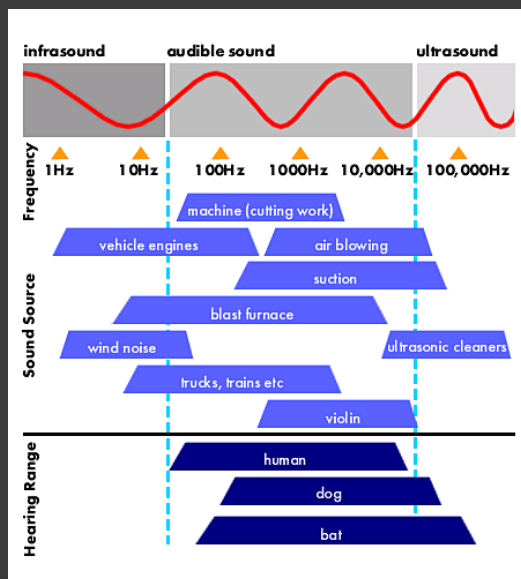


dr inż. Piotr Ody

FORMATY DŹWIĘKU

1

Parametry słuchu



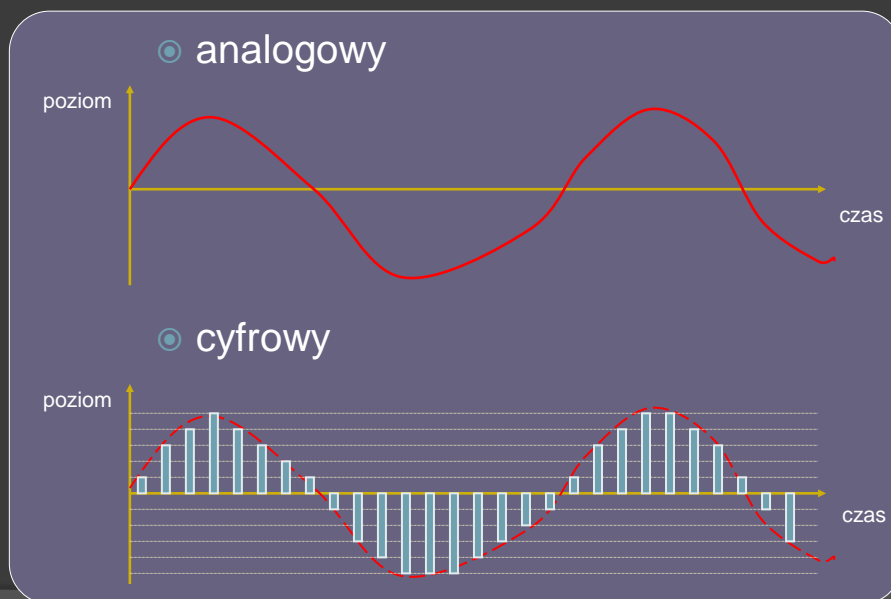
• zakres słyszanych przez człowieka częstotliwości: 20 Hz - 20 kHz;

- 10 oktaw
- 20-40-80-160-320-640-1280-2560-5120-10240-20480

• zakres dynamiki słuchu: 130 dB

2

Sygnal foniczny

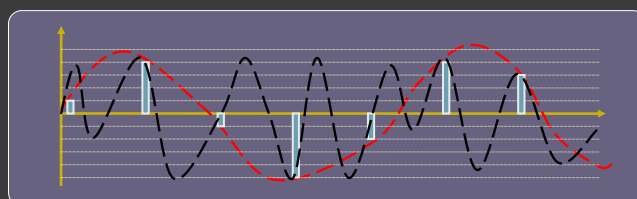


3

Cyfrowy sygnał foniczny

- ⦿ składa się z tzw. próbek pobieranych z określoną częstotliwością (szybkością) próbkowania

- częstotliwość próbkowania nie może być zbyt mała – bo nie będzie wiadomo jak naprawdę wygląda sygnał – częstotliwość próbkowania musi być dwa razy większa od maksymalnej częstotliwości sygnału
- im większa częstotliwość próbkowania, tym sygnał cyfrowy lepiej opisuje sygnał analogowy.



4

Parametry dźwięku

- ⦿ częstotliwości próbkowania (w Hz)
 - 8000 – jakość telefoniczna
 - 11025 – $\frac{1}{4}$ częstotliwości 44100
 - 16000 – stosowana w standardzie G.722
 - 22050 – $\frac{1}{2}$ częstotliwości 44100
 - 32000 – produkcja radiowa
 - **44100 – CD-Audio**
 - **48000 – częstotliwość studyjna, DVD, DVB, Blu-ray**
 - 88200 – brak typowych zastosowań
 - 96000 – 2x48000, produkcje wysokiej jakości, DVD, Blu-ray
 - 192000 – 2x96000, j.w.

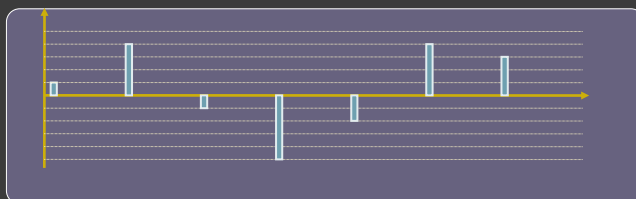
5

Cyfrowy sygnał foniczny

- ⦿ rozdzielczość bitowa – liczba bitów służąca do opisanie pojedynczej wartości pojedynczej próbki (słupka)
 - im więcej bitów służy do opisu danego dźwięku, tym dokładniej można opisać dany dźwięk
 - zakres dynamiki konwertera PCM wyraża się wzorem

$$S/N \cong 6n + 1,8 \text{ [dB]}$$

gdzie n - rozdzielczość bitowa konwertera



6

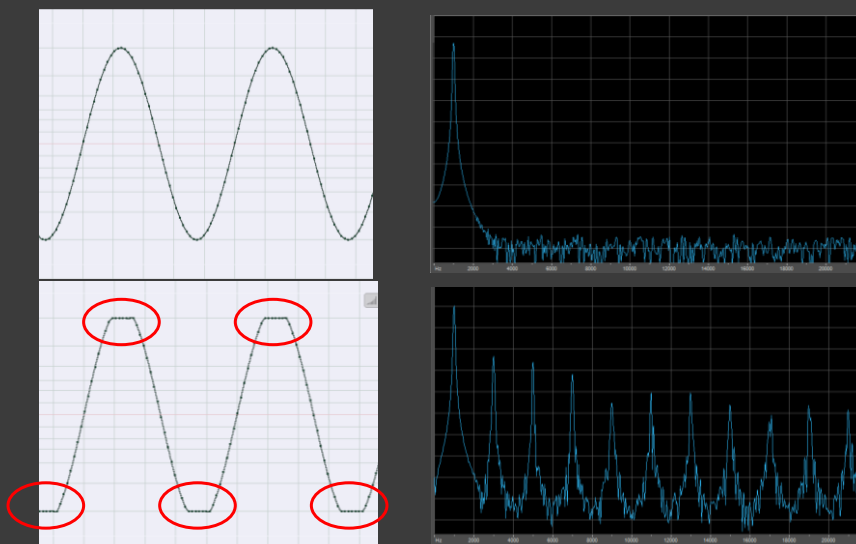
Parametry dźwięku

rozdzielczości bitowe:

- 8 bitów - czyli 2^8 możliwych wartości – 256
 - dźwięk zaszumiony, marnej jakości
- 16 bitów - czyli 2^{16} możliwych wartości – 65.536
 - najbardziej typowa rozdzielczość
 - odstęp sygnał szum rzędu 96dB
- 24 bity - czyli 2^{24} możliwych wartości – 16.777.216
 - zyskuje na popularności, używana w studiach
 - odstęp sygnał szum rzędu 144dB
- 32 bity - czyli 2^{32} możliwych wartości – 4.294.967.296
 - używana podczas wewnętrznego przetwarzania i miksowania plików (zapobieganie obcinaniu próbek)
- 32 bity float - 1 bit znaku, 8 bitów eksponenty, 23 bity mantysy
 - odstęp sygnał szum rzędu 1528dB!
 - wykorzystanie wielu stałoprzecinkowych przetworników A/C

7

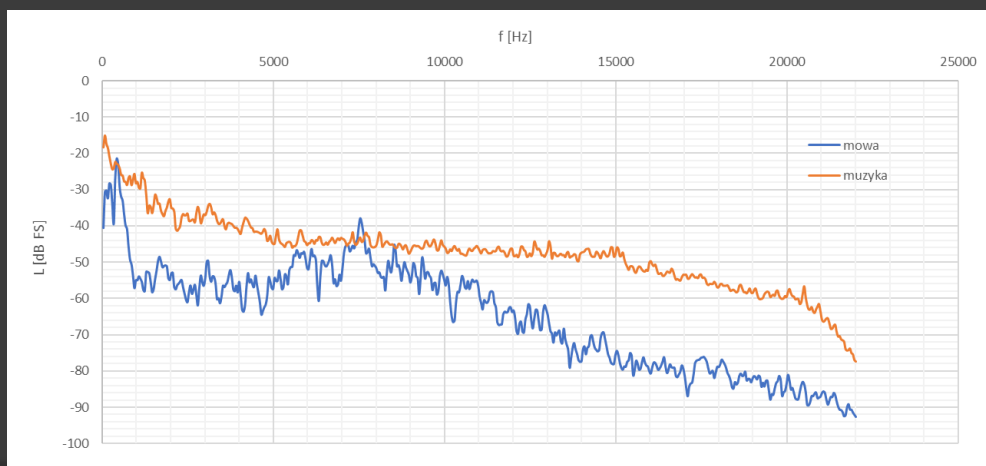
Przesterowanie dźwięku



8

Widmo dźwięku

- ◉ wizualizacja natężenia składowych dźwięku o różnych częstotliwościach



9

Kompresja

- ◉ 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.

10

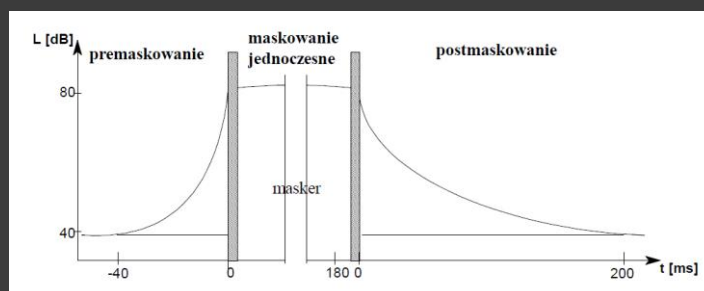
Kompresja

- ⦿ Metody bezstratne są mało efektywne
 - typowy stopień kompresji – 10-20%
 - maksymalny stopień kompresji – ok. 40-60%
- ⦿ Metody stratne charakteryzują się dużą efektywnością
 - stopień kompresji 90% przy akceptowalnej jakości dźwięku
 - wykorzystują niedoskonałości ludzkiego słuchu (kodowanie perceptualne)

11

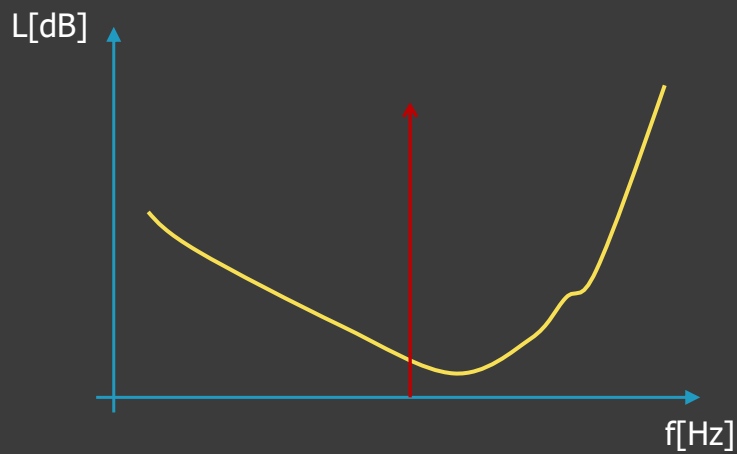
Kodowanie perceptualne

- ⦿ wykorzystuje przede wszystkim zjawisko maskowania (jednoczesnego i niejednoczesnego)
 - dźwięki o niższej amplitudzie i zbliżonej częstotliwości są „zagłuszone” przez dźwięki o wyższej amplitudzie



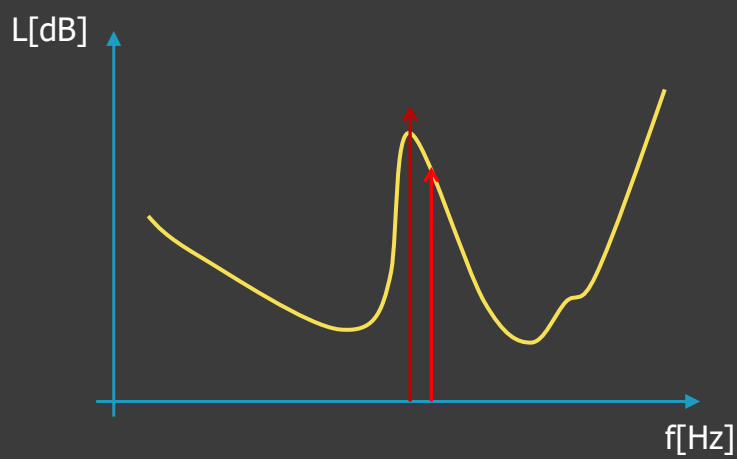
12

Ilustracja maskowania



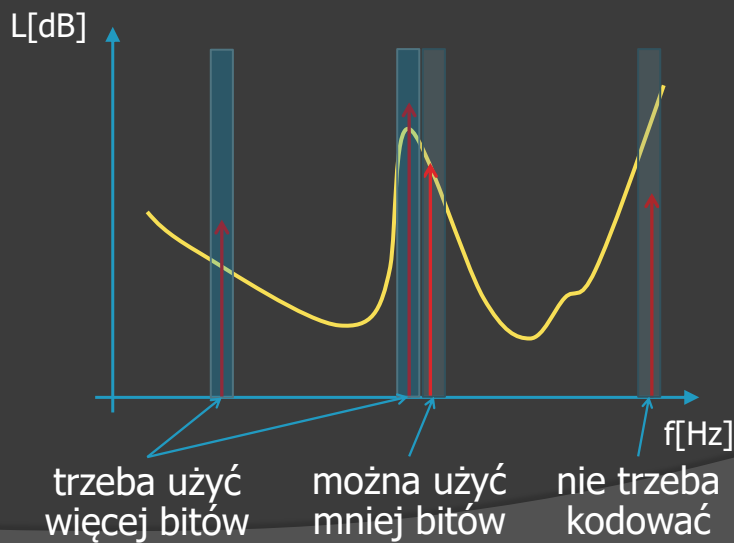
14

Ilustracja maskowania



15

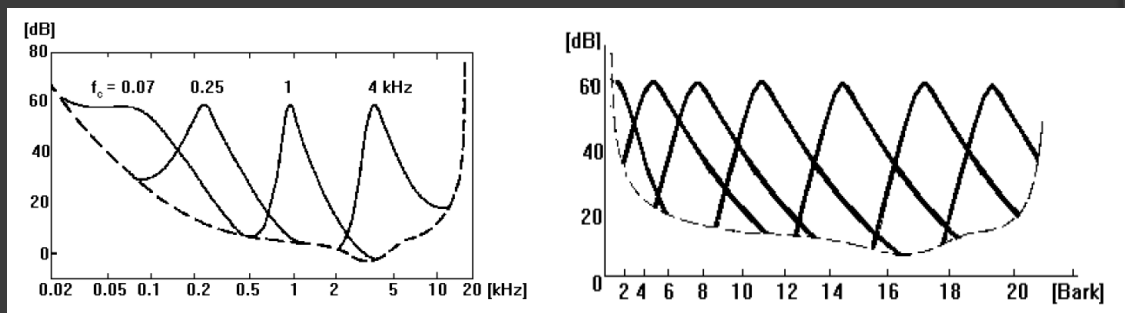
Ilustracja maskowania



16

Maskowanie

- maskowanie u każdego człowieka zachodzi nieco inaczej, dlatego kodeki używają uśrednionego modelu psychoakustycznego.
- znając składowe dźwięku maskowane w poszczególnych podpasmach, kodek usuwa je z sygnału



17

Parametry a wielkość pliku

- ⦿ 1 minuta nagrania w jakości telefonicznej
 - $60 \text{ [s]} \times 8 \text{ [bit]} \times 8000 \text{ [Sa/s]} \times 1 \text{ [kanał]} = 3,66 \text{ [Mbit]} = 468,75 \text{ [kB]}$
- ⦿ 1 minuta nagrania w jakości CD
 - $60 \text{ [s]} \times 16 \text{ [bit]} \times 44100 \text{ [Sa/s]} \times 2 \text{ [kanały]} = 80,75 \text{ [Mbit]} = 10,09 \text{ [MB]}$
- ⦿ 1 minuta nagrania w MP3/AAC z jakością zbliżoną do CD
 - $60 \text{ [s]} \times 160 \text{ [kbit/s]} = 9600 \text{ [kbit]} = 1,17 \text{ [MB]}$
- ⦿ 1 minuta nagrania 5.1 dla Blu-ray w najwyższej jakości
 - $60 \text{ [s]} \times 24 \text{ [bit]} \times 192000 \text{ [Sa/s]} \times 6 \text{ [kanałów]} = 1582 \text{ [Mbit]} = 197,75 \text{ [MB]}$

18

Przykłady formatów

19

Formaty dźwięku – kompresja bezstratna

- ⦿ wysoka jakość, ale często konieczne doinstalowanie dodatkowego oprogramowania
- ⦿ Free Lossless Audio Codec (*.flac)
 - kompresja rzędu 40-50%
 - liczba kanałów: 1 do 8
 - możliwość grupowania kanałów w celu poprawy wydajności kompresji
- ⦿ Monkey's Audio (*.ape)
 - Open Source
 - przyjmuje się, że stopień kompresji jest wyższy niż dla FLAC-a

20

Formaty dźwięku – WAVE (*.wav)

- ⦿ jeden z najpopularniejszych formatów w systemie Windows
- ⦿ **typowo** dane zapisane są w formacie PCM
 - możliwe inne formaty danych: ADPCM, u-Law, A-Law, LPC, GSM, CELP, G.721, G.723 a nawet MP3
- ⦿ obsługiwana liczba kanałów: 1, 2, 5.1 (i inne)
- ⦿ problem z plikami większymi od 4GB
 - rozwiązanie: format RF64
- ⦿ zastępowany przez format BWF (Broadcast Wave Format)

```
00000000: 52 49 46 46 54 20 E8 12|57 41 56 45 66 6D 74 20 | RIFFT 01WAVEfmt
00000010: 10 00 00 00 01 00 02 00|44 AC 00 00 10 B1 02 00 | + . 7 D- +±7
00000020: 04 00 10 00 64 61 74 61|30 20 E8 12 00 00 00 00 | J + data0 0 0
```

21

Formaty dźwięku – MPEG Layer 3 (*.mp3)

- ⦿ najpopularniejszy (?) format perceptualnej kompresji stratnej - a zarazem format pliku
 - opracowany na początku lat 90. ub. wieku
 - używa bardziej skomplikowanych modeli psychoakustycznych niż poprzednie wersje - warstwy (Layer 2, Layer 1)
 - wykorzystuje MDCT (*Modified Discrete Cosine Transform*)
- ⦿ przyjmuje się, że ucho nie dostrzeże różnicy, gdy przepływność na jeden kanał wynosić będzie 96kbit/s
- ⦿ obsługiwane przepływności: 32-320 kbit/s
- ⦿ obsługiwany także dźwięk 5.1

26

Formaty dźwięku – MPEG Layer 3 (*.mp3)

- ⦿ wykorzystywanie podobieństwa kanału lewego i prawego (np. tryb „joint stereo”) w celu poprawy wydajności kompresji
- ⦿ jakość kompresji zależy od implementacji algorytmu
 - kiedyś funkcjonowało dużo formatów pochodnych np. MP3 Pro, MP3 Surround
- ⦿ możliwość zapisu dodatkowych informacji tekstowych i graficznych (ID3 tags)

28

Formaty dźwięku – Advanced Audio Coding (*.aac)

- ⦿ następca MP3
 - rozwijany od 1997 roku
- ⦿ standard opisany w MPEG-2 Part 7 i MPEG-4 Part 3
- ⦿ nie jest zachowana kompatybilność w dół z wcześniejszymi wersjami kodeków opartych na standardach MPEG
 - pozwoliło to na osiągnięcie wyższej kompresji
 - wysoka jakość dźwięku 5.1 przy przepływnościach rzędu 320-430kbit/s
- ⦿ teoretycznie pozwala na obsługę do 48 kanałów
 - w tym mono, stereo, 5.1
- ⦿ częstotliwości próbkowania do 96kHz

30

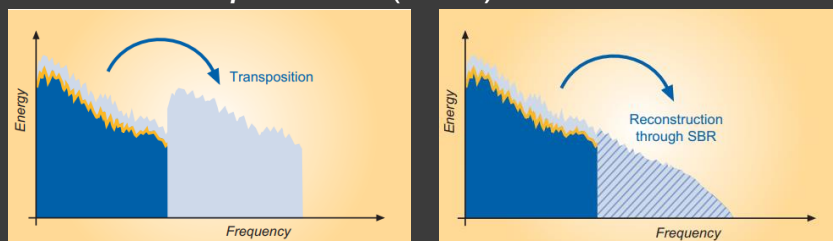
Formaty dźwięku – Advanced Audio Coding (*.aac)

- ⦿ mnóstwo wersji opracowanych pod konkretne zastosowania
 - LC- AAC – Low Complexity AAC
 - HE-AAC – High Efficiency AAC (także jako AAC+)
 - użycie *Spectral Band Replication* i *Parametric Stereo*
 - także wersje bezstratne i dopasowane do mowy
- ⦿ format typowo wykorzystywany w urządzeniach mobilnych
 - ale także w DVB i DAB+ (HE-AAC)

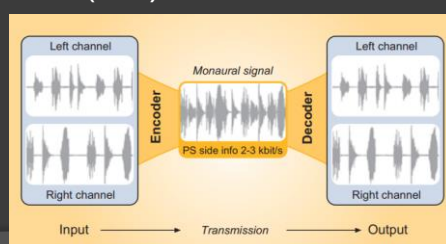
31

Formaty dźwięku – Advanced Audio Coding (*.aac)

◉ Spectral Band Replication (SBR)



◉ Parametric Stereo (PS)



źródło: Metzger Stefan, Moser Gerald (2006)
MPEG-4 HE-AAC v2 – audio coding for today's
digital media world

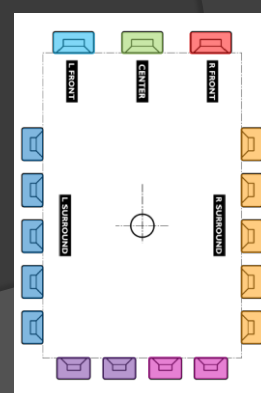
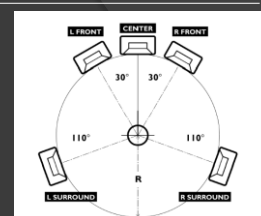
13.03.2025

32

Formaty dźwięku – formaty 5.1 / 7.1

◉ Dolby Digital (*.ac3) i DTS – Digital Theatre System (*.dts)

- dwa konkurujące ze sobą formaty kompresji
- typowo formaty stratne
 - na potrzeby Blu-ray powstały wersje z kodowaniem bezstratnym
- standardowo używane na DVD-Video, Blu-ray i UHD Blu-ray
- należy pamiętać, że oba formaty mogą być również użyte dla dźwięku monofonicznego bądź stereofonicznego

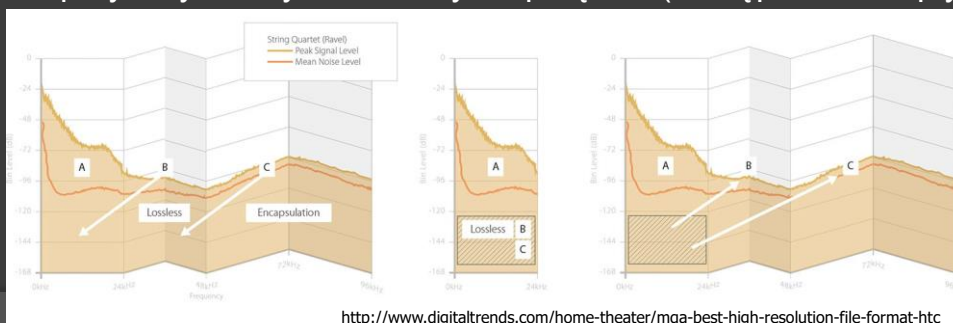


33

Przyszłość?

MQA (Master Quality Authenticated)

- format kompresji bezstratnej (?)
- polega na upakowaniu składowych wysokoczęstotliwościowych w paśmie do 20kHz
- przeznaczony do streamingu sygnału (np. Tidal)
- kompatybilny z dotychczasowym sprzętem (dostępne także płyty CD)



34

Przyszłość?

3D Audio (np. MPEG-H czy Dolby Atmos)

- kodowanie nie kanałów, a obiektów, np. dźwięk z trybun, głos komentatora itp.
- mikswanie dźwięku po stronie odbiorcy
- zwiększenie efektywności kompresji powinno umożliwić transmisję 14-18 kanałów przy przepływności rzędu 400 kbit/s (dla MPEG-H)
 - algorytmy bazują na AAC, ale nie będzie kompatybilności wstecz

13.03.2025

35

Dla zainteresowanych

- ⦿ John Watkinson, „The MPEG Handbook”, Focal Press, 2004.
- ⦿ <http://www.iis.fraunhofer.de/en/ff/amm.html>

13.03.2025