State of the Art in Perceptual Coding: MPEG-2/4 Advanced Audio Coding (AAC)
WS 2015/16
History

- 1994: Official start of AAC development

- Goal: Development of a new powerful state-of-the-art multi-channel coder without compatibility constraints

- 1997: AAC International standard (IS)

- 1999: AAC part of the MPEG-4 standard

- Today: favorite coder for many application areas like Internet radio, solid state players, High Definition TV (HDTV), satellite and terrestrial digital audio broadcasting
Overview (1)

- Next generation mono/stereo/multichannel coding
- Same quality at half the bit-rate
- International cooperation of the Fraunhofer Institute and companies like AT&T, Sony and Dolby
- Most efficient MPEG method for audio data compression up until now
- Driving force to develop AAC was the quest for an efficient coding method for surround signals, like 5-channel signals (cinemas)
Overview (2)

- Makes use of the signal masking properties of the human ear in order to reduce the amount of data

- Quantization noise is distributed to frequency bands in such a way that it is masked by the total signal

- Iterative encoder structure using Huffman coding and non-uniform quantization
  - Features found in Layer 3 and PAC

- Window type and block switching
  - Features found in AC-3, Layer 3, PAC, + new

- Temporal Noise Shaping (TNS)
  - New technique
Overview (3)

- Prediction
- Bit reservoir
- M/S stereo coding
- Intensity stereo coding
- Gain control
AAC- Encoder Overview
MPEG-AAC: Basic Features

- High frequency resolution filter bank-based coder
  (1024 subband MDCT with 50% overlap)

- 1:8 block switching (1024/128 subband MDCT)

- Non-uniform quantizer

- Noise shaping in half critical bands (scalefactor bands)

- Huffman coding of scalefactors and spectral coefficients
MPEG-2 AAC: Advanced Coding Tools

- Window shape adaptation → usually fixed (sine or KBD – Kaiser-Bessel Derived)

- Temporal noise shaping (TNS) → often used

- Gain control (SRS/ Sample Rate Scalable profile, only), not often used

- Backward adaptive prediction → not often used
Frequency response of Sine and Fielder window
Differences MPEG-2 AAC and MPEG Audio Layer-3

Filter bank

- ISO/MPEG Audio Layer-3 uses hybrid filter bank chosen for reasons of compatibility

- MPEG-2 AAC uses a plain Modified Discrete Cosine Transform (MDCT) to reduce aliasing

- Together with the increased number of subbands (1024 instead of 576 samples) the MDCT outperforms the filter banks of previous coding methods
Differences MPEG-2 AAC and MPEG Audio Layer-3

Temporal Noise Shaping TNS
- Shapes the distribution of quantization noise in time by prediction in the frequency domain
- Voice signals in particular experience considerable improvement through TNS

Prediction (in band in time domain) not in use
- A technique commonly established in the area of speech coding systems
- It benefits from the fact that stationary audio signals are predictable to a certain extend
- But requires higher computational complexity
Differences MPEG-2 AAC and MPEG Audio Layer-3

Quantization

- By allowing finer control of quantization resolution, the given bit rate can be used more efficiently

Bit-stream format

- Huffman coding of side information
- More flexibility leads to more coding efficiency
Filter Bank Details

MDCT (Princen / Bradley)

- TDAC, MLT, cosine modulated filter bank
- critical sampling
- time domain aliasing cancellation

Block switching to adjust the impulse response Window type switching

- sine window
- Kaiser Bessel Derived (KBD) window
MPEG-4
General Audio Coding
A short view into MPEG-4 Audio (1)

Very diverse requirements: no single algorithm:
- Music synthesis (Structured Audio) = kind of an extension of midi
- Very low rate parametric coding (HILN, HVXC)
- Speech coding (CELP)
- Perceptual Coding ("General Audio") over a wide range of bitrates

High quality coding done via AAC with additional coding tools:
- TwinVQ, scalability tools
- Perceptual Noise Substitution (PNS)

Backwards compatibility, no new coding paradigm for high quality audio
A short view into MPEG-4 Audio (2)

MPEG-4 General Audio Coding: The “all-round coder” in MPEG-4 audio

MPEG-4 Extensions:
- Perceptual Noise Substitution (PNS)
- Long Term Prediction (not in use)
- TwinVQ Coding Core (not in use)
Temporal Noise Shaping: TNS

Encoder Overview:

- Gain Control
- Filter Bank
- TNS (Temporal Noise Shaping)
- Intensity Coupling
- Prediction
- M/S
- Scale Factors
- Quant.
- Noiseless Coding
- Rate/Distortion Control

Input signal

Bitstream Multiplexer

Output
Temporal Noise Shaping (1)

- e.g. castanet
  - original
  - quantization noise
  - TNS

- e.g. speech
  - original
  - quantization noise
  - TNS

1 frame
Why use TNS instead of block switching?

Low number of subbands leads to higher bit rate.

Ok if it only happens occasionally. Therefore buffer can be used. Problem if there are many peaks, as in speech the glottal pulses (every few ms!). Bit rate would become too high, or the quantization noise too high.

→ Alternative approach is needed → TNS

But: TNS is not really a replacement for block switching
Temporal Noise Shaping (2)

Solution for avoiding quantization noise spread:

- Make smaller frames (works for attacks but not for speech → decrease of coding efficiency)
- Higher time resolution to shape quantization noise
- TNS

Limitation of TNS:
- Time domain aliasing
Speech Coding as Model for TNS

In frequency domain the prediction error is flat.

TNS predicts in frequency domain instead of time domain, shapes noise in time domain.

- In frequency domain the prediction error is flat.
- TNS predicts in frequency domain instead of time domain, shapes noise in time domain.
Switch roles of time and frequency domain:

→ predict not over time, but over frequencies, over the subbands

→ quantization error is shaped (after decoding) in the time domain (instead the frequency domain) like the signal

→ hopefully reduces pre-echo artifacts

→ But: aliasing in time domain limits effectiveness (peaks are mirrored over time)
Structure of TNS (encoder)

already in AAC

subbands

TNS

sequence of subbands

Audio

MDCT

1024 bands

\[ z^{-1} \]

pred

coeff

side-info

\[ Q \]

predicts from one subband to the next, starting at the lowest subband, from subband 0 it predicts 1,…
TNS Decoder

pred \( z^{-1} \)

subbands

Synth MDCT 1024 bands

Audio
Extension: Perceptual Noise Substitution (PNS)
Perceptual Noise Substitution (1)

Background:

- Parametric coding of signals gives a very compact signal representation
- Parametric coding of noise-like signal components has been used widely e.g. in speech coding
- Can similar techniques be used in perceptual audio coding?

MPEG-4:

- Perceptual Noise Substitution (PNS) permits a frequency selective parametric coding of noise-like signal components
"Perceptual Noise Substitution" (PNS): Perceptual coder + parametric represent. of noise-like signals
Perceptual Noise Substitution (3)

Principle:
- Noise-like signal components are detected on a scalefactor band basis
- Corresponding groups of spectral coefficients are excluded from quantization/coding
- Instead, only a "noise substitution flag" plus total power of the substituted band is transmitted in the bitstream
- Decoder inserts pseudo random vectors with desired target power as spectral coefficients
  → Highly compact representation for noise-like spectral components
MPEG-4
Low Delay Audio Coding
MPEG-4 Version 2 Low Delay Audio Coding

Target:
- High audio and speech quality and
- Low bitrate and Low algorithmic delay (20 ms)

Solution:
- MPEG-4 Version 2 Low Delay Audio Coder:
  - Derived from MPEG-2/4 "Advanced Audio Coding" (AAC)
  - Specific modifications for low-delay operation
Delay Sources in Perceptual Audio Coding

- Framing delay
- Filter bank delay
- Look-ahead delay for block switching
- Use of bit reservoir

→ Overall delay:

\[ t_{\text{delay}} = \frac{N_{\text{framing}} + N_{\text{filterbank}} + N_{\text{look-ahead}}}{F_s} + t_{\text{bitres}} \]
Example: Delay of AAC Codec (48 kHz / 64 kbps)

- Framing delay: 1024 samples
- Filter bank delay: 1024 samples
- Look-ahead delay for block switching: 576 samples
- Use of bit reservoir: 74.7 ms

→ Overall delay:

\[ t_{\text{delay}} = \frac{1024 + 1024 + 576}{48000} + 74.7 \text{ ms} = 129.4 \text{ ms} \]
Low Delay AAC Codec (48 kHz, min. delay mode)

- Reduced filter bank delay: 959 samples
- No block switching → no look-ahead delay: 0 samples
- Minimal bit reservoir: 0...32 bits

→ Overall delay:

\[ t_{\text{delay}} = \frac{480 + 480 + 0}{48000} + 0 \text{ ms} = 20 \text{ ms} \]
Preecho Behavior

![Graphs showing original and coded/decoded signals](image-url)
Preecho Reduction by Window Shape Adaptation
next lecture:

07.12. – Midterm Exam (60 minutes)