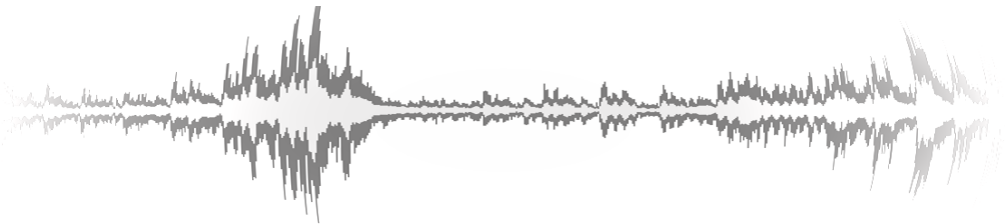
Digital Signal Processing for Music

Part 26: Perceptual Coding

alexander lerch



Georgia Tech Center for Music Technology
College of Design

perceptual coding

introduction

- **goal**:
  - encode signal in a way that the decoded signal is **perceptually** as close to the original signal as possible

- **common codecs**:
  - MPEG-1/2 Layer 2
  - MPEG-1/2 Layer 3
  - MPEG-2/4 AAC
  - AC-3/4
  - Ogg Vorbis
  - (DTS Cine/Home)
  - (ATRAC)

perceptual coding

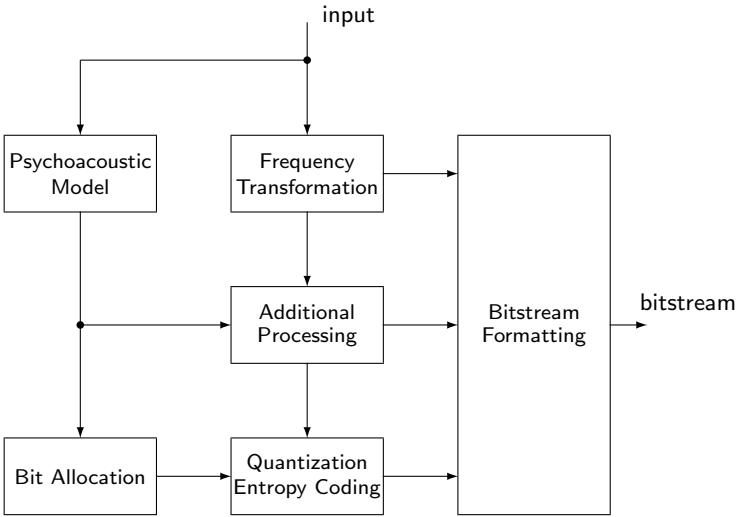
introduction

- **goal**:
  - encode signal in a way that the decoded signal is **perceptually** as close to the original signal as possible

- **common codecs**:
  - MPEG-1/2 Layer 2
  - MPEG-1/2 Layer 3
  - MPEG-2/4 AAC
  - AC-3/4
  - Ogg Vorbis
  - (DTS Cine/Home)
  - (ATRAC)

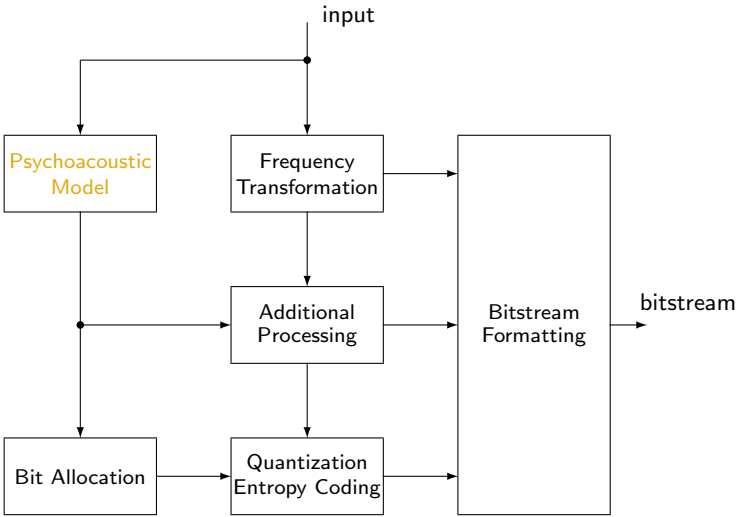
perceptual coding

overview



perceptual coding

overview



source coding

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

source coding

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

source coding

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

source coding

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

source coding

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

source coding

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

source coding

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

source coding

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

source coding

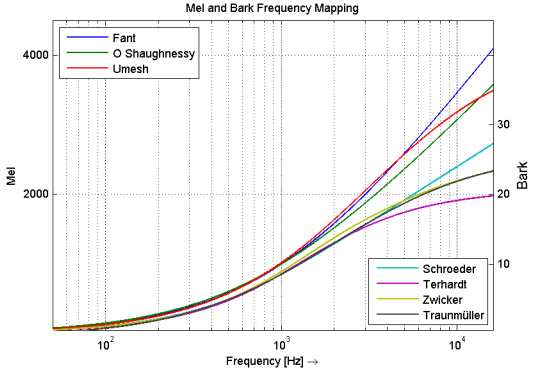
psycho-acoustic model: overview

- **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

source coding

psycho-acoustic model 1–3: frequency transform

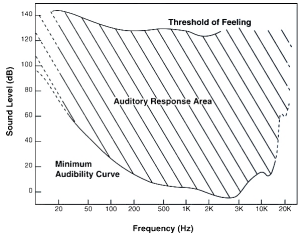
1. frequency transformation (AAC: FFT)
2. frequency warping (AAC: Bark scale)
3. group power in bands (AAC: 1/3 Bark resolution)



source coding

psycho-acoustic model 4: masking threshold 1/2

- humans are not able to perceive every possible detail in an audio signal
  1. frequency resolution (see above)
  2. sensitivity for specific frequency regions

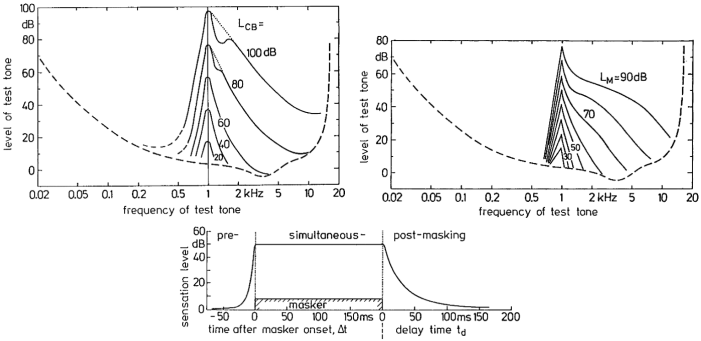


3. components masked by other components
4. masking threshold depends on
   - *frequency* of masker
   - *noisiness* of masker
   - *level* of masker
   - *duration* of masker

source coding

psycho-acoustic model 4: masking threshold 1/2

- humans are not able to perceive every possible detail in an audio signal
  - (i) frequency resolution (see above)
  - (ii) sensitivity for specific frequency regions
  - (iii) components masked by other components



- (iv) masking threshold depends on

source coding

psycho-acoustic model 4: masking threshold 1/2

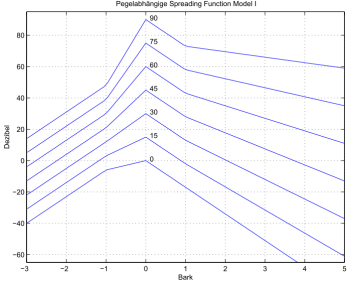
- humans are not able to perceive every possible detail in an audio signal
  - (i) frequency resolution (see above)
  - (ii) sensitivity for specific frequency regions
  - (iii) components masked by other components
  - (iv) masking threshold depends on
    - *frequency* of masker
    - *noisiness* of masker
    - *level* of masker
    - *duration* of masker

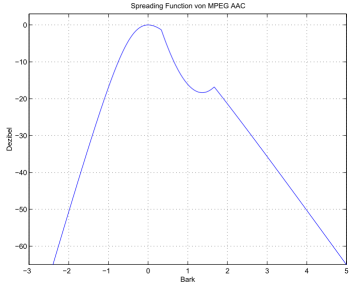
source coding

psycho-acoustic model 4: masking threshold 2/2

AAC computation of masking threshold (recommendation)

- take hearing threshold as minimum masking threshold
- convolve band spectrum with spreading function





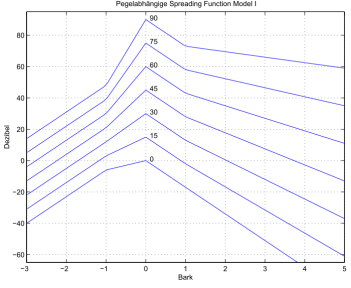
- compute tonality (with phase deviation) and apply to masking threshold (from original spectrum)

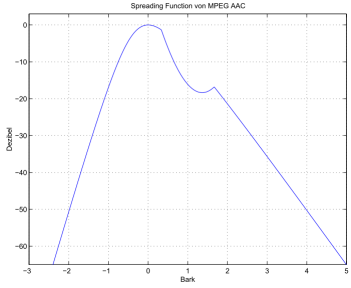
source coding

psycho-acoustic model 4: masking threshold 2/2

AAC computation of masking threshold (recommendation)

- take hearing threshold as minimum masking threshold
- convolve band spectrum with spreading function





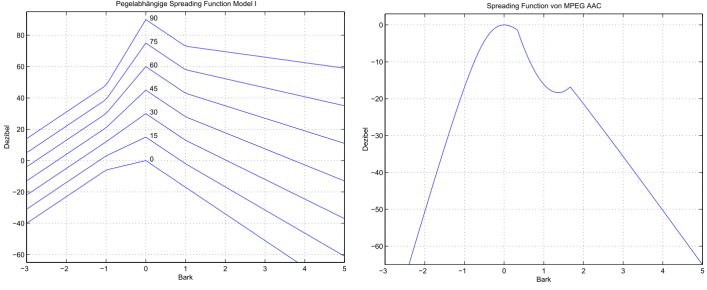
- compute tonality (with phase deviation) and apply to masking threshold (from original spectrum)

source coding

psycho-acoustic model 4: masking threshold 2/2

AAC computation of masking threshold (recommendation)

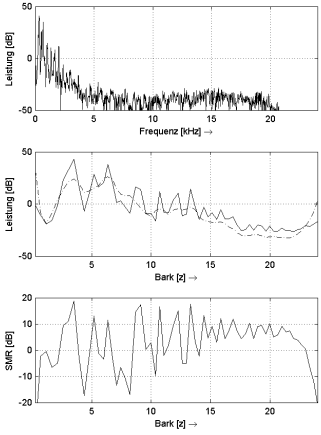
- take hearing threshold as minimum masking threshold
- convolve band spectrum with spreading function

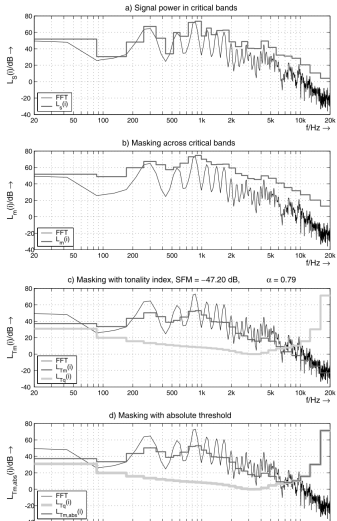


- compute tonality (with phase deviation) and apply to masking threshold (from original spectrum)

source coding

psycho-acoustic model: visualization





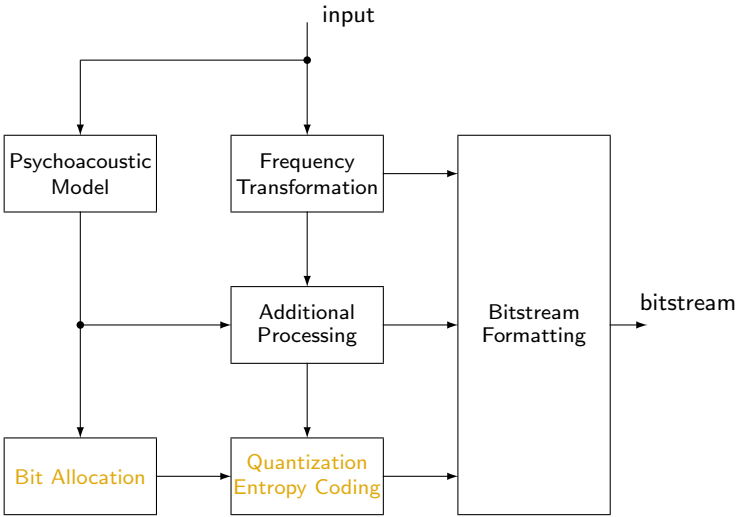
source coding

psycho-acoustic model: additionally extracted information

- control of (see below)
  - **window length switching**
  - **bit reservoir**
  - **joint stereo parameters**

perceptual coding

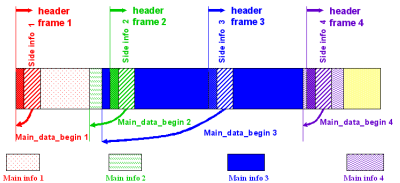
overview



perceptual 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 reservoir** ($\approx 6000 bits$, 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

perceptual coding

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?)

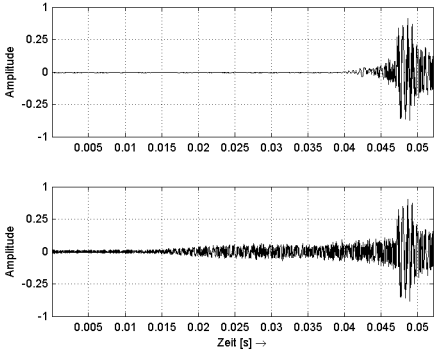
perceptual coding

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?)

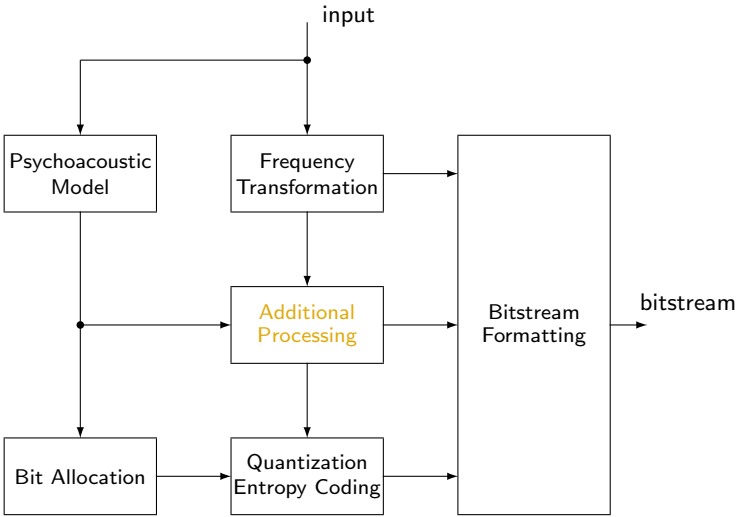
perceptual coding

artifacts: transient smearing and pre-echo



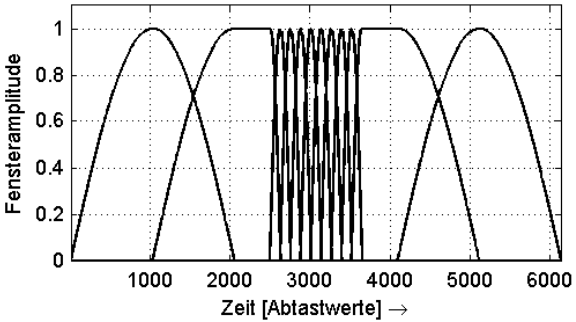
perceptual coding

overview



perceptual 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

perceptual coding

tweaks: other tools (MPEG-4 AAC, 1st generation) 1/2

- **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
- **prediction**
  - FDP (frequency domain prediction)
    - backward adaptive per band
    - increases decoder complexity
  - LTP (long term prediction)
    - time domain prediction, forward adaptive, one coefficient, large lag

perceptual coding

tweaks: other tools (MPEG-4 AAC, 1st generation) 1/2

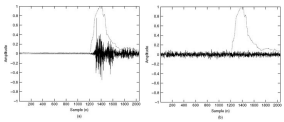
- **joint stereo coding**
  - MS (mid/side stereo)
    - exploit inter-channel *redundancy* by Mid/Side encoding
  - IS (intensity stereo)
    - remove *irrelevancy* of stereo information: replace stereo by one signal with directional information
    - works for high frequencies (per band)
    - may result in spatial distortions
- **prediction**
  - FDP (frequency domain prediction)
    - backward adaptive per band
    - increases decoder complexity
  - LTP (long term prediction)
    - time domain prediction, forward adaptive, one coefficient, large lag

perceptual coding

tweaks: other tools (MPEG-4 AAC, 1st generation) 2/2

- **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!



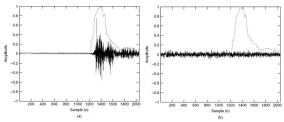
- **PNS** (perceptual noise substitution)
  - transmit noise level and inter-channel correlation instead of encoding noise subbands

perceptual coding

tweaks: other tools (MPEG-4 AAC, 1st generation) 2/2

- **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!



- **PNS** (perceptual noise substitution)
  - transmit noise level and inter-channel correlation instead of encoding noise subbands

perceptual coding

tweaks: other tools (MPEG-4 AAC, 2nd generation) 1/2

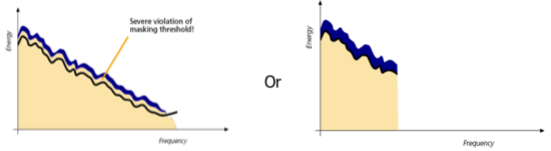
- **PS** (parametric stereo)
  - extends the IS concept:
    encode *one* channel and transmit control info to generate the other channel

perceptual coding

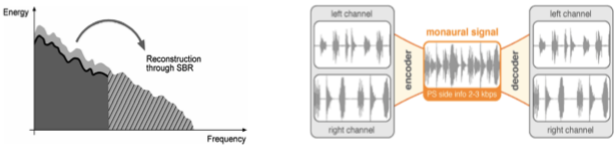
tweaks: other tools (MPEG-4 AAC, 2nd generation) 2/2

- **SBR** (spectral band replication)
  - encode *band limited* signal and transmit control info to generate high frequency content

Limitation of conventional generic audio coding (if < 32kbps/ch)



Bandwidth and Channel extended by (bitrate efficient) “Parametric represe ntation”



perceptual coding

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

perceptual coding

audio examples (MP3)

Harspichord

- original
- 256 kbps
- 128 kbps
- 96 kbps
- 64 kbps
- 32 kbps

Percussion

- original
- 256 kbps
- 128 kbps
- 96 kbps
- 64 kbps
- 32 kbps

perceptual coding

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)

perceptual coding

algorithms & properties

Name	Sampling Rates	Channels	Bit Rates
MPEG2 Layer 2	16-48k	5.1	8-160
MPEG2 Layer 3	8-96k	5.1	8-160
MPEG4 AAC	16-48k	48.16	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

perceptual coding

quality evaluation 1/2

- quality depends on:
  - bit rate
  - general coding algorithm
  - encoder implementation!
  - encoder options
  - input signal & its properties
  - listener

$\Rightarrow$ objective, technical measures for quality evaluation fail

perceptual coding

quality evaluation 1/2

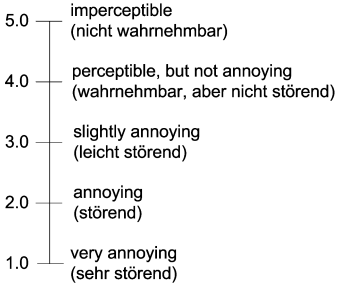
- quality depends on:
  - bit rate
  - general coding algorithm
  - encoder implementation!
  - encoder options
  - input signal & its properties
  - listener

$\Rightarrow$ objective, technical measures for quality evaluation fail

perceptual coding

quality evaluation 2/2

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

perceptual coding

quality evaluation 2/2

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

perceptual coding

requirements

- quality (see above)
- latency
  - not important for file encoding, but for real-time transmission
- complexity
  - encoder vs. decoder
- achievable bit rates
- efficiency
  - sound quality to bit rate
- availability and license
- editability, scrolling capabilities
- error resilience

perceptual coding

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