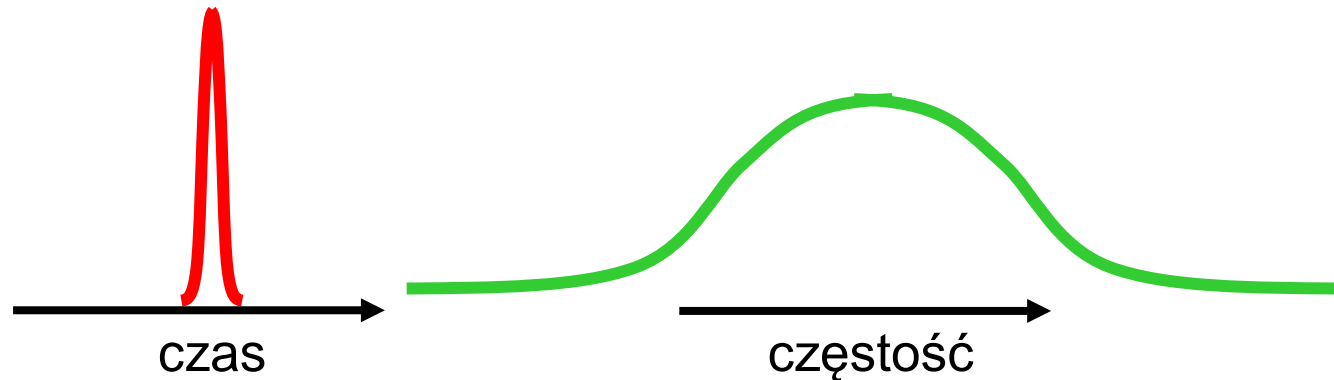
Natężenie vs. czas

Widmo

Długi impuls

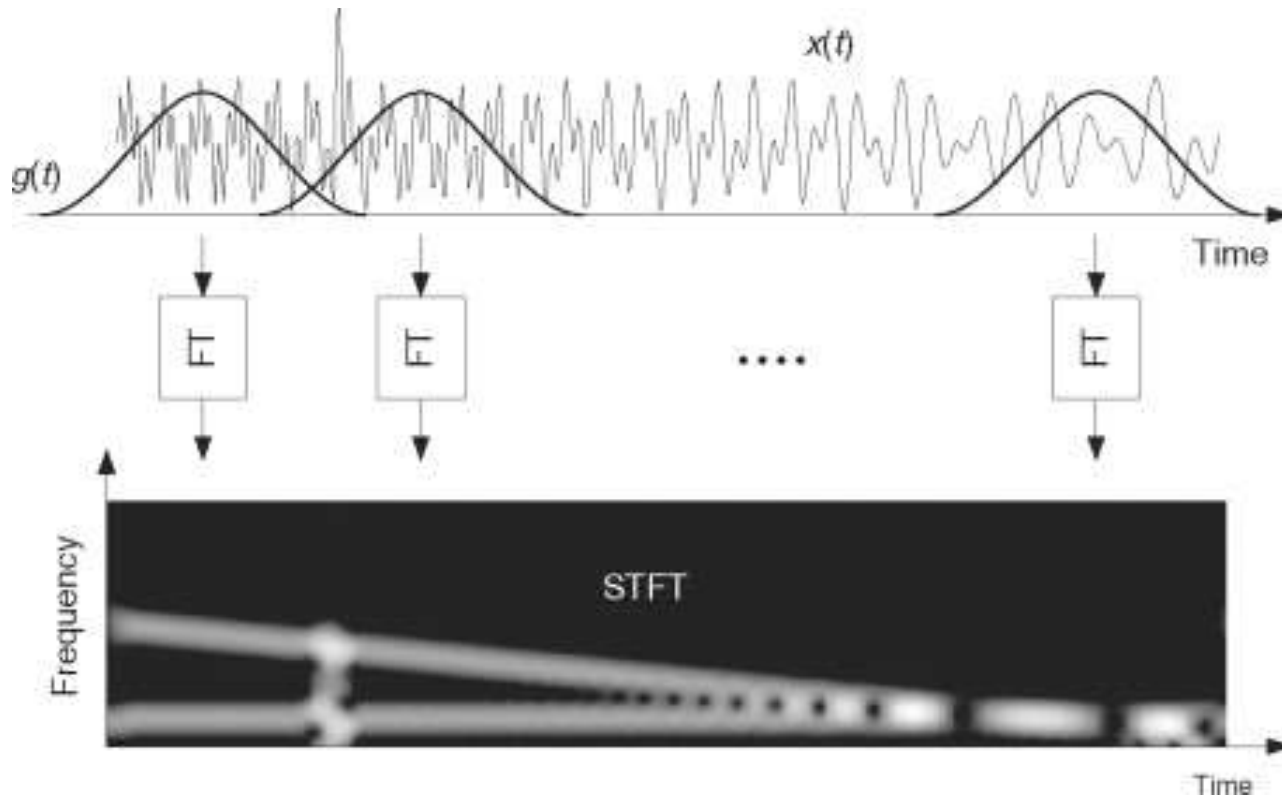


Krótki impuls



STFT

- Short-Time Fourier Transform



Zoom FFT

- Technika przetwarzania sygnału stosowana do analizy **fragmentu widma** z dużą rozdzielczością
- Istotą tej techniki jest **zachowanie tej samej rozdzielczości**, jaką można uzyskać przy pełnowymiarowym FFT na oryginalnym sygnale, obliczając małą FFT na krótszym sygnale.
- Krótszy sygnał uzyskuje się poprzez zastosowanie decymacji.
- Dzięki temu zyskuje się oszczędność złożoności obliczeniowej wynikającą z możliwości obliczenia znacznie krótszej FFT przy jednoczesnym osiągnięciu tej samej rozdzielczości.

Zoom FFT

- Przykład: dla współczynnika decymacji D nowa częstotliwość próbkowania to:

$$\mathbf{F_{sd} = F_s / D,}$$

- Nowy rozmiar ramki (i długość FFT) to:

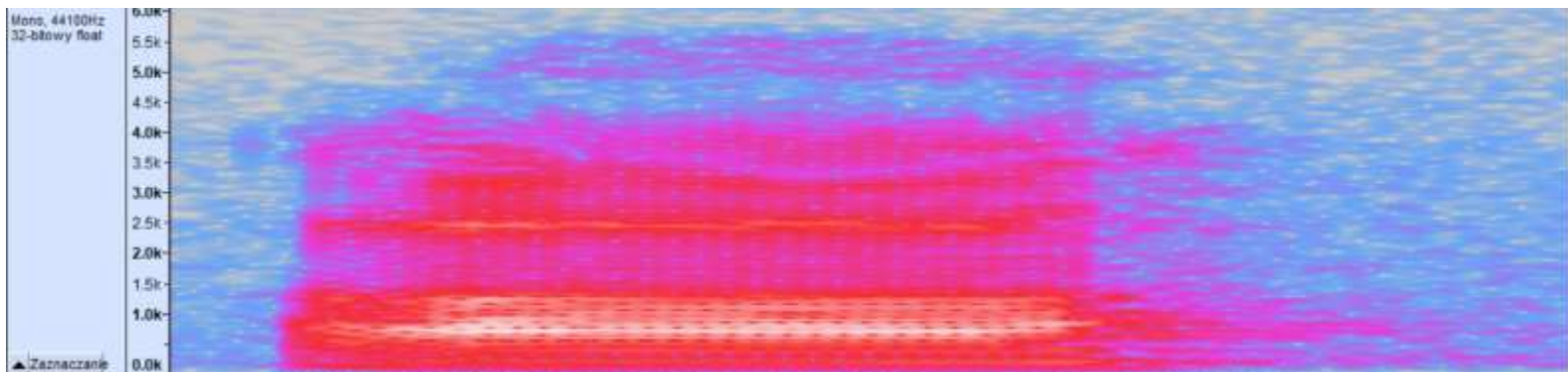
$$\mathbf{L_d = L / D,}$$

- Rozdzielczość zdecydowanego sygnału to:

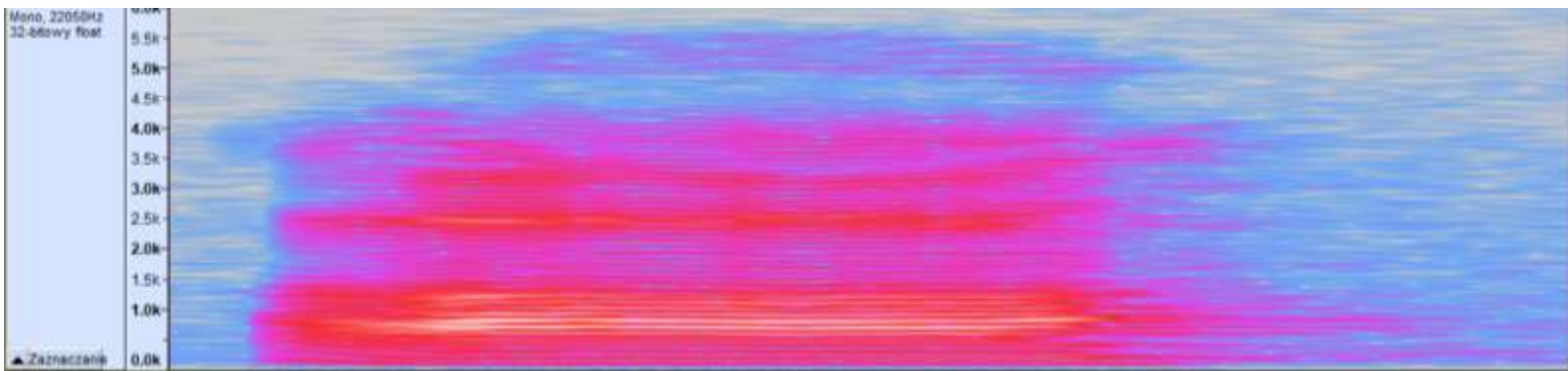
$$\mathbf{F_{sd} / L_d = F_s / L}$$

Przykład, samogłoska A

F_s
44100
NFFT
1024
 $D = 1$
 $df = 43$ Hz



F_s
22050
NFFT
1024
 $D = 2$
 $df = 21$ Hz



F_s
5512
NFFT
1024
 $D = 8$
 $df = 5$ Hz



Metoda CPB

- Dekompozycja sygnału realizowana za pomocą banku filtrów o procentowo stałej szerokości
- Szeroko stosowana w praktyce pomiarowej
- Na wyjściu danego filtru wciąż dysponujemy sygnałem w dziedzinie czasu
- Estymacja widma wymaga określenia energii sygnału na wyjściu każdego filtru

Metoda CPB

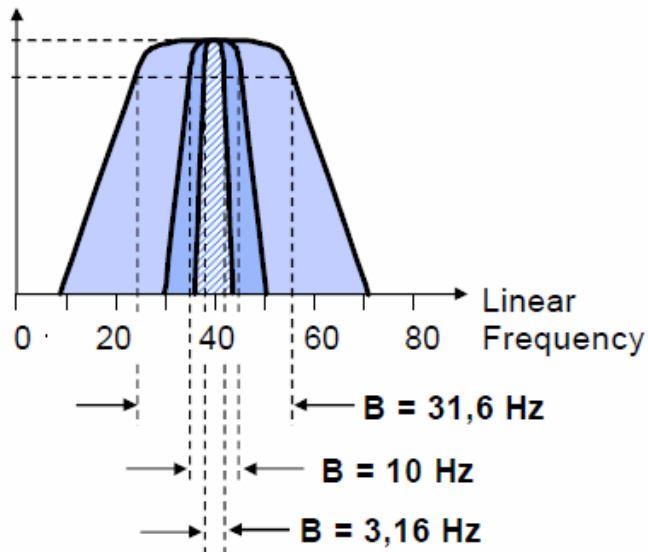
- Decydując się na analizę widma sygnału za pomocą banku filtrów należy dokonać wyboru szerokości pasma zestawu filtrów
- Rozpatrując stałą procentową szerokość pasma można wyróżnić filtry o szerokości pasma wynoszącej jedną oktawę lub jej część, czyli $1/n$ oktawy, gdzie n oznacza liczbę części oktawy, np. 3, 6, 12...

Metoda CPB

Filter Types

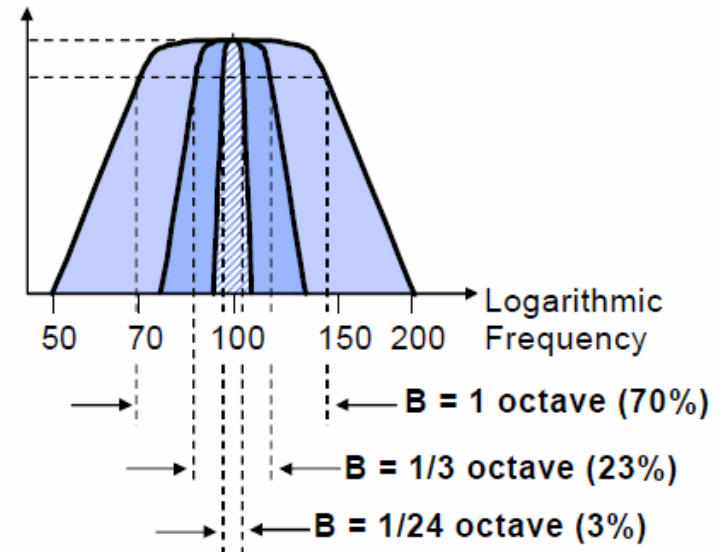
Constant Bandwidth

$$B = x \text{ Hz}$$



Constant Percentage Bandwidth
– Relative Bandwidth –

$$B = y \% = \frac{y \times f_0}{100} \text{ Hz}$$



Metoda CPB

- Ograniczenie metody CPB dla dużych wartości n (np. 12) wynika z długiego czasu ustalania się odpowiedzi na wyjściu filtru i z tego powodu użyteczność tego rodzaju analizy dla sygnału mowy jest ograniczona
- Dla filtru oktawowego przebiegi ustalają się po jednym okresie fali
- Filtr 1/12 oktawy wymaga ponad 17 okresów fali

Metoda CPB

Filtry o stałej względnej szerokości pasma – 1/1 i 1/3 oktawy

$$f_{sr} = \sqrt{f_d f_g}, \quad \frac{f_g}{f_d} = \sqrt[n]{2} = 2^{\frac{1}{n}}$$

gdzie:

f_{sr} – częstotliwość środkowa filtra,

n – część oktawy filtra, np. $n=3$ – filtr 1/3 oktawowy, $n=12$, filtr 1/12 oktawowy,

f_d, f_g – częstotliwości odpowiednio dolna i górna filtra

Metoda CPB

Względna szerokość pasma przenoszenia wyrażona w %

Zależność ogólna

$$\Delta f_w = \frac{f_g - f_d}{f_{\acute{s}r}} 100\%$$

Filtr oktawowy:

$$\Delta f_w = \frac{f_g - f_d}{f_{\acute{s}r}} 100\% = \frac{2f_d - f_d}{\sqrt{2}f_d} 100\% = \frac{1}{\sqrt{2}} 100\% = 70.7\%$$

Filtr tercjowy:

$$\Delta f_w = \frac{f_g - f_d}{f_{\acute{s}r}} 100\% = \frac{2^{\frac{1}{3}}f_d - f_d}{2^{\frac{1}{6}}f_d} 100\% = 23.1\%$$

Charakterystyka filtru a częstotliwość próbkowania

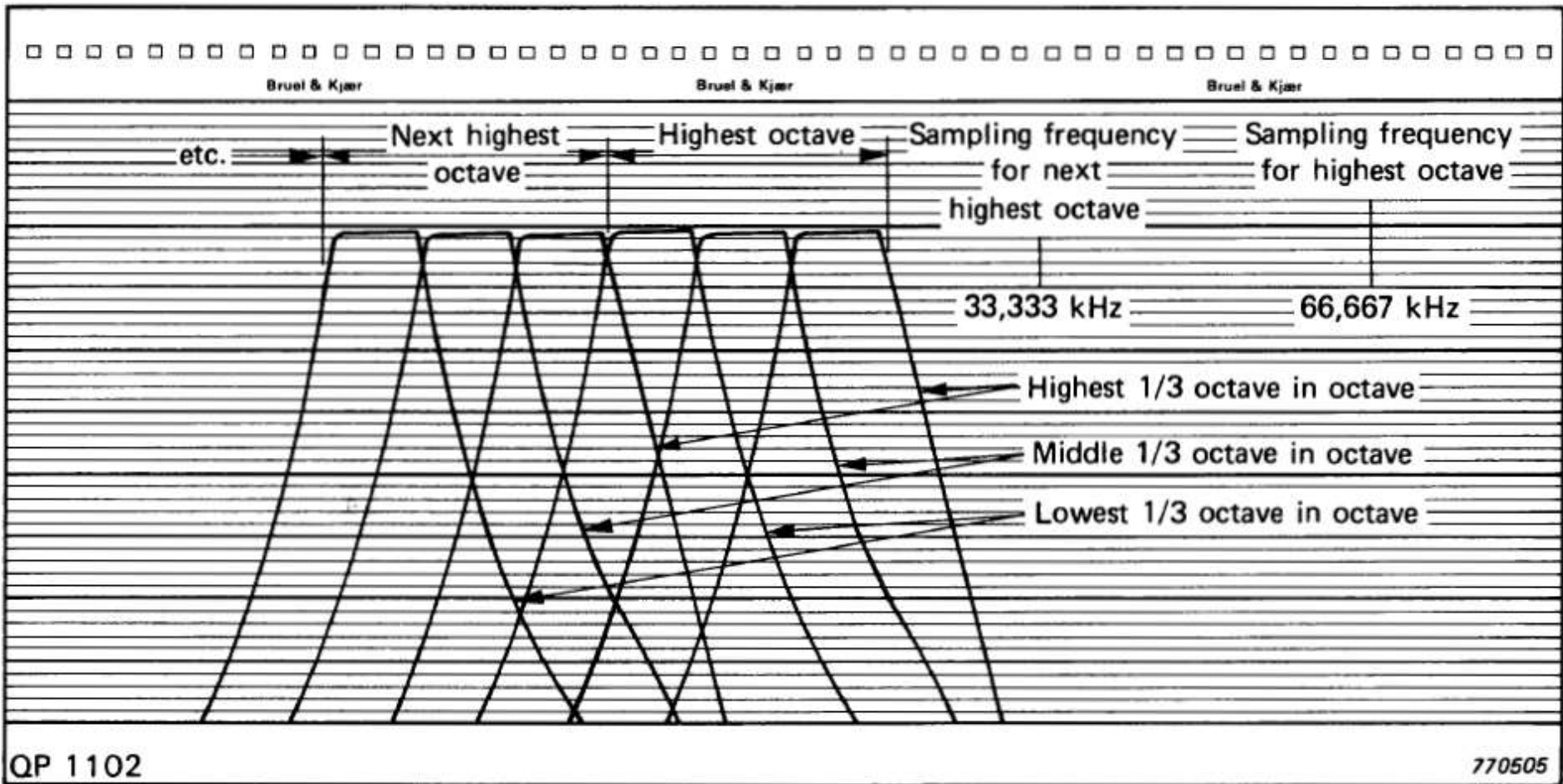
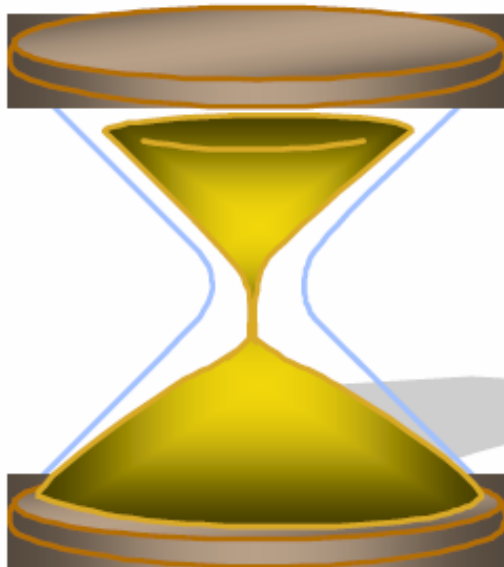


Fig. 3.39. Filter characteristics vs. sampling frequency

Most important Law in Frequency Analysis



$$\mathbf{B} \times \mathbf{T} \geq \mathbf{1}$$

B = bandwidth

T = measurement time

PORÓWNANIE METOD CPB - FFT

Zadanie domowe

- Poeksperymentować z metodami określania widma