

Podstawy kompresji treści multimedialnych



Opracował:
dr inż. Piotr Suchomski



Wprowadzenie

- Dane multimedialne to przede wszystkim duże strumienie danych liczone w MB a coraz częściej w GB;
- Mimo dynamicznego rozwoju technologii pamięci i coraz szybszych transferów danych osiąganym w systemach komunikacyjnych istnieje potrzeba stosowania wydajnych algorytmów kompresji danych multimedialnych.
- Dzięki kompresji dostęp do danych jest szybszy.
- Kompresja wpływa na jakość dostarczanej informacji



Klasyfikacja metod kompresji

■ Metody bezstratne

- Zakodowany strumień danych po dekompresji jest identyczny z oryginalnymi danymi przed kompresją,

■ Metody stratne

- W wyniku kompresji część danych (mniej istotnych) jest bezpowrotnie tracona, dane po dekompresji nieznacznie różnią się od oryginalnych danych przed kompresją.



Klasyfikacja metod kompresji

- Entropijne (ang. *entropy coding*) – nie uwzględniają korelacji występujących w kodowanym strumieniu danych;
- Źródłowe (ang. *source coding*) – w procesie kompresji wykorzystywana jest korelacja kodowanego strumienia danych;
- Hybrydowe (ang. *hybrid coding*) – kombinacja metod entropijnych i źródłowych;



Entropia

- Podatność zbioru danych na kompresję można określić obliczając entropię danych wejściowych:

$$E_{WE} = - \sum_{i=1}^n P(z_i) \times \log_2 P(z_i)$$

$P(z_i)$ – prawdopodobieństwo wystąpienia symbolu z_i w strumieniu danych.



Entropia - interpretacja

- Wartość entropii jest nieujemna,
- Maksymalna wartość gdy prawdopodobieństwa wystąpienia wszystkich symboli w strumieniu danych są takie same.
- Generalnie wartość entropii określa minimalną długość kodu do kodowania jednego symbolu w strumieniu danych.
- Metoda kompresji jest tym skuteczniejsza im minimalna długość osiągniętego kodu dla pojedynczego symbolu jest zbliżona do wartości entropii.

Współczynnik kompresji

- Charakterystyczny parametr kompresji określający stopień redukcji danych wejściowych:



- L_{WE} – rozmiar danych przed kompresją,
- L_{WY} – rozmiar danych po kompresji



Kompresja bezstratna

- Kodowanie długości serii,
- Metoda Huffmana,
- Kodowanie arytmetyczne,
- Metody słownikowe



Kodowanie długości serii

- Metoda nie potrzebuje danych początkowych o danych wejściowych,
- Polega na zastępowaniu serii takich samych danych występujących po sobie rekordem danych zawierającym kod wzorca oraz liczbę jego wystąpień.
- Przykładem takiej metody jest metoda RLE (Run-Length Encoding) stosowana do kodowania obrazów.



Przykład kompresji RLE

```
aaaaaabbbbaaaaaaa  
aaaaaabbaabbaaaaa  
aaaaabbbaaabbbaaaa
```

Bez kompresji: 3 x 18 bajtów = 54 bajty

Sposób kodowania:

a7b3a8 - 6 bajtów

a6b2a2b2a6 - 10 bajtów

a5b3a3b3a4 - 10 bajtów

Stopień kompresji: 2:1



Właściwości kodowania długości serii

- Osiągany niezbyt duży stopień kompresji,
- Bardzo szybka kompresja i dekompresja,
- Zakodowany strumień danych jest mało czuły na zakłócenia



Metoda Huffmana

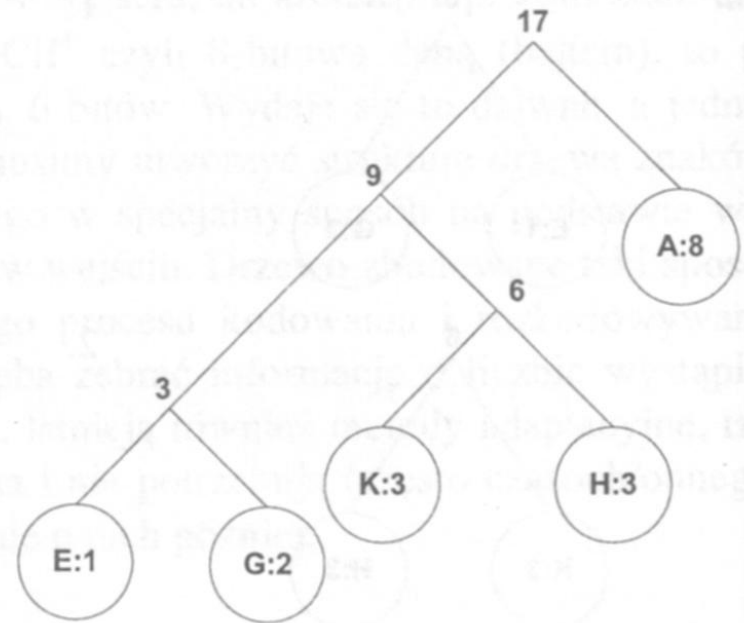
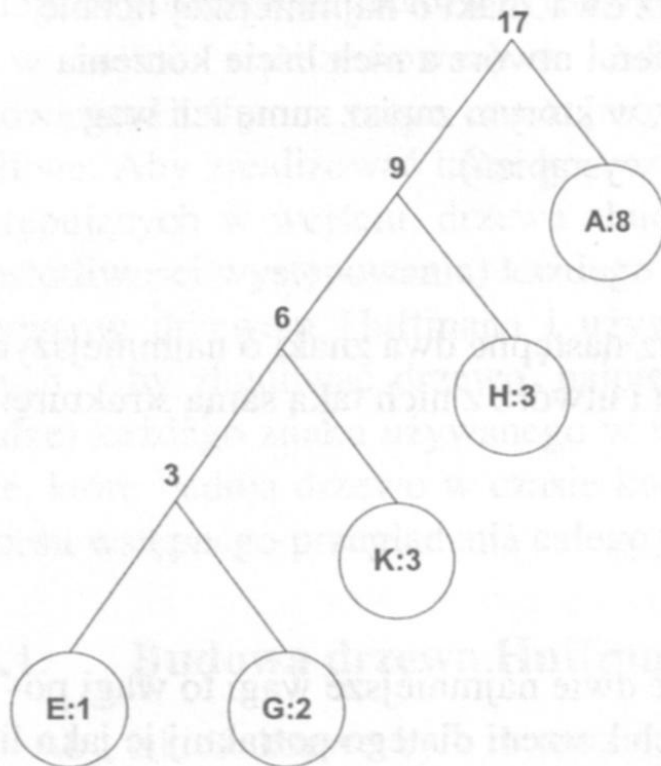
- Opracowana w 1960 roku przez Davida Huffmana, opiera się na statystycznym rozkładzie prawdopodobieństwa wystąpienia symboli w strumieniu danych wejściowych.
- Celem metody jest redukcja nadmiarowości kodu do reprezentacji poszczególnych symboli.
- Podstawą metody jest budowa drzewa Huffmana. Drzewo budowane jest od liści do korzenia. Każdy liść reprezentuje pojedynczy symbol z dziedziny danych. Wartość liścia oznacza liczbę wystąpień symbolu w strumieniu danych.



Metoda Huffmana

- Do każdego znaku prowadzi tylko jedna droga. Jeżeli rozgałęzienia na prawą stronę oznaczone zostaną 0 a na lewą 1 to dla każdego znaku można określić unikatowy kod binarny.
- Dla danej serii danych można zbudować więcej niż jedno drzewo, co oznacza, że w czasie dekompresji potrzebne jest drzewo według, którego dane zostały zakodowane.

Metoda Huffmana





Metoda Huffmana - właściwości

- Jedna z najczęściej wykorzystywanych metod kodowania (również jako element kompresji stratnej),
- Osiągany stopień kompresji rzędu 20-40%,
- Działa niezależnie od typu danych,
- Wolniejsze działanie, ze względu na konieczność budowania drzewa (zwłaszcza w wersji adaptacyjnej),
- Duża czułość na zakłócenia, wymaga specjalnych zabezpieczeń, aby nie nastąpiło przekłamanie w danych zakodowanych.



Kodowanie arytmetyczne

- Podobnie jak metoda Huffmana opiera się na rozkładzie prawdopodobieństwa wystąpienia symbolu w strumieniu danych wejściowych.
- 1 symbol może opisywać ułamkową część bitu (1 bit może nieść więcej informacji niż o 1 znaku),
- Ideą tego kodu jest przedstawienie ciągu symboli jako podprzedziału przedziału jednostkowego $[0, 1)$ wyznaczonego rekursywnie na podstawie prawdopodobieństw wystąpienia tych symboli w strumieniu danych wejściowych. Ciąg kodowy reprezentujący kodowane symbole jest binarnym zapisem wartości z wyznaczonego w ten sposób przedziału.



Kodowanie arytmetyczne

- Metoda daje lepszy stopień kompresji niż metoda Huffmana o kilka, kilkanaście procent.
- Kodowanie arytmetyczne jest wolniejsze niż kodowanie Huffmana.



Metody słownikowe

- Polega na przeglądaniu danych wejściowych w poszukiwaniu fraz identycznych z tymi, które są w słowniku. Jeśli taka fraza zostanie wykryta to w jej miejsce w strumieniu danych wyjściowych wpisywana jest sekwencja kodowa, która identyfikuje pozycję frazy w słowniku.
- Etap kompresji jest znacznie dłuższy niż etap dekompresji, bo wymaga czasochłonnych operacji przeszukiwania strumienia danych wejściowych w celu wyodrębnienia powtarzających się fraz (budowa słownika).



Metody słownikowe

- Obecnie najbardziej popularny algorytm to LZW (Lempel-Ziv-Welch). Podstawa działania takich formatów jak GIF, ZIP, ARJ, RAR.



Metody stratne

- Pozwalają uzyskać stopnie kompresji nawet do 90%
- Wykorzystują przede wszystkim charakterystykę psychofizjologiczną ludzkich zmysłów (słuchu i wzroku).
- Metody stratne pozwalają określić współczynnik kompresji powiązany z jakością kompresowanych danych (im lepsza jakość tym większy rozmiar).



Metody stratne

- Metody transformacyjne,
- Metody fraktalne.

Metody transformacyjne


- Wykorzystują różnego typu transformacje do dekorelacji danych.
- Operacje wykonywane są w dziedzinie częstotliwości.
- Najczęściej spotykane transformacje: Fouriera, kosinusowa, falkowa,
- Operacja transformacji jest procesem odwracalnym,
- Obraz dzielony jest na przylegające bloki $N \times N$,
- W każdym bloku dokonywana jest operacja transformacji,
- Uzyskane współczynniki transformacji podlegają kwantyzacji (powstają dane o b.małej entropii),
- Uzyskane dane poddawane są kompresji bezstratnej



Kompresja JPEG

- Najpopularniejszy standard kompresji i format zapisu zdjęć (Joint Photographic Experts Group),
- Wykorzystuje transformację kosinusową DCT

- DCT

$$F(u, v) = \frac{1}{4} \sum_i^{16} \sum_j^{16} C(u) \cdot C(v) \cdot \cos \frac{(2i+1)u\pi}{16} \cdot \cos \frac{(2j+1)v\pi}{16} \cdot f(i, j)$$
$$C(\xi) = \begin{cases} \frac{1}{\sqrt{2}} & \text{dla } \xi = 0 \\ 1 & \text{dla } \xi \neq 0 \end{cases} \quad u, v = 1 \dots 16$$




Kompresja JPEG - algorytm

- Obraz dzielony jest na bloki 8x8 (uproszczenie operacji, możliwość zrównoleglenia obliczeń),
- Ponieważ oko ludzkie jest bardziej wyczułone na zmianę jasności niż koloru, dlatego przestrzeń barw RGB zamieniana jest w przestrzeń YUV (luminacja i składowe chrominancji),
- Następnie w każdym bloku dokonywana jest transformacja DCT, w wyniku której otrzymywana jest macierz wartości rzeczywistych

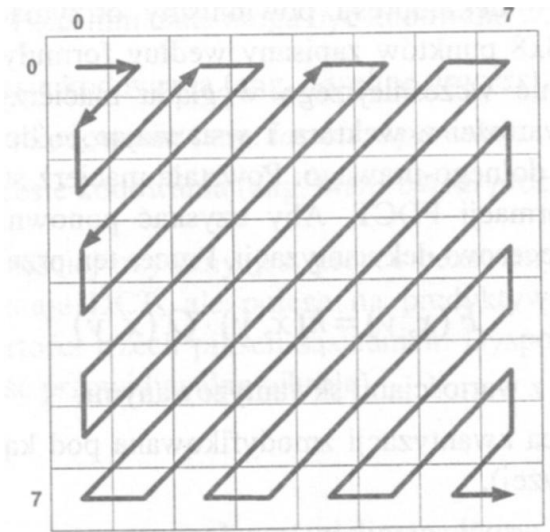


Kompresja JPEG - kwantyzacja

- Następnie dokonywana jest kwantyzacja, która odpowiada za „stratność” kompresji.
- Stosowana jest kwantyzacja równomierna połączona z progową selekcją próbek na bazie tablicy współczynników kwantyzacji. Tablica kwantyzacji ma również wymiar 8x8.
- Wyniki transformacji dzielone są przez współczynniki kwantyzacji.
- Dla każdego bloku stosowane są te same współczynniki kwantyzacji,
- W standardzie JPEG zdefiniowane są 4 tablice kwantyzacji (żadna nie jest obowiązkowa)

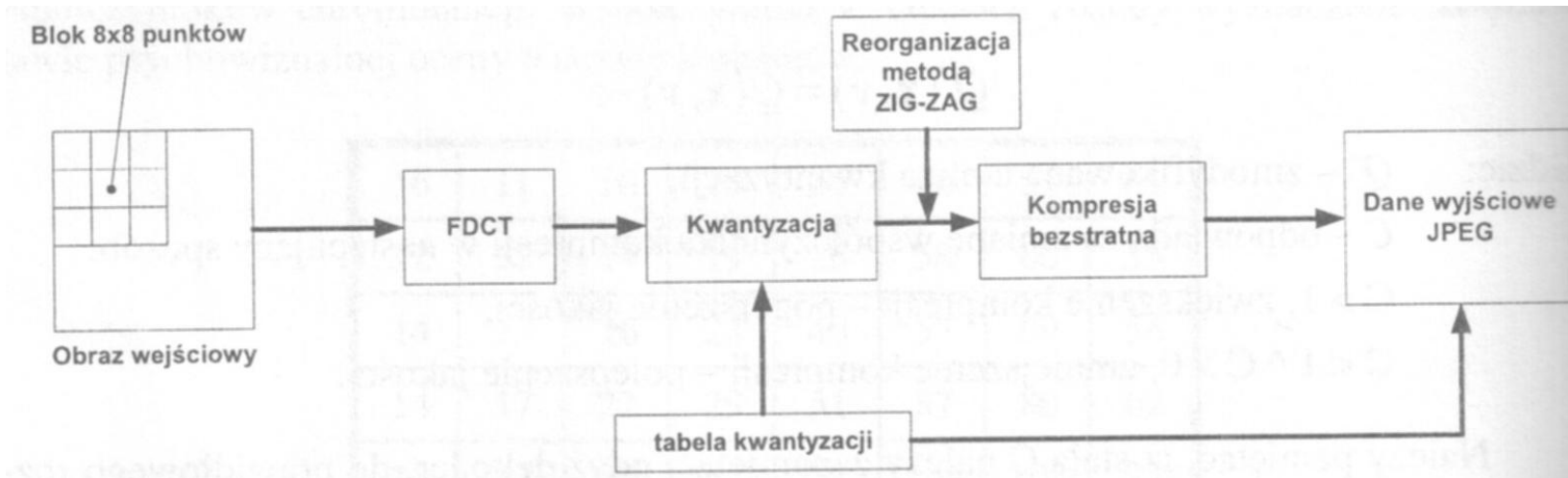
Kompresja JPEG

- Skwantowane współczynniki transformacji są komprimowane za pomocą metody bezstratnej Huffmana. Przed tym jednak współczynniki przepisywane są do 64 elementowego wektora za pomocą techniki Zig-Zag (chodzi o to aby niezerowe współczynniki były obok siebie).



Kompresja JPEG

- W pliku JPEG zapisywane są również tablice kwantyzacji oraz współczynnik skalowania. Wartości te są niezbędne do dekompresji danych.





Kompresja fraktalna

- Fraktale to zapisane matematycznie figury geometryczne, które wykazują cechy samopodobieństwa. Zazwyczaj używane do generowania obrazów.
- Od kilkunastu lat próbuje się wykonać operację odwrotną, czyli opracować metodę wyznaczania szeregu transformacji fraktalnych, które pozwalają wygenerować dany obraz.
- Można uzyskać stopnie kompresji nawet 1000:1



Kompresja fraktalna

- Do tej pory najlepsze rezultaty osiągnięto za pomocą zestawu funkcji iteracyjnych (IFS - Iterated Function System).
- Układ iterowanych odwzorowań definiuje się jako

zbiór k odwzorowań zwężających
 w_i ($i=1, 2, 3, \dots, k$), który oznacza się przez
 $\{X; w_1, w_2, w_3, \dots, w_k\}$.

w oznacza odwzorowanie afiniczne
zwężające przestrzeń (X, \square) w siebie



Kompresja fraktalna - algorytm

- Podział obrazu na rozłączne obszary R_i ,
- Dla każdego obszaru trzeba znaleźć dziedzinę D_i , która będzie jego odwzorowaniem,
- Dla każdej wybranej pary (D_i, R_j) wybrać możliwie najlepsze przekształcenie przeprowadzające D_i w R_j . Najczęstsze kryterium – minimalizacja błędu średniokwadratowego

Kompresja fraktalna

- Bardzo wolny proces kompresji, trudna automatyzacja tego procesu.
- Szybka dekompresja,
- Wysoka jakość przy bardzo dużych stopniach kompresji

24-bit true color (250 kB)



Fractal Image Format FIF (5,3kB)
st. kompresji 45:1





Kompresja obrazu ruchomego

- Obraz ruchomy (animacja, wideo) powstaje z sekwencji obrazów statycznych.
- W procesie kompresji wykorzystuje się zarówno metody kompresji obrazów statycznych jak i specjalne metody kompresji wykorzystujące właściwości ruchu obiektów w kolejnych ramkach obrazu ruchomego.
- Dwie płaszczyzny kompresji obrazu ruchomego:
 - Kompresja wewnątrzramkowa,
 - Kompresja międzyramkowa.



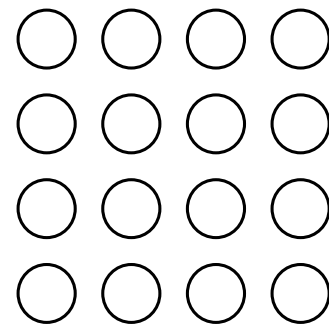
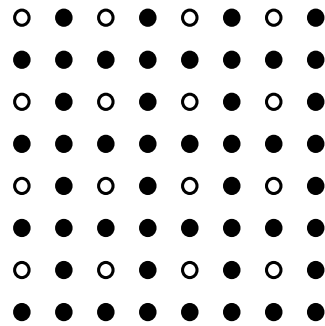
Kompresja wewnątrzramkowa

- polega na redukcji nadmiaru informacji przestrzennej w obrębie jednej ramki (ang. spatial redundancy reduction).
- metody kompresji wewnątrzramkowej:
 - podpróbkiwanie (ang. sub-sampling)
 - kwantyzacja (ang. quantization)
 - kwantyzacja wektorowa (ang. vector quantization)
 - kodowanie transformacyjne (ang. transform coding),
- Ramki, które podlegają kompresji wewnątrzramkowej to tzw. ramki kluczowe.

Podpróbkiwanie

- kodowanie metodą podpróbkiwania (ang. sub-sampling) jest stratne i polega na redukcji liczby punktów w obrazie. Skutkiem ubocznym jest równoczesne obniżenie jakości tego obrazu.
- rodzaje podpróbkiwania:
 - w kopii obrazu wykorzystana jest tylko część pikseli z obrazu oryginalnego.
 - dla grupy kilku pikseli wyznaczana jest średnia wartość piksela. Następnie, wartości z grupy pikseli są zastępowane tą wartością średnią. Algorytm jest bardziej złożony, lecz jakość obrazu jest lepsza niż w pierwszym przypadku.
- wykorzystuje zdolności ludzkiej percepcji do wypełnienia luk pomiędzy sąsiednimi pikselami.
- dekodery mogą również wstawiać piksele usunięte w wyniku podpróbkiwania (interpolacja)

Podpróbkiwanie - przykłady



Kwantyzacja

- polega na redukcji bitów reprezentujących wartość piksela. Cały zakres wartości pikseli jest podzielony na określoną liczbę przedziałów - stopni kwantyzacji.
- jeśli liczba bitów opisujących jeden piksel jest mała, to można wówczas mówić o kwantyzacji zgrubnej (ang. coarse quantization)





Kompresja międzyramkowa

- kolejne obrazy w sekwencji niewiele się od siebie różnią –możliwa jest redukcja nadmiaru informacji czasowej między kolejnymi ramkami sygnału wizyjnego (ang. temporal redundancy reduction)
- redukcja nadmiaru informacji czasowej polega na wyszukaniu różnic między kolejnymi ramkami i odpowiednim ich kodowaniu
- metody kompresji międzyramkowej:
 - podpróbkiwanie (ang. sub-sampling),
 - kodowanie różnicowe (ang. difference coding),
 - blokowe kodowanie różnicowe (ang. block based difference coding),
 - kompensacja ruchu (ang. motion compensation).



Podpróbkowanie

- podpróbkowanie można również zastosować do kompresji sekwencji ramek sygnału wizyjnego
 - wybierana i kodowana może być np.: co druga ramka.
 - po stronie dekodera następuje interpolacja, aby uzupełnić brakujące ramki.



Kodowanie różnicowe

- polega na porównaniu każdej ramki z ramką poprzednią oraz kodowaniu tylko tych pikseli, których wartość zmienia się
- jeśli kompresja ma być bezstratna, to wówczas każda zmiana wartości piksela musi być uwzględniona. Algorytm może się wtedy stać nieefektywny pod względem czasu przetwarzania, jak również objętość sygnału po kompresji może przewyższać objętość tego samego sygnału przed kompresją.
- w przypadku kompresji stratnej, nie każda zmiana wartości piksela jest brana pod uwagę. Do kodowania brane są tylko te piksele, dla których zmiany wartości są znaczne.



Blokowe kodowanie różnicowe

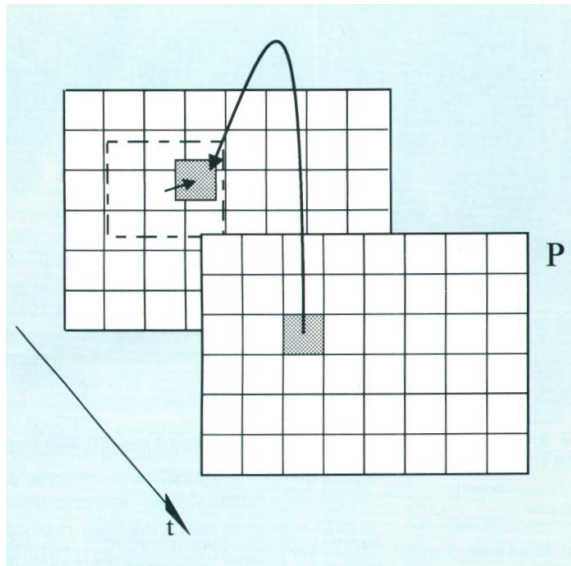
- cały obraz jest podzielony na nie zachodzące na siebie obszary - bloki i każdy blok jest porównywany ze swoim odpowiednikiem w ramce poprzedniej
- zmniejsza się czas obliczeń, gdyż aktualizacja zachodzi dla wszystkich pikseli w bloku. Dodatkowo, następuje redukcja liczby bitów koniecznych do adresacji pikseli/bloków (do aktualizacji)

Kompensacja ruchu

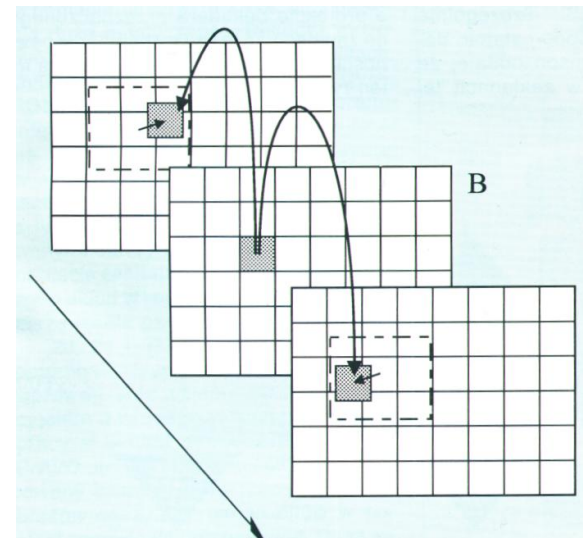
- polega na wykryciu zmian w położeniu pikseli oraz kodowaniu tylko tych zmian
- efektywność metody kompensacji ruchu zależy od stopnia, w jakim są spełnione następujące założenia:
 - ruch obiektów w obrazie zachodzi jedynie w płaszczyźnie równoległej do płaszczyzny obiektu
 - efekty powiększenia lub obrotu nie są uwzględniane
 - jasność obiektów jest czasowo i przestrzennie jednakowa (jasność poszczególnych pikseli nie zmienia się w czasie oraz w obrębie obrazu)
 - nie uwzględniane są wzajemne zakrywania się obiektów.

Kompensacja ruchu

- najczęściej używaną metodą kompensacji ruchu jest metoda dopasowania bloków
 - obraz dzielony jest na jednakowe, nie zachodzące na siebie obszary pikseli – bloki
 - metoda ta opiera się na wykryciu zmiany położenia poszczególnych bloków.



kompensacja predykcyjna



kompensacja dwukierunkowa



Kompresja dźwięku

- O jakości zapisu cyfrowego sygnału dźwięku w technice PCM decydują dwa parametry:
 - Rozdzielczość bitowa pojedynczej próbki dźwięku,
 - Częstotliwość próbkowania.
- Jakość CD to 16 bitów na próbkę, częstotliwość próbkowania 44100 Hz oraz 2 kanały dźwięku co wymaga transferu danych ok. 172 kB/s.
- Typowe algorytmy kompresji bezstratnej pozwalają uzyskać stopień kompresji rzędu 10-20%,
- Format FLAC (Free Lossless Audio Codec) wykorzystujący liniową predykcję pozwala uzyskać kompresje rzędu 40-60%



Kompresja dźwięku – kodowanie różnicowe

- Ponieważ w cyfrowym sygnale dźwięku kolejne próbki mają podobne wartości, dlatego opracowano kilka algorytmów kompresji, których podstawą jest odpowiednie kodowanie różnicy między kolejnymi próbkami. Wykorzystywane są przede wszystkim metody predykcyjne (ADPCM, G-726, LPC).

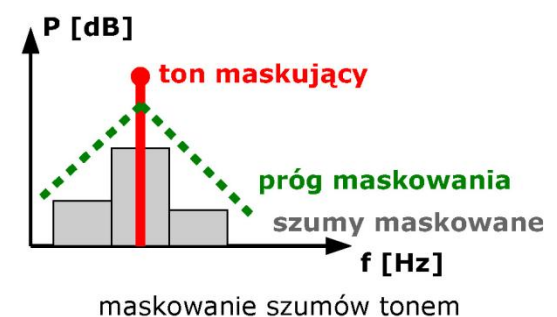
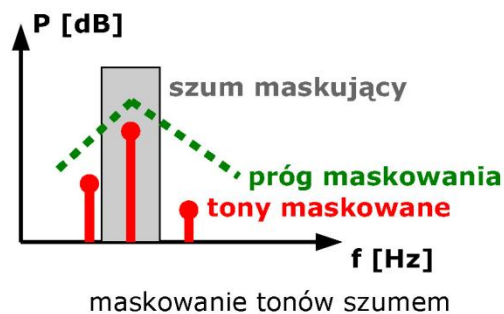


Kompresja dźwięku – kodowanie dynamiki amplitudy dźwięku

- Ważną grupę metod kompresji dźwięku stanowią kodeki dynamiki amplitudy (u-law, A-law, μ -law, G-711).
- Można zaoszczędzić na przepływności bitowej przez zastosowanie nierównomiernej kwantyzacji amplitudy, która polega na tym, że duże poziomy amplitudy mają krótszy zapis bitowy niż amplitudy o niskim poziomie. Przeznaczanie większej dokładności zapisu niskich poziomów dźwięku pozwala znacznie ograniczyć szum kwantyzacji.

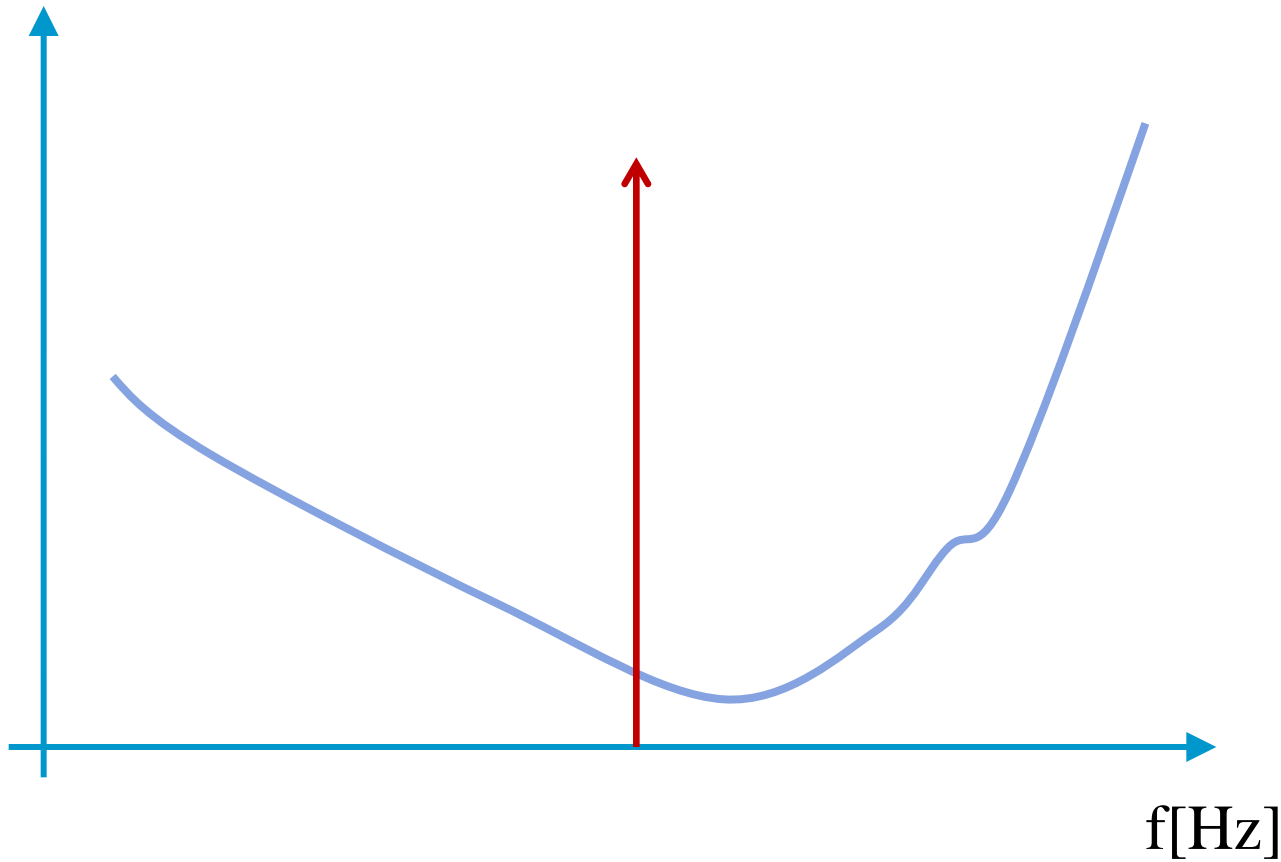
Kompresja dźwięku – kodowanie perceptualne

- Największe stopnie kompresji uzyskuje się w metodach wykorzystujących właściwości ludzkiego słuchu, nawet do 90%.
- Podstawą tych metod jest zjawisko maskowania, które polega na tym, że dźwięki o niższej amplitudzie i zbliżonej częstotliwości są „zagłuszane” przez dźwięki o wyższej amplitudzie.



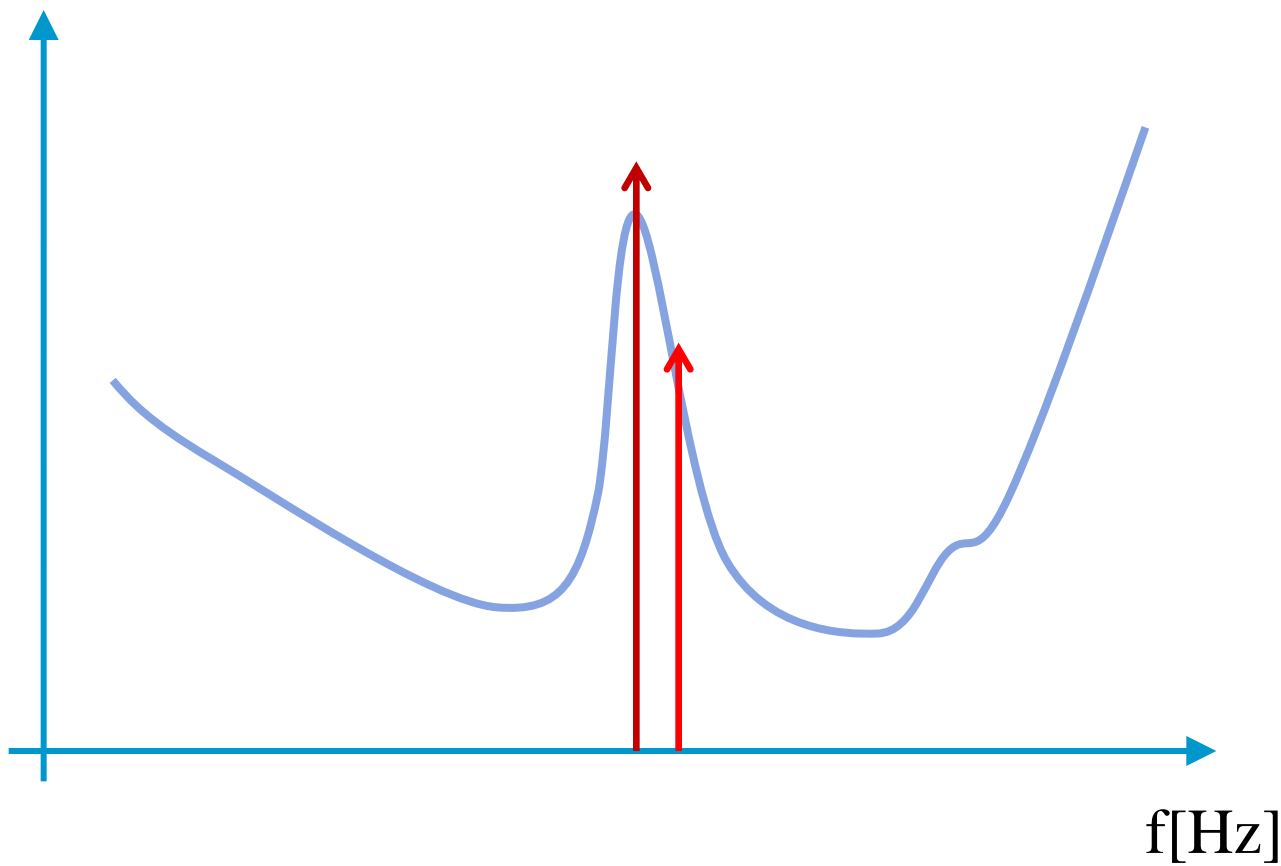
Ilustracja maskowania

L[dB]



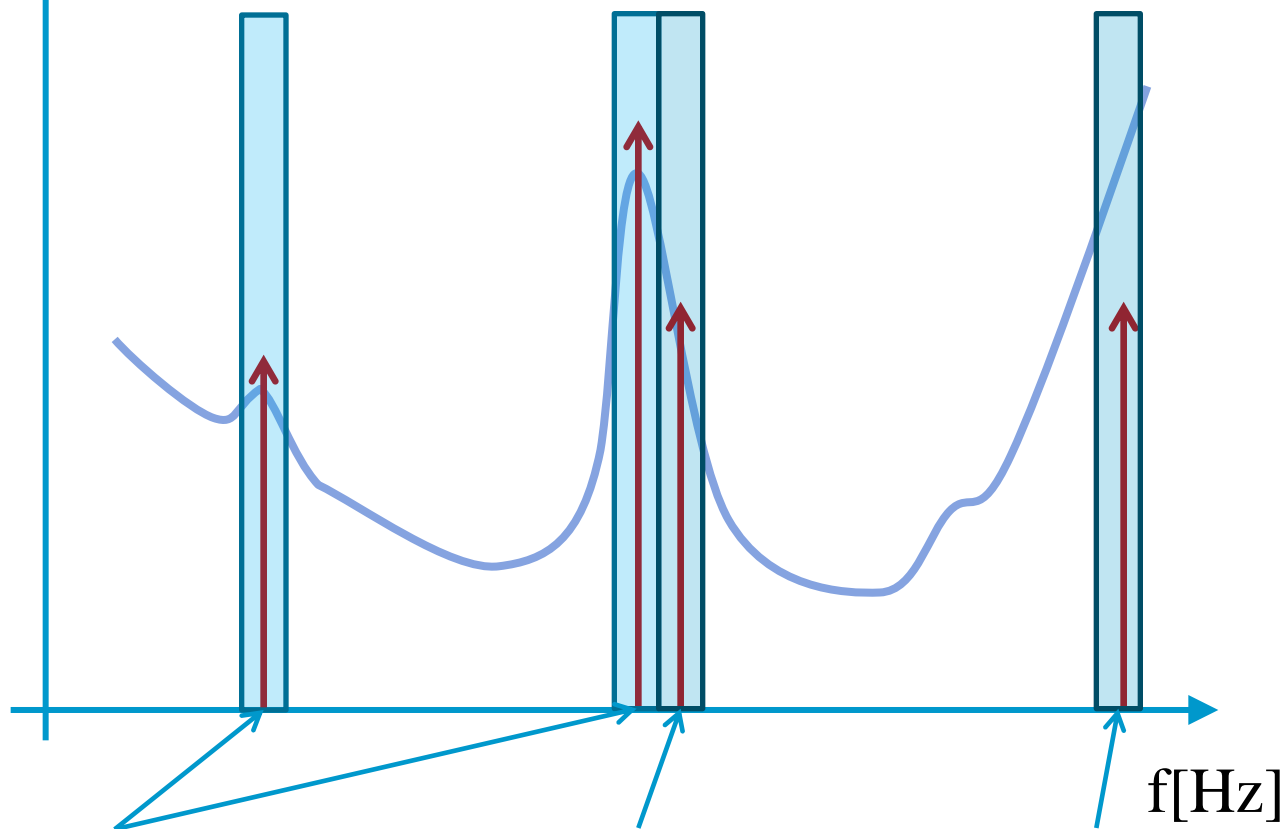
Ilustracja maskowania

L[dB]



Ilustracja maskowania

L[dB]



trzeba użyć
więcej bitów

można użyć
mniej bitów

nie trzeba
kodować



Kompresja dźwięku – kodowanie perceptualne

- Maskowanie u każdego człowieka zachodzi nieco inaczej, dlatego kodeki używają uśrednionego modelu psychakustycznego.
- Znając maskowane składowe dźwięku kodek usuwa je z sygnału.
- Najbardziej popularne kodeki wykorzystujące psychoakustyczne właściwości słuchu to: MPEG-1 warstwa 3 (MP3), kodek ATRAC, Dolby AC-3.



Dziękuję za uwagę