

Wymiana i Składowanie Danych Multimedialnych 2019

Mateusz Moderhak,
matmod@biomed.eti.pg.gda.pl,
p. 106 WETI.

Kompresja audio



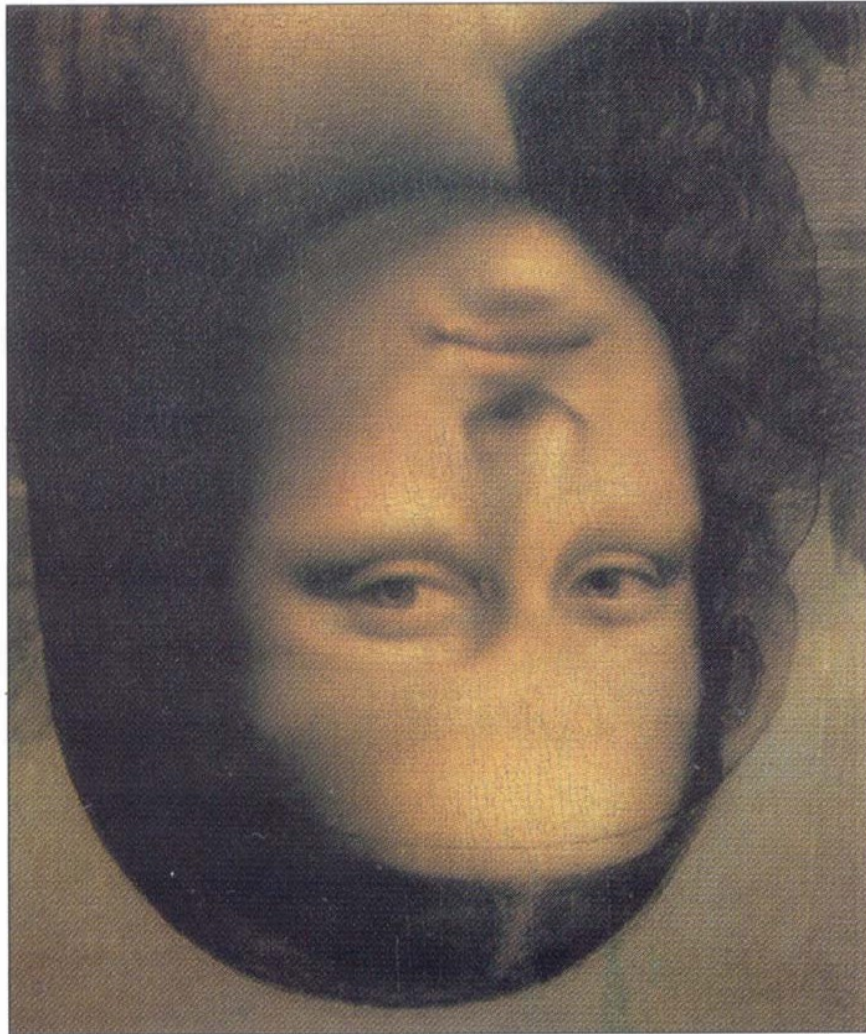


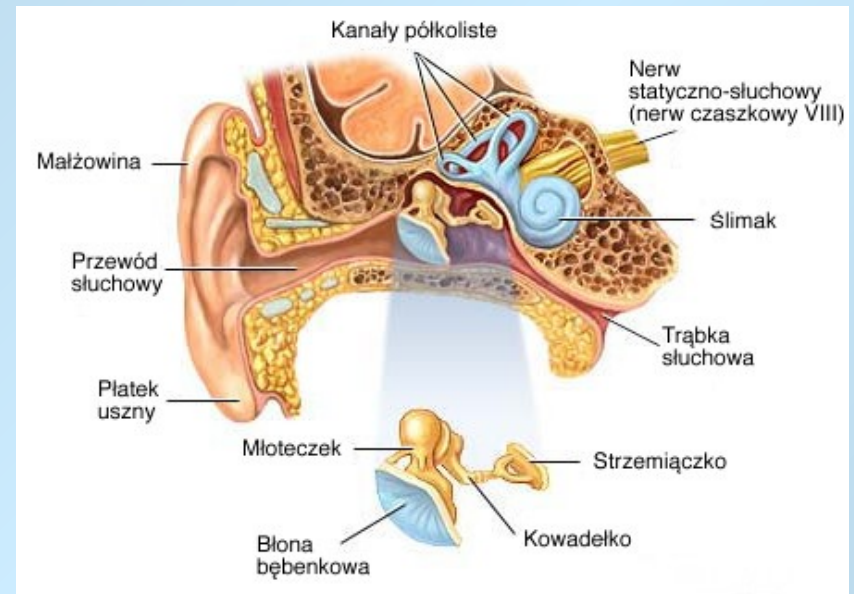
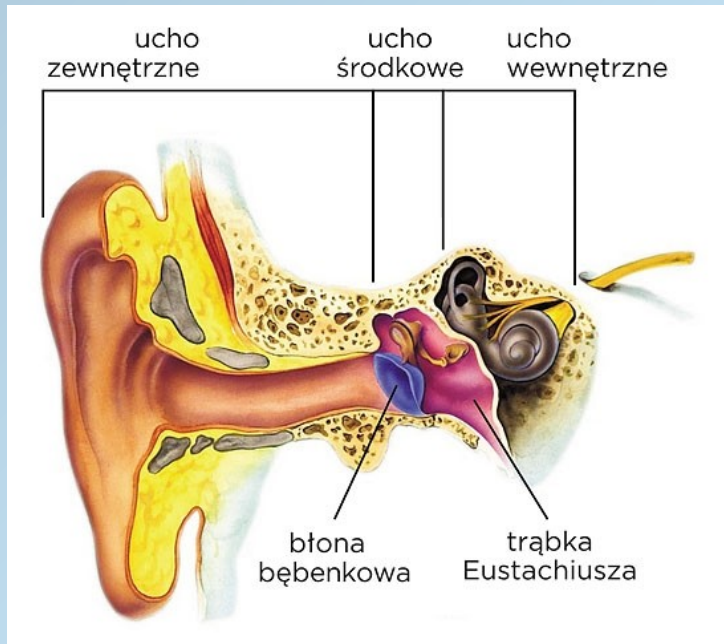
- Sygnał audio o jakości CD:
 - 2 kanały, 16 bitów bitów na próbkę, próbkowanie 44,1 kHz
- Przepływność:
- $2 \times 16 \text{ [bitów]} \times 44100 \text{ [1/s]} = 1411200 \text{ [bitów/s]}$ (ok. 172 kB/s)
- minuta muzyki o jakości płyty CD to ok. 10 MB danych

MP3

- Nazwa MP3 pochodzi od nazw: MPEG-1 lub MPEG-2 Audio Layer III (MPEG - Moving Picture Experts Group). Standard stratnej kompresji sygnału audio stosowany powszechnie w przenośnych urządzeniach audio, TV, kinie... Powstał jako owoc pracy wielu grup inżynierów: Fraunhofer IIS, University of Hannover, AT&T-Bell Labs, Thomson-Brandt, CCETT i innych w latach 1991 -1993.
- Wykorzystuje **model psychoakustyczny** percepcji dźwięku obejmujący takie elementy jak: krzywe absolutnego słyszenia, zjawiska maskowania jednoczesnego, premaskowania i postmaskowania tonów.
- Pozwala na zachowanie wysokiej subiektywnej jakości dźwięku przy dość wysokim uzyskanym współczynniku kompresji
 - Audio CD (*.wav) 1411,200 [kbit/s] , 1min ~ 10 MB danych
 - MP3 128 kbit/s, 1min ~ 0.96MB, Q= 10.5
 - MP3 196 kbit/s, 1min ~ 1.5MB, Q = 6.6
 - MP3 320 kbit/s, 1min ~ 2.4MB, Q= 4.2
- Coraz częściej zamiast stałej przepływności (CBR) stosuje się zmienną przepływność (VBR). Pozwala to uzyskać jeszcze lepsze współczynniki kompresji.

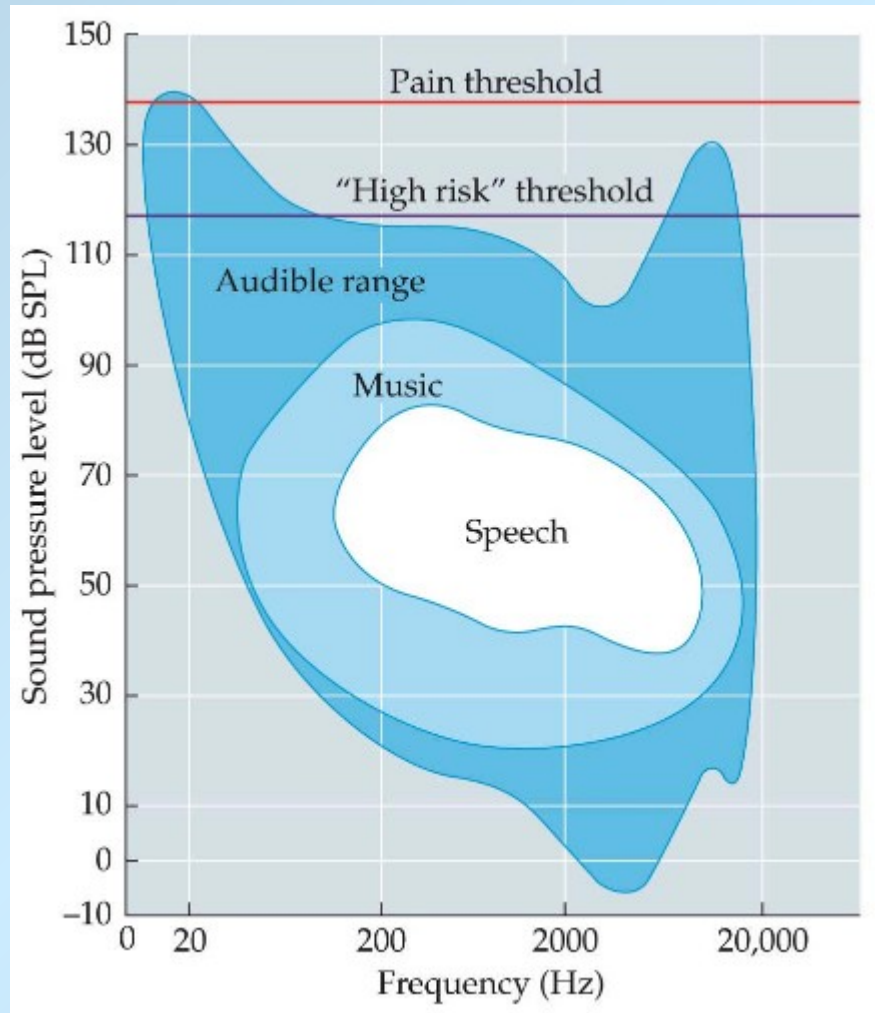
„Model psychoakustyczny”





- **Dźwięk** – wrażenie słuchowe, spowodowane falą akustyczną rozchodzącą się w ośrodku sprężystym (ciele stałym, cieczy, gazie). Częstotliwości fal, które są słyszalne dla człowieka, zawarte są w paśmie między wartościami granicznymi od ok. 16 Hz do ok. 20 kHz

Percepcja dźwięku



dB SPL

dB SPL ([sound pressure level](#)) – for sound in air and other gases, relative to 20 micropascals (μPa) = 2×10^{-5} Pa, approximately the quietest sound a human can hear. This is roughly the sound of a mosquito flying 3 meters away. For [sound in water](#) and other liquids, a reference pressure of 1 μPa is used.^[36]

An RMS sound pressure of one pascal corresponds to a level of 94 dB SPL.

dB SIL

dB [sound intensity level](#) – relative to 10^{-12} W/m², which is roughly the [threshold of human hearing](#) in air.

dB SWL

dB [sound power level](#) – relative to 10^{-12} W.

dB(A), dB(B), and dB(C)

These symbols are often used to denote the use of different [weighting filters](#), used to approximate the human ear's [response](#) to sound, although the measurement is still in dB (SPL). These measurements usually refer to noise and noisome effects on humans and animals, and are in widespread use in the industry with regard to noise control issues, regulations and environmental standards. Other variations that may be seen are dB_A or dBA. According to ANSI standards, the preferred usage is to write L_A = x dB. Nevertheless, the units dBA and dB(A) are still commonly used as a shorthand for A-weighted measurements. Compare [dBc](#), used in telecommunication.

dB HL or dB hearing level is used in [audiograms](#) as a measure of hearing loss. The reference level varies with frequency according to a [minimum audibility curve](#) as defined in ANSI and other standards, such that the resulting audiogram shows deviation from what is regarded as 'normal' hearing.^[citation needed]

dB Q is sometimes used to denote weighted noise level, commonly using the [ITU-R 468 noise weighting](#).^[citation needed]

Audio electronics [\[edit\]](#)

dBm

dB(mW) – power relative to 1 [milliwatt](#). In audio and telephony, dBm is typically referenced relative to a 600 ohm impedance,^[37] while in radio frequency work dBm is typically referenced relative to a 50 ohm impedance.^[38]

dBFS

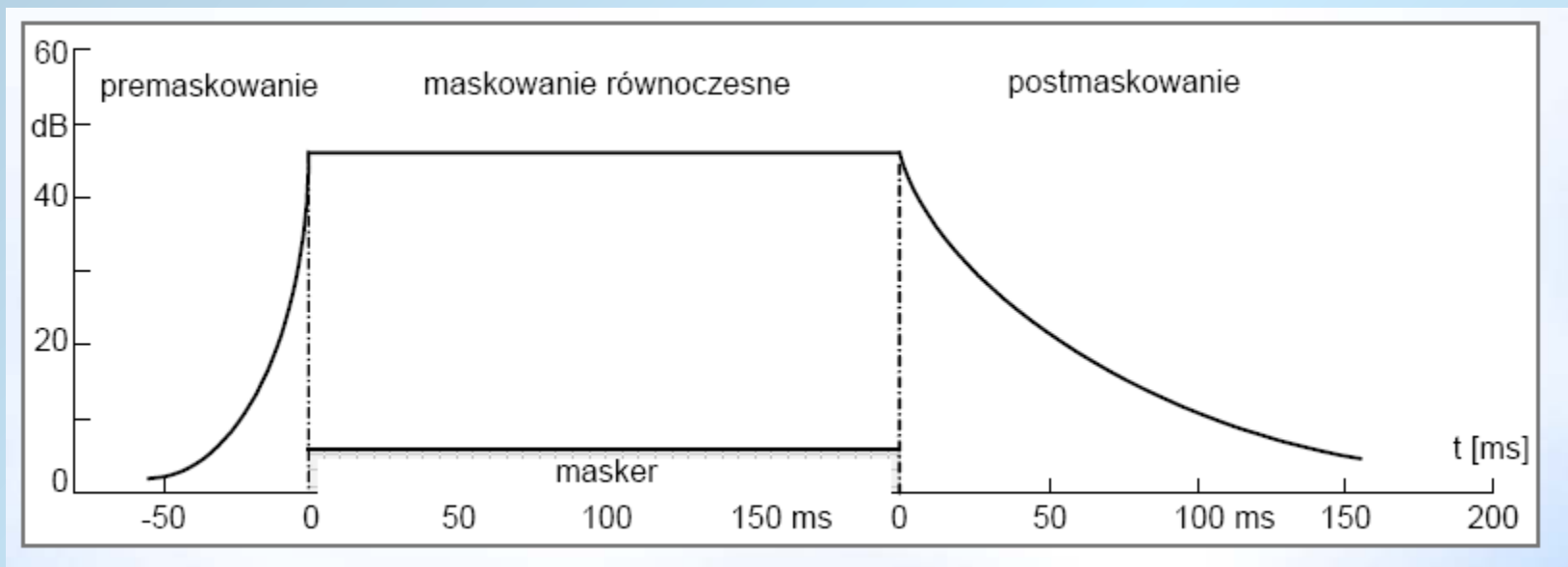
dB(full scale) – the [amplitude](#) of a signal compared with the maximum which a device can handle before [clipping](#) occurs. Full-scale may be defined as the power level of a full-scale [sinusoid](#) or alternatively a full-scale [square wave](#). A signal measured with reference to a full-scale sine-wave will appear 3dB weaker when referenced to a full-scale square wave, thus: 0 dBFS(ref=fullscale sine wave) = -3 dBFS(ref=fullscale square wave).

dBTP

dB(true peak) - [peak amplitude](#) of a signal compared with the maximum which a device can handle before clipping occurs.^[39] In digital systems, 0 dBTP would equal the highest level (number) the processor is capable of representing. Measured values are always negative or zero, since they are less than or equal to full-scale.

dB SPL ??

Maskowanie tonów (I)

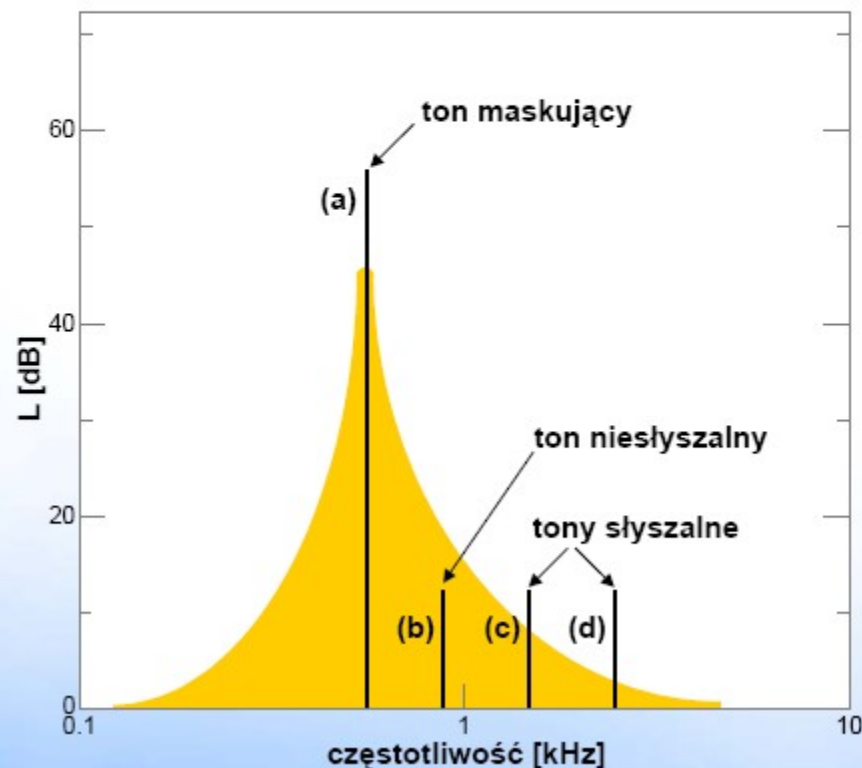


premaskowanie – dźwięk o dużym natężeniu jest w stanie zamaskować tony, które wystąpiły wcześniej. Premaskowanie trwa od 10 do 30 ms.

postmaskowanie – po wystąpieniu głośnego tonu, pozostałe dźwięki mogą nie być słyszane. Postmaskowanie trwa do ok. 200 ms. Czas ten zależy od natężenia oraz czasu trwania tonu maskującego.

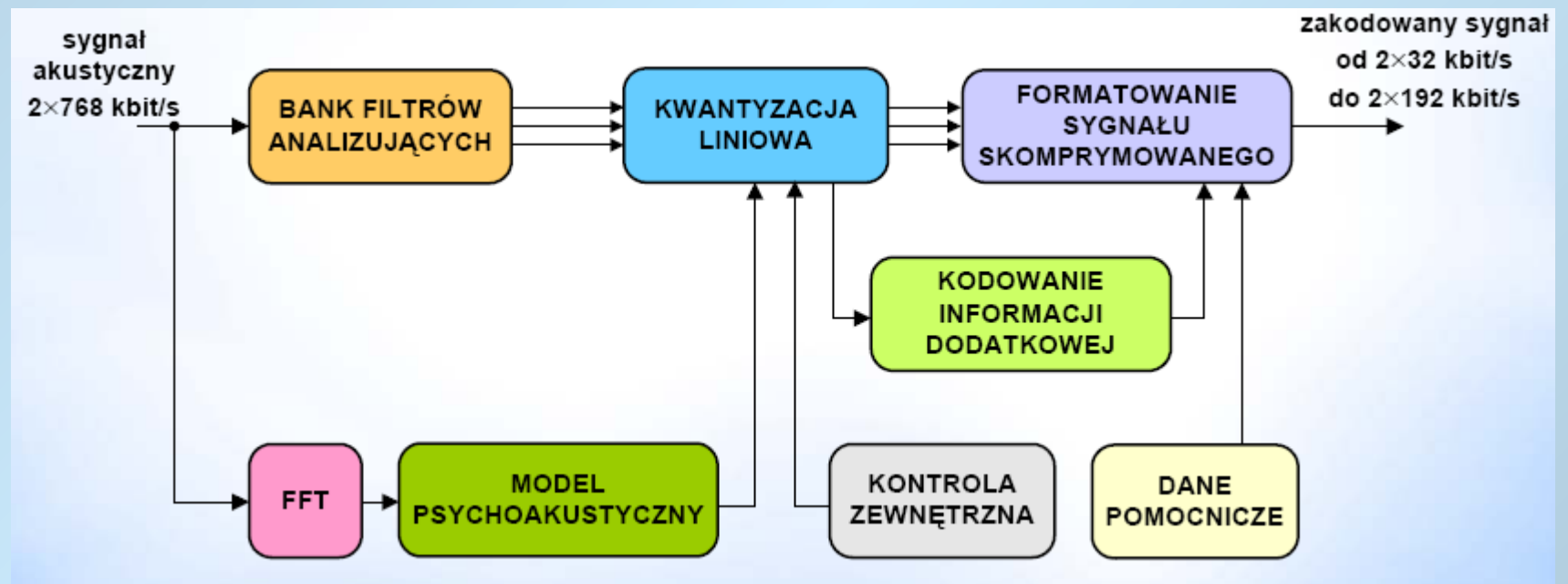
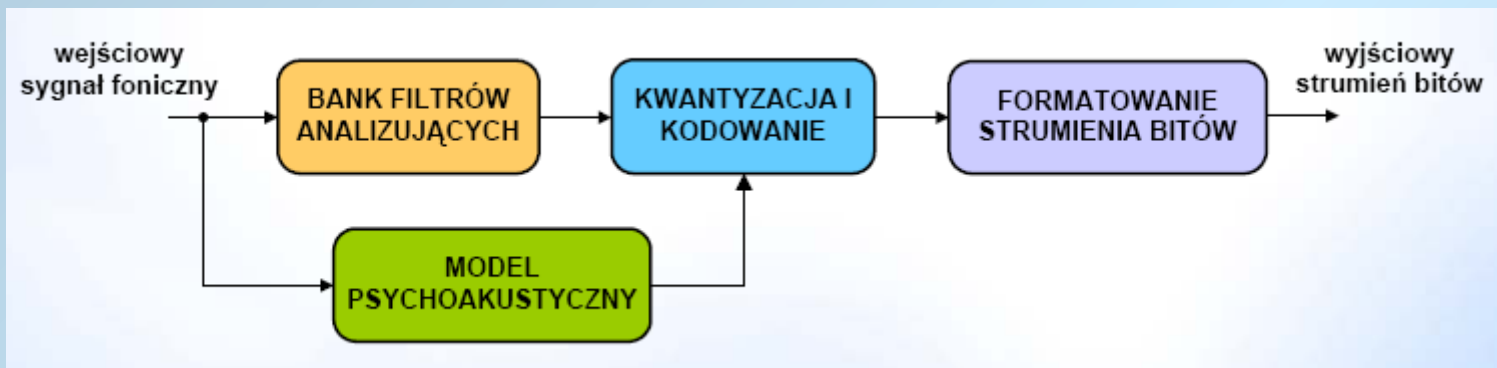
Maskowanie tonów (II)

Maskowanie równoczesne charakteryzuje się tym, że pewne tony stają się niesłyszalne w obecności innych – tzw. **maskerów**.

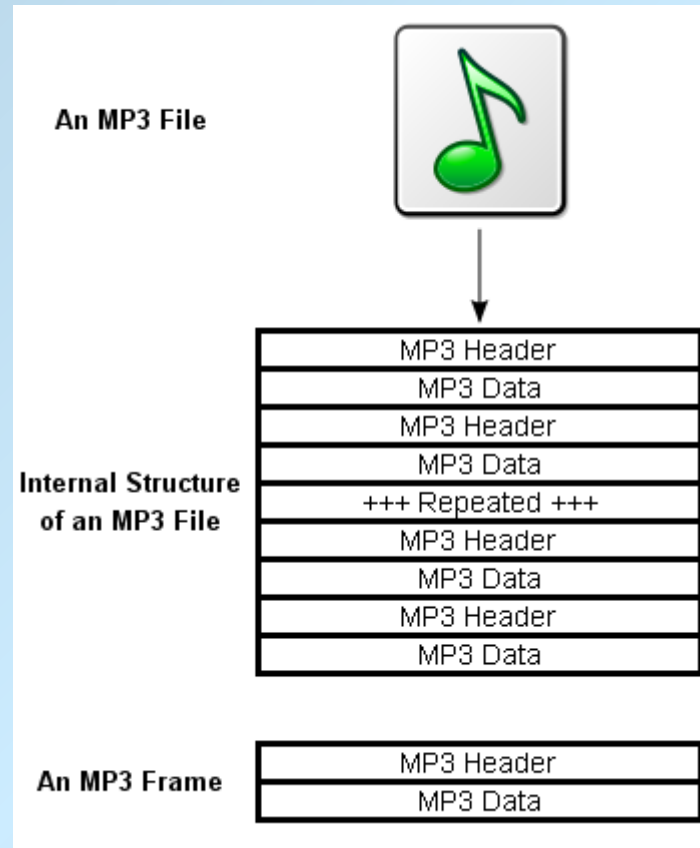


MP3 - algorytm

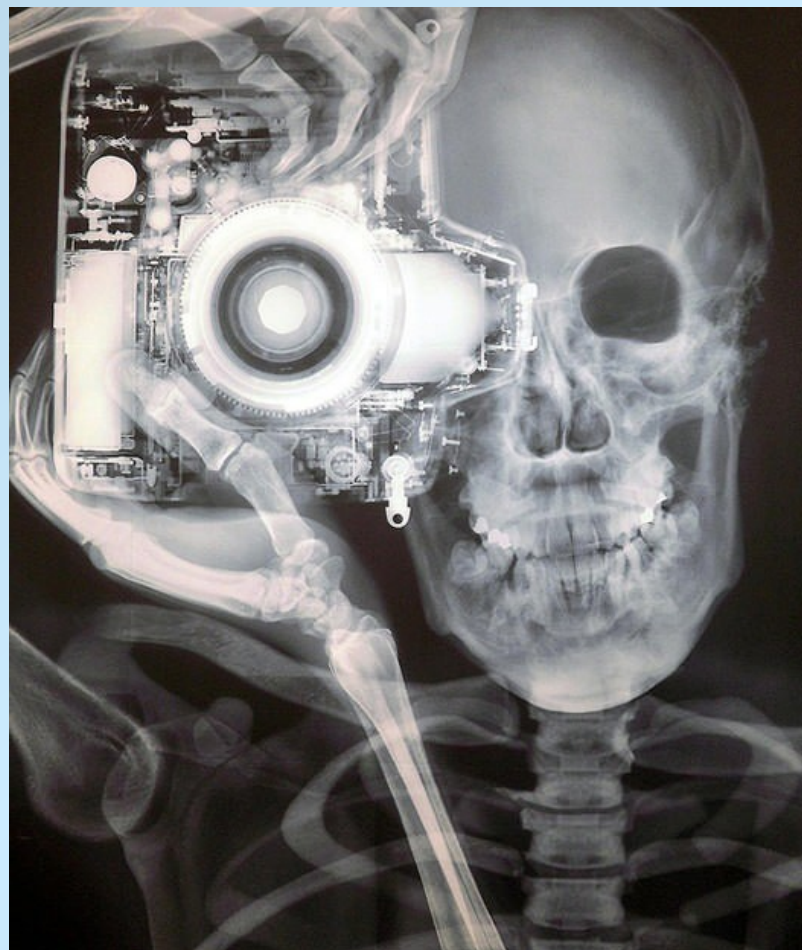
1. Wejściowy sygnał cyfrowy jest dzielony na 576, odpowiednio węższych pasm, za pomocą banku filtrów cyfrowych w procesie cyfrowego przetwarzania sygnałów. W rezultacie otrzymuje się 576 odrębnych sygnałów. Pasma każdego z nich jest 576 razy węższe od wejściowej szerokości pasma, zatem do próbkowania można użyć odpowiednio niższej częstotliwości.
2. Kodek MP3 analizuje poszczególne pasma i na podstawie modelu psychoakustycznego określa z jaką dokładnością należy zakodować każde z nich i czy jest potrzeba uwzględnienia danego pasma. W ten sposób znacznie ograniczono ilość bitów potrzebnych do zakodowania, a szum kwantyzacji, mimo że znacznie większy niż w oryginale, jest tak ukształtowany, że jest niesłyszalny. Subiektywna jakość dźwięku zależy ponadto od zastosowanego w kodeku modelu psychoakustycznego oraz od samego słuchacza. Jeśli jego słuch będzie bardzo odbiegał od modelu użytego w kodeku, oceniona jakość dźwięku może być odebrana jako znacznie różniąca się od oryginału
3. Dane zapisuje się w postaci zmiennoprzecinkowej, co dodatkowo zmniejsza ich rozmiar.
4. Tak przygotowane dane łączy się, dodając dane sterujące umożliwiające odkodowanie dźwięku i na końcu poddaje się je kompresji bezstratnej – powstaje plik dźwiękowy w formacie MP3.

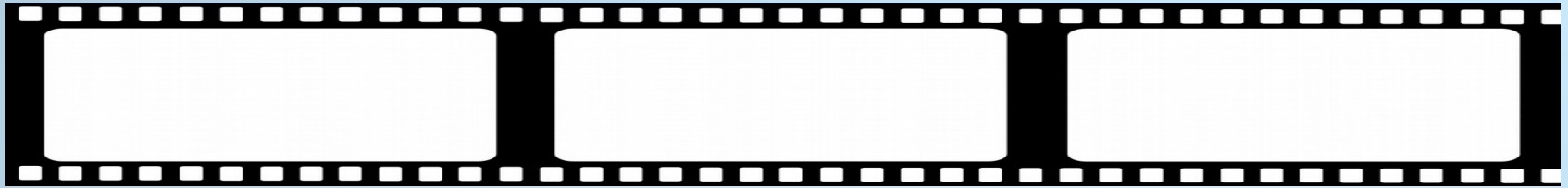


Plik MP3



Kompresja ruchomych obrazów





- Połączenie mediów dźwięku i obrazu (24 fps kino, 86 Hz CRT, <60 Hz LCD)
- Ogromne potrzeby dotyczące pasma oraz przestrzeni dyskowej celem ich wymiany oraz składowania
- Przykład: strumień HDTV w formacie panoramicznym 16:9, 920 × 1080 pikseli, 30 klatek na sekundę
 - $920 \times 1080 \times 3 \times 8 \times 30 = 715\,392\,000$ kbit/s = 89424000 kB/s
 - Godzinny program HDTV: $\times 60 \times 60 \sim 322$ TB bez audio (!!)

Kompresja Video



Standardy kompresji video

- **MJPEG**

- Mało zaawansowany, stosowany w transmisji IP, HTTP, kamery video, konsole gier, standard DV (Digital Video), QuickTime Player, lustrzanki cyfrowe
- Tylko kompresja „wewnątrz klatkowa”
- Maksymalny współczynnik kompresji $< 1:20$
- Wygodny proces kodowania oparty na JPEG

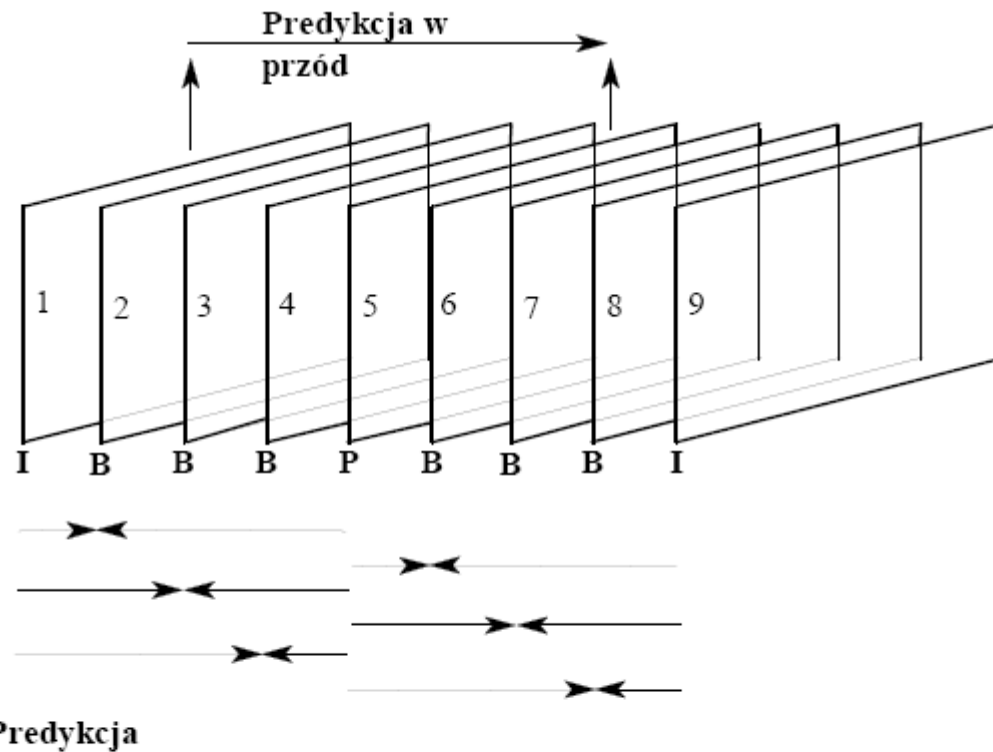
- **MPEG 1,2,4**

- Powstały w latach 1991, 1995 i 1999.
- Oferuje współczynnik kompresji $> 1:50$ dzięki zastosowaniu kompresji międzyklatkowej i algorytmów predykcji
- Transmisja danych, telewizja cyfrowa, elektronika użytkowa...
- Skomplikowany i wymagający proces kodowania i dekodowania

Ramki w MPEG

- **Ramki „I”** (intra frames)
 - Zakodowane przy pomocy DCT (JPEG)
 - Najmniejszy współczynnik kompresji
 - Zaczynają każdą sekwencję ramek (GOP)
 - Swobodny dostęp
- **Ramki typu „P”** (predicted frames)
 - Kodowane przy pomocy predykcji w przód
 - Nie muszą być poprzedzone przez ramkę I lub P
 - Duży współczynnik kompresji
- **Ramki typu „B”** (bidirectional frames)
 - Zakodowane przy pomocy dwóch ramek referencyjnych: byłej i przyszłej (I lub P)
 - Największy współczynnik kompresji
 - Konieczna zmiana sekwencji transmisji obrazów – najpierw ramki odniesienia

Struktura ramek w MPEG



- kolejność dekodowania: 1, 5, 2, 3, 4, 9, 6, 7, 8
- typowa sekwencja: IBBPBBPBBPBBPBB

MPEG 1 i 2 cechy wspólne

- Zgrupowane ramki
- Trzy rodzaje ramek I, P, B
- Dopuszczalne różne formaty obrazu
- Podobna budowa nagłówka
- Możliwość odtwarzania uszkodzonych plików jak również plików z „uciętym końcem”
- Kłopotliwa edycja, wymaga przeliczania ramek lub stosowania jedynie ramek „I”

MPEG -1

- Opracowany z myślą o zapisie/odczycie materiału na płycie CD-ROM 150kbit/s
- Maks rozdzielczość 352x288 -> 4096x4096
- Brak ramek „B”
- Format 4:2:0
- Dźwięk: 32kHz, 44.1kHz, 48kHz

MPEG -1 zastosowania

- VCD
 - PAL, NTSC, XVCD
- Cyfrowa TV w USA
- DVD

MPEG - 2

- Zoptymalizowany dla przepływności 4Mbit/s
- PAL NTSC, SECAM, MAC, HDTV
- Formaty: 4:2:0, 4:2:2, 4:4:4
- Rozdzielczość 16384x16384
- Wielokanałowy dźwięk
- Zastosowanie „profilu”

MPEG – 2, profile

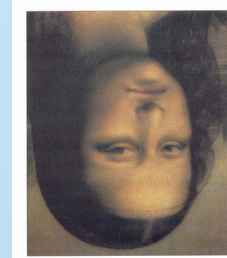
- **Prosty:** tylko ramki I lub P, eliminacja ramek B pozwoliła na zmniejszenie ilości pamięci w dekodernach
- **Główny:** ramki I, P, B, sygnał nie może być dzielony na warstwy
- **Skalowanie przestrzenne:** w strumieniu danych zapisane są dwie warstwy dla różnych rozdzielczości obrazu
- **Skalowane szumowo:** dwie warstwy różniące się jakością; w zależności od SNR wykorzystywane są obie warstwy lub tylko niższa,
- **Wysoki:** zastosowane są wszystkie typy obrazów i wszystkie warstwy

MPEG2 zastosowanie

- Telewizja cyfrowa od 1.5Mbit/s (VHS) do 9Mbit/s (jakość studyjna)
- DVD do 9.8Mbit/s, 720x576, 352x576; do 8 ścieżek dźwiękowych
- SVCD: 1500-2500 Mbit/s, 480x576, fonia w MPEG (też 5.1)
- SXVCD
- CVD
- DIVX

MPEG 4

- Projektowany z myślą o niskich przepływnościach – transmisje sieciowe
- Zastosowanie globalnej kompensacji ruchu, analiza przesunięcia z dokładnością do $\frac{1}{4}$ piksela
- Modelowanie psychowizualne
- Pojęcie sceny składające się z obiektów które powiązane są ze sobą relacjami w czasie i przestrzeni



MPEG 4

- Specyfikacja:
 - Systemowa (m.in. opis sceny)
 - Wizualna (metody kodowania naturalnych i sztucznych obiektów wizualnych)
 - Dźwiękowa
- Zdefiniowano 50 profili podzielonych na poziomy
- Rozdzielczość: 8x8 -> 2048x2048
- Formaty: 4:0:0, 4:2:0, 4:2:2
- Możliwość przeplotu i skalowania
- 5kbit/s do 1Gbit/s

Kompensacja ruchu w MPEG4

- Standardowa na blokach 8x8 lub 16x16, dokładność do $\frac{1}{4}$ piksela
- Globalna kompensacja ruchu (GMC) kodująca trajektorię ruchu za pomocą niewielkiej liczby parametrów
- GMC dla dużych statycznych obrazów opisująca tło, w każdej ramce zdefiniowano 8 parametrów opisujących ruch kamery, wykorzystywanych do rekonstrukcji tła przesłanego w pierwszej ramce

Pytania na kolokwium

- Wymień i opisz ramki jakie występują w standardzie MPEG1 i 2, narysuj przykładową strukturę ramek.
- Przedstaw algorytm kompresji dźwięku MP3
- Jakie właściwości percepcji dźwięku przez człowieka wykorzystano w standardzie MP3?
- Co oznacza pojęcie „format” i czym się różnią formaty: 4:0:0, 4:2:0, 4:2:2?
- Czym się różni standard MPEG4 od MPEG1 i 2?
- Scharakteryzuj standardy zapisu obrazów MJPEG, MPEG 1-4.

Kollokwium/Test

- Pytania:
 - Na czym polega idea transformaty falkowej?
 - Co to jest falka?
 - Narysuj dowolną falkę i na jej przykładzie zademonstruj operację skalowania i przesuwania falki
 - Co to jest falkogram/ skalogram?
 - Zasada działania/algorytm dyskretnej transformaty falkowej
 - Wymień zastosowania transformaty falkowej
 - Algorytm JPEG2000

Dziękuję za uwagę

