MODERN HYBRID AUDIO CODING BANDWIDTH EXTENSION METHODS

142nd AES Convention Berlin, May 22nd 2017

Andreas Niedermeier, Fraunhofer IIS Erlangen

Fraunhofer IIS

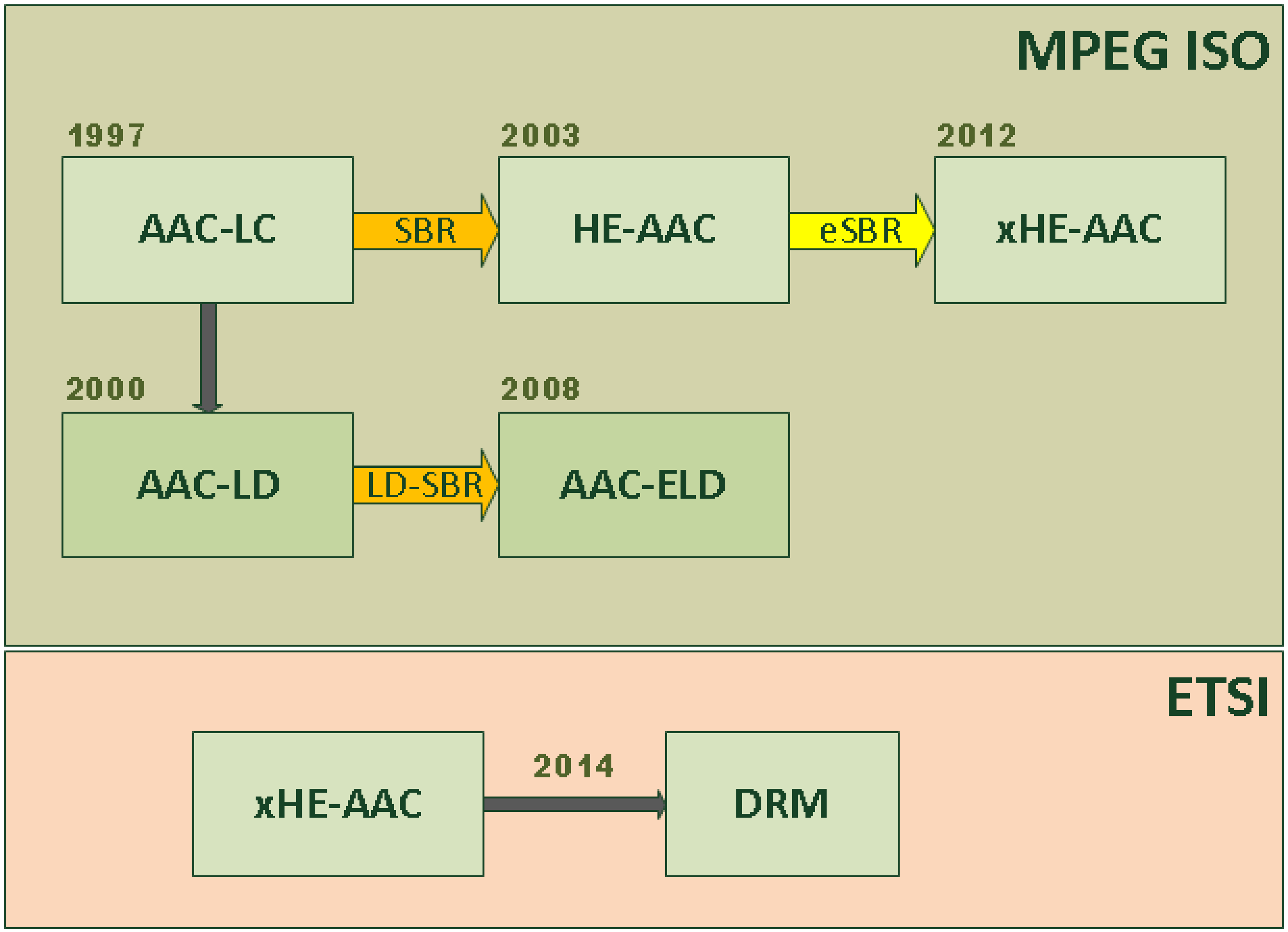
IMPROVED TECHNOLOGY AFTER SBR OVERVIEW

- **Enhanced SBR (eSBR)**
  - Harmonic Bandwidth Extension (HBE)
  - Predictive Vector Coding (PVC)
  - eSBR tools
- **Intelligent Gap Filling (IGF)**
  - Motivation
  - Decoding process
  - IGF tools
- **Summary**

Waveform Preserving -, Parametric - and Hybrid Audio Coding

- **Waveform preserving audio coding**
  - Usally done with a transform coder
  - Scales up to transparency
  - Has a relatively high bit-consumption
  - Example: AAC-LC
- **Parametric audio coding**
  - Does usually not scale up to transparency
  - Has a relatively low bit-consumption
  - Example: parametric bandwidth extensions
- **Hybrid audio coding**
  - combines both approaches

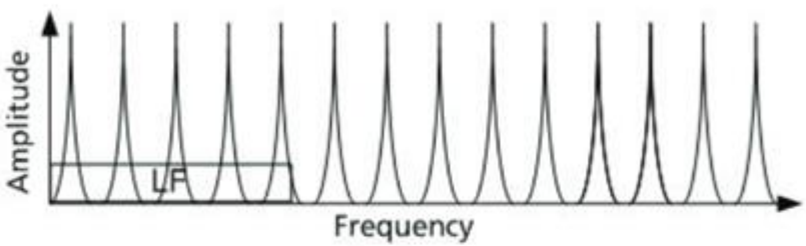
Enhanced SBR (eSBR)



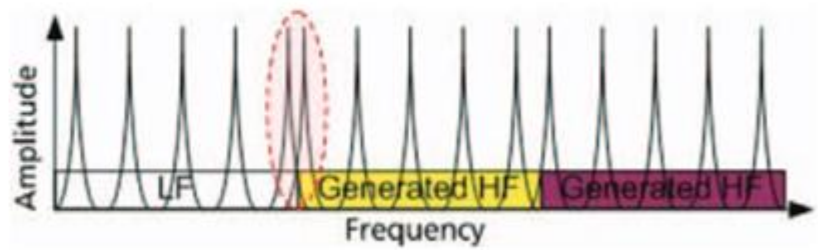
Fraunhofer IIS

Harmonic Bandwidth Extension (HBE) in eSBR

- SBR produces roughness and tonal beating:



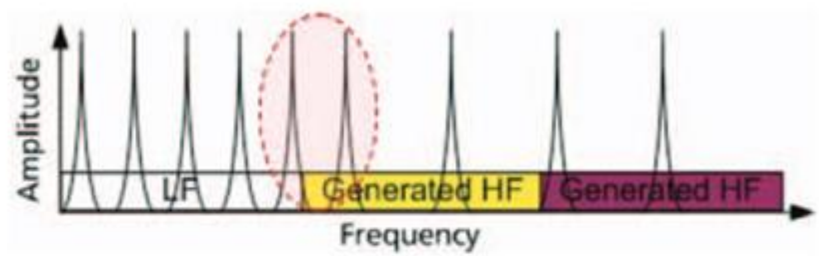
(a) Original



(b) SBR

- HBE preserves the harmonic continuation of spectral lines:

(c) HBE



Predictive Vector Coding (PVC) in eSBR

- With speech content SBR is not performing optimally at low bitrates (8 kbps mono,12 kbps mono, 16 kbps stereo, 24 kbps stereo)
- **Predictive Vector Coding (PVC)**
  - Exploits the correlation between LF and HF with speech content
  - Offers a higher time resolution compared to SBR
  - Uses a predefined table of HF envelopes instead transmitting the HF envelope as side information

Further eSBR tools

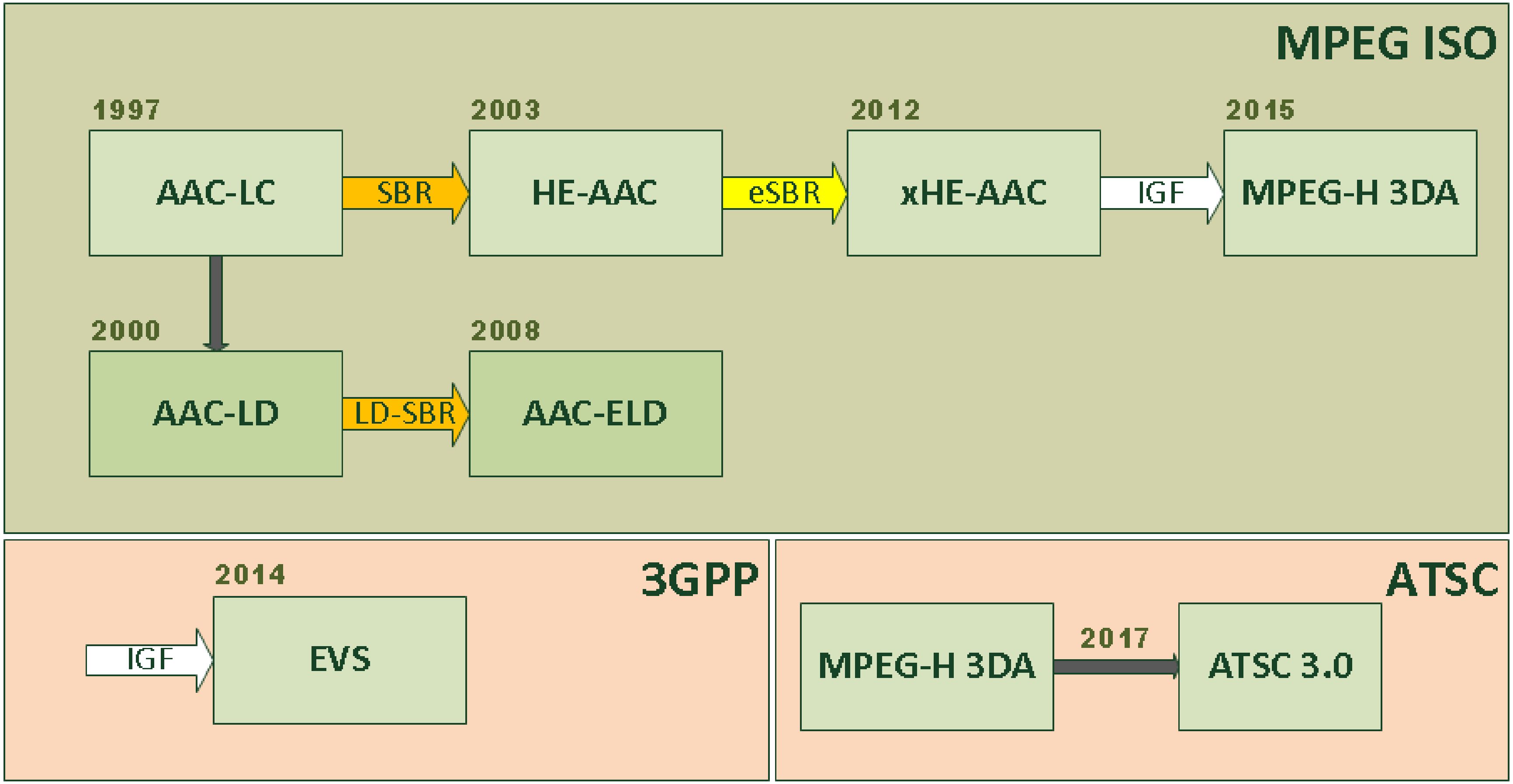
- **interTES**
  - Is a low cost time-domain envelope shaping tool
  - Helps to suppress pre- and post echoes
- **Improved time-resolution** of the SBR grid
  - Helps in non-stationary signal parts
- **New up-sampling ratios** 1:4 and 3:8 ratio
  - Enhance perceptual quality with very low bitrates

Discussion of (e)SBR

- **Upside**
  - Allows coding with very low bitrates, e.g. 12 kbps stereo (xHE-AAC)
- **Downsides**
  - **High complexity and delay**
    - AAC core coder uses a MDCT filter bank whereas (e)SBR uses an additional QMF filter bank
    - This adds to complexity and delay of a codec using (e)SBR
  - **Only parametric coding in (e)SBR range**
    - (e)SBR uses only parametric schemes to generate HF content
    - Waveform preserving coding in HF range would enable highest quality
  - **New research shows: both concepts can be combined seamlessly!**

Fraunhofer IIS

Intelligent Gap Filling (IGF)

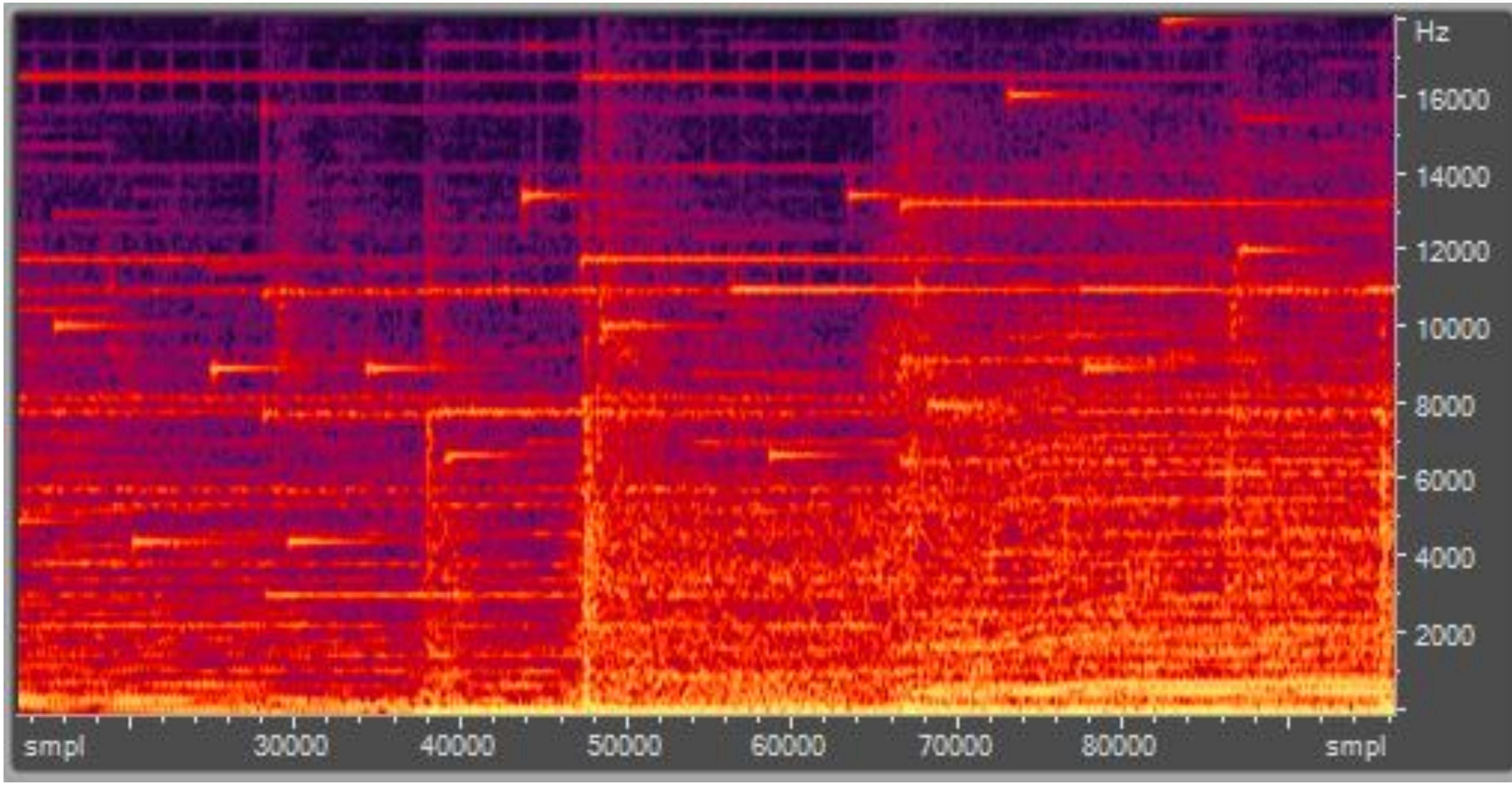


Intelligent Gap Filling (IGF)

- **Low complexity and no additional delay**
  - IGF shares the MDCT domain with the core coder
- **Universality**
  - Allows for HF waveform preserving coding through the core coder
- **Flexibility**
  - Almost free choice of start- and stop-frequency
- **Adaptivity**
  - Allows for a signal adaptive choice of high band content

Fraunhofer IIS

Intelligent Gap Filling (IGF), Principle

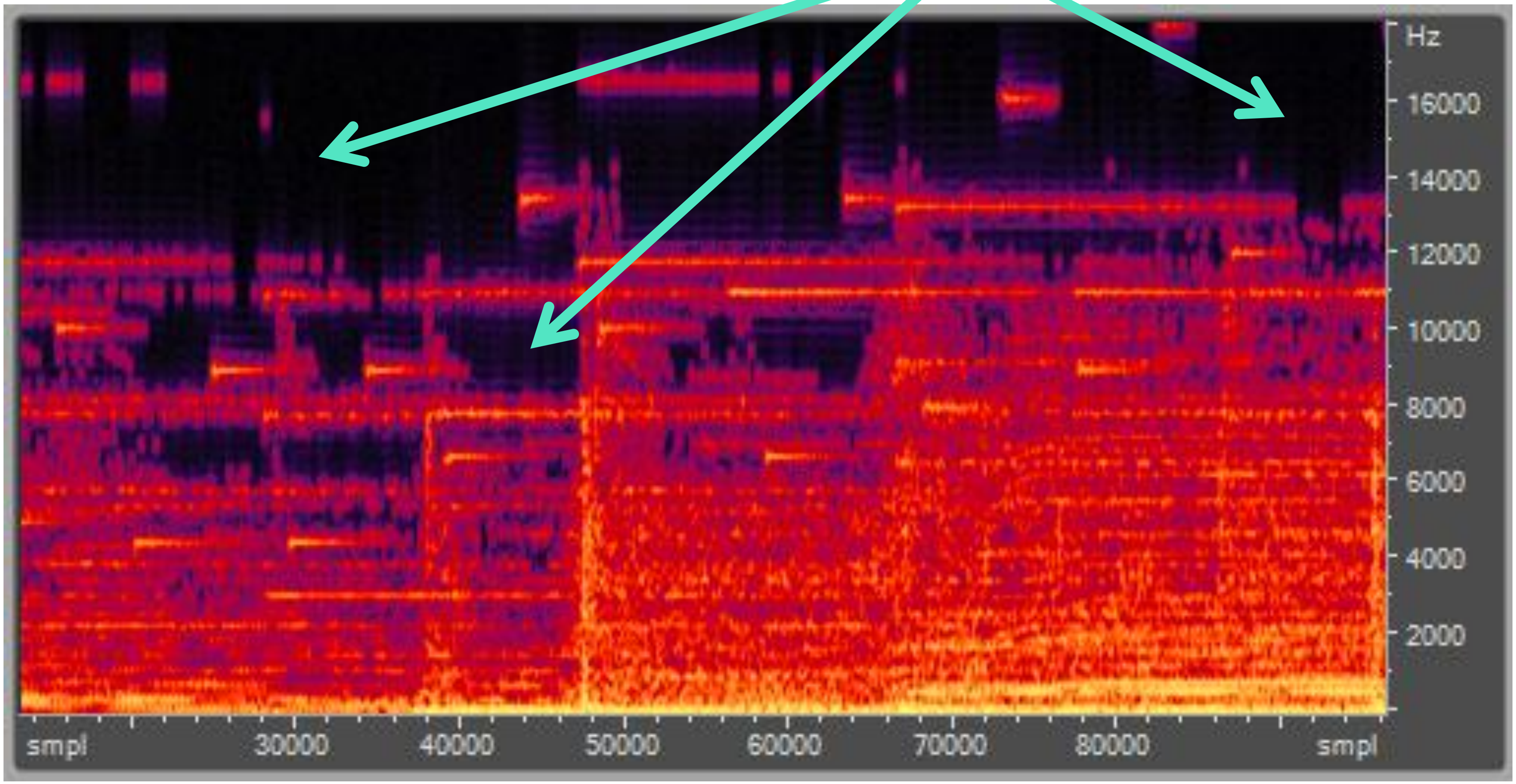


spectrogram of original audio

Fraunhofer IIS

Intelligent Gap Filling (IGF), Principle

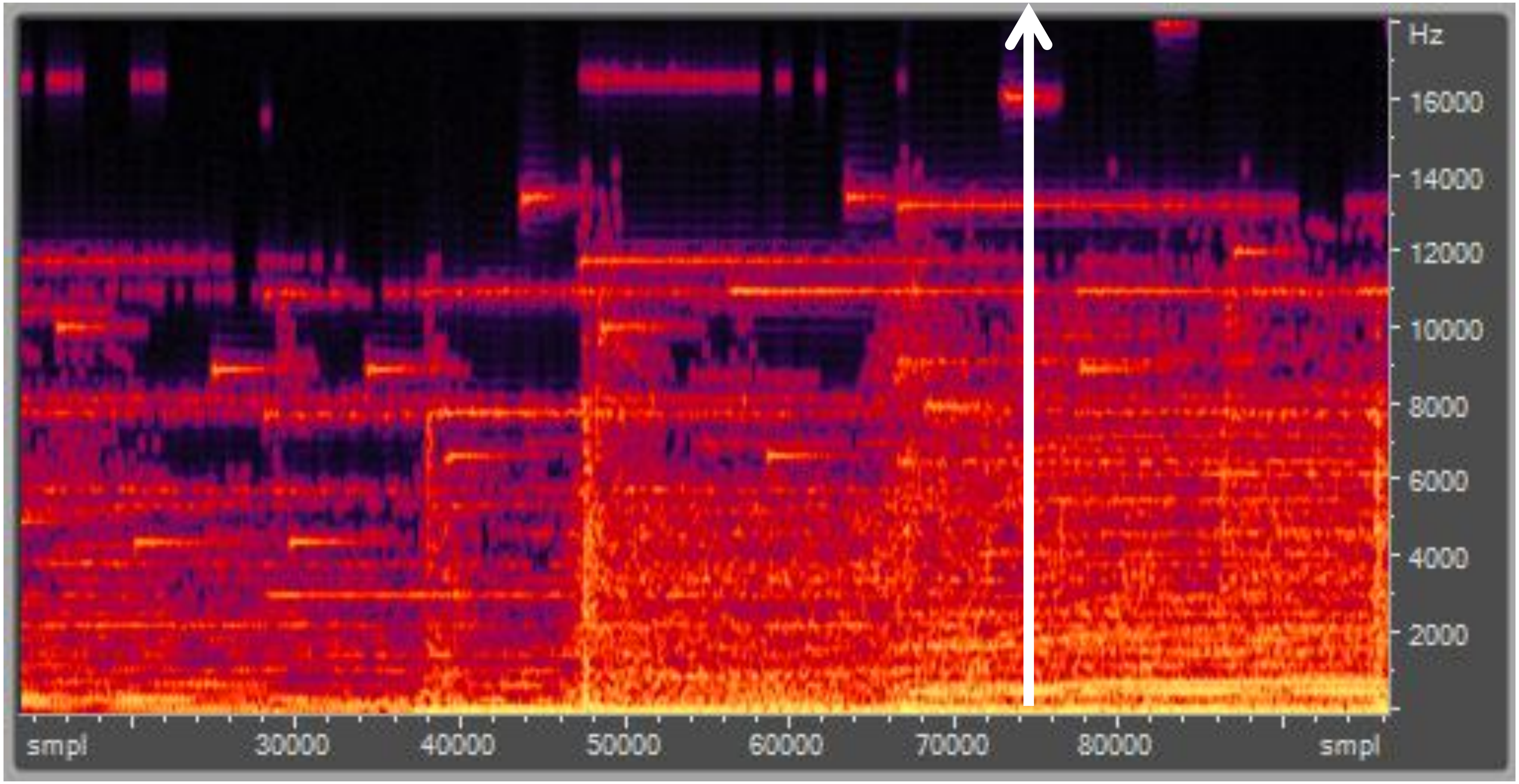
- Coarse quantization leads to spectral gaps:



spectrogram of quantized audio

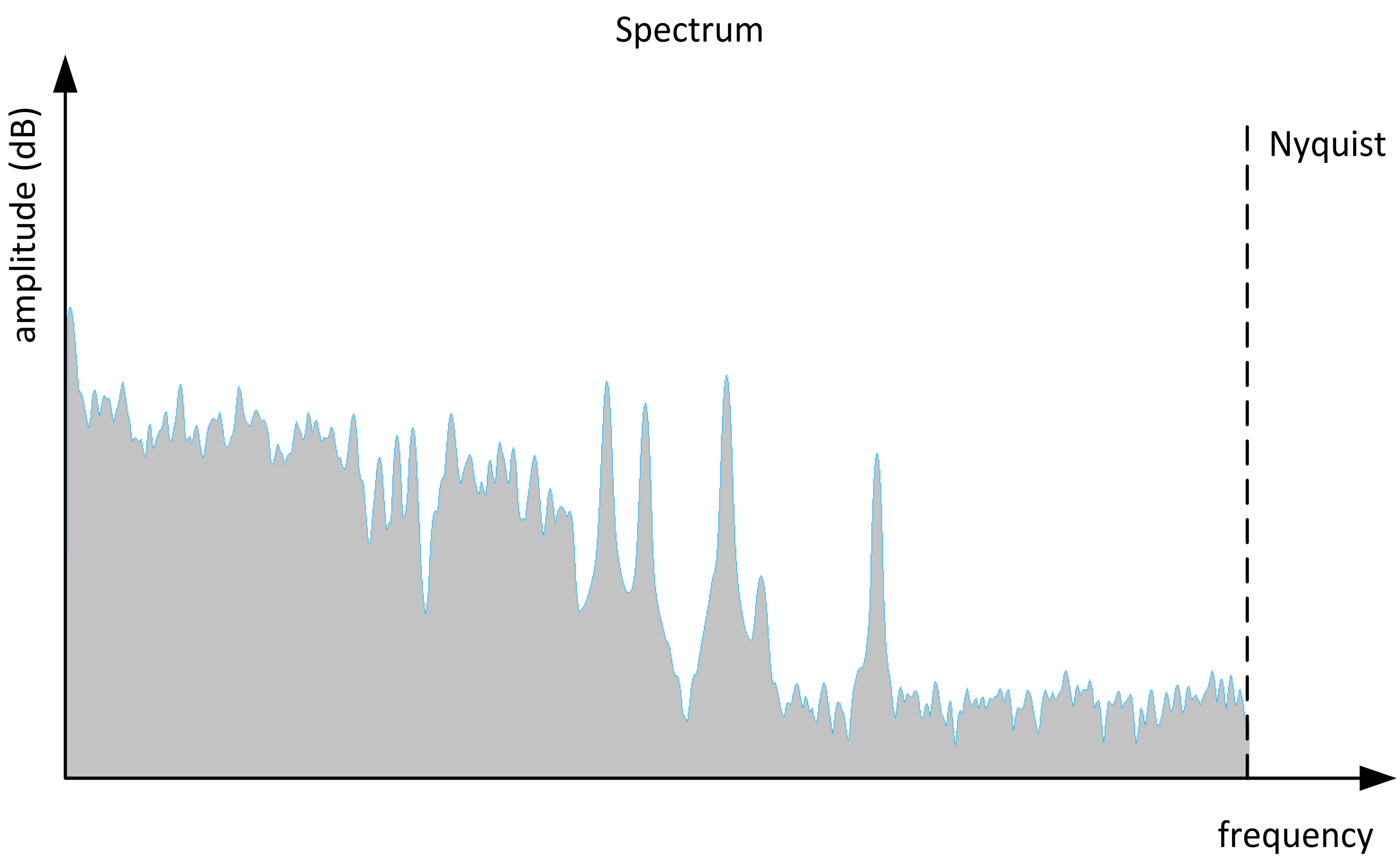
Intelligent Gap Filling (IGF), Principle

- Point in time for subsequent slides:

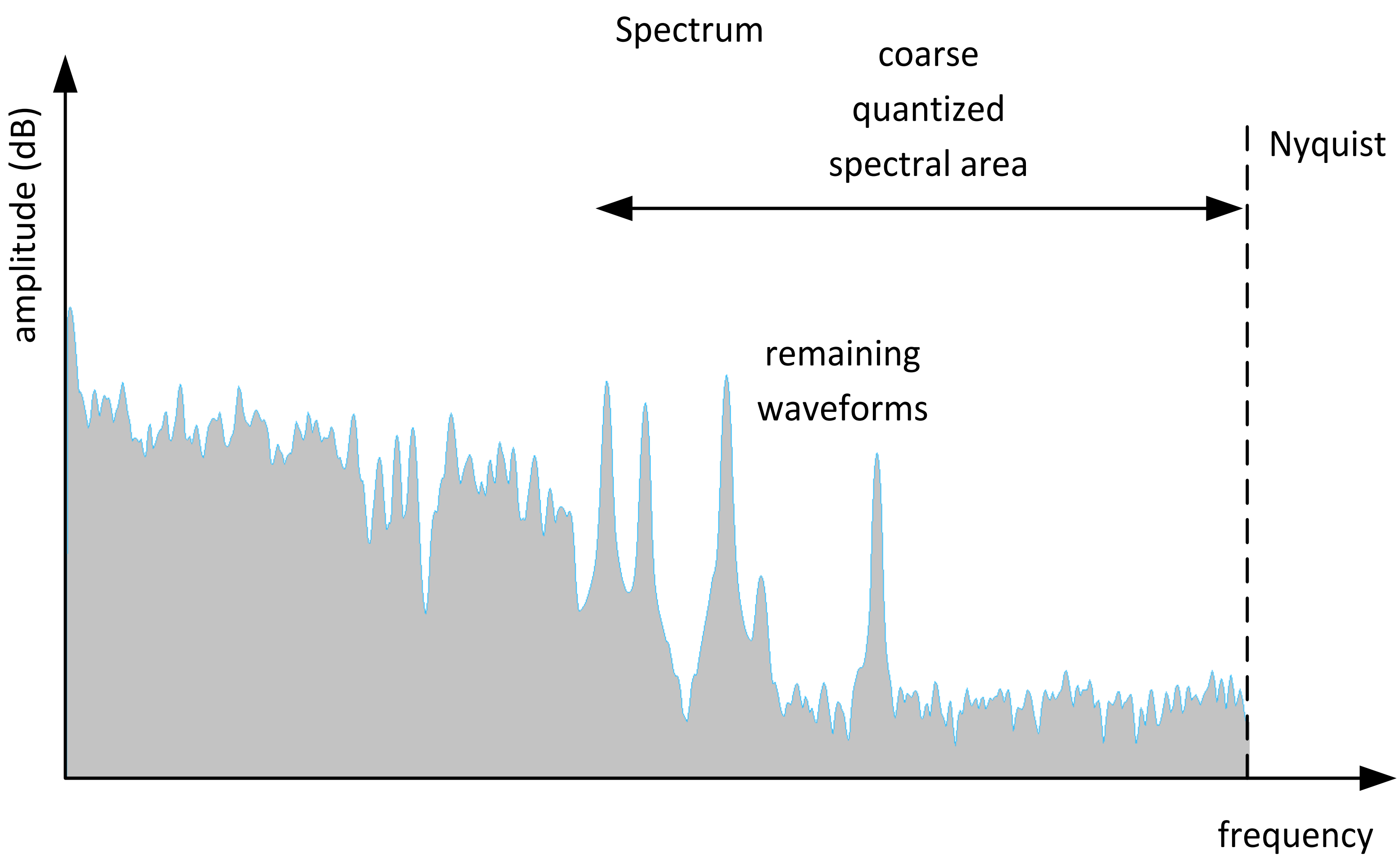


spectrogram of quantized audio

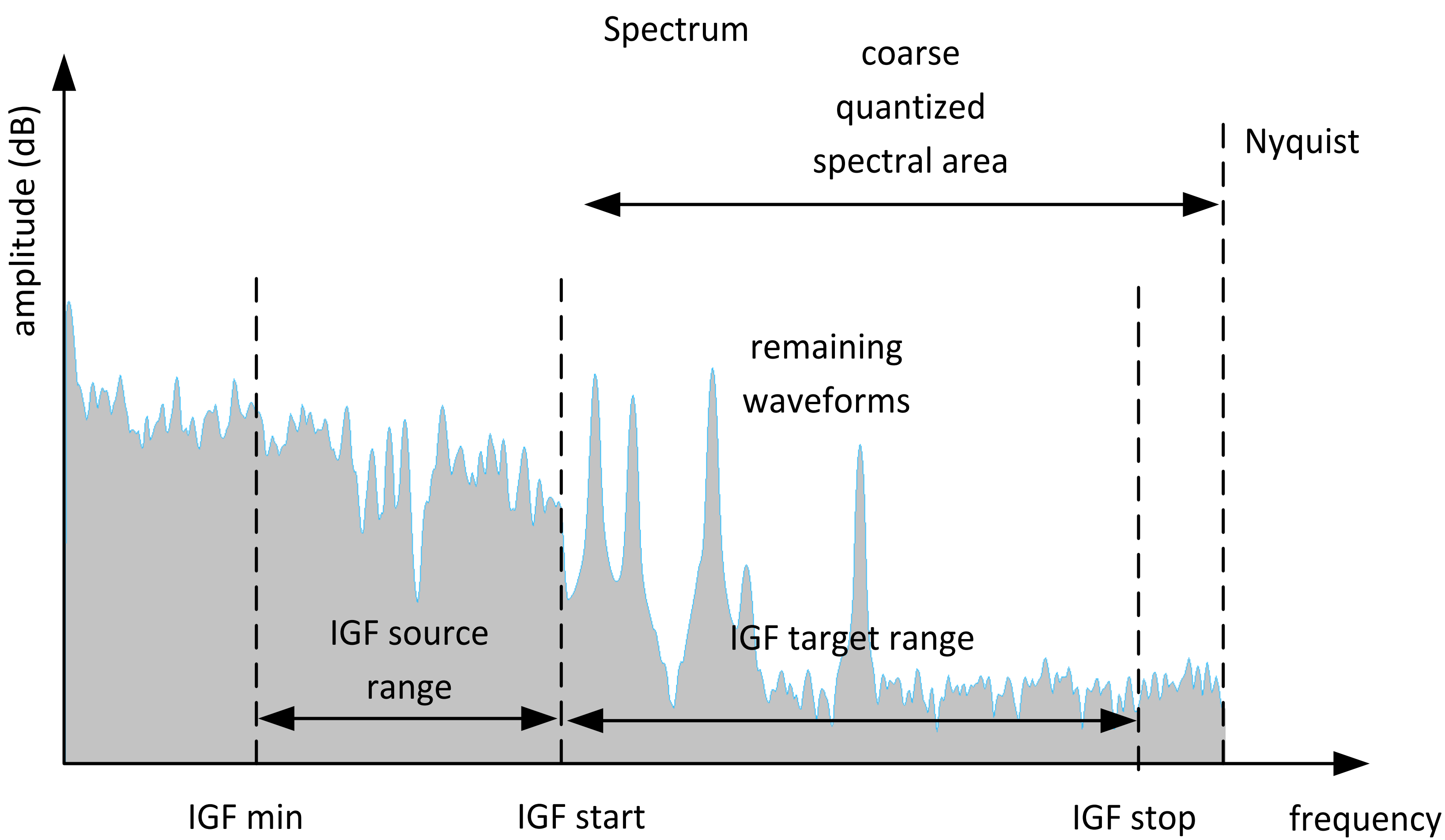
Transmitted Spectrum After De-Quantization



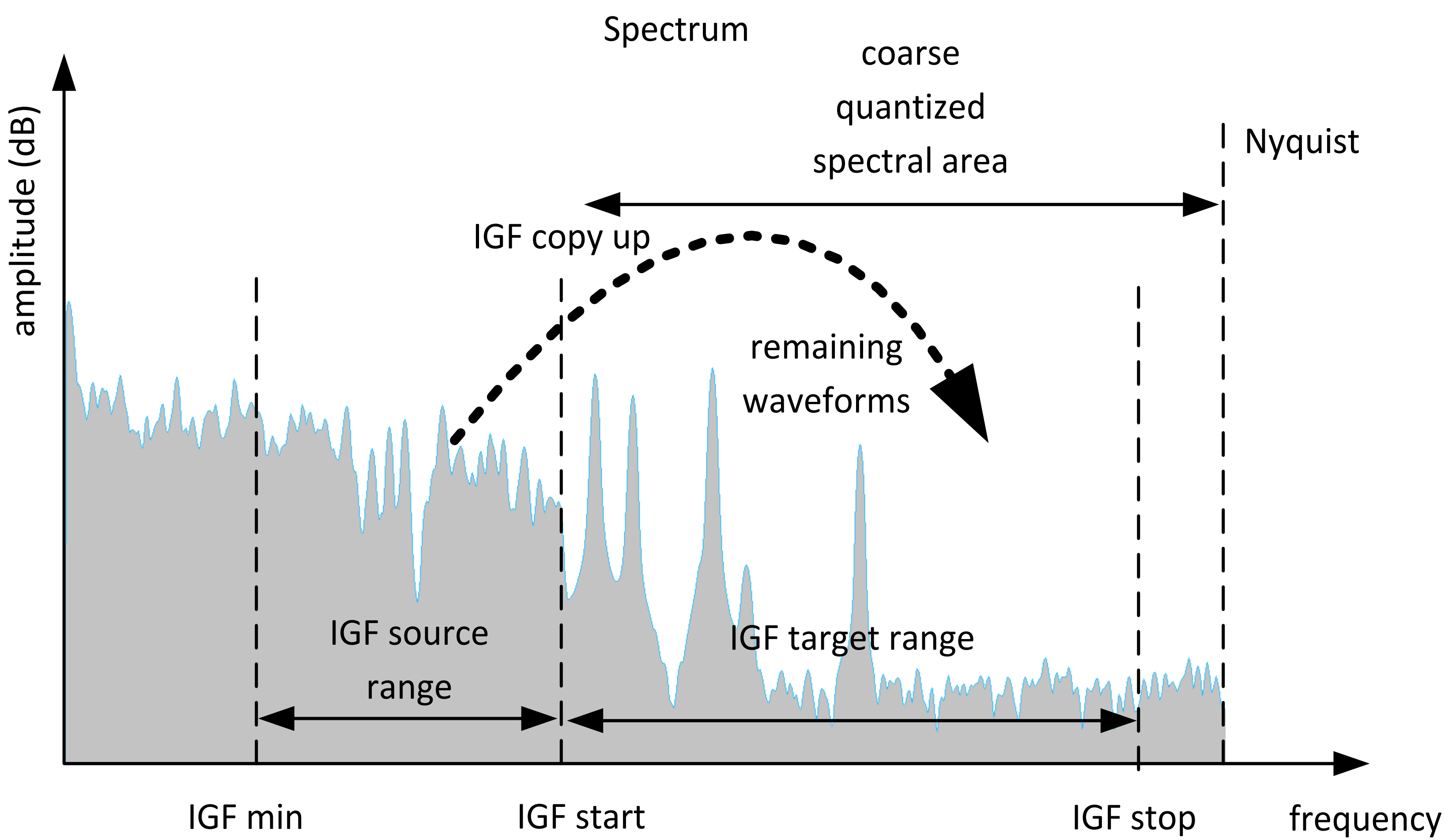
Transmitted Spectrum After De-Quantization



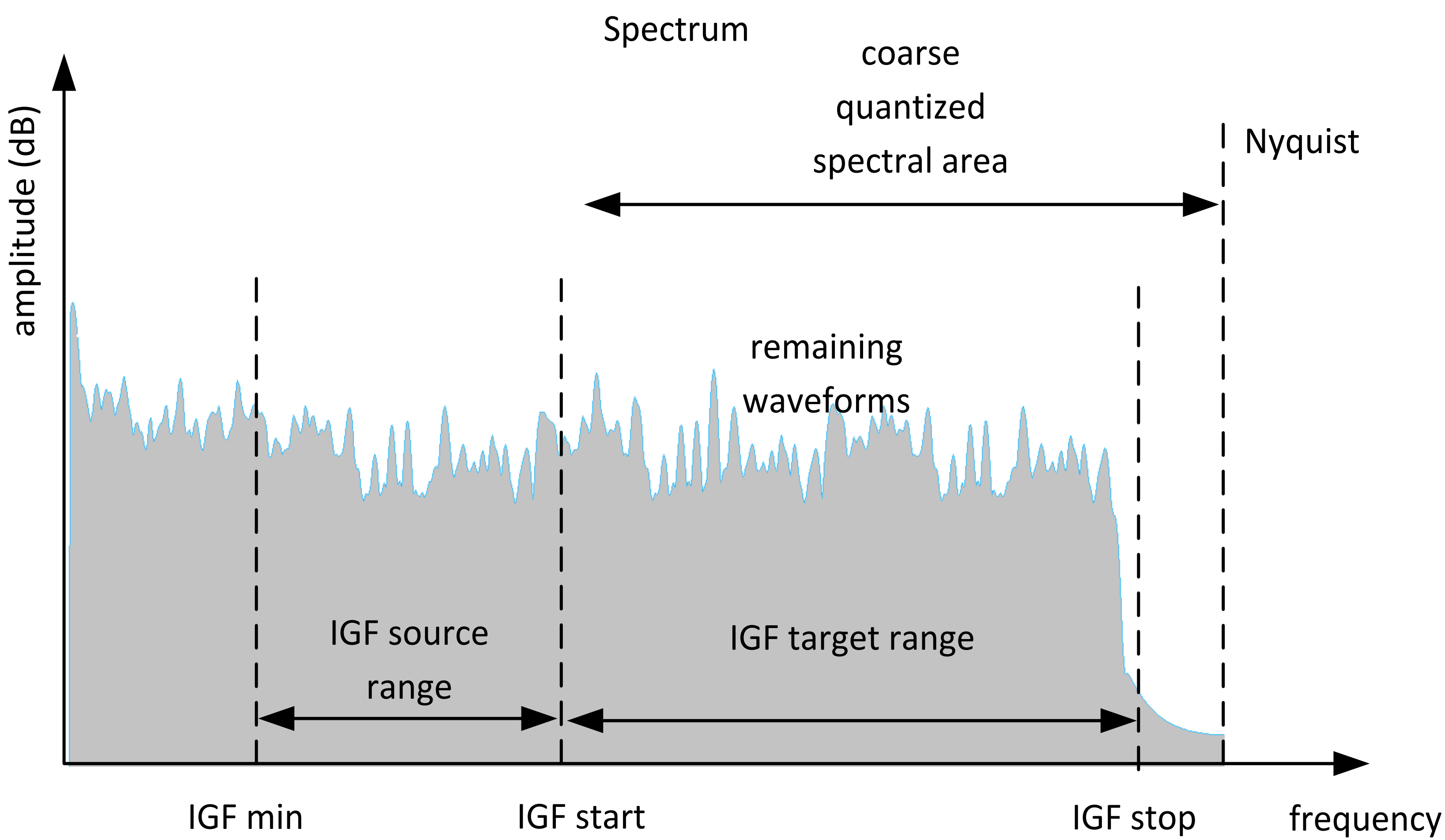
Transmitted Spectrum After De-Quantization



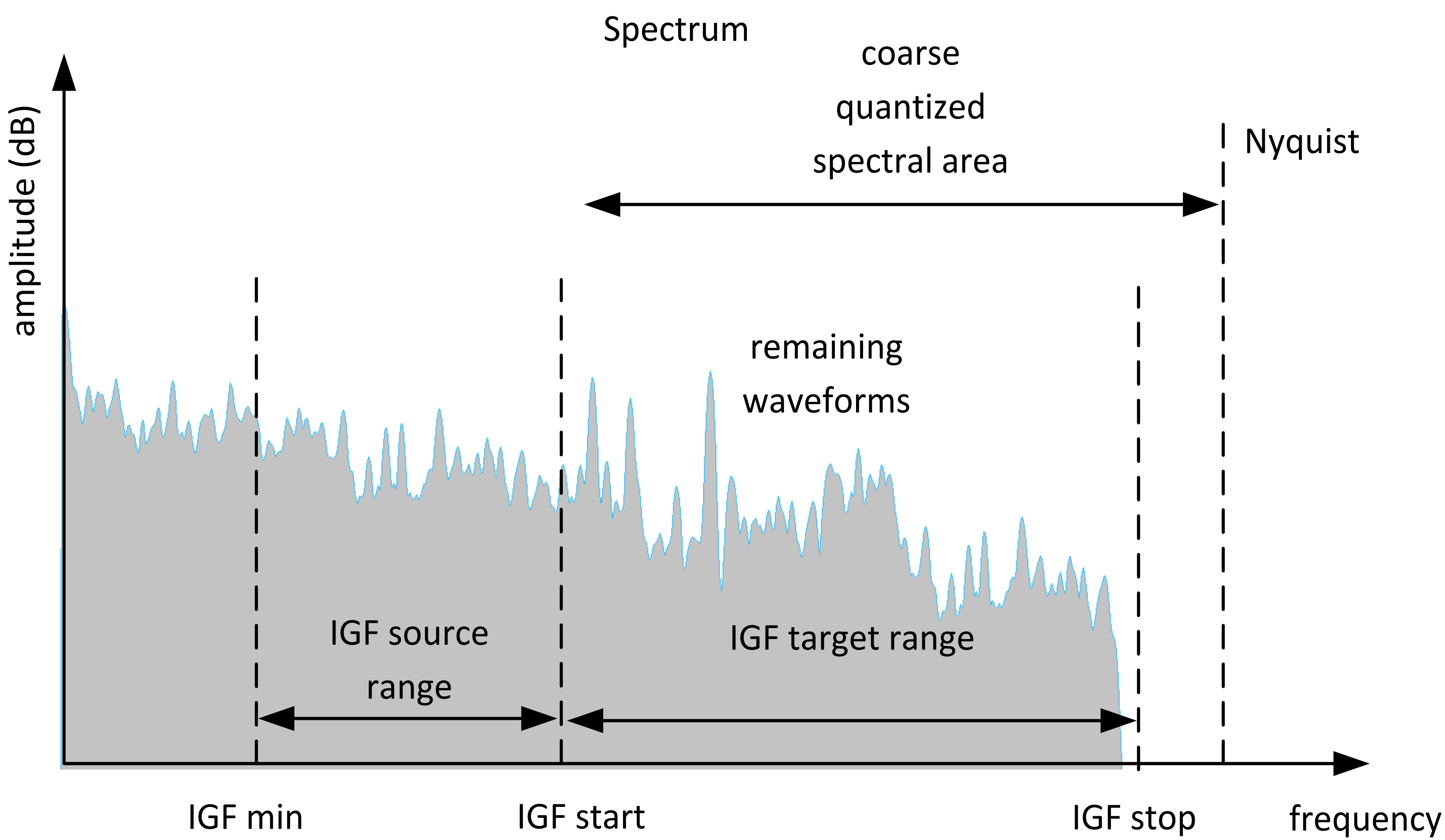
The Copy-Up Process



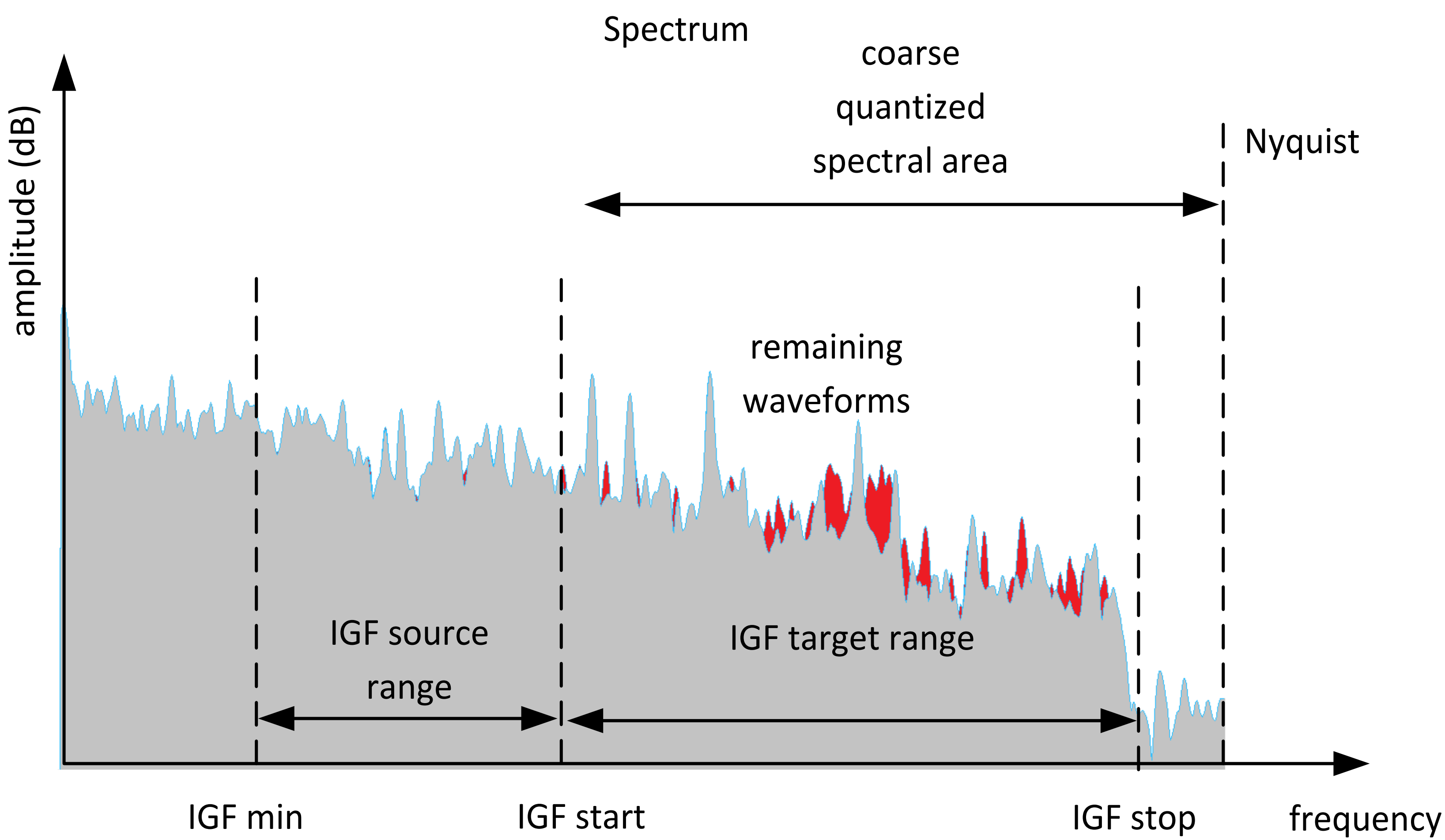
The Copy-Up Process



The Copy-Up Process, Envelope Shaping



The Copy-Up Process, Original (grey) vs Decoded (red)



Reviewing The Bandwidth Extension Copy-Up Process

- **Assumptions of a HF bandwidth extension**
    - HF content is similar to LF content
    - Both, LF and HF are tonal
    - Both, LF and HF are noise-like

- **What if HF content is not similar to LF content?**
  - By only using copy-up, perceivable artifacts will occur:
    - Wrong timbre
    - Buzzy tones
    - Rough noise
    - Sizzling noise

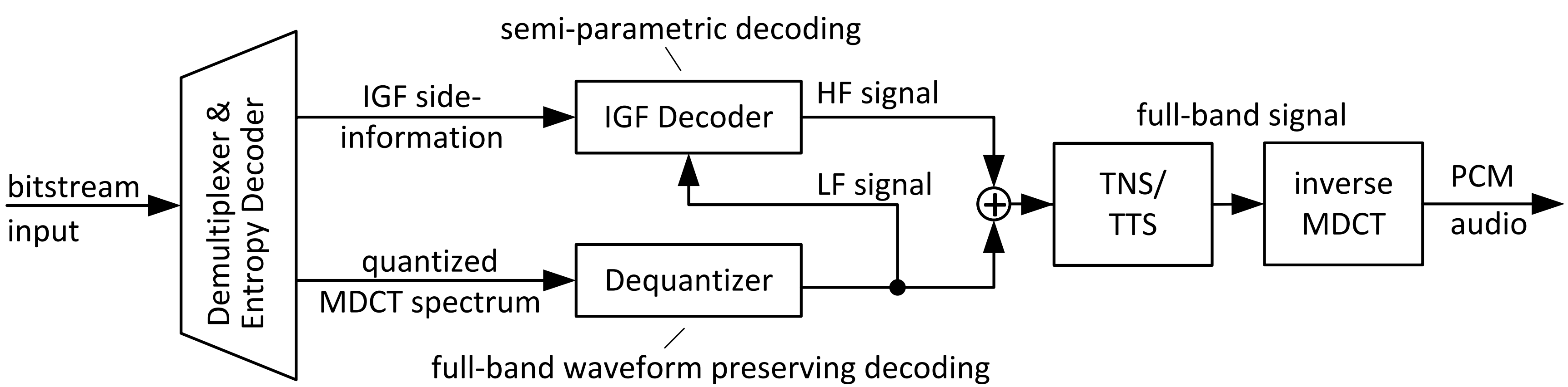
Decoder Tools To Modify IGF Target Range

- **Spectral whitening**
  - Changes tonal content in target range into noise-like content
- **Temporal Tile Shaping (TTS)**
  - Trims time domain envelope of frequencies in the IGF target range
- **Remaining waveforms**
  - Allows waveform preserving coding up to Nyquist
  - Spectral gaps are still filled with IGF

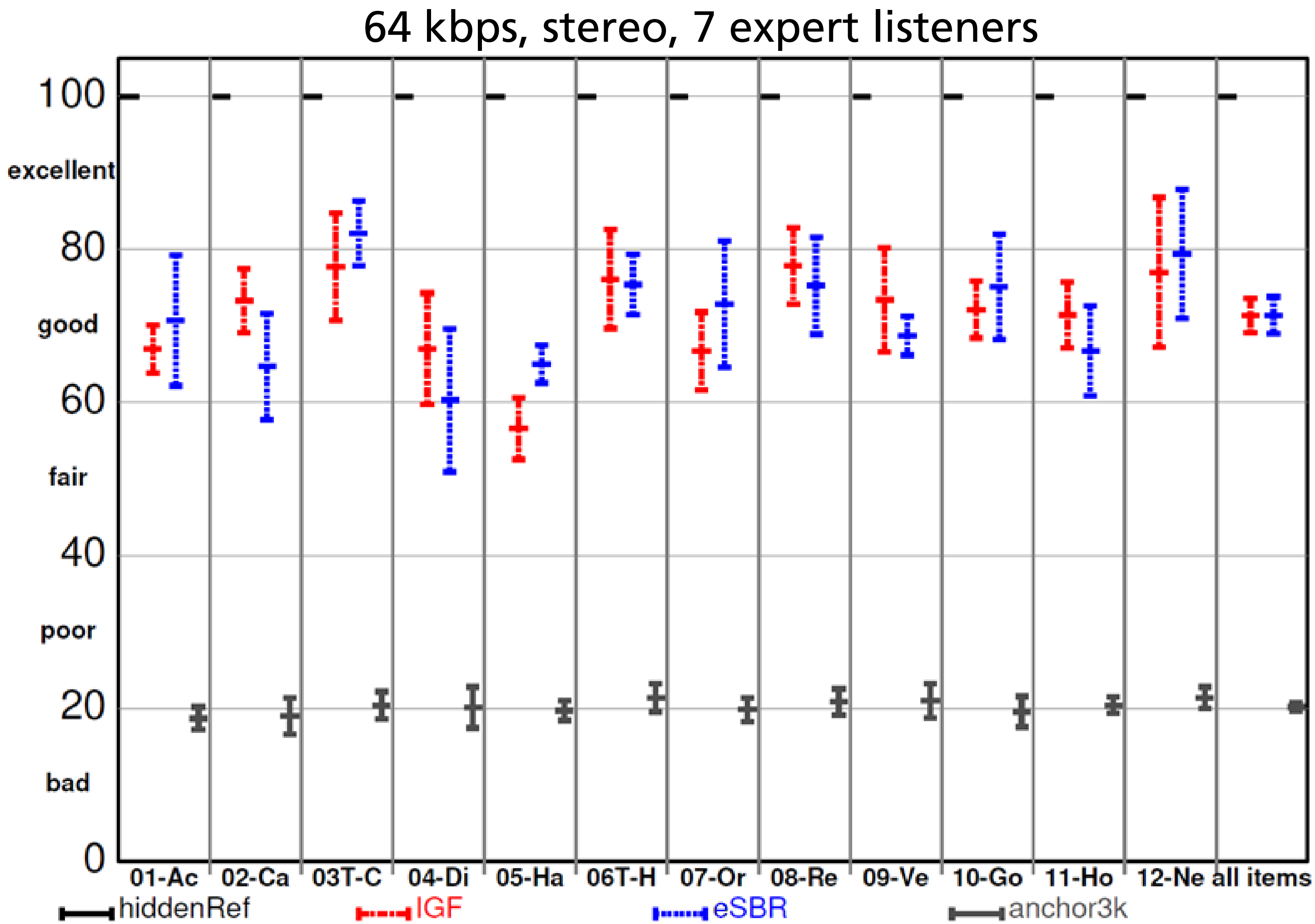
Decoder Tools To Modify IGF Target Range

- **Envelope Noise Flattening (ENF)**
  - Reduces sizzling noise with random noise substitution and time domain envelope shaping
- **Flexible tiling**
  - Selects for every frame a best matching source for the copy-up process
- **Global tuning choices**
  - Start and stop frequency are almost arbitrarily selectable
  - Low or high resolution for the target range envelope
  - Stereo tools

Decoder Signal Flow with IGF



Comparison IGF vs. eSBR in MPEG-H 3D Audio



See: Disch, Sascha, et al. "Intelligent Gap Filling in Perceptual Transform Coding of Audio." *141nd AES,* 2016.

Summary

- **eSBR enhances SBR** to improve perceptual quality with low bitrates
  - **HBE** avoids artifacts that can occur at the transition between the core coder and the SBR frequency range. Boosts performance for tonal signals.
  - **PVC** has a higher time resolution and adapts to speech content. Boosts performance for speech signals with low bitrate coding.

- **IGF** does not require an additional filter bank
  - **Complexity saving ~40%** from xHE-AAC to MPEG-H per channel
  - **Codec Delay reduced**
  - **Offers comparable perceptual quality** to SBR at stereo 64 kbps

- Modern hybrid bandwidth extension tools enable high perceptual quality at medium and low bit rates and can be scaled seamlessly up to full waveform preserving coding and transparency.

Thank you for your attention!

andreas.niedermeier@iis.fraunhofer.de

www.iis.fraunhofer.de