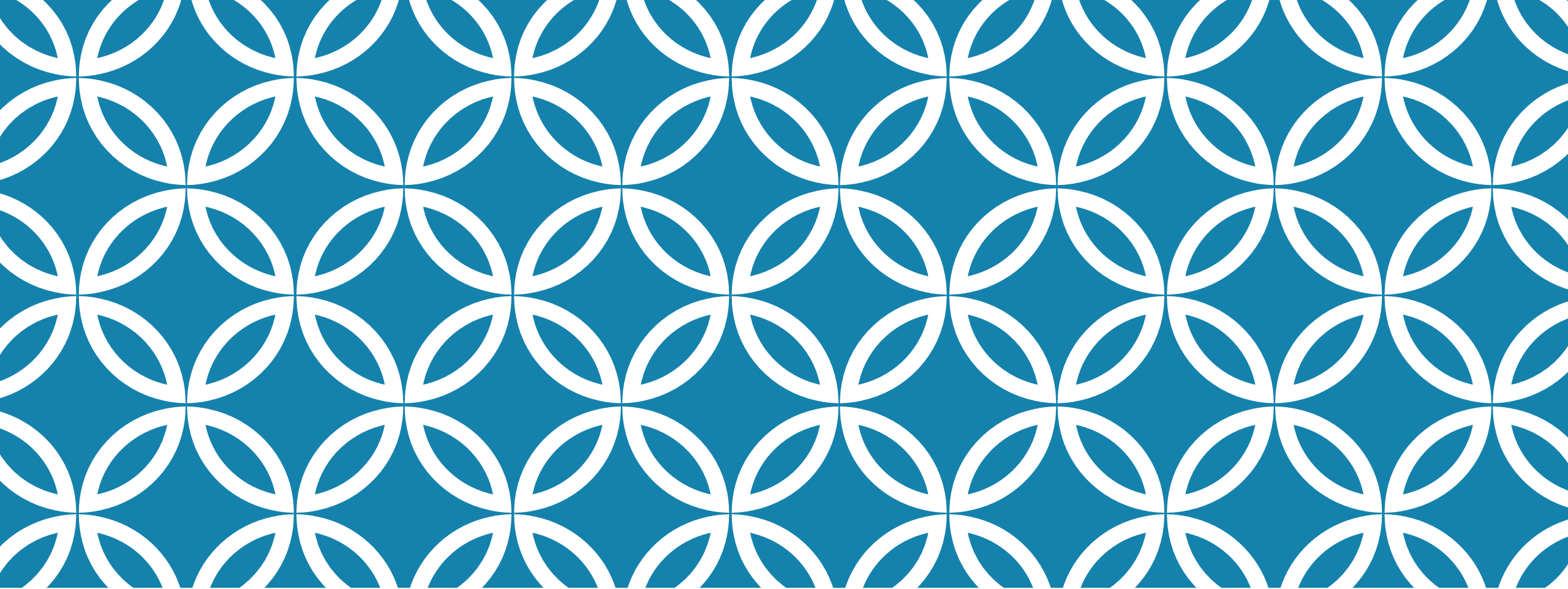


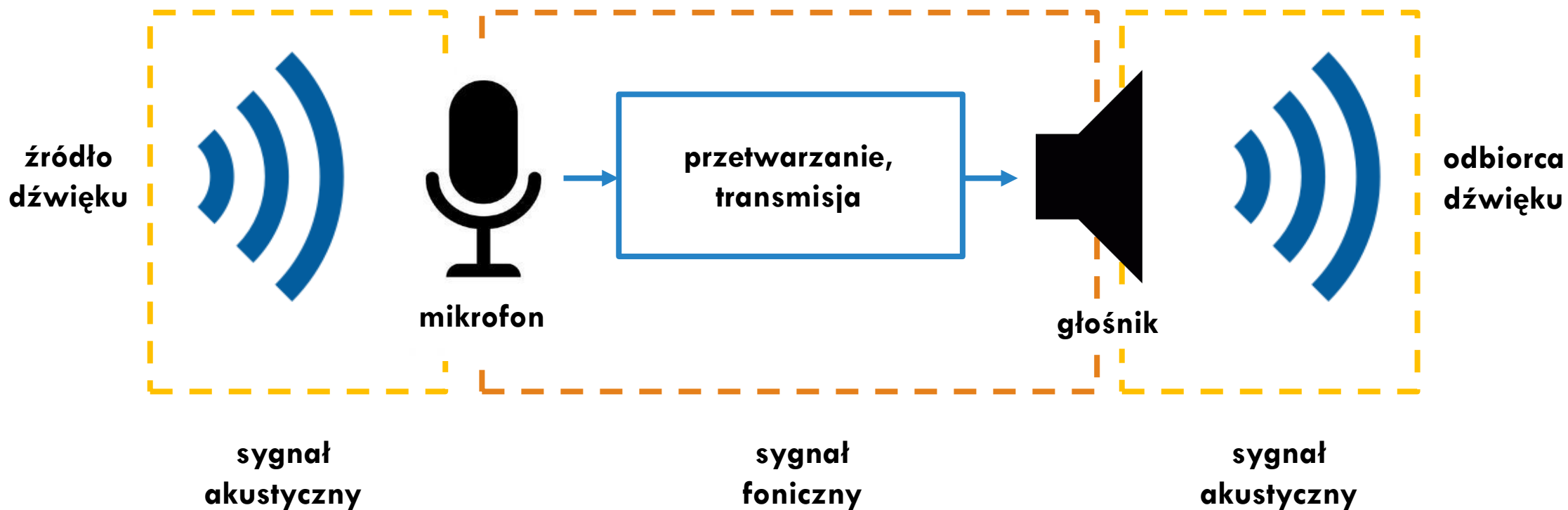


WSTĘP DO MULTIMEDIÓW

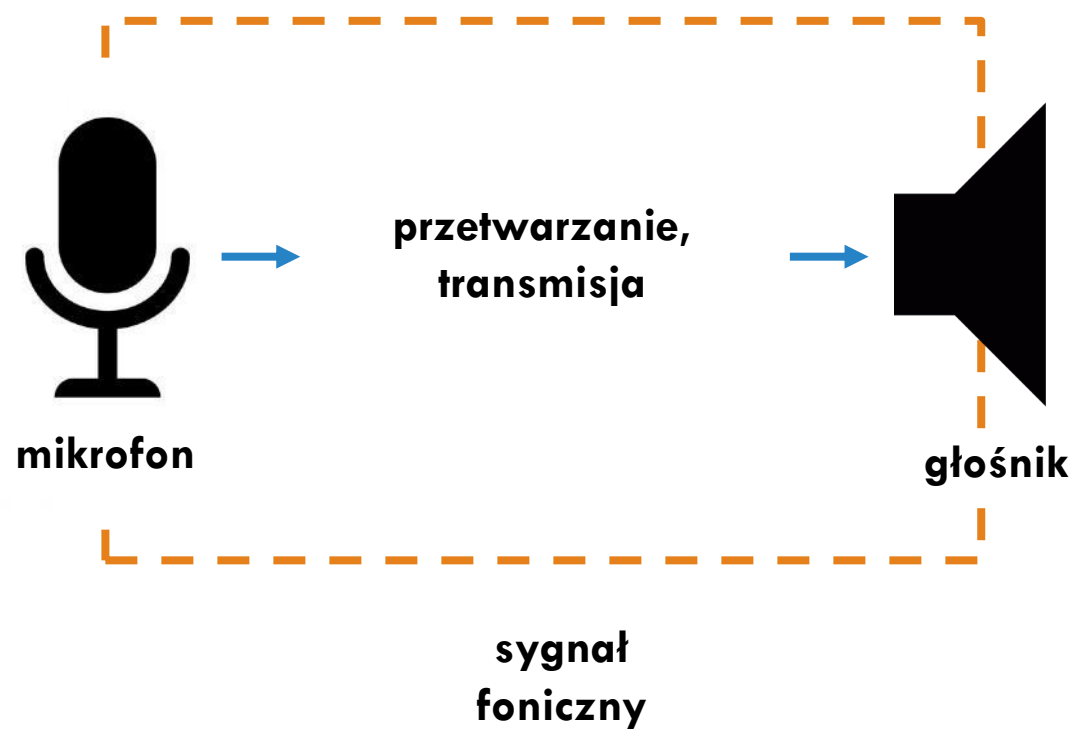
WYKŁAD 4 – CYFROWE PRZETWARZANIE DŹWIĘKU

A. Paula Pietrzak
A.Pietrzak@ire.pw.edu.pl

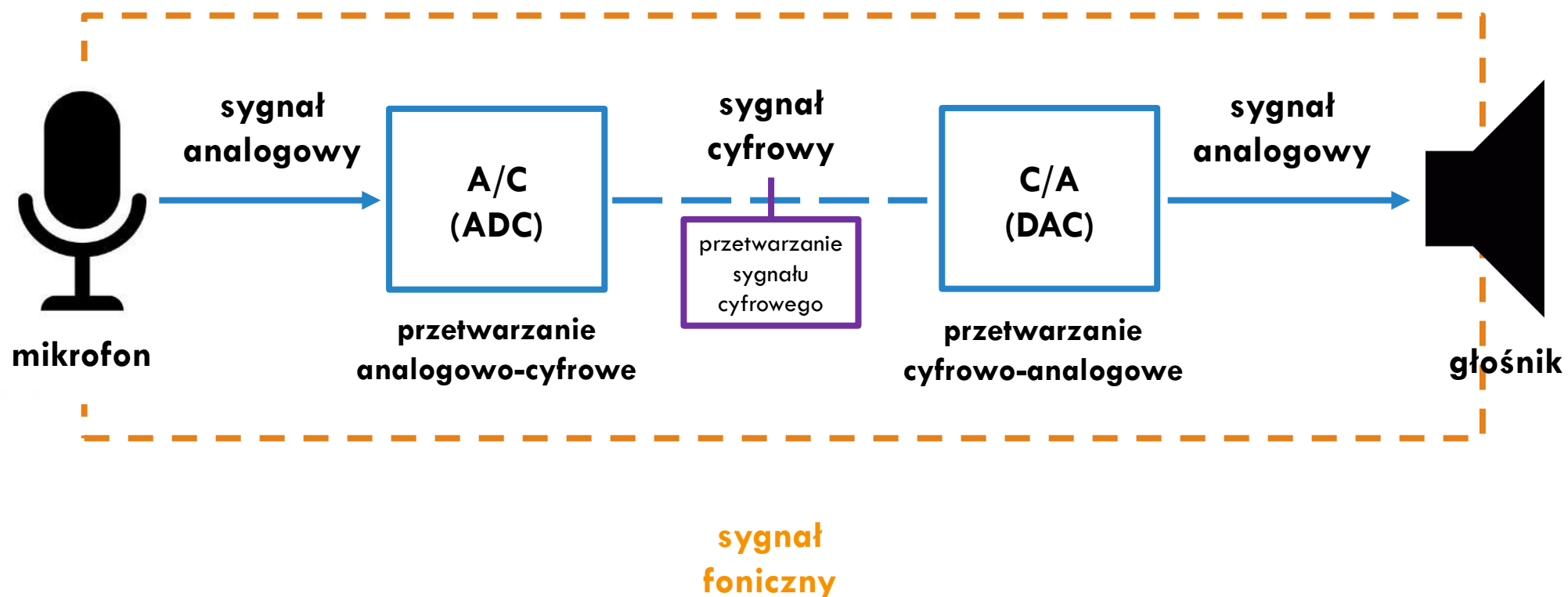
SYGNAŁ AKUSTYCZNY / SYGNAŁ FONICZNY



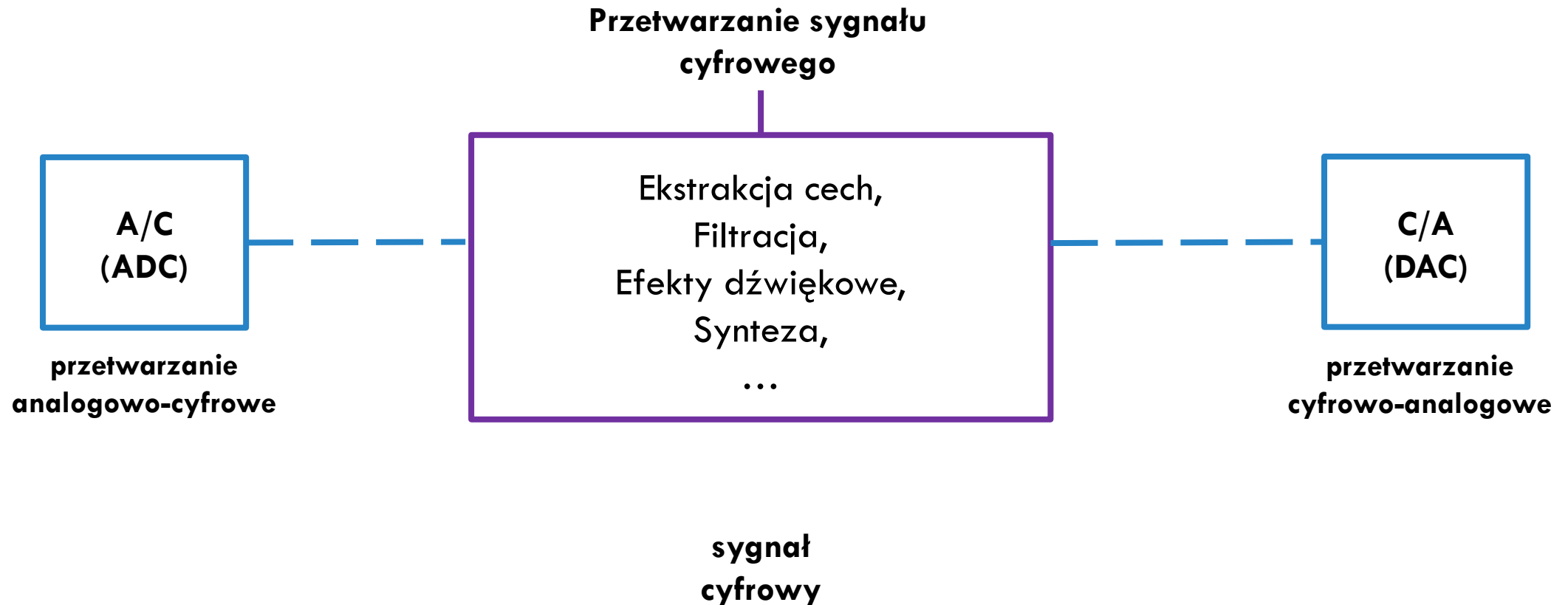
SYGNAŁ AKUSTYCZNY / SYGNAŁ FONICZNY



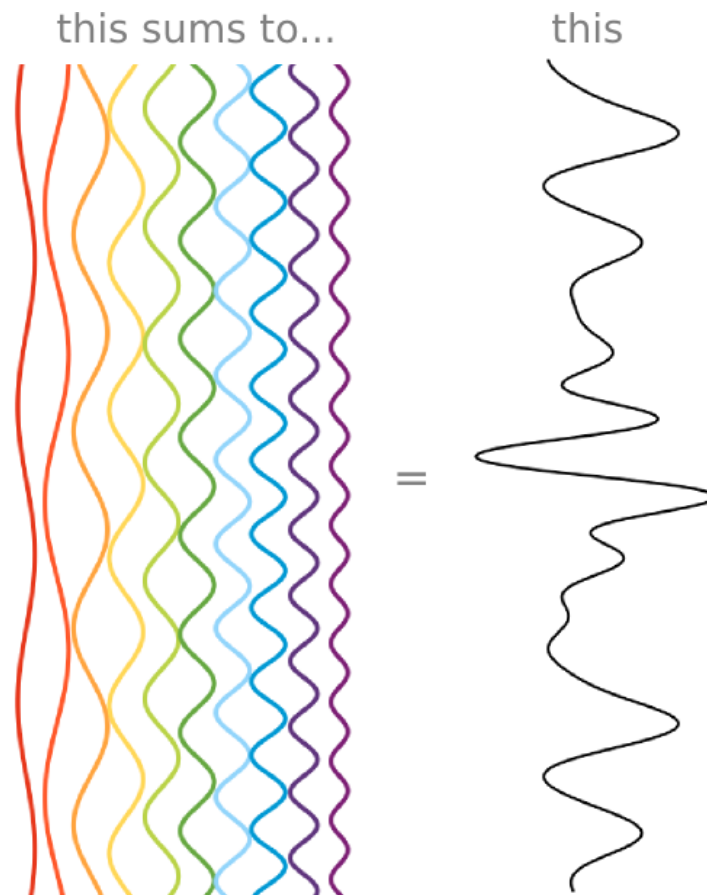
CYFROWY SYGNAŁ FONICZNY

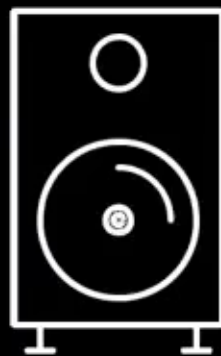
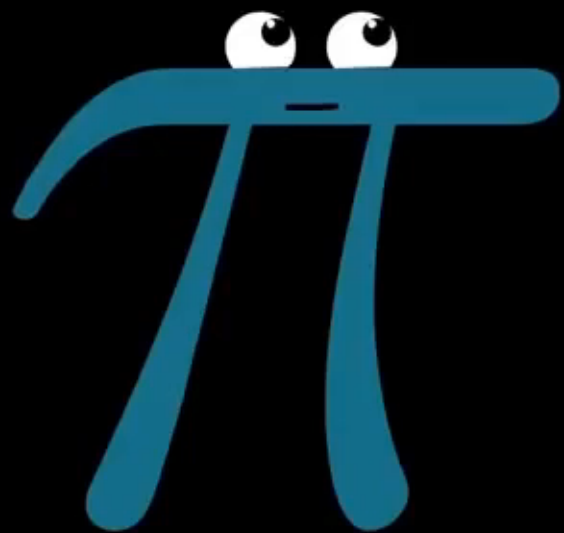


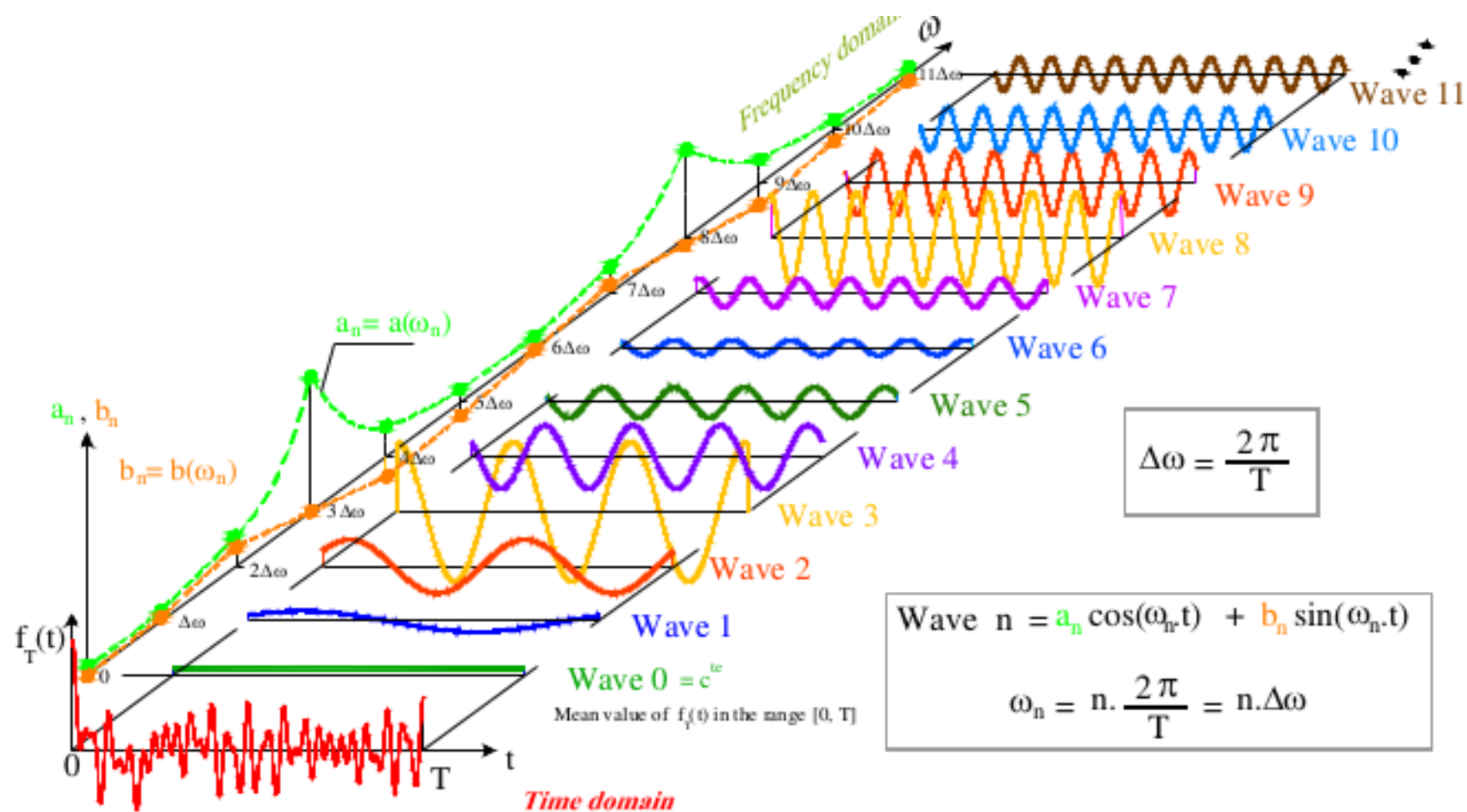
CYFROWE PRZETWARZANIE SYGNAŁÓW FONICZNYCH



DZIEDZINA CZĘSTOTLIWOŚCI







Fourier series approximation of the function: $f_T(t)$, $t \in [0, T]$

$$f_T(t) = \underset{\downarrow v_m}{c^{lc}} + \text{Wave 1} + \text{Wave 2} + \text{Wave 3} + \text{Wave 4} + \dots$$

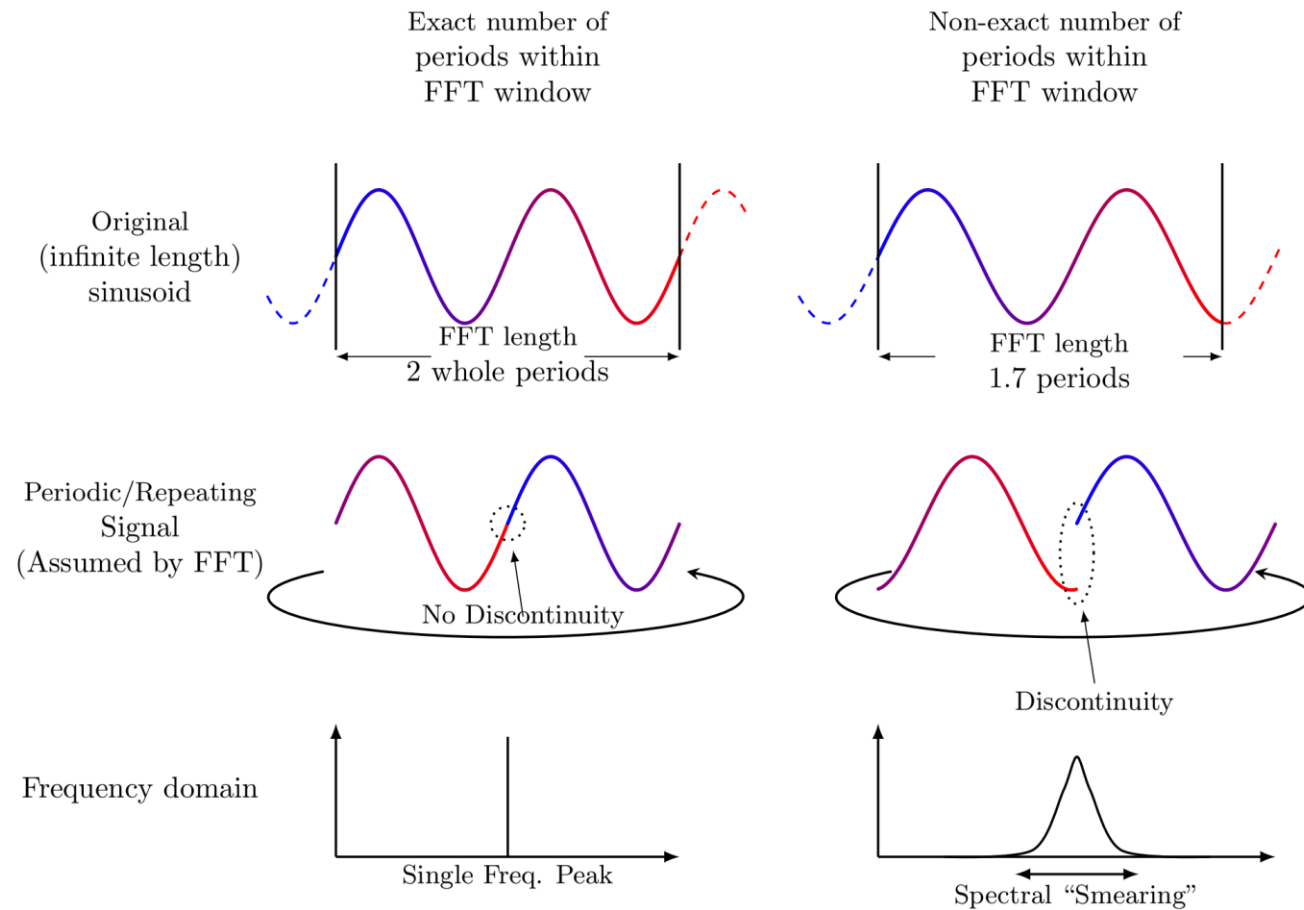
$\downarrow$
 $a_1 \cos(\omega_1 t) + b_1 \sin(\omega_1 t)$
 $\omega_1 = 1 \cdot \Delta\omega$

$\downarrow$
 $a_2 \cos(\omega_2 t) + b_2 \sin(\omega_2 t)$
 $\omega_2 = 2 \cdot \Delta\omega$

$\downarrow$
 $a_3 \cos(\omega_3 t) + b_3 \sin(\omega_3 t)$
 $\omega_3 = 3 \cdot \Delta\omega$

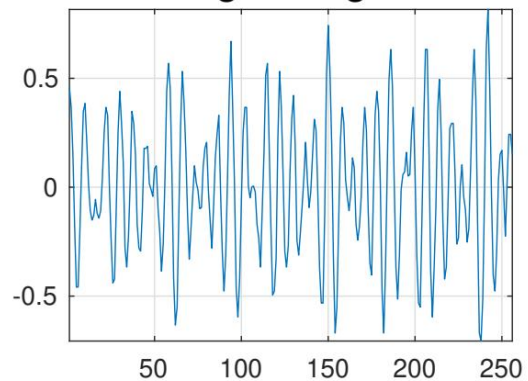
...

OKNO ANALIZY FFT

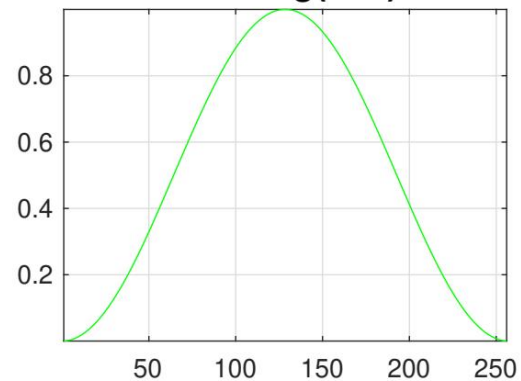


OKNO ANALIZY FFT

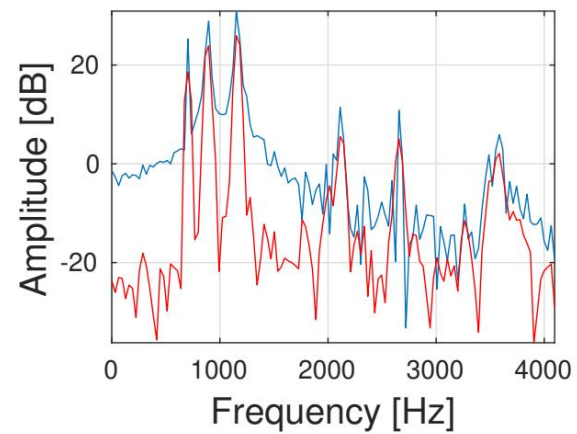
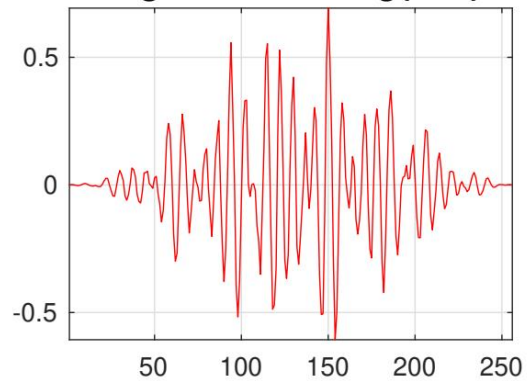
Original Signal



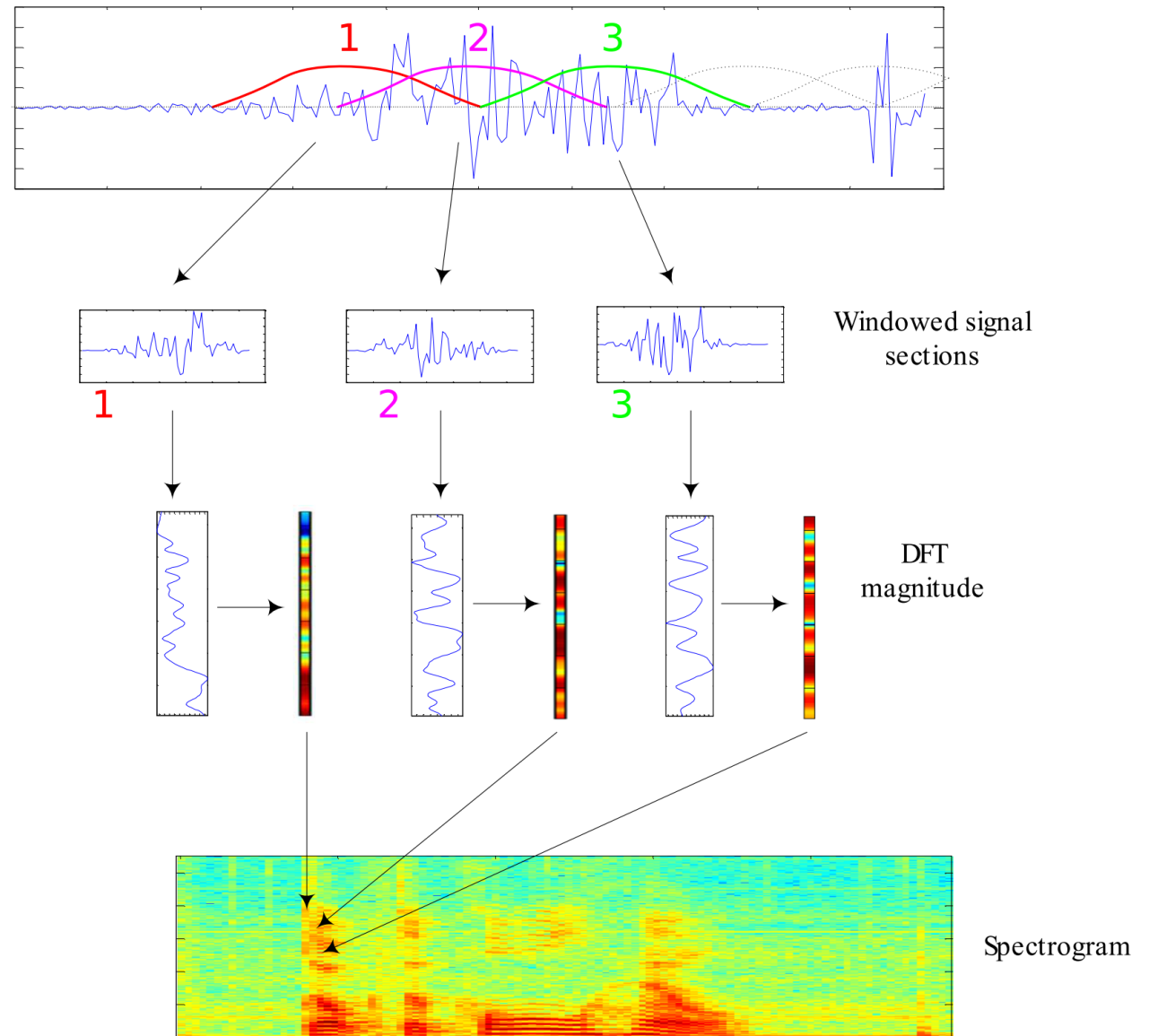
`hanning(256)`



`Signal .* hanning(256)`

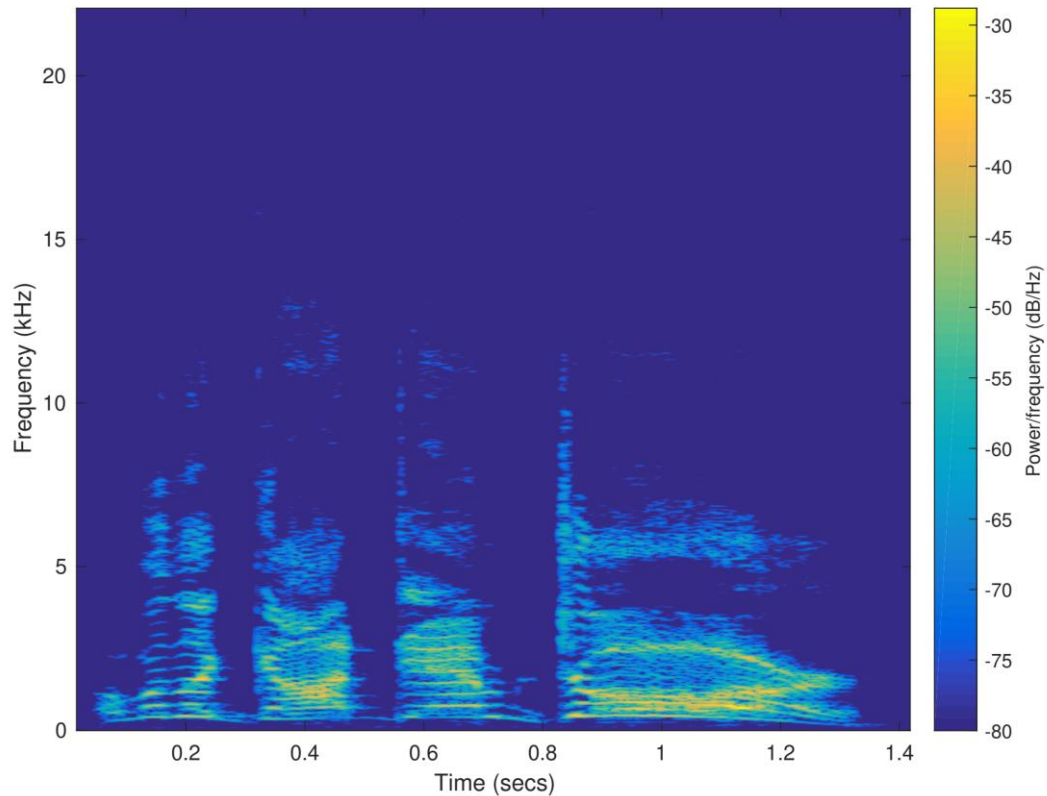


SPEKTROGRAM

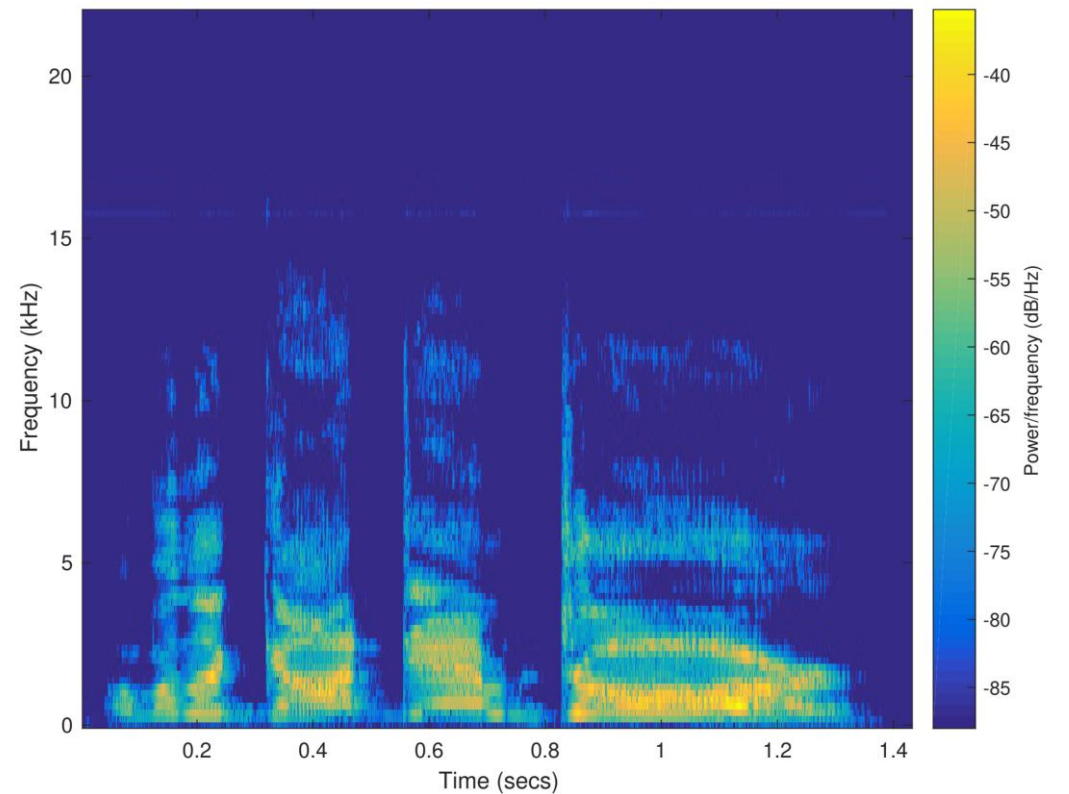


ROZMIAR OKNA

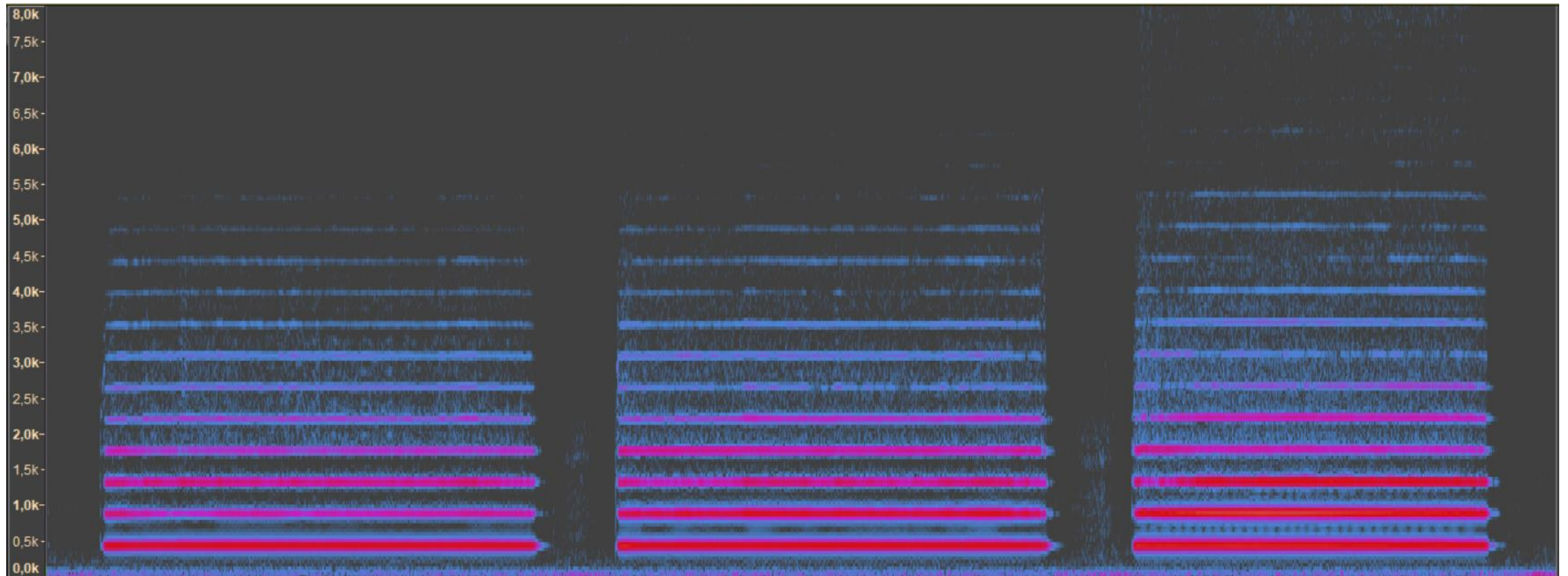
dłuższe okno - większa liczba próbek:
+ lepsza rozdzielczość częstotliwości
- gorsza rozdzielczość czasowa



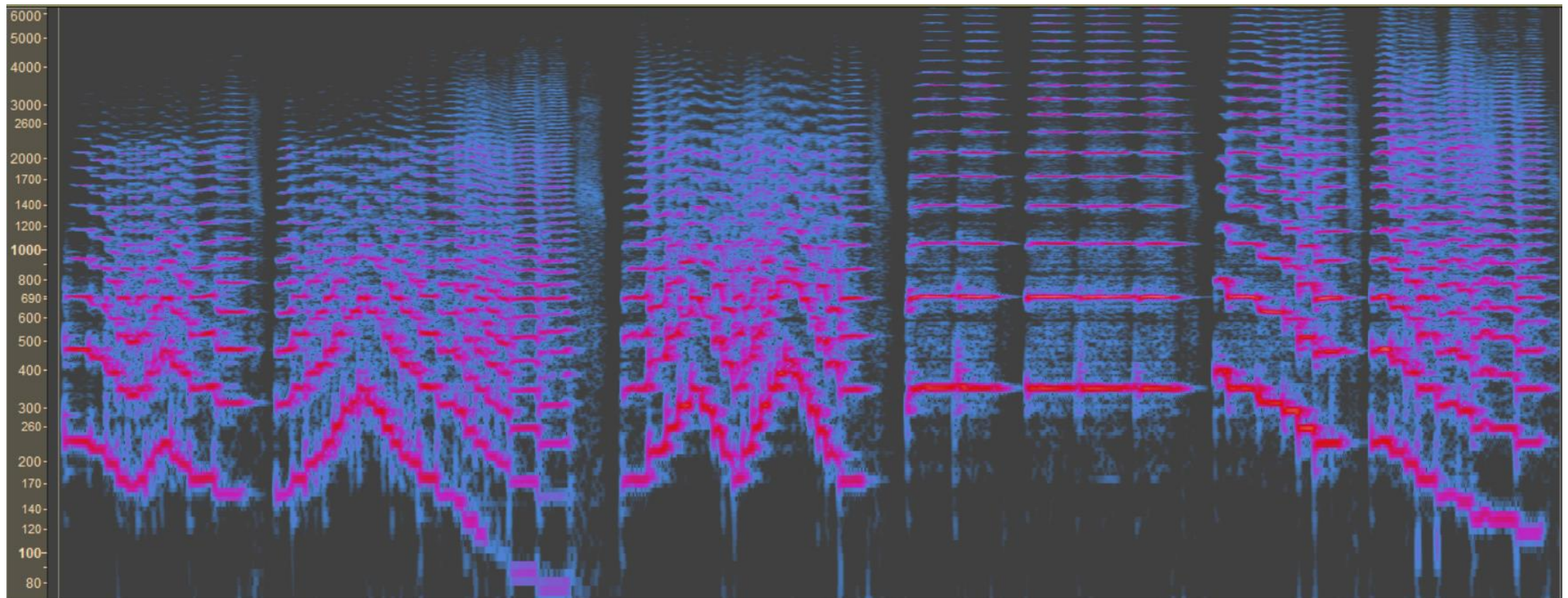
krótsze okno - mniejsza liczba próbek:
+ lepsza rozdzielczość czasowa
- gorsza rozdzielczość częstotliwości



SPEKTROGRAM

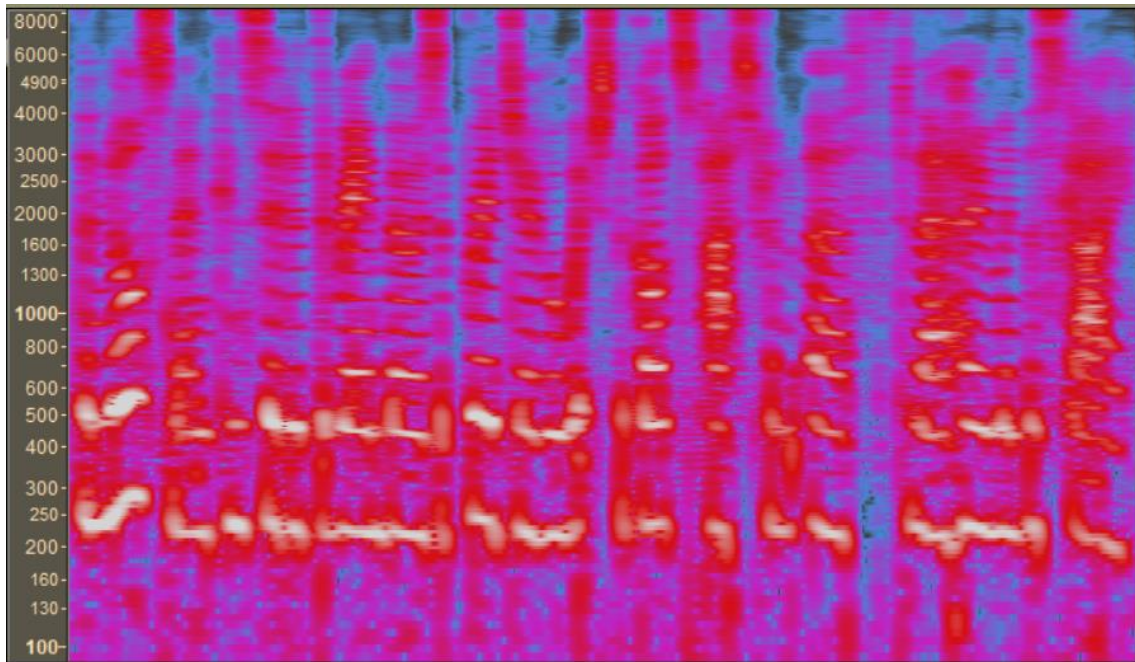


MUZYKA - MELODIA

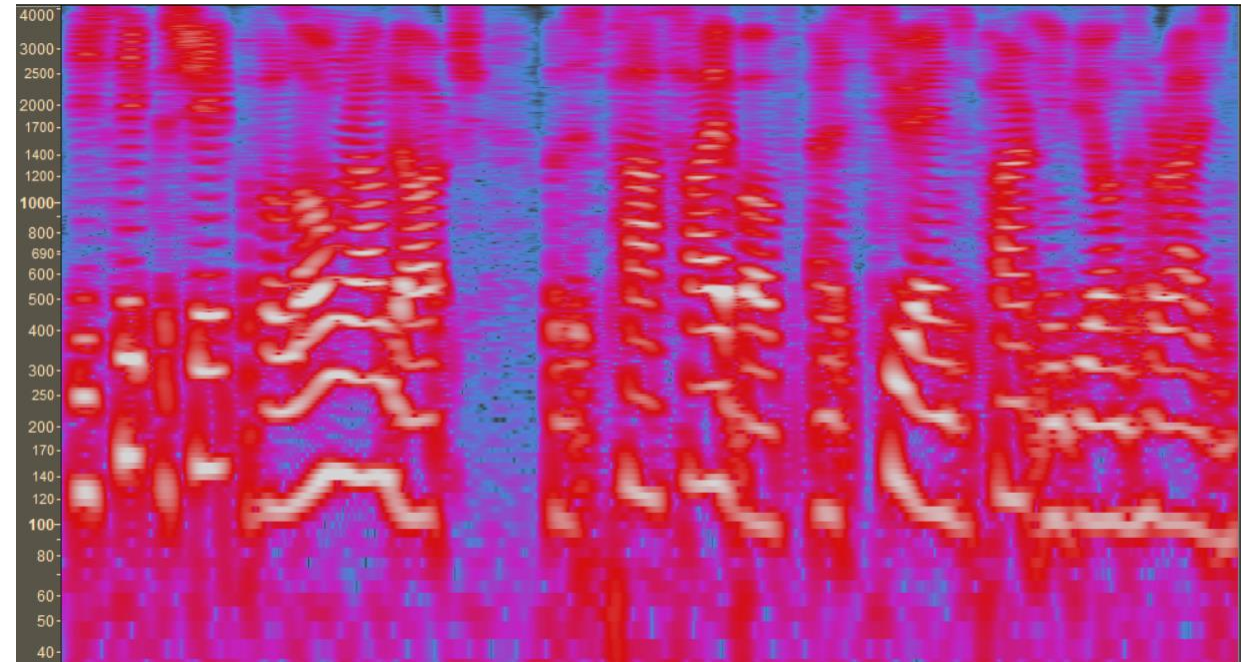


MOWA

KOBIETA

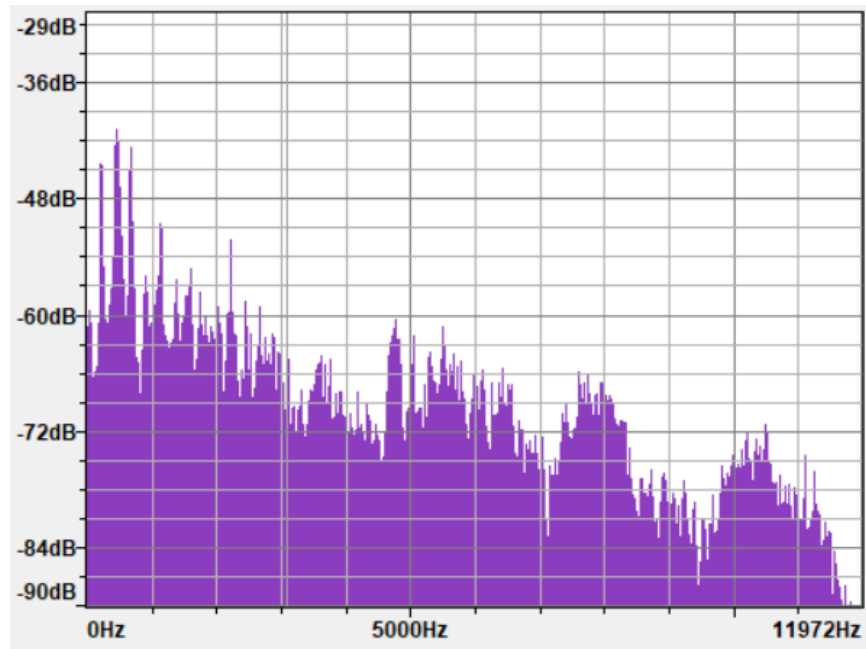


MĘŻCZYŻNA

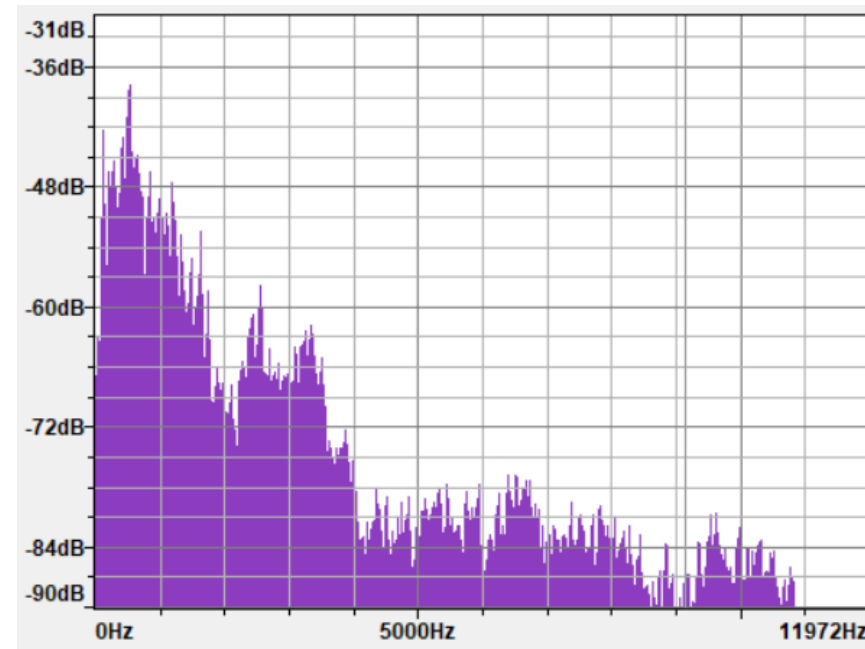


MOWA

KOBIETA



MĘŻCZYŻNA



EKSTRAKCJA CECH DŹWIĘKU

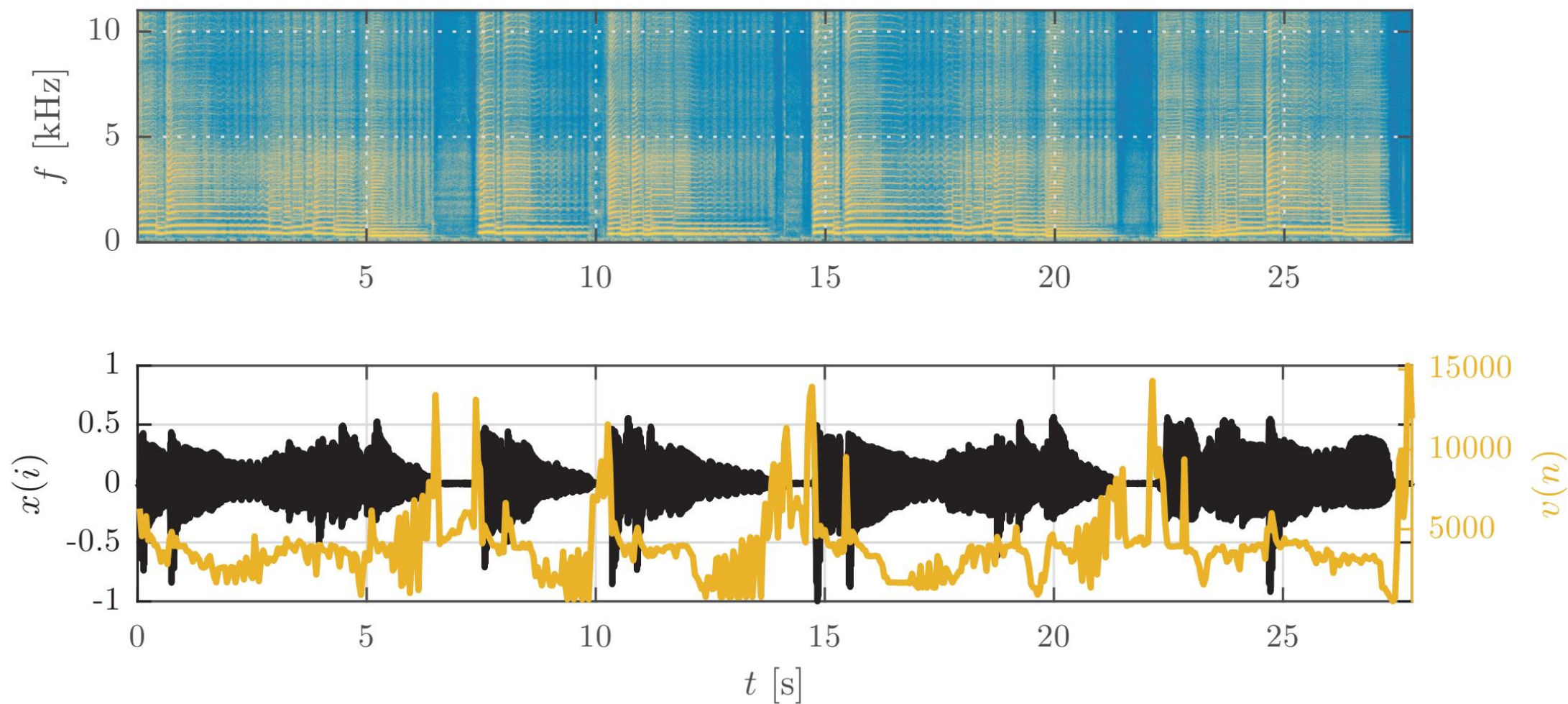


SPECTRAL ROLL-OFF

- parametr „opadania widma”
- częstotliwość poniżej której znajduje się określony procent całej energii, np. 85%

$$v_{\text{SR}}(n) = i \left| \sum_{k=0}^i |X(k,n)| = \kappa \cdot \sum_{k=0}^{\kappa/2-1} |X(k,n)| \quad \kappa = 0.85 \quad \kappa = 0.95 \right.$$

SPECTRAL ROLL-OFF

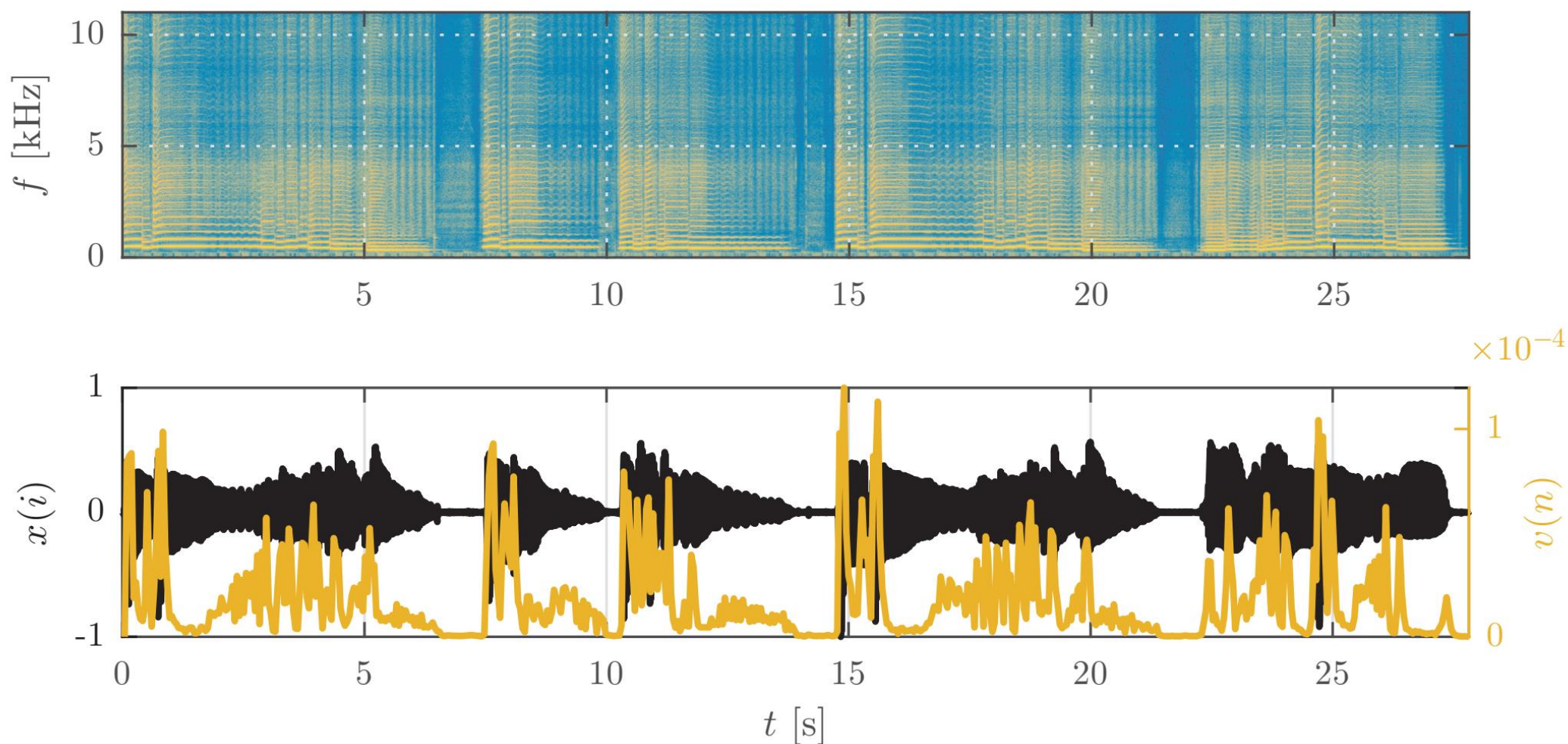


SPECTRAL FLUX

- parametr określający szybkość zmian widma
- porównanie widma z bieżącego okna do widma z okna poprzedniego
- wykorzystywany do opisu barwy dźwięku
- do wykrywania **początków nut** („onset detection”) lub słów

$$v_{\text{SF}}(n) = \frac{\sqrt{\sum_{k=0}^{\kappa/2-1} (|X(k, n)| - |X(k, n-1)|)^2}}{\kappa/2}.$$

SPECTRAL FLUX

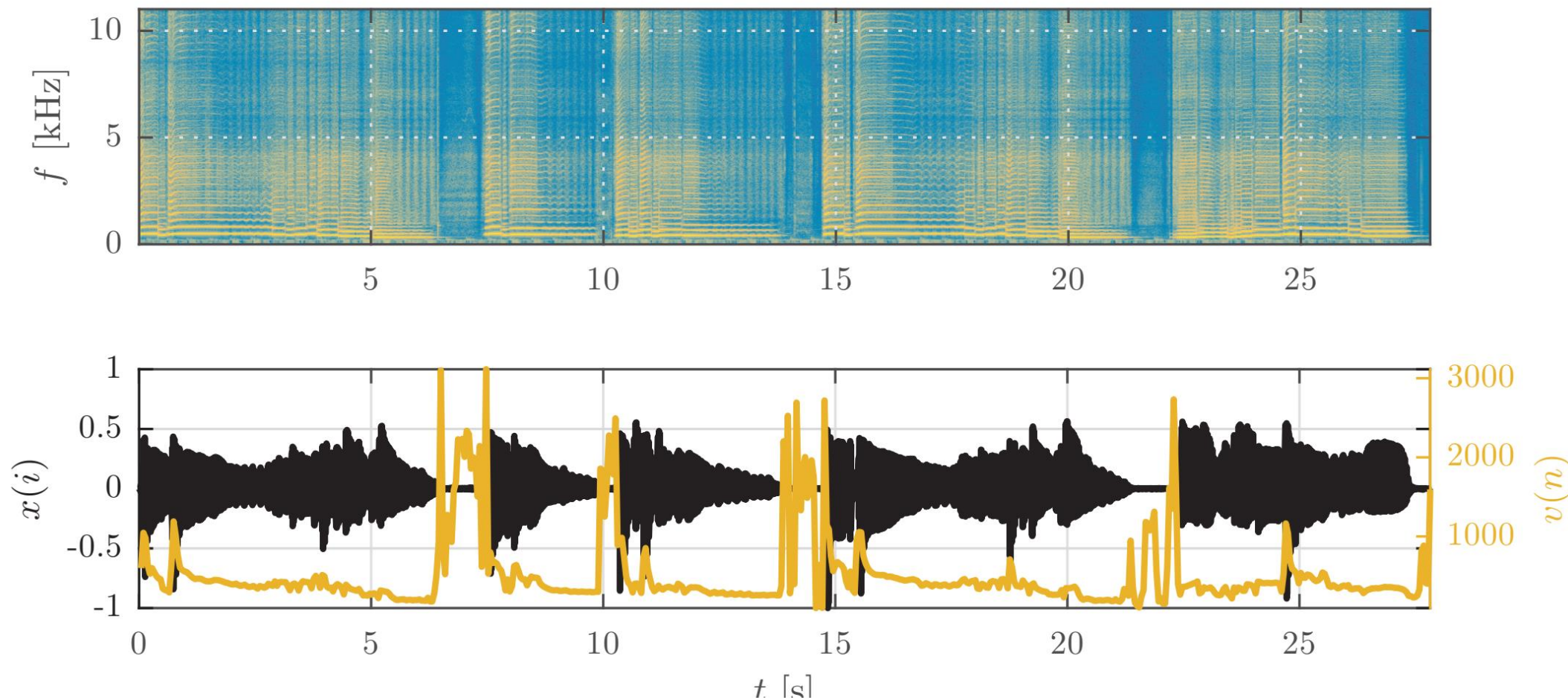


SPECTRAL CENTROID

- „środek ciężkości” widma
- określa jasność brzmienia
- im większa wartość, tym jaśniejsza barwa dźwięku

$$v_{\text{SC}}(n) = \frac{\sum_{k=0}^{\mathcal{K}/2-1} k \cdot |X(k, n)|^2}{\sum_{k=0}^{\mathcal{K}/2-1} |X(k, n)|^2}.$$

SPECTRAL CENTROID

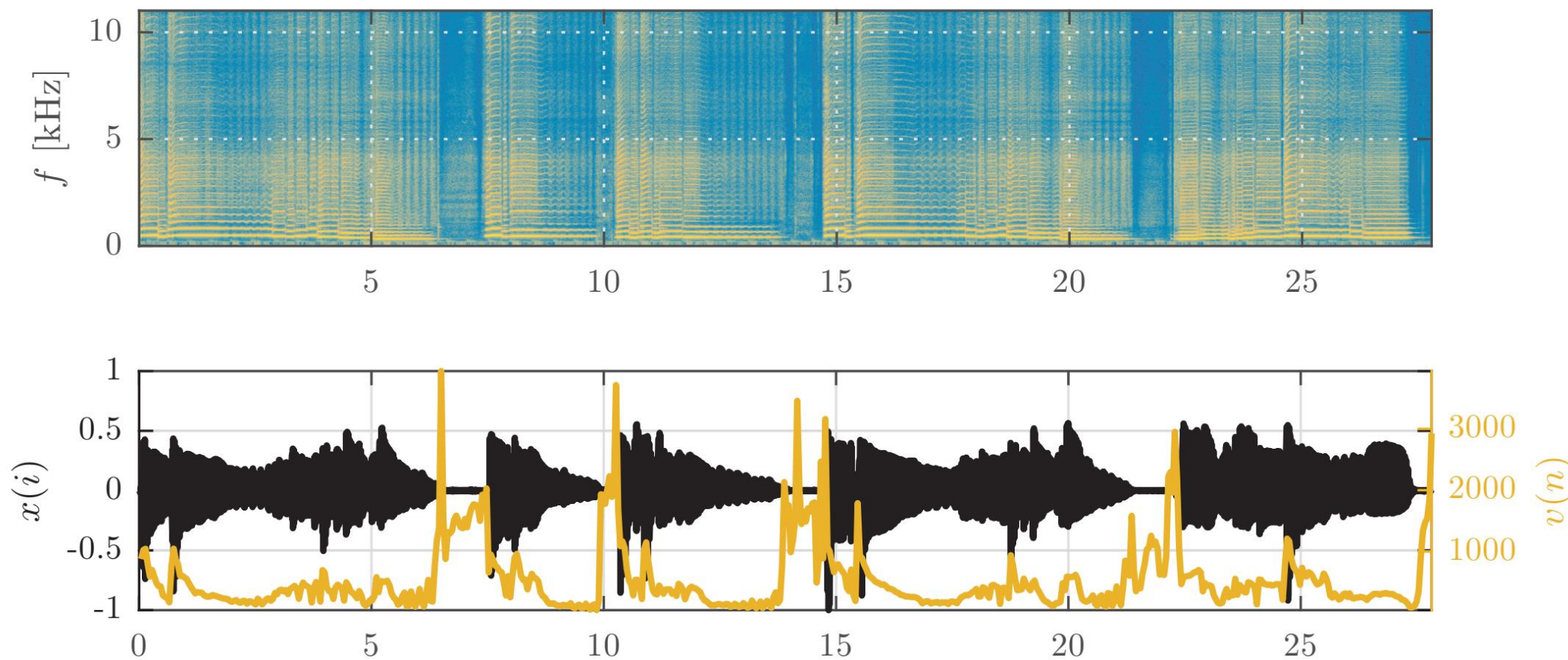


SPECTRAL SPREAD

- opisuje średnie odchylenie od centroidy
- dla szumów duża wartość
- dla dźwięków tonalnych niska wartość

$$v_{SS}(n) = \sqrt{\frac{\sum_{k=0}^{\mathcal{K}/2-1} (k - v_{SC}(n))^2 \cdot |X(k, n)|^2}{\sum_{k=0}^{\mathcal{K}/2-1} |X(k, n)|^2}}.$$

SPECTRAL SPREAD

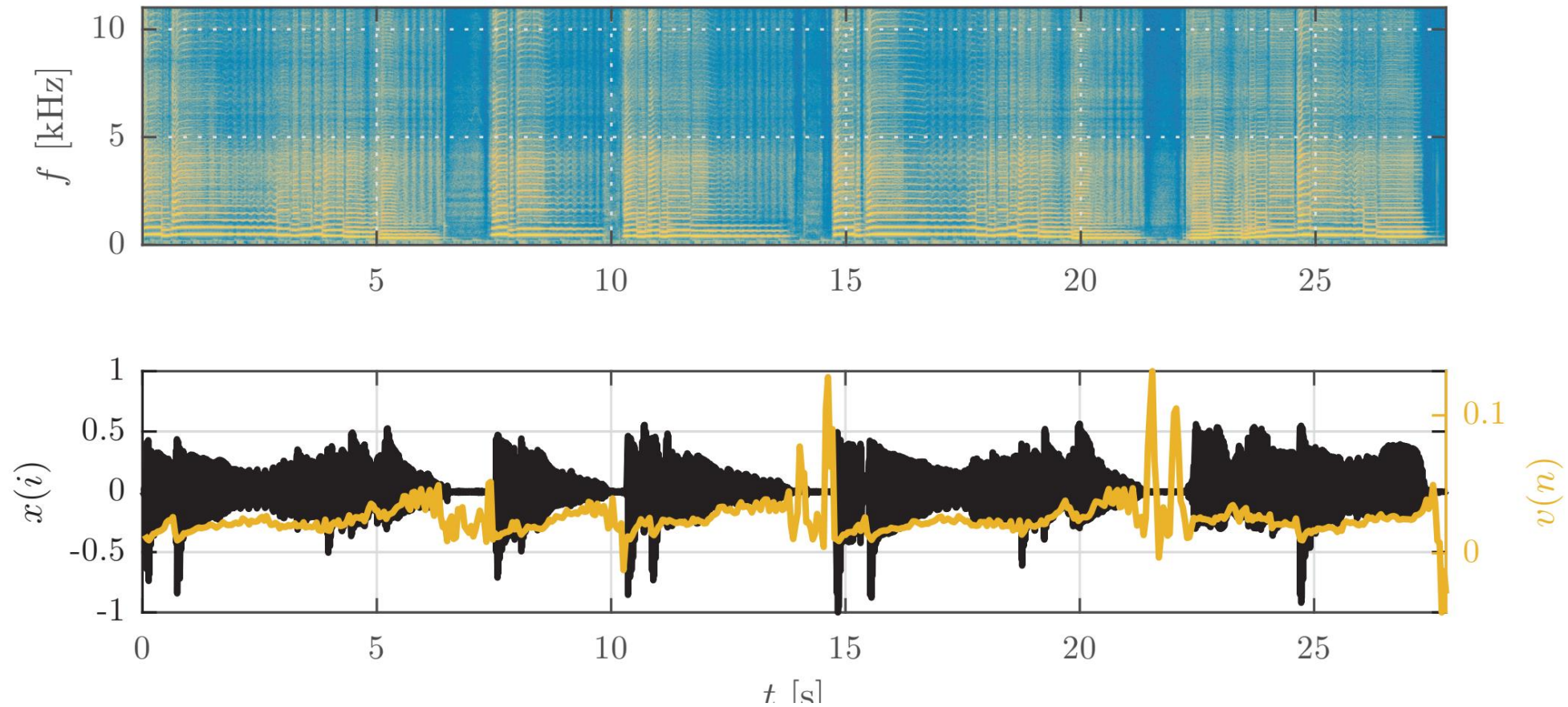


SPECTRAL DECREASE

- opisuje stromość zbocza widma
- niskie wartości wskazują na koncentrację energii dla niskich częstotliwości

$$v_{SD}(n) = \frac{\sum_{k=1}^{\kappa/2-1} \frac{1}{k} \cdot (|X(k, n)| - |X(0, n)|)}{\sum_{k=1}^{\kappa/2-1} |X(k, n)|}.$$

SPECTRAL DECREASE



MFCC - MEL-FREQUENCY CEPSTRAL COEFFICIENTS

- cepstrum sygnału przedstawione w skali melowej
- wektor parametrów mel-cepstralnych to wektor współczynników cepstrum w odpowiednich pasmach melowych
- mają za zadanie odzwierciedlać naturalną odpowiedź układu słuchowego na pobudzenie dźwiękami mowy

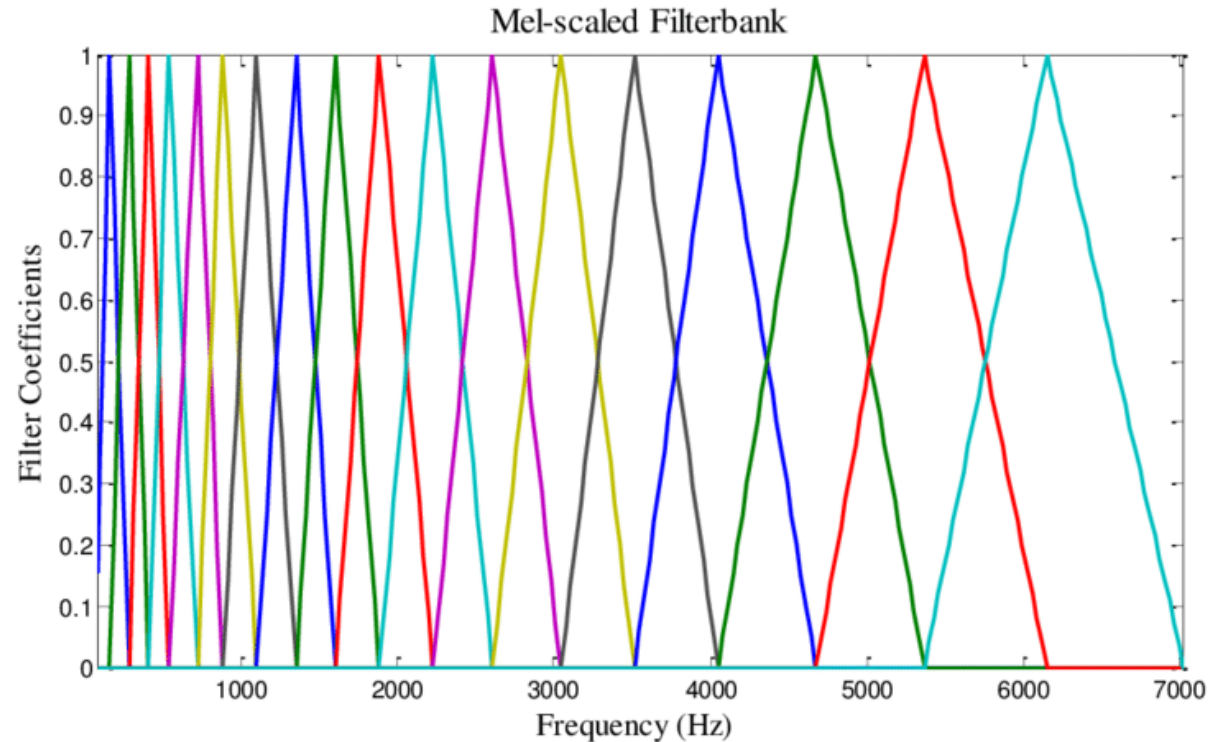
$$v_{\text{MFCC}}^j(n) = \sum_{k'=1}^{\mathcal{K}'} \log (|X'(k', n)|) \cdot \cos \left(j \cdot \left(k' - \frac{1}{2} \right) \frac{\pi}{\mathcal{K}'} \right)$$

SKALA MELOWA

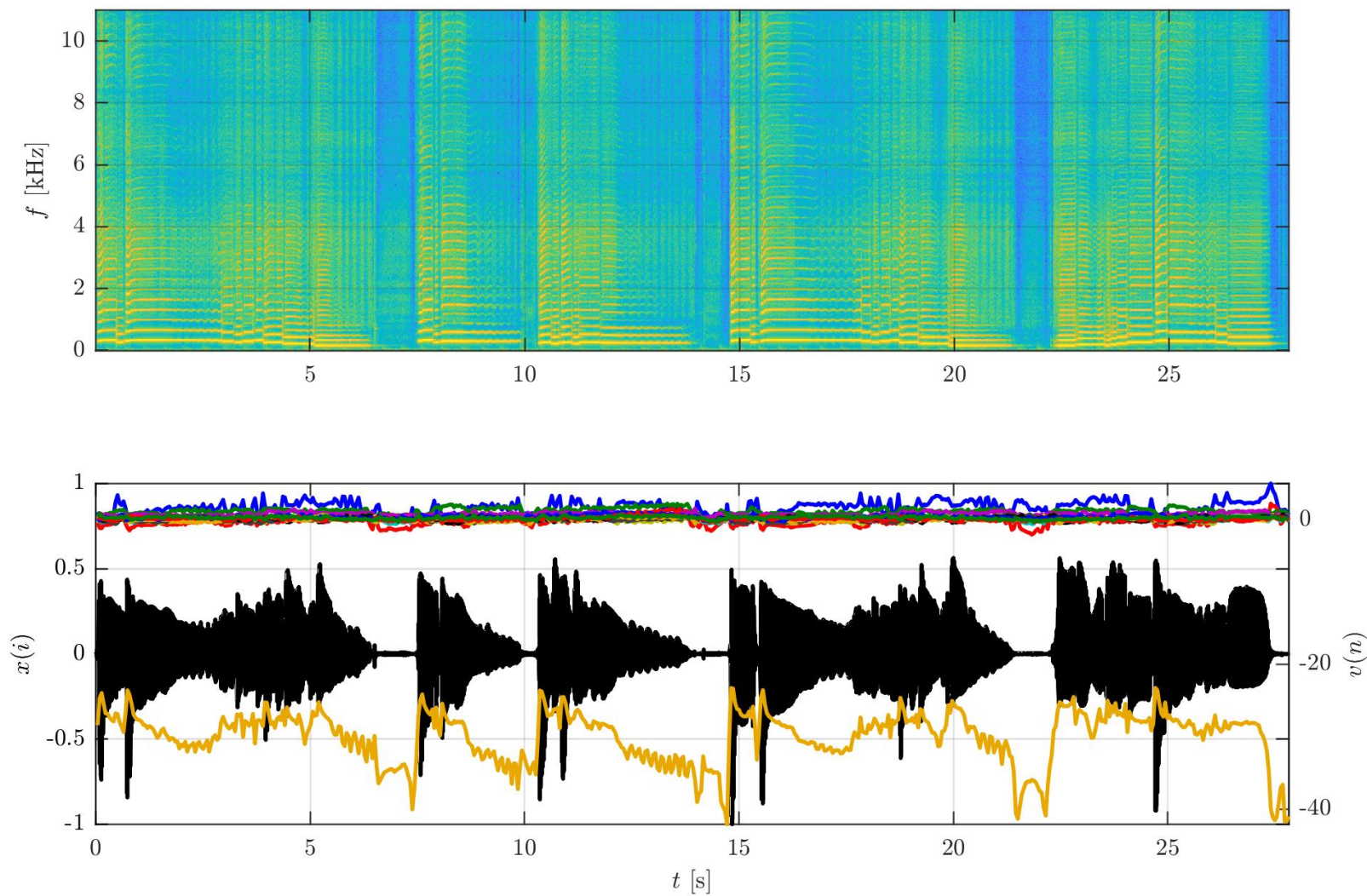
Skalę melową uzyskuje się poprzez filtrację sygnału bankiem filtrów o charakterystyce trójkątnej.

K-ty współczynnik mel-cepstralny odpowiada zawartości k-tego pasma.

Zazwyczaj liczba pasm wynosi od 12 do 20.



MFCC

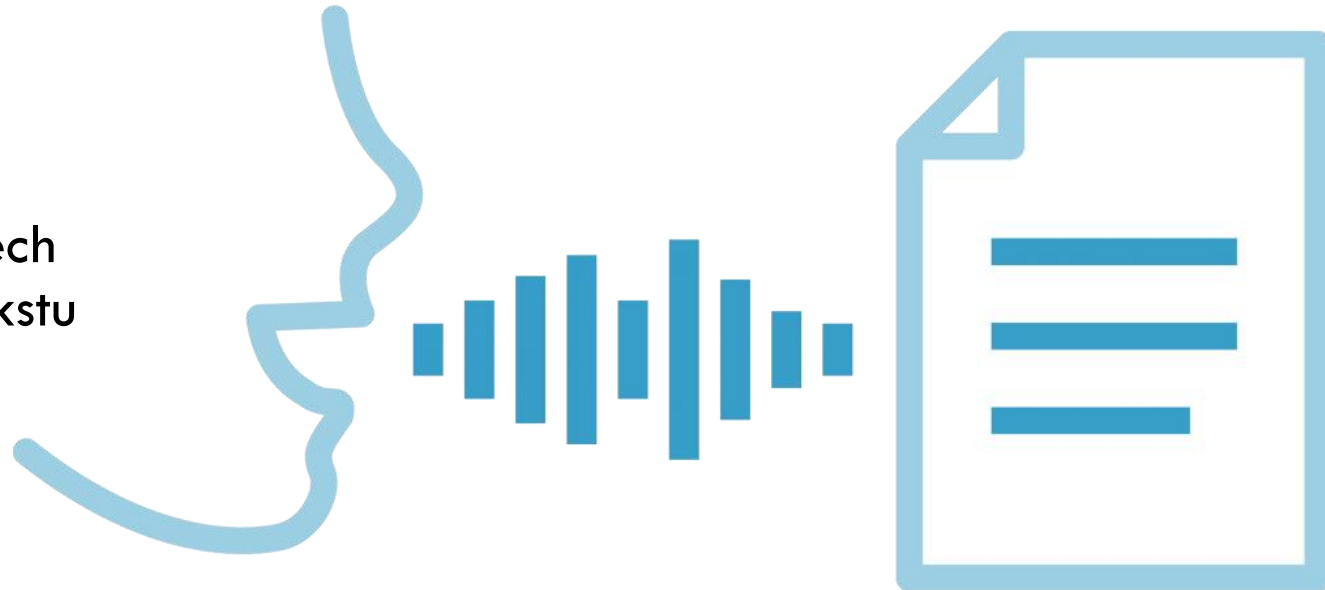


ZASTOSOWANIA - ASR

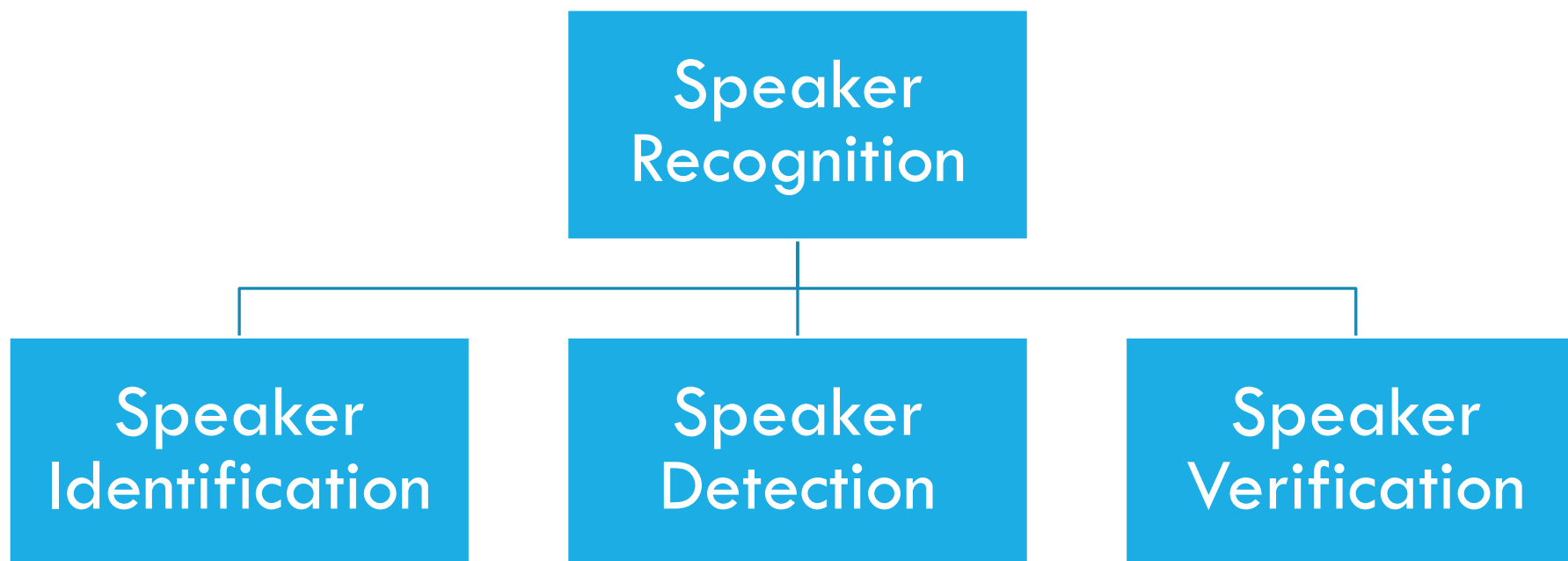
ASR – Automatic Speech Recognition

STT – Speech-To-Text

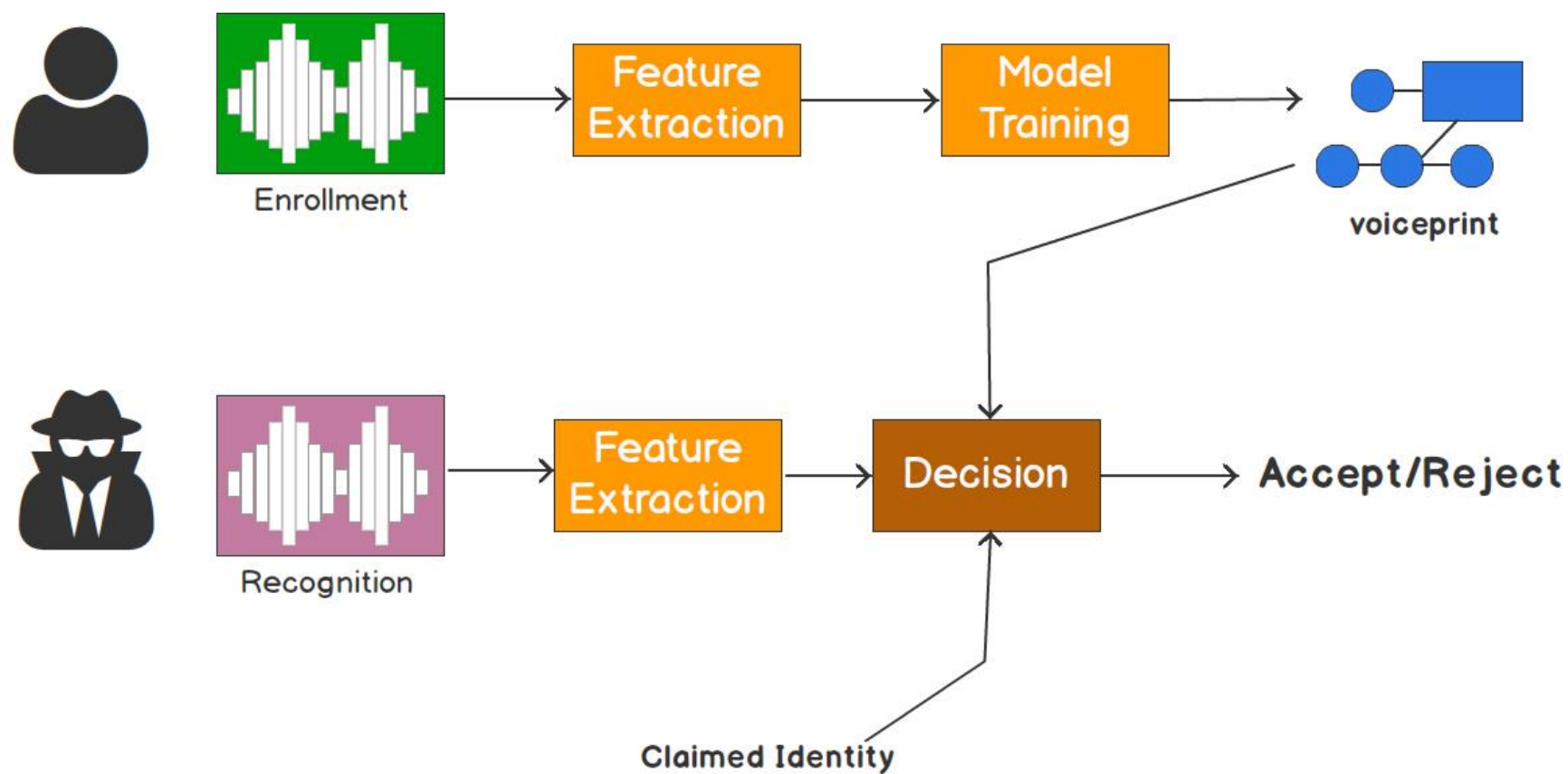
LVCSR – Large Vocabulary Continuous Speech
mowa ciągła, dyktowanie tekstu



ZASTOSOWANIA – ROZPOZNAWANIE MÓWCY



WERYFIKACJA MÓWCY

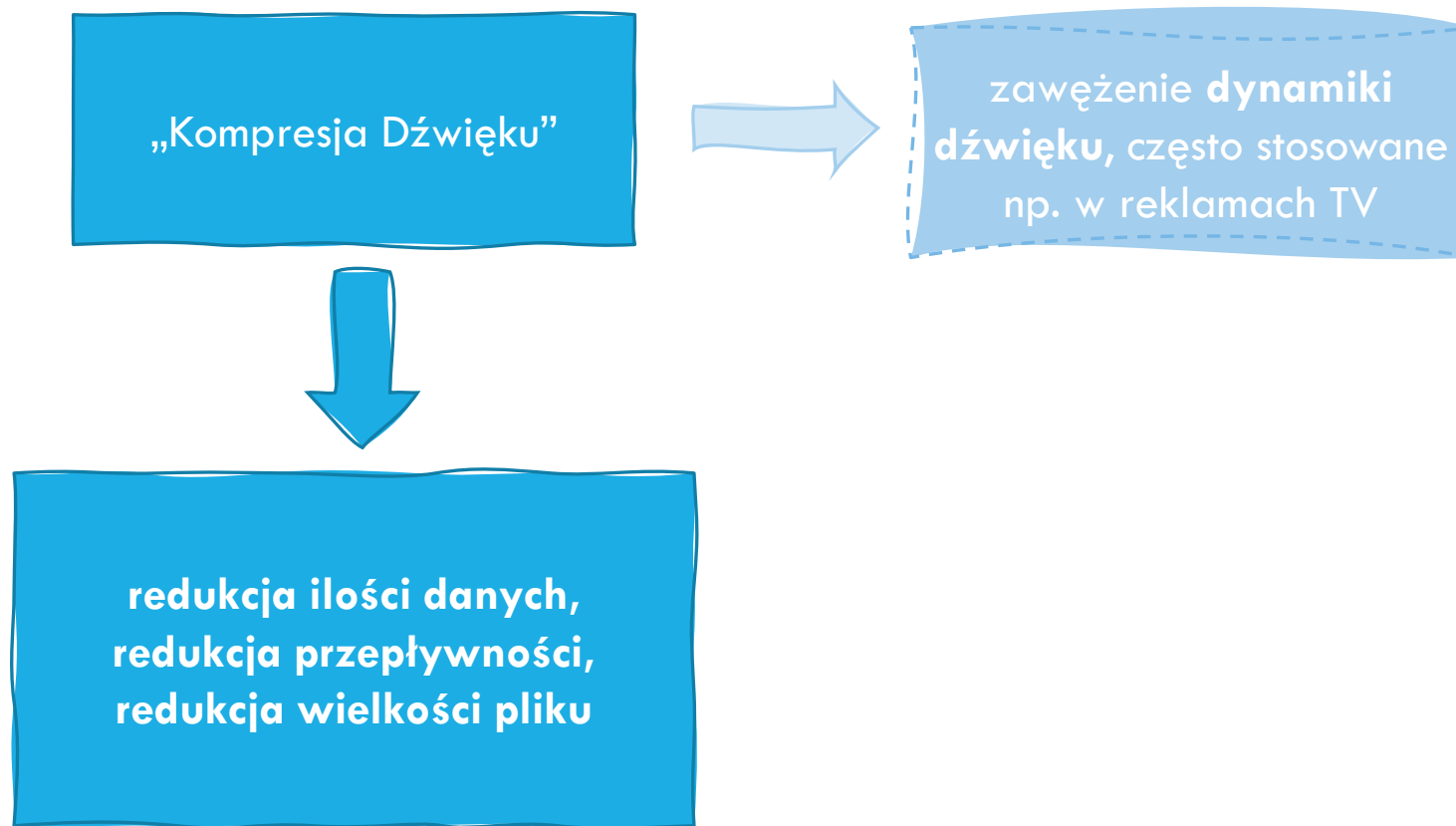


ROZPOZNAWANIE EMOCJI W GŁOSIE



Emocja [%]	ASC	ASF	Roll-off	ZC	RMS
Neutralna	98,52	48,26	34,93	34,53	68,23
Radość	97,87	52,13	38,66	31,81	67,24
Smutek	95,69	47,42	37,84	30,15	62,51
Złość	70,47	27,53	30,92	18,32	47,83
Smutek	93,36	53,77	33,36	24,36	53,27
Strach	96,55	43,29	32,67	21,84	56,68
Wszystkie	79,39	41,23	35,73	27,27	62,26

KOMPRESJA DŹWIĘKU





KOMPRESJA DŹWIĘKU

1. Kompresja umożliwia zwiększenie ilości danych zapisanych na danym nośniku.
2. W systemach transmisyjnych kompresja umożliwia zmniejszenie szerokości pasma, które generalnie skutkuje obniżeniem kosztów transmisji.
3. Kompresja umożliwia szybszą transmisję w danym paśmie.

RODZAJE KOMPRESJI

BEZSTRATNA
(.wav, .flac)

STRATNA
(.mp3, .ogg, .wma)

KOMPRESJA BEZSTRATNA

- bez utraty informacji o sygnale pierwotnym
 - kodowanie ciszy poprzez czas jej trwania, a nie wartość poszczególnych próbek
 - kodowanie mono z wyodrębnieniem i oddzielnym kodowaniem różnic między kanałami
-
- FLAC – Free Lossless Audio Codec – redukcja do 50-70% oryginalnego pliku
 - WMA (Windows Media Audio) 9 Lossless – kompresja 2:1 lub 3:1
 - MLP – Meridian Lossless Packing – kompresja 1,5:1
 - M4A – Apple Lossless – kompresja 2:1
 - APE – Monkey's Audio – redukcja do 40-50% rozmiaru oryginalnego pliku

KOMPRESJA STRATNA

- utrata informacji o sygnale pierwotnym (czasem znacząca)
 - usuwanie skrajnie wysokich i niskich częstotliwości
 - usuwanie dźwięków o niskim poziomie
 - wykorzystywanie psychoakustycznych modeli
-
- MP1, MP2, MP3
 - AAC
 - MPC (Musepack)
 - OGG (Vorbis)
 - ATRAC (używany w Minidisc)