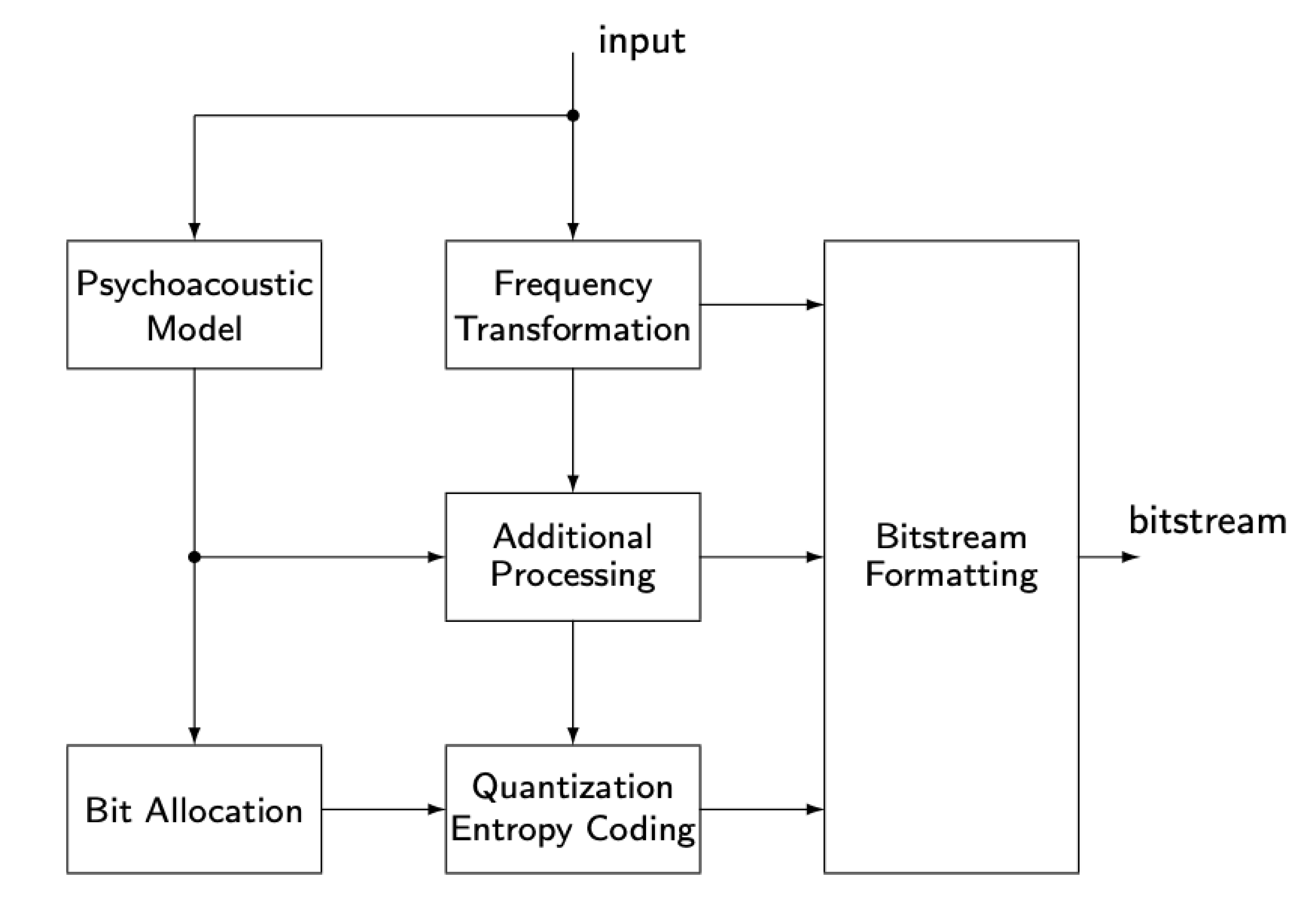
Digital Signal Processing for Music

Part 26: Perceptual Coding

Andrew Beck

Introduction

Overview


input
Psychoacoustic Model
Frequency Transformation
Additional Processing
Bit Allocation
Quantization Entropy Coding
Bitstream Formatting
bitstream

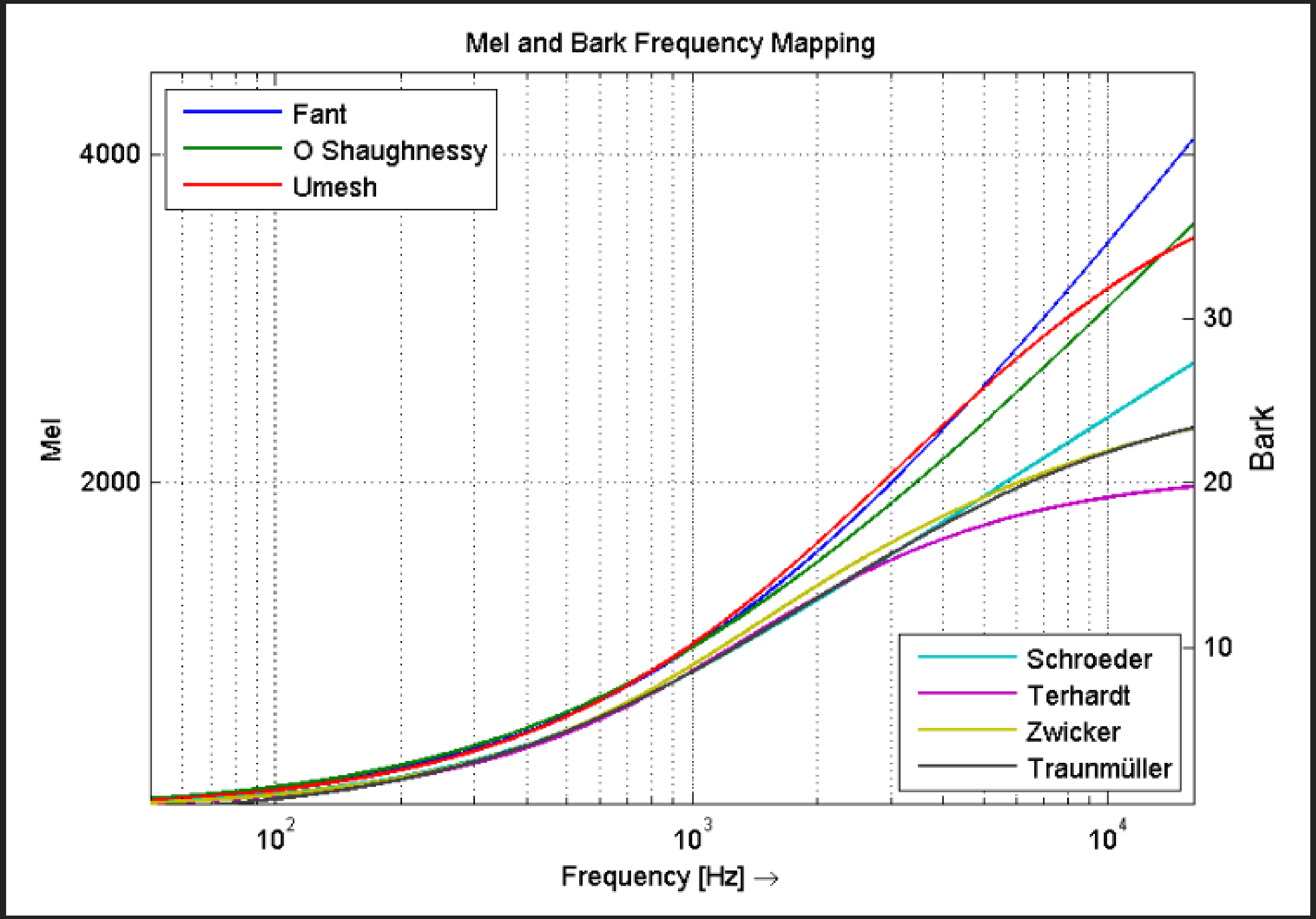
Overview

Psycho-Acoustic Model: Overview

- **Objective**:
  - Identify components perceptible and imperceptible by humans
- **Approach**:
  - Build model of human sound perception (*analysis only!*)
- **Processing Steps**:
  1. Transform to **frequency domain**
  2. Map to **perceptual frequency scale**
  3. Group into bands
  4. Compute (perceptual) **masking threshold**
  5. Compute **Signal-to-Mask Ratio** (SMR)
  6. Compute additional analysis results
- **Recommendation only!** No standardization of implementation

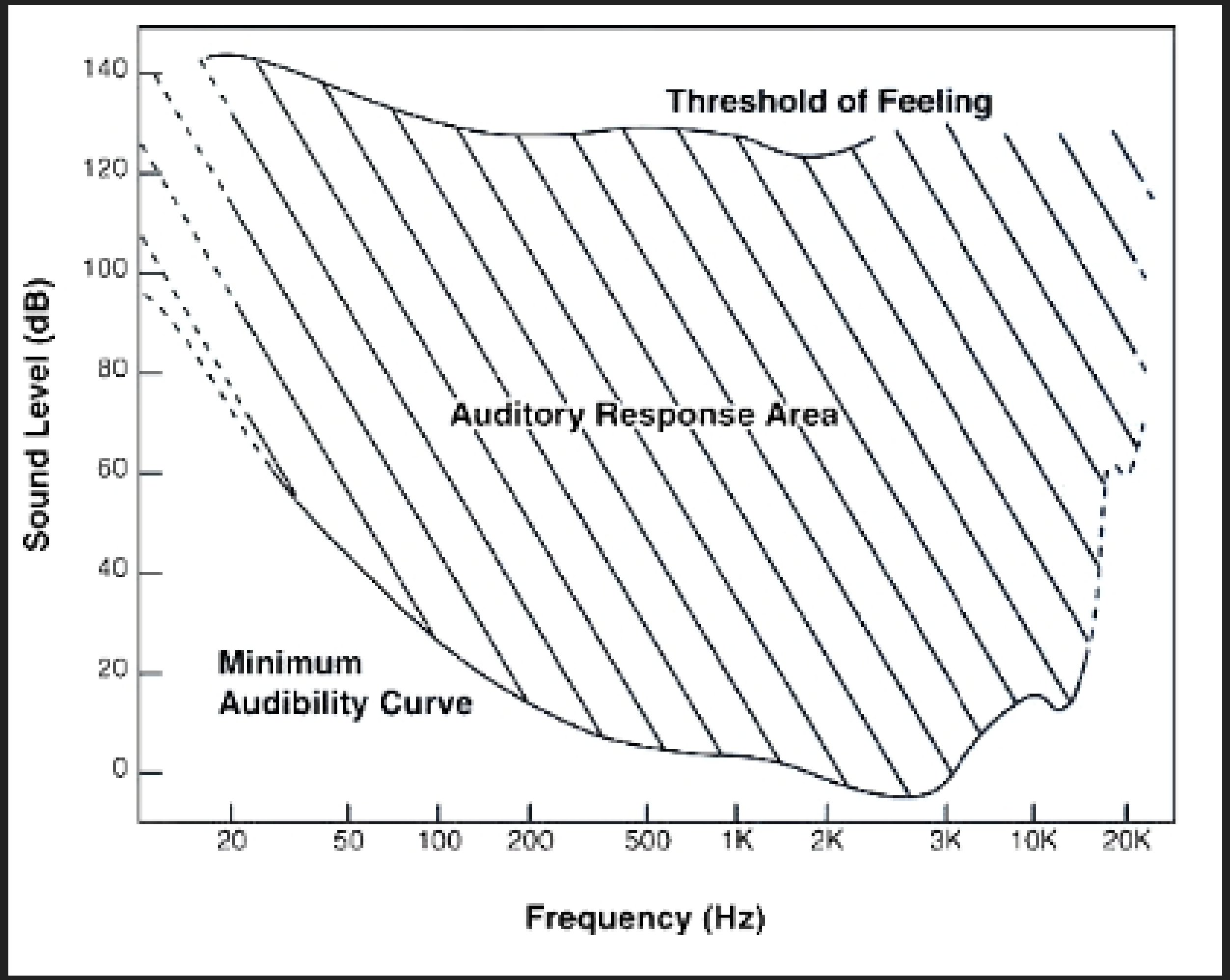
Psycho-Acoustic Model 1-3: Frequency Transform

1. Frequency transformation (AAC: FFT)
2. Frequency Warping (AAC: Bark Scale)
3. Group power in bands (AAC $\frac{1}{3}$ Bark resolution)



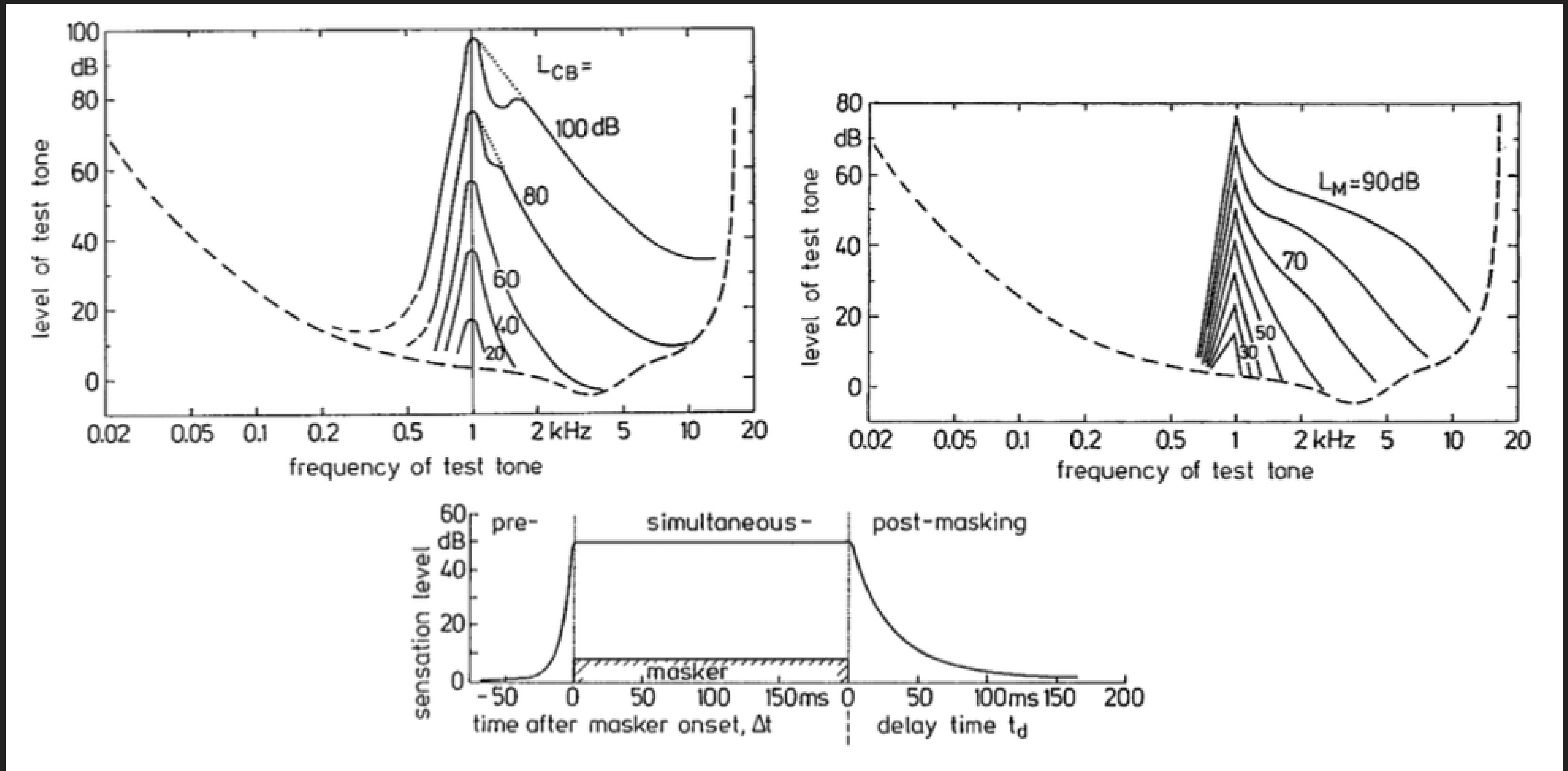
Psycho-Acoustic Model 4: Masking Threshold

- Humans are not able to perceive every possible detail in an audio signal
  - Frequency resolution (see above)
  - Sensitivity for specific frequency regions



Psycho-Acoustic Model 4: Masking Threshold

- Humans are not able to perceive every possible detail in an audio signal
  - Frequency resolution (see above)
  - Sensitivity for specific frequency regions
  - Components masked by other components



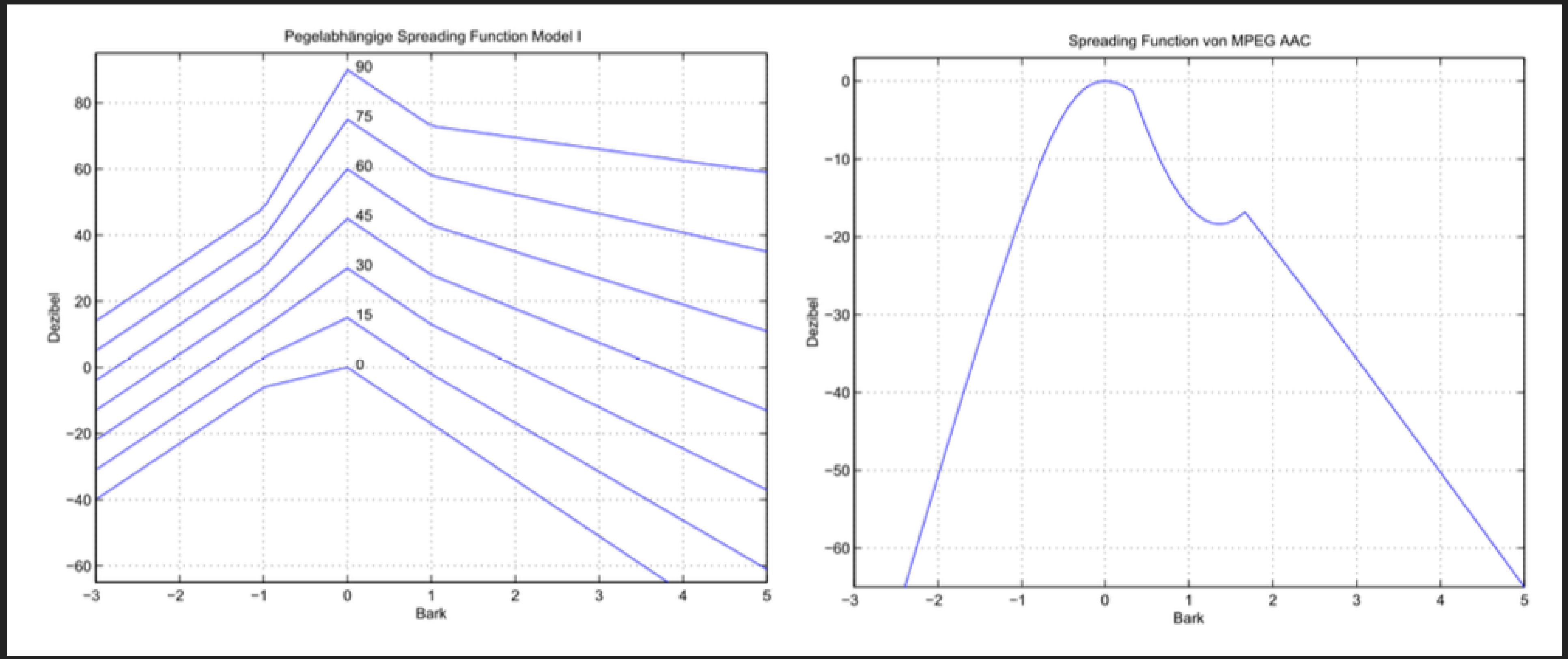
Psycho-Acoustic Model 4: Masking Threshold

- Humans are not able to perceive every possible detail in an audio signal
  - Frequency resolution (see above)
  - Sensitivity for specific frequency regions
  - Components masked by other components
  - Masking threshold depends on
    - *Frequency* of masker
    - *Noisiness* of masker
    - *Level* of masker
    - *Duration* of masker

Psycho-Acoustic Model 4: Masking Threshold

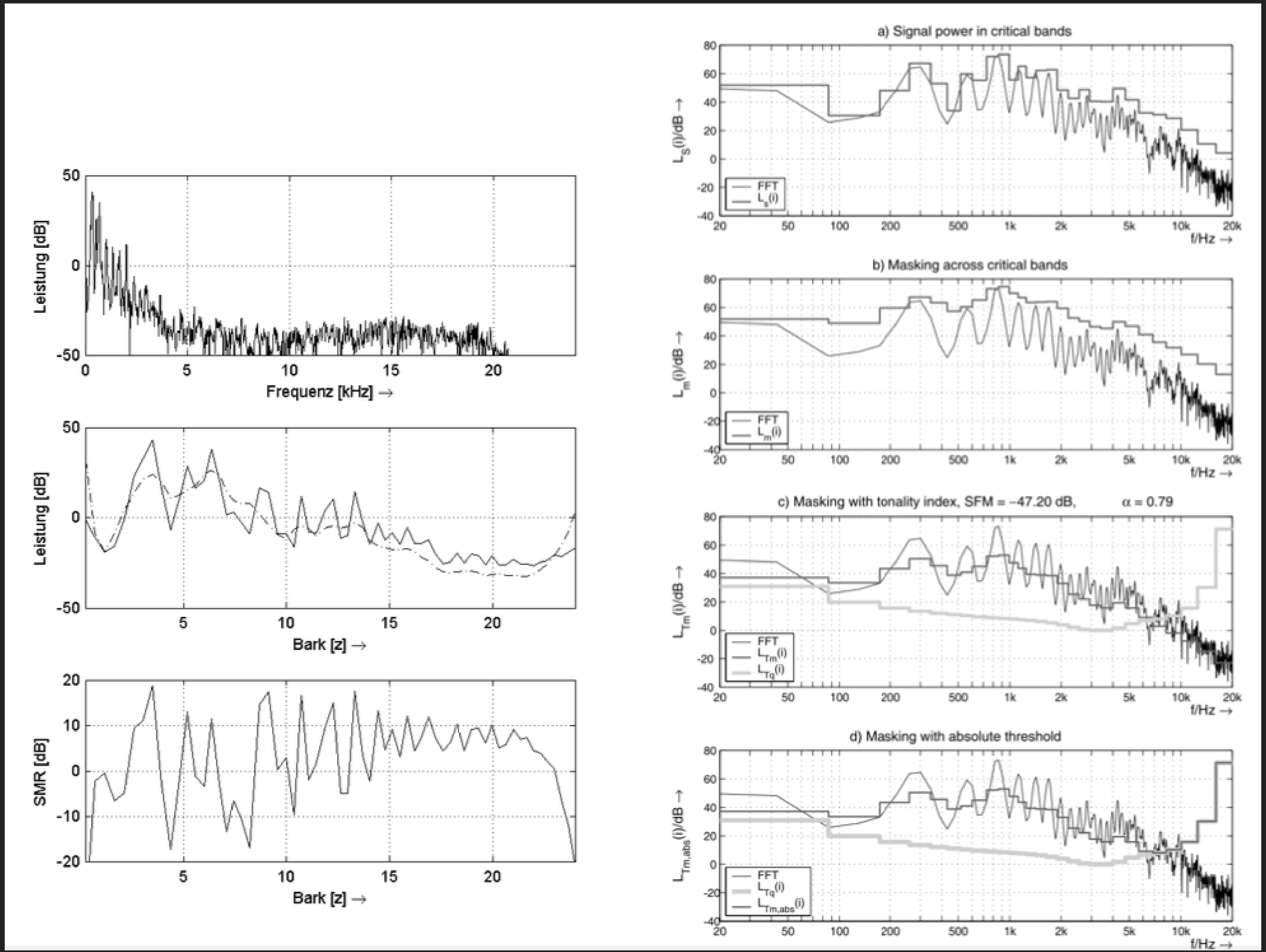
AAC computation of masking threshold (recommendation)

- Take hearing threshold as minimum masking threshold
- Convolve band spectrum with spreading function



- Compute tonality (with phase deviation) and apply to masking threshold (from original spectrum)

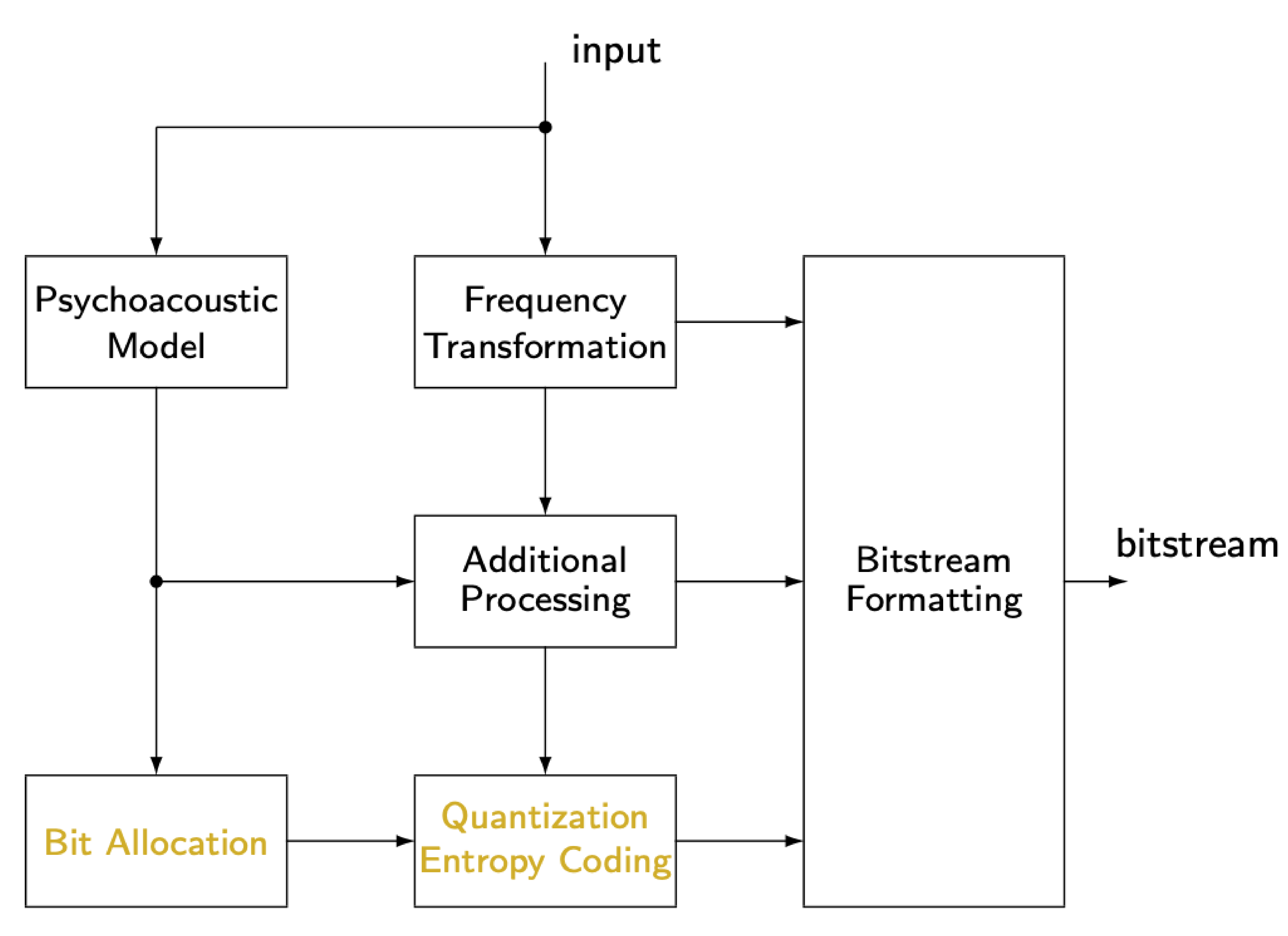
Psycho-Acoustic Model: Visualization



Psycho-Acoustic Model: Additionally extracted information

Control of:

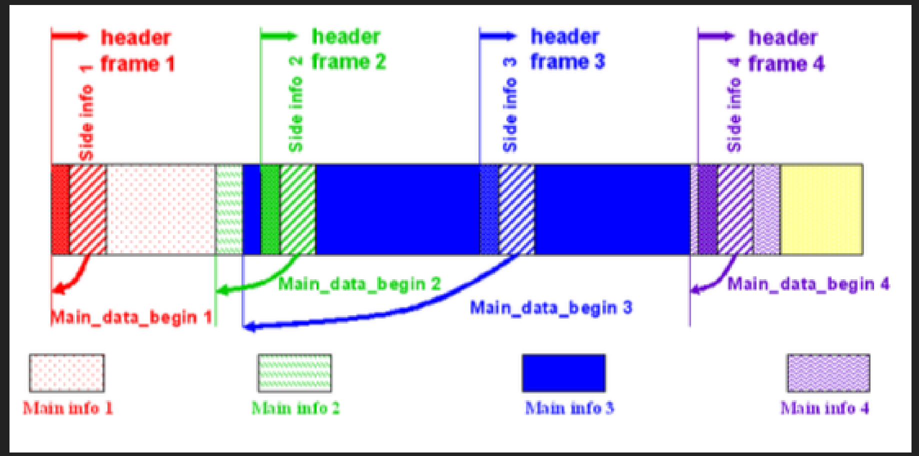
- **Window length switching**
- **Bit reservoir**
- **Joint stereo parameters**


input
Psychoacoustic Model
Frequency Transformation
Additional Processing
Bitstream Formatting
bitstream
Bit Allocation
Quantization
Entropy Coding

Bit Allocation

Bit Allocation

- How many bits are **required** (SMR)?
  Exact output rate is unknown (entropy coding)
- How many bits are **available** per block?
- Are there bits available in the **bit reservor** (~6000*bits*, bit rate dependent)
  - Actual rate must never exceed channel capacity
  - Some frames might need more bits to properly encode
  - Allow deviation from constant bitrate
  - Has to be allocated from previous frames
  - Causes additional decoder delay

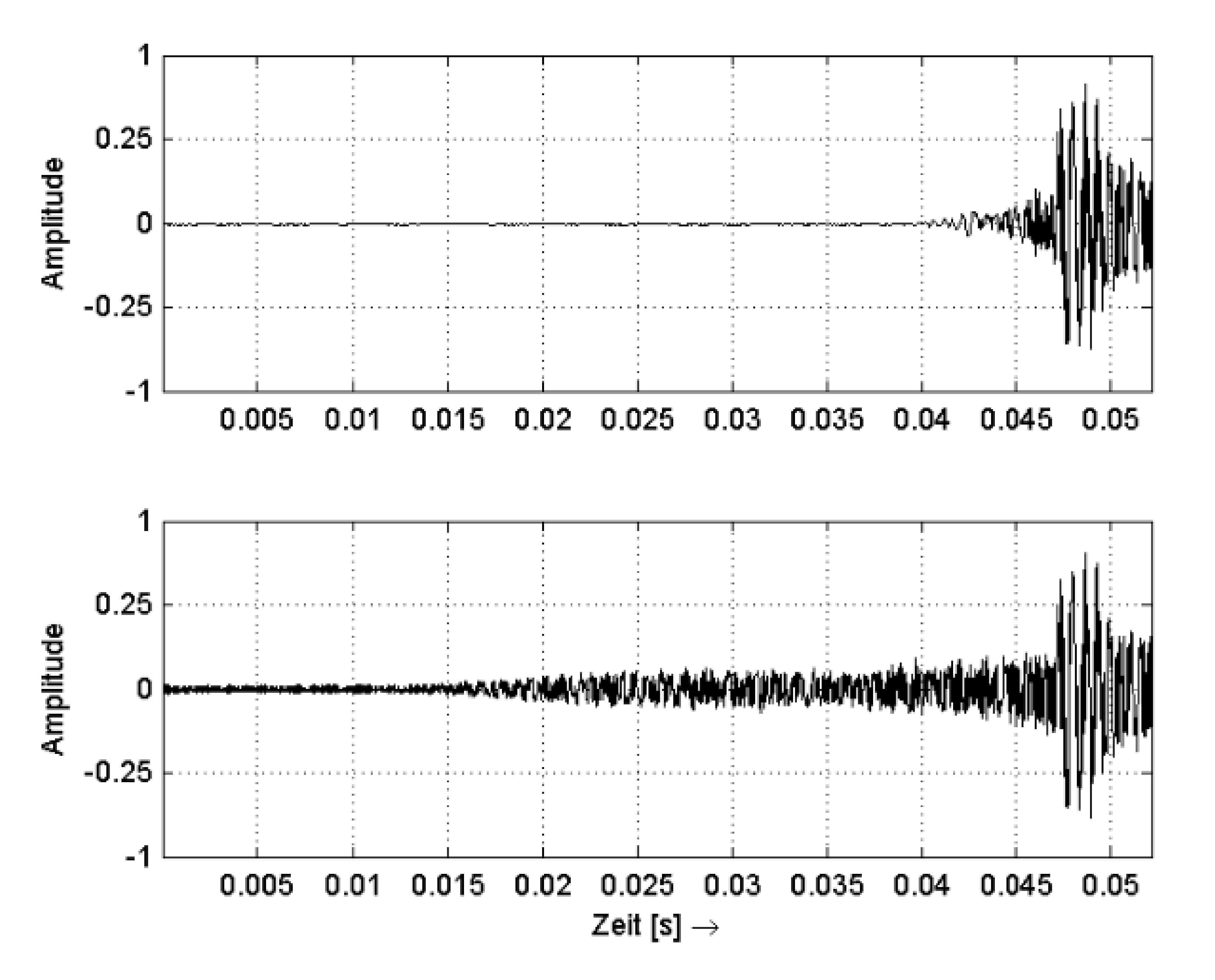


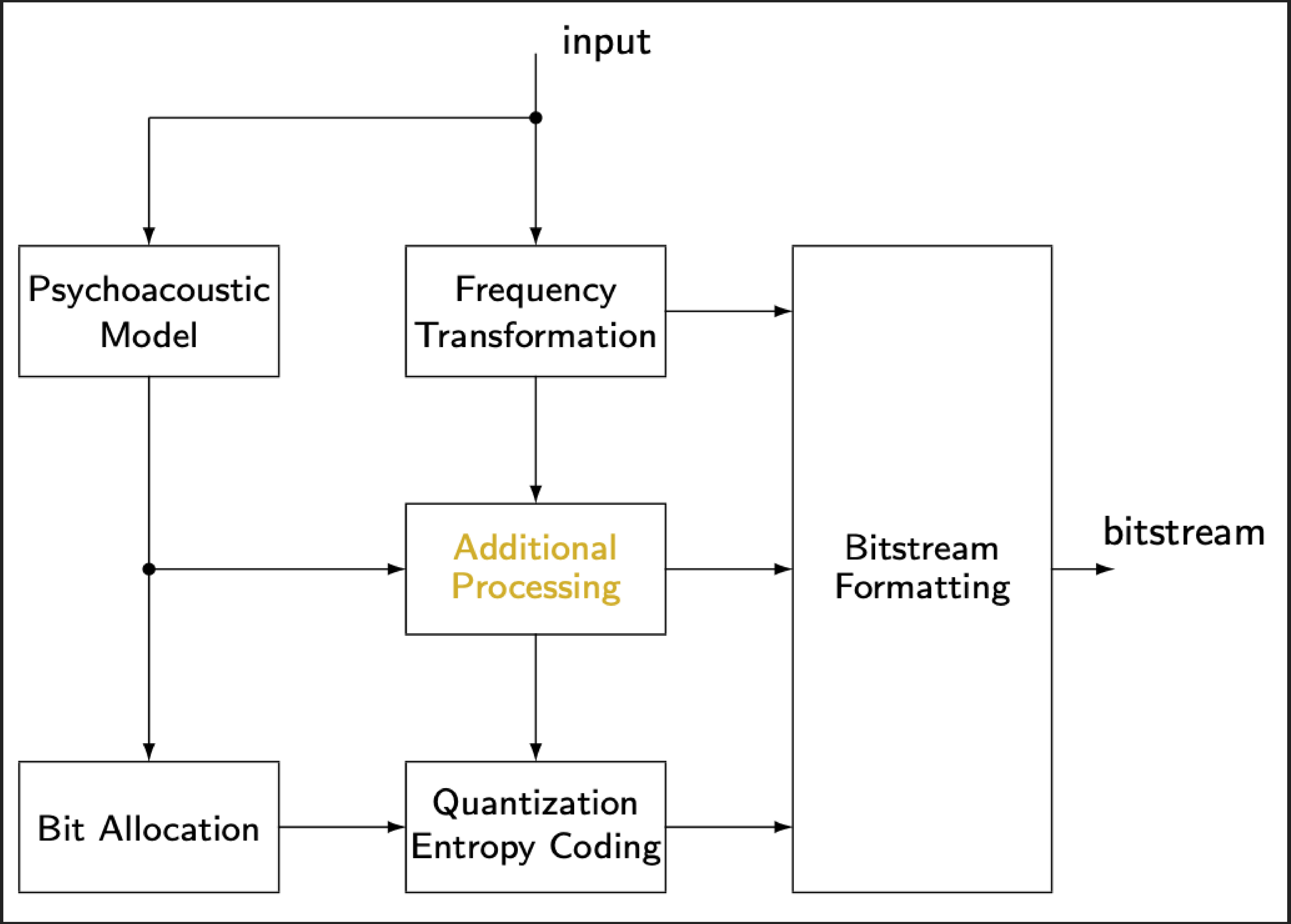
- Intelligently distribute available bits over bands

Quantization & Entropy Coding

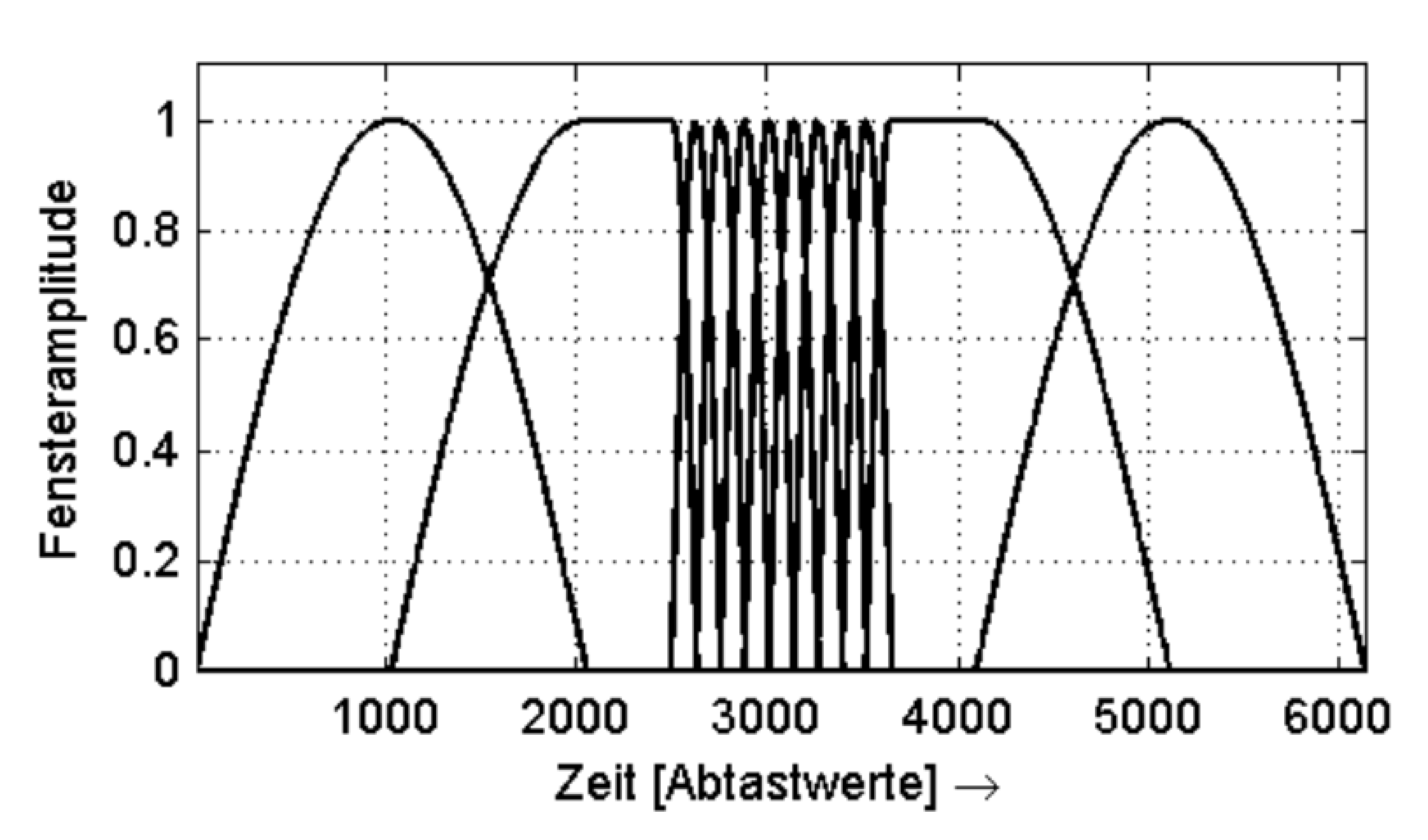
- **Quantization**
  - Re-quantize the spectrum per band
  - Each band has different *scaling factor* and *word length*
  - Non-uniform quantization
- **Entropy Coding**
  - Apply lossless coding (multiple dictionaries)
  - Submit the gained bits to bit allocation (re-iterate?)

Artifacts: Transient Smearing and Pre-Echo




input
Psychoacoustic Model
Frequency Transformation
Additional Processing
Bitstream Formatting
bitstream
Bit Allocation
Quantization Entropy Coding

Tweaks: Block Switching



- AAC: transients are encoded by 8 short frames (256) instead of 1 long frame (2048)
- Introduces additional encoding delay because of different start window shape

Tweaks: Other tools (MPEG-4 AAC, 1st generation)

- **Joint Stereo Coding**:
  - MS (Mid/Side stereo)

    Exploit inter-channel *redundancy* by mid/side encoding
  - IS (Intensity Stereo)

    Remove *irrelevancy* of stereo information: replace stereo by one signal with directional information

    Works for high frequencies (per band)

    May result in spatial distortions

Tweaks: Other tools (MPEG-4 AAC, 1st generation)

- **Prediction**
  - FDP (Frequency Domain Prediction)

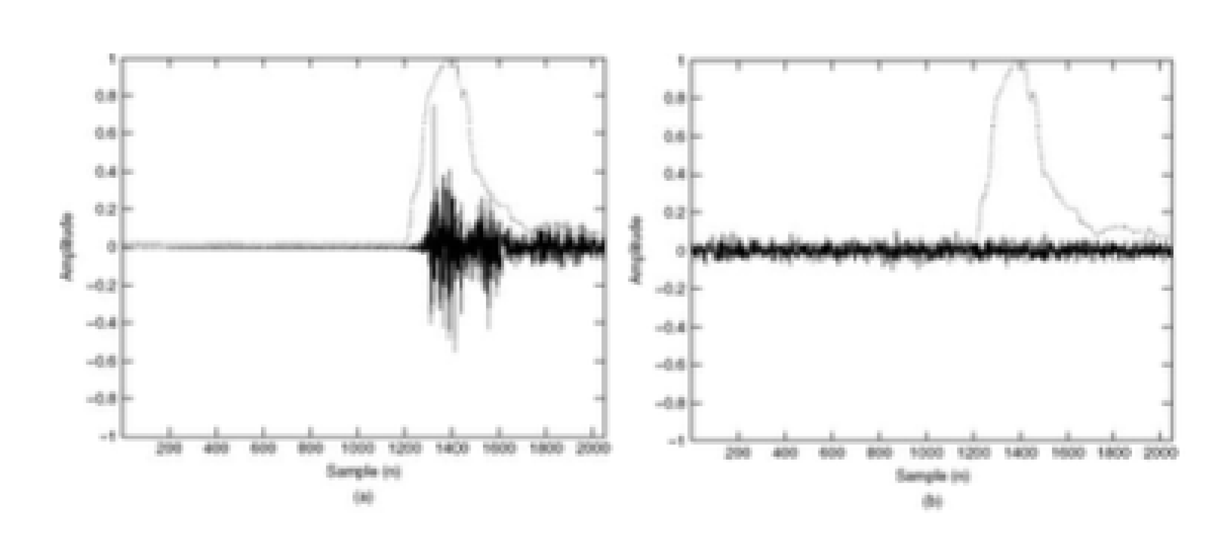
    Backward adaptive per band

    Increases decoder complexity
  - LTP (Long term prediction)

    Time domain predictor, forward adaptive, one coefficient, large lag

Tweaks: Other tools (MPEG-4 AAC, 1st generation)

- **TNS** (Temporal Noise Shaping)
  - Transient artifacts remain problematic
  - D*PCM in the frequency domain → time-domain envelope of the error shaped after signal envelope
  - Shift quantization error power to high amplitude regions



Tweaks: Other tools (MPEG-4 AAC, 2nd generation)

- **PNS** (Perceptual Noise Subsitution)
  - Transmit noise level and inter-channel correlation instead of encoding noise subbands
- **PS** (Parametric Stereo)
  - Extends the IS concept:
  - Encode *one* channel and transmit control info to generate the other channel

Artifacts

- **Transient Smearing**
  Transients are smoothed out
- **Musical Noise** (ringing)
  Switch high frequency bands on and off
- **Stereo Imaging**
  Changing localization and spatial impression
- **Roughness**
  Time-variant granular quantization noise

Audio Examples (MP3)

Harpsichord		Percussion	
Original	0:00 / 0:41	Original	0:00 / 0:40
256 kpbs	0:00 / 0:41	256 kpbs	0:00 / 0:40
128 kpbs	0:00 / 0:41	128 kpbs	0:00 / 0:40
96 kpbs	0:00 / 0:41	96 kpbs	0:00 / 0:40
64 kpbs	0:00 / 0:41	64 kpbs	0:00 / 0:40
32 kpbs	0:00 / 0:41	32 kpbs	0:00 / 0:40

Bitrate Models

- **Constant Bit Rate** (CBR):
  - Bit rate constant over time
  - Quality changes over time
- **Variable Bit Rate** (VBR):
  - Bit rate changes over time
  - Quality constant over time (Depends on psychoacoustic model)

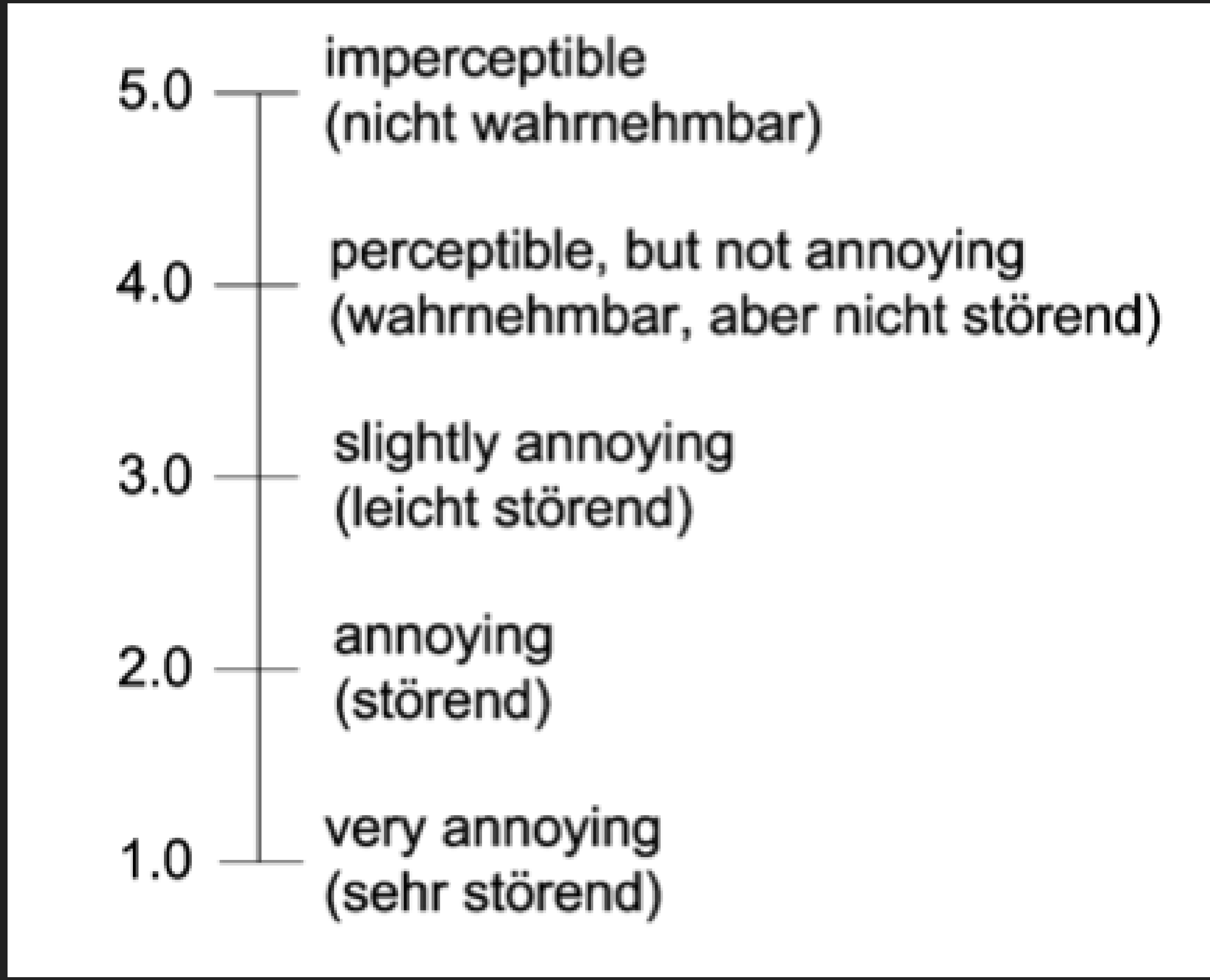
Algorithms & Properties

Name	Sampling Rates	Channels	Bit Rates
MPEG2 Layer 2	16-48k	5.1	8-160
MPEG2 Layer 3	8-96k	5.1	8-320
MPEG4 Layer AAC	16-48k	16	8-320
ATRAC1	44.1k	2	146
ATRAC3	44.1k	2	66,33
SDDS	44.1k	7.1	146
AC-3	32-48k	5.1	32-640
E-AC-3	32-48k	13.1	32-6144
DTS (Cine)	44.1k	5.1 / 6.1	192
DTS (Home)	32-96k	8	8-512

Quality Evaluation

- Quality depends on:
  - Bit Rate
  - General coding algorithm
  - Encoder implementation
  - Encoder options
  - Input signal & its properties
  - Listener
- Objective, technical measures for quality evaluation fail

Blind listening tests with hidden reference



Example Results

approach	SDG (app.)
AAC/128, AC-3/192	-0.5
PAC/160	-0.8
PAC/128, AC-3/160, AAC/96, Layer 2/192	-1.2 ... -1.0
ITIS/192	-1.4
Layer 3/128, Layer 2/160, PAC/96, ITIS/160	-1.8 ... -1.7
AC-3/128, Layer 2/128, ITIS/128	-2.2 ... -2.1
PAC/64	-3.1
ITIS/96	-3.3

Requirements

- Quality (see above)
- Latency (not important for file encoding, but for real-time transmission and real-time systems)
- Complexity (encoder vs decoder)
- Achievable bit rates
- Efficiency (sound quality to bit rate)
- Availability & licensing
- Editability, scrolling capabilities
- Error resilience

Summary

- Perceptual codecs take advantage of properties of human hearing and combine this with principles of redundancy coding
- (MPEG) encoders are only specified by their output stream $\Rightarrow$ different encoders have different quality
- Bitrate/quality tradeoff cannot be completely overcome, however, synthesis-based approaches are more and more successful at low bitrates