

Kóder a dekóder audio signálov

1.1 Úvod

Za účelmi šírenia a archivovania audio signálov v digitálnej podobe je často žiadúce aby množstvo dát bolo čo najmenšie . Dôvody sú veľmi jednoduché , menšie náklady , úspora prenosového pásma , praktickejšie archivovanie , “viac muziky za menej peňazí” . Na tento účel sa používa metóda takzvanej stratovej kompresie , pri ktorej dochádza k bez stratovej kompresii ale len ohraničeného informačného obsahu vstupných dát . Pre kódovanie (stratovú kompresiu) audio signálu sa používa tzv. “perceptual audio coding“ pri ktorom sa využíva fakt že ľudský zvukový systém vie zachytiť len obmedzené množstvo informácii . Je to spôsobené z veľkej časti fyziologicky , a rovnako aj z dôvodu aby nedošlo k zahltenu informáciami ktoré nie sú potrebné . Podobne ako zrak keď človek vníma cele zorné pole zdanlivo rovnako ostro ale v skutočnosti je celá scenéria skladaná v mozgu , alebo keď vidíme z veľkej diaľky ľubovoľný predmet sme ho schopný rozpoznať len podľa jeho špecifických črtí a myslíme si že ho vidíme do väčších detailov , ale v skutočnosti veľa nevidíme len vieme že to tam niekde je . Pri “perceptual audio coding“ je snahou aby zakódované audio bolo vnímané rovnako ako pôvodné a to pri strate alebo znehodnotení niektorých jeho informácii .

1.2.1 Porovnanie stratových kompresných algoritmov

V tejto časti upriem pohľad predovšetkým na algoritmy “MPEG-1 layer3“ ďalej mp3 , a “MPEG-4 Advanced Audio Coding“ ďalej “AAC“ ktoré boli vyvinuté organizáciou Mpeg [Mp3aac] . Kompresný algoritmus AAC (prípadne AAC Plus , vid' ďalej) sa v súčasnosti považuje za state-of-art v stratovom kódovaní audia . Hoci AAC predstavuje následníka algoritmu mp3 , predsa majú dosť veľa spoločného.

Definícia PCM (Pulse Code Modulation) :

Nech je daná funkcia $s(t)$, spojitá v čase t , a konštanta T . Potom PCM je postupnosť hodnôt $\{ s_i \}$ pre $i = -\infty .. 0 .. \infty$, taká že platí $s_i = s(i.T)$. Každý prvok postupnosti $\{ s_i \}$ nazveme vzorkou , a konštantu T nazývame periódou vzorkovania.

Mp3 aj AAC používajú MDCT [Mdct] (Modified Discrete Cosine Transform) transformáciu ktorá sa používa na zmenu reprezentácie signálu z PCM do spektrálnej reprezentácie . MDCT sa neaplikuje signál ako celok , ale tak že sa PCM postupnosť roztrhá na fragmenty dĺžky n a tieto sa potom transformujú na postupnosť hodnôt dĺžky n (teda MDCT sa vykonáva na n -ticiach hodnôt) . Abstraktne sa na MDCT možno pozerat' ako na regulárnu transformáciu dimenzie n ktorá je ortogonálna , pričom táto transformácia je zmenou bázy (lineárna algebra) kde báзовé funkcie / vektory sú sínusoidy o rôznych vlnových dĺžkach .

Mp3 aj AAC umožňujú použiť dva rôzne spôsoby kódovania stereo informácie (nad spektrálnou reprezentáciou) . Prvý spôsob je tzv. “Intensity Stereo“ (prípadne Joint Stereo) . Princíp spočíva v tom že signál sa ako dva plnohodnotné kanály kóduje len v frekvenčnom pásme od 0 Hz .. 2000 Hz (2000 je len približný údaj) . Zvyšné pásmo 2000 Hz .. 16000 Hz , je prenášané len ako mono signál (súčet kanálov) spolu s niekoľko málo informáciami o energiách v jednotlivých kanáloch tohto pásma . Pri dekódovaní je potom tento mono signál

umiestnený medzi kanály tak aby bol zachovaný pomer energii tak ako boli namerané v zdrojovom signále .

Druhý spôsob kódovania sterea je "Middle/Side Coding" . Na rozdiel od predošlého toto kódovanie je v zmysle sterea bezstratové . Využíva sa predpoklad že obsah kanálov je značne korelovaný . Preto namiesto kódovania hodnôt každého kanála sa kóduje súčet (middle) a rozdiel (side) kanálov . Dá sa očakávať že väčšia časť zvuku je v strede ľavo-pravej bázy a preto bude "side" informácia celkovo nadobúdať menšie hodnoty koeficientov ako "middle" .

Ďalšími spoločnými črtami je použitie kvantizácie spektrálnych koeficientov , čo znamená že za pomoci "Psychoacoustic Model"-u je určená potrebná presnosť (kvantizér) kódovania spektrálnych koeficientov aby bol zakódovaný zvuk vnímaný rovnako , tieto koeficienty sú potom kódované ako zaokrúhlené celočíselné násobky kvantizéra .

Na kódovanie postupností celých čísel sa používa kódovanie z premenlivou dĺžkou slova využívajúc entropiu-u .

1.2.2 Čo je nové v AAC ?

AAC obsahuje všetky techniky ktoré boli použité v mp3 , obsahuje ale aj techniky nové . AAC v porovnaní mp3 , dokáže o niečo lepšie využiť aj štruktúru pre človeka bežného audio signálu . Ďalšie vylepšenia sú v spôsobe kódovania šumu a maskovanie nežiadúcich artefaktov spôsobených kvantizáciou spektrálnych koeficientov . AAC používa tieto nové techniky [Aac] :

"Long Term Prediction" (ďalej LTP) . Algoritmus mp3 kóduje navzájom nezávislé bloky dlhé približne 26ms (38 blokov za sekundu) . Preto ak sa v signále nachádza zvuk ktorý je čase kvázi statický (veľmi málo meniaci sa tvar signálu) mp3 nedokáže túto vlastnosť využiť mimo rámec 26ms . AAC tak isto používa kódovanie v blokoch , ale umožňuje využívať vzájomné podobnosti susedných blokov , a to tak že pomocou LTP algoritmus predpovedá tvar spektra práve kódovaného bloku , z už zakódovaného predošlého bloku . Následne stačí zakódovať len rozdiel medzi práve kódovaným blokom a zmienenej predpovede . Táto technika je veľmi účinná hlavne pri zvukoch ako sú sláčikové nástroje , píšťala , orchestrálne nástroje , dlho znejúce samohlásky vokálov .

"Perceptual Noise Substitution" (ďalej PNS) . Predstavme si že signál ktorý chceme kódovať je biely šum (šum ktorý má v celej oblasti spektra rovnomerne rozdelenú energiu) . Tento signál asi najchúlostivejší na kódovanie cestou mp3 , pretože je vybudené celé spektrum (biely šum možno jednoducho generovať ako postupnosť náhodných čísel s rovnomerným rozdelením) . AAC využíva fakt že dva vygenerované šumy (teda odlišné signály) znejú pre človeka rovnako . Nieje teda dôležitý samotný tvar šumového signálu ale len fakt že ide o šum , a aby bol vnímaný rovnako stačí dodržať jeho energetické rozdelenie v spektre (spektrálna obálka) . PNS je nástroj ktorého snahou je detekovať v signále zvuky s charakteristikou šumu a nekódovať ho cez koeficienty spektra , ale zakódovať len jeho spektrálnu obálku podľa ktorej sa pri dekódovaní bude šum v spektre generovať .

"Temporal Noise Shaping" (ďalej TNS) . Pri kvantizovaní spektrálnych koeficientov dochádza k tomu že vnesená chyba sa prejavuje v PCM reprezentácii v celom príslušnom bloku rovnomerne . Teda ak sa v práve kódovanom bloku nachádza prudká zmena dynamiky signálu , napr. ak v jednej polovici je ticho a v druhej začína ľubovoľný zvuk . Tak kvantizovanie pri dekódovaní signálu spôsobí že v čase ktorom bolo pôvodne ticho bude už

nežiadúci signál . Toto sa môže posluchovo prejavíť ako rozmazávanie detailov zvuku v čase tzv. “Post/Pre Echo“. Zabrániť je tomu možné tak že môžeme zvoliť jemnejšiu kvantizáciu (do určitej miery musíme) . TNS využíva “dualitu“ MDCT , ktorá je daná podobnosťou inverznej transformácie k MDCT (iMDCT) . Spočíva v tom že ak je napr. v PCM reprezentácii len v jednom bode signál nenulový tak v spektrálnej je vybudené celé spektrum ale hodnoty sú veľmi korelované , alebo naopak ak je len jeden spektrálny koeficient nenulový tak PCM reprezentácia bude mať všetky prvky nenulové (sínusoida) . Teda ak sú korelované hodnoty pri jednej reprezentácii , sú korelované aj pri druhej . TNS je technika predikovania spektrálnych koeficientov z predošlých , teda ak už je zakódovaných j koeficientov spektra tak koeficient j +1 je odhadnutý z hodnôt predošlých (predošlého) koeficientov a kódovaný je len rozdiel odhadu a skutočnej hodnoty . Keďže tieto kódované hodnoty sú už menšie , môžeme zvoliť jemnejšiu kvantizáciu a zmierniť tak Pre/Post Echo .

1.2.3 Spectral Band Replication (ďalej SBR)

SBR je technológia vyvinutá spoločnosťou “Coding Technologies“ [Sbr] ktorá výrazne zlepšuje účinnosť perceptualnej stratovej kompresie . SBR je tiež nazývaná “Bandwidth Extension Technique“ . Myšlienka spočíva v tom že sa nejakým štandardným spôsobom kóduje signál len s dolnou polovicou celého frekvenčného pásma (spektra) ktoré chceme kódovať , a horná polovica frekvenčného pásma sa “rekonštruuje“ z dolnej polovice pri dekódovaní signálu .

Treba dodať že to nefunguje celkom automaticky ako sa môže zdať z uvedenej vysvetlivky , ale využíva sa že tieto polovice spektra sú v bežnom signály korelované . Dolná polovica obsahuje väčšinu energie celého signálu a je teda podstatne dôležitejšia ako horná polovica , navyše väčšina melodiky (základných tónov) sa nachádza práve v tomto pásme . Väčšina audio signálov obsahuje v hornej polovici spektra už len harmonické kmitočty (Fourierov rad) tónov z dolnej polovice , a tieto sú počas znenia príslušného melodického zvuku často celú dobu svojej prítomnosti vo vzťahu so svojim základným tónom . Teda harmonické tóny podliehajú rovnakým dynamickým zmenám (zmeny hlasitosti) .

SBR teda na strane kódovania audio signálu nachádza dané vzťahy a pridáva ich kód pridáva k už zakódovanému dolnému pásmu . Faktom je ale že existujú signály ktoré do tohto modelu nezapadajú , preto ma SBR ešte ďalšie prídavné techniky ako je “Sinusoidal Regeneration“ (ďalej SR) a “Adaptive Noise Addition“ (ďalej ADA) .

SR slúži nato keď je v hornom pásme signálu tón ktorý nieje harmonickým k žiadnemu tónu z dolného pásma a teda nieje ho možné vyjadriť vzťahom ale je potrebné ho zakódovať . ADA slúži na podobný účel ako PNS v AAC . Jedná sa o pridávanie generovanie šumu v hornej polovici spektra na strane dekódera , ak tak aby doplnil energiu tohto spektra to energie akú mal pôvodný kódovaný signál teda je nutné informácie o energii taktiež kódovať .

Technológia SBR sa používa v spojení mp3 aj AAC kóderov . známych ako “mp3 Pro“ a “AAC Plus“ . SBR umožňuje zmenšenie veľkosti kódu na 70 % oproti kódovaniu bez použitia SBR pri rovnakej subjektívnej kvalite .

1.2.4 Čo tvrdia tvorcovia o mp3 vs. AAC ?

...

1.3 Čo je cieľom tejto práce

...

Referencie

[Mdct]

<http://ccrma-www.stanford.edu/~bosse/proj/node27.html>

http://www.ifn.et.tu-dresden.de/MNS/veroeffentlichungen/2003/Nikolajevic_V_Kluwer_03.pdf

[Mp3aac]

<http://mia.ece.uic.edu/~papers/WWW/MultimediaStandards/chapter4.pdf>

[Aac]

<http://faac.sourceforge.net/wiki/index.php?page=AAc>

[Sbr]

<http://www.esat.kuleuven.ac.be/~spch/mpca/papers/ekstrand:mpca02.pdf>

<http://www.codingtechnologies.com/products/sbr.htm>