

PDiM FAQ

7 czerwca 2009

Spis treści

1	Percepcja dźwięku	2
2	Reprezentacje czasowo-częstotliwościowe	3
2.1	Metody poprawy czytelności spektrogramu – wymienić i krótko scharakteryzować (2 zdania + rysunek) co najmniej 3 takie metody.	3
2.2	Co to jest cepstrum i jak się je oblicza?	3
2.3	Zastosowania cepstrum	4
3	Modelowanie	4
4	Poprawa jakości dźwięku	5
4.1	W jakich sytuacjach sygnał muzyki może być zniekształcony zakłóceniami impulsowymi?	5
4.2	Wymienić i krótko opisać poznane metody odszumiania dźwięku	5
4.3	Kiedy zadanie usuwania szumu jest najłatwiejsze, a kiedy najtrudniejsze?	6
4.4	W jaki sposób działa odejbowanie widmowe?	6
5	Kodowanie mowy	6
5.1	Jak działa LPC?	6
5.2	Proszę podać krótki opis (2 zdania + rysunek) sposobu wyznaczania współczynników LP. Co opisują w dziedzinie częstotliwości współczynniki predyktora a_k ?	7
5.3	Co jest przedmiotem adaptacji w technikach kodowania mowy ADPCM? .	7
5.4	Jakie typy adaptacji można stosować?	7
5.5	Jaki jest zysk z adaptacji w ADPCM?	7
5.6	Jaka jest zasadnicza różnica pomiędzy ADPCM a LPC?	7
5.7	Jakie dane są przesyłane do dekodera w tych metodach?	8
5.8	Omówić zasadę działania techniki kodowania mowy CELP	8
6	Kodowanie dźwięku szerokopasmowego	8
6.1	Co to jest QMF?	8
6.2	Co to jest TDAC?	8
6.3	Jaką funkcję pełni przekształcenie MDCT w kodowaniu psychoakustycznym?	9

6.4	Jak działa model psychoakustyczny?	9
6.5	Wymienić etapy analizy sygnału dla modelu psychoakustycznego 1. lub 2. stosowanego w standardach MPEG	10
6.6	Do czego służy informacja uzyskiwana z modelu psychoakustycznego?	11
6.7	Narysować schemat blokowy kodera MPEG-4 AAC. Krótko (1 zdaniem) opisać elementy nowe w stosunku do techniki MPEG-1 L3	11
6.8	Do czego służy i na czym polega TNS?	11
6.9	Do czego służy i na czym polega PNS?	12
6.10	Podać schemat ideowy i opisać zasadę kodowania dźwięku szerokopasmowego w standardzie MPEG-4 AAC HE v2, z uwzględnieniem nowych narzędzi kodowania	12
6.11	Narysować schemat ideowy i opisać zasadę kodowania PS	12
6.12	Narysować schemat ideowy i opisać zasadę działania SBR	13
7	Kodowanie bezstratne i skalowalne	13
7.1	Opisać zasadę działania metod bezstratnej kompresji dźwięku	13
7.2	Jaki stopień kompresji jest typowo uzyskiwany w kompresji bezstratnej? Od czego on zależy?	14
8	Parametryczne metody kodowanie fonii	14
8.1	Jaką specjalną funkcjonalność, niedostępną w innych technikach, oferuje kodowanie parametryczne?	14
8.2	Jak przebiega proces modelowania sinusoidalno-szumowego?	15
8.3	Jak działa parametryczny koder dźwięku MPEG-4 SSC?	15
8.4	Jakie składniki sygnału są wyróżnione i w jaki sposób są modelowane?	16
8.5	A jak są reprezentowane?	16
8.6	Wymienić i krótko wyjaśnić metody stosowane do estymacji częstotliwości i amplitudy składowych sinusoidalnych dźwięku	16
9	Synteza dźwięków muzycznych	16
9.1	Wymienić poznane techniki syntezy dźwięku muzycznego.	16
9.2	Omówić zasadę działania syntezy FM.	17
9.3	Jaka metoda jest tania, oszczędza RAM, pozwala na uzyskanie ogromnej palety dźwięków, ale ma jedną poważną wadę, która spowodowała, że dzisiaj nie jest już powszechnie stosowana? Jaka to wada?	17

1 Percepcja dźwięku

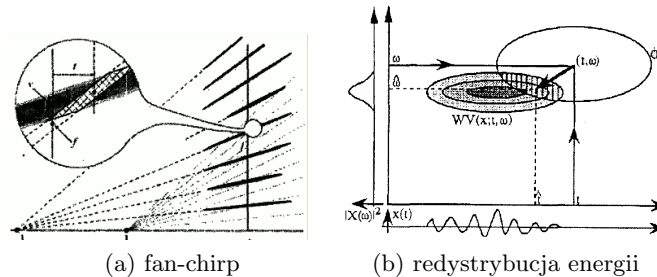
Póki co nic :(

2 Reprezentacje czasowo-częstotliwościowe

2.1 Metody poprawy czytelności spektrogramu – wymienić i krótko scharakteryzować (2 zdania + rysunek) co najmniej 3 takie metody.

przełączanie okien (z adaptacją) w zależności od lokalnych właściwości sygnału dobiera się optymalne okno wg pewnej miary, np lokalnego kontrastu lub koncentracji energii. W praktyce generuje się po prostu zbiór spektrogramów i dla każdego punktu wyznaczamy miarę w pewnym otoczeniu, dzięki czemu można dokonać wyboru, z którego z wyliczonych spektrogramów skorzystać

redystrybucja energii *time-frequency reassignment* polega na korekcji lokalizacji energii na płaszczyźnie czas-częstotliwość. Wyliczane są „środki ciężkości”, a energia jest przesuwana w ich stronę



(a) fan-chirp (b) redystrybucja energii
Poprawa czytelności spektrogramu

dystrybucja Wignera-Ville’a dystrybucja energii na płaszczyźnie TF wyznaczana dla sygnału analitycznego o jednostronnym widmie. Wysoka rozdzielczość wynika ze zdolności określanej jako de-chirping: analiza sygnału o liniowej zmianie częstotliwości daje rezultat o maksymalnym skupieniu. Istotnym problemem jest pojawianie się tzw. składników skośnych, które trzeba zidentyfikować.

transformacja fan-chirp funkcje bazowe są rozchodzącymi się „wachlarzowo” pod różnymi kątami chirpami. Trzeba wyznaczyć dla każdego okna częstotliwość podstawową i tempo jej zmian, żeby wyznaczyć optymalny współczynnik zakrzywienia

2.2 Co to jest cepstrum i jak się je oblicza?

Cepstrum jest rezultatem obliczenia transformaty Fouriera decybelowego widma sygnału. Logarytm z widma jest podawany na wejście transformacji. Tzw. cepstrum zespolone ma postać

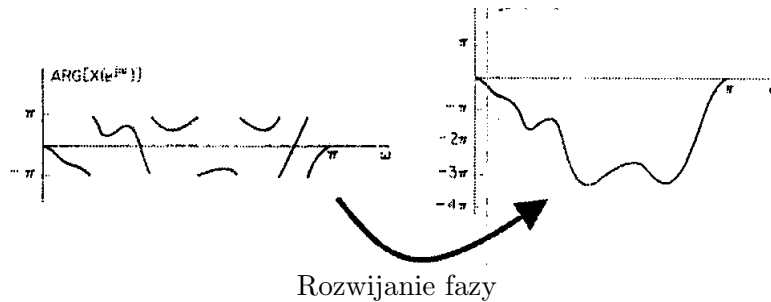
$$X(\tilde{t}) = \mathcal{F}^{-1} \left\{ \ln (\mathcal{F} \{x(t)\}) \right\}$$

gdzie \ln to logarytm zespolony, uogólnienie logarytmu naturalnego. Tak liczone cepstrum jest operacją odwracalną – teoretycznie możliwa jest całkowita rekonstrukcja sygnału dzięki przenoszonej informacji fazowej, choć w praktyce to trudne do realizacji. Trudność

polega na konieczności rozwinięcia fazy – patrz rysunek. Jest też cepstrum rzeczywiste, liczone na podstawie widma amplitudowego

$$X(\tilde{t}) = \mathcal{F}^{-1}\left\{\ln|\mathcal{F}\{x(t)\}|\right\}$$

Nie przenosi ono informacji fazowej, a zatem rekonstrukcja nie jest możliwa. Niektóre istotne własności:



- splot sygnałów w dziedzinie czasu jest równoważny sumie w dziedzinie pseudo-czasu
- okresowość w czasie skutkuje pojawieniem się wielu prążków w widmie, natomiast w dziedzinie cepstrum występuje tylko jeden prążek
- operację można odwrócić, jeśli sygnał jest minimalnofazowy
- cepstrum sygnału rzeczywistego, jest również rzeczywiste
- nawet dla rzeczywistych i minimalnofazowych sygnałów cepstrum jest nieskończenie długie
- przy praktycznym obliczaniu przez DFT jest cykliczne, co może powodować aliasing pseudo-czasowy (minimalizacja przez *zero padding*)

2.3 Zastosowania cepstrum

- rozpoznawanie dźwięku i mowy
- estymacja parametrów sygnału (np. częstotliwości podstawowej, charakterystyki)
- wygładzanie widmowe

Stosowane jest do rozpoznawania dźwięku i mowy oraz do estymacji parametrów sygnału (np. częstotliwości podstawowej, charakterystyki), wygładzanie widmowe.

Niezależnie od tego które cepstrum, wynik jest w dziedzinie pseudo-czasu.

3 Modelowanie

Póki co nic :(

4 Poprawa jakości dźwięku

4.1 W jakich sytuacjach sygnał muzyki może być zniekształcony zakłóceniami impulsowymi?

Trzaski z czarnej płyty, ze styków i gniazd, np przy łączeniu, etc.

4.2 Wymienić i krótko opisać poznane metody odszumiania dźwięku

Zakłócenia impulsowe

filtracja medianowa chyba nikomu nie trzeba mówić o co chodzi

algorytmy miękkodecyzyjne oparte np. o przetwarzanie wartości błędu predykcji (metody auto-regresyjne)

Redukcja szumu szerokopasmowego

filtracja Wienera estymacja zakłócenia na podstawie sygnału, filtracja odbywa się przy założeniu braku korelacji zakłócenia i sygnału użytecznego

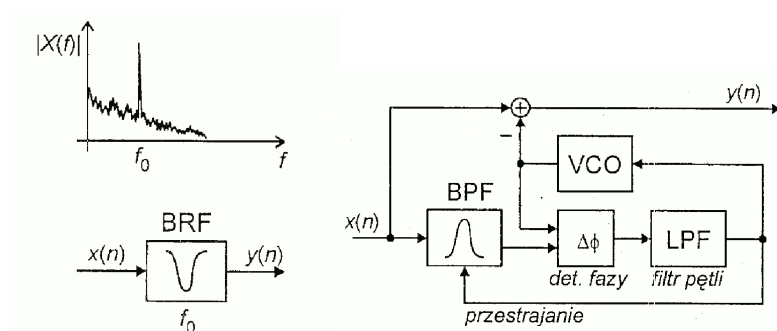
filtracja Kalmana jeżeli możemy założyć, że zakłócenie jest szumem gaussowskim; działa w oparciu o minimalizację autokowariancji błędu

odejmowanie widmowe jeżeli możemy założyć stacjonarność szumu, a w dodatku mamy próbkę samego szumu – jednak dla maksymalnej skuteczności próbka powinna być jak najdłuższa

metody statystyczne oparte na modelu statystycznym sygnału

filtracja falkowa zaletą jest możliwość dobrania zestawu falek do charakteru zakłócenia

Zakłócenia wąskopasmowe



(a) filtracja wąskopasmowej (b) resynteza zakłócenia z zachowaniem zgodności fazy

tzw. metody „inżynierskie”

Maskowanie brakujących fragmentów

metoda Burga polega na zastosowaniu predykcji liniowej wprzód i wstecz na próbkach sprzed i zza uszkodzenia. Polega na wyznaczeniu współczynników predykcji dla odpowiednich fragmentów (poprzedzającego i następującego, z odwróceniem czasu), a następnie na ekstrapolacji fragmentu przez kontynuację odpowiedzi filtru z tym samym pobudzeniem.

4.3 Kiedy zadanie usuwania szumu jest najłatwiejsze, a kiedy najtrudniejsze?

Najtrudniej jest w przypadku zakłócenia szumem addytywnym niestacjonarnym – kiedy widma sygnału i szumu nakładają się, a w dodatku widmo szumu jest zmienne w czasie, a więc nic nie da posiadanie próbki.

Względnie łatwe jest usuwanie zakłóceń impulsowych (filtracja medianowa), jednolitego, stacjonarnego szumu o znanym widmie (bardzo duża skuteczność odejmowania widmowego).

4.4 W jaki sposób działa odejmowanie widmowe?

5 Kodowanie mowy

5.1 Jak działa LPC?

LPC *linear predictive coding* kodowanie z użyciem predykcji liniowej. Sygnał mowy jest modelowany w dziedzinie następujących parametrów:

- współczynniki filtru a_k
- okres drgań
- wzmocnienie
- rodzaj pobudzenia

Parametry są kodowane i przesyłane do dekodera (serio!). W odbiorniku sygnał z generatora ciągów impulsów (źródła głosek dźwięcznych) lub z generatora szumu (źródła głosek bezdźwięcznych) podlega wzmocnieniu i jest przepuszczany przez filtr syntezujący o współczynnikach a_k odebranych z dekodera. W technice LPC wykorzystujemy fakt zależności statystycznej pomiędzy próbkami sygnału. Współczynniki a_k opisują odpowiedź impulsową przyczynowego inercyjnego systemu produkującego próbki sygnału.

Przy założeniu lokalnej stacjonarności sygnału w obrębie bloku i uśredniając po próbkach z jego wnętrza estymujemy $R_{xx}()$. By uniknąć efektów warunków brzegowych, modyfikujemy sygnał funkcją okna. Współczynniki a_k można wyznaczyć korzystając z:

- metody autokorelacyjnej, z uśrednianiem próbek w nieruchomym oknie
- metody kowariancyjnej, z uśrednianiem próbek ruchomym oknem o stałej długości

Dodatkowo to, co jest w 5.7.

5.2 Proszę podać krótki opis (2 zdania + rysunek) sposobu wyznaczenia współczynników LP. Co opisują w dziedzinie częstotliwości współczynniki predyktora a_k ?

Perwsze jest wyżej, a drugie to bieguny transmitancji.

5.3 Co jest przedmiotem adaptacji w technikach kodowania mowy ADPCM?

Adaptacyjnie modyfikowane są współczynniki predyktora. Dodatkowo adaptacyjnie może być dobierany także schemat kwantyzacji. Dostępne są takie duże rysunki w slajdach z wykładu, ale tu ich nie będzie. Choć umieć pewnie warto.

5.4 Jakie typy adaptacji można stosować?

Predykcja adaptacyjna:

AFP – *adaptive forward prediction*, adaptacja w przód. Wymaga przesyłania do dekodera aktualizacji współczynników predyktora.

ABP – *adaptive backward prediction*, adaptacja wstecz

Kwantowanie adaptacyjne:

AFQ – *adaptive forward quantization*, adaptacja w przód. Wymaga przesyłania do dekodera aktualizacji współczynników predyktora.

ABQ – *adaptive backward quantization*, adaptacja wstecz

5.5 Jaki jest zysk z adaptacji w ADPCM?

Adaptacja pozwala lepiej dostosować koder do własności sygnału. Przez własności sygnału rozumiemy nie tylko własności źródła (mówcy) i warunków pracy, np. szumów, ale także uwzględnienie niestacjonarności tych warunków, jak i samego sygnału. Zysk predykcji rośnie średnio z 10 do 14dB po zastosowaniu predyktora adaptacyjnego zamiast stałego (dla wyższych rzędów predyktora, rysunek w [3] #5, s. 7).

5.6 Jaka jest zasadnicza różnica pomiędzy ADPCM a LPC?

W ADPCM sygnał, a dokładnie rzecz biorąc kształt fali, reprodukowany jest na podstawie błędu predykcji, natomiast w LPC jest syntezerowany w dekoderyze na podstawie przesłanych współczynników filtru i innych parametrów.

5.7 Jakie dane są przesyłane do dekodera w tych metodach?

ADPCM	LPC
<ul style="list-style-type: none">• błąd predykcji• współczynniki predyktora (opcja)• schemat kwantyzacji (opcja)	<ul style="list-style-type: none">• informacja V/U ¹• okres pobudzenia impulsowego• wzmacnienie• bieguny transmitancji

5.8 Omówić zasadę działania techniki kodowania mowy CELP

CELP *code excited linear prediction*, zamiast modelowania sygnału pobudzenia jest książka kodowa, zawierająca typowo 512 lub 1024 pobudzenia (kodowanie wektorowe). Wybierana jest sekwencja o najmniejszej średniokwadratowej różnicy między ramką oryginalną a „książkową”. Wyznaczony sygnał błędu jest perceptualnie ważony w celu wykorzystania zjawiska maskowania równoczesnego. CELP umożliwia uzyskanie małych przepływności (2–4kbps) przy zachowaniu dużej naturalności syntetyzowanego sygnału. Wadą jest duża złożoność obliczeniowa – każdym kroku należy spróbować resyntezy 512/1024 razy.

6 Kodowanie dźwięku szerokopasmowego

6.1 Co to jest QMF?

QMF *quadrature mirror filter*, to zestaw filtrów o symetrycznych charakterystykach, zaprojektowany tak, by spełnić tzw. warunki idealnej rekonstrukcji, co umożliwia znoszenie się składowych aliasowych. Pomimo, że w poszczególnych zdecydowanych składnikach subpasmowych występują składowe aliasowe, to po ich ponownym zsumowaniu okazuje się, że mają parami przeciwne znaki i się znoszą. Jest to efekt FDAC (*frequency domain aliasing cancellation*), analogicznie do TDAC.

6.2 Co to jest TDAC?

TDAC *time domain aliasing cancellation*, efekt kompensacji aliasingu w dziedzinie czasu. DCT przekształca wektor N próbek w wektor $N/2$ współczynników, co grozi aliasingiem w dziedzinie czasu. Jednak symetryczne własności DCT powodują, że w każdej z połówek okna aliasing ma przeciwny znak – a więc składowe aliasowe ponownie się znoszą.

¹(voiced/unvoiced), możliwe stany pośrednie, „mieszane”

6.3 Jaką funkcję pełni przekształcenie MDCT w kodowaniu psychoakustycznym?

MDCT jest kluczowym elementem odróżniającym warstwę 3. MPEG-1 od 1. i 2. Realizuje ideę tzw. transformacji zakładkowej, co pozwala uniknąć efektu blokowego znanego chociażby z JPEG, który słyszalny byłby co kilka milisekund (tyle trwa jeden blok) – a więc przy praktycznie takiej samej złożoności obliczeniowej i wymaganej przepływności uzyskujemy dużo lepszą jakość.

W porównaniu do DCT MDCT jest niewrażliwy na skoki fazy i daje czystsza reprezentację tonalną (rysunek z sinusem przesuniętym o 1 próbkę ([3] #7, s.10) – czego byśmy z sygnały nie wycięli, nie pojawiają się listki boczne (ponownie rysunek).

Przekształcenie można zinterpretować podobnie do DFT z doklejonym odbiciem lustrzanym sygnału i decymacją razy dwa. W efekcie uzyskujemy sygnał parzysty bez skoku wartości na końcu przedziału. W zwykłym DFT było milczące założenie, że sygnał jest okresowy, a nie był, stąd skok na końcach.

Jest liczone w blokach o długości N , a transformata ma długość $N/2$. Taka transformacja jest nieodwracalna. Jednak w związku z tym, że mamy TDAC, możemy dokonać idealnej rekonstrukcji. Warunkiem jest odpowiedni kształt funkcji okna $h(n)$ spełniający warunek:

$$h^2(n) + h^2(n + \frac{N}{2}) = 1$$

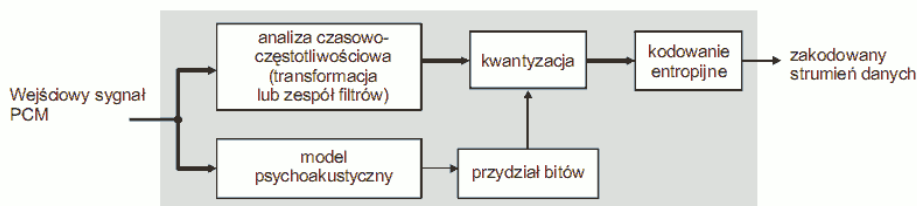
Jest on związany po pierwsze z tym, że gdyby zakładki nie sumowały się do 1, to wyszłaby z tego modulacja AM. Po drugie zaś, okno jest stosowane w obliczeniach dwa razy – stąd kwadraty.

Ważne jest też to, że MDCT można potraktować jako drugi, po analizie subpasmowej przy pomocy zespołu filtrów QMF, etap analizy częstotliwościowej. Dzięki dużo wyższej rozdzielczości ($32 \times 36 = 576$ pasm) takiej hybrydowej analizie uzyskujemy dokładniejsze kształtowanie błędu kwantyzacji w dziedzinie częstotliwości.

6.4 Jak działa model psychoakustyczny?

W modelach psychoakustycznych wykorzystywane są takie własności ludzkiego słyszenia, jak:

- zależność progu słyszenia od częstotliwości
- maskowanie
 - jednoczesne, tj. w dziedzinie częstotliwości
 - w przód
 - wsteczne
- niedokładność percepcji częstotliwości
- niedokładność percepcji głośności



Model psychoakustyczny, ogólny schemat

6.5 Wymienić etapy analizy sygnału dla modelu psychoakustycznego 1. lub 2. stosowanego w standardach MPEG

Model psychoakustyczny w MPEG-1, warstwy 1,2

Wykorzystuje fakt, że szum maskuje około 10–20dB lepiej niż ton, a funkcja maskowania jest nieliniową funkcją czasu i częstotliwości.

1. Wyliczamy chwilowe widmo w bloku
2. Znajdujemy tony powyżej progu (prosty warunek amplitudowo-szczytowy) – pozostałe składniki są interpretowane jako szumowe
3. Dla każdego maskera konstruowana jest funkcja maskowania przez splot z predefiniowaną nieliniową funkcją
4. Superpozycja „trójkątów” w dziedzinie mocy i mamy gotowy próg maskowania (wszystko na raz, zamiast iteracyjnie, dla uproszczenia algorytmu, żeby działał w latach 80’)

Wiemy już, które subpasma są „zamaskowane”. Następnie wyznaczamy minimum progu w danym subpasmie \approx maksimum dopuszczalne szumu kwantyzacji (plus mały zapas). Dysponujemy skończoną liczbą bitów na kwantyzację, przydzielamy je subpasmom z uwzględnieniem progu maskowania wg algorytmu:

1. Wybieramy subpasmo w którym S/N_Q jest najgorszy
2. Poprawiamy kwantyzację w tym pasmie o 1 bit
3. Powrót do 1., chyba że nie ma już bitów do przydzielenia.

Model psychoakustyczny w MPEG-1, warstwy 3

Podstawowa różnica: brak ostrej klasyfikacji ton/szum. Zdefiniowana jest miara tonalności

$$t(b) = -0.299 - 0.43 \ln(\tilde{c}_s(b))$$

określająca tonalność w zakresie 0–1, zależną od miary niedeterminizmu $\tilde{c}_s(b)$. Już bez zabawy we wzory, działa tak, że jeżeli sygnał bardziej niedeterministyczny, tym mniej stacjonarny i rośnie prawdopodobieństwo, że jest szumem.

6.6 Do czego służy informacja uzyskiwana z modelu psychoakustycznego?

Właściwie jest to wyjaśnione już powyżej.

6.7 Narysować schemat blokowy kodera MPEG-4 AAC. Krótco (1 zdaniem) opisać elementy nowe w stosunku do techniki MPEG-1 L3

TNS *temporal noise shaping*, kształtowanie obwiedni czasowej szumu kwantyzacji w zależności od chwilowej amplitudy sygnału

predykcja predykcja w dziedzinie czasu współczynników MDCT

PNS *perceptual noise substitution*, percepcyjne zastąpienie szumu – części widma zidentyfikowane jako szum są pomijane w koderze, a potem rekonstruowane pseudolosowym szumem syntetycznym

LTP *long term prediction*, mechanizm alternatywny dla złożonych obliczeniowo predyktorów z MPEG-2 AAC

TwinVQ *vector quantization*, alternatywny mechanizm kwantyzacji, oparty na kwantowaniu wektorowym

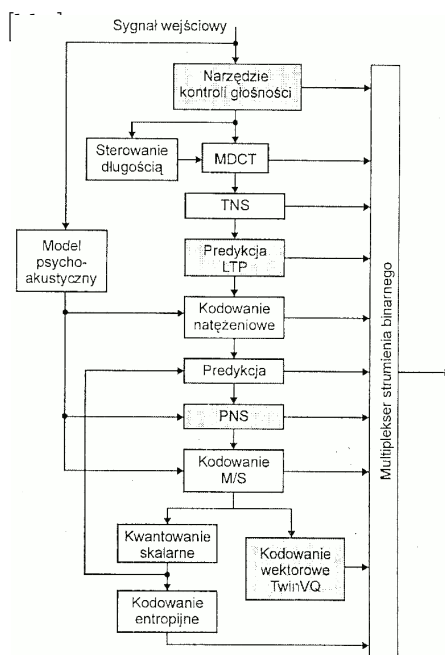
kontrola głośności globalne sterowanie natężeniem dźwięku

Poza występującymi w schemacie blokowym jest też wiele innych nowości, m.in.:

- lepszy model psychoakustyczny
- lepsze kodowanie entropijne
- większy zakres częstotliwości próbkowania
- jednostopniowa analiza częstotliwościowa w postaci MDCT o dużej rozdzielczości (1024 próbki lub 8×128)
- obsługa do 48 kanałów (zamiast 2/6)
- dostępna alternatywa dla okna Hanna – tzw. okno Keysera-Bessela, nie opisane analitycznie, a przez tablicę próbek

6.8 Do czego służy i na czym polega TNS?

TNS jest metodą kształtowania obwiedni czasowej szumu kwantyzacji, ma na celu przede wszystkim redukcję pre-echa, co wpływa na poprawę jakość kodowania transientów. Kształtowanie szumu



Nie pomylić z MPEG-2 AAC, gdzie nie ma LTP, PNS, TwinVQ i kontroli głośności!

MPEG-4, general audio

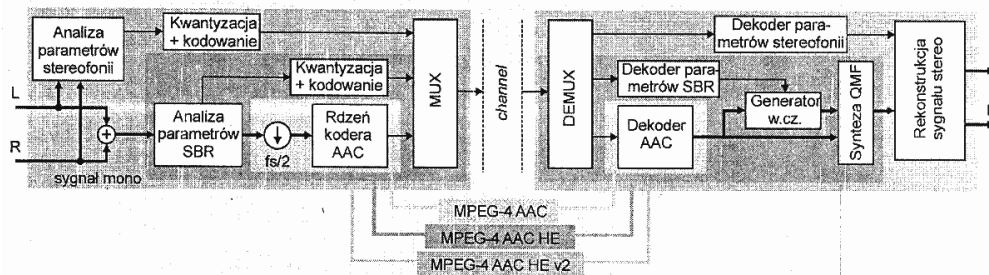
kwantyzacji opiera się na dualizmie DCT: w stanach nieustalonych bezpośrednio kwantowanie współczynników zastąpione jest przez kwantowanie błędu predykcji w kierunku od najniższych do wyższych częstotliwości, co powoduje, że dystrybucja błędu kwantyzacji w dziedzinie czasu przypomina obwiednię czasową sygnału.

6.9 Do czego służy i na czym polega PNS?

Grupy prążków widmowych zidentyfikowane jako szum pomijane są w koderze i zrekonstruowane z wartości losowych w dekoderze, do którego przesyłana jest tylko informacja o jego obwiedni i energii. Rekonstrukcja sygnału polega na generacji przebiegu pseudolosowego o równoważnej energii i obwiedni widma, który dla słuchacza jest nieodróżnialny od oryginalnego szumu.

6.10 Podać schemat ideowy i opisać zasadę kodowania dźwięku szerokopasmowego w standardzie MPEG-4 AAC HE v2, z uwzględnieniem nowych narzędzi kodowania

HE dodaje
SBR, HEv2
dodaje PS



Ogólna idea profilu *high efficiency*

HEv2 używa nowych narzędzi poprawiających jakość dźwięku przy niskich przepływnościach

PS *parametric stereo*, stereofonia parametryczna.

SBR *spectral band replication*, powielanie widma.

Istotne (także w kontekście poniższych pytań) jest to, że dekodery obsługujące tych narzędzi odtworzy pasmowo obcięty (SBR) sygnał mono (PS). Na schemacie widać, że osobno przesyłane jest „normalne” AAC, a osobno informacje dodatkowe. Dziwnym nie jest zatem, że w praktyce narzędzia te nadają się do zastosowania z dowolną inną niż AAC techniką.

Poza tym wymagania na zespoły filtrów QMF dla obu technik są bardzo podobne, przy jednoczesnym ich zastosowaniu można wszystko liczyć w tej samej dziedzinie subpasmowej – oszczędność na złożoności.

6.11 Narysować schemat ideowy i opisać zasadę kodowania PS

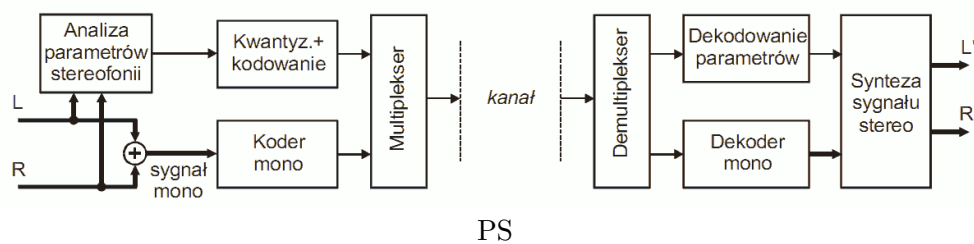
Metoda przesyłania sygnału stereofonicznego jako mono + parametryczna informacja różnicowa o stereo (niewielka, koło 2kbps). Przesyłane parametry to:

IPD *interchannel phase difference*, opóźnienie aka różnica czasu/fazy między kanałami

IPD *overall phase difference*, summaryczna różnica fazy (?)

IID *interchannel intensity difference*, różnica natężenia między kanałami.

IC *interchannel coherence*(?), stopień korelacji między kanałami



W dekodery jest analiza subpasmowa sygnału mono, a sygnał jest poddawany de-korelacji przez splot z szumem białym – powstają w ten sposób dwa nieskorelowane kanały, a de-korelacja jest potrzebna do uzyskania szerokości panoramy stereo. Następnie dodawane są parametry, a na koniec kolejny zespół filtrów syntezuje z subpasm lewy i prawy kanał.

6.12 Narysować schemat ideowy i opisać zasadę działania SBR

Ogólnie rzecz biorąc, lepiej zmniejszyć szerokość pasma, częstotliwość próbkowania niż żeby zabrakło bitów na środek i dół pasma, gdzie jest najistotniejsza część dźwięku. SBR stosuje się, aby zminimalizować subiektywne wrażenie utraty wysokich tonów po ograniczeniu widma poprzez rekonstrukcję po stronie dekodera brakującej części widma.

Metoda wykorzystuje podobieństwa między górą a dołem widma (np. regularne szeregi harmoniczne) oraz ciągłość widma chwilowego. Po zdekodowaniu sygnału odtwarza się górę jako kopię dołu skalowaną energetycznie zgodnie ze współczynnikami estymowanymi w koderze. W krytycznych przypadkach, gdy nie da się odtworzyć informacji na podstawie replikacji, brakujące składowe są syntetyzowane.

Z uwagi na manipulację subpasmami konieczne jest udoskonalenie metod analizy subpasmovej – teraz aliasing już się nie kasuje. Stosuje się dwa rozwiązania:

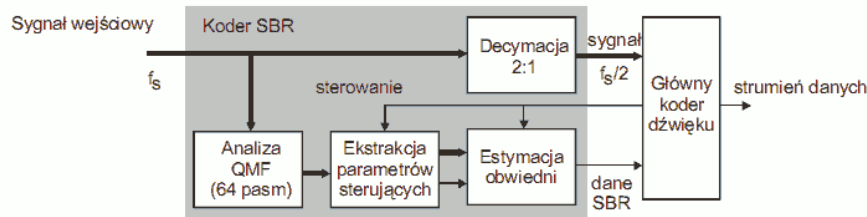
- tryb **HQ** opierający się na zespolonej reprezentacji sygnałów subpasmych
- tryb **LC** pozostaje przy mniej złożonej reprezentacji rzeczywistej, jednak konieczny jest dodatkowy blok korekcji subpasm, minimalizujący zniekształcenia aliasowe

7 Kodowanie bezstratne i skalowalne

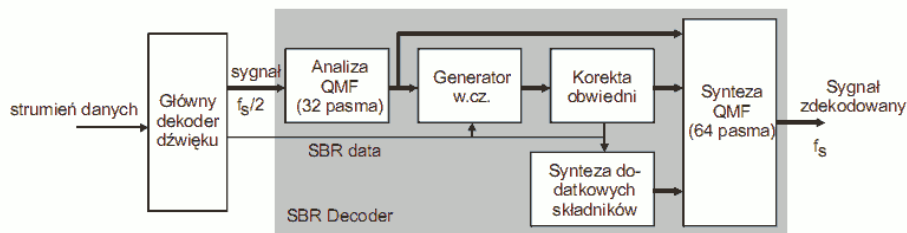
7.1 Opisać zasadę działania metod bezstratnej kompresji dźwięku

Kodowanie bezstratne polega w ogólności na de-korelacji sygnału i kodowaniu entropijnym.

Dwie filozofie kodowania bezstratnego:



(a) koder



(b) dekodek
SBR

- kodowanie oparte na bezstratnej predykcji, z predyktorem opartym na arytmetyce całkowitoliczbowej lub kwantyzacji wyniku predykcji. Błąd predykcji jest liczbą całkowitą i może być wtedy zakodowany entropijnie, kodami o zmiennej długości (np. Huffman)
- kodowanie stratne + kodowanie reszty, które wykorzystuje dowolną technikę dekorrelacji (predykcja lub kodowanie transformatowe) oraz kodowanie błędu aproksymacji.

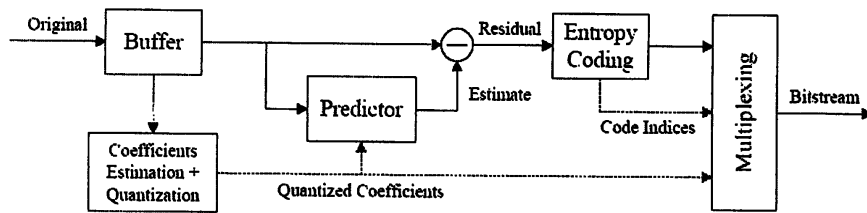
7.2 Jaki stopień kompresji jest typowo uzyskiwany w kompresji bezstratnej? Od czego on zależy?

Na ogół 1:1.5 do 1:2, góra 1:3. Zależy od własności konkretnego dźwięku – od jego złożoności statystycznej. W związku z tym nie ma też za bardzo możliwości regulacji – tzn. można zrobić trochę gorszą kompresję kosztem mniejszej złożoności obliczeniowej, ale to wszystko. Niezależnie od tego, wyższe częstotliwości próbkowania oznaczają wyższą korelację między próbkami, a więc możliwość większej efektywności.

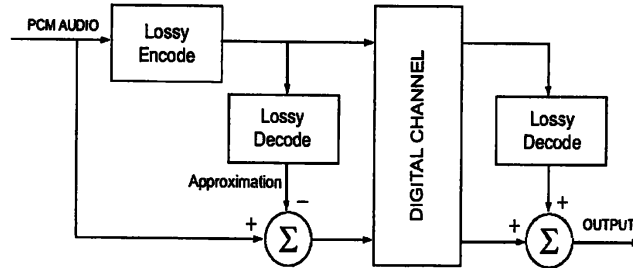
8 Parametryczne metody kodowanie fonii

8.1 Jaką specjalną funkcjonalność, niedostępną w innych technikach, oferuje kodowanie parametryczne?

Możliwość zmiany wysokości i długości dźwięków bez strat jakości, przepróbkowywania, etc.



(a) bezstratna predykcja



(b) stratne + reszta

filozofie kodowania bezstratnego

8.2 Jak przebiega proces modelowania sinusoidalno-szumowego?

Sygnal aproksymowany jest przez sumowanie sinusoid o zmiennej częstotliwości i amplitudzie oraz szum od zmiennej wariancji i obwiedni widma. Składniki harmoniczne wyróżniamy kryterium amplitudowym czyli zakładamy, że szum ma mniejszą energię, prążki przekraczające wygładzoną funkcję $X(f)$ to składniki tonalne. Kryterium harmoniczności czyli składniki tonalne wielokrotnością podstawowej. Szum modelowany metodą LCP z residuum czyli pozostałości po odjęciu składników tonalnych.

8.3 Jak działa parametryczny koder dźwięku MPEG-4 SSC?

Koder dokonuje dekompozycji sygnału na szerokopasmowe transjenty, indywidualne składowe sinusoidalne oraz szum o określonej obwiedni widmowej i amplitudowej.

Dekompozycja sygnału oryginalnego podzielonego na bloki próbek (ramki) postępuje wg scenariusza analizy przez syntezę: zidentyfikowane składniki są stopniowo odejmowane, a pozostałość jest przekazywana do dalszych etapów modelowania.

1. Wykrywanie transientów, klasyfikowane jako skokowe (modelowane prostym zbroczem liniowym) lub impulsowe (modelowane tzw. funkcją Meixnera).
2. W pozostałym sygnale wykrywane są składowe quasi-sinusoidalne w oparciu o analizę FFT o dużej rozdzielczości.
3. Pozostałość sygnału po odjęciu zakodowanych sinusoid jest traktowana jako niestacjonarny szum. Obwiednia widmowa tego szumu jest modelowana przy pomocy współczynników filtra rekursywnego kształtującego widmo szumu białego.
4. Ostatnim etapem modelowania jest parametryzacja obwiedni amplitudowej szumu

Do kodowania wybierane są wyłącznie składowe istotne percepcyjnie, czyli takie, których amplitudy przekraczają próg słyszenia wyznaczony przez model psychoakustyczny.

Amplitudy i częstotliwości składowych sinusoidalnych połączone w trajektorie są skwantowane nierównomiernie z dokładnością odpowiadającą przeciętnym granicom percepcji i zakodowane różnicowo.

8.4 Jakie składniki sygnału są wyróżnione i w jaki sposób są modelowane?

Transjenty, tony, szum. Szczegóły powyżej.

8.5 A jak są reprezentowane?

Mniej więcej powyżej.

8.6 Wymienić i krótko wyjaśnić metody stosowane do estymacji częstotliwości i amplitudy składowych sinusoidalnych dźwięku

metoda McAulay-Quatieriego estymacja składników sinusoidalnych na podstawie widma chwilowego (FFT) kolejnych bloków sygnału (z pewnym umownym progiem detekcji) Wymagana jest duża rozdzielczość DFT, a i tak bliskie prążki przesuwają swoje szczyty (bo mają niezerową szerokość).

analiza przez resynteze dokładniejsza wersja, choć bardziej złożona obliczeniowo. Wykrywamy składniki po jednym i odejmujemy od widma. Wbrew pozorom DFT liczymy tylko raz.

Dodatkowo wpływają na to trajektorie, bo staramy się zachować ich gładkość, np. predykcją czy metodami probabilistycznymi (łańcuchy Markowa). Można też dopisać coś z SSC/HILN.

9 Synteza dźwięków muzycznych

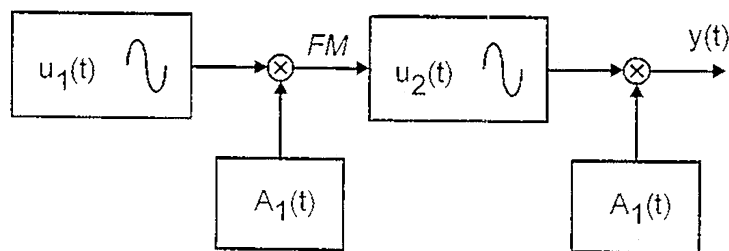
9.1 Wymienić poznane techniki syntezy dźwięku muzycznego.

- Metody widmowe
 - synteza addytywna
 - synteza subtraktywna
- Metody przetwarzania sygnału zarejestrowanego
 - synteza tablicowa (*wavetable*)
 - synteza przez sampling
 - synteza granularna
- Metody abstrakcyjne
 - modulacja częstotliwości

- nieliniowe kształtowanie fali
- synteza fraktalna
- Metody wykorzystujące modelowanie
 - modelowanie fizyczne
 - metoda falowodowa

9.2 Omówić zasadę działania syntezy FM.

Sygnal powstaje przez modulację kąta (*aka* częstotliwości, ale modulacja fazowa jest łatwa do realizacji cyfrowej) nośnej kosinusoidalnej o częstotliwości f_n zgodnie z wartościami modulatora, który, podobnie jak nośna, pracuje w zakresie akustycznym.



PS

Przy ustalonym stosunku $\frac{f_0}{f_n}$ zmieniając wyłącznie indeks modulacji $A_2(t)$ uzyskujemy zmianę widma w szerokim zakresie – dobry parametr do manipulacji dźwiękiem wynikowym.

Strukturę syntezy FM tworzą tzw. operatory składające się z oscylatora, wzmacniacza oraz generatora obwiedni. Ich ilość i rodzaj określa możliwości brzmieniowe instrumentu – można uzyskać właściwie dowolnie skomplikowane widma stosując różne połączenia.

9.3 Jaka metoda jest tania, oszczędza RAM, pozwala na uzyskanie ogromnej palety dźwięków, ale ma jedną poważną wadę, która spowodowała, że dzisiaj nie jest już powszechnie stosowana? Jaka to wada?

Subtraktywna? Jest prosta w implementacji, *chyba* zużywa mało pamięci, generuje dużo dźwięków. Ale trudno uzyskać dobrą symulację zwykłych, akustycznych instrumentów i naturalnego ich brzmienia.

Synteza addytywna, która w dziedzinie cyfrowej jest łatwa w implementacji, generuje realistyczne dźwięki instrumentów, pamięci raczej za dużo nie używa, jednakże posiada ogromny zestaw parametrów wpływających na barwę dźwięku – co powoduje problemy ze sterowaniem.

Rozpoznawanie i synteza mowy

Póki co nic :(

Od redakcji :P

Opracowanie „zawiera” (bedzie zawierać) pytania z egzaminów z lat 2005, 2006 oraz 2008, *chyba* wszystkie, za wyjątkiem tych z adnotacjami typu „p. . . , nie robię”, oraz *być może* pewną ilość pytań dodatkowych. Jest „under development” i pewnie nigdy nie będzie inne, bo przed egzaminem nie trafi do niego wszystko co trzeba, a po egzaminie nie będzie mi się za bardzo chciało nic dopisywać.

Literatura

- [1] A. Spanias, T. Painter, V. Atti *Audio signal processing and coding*, Wiley Inter-Science, 2007.
- [2] M. Bartkowiak *State-of-the-art in audio coding*, Politechnika Poznańska, 2006.
- [3] M. Bartkowiak, *Wykład i materiały do wykładu „Przetwarzanie dźwięku i mowy”*, Politechnika Poznańska, 2008/2009.
- [4] J. Wales, et al. *Wikipedia*.
- [5] M. Bartkowiak, *Wysokorozdzielcze i adaptacyjne metody analizy widmowej sygnałów fonicznych*