SAOC and USAC

Spatial Audio Object Coding / Unified Speech and Audio Coding

Lecture "Audio Coding"
WS 2014/15

Prof. Dr.-Ing. Gerald Schuller

SAOC – Spatial Audio Object Coding

Outline

- Introduction
- From Spatial Audio Coding to SAOC
- Audio objects
- SAOC Decoding
- Applications
- Performance Evaluation
- Conclusion

Gerald.schuller@tu-ilmenau.de

SAOC - Introduction

- Perceptual audio coding for multichannel signals is widely used
  - “Spatial Audio Coding”, for instance MPEG Surround

- Existing Spatial Audio Coders are channel-based
  - Designed for a specific reproduction setup

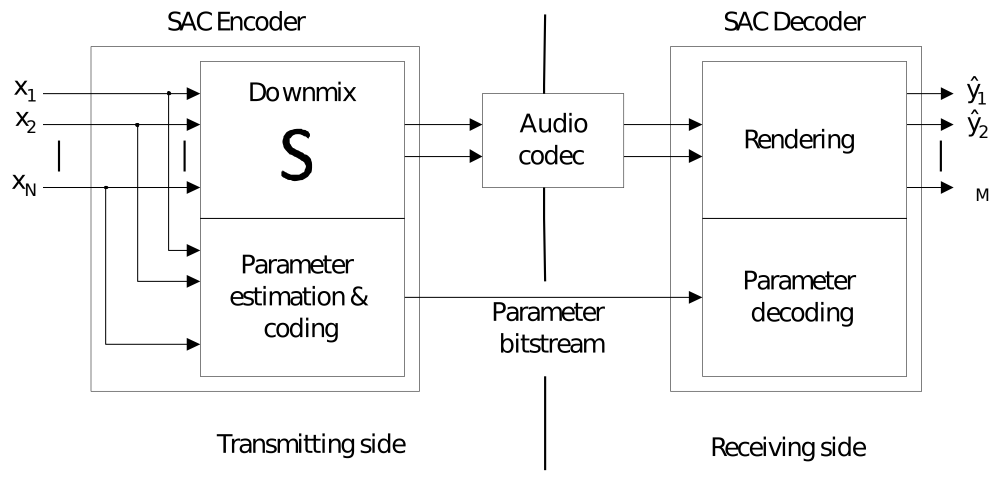
- Spatial Audio Object Coding
  - Continuation of the “Spatial Audio Coding” paradigm
  - Transmit audio objects instead of channel signals
  - ISO/IEC 23003-2:2010 Standard

Gerald.schuller@tu-ilmenau.de

SAOC – From Spatial Audio Coding to SAOC

Spatial Audio Coding (e.g., MPEG Surround)

- Channel-oriented
- Downmix (mono or stereo)
- Transmit downmix using standard audio codec (AAC)
- Additional parameter data (parametric coding)
- Output channels for specific reproduction setup
  - 5.1, 7.1

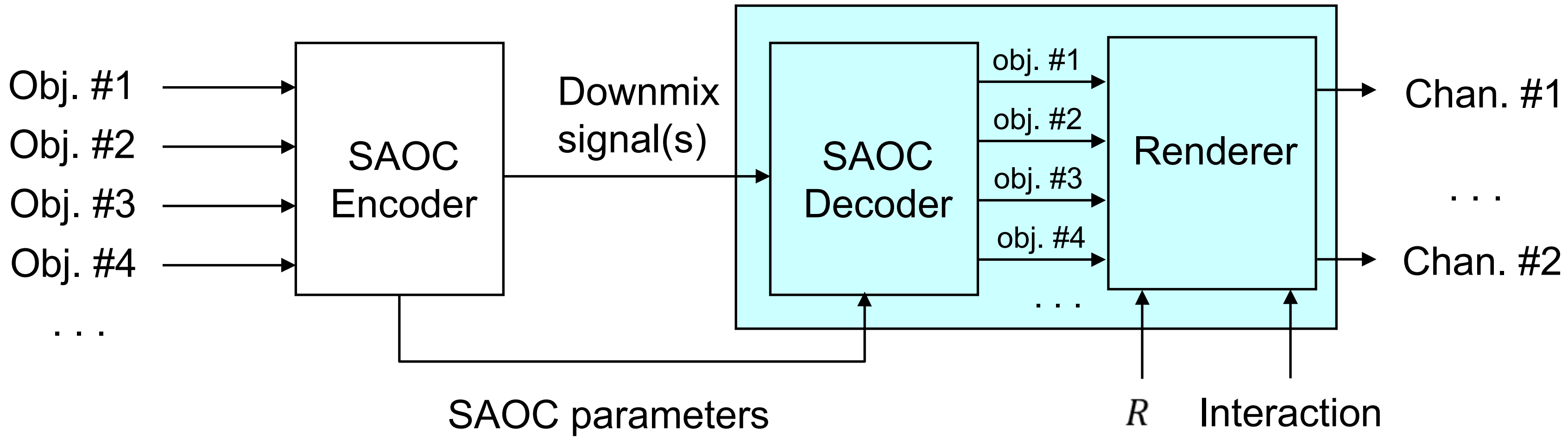


[1] Herre et.al 2012: “MPEG Spatial Audio Object Coding, J. Audio Eng. Soc. 60:9, 2012

Gerald.schuller@tu-ilmenau.de

SAOC – Audio Objects

- Audio objects instead of channels
- SAOC encoder: Stereo or mono downmix plus SAOC parameters
- SAOC decoder: Use SAOC parameters to transform downmix into audio objects
- Rendering to loudspeaker configuration (Rendering matrix)



Gerald.schuller@tu-ilmenau.de

SAOC – Audio Objects

Advantages of object-based processing

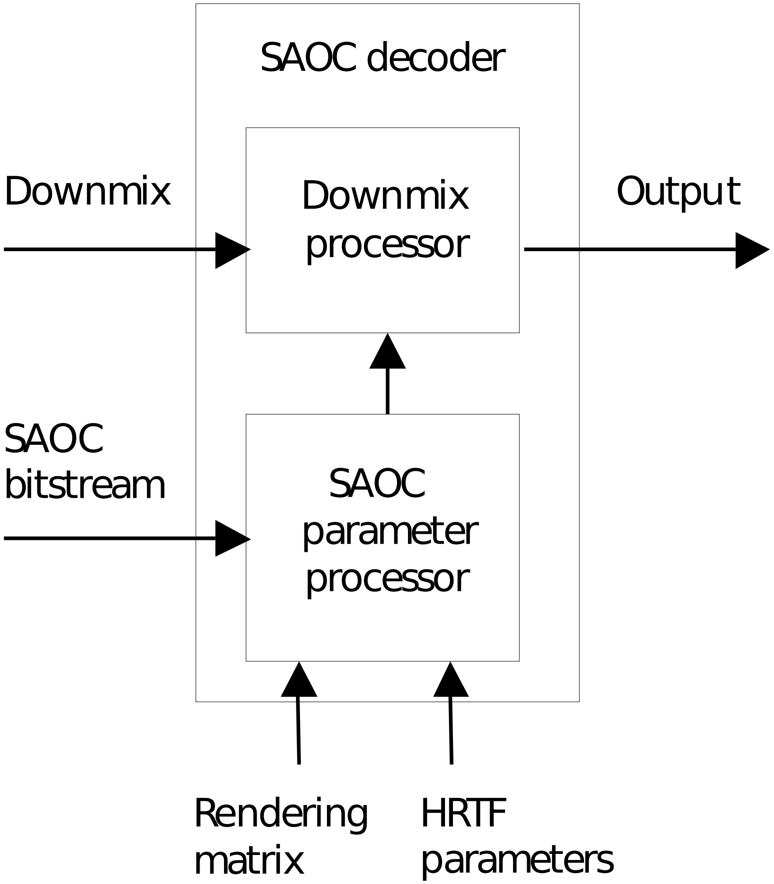
- Coding efficiency: SOAC parameters only a few kbit/s per audio object
- Coding and transmission independent of reproduction setup
- Rendering on arbitrary loudspeaker setups
    - 5.1, 7.1, 10.2, 22.2, Binaural reproduction, Wave field synthesis, …
    - Rendering controllable (real-time user interaction)
  - Control over individual audio objects
    - Change gain, equalization, effects, …

Gerald.schuller@tu-ilmenau.de

SAOC Decoding Modes

Decoder Processing Mode

- Rendering integrated into decoding (efficient)
- For mono and stereo output, incl. binaural reproduction
- Rendering matrix: realtime control of rendering
- HRTF parameters for binaural
  - Open SAOC interface
  - Enables use of individual HRTFs
  - Efficient parametric representation

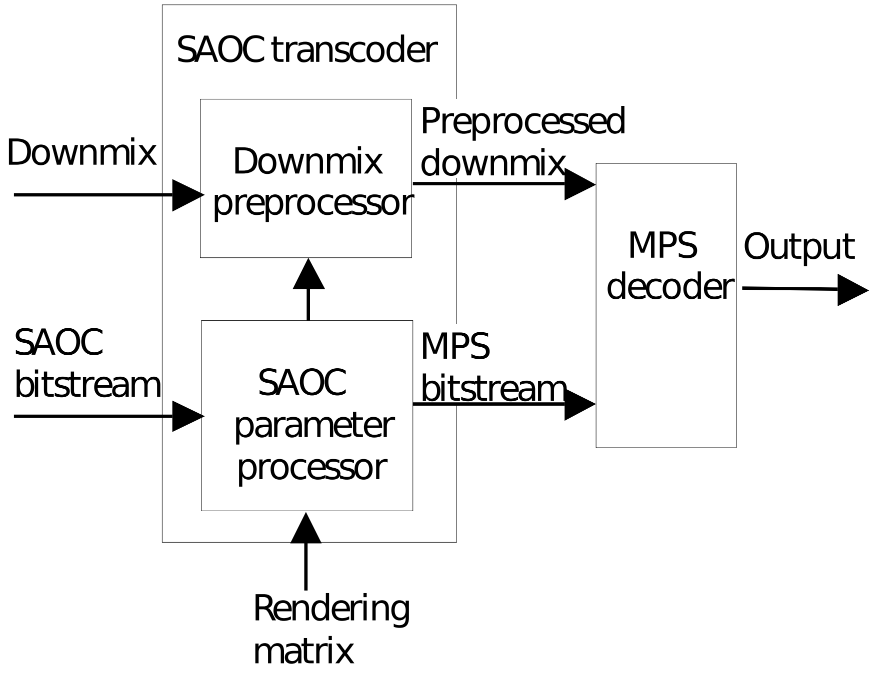


Gerald.schuller@tu-ilmenau.de

SAOC Decoding Modes

Transcoder Processing Mode

- For multichannel output (MPEG Surround - MPS)
- SAOC encoder works as transcoder
  - Transcoding of SAOC parameters to MPS bitstream
  - Adjustment of downmix panning (only for stereo downmix)
  - Highly Efficient
    - Operates in transform domain
    - Avoid unnecessary (de)quantization and decoding



Gerald.schuller@tu-ilmenau.de

SAOC Bitstream

- Contains parametric description of audio objects: SAOC parameters
- Typical: 2-3 kbit/s per audio object (plus 3 kbit/s per audio scene)
- SAOC bitstream embedded in ancillary data of core audio coder
  - Enables backward compatibility
- Parameters transmitted in flexible time/frequency grid
  - Adaptation to bitrate demands and/or signal characteristics
  - Same time/frequency grid as in MPEG Surround
    - Lossless, efficient transcoding

Gerald.schuller@tu-ilmenau.de

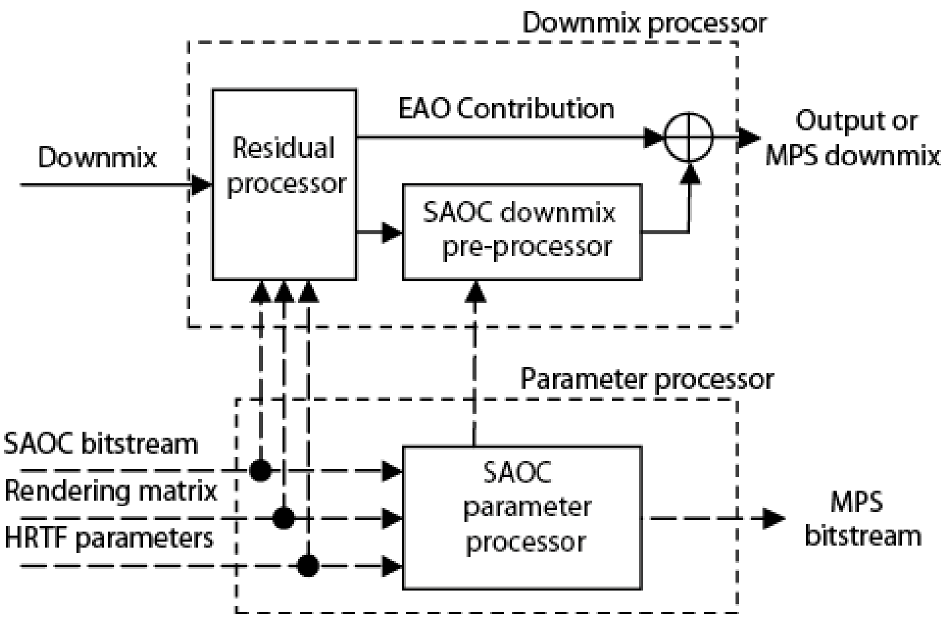
SAOC Parameters

- **Object Level Differences (OLD)**: Level relative to loudest object
- **Inter-Object Cross Correlations (IOC):** Similarity between pairs of objects
- **Downmix Gain (DMG):** Gains used in the downmix of individual objects
- **Object Energies (NRG):** Absolute energy of loudest object. Optional, enables merging of multiple SAOC streams

Gerald.schuller@tu-ilmenau.de

SAOC – Enhanced Audio Objects

- Allow arbitrary attenuation or amplification of objects
  - Karaoke
  - Solo voices,…
- SAOC bitstream contains residual signal
- Reconstruction from downmix and residual
- Efficient transmission of residual signal (AAC)



Gerald.schuller@tu-ilmenau.de

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

SAOC - Applications

Interactive Remix / Karaoke

- Interactive remixes
- Equalization, room simulation,… for individual objects
  - For channel-based formats, only applicable to whole scene
- Modification of specific audio objects (instruments, voices,…)
- Karaoke, vocal solo
  - Suppress main voice or background music
  - Advantageous: Enhanced Audio Objects
- Future extensions of digital broadcasting
  - Clean-audio dialogs
  - Additional objects for interactivity

Gerald.schuller@tu-ilmenau.de

SAOC - Applications

Teleconferencing

- Today: Mainly monophonic reproduction
  - Suboptimal for multi-user scenarios

- Key benefits of SAOC
  - Adjustment of individual speaker signals
  - Spatial representation of audio scene
  - Improved intelligibility and listening comfort
  - Match between visual and audio scene
  - Transmission efficiency
  - Backward compatibility

SAOC - Applications

Rich Media / Gaming

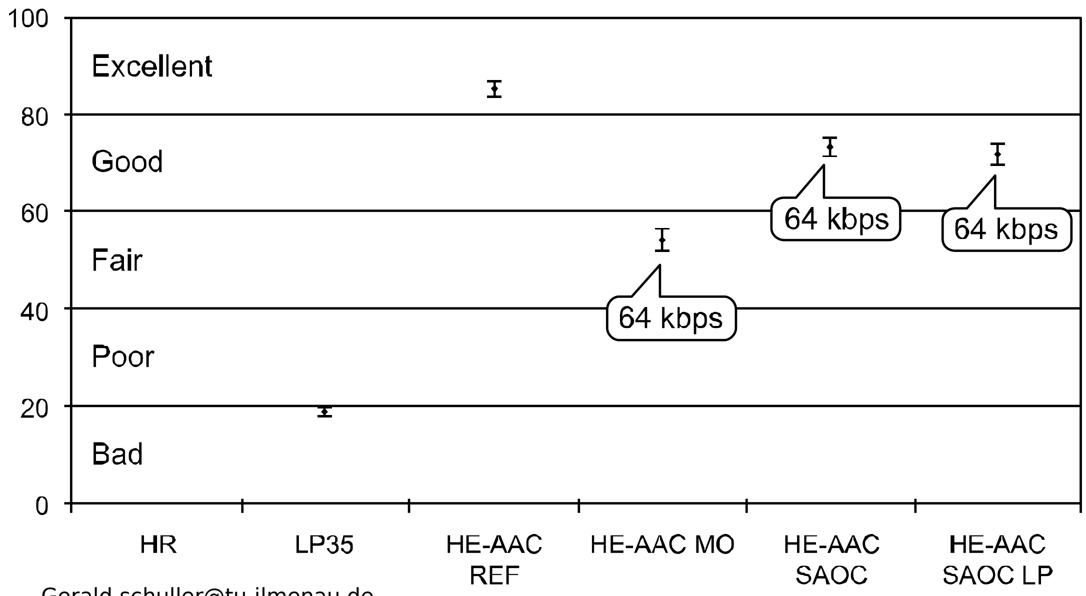
- Applications of Rich Media
    - Interactive audio-visual interfaces
    - Games
  - Platforms
    - Mobile
    - Flash- or Java-Based
    - Limited audio rendering capabilities (audio scene size)
- Key advantages
    - Low complexity (number of output channels instead of scene size)
    - Interactivity (adjust level of objects and background music)
    - Efficient transmission, backward compatibility

Gerald.schuller@tu-ilmenau.de

SAOC – Performance Evaluation

Listening Test – Remix scenario

- Part of MPEG verification tests (5 sites, 125 participants)
- MUSHRA test (ITU BS.1534-1)
- Simulate adjustments to a mix of audio objects
- Core coder: High Efficiency AAC (HE-AAC)



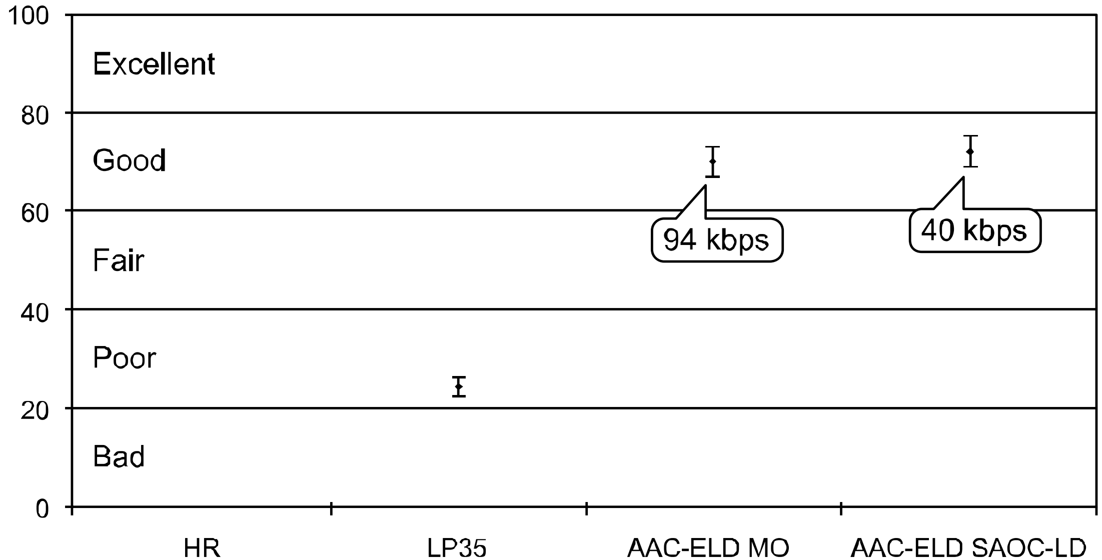
HR: Hidden reference
LP: 3.5 kHz Lowpass
HE-AAC REF: Individual objects, high bitrate
HE-AAC MO: Individual objects, same bitrate
HE-AAC SAOC: standard SAOC
HE-AAC SAOC LP: low power

Gerald.schuller@tu-ilmenau.de

SAOC – Performance Evaluation

Listening Test – Teleconferencing

- Part of MPEG verification tests
- Teleconferencing application: Simulate adjustments of a participant
- Core coder: MPEG-4 Enhanced Low Delay AAC (AAC-ELD)



HR: Hidden reference
LP: 3.5 kHz Lowpass
AAC-ELD MO: Individual objects
AAC-ELD SAOC-LD: low delay

Gerald.schuller@tu-ilmenau.de

SAOC - Summary

- Highly efficient transport/storage of audio objects **and** flexible/interactive audio scene rendering
- Backwards compatible downmix for reproduction on legacy devices
- Flexible rendering configurations (loudspeaker setups)
- ISO/MPEG standard
- Very interesting applications, e.g.:
  - Remixing / Karaoke
  - Gaming / Rich media
  - Teleconferencing
  - Interactivity for broadcast applications

USAC – Unified Speech and Audio Coding

Outline

- Introduction
- Differences between speech and audio coding
- Codec structure
- Improvements to coding
- Performance Evaluation
- Applications
- Summary

Gerald.schuller@tu-ilmenau.de

USAC - Introduction

Status quo:

- General audio coding and speech coding are largely separate worlds

Problem:

- Increased demand for audio coders that handle all types of inputs
    - Broadcasting
    - Audio books, multimedia
    - Mobile devices for all types of content (often low bandwidths)
  - Objective (initiated by MPEG)
  - Universal codec that handles all types of content at least as well
    - as the best current speech or audio codec

Gerald.schuller@tu-ilmenau.de

USAC – Differences Between Speech and Audio Coding

Audio Coding

- "Information sink model"
    - Characteristics of human hearing
  - Typically transform- or subband-based approaches
    - Divide signal in multiple bands and apply psychoacoustics

  - Not well-suited for speech

(at bit rates typically used by speech coders …)

Speech Coding

- "Information source model"
    - Characteristics of vocal tract
  - Typically based based on prediction coding
    - Predicted filter for the vocal tract and an excitation signal

- Poor quality for music

Gerald.schuller@tu-ilmenau.de

USAC – Hybrid coding approach

- Combine state-of the art speech and audio coder
  - HE-AACv2
  - AMR-WB+
- Switch between coders based on content
  - Signal classification
- Share common functionality
- Take care of artifacts due to switches …

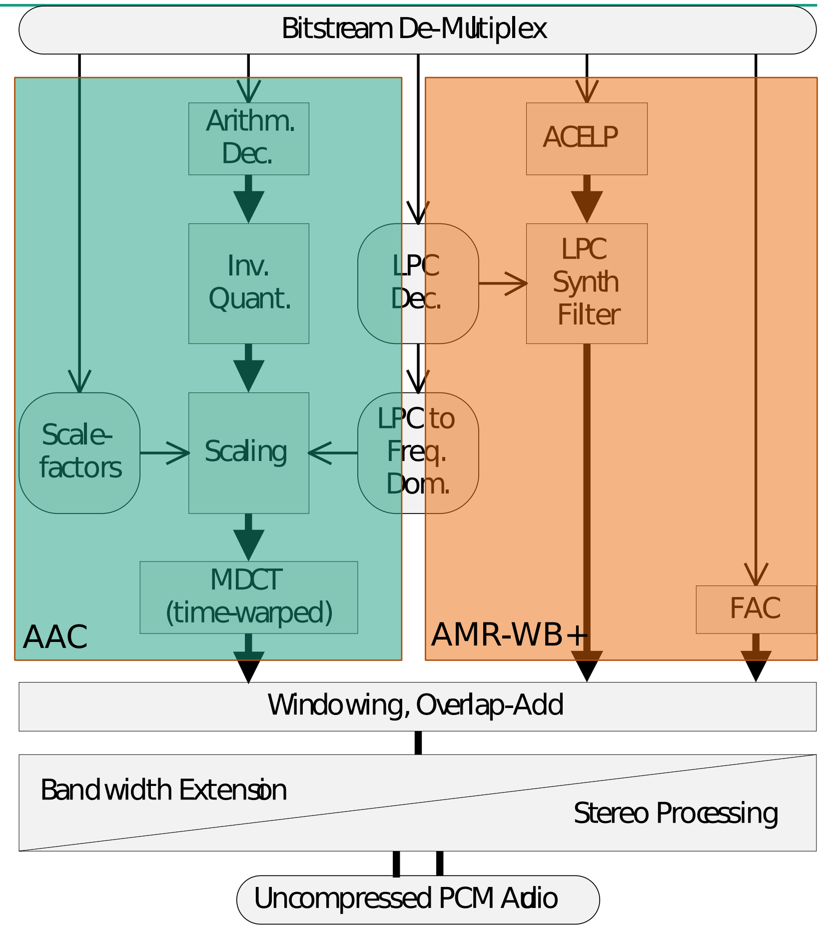


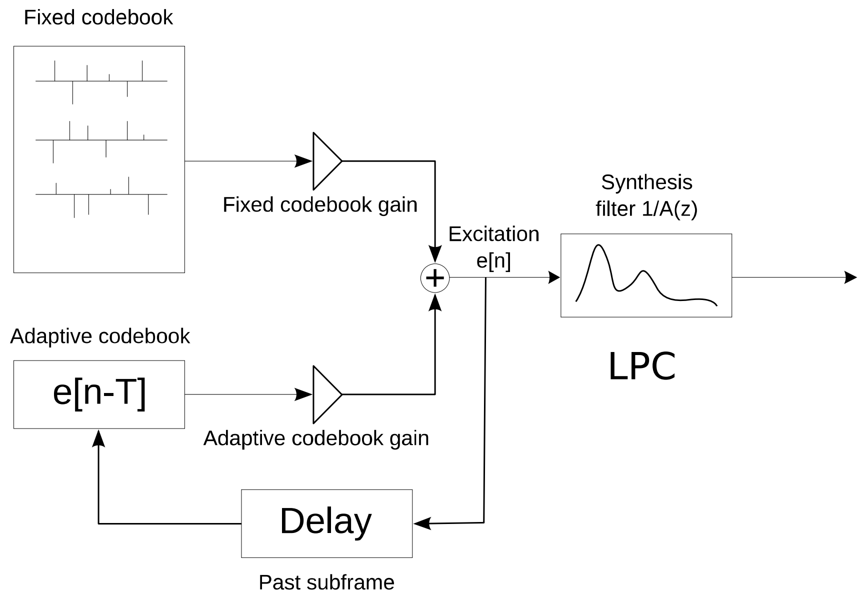
Figure: Neuendorf et.al: MPEG Unified Speech and Audio Coding, J. Audio Eng. Soc., 61:12, Dec. 2013

Gerald.schuller@tu-ilmenau.de

USAC – AMR-WB+ (1)

Adaptive Multi-Rate Wideband

- State-of-the-art speech coder
- Based on ACELP (Algebraic code-excited linear prediction)
- CELP: Encode signal by
    - LPC coefficients
    - LTP coefficients: “long term prediction” (delay and gain)
    - “Innovation codebook”: excitation signal, sparse pulses
  - ACELP: Algebraic representation of innovation codebook



CELP encoder

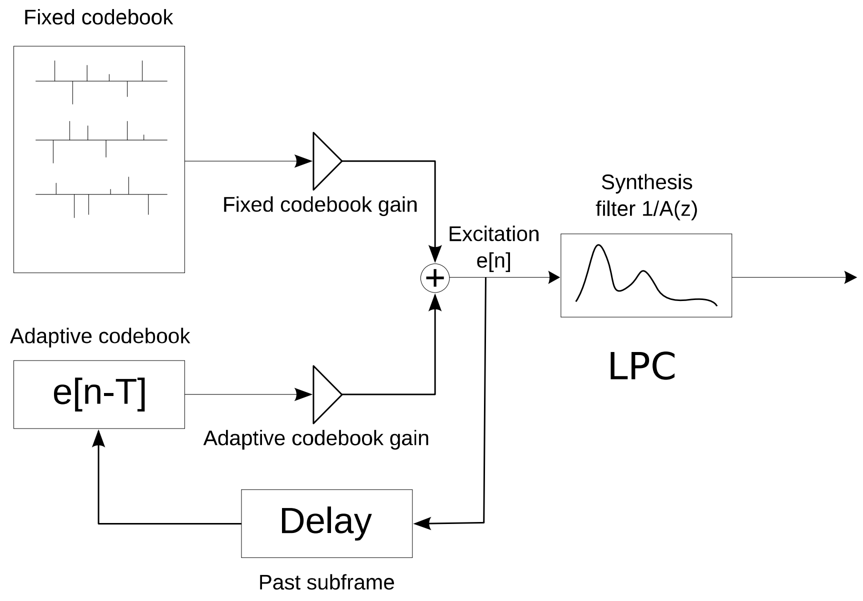
Figure: A. Valin: Speex: A Free Coder for Free Speech, 2006

Gerald.schuller@tu-ilmenau.de

USAC – AMR-WB+ (2)

Extended Adaptive Multi-Rate Wideband

- The "+" in AMR-WB+
  - Additional transform-domain coder for music signals
  - Parametric high frequency extension
  - Parametric stereo extensions
- But: For music, still inferior to good audio coders



CELP encoder

Figure: A. Valin: Speex: A Free Coder for Free Speech, 2006

Gerald.schuller@tu-ilmenau.de

USAC – Coder/Encoder Structure

General structure of modern audio codecs

- Encoder
  - Spatial coding
  - Parametric bandwidth extension
  - Core coding
- Decoder
  - In opposite order

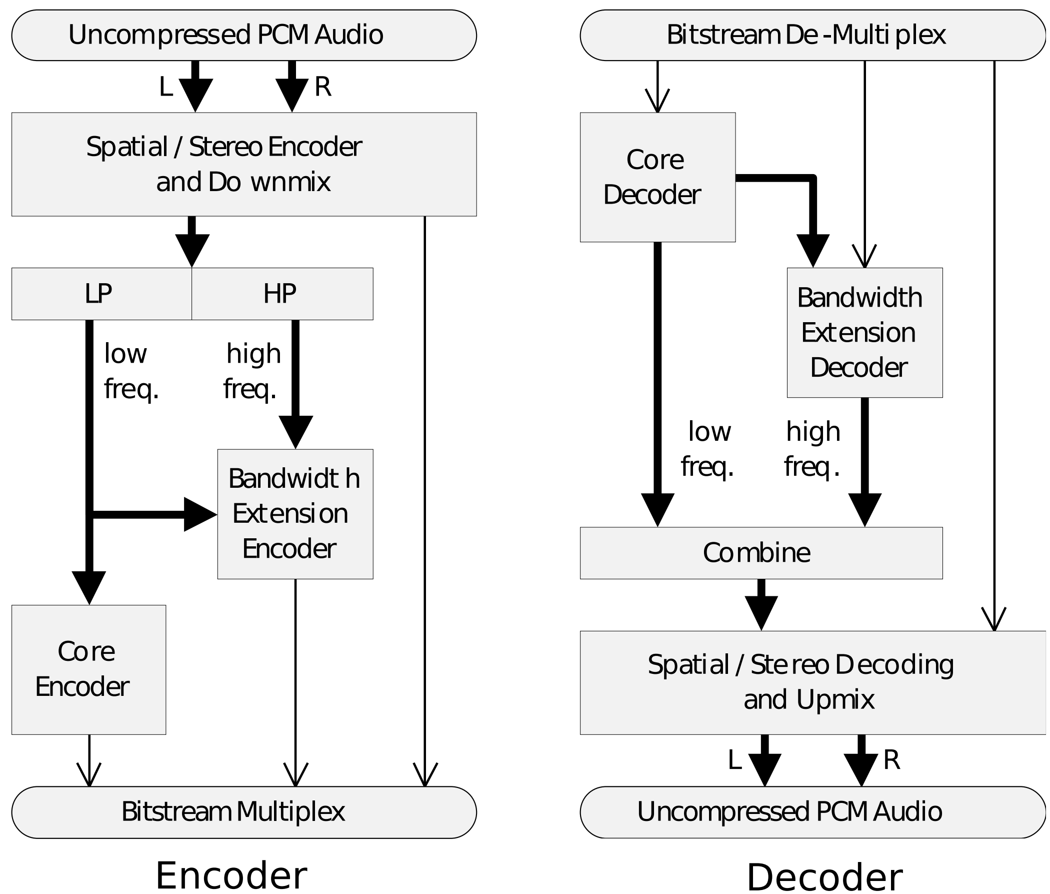


Figure: Neuendorf et.al: MPEG Unified Speech and Audio Coding, J. Audio Eng. Soc., 61:12, Dec. 2013

USAC – Decoder Structure

- Here: Focus on decoder
  - Only decoder is standardized
- Follows general codec structure
- Left part: audio coder
- (HE-AACv2)
- Right part: speech coder
- (AMR-WB+)
- Some tools shared (LPC decoding)
- Challenge: Switching between modes

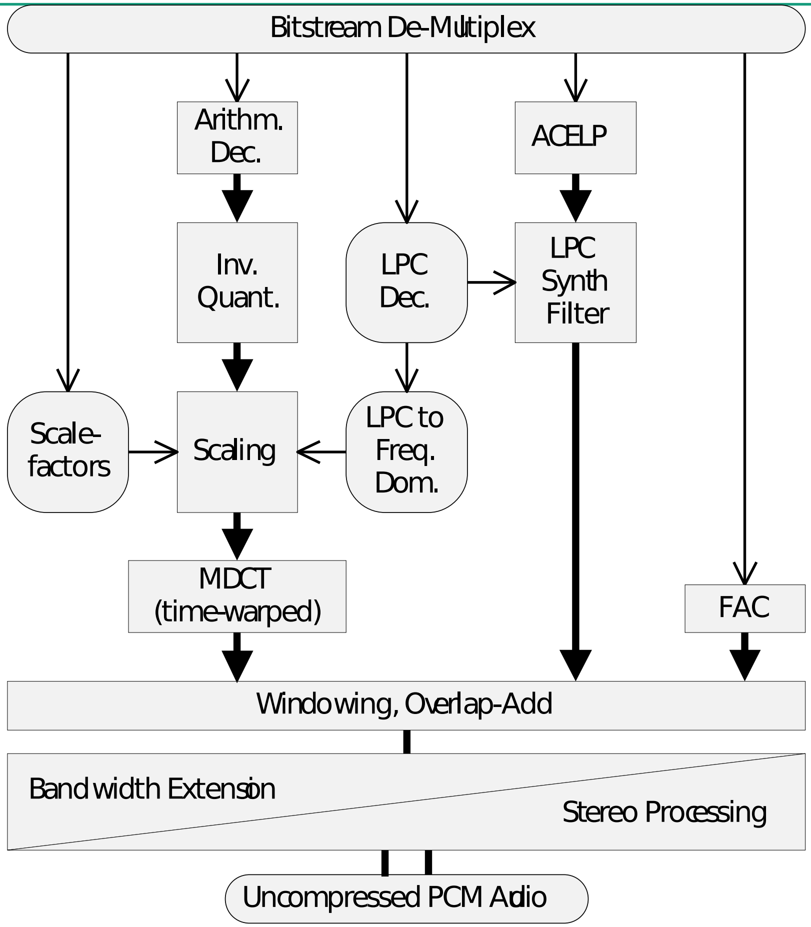
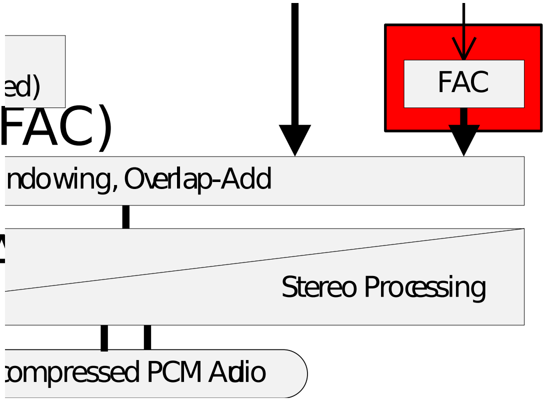


Figure: Neuendorf et.al: MPEG Unified Speech and Audio Coding, J. Audio Eng. Soc., 61:12, Dec. 2013

Gerald.schuller@tu-ilmenau.de

USAC – Transition Handling

- Encoder switches between two modes
    - Signal classifier (speech or music)
- Transition handling without audible errors or loss of coding efficiency
- HE-AAC: MDCT:
    - Transform (frequency) domain, overlapping windows, time-domain alias cancellation (TDAC)
- AMR-WB+
    - Time-domain, no overlap
- Solution: Forward Aliasing Cancellation (FAC)
- In case of transitions, transmit the
    - “alias cancellation” information for TDA



Gerald.schuller@tu-ilmenau.de

USAC – Improvements to Coding Tools

- USAC is not just a combination of HE-AAC and AMR-WB+
- Multiple improvements to both parts
  - Context-adaptive arithmetic coder for transform coding
  - Additional quantization modes
  - Alternate LPC-based noise shaping
  - Additional MDCT window sizes
  - Time-Warped MDCT
  - Enhanced Spectral Bandwidth Replication
  - Unified Stereo Coding
  - …

Gerald.schuller@tu-ilmenau.de

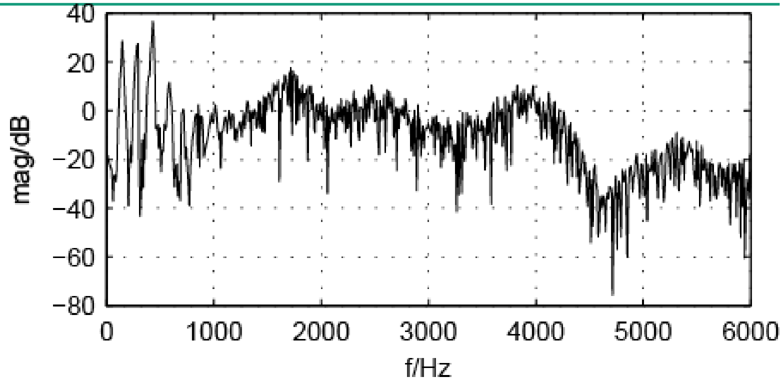
USAC – Time-Warped MDCT (1)

- Transform coding good for stationary tonal signals
    - Sparse spectrum, few nonzero spectral coefficients to code
    - High coding gain
- Problematic: Pitch changes within signal
    - Typical signal: Voiced speech
    - Smearing of energy over many spectral coefficients
    - Decreased coding efficiency

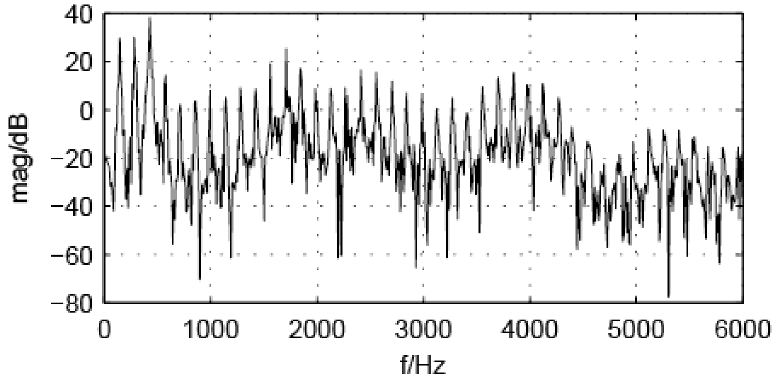
Gerald.schuller@tu-ilmenau.de

USAC – Time-Warped MDCT (2)

- Solution: Time-Warped MDCT
  - Reduce variations of fundamental frequency
- Basic algorithm (encoder side)
  - Apply a time-variant resampling prior to the MDCT
  - Adjust MDCT windows to preserve TDAC
  - Transmit resampling ratio as side information



Original spectrum



Time-warped spectrum

Figure: Edler et.al: A time-warped MDCT approach to speech transform coding, AES 126th Convention, May 2009

USAC – Enhanced Spectral Band Replication (1)

State of the Art – Spectral Band Replication

- Basis: SBR of HE-AAC
  - Operates in QMF domain
  - Copy low-frequency spectrum to higher frequencies
  - Adjust HF copies based on
  - parameters (side info)
    - Tonality
    - Envelope
    - Additional noise, sinusoids

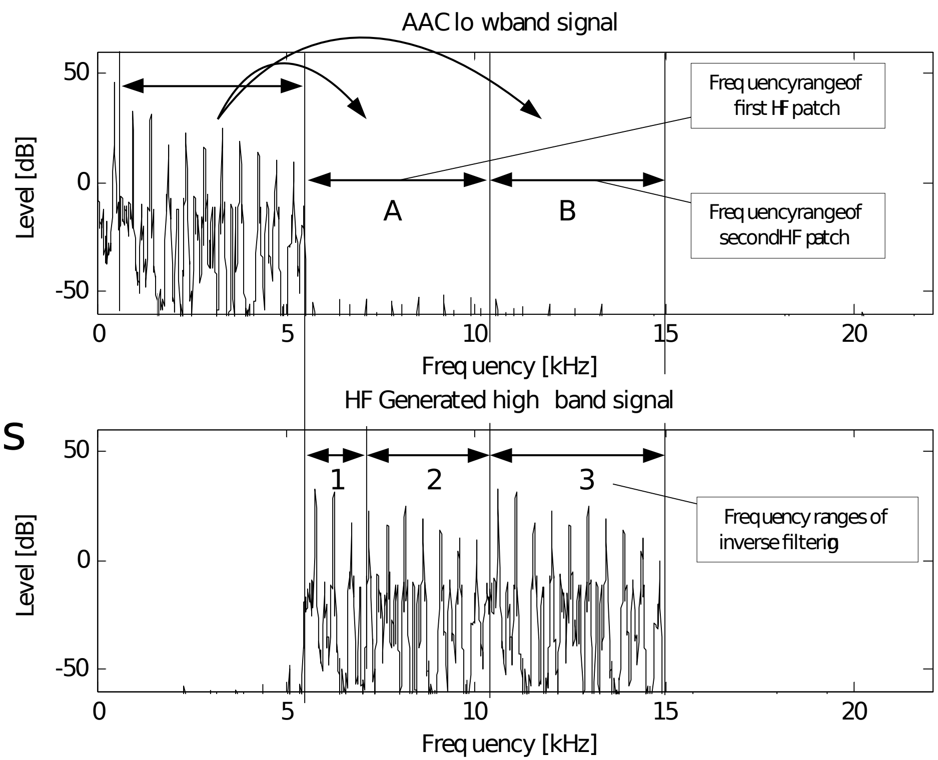
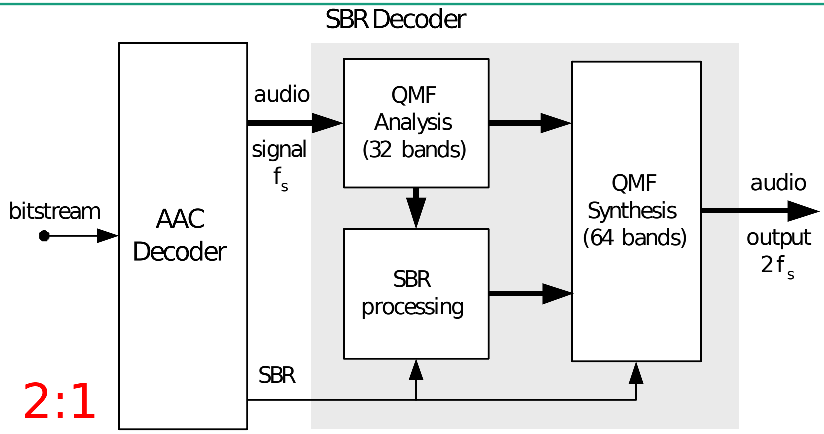


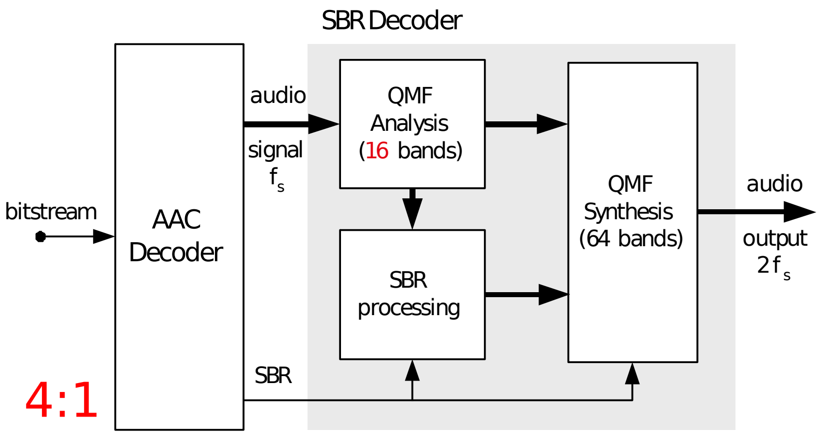
Figure: Neuendorf et.al: MPEG Unified Speech and Audio Coding, J. Audio Eng. Soc., 61:12, Dec. 2013

USAC – Enhanced Spectral Band Replication (2)

Alternative Sampling Rate Ratios

- HE-AAC SBR performs a 2:1 upsampling in QMF domain
  - Bandwidth doubled
- USAC: Additional ratios
- 4:1 (16 QMF analysis bands)
  - Four times the core bandwidth
  - Good for very low bit rates
- 8:3 (24 QMF analysis bands)
  - Halfway between 2:1 and 4:1
  - Best tradeoff for medium bit rates (~ 24 kbit/s)





Gerald.schuller@tu-ilmenau.de

USAC – Enhanced Spectral Band Replication (3)

Harmonic Transposition

- HE-AAC SBR: Spectral copies
    - Frequency shifts
    - Bad match for harmonics of tonal signals (integer multiples)
- USAC eSBR: Harmonic transposer
    - Map sinusoid with frequency to sinusoid with frequency , integer
    - Supported orders
    - Frequency shifts for higher orders

- Other improvements in eSBR (not covered here)
    - Predictive vector coding for SBR spectral envelopes
    - …

Gerald.schuller@tu-ilmenau.de

Stereo Coding in USAC – Unified Stereo Coding (1)

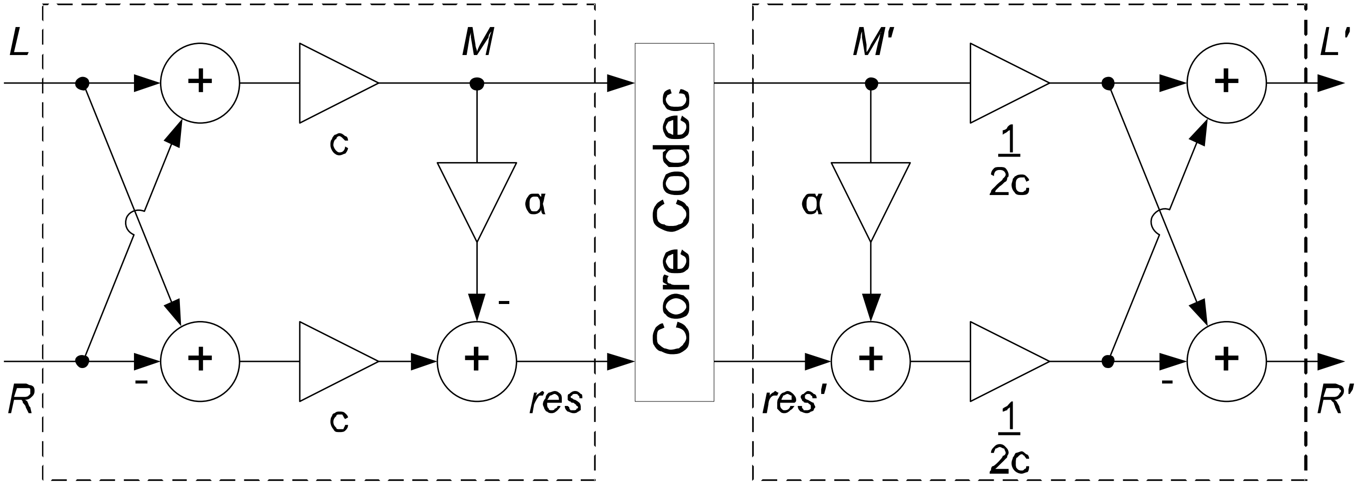
Discrete Stereo Coding
- Strives to preserve waveforms
- Joint coding techniques (e.g., M/S)
- Used with higher bit rates

Parametric Stereo Coding
- Mono downmix and side info (parameters)
- Typically used with low bit rates

Unified Stereo Coding
- Extends and combines discrete and parametric stereo coding
- Additional parameter: IPD (inter-channel phase difference)
- Transmit parameters and residual signal
- Use parameters to minimize residual

Gerald.schuller@tu-ilmenau.de

Stereo Coding in USAC – Unified Stereo Coding (2)



- Prediction factor α (complex-valued)
- Gain normalization c
- c and α determined from parametric stereo parameters

Figure: Neuendorf et.al: MPEG Unified Speech and Audio Coding, J. Audio Eng. Soc., 61:12, Dec. 2013

Gerald.schuller@tu-ilmenau.de

USAC – Performance Evaluation

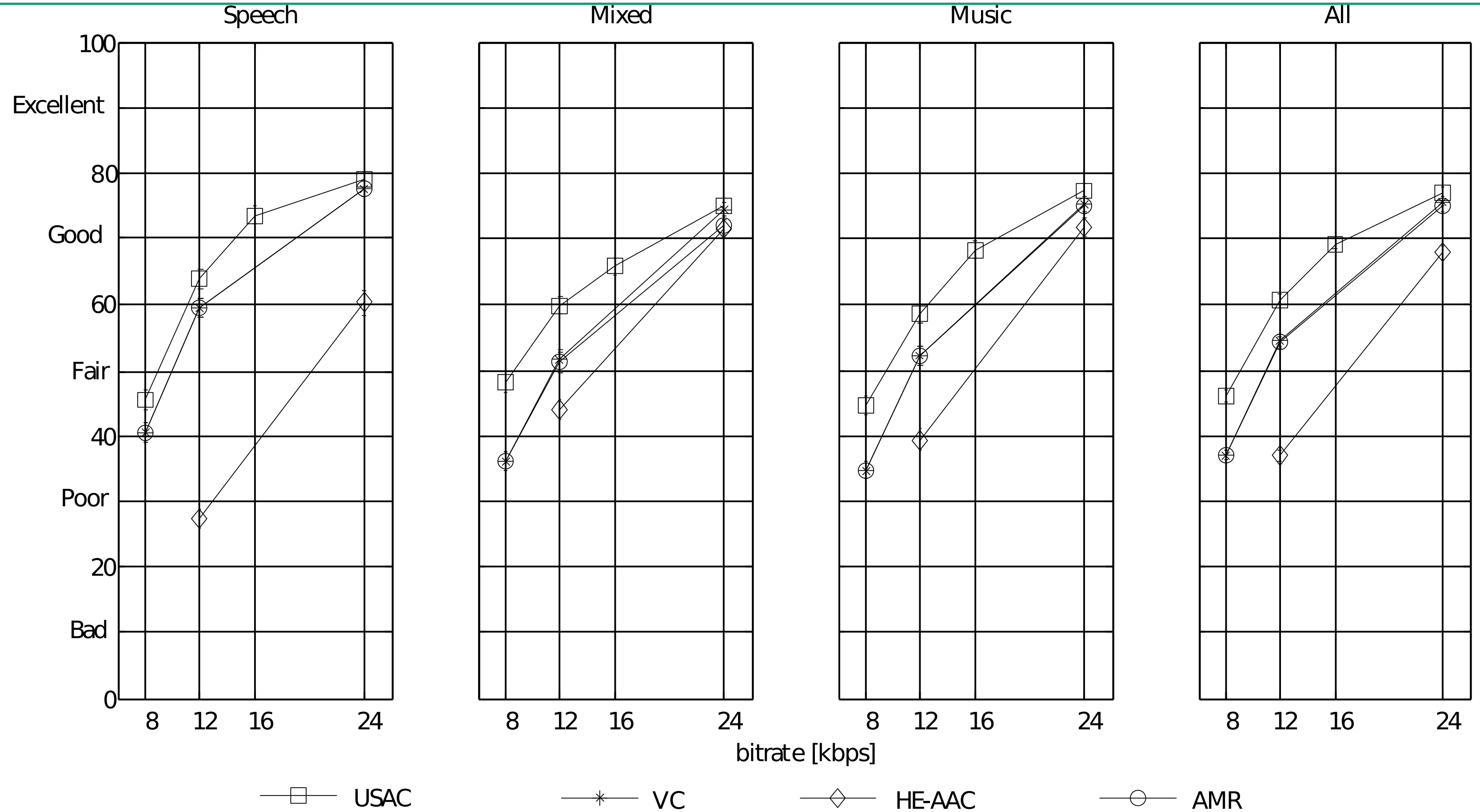
Verification Test Results

- Question: Whether USAC performs as least as good as the better of the best speech or audio coder
- Part of verification tests for approval by ISO/IEC
- 3 tests, 13 test sites, 60-25 participants
- MUSHRA methodology
- Test subjects
  - USAC
  - HE-AACv2
  - AMR-WB+
  - Virtual coder (VC): The better of HE-AACv2 and AMR-WB+
    - Determined separately for each test item and bit rate

Gerald.schuller@tu-ilmenau.de

USAC – Performance Evaluation

Listening Test – Mono, Low Bit Rates



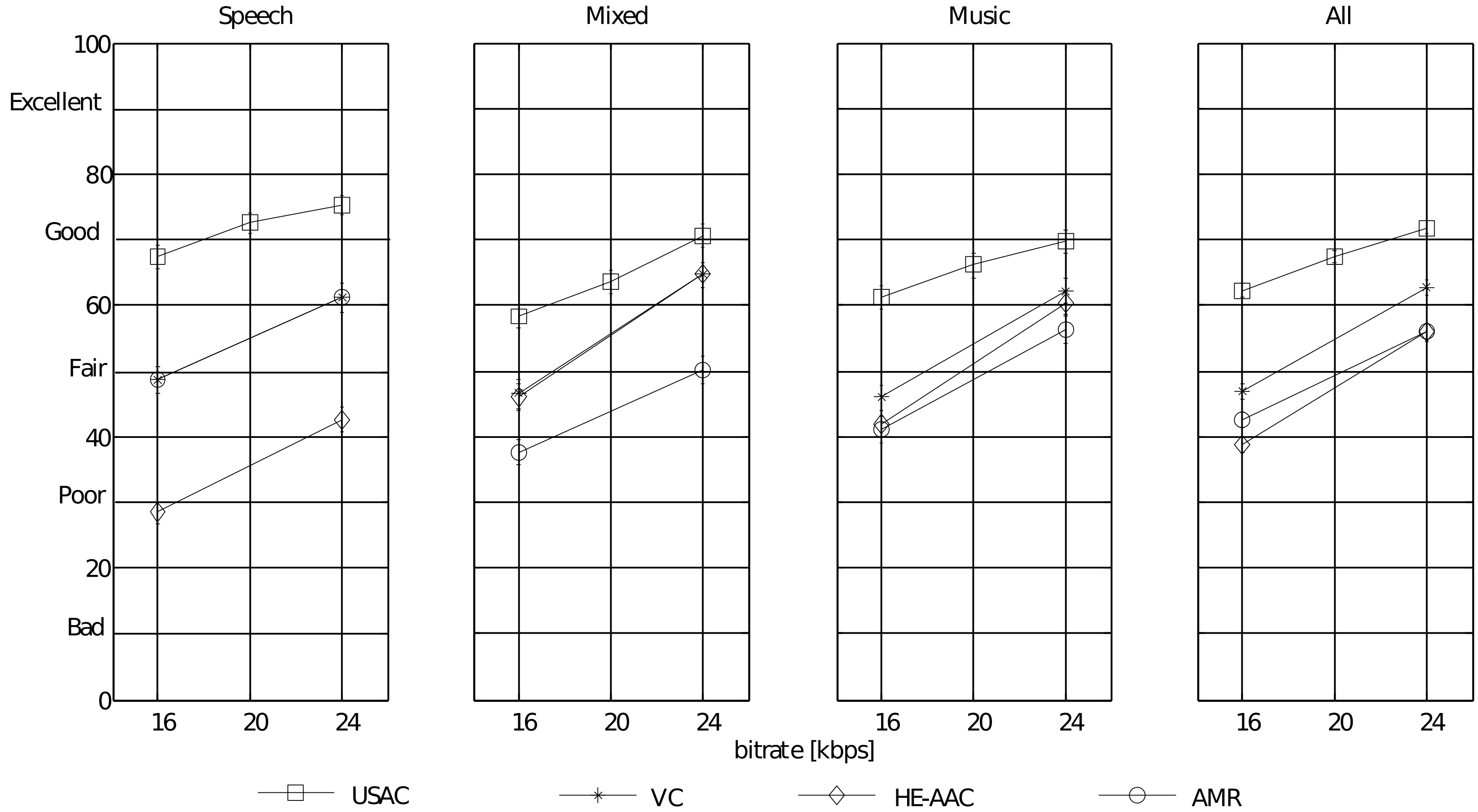
Gerald.schuller@tu-ilmenau.de

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

USAC – Performance Evaluation

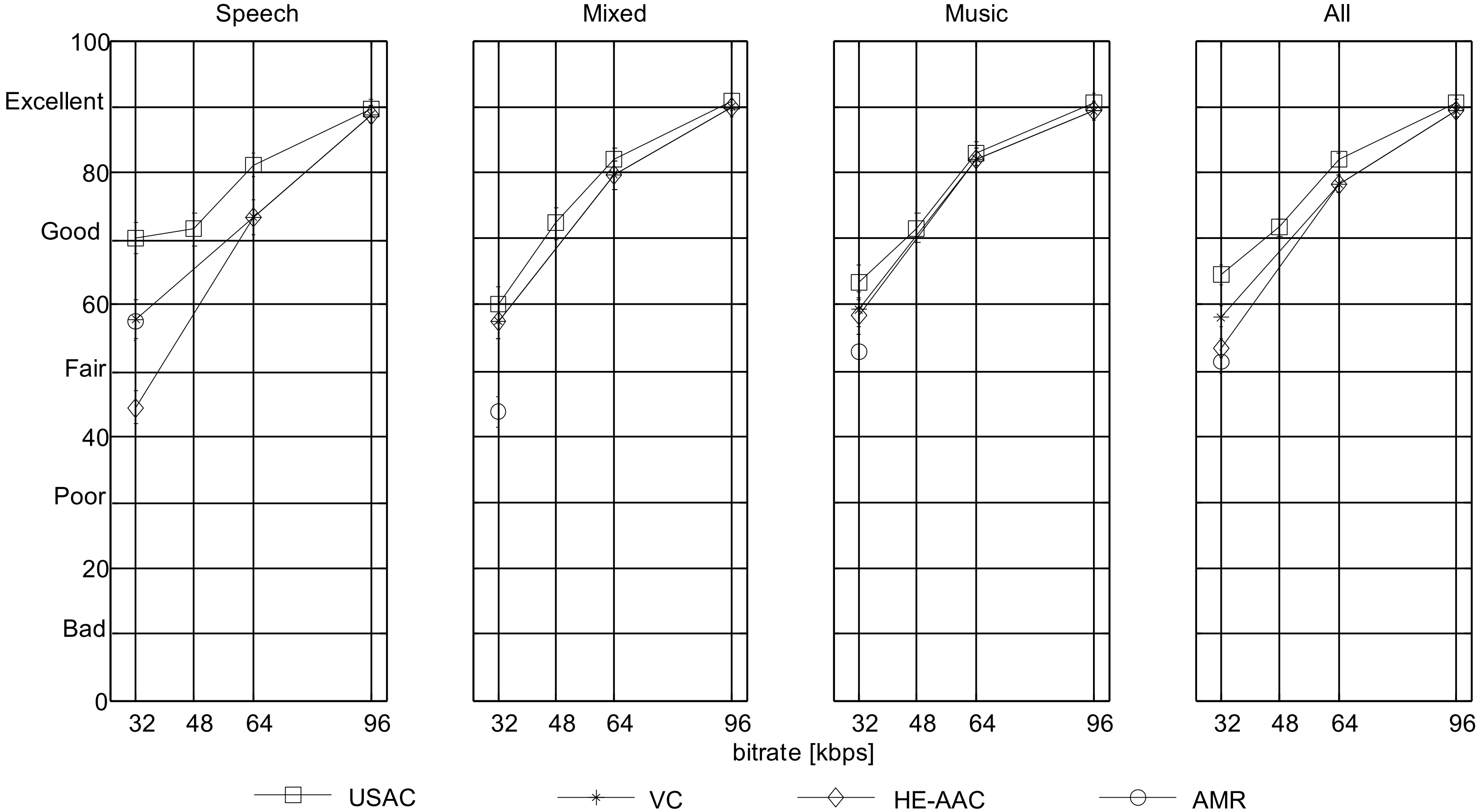
Listening Test – Stereo, Low Bit Rates



Gerald.schuller@tu-ilmenau.de

USAC – Performance Evaluation

Listening Test – Stereo, High Bit Rates



Gerald.schuller@tu-ilmenau.de

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

USAC - Applications

- Multimedia streaming
  - Mobile devices
  - Scalability is key feature
  - Significant quality improvements for low bit rates
- Broadcasting applications
  - Coding efficiency saves bandwidth
- Audio books
  - Mainly speech
  - Guarantees good quality for music and effects

USAC – Unified Speech and Audio Coding

Summary

- First audio codec that successfully merges general audio and speech coding
- For music signals, improved quality especially at low to very low bit rates
- Moderately increased computational complexity
- Standardized as ISO/IEC 23003-3:2012 MPEG-D Unified Speech and Audio Coding
- Applicable as general-purpose codec at bit rates

Gerald.schuller@tu-ilmenau.de