Grafika Komputerowa i Multimedia

Wykład 13
Stratna kompresja dźwięku

Damian Grela
e-mail: **dgrela@pk.edu.pl**
http://www.dgrela.pl

Stratna kompresja dźwięku

- Metody
  - Modulacja Delta
  - DPCM
  - Metody Transformacyjne
  - Kodowanie podpasmowe
  - Schemat analiza — synteza

Modulacja Delta

- Modulacja Delta, (modulacja DM) - jest to próbkowanie sygnału informacyjnego z dużą częstotliwością.
- Koncepcja Modulacji Delta polega na zmniejszeniu dynamiki zmian wartości kolejnych próbek, a zarazem na zwiększeniu korelacji między tymi próbkami.

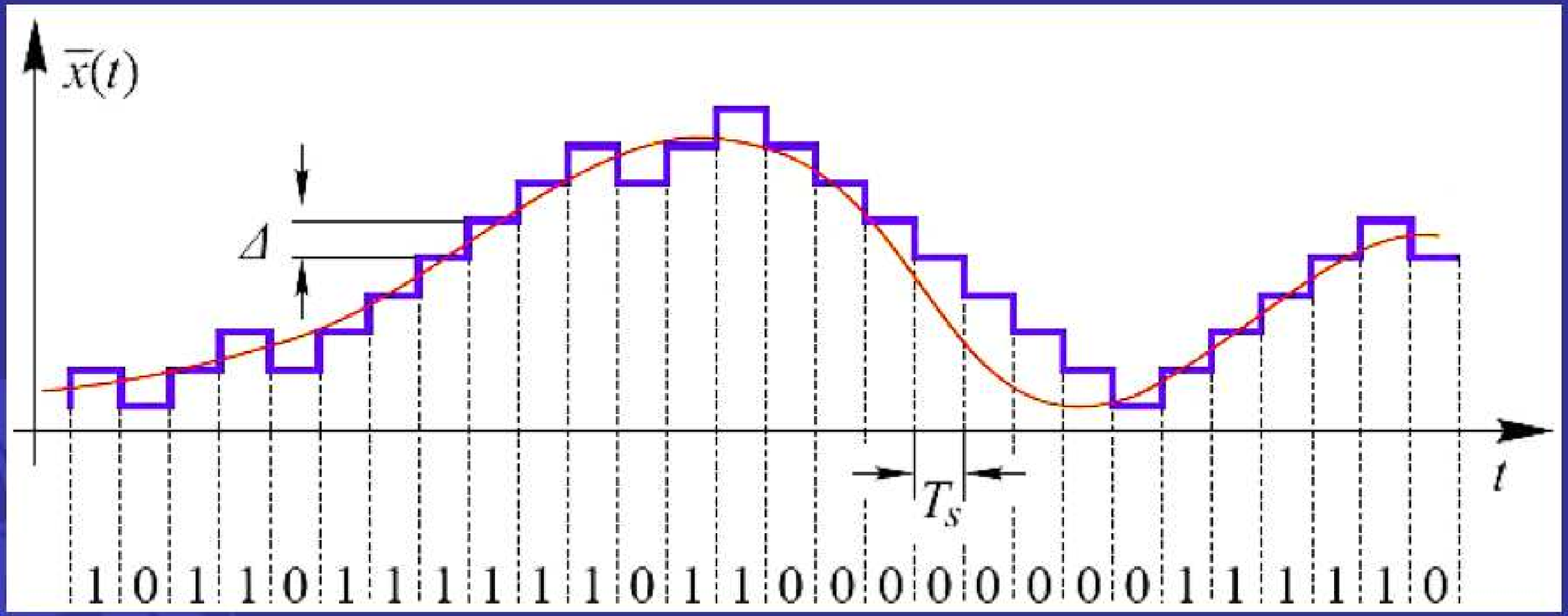
Modulacja Delta

- Częstotliwość DM jest znacznie większa od częstotliwości Nyquista, co skutkuje tym, że nie ma wyraźnych zmian sygnału między dwoma sąsiadującymi próbkami, a także różnica między wartościami tych próbek jest nieznaczna.
- Zmiany wartości sygnału od próbki do próbki są wówczas na tyle nieznaczne, że informacja o nich (a dokładniej o tym, czy sygnał wzrósł, czy zmalał) może być zakodowana za pomocą tylko jednego znaku binarnego.

Częstotliwość Nyquista

- Częstotliwość Nyquista jest to maksymalna częstotliwość składowych widmowych sygnału poddawanego procesowi próbkowania, które mogą zostać odtworzone z ciągu próbek bez zniekształceń.
- Składowe widmowe o częstotliwościach wyższych od częstotliwości Nyquista ulegają podczas próbkowania nałożeniu na składowe o innych częstotliwościach (zjawisko aliasingu), co powoduje, że nie można ich już poprawnie odtworzyć.
- Inaczej mówiąc, częstotliwość Nyquista jest równa połowie częstotliwości próbkowania, $f_N = f_s / 2$

Modulacja Delta



Modulacja Delta

x_n — próbka n

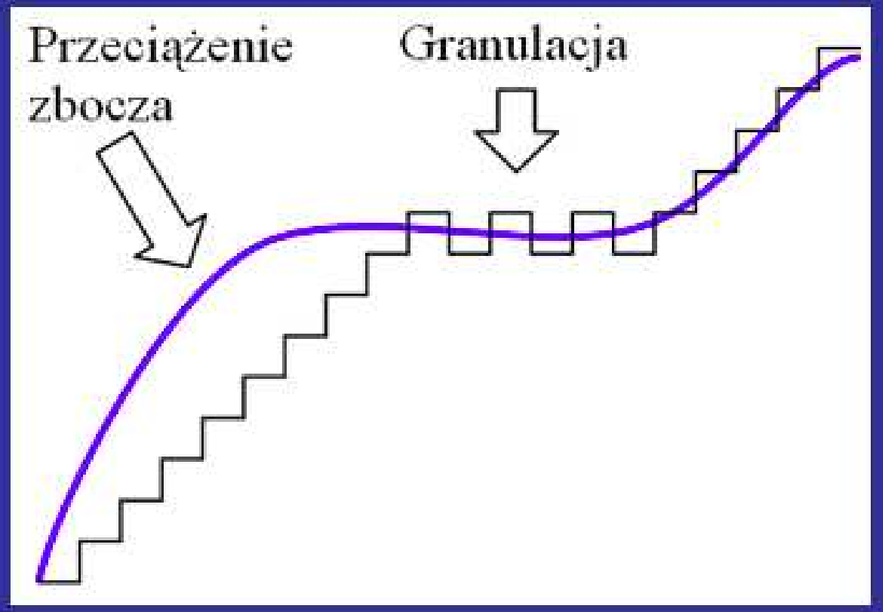
r_n — zrekonstruowana (zdekodowana) próbka n

$e_n = x_n - r_{n-1}$ — sygnał różnicowy (błąd predykcji próbki n)

$d_n = \mathrm{sgn}(e_n)$ — skwantowany sygnał różnicowy transmitowany do odbiorcy

dekodowanie (również kodowanie) sygnału z krokiem kwantyzacji Δ

$$r_n = r_{n-1} + e_n \cdot \Delta$$



DPCM

- Różnicowa modulacja kodowo-impulsowa (DPCM) opiera się na zasadach kodowania sygnału zastosowanych w modulacji kodowo-impulsowej (PCM).
- Różnica polega na tym, że nadajnik DPCM próbkuje otrzymany sygnał a następnie koduje jedynie różnicę pomiędzy próbką rzeczywistą a przewidywaną.

- Słowa kodowe metody DPCM reprezentują różnice pomiędzy próbkami natomiast słowa kodowe metody PCM konkretne wartości próbek.
- Odbiornik odtwarza oryginalny sygnał na podstawie przewidzianej przez siebie wartości oraz otrzymanej różnicy.

- Efektywność DPCM zależy od zdolności algorytmu predykcji (przewidywania), który może być bardziej lub mniej złożony jednak jego wynik zawsze uzyskuje się w oparciu o poprzednie wartości próbek.

- DPCM — Differential Pulse Code Modulation

x_n — próbka n

r_n — zrekonstruowana (zdekodowana) próbka n

p_n — predykcja wartości próbki n

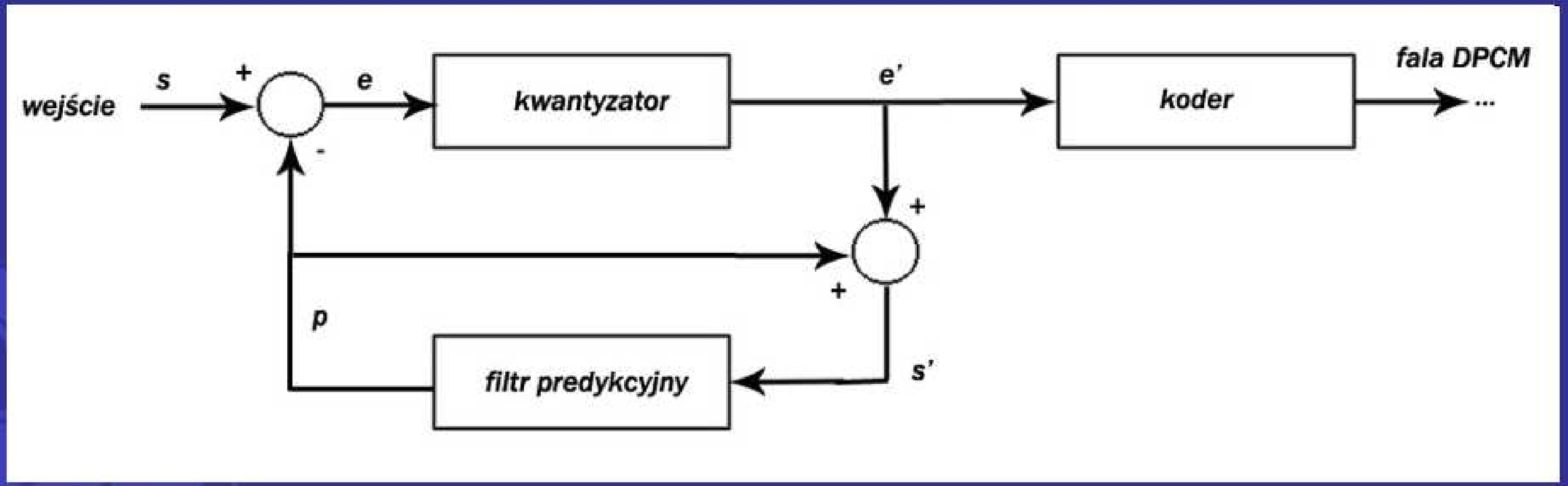
(α_i — współczynniki predykcji):

$$p_n = \sum_{i=1}^{N} \alpha_i r_{n-i}$$

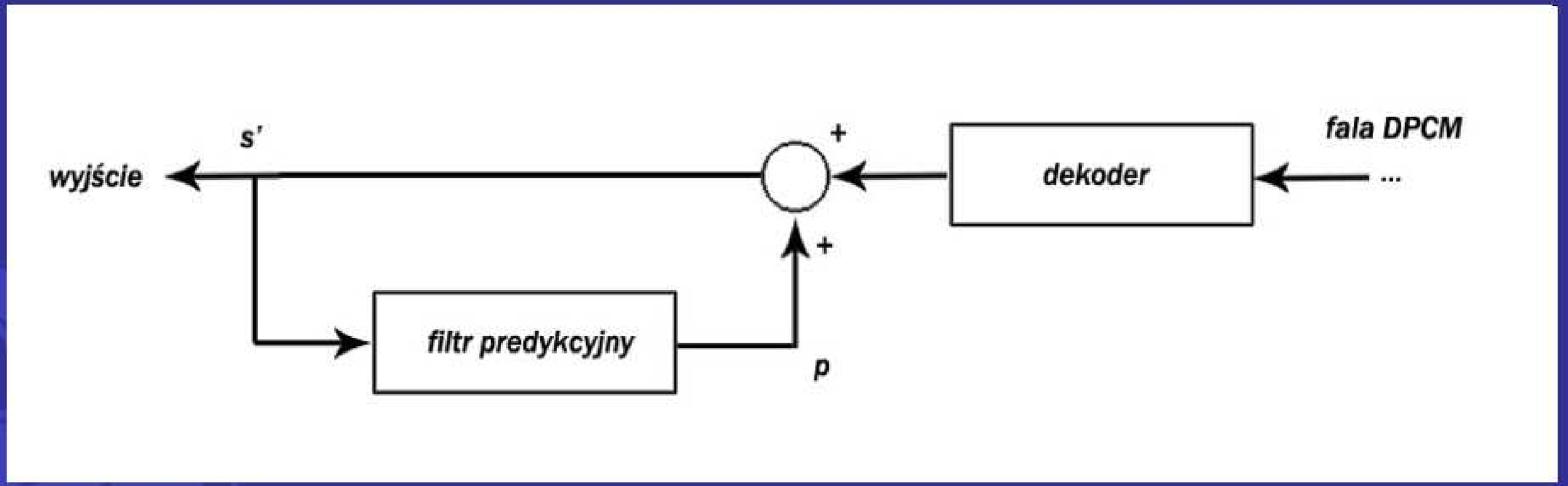
e_n — błąd predykcji próbki n: $e_n = x_n - p_n$

$d_n = Q(e_n)$ — skwantowany błąd predykcji

- Nadajnik DPCM



- Odbiornik DPCM



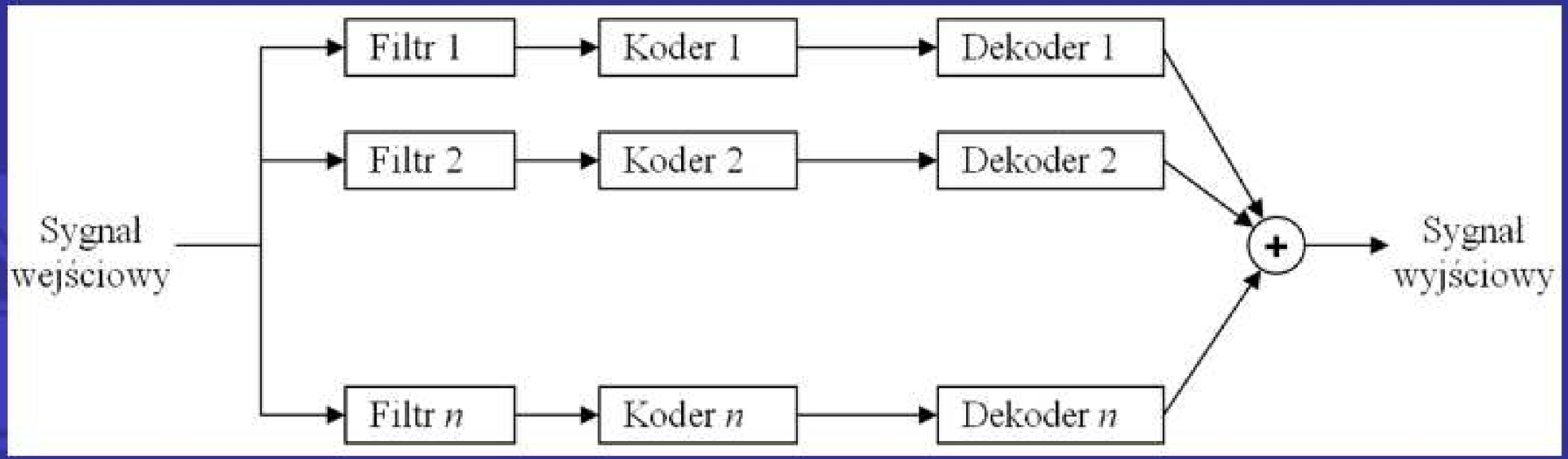
Kodowanie podpasmowe

Ogólny schemat kodowania podpasmowego:

1. Wybierz zbiór filtrów.
2. Używając filtrów, rozłóż sygnał wejściowy na sygnały podpasm.
3. „Rozrzedzanie” (dziesiątkowanie) sygnałów (jeśli częstotliwość większa od tej która wystarcza do odtworzenia rozważanych częstotliwości, wybieramy tylko „co p-tą” próbkę)
4. Kwantyzacja sygnałów, lub kodowanie różnicowe i kwantyzacja:
   - niektóre sygnały mogą być mniej zauważalne (np. te o większych częstotliwościach): stosujemy dla nich mniej dokładną kwantyzację;
   - wynikowe sygnały mogą mieć różne własności statystyczne: możemy każdy z nich inaczej kodować;
   - efekt maskowania: niektóre składowe możemy pomijać w określonych przedziałach.
5. Kodowanie wynikowych sygnałów.

Kodowanie podpasmowe

- Dekompozycja sygnału na składowe za pomocą banku filtrów
  - niżej filtry analityczne, (istnieją również syntetyczne)



Kodowanie podpasmowe

Schemat dekodowania:

1. Dekodowanie poszczególnych ciągów skwantyzowanych (algorytm bezstratny);
2. Odtworzenie wartości (przybliżonych) poszczególnych podpasm (dekodowanie DPCM lub dekodowanie odpowiedniego algorytmu kwantyzacji);
3. Zagęszczenie (uzupełnienie ciągów zerami) tak, aby uzyskać częstotliwość ciągu wejściowego;
4. Zastosowanie filtrów syntezujących (odpowiednio dopasowanych do filtrów stosowanych w trakcie kodowania).

Główna idea

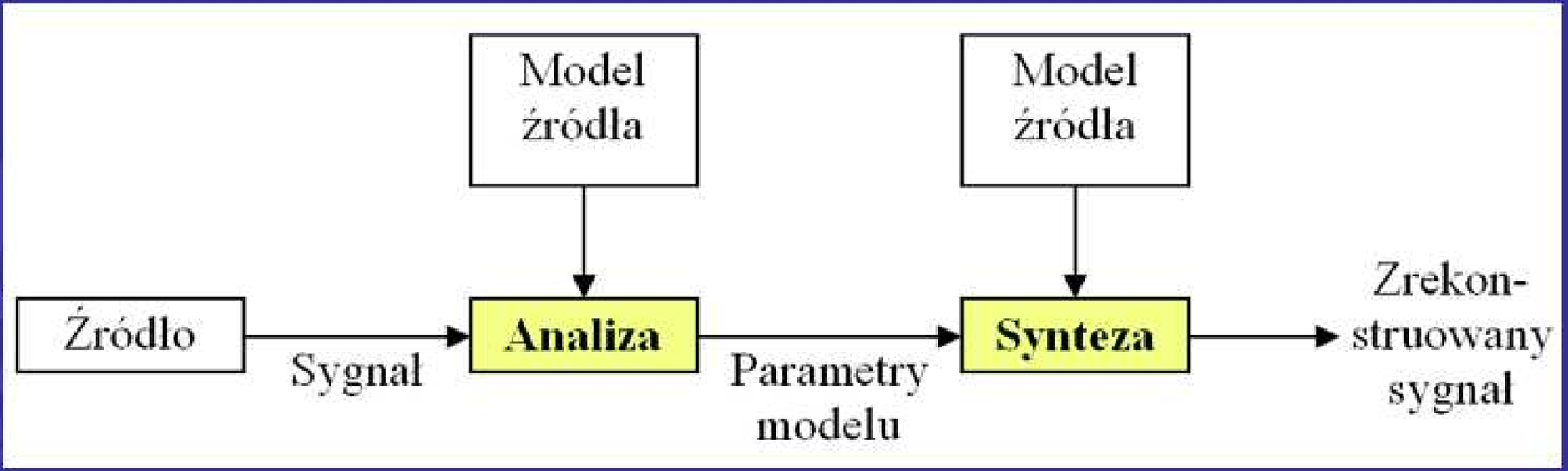
- Zamiast otrzymywać przybliżenia kolejnych wartości sygnału źródłowego modelujemy sygnał wejściowy i przesyłamy parametry modelu.
- Odbiornik na podstawie otrzymanych parametrów syntetyzuje sygnał

Schemat analiza-synteza

- Metoda stosowana najczęściej do kompresji mowy.
- Stosujemy model symulujący działanie narządu mowy wysyłając takie parametry jak dźwięczność czy bezdźwięczność, szybkość przepływu powietrza czy naprężenie strun głosowych.

Schemat analiza-syntezy

- Odmienne podejście: nie kodujemy sygnału
- zamiast tego analizujemy sygnał i na jego podstawie wyznaczamy/szacujemy parametry modelu



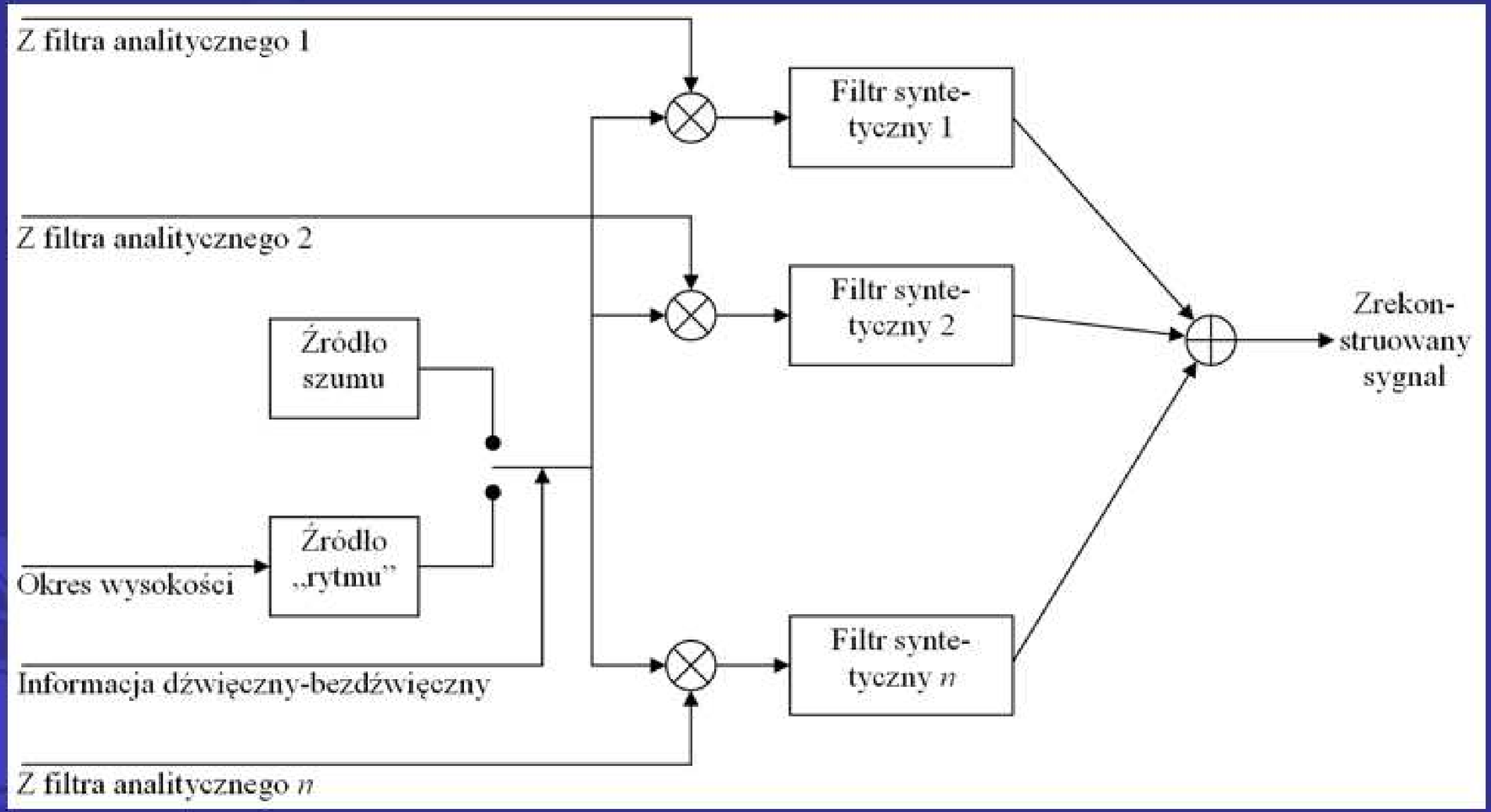
(na podstawie: K. Sayood, Kompresja danych. Wprowadzenie, RM, W-wa, 2002)

- Analiza
  - zastosowanie banku filtrów środkowoprzepustowych (jak w kodowaniu podpasmowym), mierzona jest energia sygnału w pasmach odpowiednich filtrów
  - detekcja głównej składowej harmonicznej (tzw. okres wysokości dźwięku)

Wokoder kanałowy

- detekcja głosek dźwięcznych (dominujące składowe harmoniczne) i bezdźwięcznych (dominujący szum)
- analiza przeprowadzana i parametry przekazywane do dekodera z pewną częstotliwością (np. 50 Hz)
- (algorytm obecnie ma znaczenie historyczne)

Wokoder kanałowy - synteza



(na podstawie: K. Sayood, Kompresja danych. Wprowadzenie, RM, W-wa, 2002)

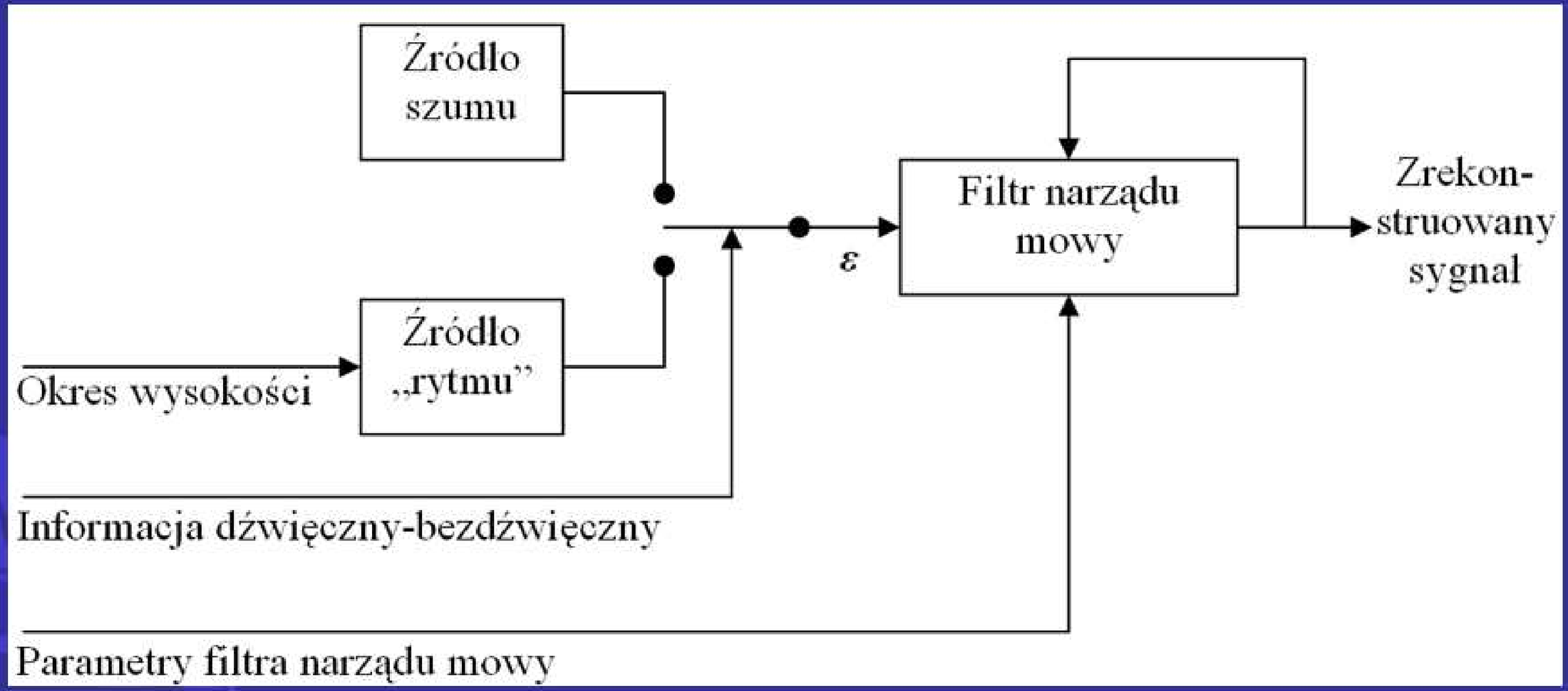
- LPC-10
  - Oparty o schemat analiza — synteza oraz o predykcję (LPC — linear predictive coder)
    (istnieją również algorytmy oparte wyłącznie o predykcję)
  - Standard rządowy (USA) dla kodowania dźwięku z prędkością 2.4 kbps
  - Kodujemy dźwięk próbkowany 8000 razy na sekundę
  - Kodujemy bloki po 180 próbek (22.5 ms)

- Analiza
  - detekcja głównej składowej harmonicznej (tzw. okres wysokości dźwięku)
  - detekcja głosek dźwięcznych i bezdźwięcznych (na podstawie częstości przejść przez 0)
  - zastosowanie pojedynczego filtra (filtr narządu mowy)

$$y_n = \sum_{i=1}^{M} b_i y_{n-i} + G\varepsilon$$

y_j — j-ta próbka, b_m — m-ty parametr filtra,
ε — sygnał z generatora, G — tzw. wzmocnienie filtra

LPC-10 - dekoder



- Wady algorytmu
  - Wyraźna mowa przy 2.4 kbps, ale sztuczna barwa głosu
    - przyczyną jest użycie tylko dwóch generatorów
    - można zastosować kilka generatorów rytmów (algorytm CELP)
  - Szum tła może wprowadzić w błąd koder, co powoduje utratę informacji o składowych harmonicznych dźwięku i w konsekwencji niezrozumiałość dekodowanej mowy
    (w LPC-10 zastosowano prosty detektor dźwięczności głosek)

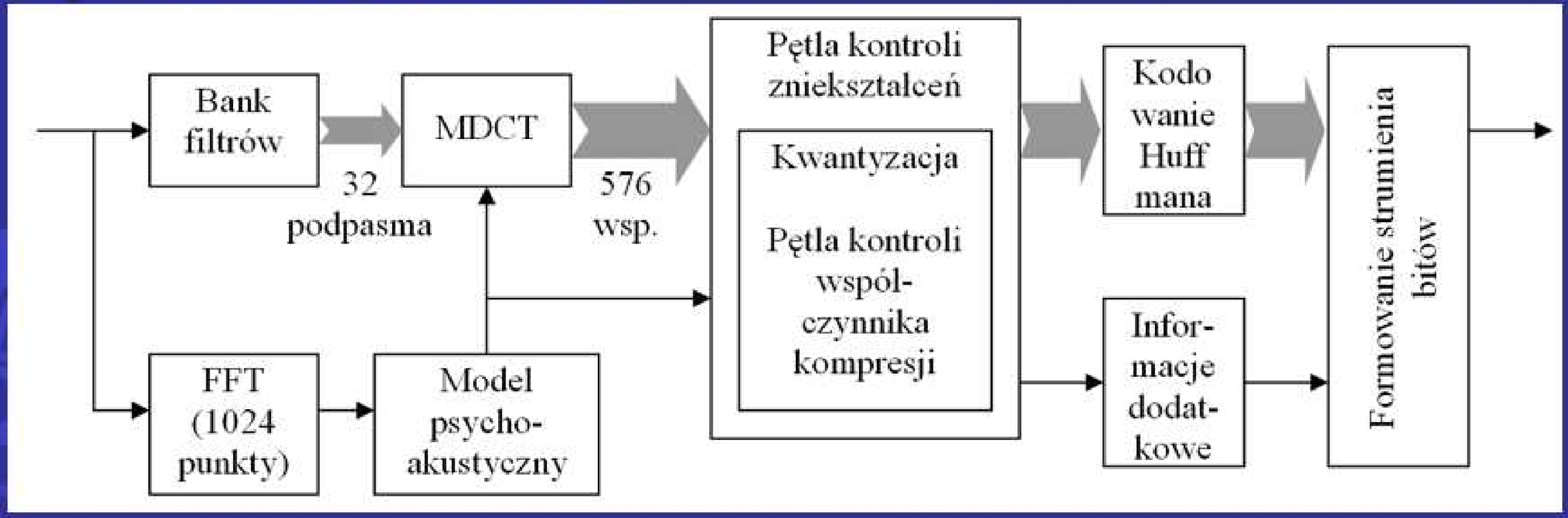
- Wady algorytmu
  - W niektórych zastosowaniach opóźnienie 20ms może być zbyt duże
    - standard CCITT G.728 (na bazie CELP):
      - opóźnienie 2ms — blok zawiera 5 próbek, dźwięk 8000 próbek na sekundę, 16 kbps
      - zastosowanie adaptacji wstecz — współczynniki filtra dla danego bloku są obliczane na podstawie poprzedniego bloku

- MP3 to MPEG-1/2 Layer 3
  - element standardu kompresji wideo MPEG 1/2 (kompresja wideo na następnym wykładzie)
  - MPEG 1 (1992) zawiera specyfikacje MPEG 1 Audio:
    - Layer 1 i Layer 2 — niższa złożoność i niższa jakość
    - Layer 3 — większa złożoność i wysoka jakość,
      - optymalizowana dla przepływności ok. 128 kbps (dla sygnału stereo)
        - » dostępne przepływności od 32 do 320 kbps
      - tryby mono, stereo, joint stereo i dual channel (2x mono, np. wersje językowe)
      - dla dźwięku próbkowanego z częstotliwościami 32 kKz, 44.1 kHz i 48 kHz

- MP3 to MPEG-1/2 Layer 3
  - MPEG 2 (1994) zawiera rozszerzoną specyfikację Audio Layer 3
    - niższe (o połowę) częstotliwości próbkowania (16, 22.05 i 24 kHz)
      - oraz niższe przepływności 8 do 160 kbps
    - dźwięk w formacie 5.1

(na podstawie: K. Brandenburg, MP3 and AAC explained. AES 17th Int. Conf. on High Quality Audio Coding)

- Przykładowa struktura kodera

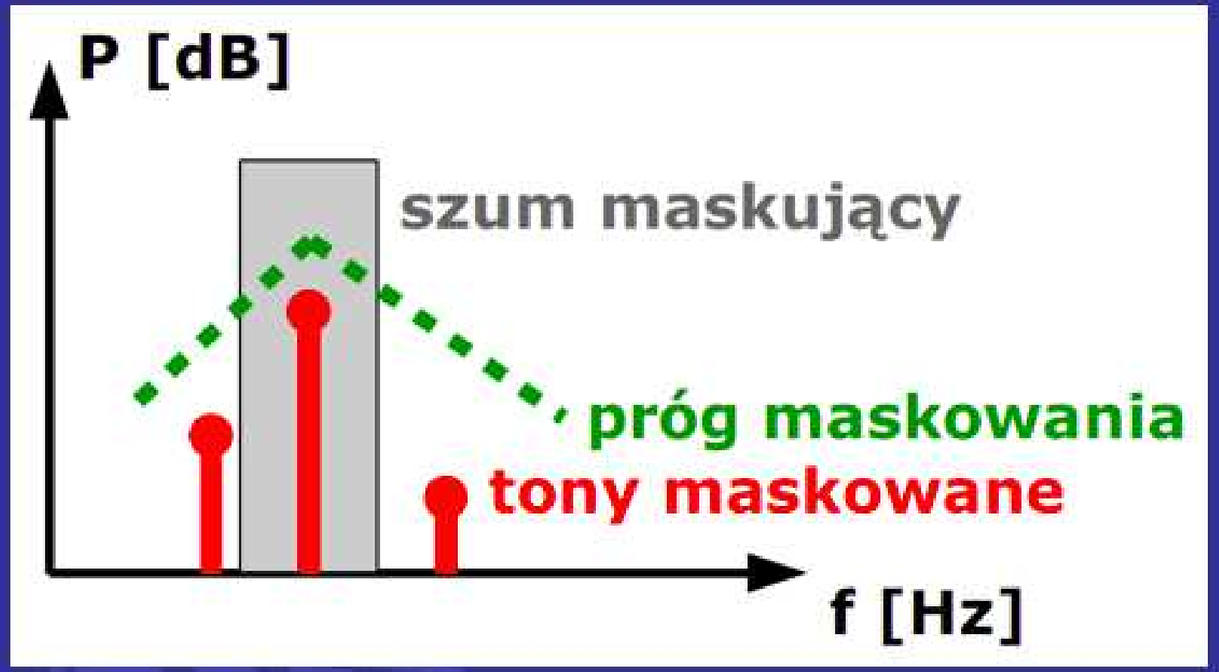


MP3 – bank filtrów i MDCT

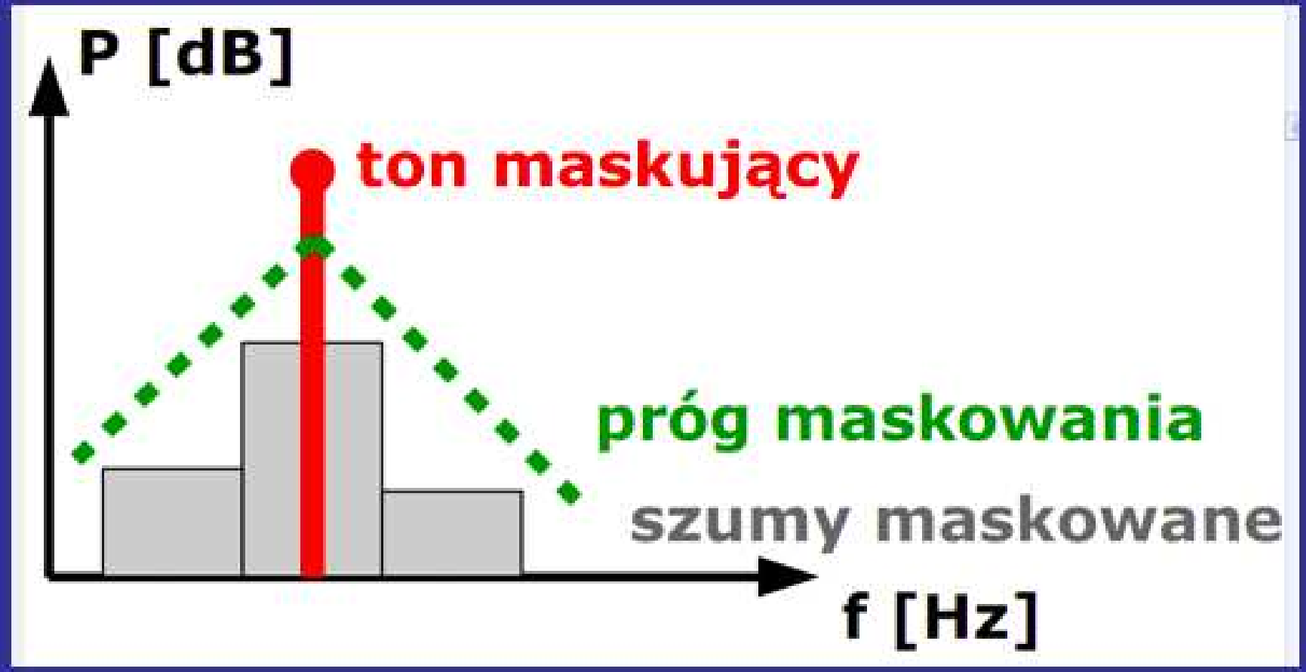
- Bank filtrów
  - 32 filtry stosowane również w Layer 1 i 2
- Podpasma uzyskane z banku filtrów przekształcane są zmodyfikowaną transformatą kosinusową
  - Modified Discrete Cosint Transform (MDCT)
  - MDCT generuje 18 współczynników dla każdego podpasma
    - 32x18=576

- W algorytmie mp3 zastosowano model psychoakustyczny słuchu ludzkiego oparty o zjawisko maskowania składowych dźwięku.
  - w dziedzinie częstotliwości
  - w dziedzinie czasu
- Polega ono na tym, że nie słyszymy pewnych słabych dźwięków, w obecności silniejszych, maskujących je.

Model psychoakustyczny



maskowanie tonów szumem



maskowanie szumów tonem

Model psychoakustyczny

- Zjawisko maskowania psychoakustycznego może zostać w koderze wykorzystane w dwojaki sposób:

1. Korzystając z wiedzy o zamaskowanych składowych, można je usunąć z sygnału.

2. Korzystając z wiedzy o charakterze wprowadzanych przez kodek zniekształceń, można je uczynić niesłyszalnymi (noise shaping)

Model psychoakustyczny

- Zjawisko maskowania jest dość złożone.
- Kształt krzywej maskowania zależy od częstotliwości składowych maskujących i maskowanych i charakteru składowych (szum/sygnał).
- Oprócz tego mamy do czynienia ze zjawiskiem maskowania czasowego (sygnał może być maskowany jakiś czas po sygnale maskującym, a nawet przed (!)).
- Oprócz tego u każdego człowieka maskowanie zachodzi w nieco inny sposób.
- Uśredniony i zapisany matematycznie model zjawiska maskowania nosi nazwę modelu psychoakustycznego.

Analiza psychoakustyczna

- Ponieważ model psychoakustyczny jest zdefiniowany w dziedzinie częstotliwości, koder musi dokonać transformacji sygnału wejściowego w tę dziedzinę.
- Odbywa się to za pomocą szybkiej, 1024-punktowej transformaty Fouriera.
- Na bazie transformaty dokonywane jest wydzielenie składników szumowych i tonalnych i obliczenie progów maskowania.

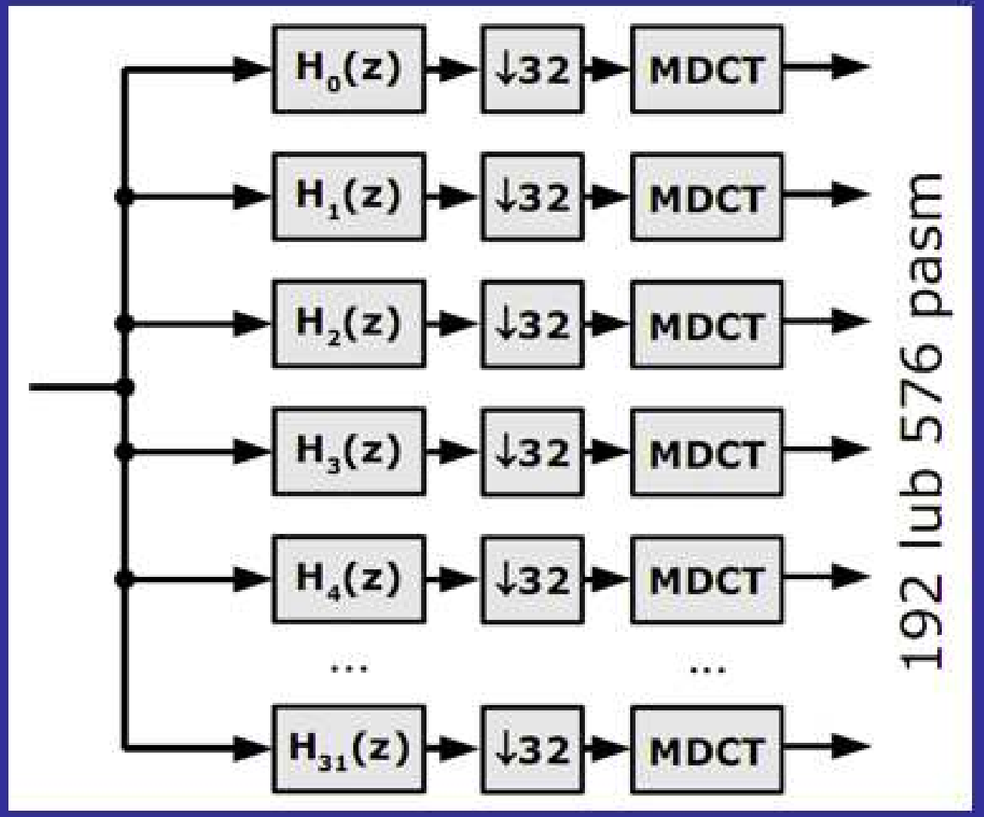
Analiza psychoakustyczna

- Na ich bazie określana jest rozdzielczość transformaty cosinusowej: 6 próbek (duża rozdzielczość czasowa kosztem częstotliwościowej dla przebiegów szybkozmiennych) lub 18 próbek (duża rozdzielczość częstotliwościowa, dla przebiegów wolnozmiennych).
- Progi maskowania określają też współczynniki skalujące i siłę kwantowania poszczególnych współczynników transformaty przed kompresją Huffmana tak, aby z jednej strony zmieścić się w narzuconym bitrate, z drugiej zaś uzyskać możliwie najlepszą jakość sygnału.

- Kodowanie podpasmowe jest esencją standardów MPEG Audio.
- Rozdzielenie sygnału na pasma odbywa się za pomocą banku filtrów.
- Bank filtrów użyty w MP3 posiada 192 albo 576 podpasm (w zależności od wymaganej rozdzielczości czasowej).
- Osiągnięcie takiej ilości pasm za pomocą zwykłych filtrów FIR wymaga zastosowania ogromnej ilości długich filtrów, co oznacza dużą ilość obliczeń.
- Dlatego MP3 używa hybrydowego banku filtrów, pierwszy stopień to typowy bank filtrów FIR o 32 pasmach.

Bank filtrów

- Pierwszym stopniem filtru hybrydowego są 32 filtry FIR. Wszystkie używają tego samego rodzaju okna – okna sinusoidalnego. To okno daje całemu filtrowi odwracalność, to znaczy, że możliwa jest idealna rekonstrukcja sygnału, przez odwrotny filtr syntezujący.



Bank filtrów

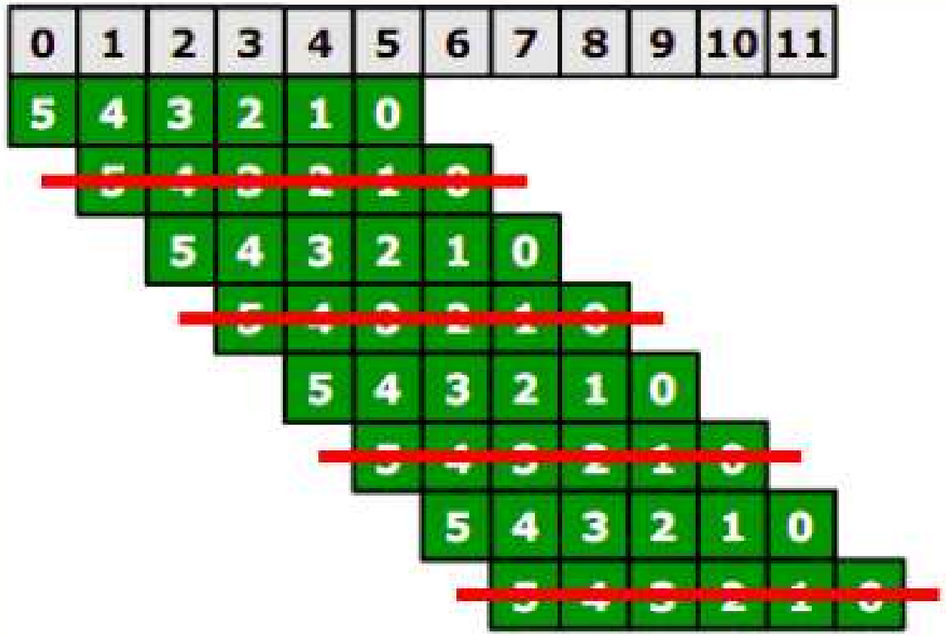
- Po przejściu przez filtr pasmowy, z podpasm usuwane są próbki, tak, że zostaje tylko co 32 (następuje decymacja).
- Z tego względu, każdy filtr oblicza tylko co 32 próbkę, a więc jest filtrem polifazowym.
- Zdecymowane sekwencje próbek są poddawane zmodyfikowanej transformacie cosinusowej typu IV, co zwiększa ilość pasm do 192 lub 576 (długość transformaty jest wybierana przez blok analizy psychoakustycznej).

- Modyfikacja transformaty polega na tym, że obejmuje ona dwa bloki próbek (a więc 12 lub 36), kolejne transformaty zazębiają się o 50%. Dzięki takiemu zazębieniu się, unika się zniekształceń na granicy ramek.

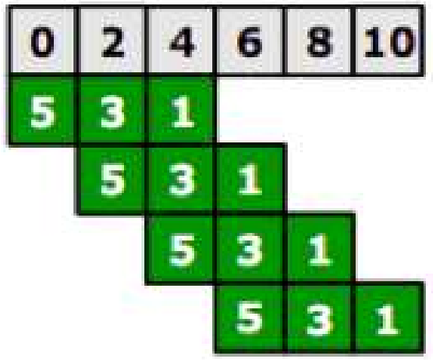
Filtr polifazowy

- Polifazowy filtr FIR to taki, w którym interesuje nas jedynie co któraś próbka sygnału wyjściowego.
- Najczęściej z filtrami polifazowymi mamy do czynienia przy interpolacji, a więc cyfrowym zwiększaniu częstotliwości próbkowania, oraz w bankach filtrów.

Filtr polifazowy

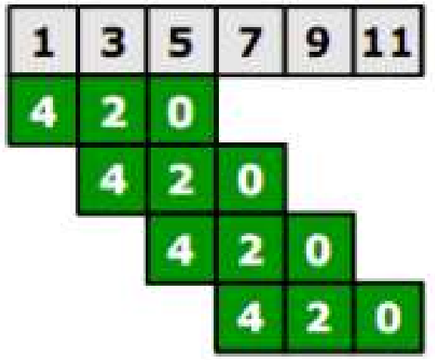


=



faza 0°

+



faza 180°

- Na podstawie modelu określa się dopuszczalny szum kwantyzacji (a właściwie błąd/krok kwantyzacji) danej składowej częstotliwości w danym momencie czasu
  - niektóre składowe mogą być odrzucone w całości, gdyż i tak nie docierają do świadomości odbiorcy
  - inne mogą być zakodowane z małą dokładnością, gdyż błąd/szum kwantyzacji jest poniżej progu percepcji

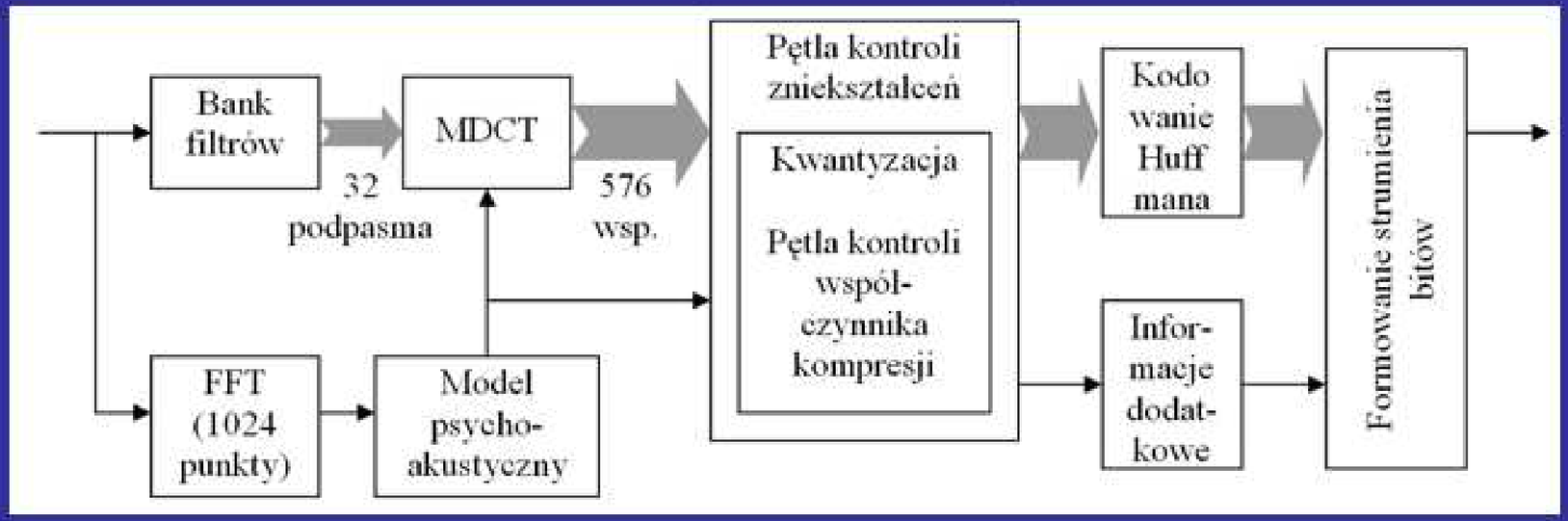
- (w mp3 stosujemy skalarny kwantyzator nierównomierny — skok kwantyzacji rośnie ze wzrostem wartości sygnału)
- model używający FFT to tylko jedna z możliwości; standard definiuje jedynie semantykę i syntaktykę strumienia bitów (dekoder nie używa modelu)
  - model można zbudować w oparciu o bank filtrów, lub zupełnie inaczej
  - istnieje wiele różnych koderów zgodnych z mp3; zgodność z mp3 nie gwarantuje jakości kodowania — istnieją kodery lepsze i gorsze

MP3 - kodowanie

- Oparte o algorytm Huffmana
  - Ze względu na kwantyzację małe wartości występują z dużymi prawdopodobieństwami
  - Kodowane są grupy po 2 (a dla małych wartości po 4) wartości po kwantyzacji
  - Do kodowania różnych podpasm można stosować różne kody (tablice kodów) Huffmana
  - Poszczególne bloki kodowane są niezależnie
    - dopuszczalne są zmiany przepływności (VBR — Variable BitRate)
    - różny krok kwantyzacji → duży zakres dynamiczny (>24 bit)

MP3 - kwantyzacja

- Dobieramy
  - indywidualne współczynniki kwantyzacji dla każdego pasma z osobna
  - oraz globalny mnożnik dla wszystkich współczynników kwantyzacji
  - (stosowana jest kwantyzacja nieliniowa)
  - (to tylko przykładowa metoda doboru parametrów kwantyzacji)



- Pętla kontroli współczynnika kompresji
  - dla poszczególnych pasm przeprowadzana jest kwantyzacja
  - symulowane jest kodowanie skwantowanych współczynników
  - jeżeli wynik kodowania przekracza zadane ograniczenie przepływności to globalny mnożnik jest zwiększany i pętla wykonywana jest ponownie

- Pętla kontroli zniekształceń
  - Rozpoczynamy od ustawienia mnożników indywidualnych współczynników na 1
  - Jeżeli błąd kwantyzacji dla danego pasma przekracza oszacowany przez model próg percepcji dla tego pasma to odpowiednio zmieniamy jego indywidualny współczynnik kwantyzacji

MP3 - kwantyzacja

- Nie zawsze możliwe jest jednoczesne uzyskanie zadanej przepływności i spełnienie wymagań narzuconych przez model psychoakustyczny
  - pętle mogłyby się wykonywać w nieskończoność, aby do tego nie dopuścić pętla kontroli zniekształceń może być przerwana mimo nie spełnienia wymagań modelu
  - niekiedy możliwe jest spełnienie obu wymagań jednocześnie i to zapasem → VBR

Tryby kompresji

- Każda ramka MP3 zawiera 1152 próbki (2304 przy stereo).
- Długość ramki w bajtach zależy od założonego stopnia kompresji (wyrażanego jako przepływność skompresowanego strumienia w kbit/s).
- Ze względu na to, że każda ramka może mieć inną docelową przepływność, możliwe są 3 tryby kompresji.

CBR (Constant BitRate)

- Wszystkie ramki mają tę samą przepływność, równą średniej przepływności całego strumienia.
- Rozmiar pliku jest proporcjonalny do czasu trwania utworu.

ABR (Average BitRate)

- Ramki mają różną przepływność dobraną tak, aby skompresować utwór z możliwie małymi stratami, zachowując zadaną średnią przepływność.
- Rozmiar pliku jest w przybliżeniu proporcjonalny do czasu trwania utworu.

VBR (Variable BitRate)

- Ramki mają różną przepływność dobraną tak, aby zachować stałą jakość kompresji.
- Rozmiar pliku nie jest znany przed zakończeniem kompresji i nie jest proporcjonalny do czasu trwania utworu.

- MP3 AAC to MPEG-2 Layer 3 AAC (Advanced Audio Coding)
  - Rozszerzenie standardu MEPG 2 z roku 1997
  - Zastosowanie dodatkowo predykcji (wstecznej)
  - Udoskonalony tryb joint-stereo
  - Udoskonalone kodowanie (częstsze kodowanie czwórek symboli)

- Większa rozdzielczość w dziedzinie częstotliwości i czasu
  - dekompozycja składowych bankiem filtrów MDCT generującym 1024 współczynniki
  - poprawa odpowiedzi impulsowej filtra (dla krótkich bloków i 48 kHz) z 18.6 ms do 5.3 ms (redukcja efektu pre-echa)
- Technika TNS (Temporal Noise Shaping)
  - kontrola błędu kwantyzacji w dziedzinie czasu dająca przede wszystkim poprawę jakości rekonstrukcji mowy dla małych przepływności
- W porównaniu do mp3, AAC daje taką samą jakoś przy przepływności mniejszej o 30% (za K. Brandenerburg)

Dziękuję za uwagę...

Damian Grela
e-mail: **dgrela@pk.edu.pl**
http://www.dgrela.pl