

## How to Compare?

Part II Compression

- ❑ performance tradeoffs
  - ❑ encoding/decoding time and complexity
  - ❑ image size and frames per second
  - ❑ compression factor
  - ❑ image quality
- ❑ functionality tradeoffs
  - ❑ multiple resolutions
  - ❑ constant vs. variable bit rate
  - ❑ ease of editing

## MPEG System Standard

- ❑ stream specification—interleaving of audio and video, packetizing, timing
- ❑ MPEG-1 program streams—variable length packets, 1-2kB up to 64 kB
- ❑ MPEG-2 transport streams—fixed length packets, 188bytes/ packet=> 4 ATM packets per transport stream packet
- ❑ timing/synchronization specified in streams (27 MHz virtual clock)

## MPEG Performance

- ❑ decoding is relatively easy
  - ❑ real-time sw decoding possible on current platforms
- ❑ encoding is expensive
  - ❑ sw encoders on current workstations are still too slow for good quality and high compression factors

## MPEG-2 Video Standard Overview

- ❑ concept similar to MPEG-1, but includes extensions to cover a wider range of applications
  - ❑ all digital transmission of broadcast TV quality video at coded bit rates between 4 and 9 Mbps
  - ❑ efficient for HDTV bit rates
- ❑ enhancements
  - ❑ addition of syntax for efficient coding of interlaced video (e.g., 16x8 block size motion compensation)
  - ❑ 10-bit DCT DC coefficients (MPEG-1 8 bits)
  - ❑ scalable extensions which permit the division of a continuous video signal into 2 or more coded bit streams representing the video at different resolutions, picture quality or picture rates

## MPEG-2 Video Standard Overview

- ❑ MPEG-2 addresses applications using CCIR-601 recommendation: more than 15 million samples/sec
- ❑ MPEG-2 add to MPEG-1
  - ❑ more aspect ratios
  - ❑ 4:2:2, 4:4:4 macroblocks
  - ❑ progressive and interlaced frame coding
  - ❑ additional prediction modes (16x8 MC, field MC)
  - ❑ four scalable modes
  - ❑ improved picture quality (quantization, zig-zag scan)

## Scalable Bit Streams

four scalable modes—break MPEG-2 video into different layers mostly for purpose of prioritizing video data:

- ❑ spatial scalability—codes a base layer at lower sampling dimensions (resolutions) than the upper layers
- ❑ data partitioning—breaks the block of 64 quantized coefficients into 2 bit streams. The first, higher priority bit stream contains the more critical lower frequency coefficients and side information (DC values, motion vectors)

## MPEG-2 Video Standard

MPEG soon realized:

- ❑ no reason to restrict maximum coded bit rate to 10 Mbits/s; MPEG could support higher bit rates (80 to 100 Mbit/s)
- ❑ impossible to define a single standard satisfying all requirements
- ❑ most applications would only use a small subset of features offered. Hence, MPEG decided to adopt a toolkit-like approach: MPEG-2 is a collection of tools defined in such a way as to satisfy the requirements of specific major applications.

## Scalable Bit Streams

- ❑ SNR scalability—channels are coded at identical sample rates, but with different picture quality (through quantization)
- ❑ temporal scalability—higher priority bit stream codes video at a lower frame rate, intermediate frames can be coded in a second bit stream using the first bit stream reconstruction as prediction

## Profiles and Levels

- ❑ the range of coding support is divided into *profiles* and *levels*.  
For each profile/level, MPEG-2 provides the syntax for the coded bit stream and the decoding requirements.
- ❑ A profile is a defined subset of the entire bit stream syntax.  
Within a profile, a level is defined as a set of constraints imposed on the parameters of the bit stream (e.g., resolution, max. bit rate)
- ❑ profiles: simple, main, 4:2:2, SNR, spatial, high, multiview
- ❑ levels: low (SIF), main (CCIR 601), high-1440, high (HDTV)

## Profiles and Levels

- ❑ tables show upper bounds for picture resolution, frame rate and bit rates. Bitrate data refer to the max. compressed bit rate supported by the input buffers of a decoder. Areas with N/A indicate no conformance restrictions for these variables.
- ❑ early implementations only support the main profile at the main level. The simple profile is a low-cost version of the main profile (without bidirectional prediction). Levels are related to resolution: main level (CCIR 601), high-1440 and high level (HDTV), low level (SIF).

## Profiles and Levels, Nonscalable Modes

Levels		Profiles (nonscalable)		
		simple 4:2:0 (I, P)	main 4:2:0 (I, B, P)	4:2:2 (I), (I, B, P)
High	res. / rate	N/A	1920 x 1152 / 60	N/A
	Mbits/s	N/A	80	N/A
High-1440	res. / rate	N/A	1140 x 1152 / 60	N/A
	Mbits/s	N/A	60	N/A
Main	res. / rate	720 x 576 / 30	720 x 576 / 30	720 x 608 / 30
	Mbits/s	15	15	50
Low	res. / rate	N/A	352 x 288 / 30	N/A
	Mbits/s	N/A	4	N/A

## Profile and Levels

- ❑ profiles and levels have a hierarchical relationship. The syntax supported by a higher profile/level includes all the syntactical elements of lower profile/levels. Decoders for a specific profile/level should be able to decode also bit streams of lower profiles/levels.
- ❑ Exception:  
Decoders of simple profile at main level are also required to decode main profile at low level bit streams (MPEG-1).

## MPEG-2 4:2:2 Profile

### Features:

- ❑ DC precision of intracoded blocks can be 8, 9, 10 or 11 bits.
- ❑ Number of bits in a macroblock is unconstrained.
- ❑ Excellent multigeneration performance (up to 8 generations).
- ❑ GOP selection is flexible and permits all encoded pictures to be I-pictures.
- ❑ For the same bit rate across several generations, MPEG 4:2:2 provides better image quality than Motion-JPEG.

## Profiles and Levels, Scalable Modes

Levels		Profiles (scalable)			
		SNR 4:2:0	spatial 4:2:0	high 4:2:0, 4:2:2	multiview 4:2:2
Main	enhancem. (auxiliary)	720 x 576 / 30	N/A	720 x 576 / 30	720 x 576 / 30
	lower (base)	-	N/A	352 x 288 / 30	720 x 576 / 30
	Mbits/s	15 (all layers) 10 (base layer)	N/A	20 (all layers) 15 (base + mid) 4 (base layer)	25 (all layers) 10 (auxiliary) 15 (base layer)
Low	enhancem. (auxiliary)	352 x 288 / 30	N/A	N/A	352 x 288 / 30
	lower (base)	-	N/A	N/A	352 x 288 / 30
	Mbits/s	4 (all layers) 3 (base layer)	N/A	N/A	8 (all layers) 4 (auxiliary) 4 (base layer)

## Profiles and Levels, Scalable Modes

Levels		Profiles (scalable)			
		SNR 4:2:0	spatial 4:2:0	high 4:2:0, 4:2:2	multiview 4:2:2
High	enhancem. (auxiliary)	N/A	N/A	1920 x 1152 / 60	1920 x 1152 / 60
	lower (base)	N/A	N/A	960 x 576 / 30	1920 x 1152 / 60
	Mbits/s	N/A	N/A	100 (all layers) 80 (base + mid) 25 (base layer)	130 (all layers) 50 (auxiliary) 80 (base layer)
High-1440	enhancem. (auxiliary)	N/A	1440 x 1152 / 60	1440 x 1152 / 60	1920 x 1152 / 60
	lower (base)	N/A	720 x 576 / 30	720 x 576 / 30	1920 x 1152 / 60
	Mbits/s	N/A	60 (all layers) 40 (base + mid) 15 (base layer)	80 (all layers) 60 (base + mid) 20 (base layer)	100 (all layers) 40 (auxiliary) 60 (base layer)

## MPEG Audio

- ❑ exponential growth of MPEG coded audio material on the internet since 1995
- ❑ “mp3” the most searched for term in early 1999
- ❑ everybody is using MP3, not many know details of MPEG audio coding
- ❑ *MPEG-1/2 Layer-3 is called MP3*

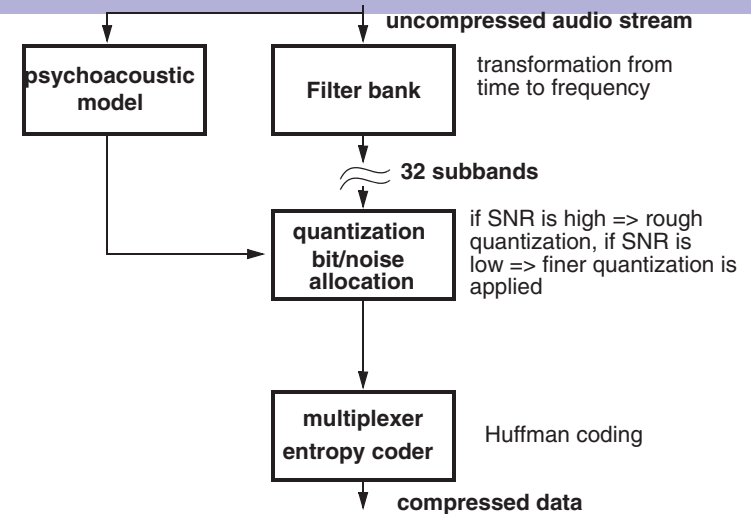
## Why MPEG-1/2 Layer-3?

- ❑ open standard
  - ❑ specification available for a fee to everybody
  - ❑ patents licensed on fair and reasonable terms
  - ❑ no single company “owns” the standard
  - ❑ public example source code available to implementers
  - ❑ format well defined
- ❑ availability of encoders and decoders
- ❑ supporting technologies—widespread use of sound cards; computers fast enough for audio decoding and even encoding, fast internet access; the spread of CD-ROM and DVD writers

## High Quality Audio Coding

- ❑ MPEG-1 Layer-3 has been defined in 1991
- ❑ since then, research on perceptual audio coding has progressed and codes with better compression efficiency became available: MPEG-2 Advanced Audio Coding (AAC) and other proprietary compression systems.
- ❑ basic task is to compress audio data in a way that
  - ❑ the compression is as efficient as possible (file size small)
  - ❑ the reconstructed (decoded) audios sounds exactly (or as close as possible) to the original audio data
  - ❑ requires low complexity (sw or inexpensive hw)
  - ❑ offers flexibility for different application scenarios.

## MPEG Audio Standard (Encoding)

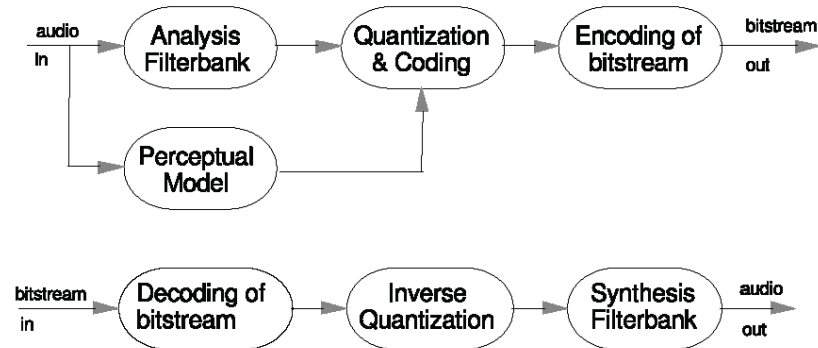


## A Basic Perceptual Audio Encoder

consists of the following building blocks:

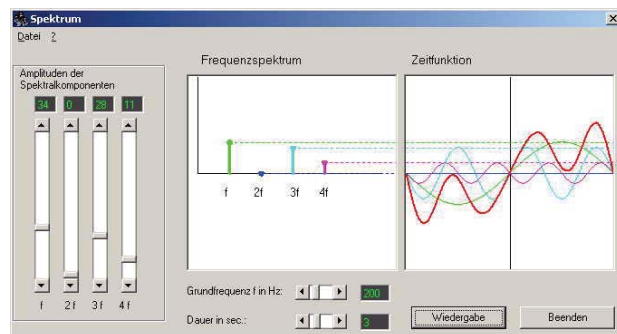
- ❑ **filter bank**—is used to decompose the input signal into subsampled spectral components (time/frequency domain)
- ❑ **perceptual model**—using either the domain input signal and/or the output of the analysis filter bank, an estimate of the actual masking threshold is computed by rules from psychoacoustics
- ❑ **quantization and coding**—the spectral components are quantized and coded; noise introduced should be kept below threshold
- ❑ **encoding of bitstream**

## A Basic Perceptual Audio Coder



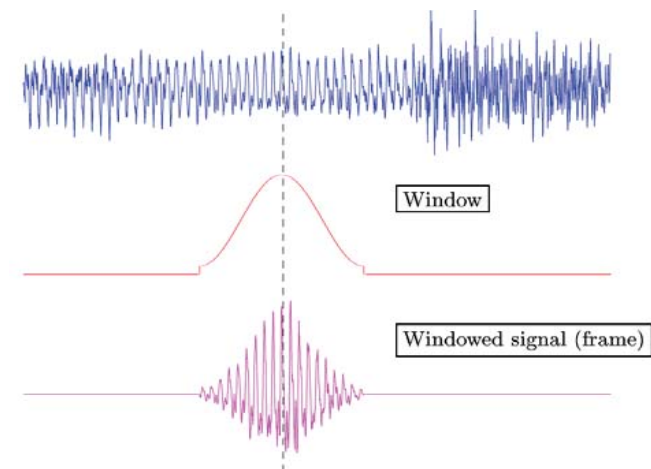
## Einschub Spektrogramm

## Signaldarstellung



- ☐ Frequenzdomäne (Spektrum, links)
- ☐ Zeitdomäne (rechts)

## Frames



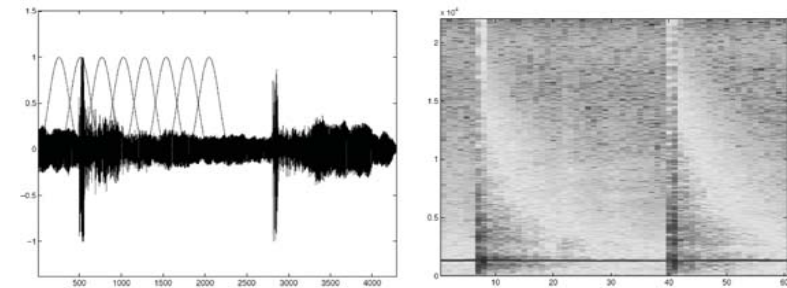
## Spektrogramm

- ❑ einfache Darstellungen haben Grenzen:
  - ❑ Zeit-Domäne zeigt Frequenz-Anteile eines Signals nicht
  - ❑ Frequenz-Domäne zeigt nicht, wann Frequenzen auftreten
- ❑ kombinierte Darstellung - Spektrogramm
  - ❑ x-Achse: Zeit, y-Achse: Frequenzanteile
  - ❑ Schwärzung (Farbe) eines Punkts: Energie der Frequenz zu dieser Zeit
- ❑ Analysen
  - z.B. Regelmäßigkeit des Auftretens von Frequenzen, Musik vs. Geräusch

## Vorgangsweise

- ❑ Eingangssignal wird blockweise verarbeitet
- ❑ überlappende Segmente des Signals werden verwendet
- ❑ sinusoide Fensterfunktionen in Abb. deuten Signalausschnitte an, auf die sich die Analyse in einem Schritt „konzentriert“
- ❑ punktweise Multiplikation von Signalblock mit Fensterfunktion
- ❑ Fenster können verschiedene Formen haben (z. B. Hamming, Bartlett, Dreieck, Rechteck)
- ❑ resultierendes Signal wird Fouriertransformiert
- ❑ Aneinanderreihung der Spektralvektoren liefert Zeit-Frequenzdarstellung des Signals

## Zeit-Frequenzdarstellung



## Short Time FT (STFT)

- ❑ Fenster (Frame, Windowed Signal):

$$x'_t(\tau) = x(\tau)w(\tau - t)$$

- ❑ Fenster  $w$ , Zeit  $t$ ,  $t$  Fensterposition
- ❑ Fourier Transformation von  $x'_t$  ist STFT von  $x$  zur Zeit  $t$  mit Fenster  $w$

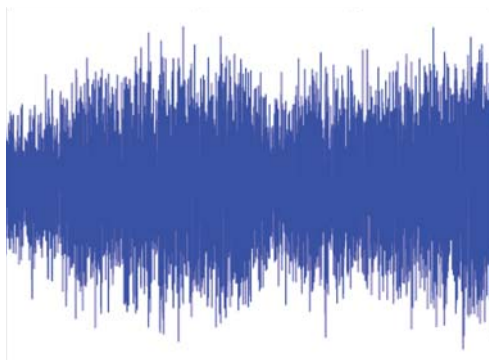
## Spektrogramm

- Spektrogramm von  $x$  ist das Quadrat der STFT:

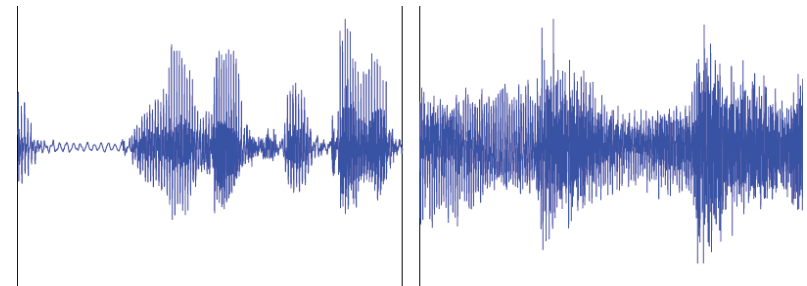
$$\begin{aligned}\text{Spect}_x(t, f) &= \left| \int x(\tau) w(\tau - t) e^{-j2\pi f\tau} d\tau \right|^2 \\ &= |\text{STFT}_x(t, f)|^2\end{aligned}$$

- häufig statistische Interpretation als nicht normalisierte Dichtefunktion (pdf) über die Frequenz
- erlaubt Berechnung statistischer Parameter, z.B. Lagemaße, Streuung

## Beispiel Tierlaut

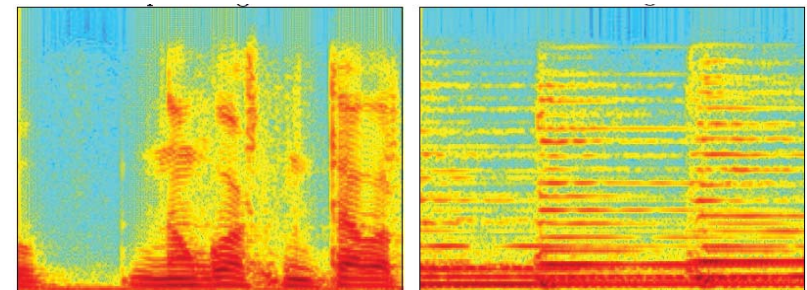


## Beispiele: Sprache / Musik



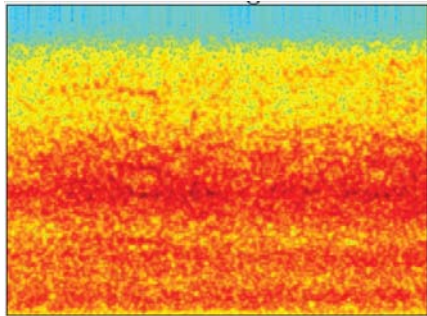
## Beispiel-Spektrogramme

- Sprache / Musik (Zeit 0-1.14 s, Frequenz 0-5512 Hz)

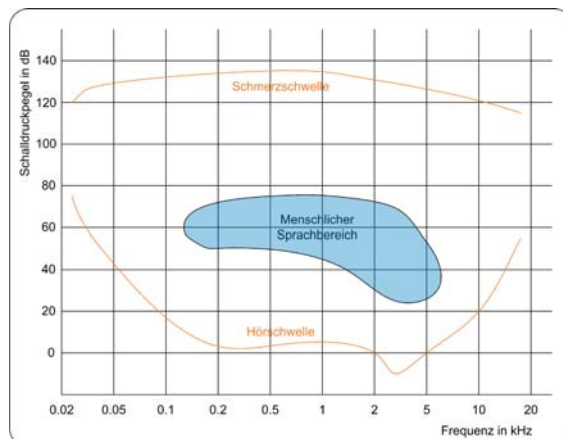


## Beispiel-Spektrogramme

### ❑ Tiergeräusch



## Hörbereich



# BG: Grundlagen der MPEG Audiokodierung

## Hörbereich

- ❑ Schmerzschwelle und Sprachbereich sind für MPEG Audiokodierung nicht relevant
- ❑ Hörschwelle – essentiell für MPEG Audiokodierung
  - ❑ Ton muss gewissen Mindestschallpegel aufweisen, um wahrgenommen zu werden
  - ❑ Mindestschallpegel ist frequenzabhängig
  - ❑ Hörschwellenkurve: Funktion Mindestschallpegel/Frequenz
  - ❑ eingezeichnete Hörschwelle ist Ruhehörschwelle

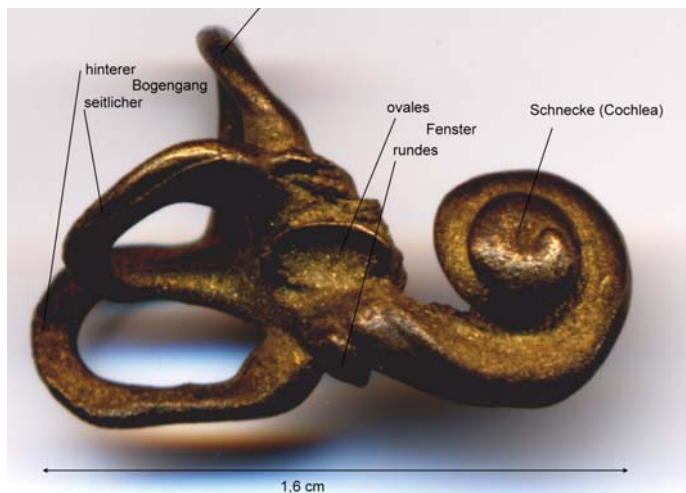
## Ruhehörschwelle

- ❑ Messung der Ruhehörschwellenkurve
  - ❑ Einzeltöne unterschiedlicher Frequenz und Intensität werden hintereinander abgespielt
  - ❑ Testperson gibt an, wann ein Ton gerade noch wahrnehmbar ist
  - ❑ daraus ergibt sich Diagramm Mindestschallpegel/Frequenz
  - ❑ Ruhehörschwellenkurve – außer gespieltem Einzelton vollkommene Ruhe
- ❑ sind während Messung zusätzliche Töne oder Geräusche zu hören, ergibt sich eine andere Hörschwellenkurve

233

Christian Breiteneder

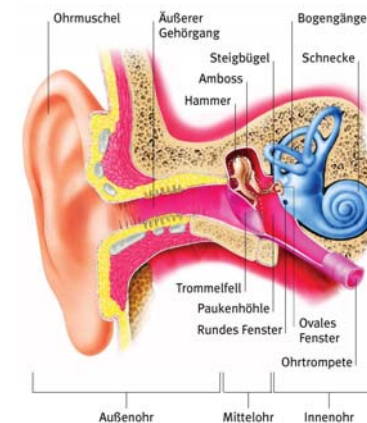
## Innenohr



235

Christian Breiteneder

## Das menschliche Ohr



234

Christian Breiteneder

## Das Innenohr

- ❑ komplex gestalteter Hohlraum im Felsenbein
- ❑ Hohlraum ist mit Flüssigkeit gefüllt (Perilymphe), darin dünnwandiges häutiges Labyrinth mit Endolymphe
- ❑ 2 Öffnungen zum Mittelohr: ovales und rundes Fenster,
  - ❑ Ovales F.: Übertragung der Schallwellen
  - ❑ Rundes F.: Dämpfung der Schwingungen in der Cochlea

236

Christian Breiteneder

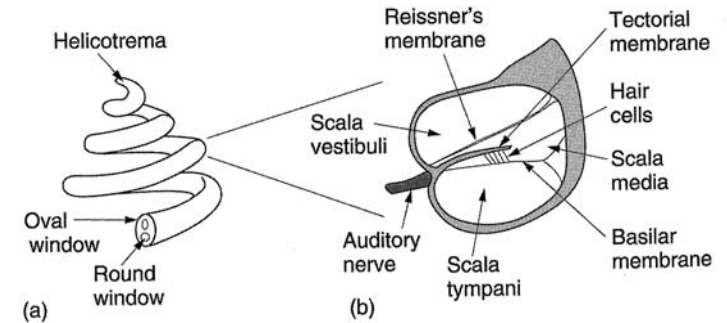
## Hören

- ☐ Schallwellen treffen auf Trommelfell
- ☐ Trommelfell überträgt Schwingung auf die Gehörknöchelchenkette im Mittelohr;
- ☐ als deren letztes Glied wird Fußplatte des Steigbügels und
- ☐ das Ovale Fenster bewegt.
- ☐ Dadurch Schwingung der Endolymphe in Vorhoftreppe

## Hörschnecke

- ☐ 3 übereinanderliegende flüssigkeitsgefüllte Gänge:
  - ☐ *Scala vestibuli* (Vorhoftreppe)
  - ☐ *Scala media* oder Ductus cochlearis (Schneckengang) und
  - ☐ *Scala tympani* (Paukentreppe)
- ☐ Basilarmembran trennt Scala Media und Paukentreppe
- ☐ auf Basilarmembran ca. 15.000 Haarzellen
- ☐ *Äußere Haarzellen* (3 Reihen) dienen der Verstärkung der Schallwanderwellen innerhalb der Cochlea; Vorfilter
- ☐ *Innere Haarzellen* (1 Reihe) leisten Umwandlung mechanischer Schwingungen in Nervenimpulse (Transduktion)

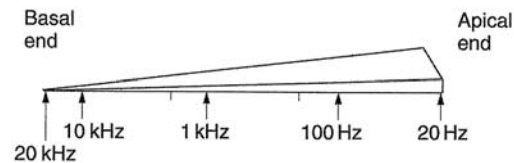
## Hörschnecke (Cochlea)



## Basilarmembran

- ☐ Breite der Basilarmembran nimmt vom ovalen Fenster zum Helicotrema hin zu.
- ☐ Dadurch ändern sich die mechanischen Eigenschaften (Massenbelag, Steife, Dämpfung) und damit auch die Schwingungseigenschaften des Systems in Abhängigkeit vom Abstand zum Helicotrema.
- ☐ Basilarmembran gerät für unterschiedliche Frequenzen an unterschiedlichen Stellen in Resonanz.
- ☐ Hohe Frequenzen besitzen in der Nähe des ovalen Fensters ein Auslenkungsmaximum,
- ☐ tiefe Frequenzen dagegen erst in der Nähe des Helicotrema.

## Basilarmembran



241

Christian Breiteneder

## Frequenzgruppen

- ❑ Gehör teilt die Basilarmembran in *24 gleich lange Abschnitte* auf (Frequenzgruppen, critical bands, kritische Bandbreiten)
- ❑ Nervenimpulse aus einer Frequenzgruppe werden gemeinsam ausgewertet ([Tonhöhe], Lautstärke, Klang, Richtung)
- ❑ Breite einer Frequenzgruppe beträgt ca. 100 Hz bei Frequenzen bis 500 Hz und eine kleine Terz oberhalb von 500 Hz. (entspricht ca. 1 Bark oder 100 Mel)

243

Christian Breiteneder

## Haarzellen

- ❑ An der Resonanzstelle werden die Schwingungen verstärkt und dadurch die inneren Haarzellen stärker stimuliert.
- ❑ Jenseits der Resonanzstelle werden die Schwingungen stark gedämpft, die entsprechende Frequenz breitet sich kaum weiter aus (Frequenzselektivität)
- ❑ Hohe Frequenzen (Resonanzstelle in der Nähe des ovalen Fensters) bewirken keine Anregung der inneren Haarzellen für tiefe Frequenzen,
- ❑ tiefe Frequenzen erregen aber auch die für hohe Frequenzen zuständigen Haarzellen.

242

Christian Breiteneder

## Frequenzgruppen

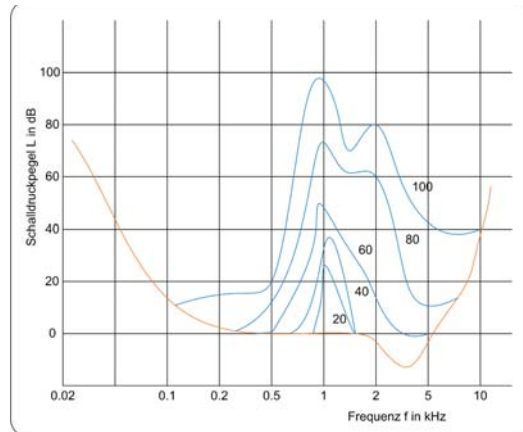
- ❑ Psychoakustischer Effekt des Gehörs
- ❑ Frequenzen gleichzeitig erklingender Töne müssen bestimmten Mindestabstand haben, um als Töne verschiedener Tonhöhe wahrgenommen zu werden
- ❑ Frequenzgruppe umfasst einen Frequenzbereich, innerhalb dessen unser Gehör Töne nicht differenzieren kann
  - ❑ bezieht sich nur auf gleichzeitig gespielte Töne (bei sequenziell gespielten Tönen ist Frequenzauflösung des Gehörs wesentlich höher)

244

Christian Breiteneder

## Maskierungseffekt

Hörschwellenkurve bei 1 kHz Maskierer (unterschiedliche Schallintensitäten)



245

Christian Breiteneder

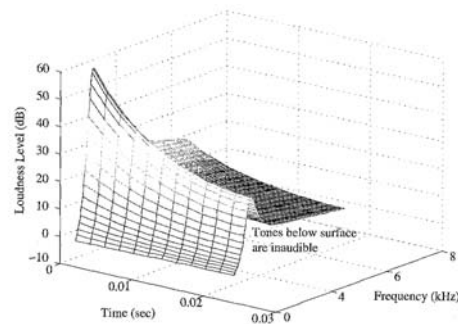
## Maskierungseffekt

- ❑ zwei Töne erklingen gleichzeitig
  - ❑ mit unterschiedlicher Lautstärke und
  - ❑ (ähnlicher Frequenz)
- ❑ Maskierungseffekt – Ohr hört nur den lauten Ton
  - ❑ leiser Ton wird "maskiert"
  - ❑ der laute Ton mindert Empfindlichkeit des Gehörs im umliegenden Frequenzbereich
  - ❑ leiser Ton liegt unter neuer Hörschwellenkurve
- ❑ Maskierungsschwellwert ist minimale Lautstärke, die der leise Ton aufweisen müsste, um nicht maskiert zu werden

246

Christian Breiteneder

## Maskierung

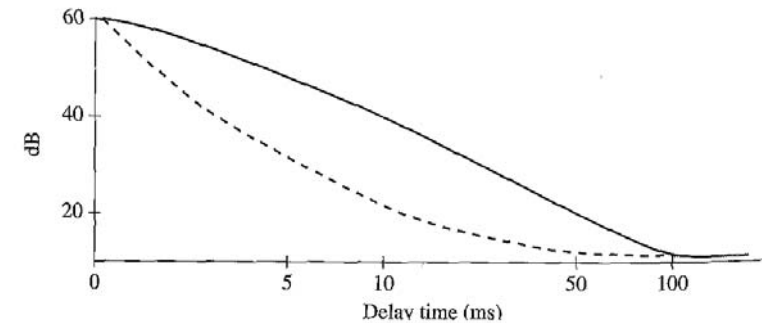


- ❑ Maskierung abhängig von Lautstärke, Frequenz, Zeitintervall und Dauer des maskierenden Tons ("Maskierers")

247

Christian Breiteneder

## Zeitliche Maskierung



- ❑ Spieldauer des Maskierers: 200 msec / 100 msec
- ❑ Pre-masking (2-5 msec) und Post-masking (50-200 msec)

248

Christian Breiteneder

## Maskierung und MPEG-Audio

- ❑ allgemeines Audiosignal besteht aus vielen Frequenzanteilen
  - ❑ jeder Frequenzanteil beeinflusst die Hörschwellenkurve
  - ❑ Audiosignal variiert mit der Zeit
    - ❑ Hörschwelle zu jedem Zeitpunkt verschieden
- ❑ Psychoakustisches Modell
  - ❑ Berechnet jeweils aktuelle Hörschwellenkurve
- ❑ Variable Quantisierung
  - ❑ Frequenzkomponenten so quantisiert, dass Quantisierungsrauschen gerade unter aktueller Hörschwellenkurve
  - ❑ Wo maskiert wird, grobe Quantisierung, kleine Wortlängen

## MPEG Polyphase Filterbank

- ❑ MPEG Polyphase Filterbank
  - ❑ Audiosignal wird in 32 Subbands unterteilt
  - ❑ Bandbreite für jedes Subband gleich
- ❑ Vorteil
  - ❑ einfacher Aufbau
- ❑ Nachteil
  - ❑ untere Subbands erfassen mehrere Frequenzgruppen
  - ❑ überlappende Subbands bringen Informationsverlust

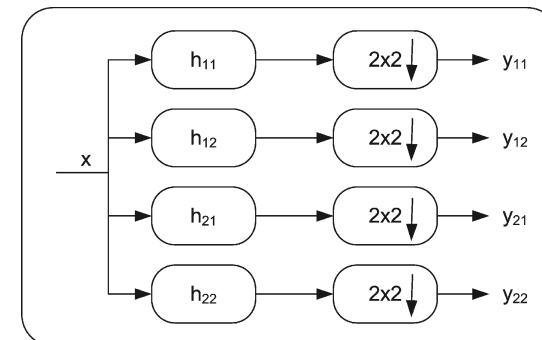
## Subbandkodierung

MPEG-Audio kodiert Frequenzkomponenten

- ❑ Filterbank
  - ❑ teilt Audiosignal in Frequenzbänder auf (Subbands)
- ❑ Ideale Filterbank
  - ❑ Subbands identisch mit Frequenzgruppen des menschlichen Gehörs

## Filterbank – Kritische Abtastung

Beispiel Filterbank mit 4 Subbands

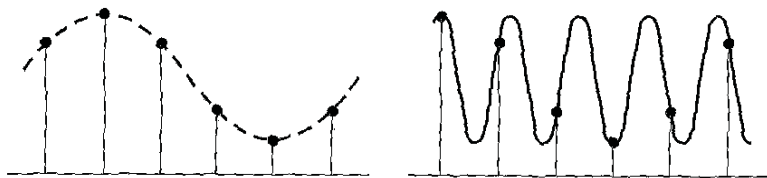


## Filterbank – Kritische Abtastung

Beispiel Filterbank mit 4 Subbands

- ❑ Eingang – 1 x PCM: Frequenzbereich 0 - 20kHz, Abtastrate 48 kHz; Wortlänge 16 Bit pro Abtastwert
- ❑ Ausgang – 4 x PCM
  - ❑ Subband 1: 0Hz - 5kHz; Abtastrate 12kHz (!); 16 Bit
  - ❑ Subband 2: 5kHz - 10 kHz; Abtastrate 12kHz (!); 16 Bit
  - ❑ Subband 3: 10kHz - 15 kHz; Abtastrate 12kHz (!); 16 Bit
  - ❑ Subband 4: 15 kHz - 20 kHz; Abtastrate 12kHz (!); 16 Bit

## Filterbank – Kritische Abtastung



- ❑ Samples / Signal ohne Angabe des Frequenzbereichs (links)
- ❑ Samples / Signal mit Angabe des Frequenzbereichs (rechts)

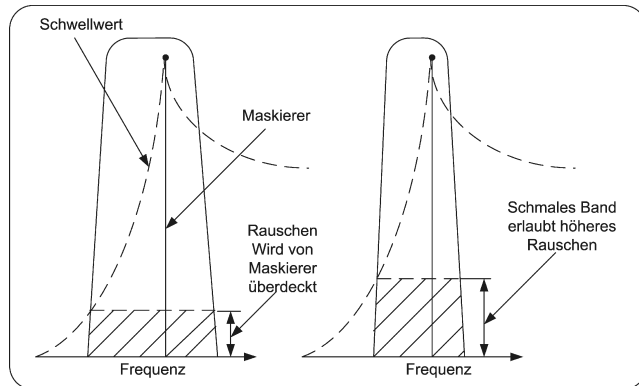
## Filterbank – Kritische Abtastung

- ❑ Kritische Abtastung digitaler Filterbanken
  - ❑ Wortlängen bleiben in Subbands unverändert
  - ❑ Abtastrate in Subbands reziprok zur Anzahl der Subbands
  - ❑ Reduzierte Abtastrate in Subbands wird als kritische Abtastrate bezeichnet
- ❑ Kritische Abtastung der Subbands gewährleistet, dass Filterbank keinen Einfluss auf Datenmenge hat
  - ❑ Es werden zwar mehr Kanäle erzeugt, aber deren Abtastfrequenzen sind dementsprechend geringer.

## Subbands im MPEG Standard

- ❑ MPEG-1/Layer 1 und Layer 2 – 32 Subband Filter Bank
- ❑ MP3
  - ❑ Hybridfilterbank – 32 Filterbank wird MDCT nachgeschaltet
    - ❑ Frequenzzersetzung entspricht 576 Subbands
- ❑ MPEG-2 – AAC
  - ❑ Verlustlose MDCT – Frequenzzersetzung entspricht 1024 Subbands
- ❑ Unterschied Subband/MDCT (In der Praxis aber gleichgesetzt)
  - ❑ Subband – Zeitwerte in einem bestimmten Frequenzband
  - ❑ MDCT – Reine Frequenzkomponenten

## Subbandanzahl / Kompression



## MPEG Psychoakustisches Modell

- ❑ analysiert Audiosignal abschnittsweise auf Maskierungseffekte
  - ❑ Algorithmen simulieren menschliches Gehör
- ❑ berechnet Maskierungsschwellwert / Subband
  - ❑ berechnet maximal erlaubtes Quantisierungsrauschen/ Subband
- ❑ Psychoakustisches Modell nicht Teil von MPEG
  - ❑ andere Algorithmen können verwendet werden
  - ❑ Qualität des psychoakustischen Modells hauptverantwortlich für Gesamtqualität der Kodierung

## Subbandanzahl / Kompression

- ❑ Für jedes Subband wird untersucht, wie hoch Quantisierungsrauschen sein darf, das noch maskiert wird
- ❑ Schmale Subbands
  - ❑ hoher Maskierungsschwellenwert für Rauschen
  - ❑ Frequenzkomponenten benötigen nur geringe Wortlänge, das entspricht hoher Kompression
- ❑ Allgemein gilt
  - ❑ je höher die Subbandanzahl desto schmaler die Subbands
  - ❑ je schmaler die Subbands desto höher die Kompression

## Modelle des MPEG Audiostandards

MPEG bietet zwei exemplarische Implementationen von psychoakustischen Modellen

- ❑ Modell 1 – optimiert für schnelles Rechnen
- ❑ Modell 2 – optimiert für genaues Rechnen, Spezifikationen für MP3

Arbeitsweise (beide Modelle)

- ❑ Berechnung mehrerer Maskierungsschwellwerte pro Subband
- ❑ aus Berechnung 1 wird dann erst endgültiger Wert ermittelt
- ❑ Modell 2 verwendet dazu komplexere Algorithmen

## Tonale / Atonale Komponenten

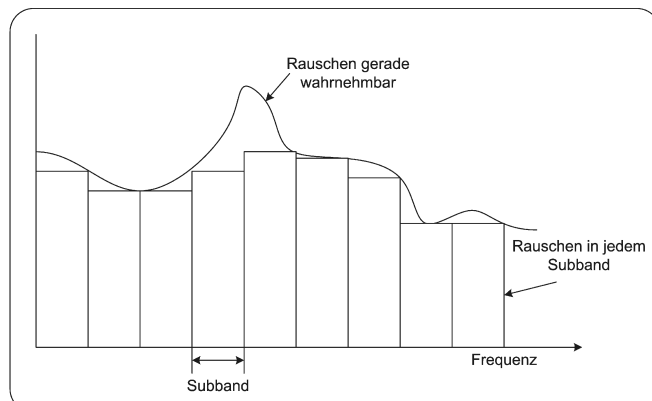
- ❑ tonale und geräuschähnliche akustische Ereignisse haben unterschiedliche Maskierungseigenschaften
- ❑ Psychoakustisches Modell benötigt Algorithmen, um zwischen Ton und Geräusch unterscheiden zu können
- ❑ Modell 1
  - ❑ starke Spitzen im Frequenzspektrum als tonale Komponenten gewertet
  - ❑ verbleibende Anteile als Geräusche interpretiert
- ❑ Modell 2
  - ❑ komplexere Algorithmen, Berücksichtigung Phasen, etc.

261

Christian Breiteneder

## Bsp: Ergebnis psychoakustisches Modell

- ❑ kontinuierliche Kurve – Berechnete Hörschwellenkurve
- ❑ Treppenkurve – Maximales Rauschen/Subband



263

Christian Breiteneder

## Fast Fouriertransformation

- ❑ zur Signalanalyse ist FFT besonders gut geeignet
  - ❑ liefert feine Frequenzauflösung des Signals
  - ❑ liefert Phaseninformationen der Frequenzkomponenten (wichtig für Modell 2)
  - ❑ sehr genaue Berechnung der Maskierungsschwellwerte
    - ❑ hohe Kompression möglich
- ❑ Technik zur Minimierung des Rechenaufwandes
  - ❑ Zusammenfassung der Werte in Frequenzgruppen
  - ❑ Frequenzgruppen entsprechen kritischen Bandbreiten des menschlichen Gehörs

262

Christian Breiteneder

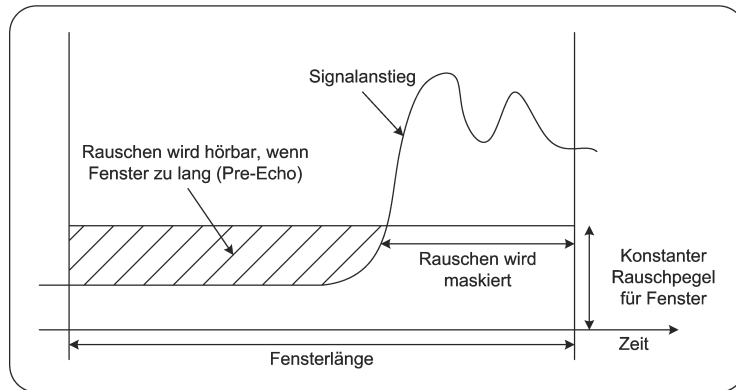
## Bsp: Ergebnis psychoakustisches Modell

- ❑ Psychoakustisches Modell errechnet Hörschwellenkurve
  - ❑ Rauschen unterhalb Kurve ist nicht hörbar
  - ❑ Rauschen oberhalb Kurve ist hörbar
  - ❑ für rauschfreie Kodierung darf hörbares Rauschniveau an keiner Stelle überschritten werden
- ❑ Berechnung der Maskierungsschwellwerte
  - ❑ für jedes Subband wird der Maskierungsschwellwert berechnet
  - ❑ je kleiner Subbands, desto besser die Anpassung der Stufenfunktion an Hörschwellenkurve (hohe Kompression)

264

Christian Breiteneder

## Kompression bei steilem Signalanstieg



265

Christian Breiteneder

## Frequenzauflösung / Zeitfenster

- ❑ hohe Frequenzauflösung bei langem Zeitfenster
  - ❑ + : viele Subbands, hohe Kompression
    - ❑ Subbands gut an Hörschwellenkurve anpassbar
  - ❑ - : Quantisierungsrauschen über lange Zeit nicht änderbar
    - ❑ schlecht bei Signalanstiegen
- ❑ geringe Frequenzauflösung bei kurzem Zeitfenster
  - ❑ + : Quantisierungsrauschen in kurzen Zeitabständen änderbar
    - ❑ Rauschen kann gut an Zeitkurve angepasst werden
  - ❑ - : weniger Subbands, mäßige Kompression

267

Christian Breiteneder

## Kompression bei steilem Signalanstieg

- ❑ Signalabschnitte (Fenster) mit steilen Signalanstiegen sind sehr schwer zu komprimieren
  - ❑ Quantisierungsrauschen über gesamte Fensterlänge gleich
  - ❑ Rauschen vor Signalanstieg als Pre-Echo hörbar
- ❑ MPEG Strategien
  - ❑ Layer1/Layer2 – keine artefaktfreie Kompression möglich
  - ❑ MP3 – verwendet dafür kurzes Fenster, Pre-Echo Effekt wird vermindert, mäßige Kompression, da bei kurzem Fenster weniger Subbands erzielbar
  - ❑ AAC – kurzes Fenster, TNS Modul, gute Kompression

266

Christian Breiteneder

## Skalierfaktor

- ❑ Skalierung ist eine Abwandlung der Bandrausch-  
unterdrückungssysteme für Tonbandgeräte (DolbyA, B, C)
  - ❑ kleine Signalanteile vor Bandaufzeichnung verstärkt
  - ❑ bei Wiedergabe um selben Faktor gedämpft
  - ❑ Bandrauschen mit selben Faktor gedämpft
- ❑ Skalierung bei MPEG Kodierung
  - ❑ kleine Werte vor Kodierung mit Skalierfaktor multipliziert
  - ❑ bei Dekodierung mit selben Faktor dividiert
  - ❑ Quantisierungsrauschen um Skalierfaktor gedämpft
- ❑ Skalierung verbessert SNR für leise Signale

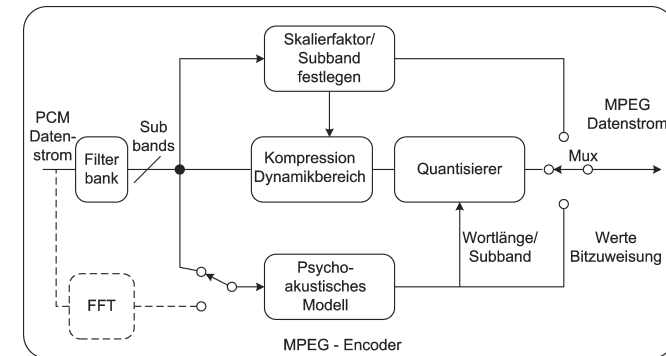
268

Christian Breiteneder

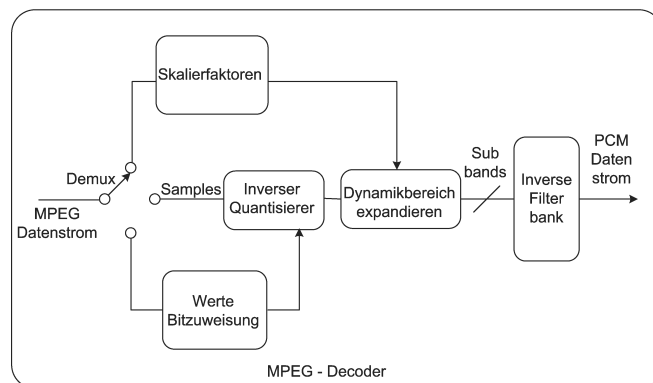
## Bitzuweisung

- ❑ Benutzer(in) kann bei MPEG gewünschte Datenrate wählen
  - ❑ Psychoakustisches Modell berechnet Werte, dass kein Quantisierungsrauschen hörbar
  - ❑ bei niedrigen Datenraten reichen meist Bitstellen nicht aus
  - ❑ Algorithmen zur Bitzuweisung
    - ❑ hohe Frequenzen werden entfernt
    - ❑ Stereo wird räumlich ungenauer aufgelöst
    - ❑ wohldosierte Einführung von Quantisierungsrauschen
- ❑ Algorithmen bleiben Softwareentwickler überlassen
- ❑ MPEG/Audio Standard bietet Tabellen als Empfehlung

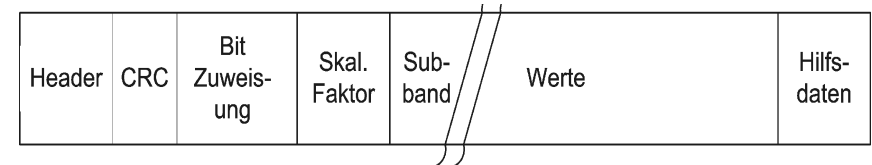
## Einfacher MPEG Encoder



## Einfacher MPEG Decoder



## Einfacher MPEG Datenstrom



- ❑ Header– Beschreibung des Datenstroms
- ❑ Redundancy Code (CRC) – Erkennung von Datenstromfehlern
- ❑ Bitzuweisung – Wortlänge der folgenden Subbandwerte
- ❑ Skalierfaktor – 6-bit Skalierfaktor für folgende Subbandwerte
- ❑ Hilfsdaten – Beliebige Daten können eingefügt werden

## MPEG-1 Audio Standard

- ❑ Precision: 16 Bits, sampling frequencies 32kHz, 44.1kHz, 48kHz
- ❑ audio coding: single channel, 2 independent channels, 1 stereo channel
- ❑ 3 compression methods: Layer 1, Layer 2, Layer 3 (hierarchical decoding)
  - ❑ Layer 3: decoder is accepting Layer 2 and Layer 1 (32kbps-320kbps, target 64 kbps)
  - ❑ Layer 2: decoder is accepting Layer 1 (32 kbps-384kbps, target 128 kbps)
  - ❑ Layer 1: 32 kbps-448kbps, target 192 kbps

## MPEG-2 Audio

original (finalized 1994) MPEG-2 Audio standard consists of 2 extensions to MPEG-1:

- ❑ backwards compatible multichannel coding adds option of forward and backwards compatible coding of multichannel signals (including 5.1 channel configurations)
- ❑ coding at lower sampling frequencies of 16kHz, 22.05 kHz and 24 kHz

verification tests in 1994 showed that introducing new coding algorithms and giving up backwards compatibility promised a significant improvement in coding efficiency => AAC

## MPEG Layer-3 Audio Coding

- ❑ operating mode
  - ❑ single channel                      ❑ stereo
  - ❑ dual channel                        ❑ joint stereo
- ❑ sampling frequency
  - ❑ MPEG-2.5 (Frauenhofer): 8 kHz, 11.05 kHz, 12 kHz
- ❑ Bit-rate
  - ❑ within limits left to the implementer; standard defines a range from 32 kbit/s (MPEG-1) or 8 kbit/s (LSF, MPEG-2) to 320 kbit/s (resp. 160 kbit/s for LSF).

## Layer-3 Encoding Algorithm

basic building blocks

- ❑ filterbank—actually 2 cascading filterbanks:
  - ❑ polyphase filterbank (as in Layer-1 and 2)
  - ❑ additional Modified DCT (MDCT); the subdivision of each polyphase frequency band into 18 finer subbands increases the potential for redundancy removal;
- ❑ perceptual model
  - ❑ mainly determining the quality of a given encoder implementation; uses either a separate filterbank or combines the calculation of energy values (for masking calculations) and main filter bank

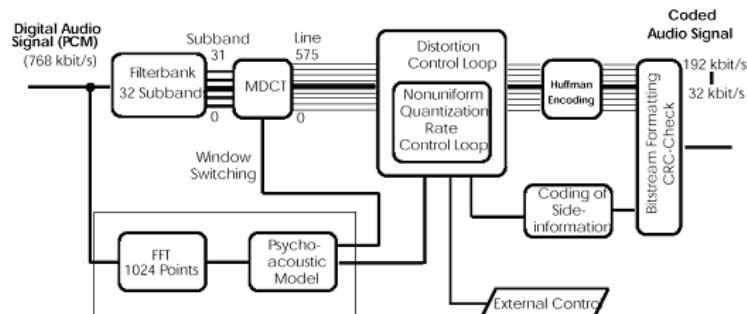
## Layer-3 Encoding Algorithm

- ❑ perceptual model (continued)
  - ❑ output of the perceptual model consists of values for the masking threshold or allowed noise for each coder partition (roughly equivalent to the critical bands of human hearing)
- ❑ quantization and coding
  - ❑ two nested iteration loops (rate loop and Noise control loop)
  - ❑ quantization is done via a power-law quantizer; larger values are automatically coded with less accuracy
  - ❑ quantized values are coded by Huffman coding

277

Christian Breiteneder

## Layer-3 Encoding Algorithm



279

Christian Breiteneder

## Layer-3 Encoding Algorithm

- ❑ quantization and coding (continued)
  - ❑ inner iteration loop (rate loop)
 

If number of bits resulting from coding operation exceeds the number of bits available to code a block of data, a larger quantization step is applied.
  - ❑ outer iteration loop (noise control loop)
 

If the quantization noise in a given band is found to exceed the masking threshold (allowed noise) as supplied by the perceptual model, the scalefactor for this band is adjusted to reduce the quantization noise.

278

Christian Breiteneder

## Quality Considerations

- ❑ common types of artifacts
  - ❑ loss of bandwidth
  - ❑ pre-echoes
 

name somewhat misleading: the basic coding artifact is noise spread out over some time even before the music event causing the noise (see next slide)
  - ❑ roughness, double-speak
 

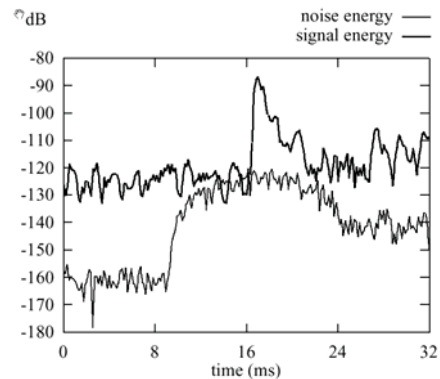
at low bit-rates and lower sampling frequencies there is a mismatch between time resolution of the coder and the requirements to follow the time structure of some signals.

280

Christian Breiteneder

## Quality Considerations

### ❑ pre-echoes



## ABC/HR Test

- ❑ listening sequence is "ABC", with A = original, BC = pair of original / coded signal with random sequence,
- ❑ listener must first distinguish the encoded version from the original (Hidden Reference, "HR" )
- ❑ then assign a score as a subjective judgment of the quality.

## How to Measure Codec Quality

### listening tests

- ❑ large scale and well-controlled listening tests are still the only method available
- ❑ the aim of these tests is to stress the encoders under worst case conditions
- ❑ During the ISO-MPEG process: a number of international listening tests with trained listeners.
- ❑ tests used the "triple stimulus, hidden reference" (ABC/HR) method and the "CCIR impairment scale"

## CCIR Impairment Scale

- ❑ the listener has to evaluate both B and C with a number between 1.0 and 5.0.
- ❑ meaning of these values:
  - 5.0 = transparent (this should be the original signal)
  - 4.0 = perceptible, but not annoying (first differences noticable)
  - 3.0 = slightly annoying
  - 2.0 = annoying
  - 1.0 = very annoying

## How to Measure Codec Quality

- ❑ simple objective measurement techniques
  - ❑ *misleading; these methods defy the whole purpose of perceptual coding*
- ❑ example measures: signal-to-noise-ratio, bandwidth of decoded signal

## How to Measure Codec Quality

- ❑ perceptual measurement techniques (pmt)
  - ❑ state of the art not yet sufficient to make large scale and well-prepared listening tests obsolete
  - ❑ however, pmt have progressed to the point where they are a very useful supplement to listening tests and can replace them in some cases.
  - ❑ ITU-R Task Group 10/4 (Int. Telecommunications Union, Radiocommunications sector) produced a recommendation on a system called PEAQ (Perceptual Evaluation of Audio Quality)