
MODERN HYBRID AUDIO CODING PARAMETRIC SPATIAL CODING

AES 142nd Convention Berlin, May 22nd 2017

Sascha Dick, Fraunhofer IIS Erlangen



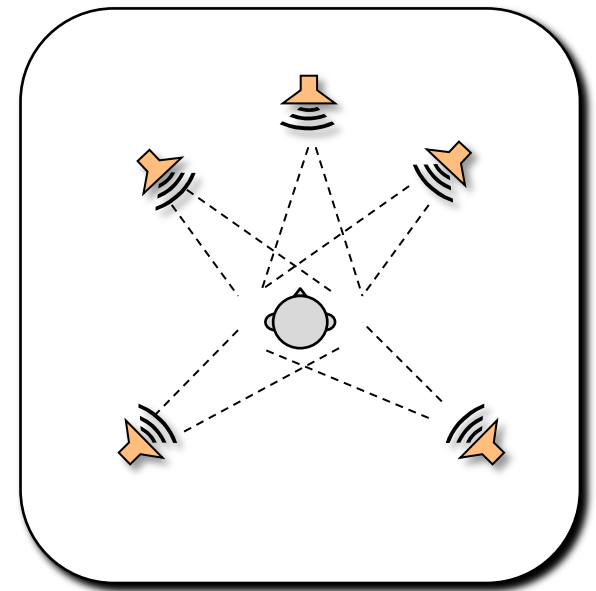
Fraunhofer
IIS

OVERVIEW

- Introduction
- Stereo Coding Techniques
- Multichannel Coding Techniques
- Recent Developments in MPEG-H 3D Audio
- Summary

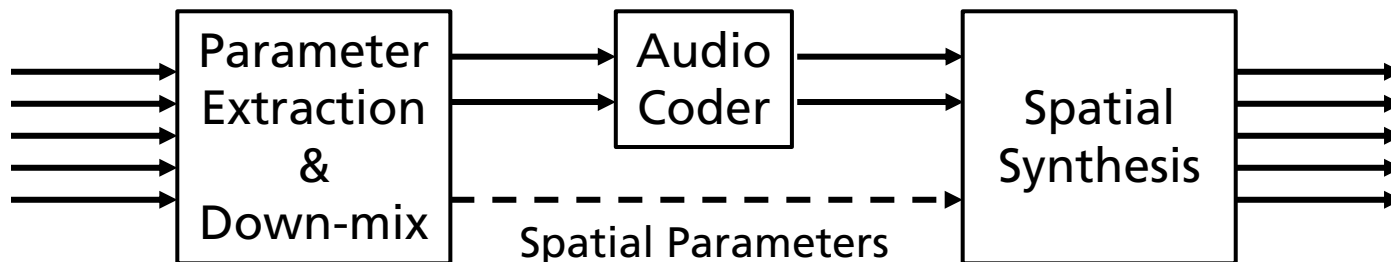
Introduction

- Stereo, surround and immersive loudspeaker setups add the dimension of space to the listening experience
- “We have only two ears...”
- Superposition of loudspeaker signals at ears
- Spatial hearing is determined by binaural cues with limited resolution, e.g.
 - Level Difference
 - Time Difference
 - Coherence




Introduction

- Parametric spatial coding approach:
 - Code fewer audio signals with transform coder
 - Distribute to multiple channels using parameters describing “space”
- Parameters based e.g. inter-channel cues derived from binaural cues
- Allows scaling of spatial accuracy depending on bit rate

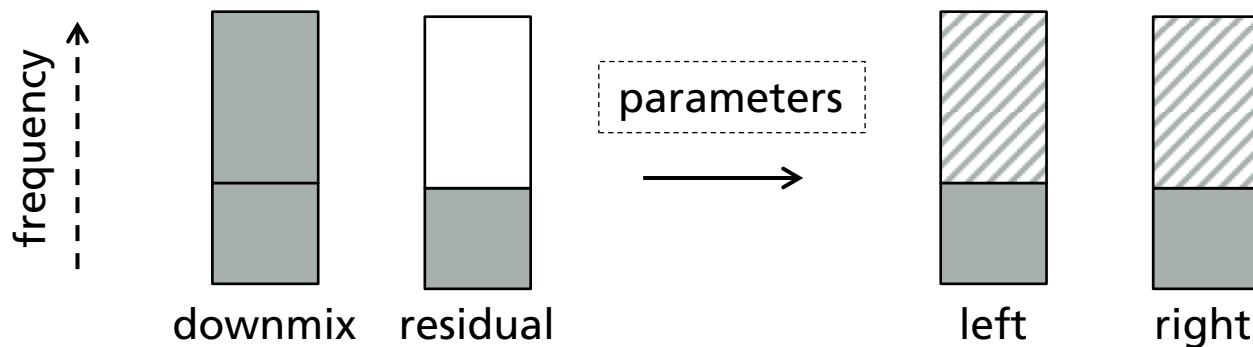


History of Stereo Coding

- 
- ~1930s ■ Mid/Side (MS) - Stereo (analog radio)
 - Transmission of sum and difference signals
 - ~1950s ■ "Pseudo Stereo" (Lauridsen, Schroeder)
 - Mono to stereo expansion without parameters
 - Playback-side postprocessing, e.g. by adding reverberation
 - 1992 ■ Intensity Stereo (MPEG-1)
 - Transmission of mono downmix and measured level differences
 - Mono to stereo upmix using panning parameters
 - 2004 ■ Parametric Stereo (HE-AAC v2)
 - Advanced stereo parameters including correlation and phase
 - Spatial processing in complex-valued filter bank
 - De-correlation to widen stereo signal

Residual Coding

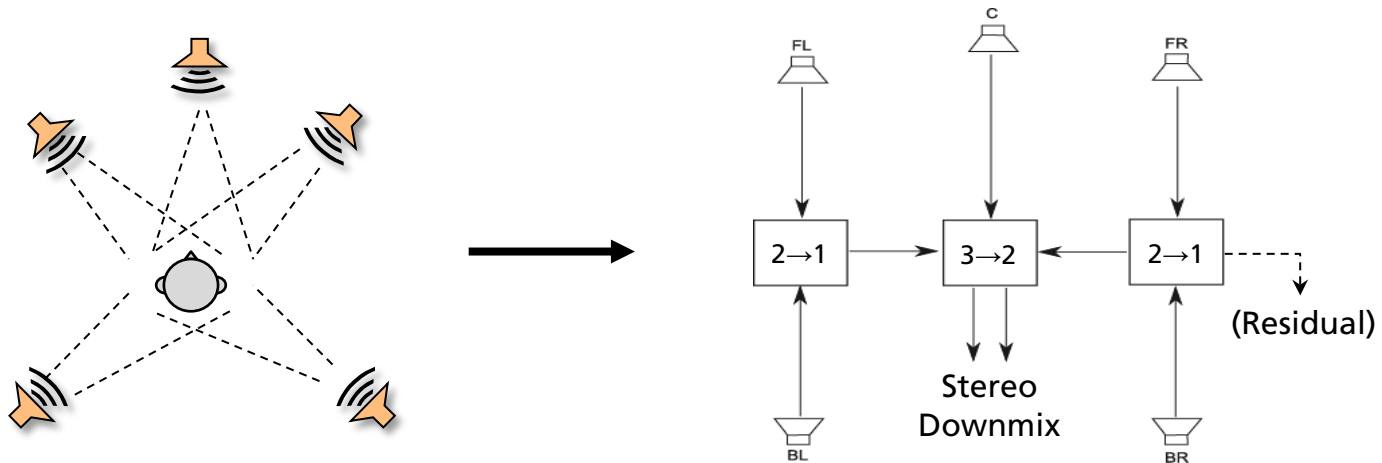
- Quality of parametric coding saturates
→ gap between parametric and non-parametric coding
- Hybrid approach
 - Lower frequencies: discrete coding of residual (= “difference-signal”)
 - Upper frequencies: parametric coding
- Crossover frequency is scalable dependent on the bit rate
- E.g. in MPEG Surround (2007), MPEG-D USAC (“Unified Stereo”, 2012)



Multichannel Coding

MPEG-D MPEG Surround (2007)

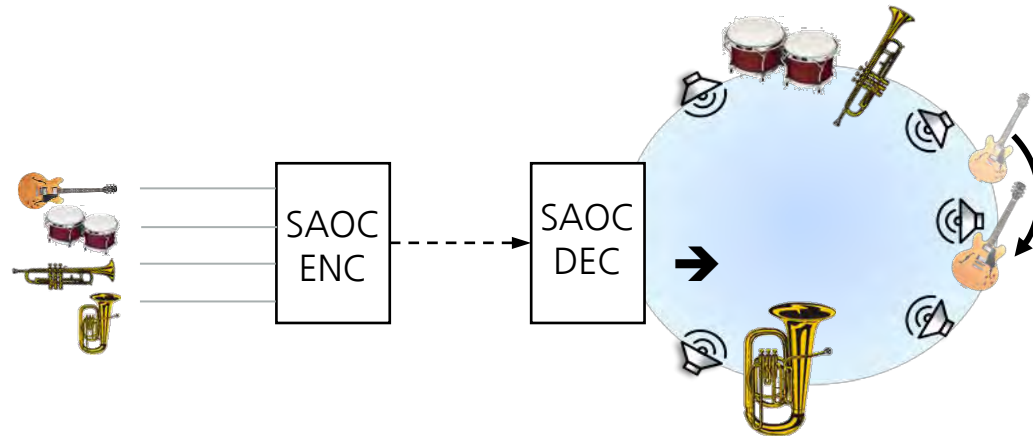
- Extension of parametric stereo coding principles to surround sound
- Tree structure of parametric coding boxes
- Residual Coding possible
- Transmission of stereo downmix plus parameters
 - ➔ Enables playback on legacy stereo devices



Multiobject Coding

MPEG-D Spatial Audio Object Coding (SAOC) (2010)

- Generalized approach: parametric coding of audio objects, independent of loudspeakers
- Enables user interaction, e.g. Dialogue Enhancement
- Decoder renders objects to loudspeakers using parameters and user interaction data

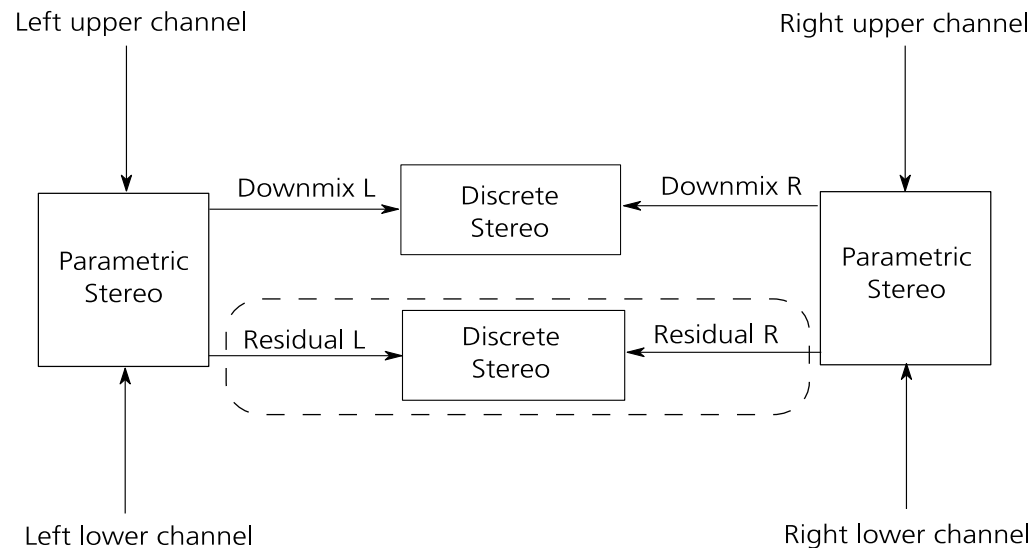


Immersive Multichannel 3D-Audio Coding



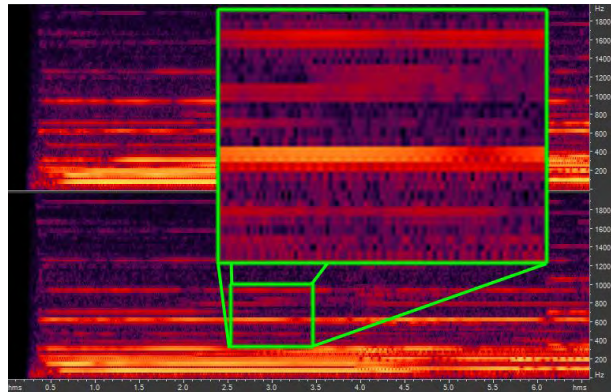
MPEG-H Quad Channel Element

- Recent developments: Immersive 3D audio including height speakers
- MPEG-H 3D Audio (2015): Quad Channel Element (QCE)
- Stereo pair in 2D → 4 channels in 3D
- Parametric coding of height (optional residual), discrete left/right coding

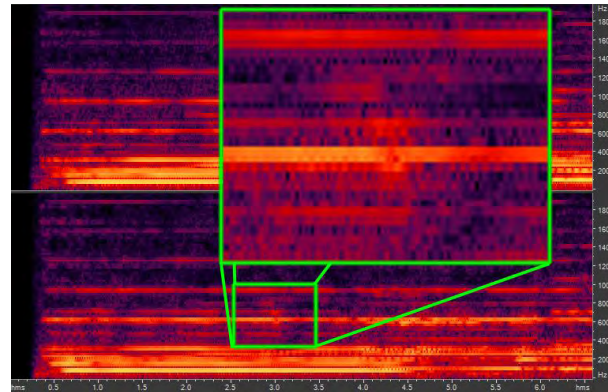


MPEG-H Hybrid Residual Coding

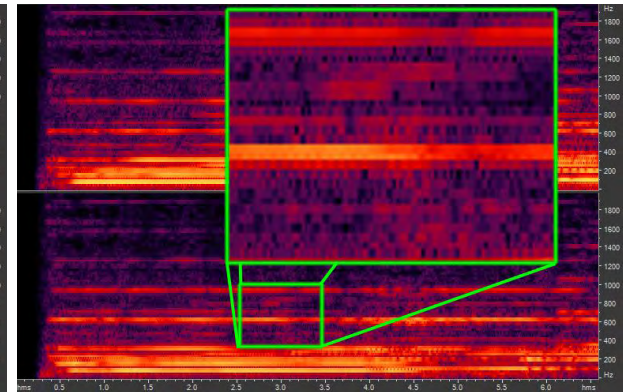
- For low bit rates, QCE operates with parametric height coding
- Phase cancellations can lead to artifacts in downmix of tonal signals
- Hybrid Residual Coding:
 - Transmit residual signal only where downmix artifacts would occur
 - Decoder measures energy, blends between decorrelator and residual



original



purely parametric



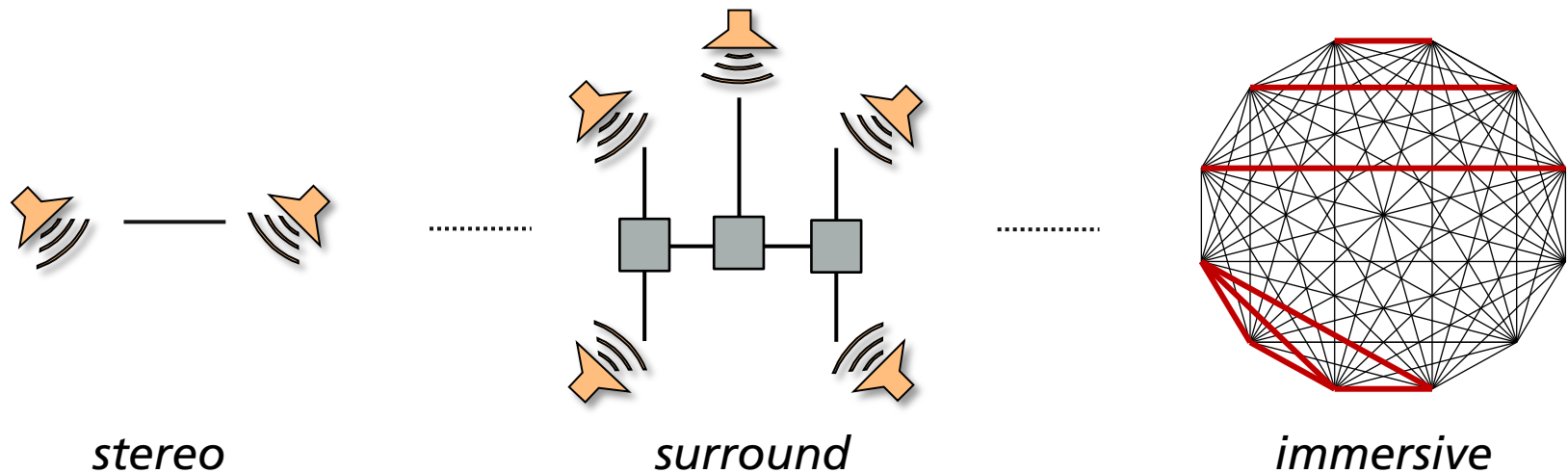
hybrid residual

MPEG-H Stereo Filling

- Enhancement for discrete stereo tools (MS-Stereo + prediction based)
- Semi-parametric operation mode for low bit rates
- Fill spectral holes caused by coarse quantization of the side signal with the previous frame's downmix signal instead of random noise
- Operates in the MDCT Domain
 - No additional filter bank (e.g. QMF) needed
 - No additional delay and low complexity

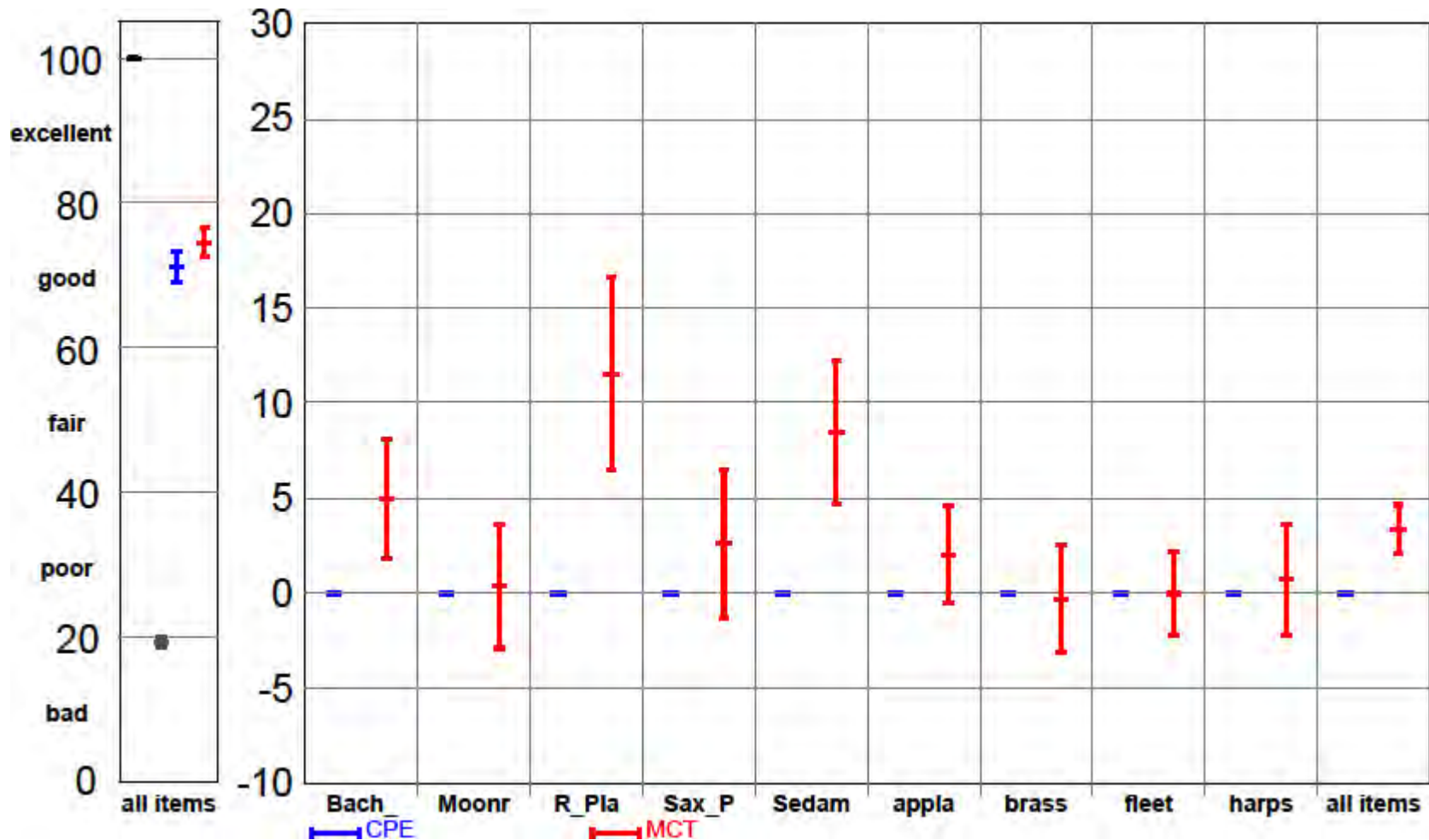
MPEG-H Multichannel Coding Tool (MCT)

- 3D Audio: many channels with time-variant inter-channel relations
 - One predefined topology cannot cover all possible relations!
- Solution: MCT enables time variant and signal dependent
 - selection and concatenation of stereo coding pairs
 - semi-parametric operation using Stereo Filling
- All processing is done in MDCT Domain



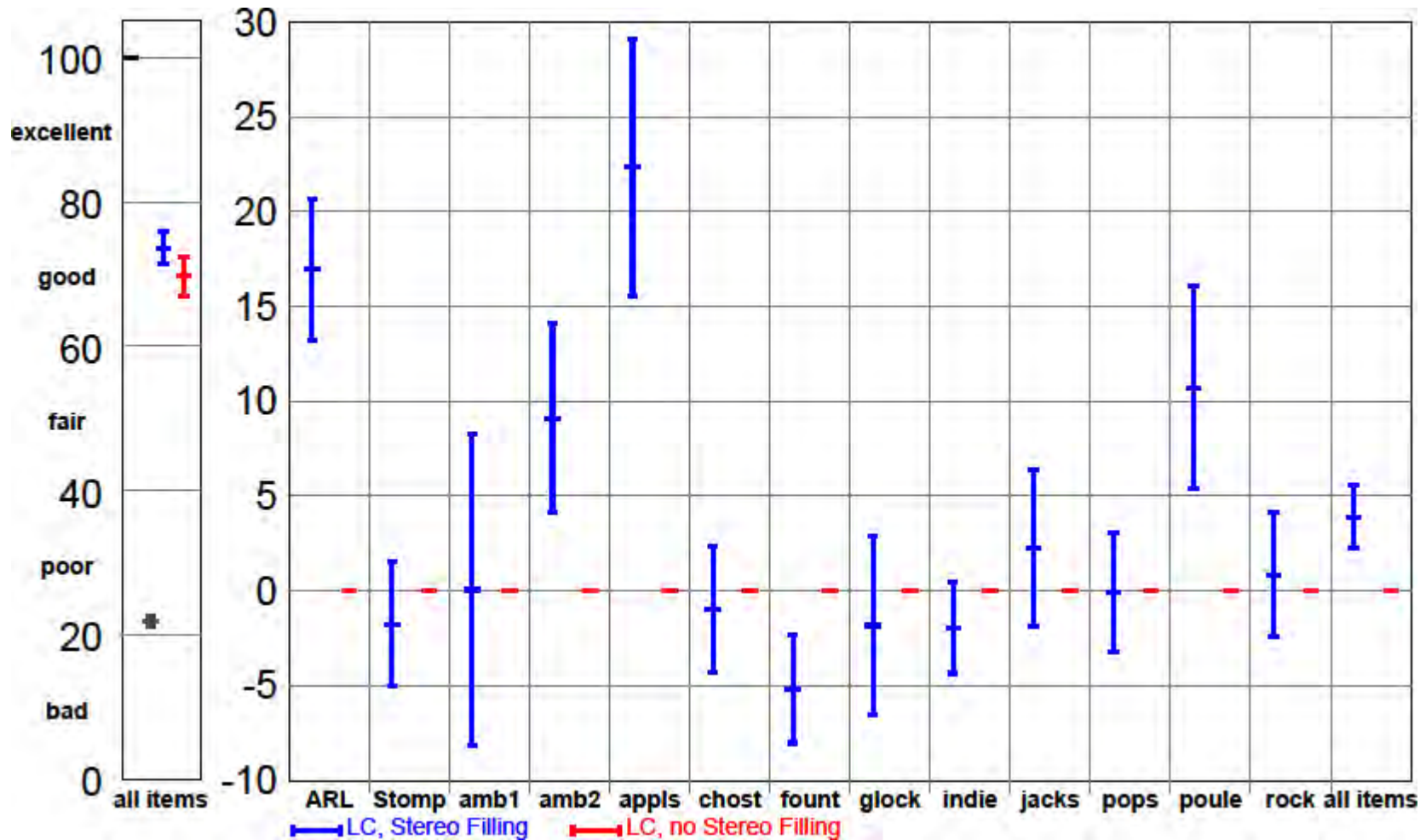
Listening Test Results: Channel Pairs vs MCT

5.1 at 144 kbps, average and 95% confidence interval, overall absolute and detailed differential scores



Listening Test Results: Stereo Filling

5.1 at 80 kbps, average and 95% confidence interval, overall absolute and detailed differential scores



Summary

- Parametric “spatial audio coding” enables multichannel audio at very low bit rates (e.g. ≤ 32 kbps stereo, 48 kbps 5.1)
- This is in fact a *Hybrid Coding* approach:
 - Carries basic (downmix) waveforms using waveform coding
 - Represents spatial sound image by parametric means (i.e. perceptually motivated spatial parameters)
 - Optionally, some schemes allow transmission of residual signals
- Modern hybrid spatial coding tools enable seamless scaling between parametric and waveform preserving operation modes
 - Highly efficient modes for intermediate quality at lowest bitrates
 - Up to traditional highest quality coding
 - Depending on bit rate *and* signal characteristics

Thank you for your attention!

sascha.dick@iis.fraunhofer.de

www.iis.fraunhofer.de

Referenced MPEG Standards

■ AAC

ISO/IEC 13818-7:1997 - Information technology - Generic coding of moving pictures and associated audio information - Part 7: Advanced Audio Coding (AAC)

ISO/IEC 14496-3:2009 - Information technology - Coding of audio-visual objects - Part 3: Audio

■ MPEG Surround

ISO/IEC 23003-1:2007 - MPEG audio technologies – Part 1: MPEG Surround

■ SAOC

ISO/IEC 23003-2:2010 - Information technology - MPEG audio technologies - Part 2: Spatial Audio Object Coding (SAOC)

■ USAC

ISO/IEC 23003-3:2012 - Information technology - MPEG audio technologies - Part 3: Unified speech and audio coding

■ MPEG H 3D Audio

ISO/IEC 23008-3:2015 - Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio