C I & S S L a b

UNIVERSITÀ PARTHENOPE
DiST
DIPARTIMENTO DI SCIENZE E TECNOLOGIE

Intelligent Signal Processing

Audio Compression

Angelo Ciaramella

Introduction

- To be transmitted over the network multimedia content
  - must be digitized and compressed

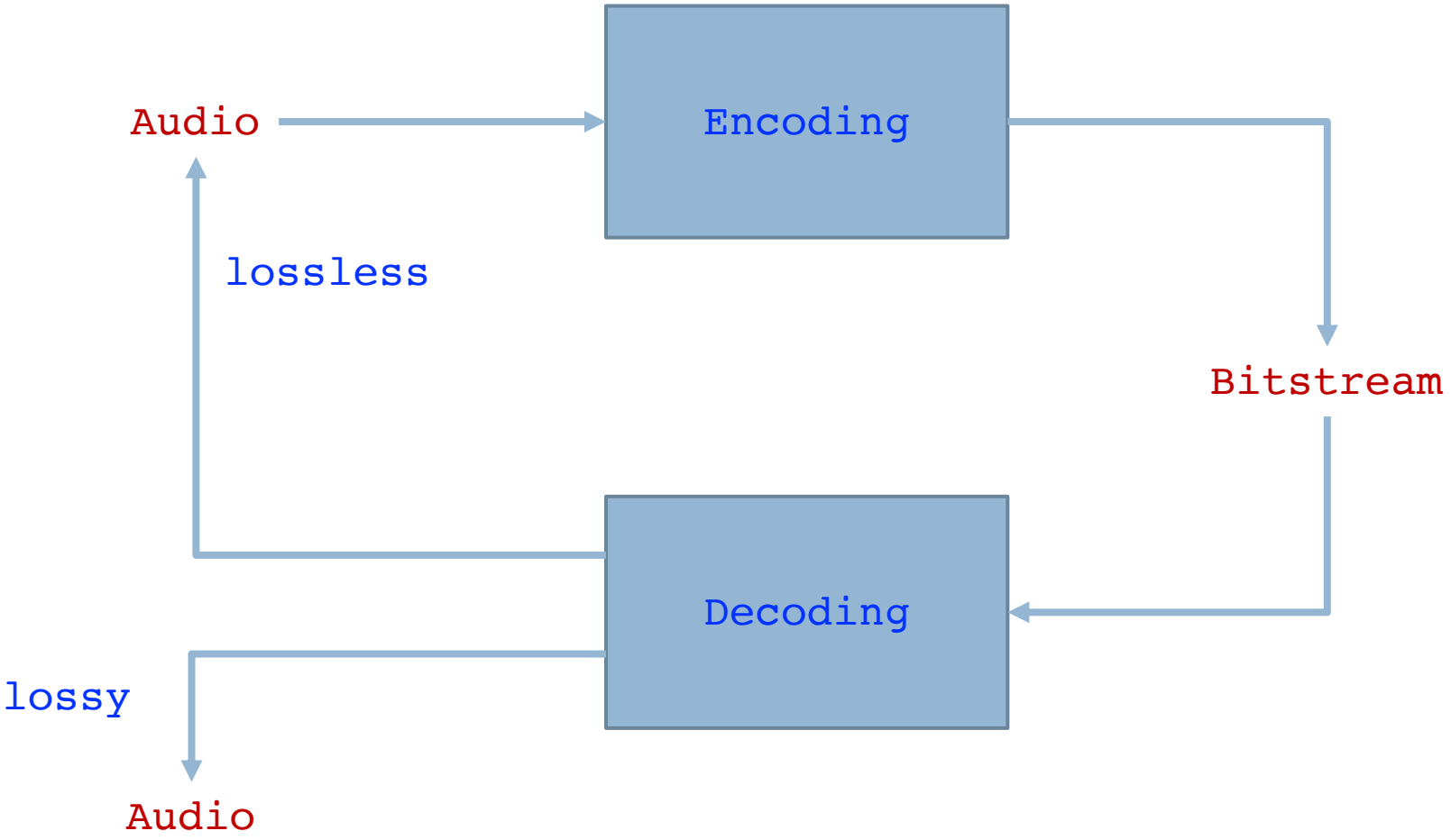
- Image
  - uncompressed 1024 x 1024 image
    - 8 bits for each color (RGB)
  - 3 Mbyte of memory
  - the transmission on a 64 Kbps channel needs of 7 minutes

Pulse Code Modulation

- Pulse Code Modulation (PCM)
  - sampling frequency
  - quantization bits

- Examples
  - voice
    - 8000 Hz
    - 8 bits
  - CD audio
    - 44100 Hz
    - 16 bits

Encoding-Decoding



Compression scheme

- Two kinds of signal processing
  - redundant data elimination
    - Run Length Encoding
    - RLE-Huffman
    - LZ(W)
  - Percettive based approaches
    - PASC
    - Mpeg Layer 3 (mp3)

Silence compression

- Silence compression
  - signals with pauses
  - near zero intensity

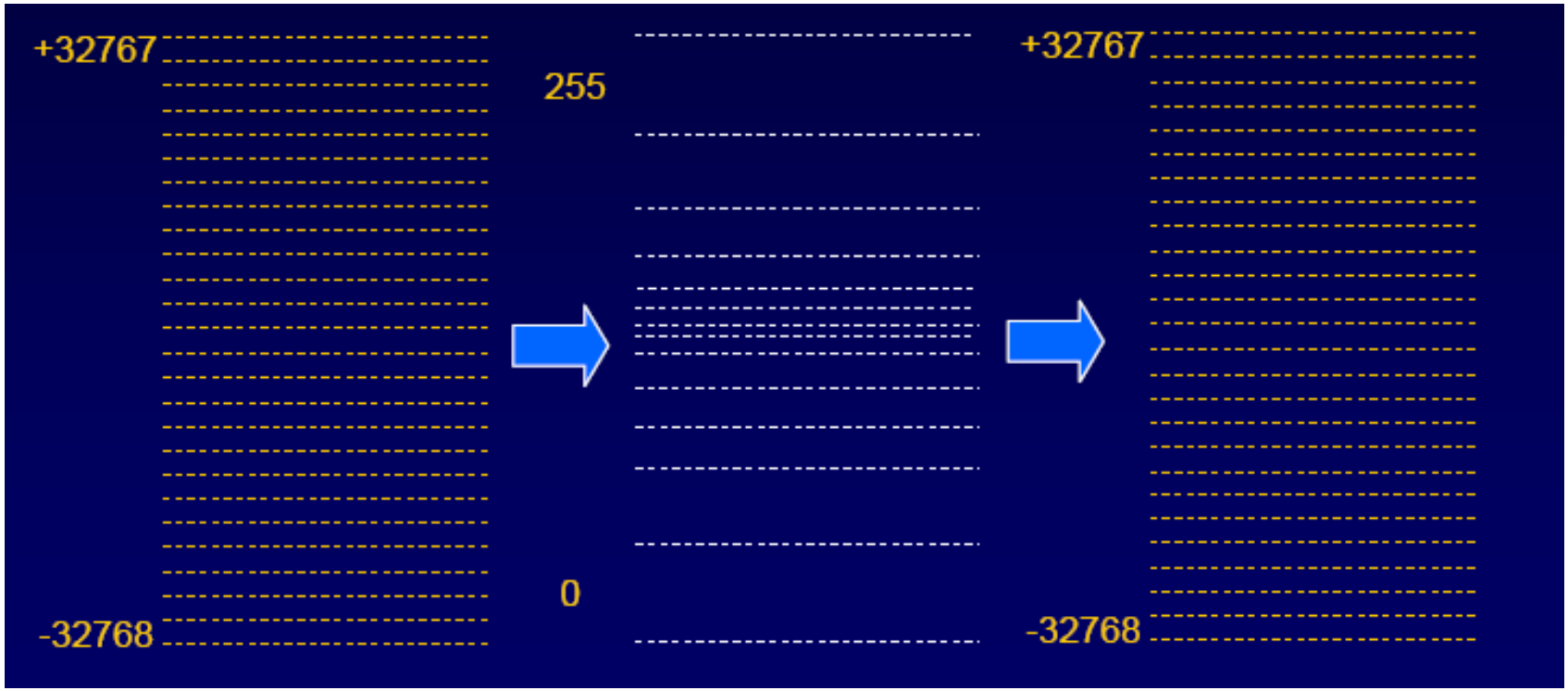
- RLE

  ddddddddhhhhhhhhyyyyyyyyy → 7d8h9y

μ-law and A-law compressions

- μ-law coding
  - North America and Japan
  - digital phone on ISDN
- A-law
  - Europe
  - International traffic on ISDN
- Both uses a 8 bits for quantization
- It is a lossy compression

μ-law and A-law compressions



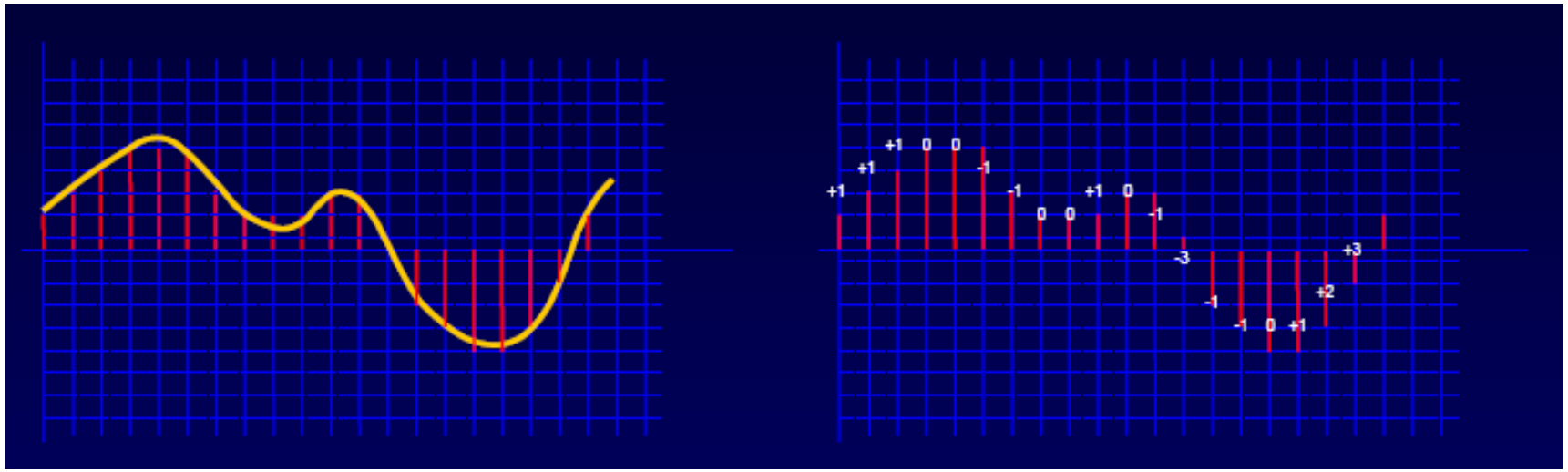
16 bits 8 bits 16 bits

ln exp

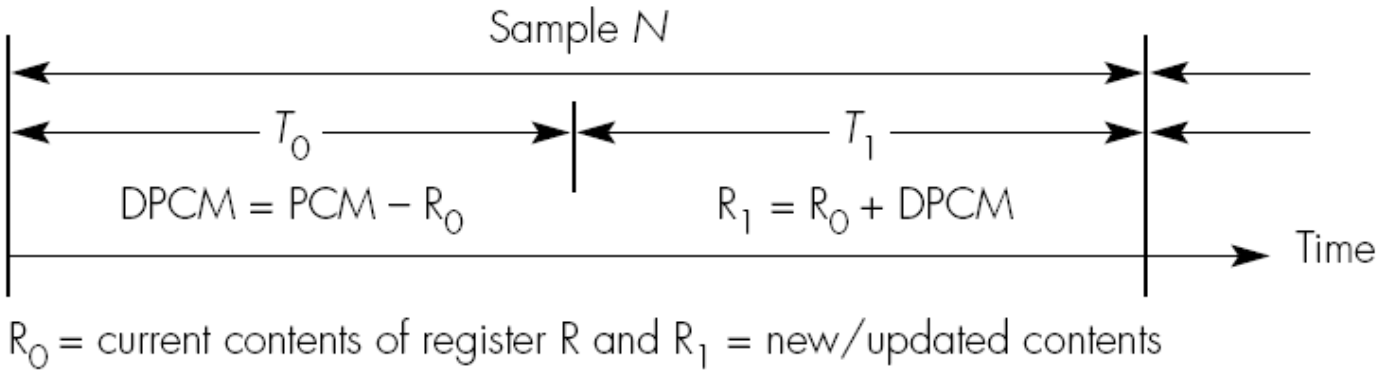
Aim – to use 8 bits instead of 8 bits

Differential Pulse Code Modulation

- Differential Pulse Code Modulation (DPCM)
  - derives from PCM
  - difference between two consectuive samples

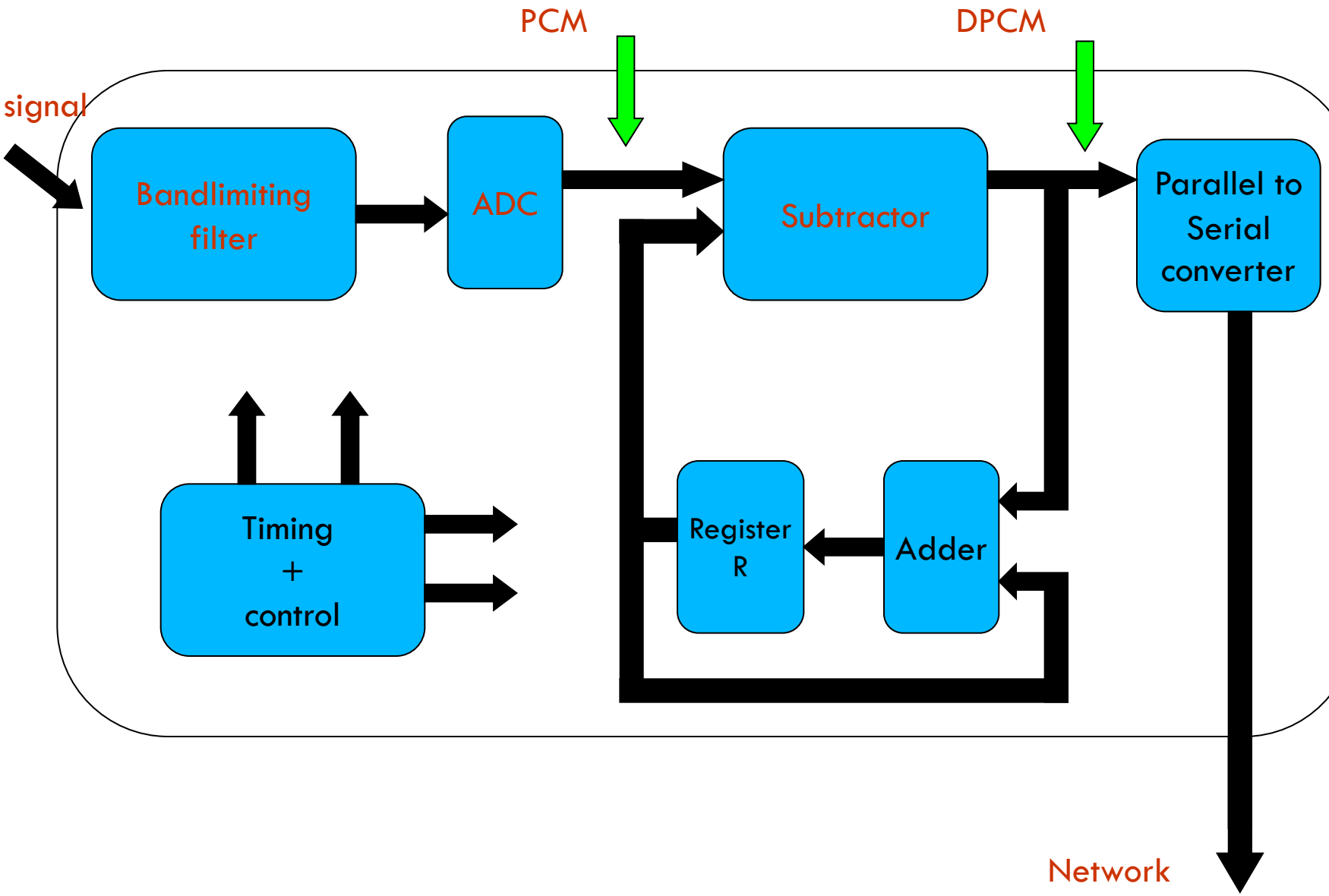


DPCM

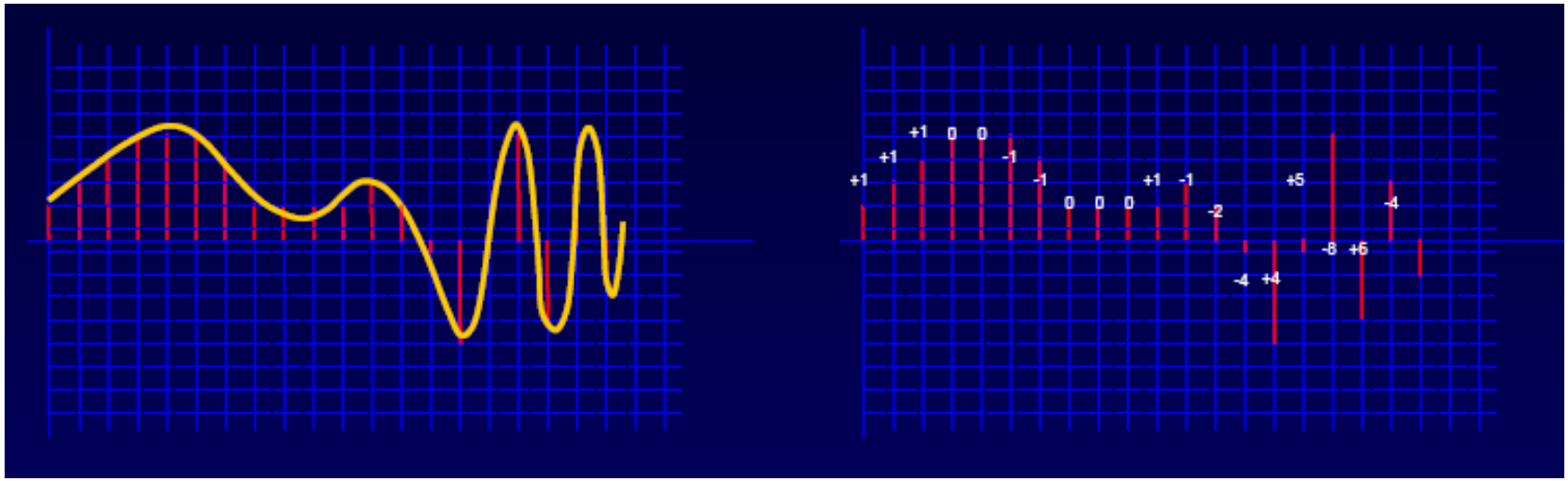


Timing phase. Two registers are used.

DPCM encoder

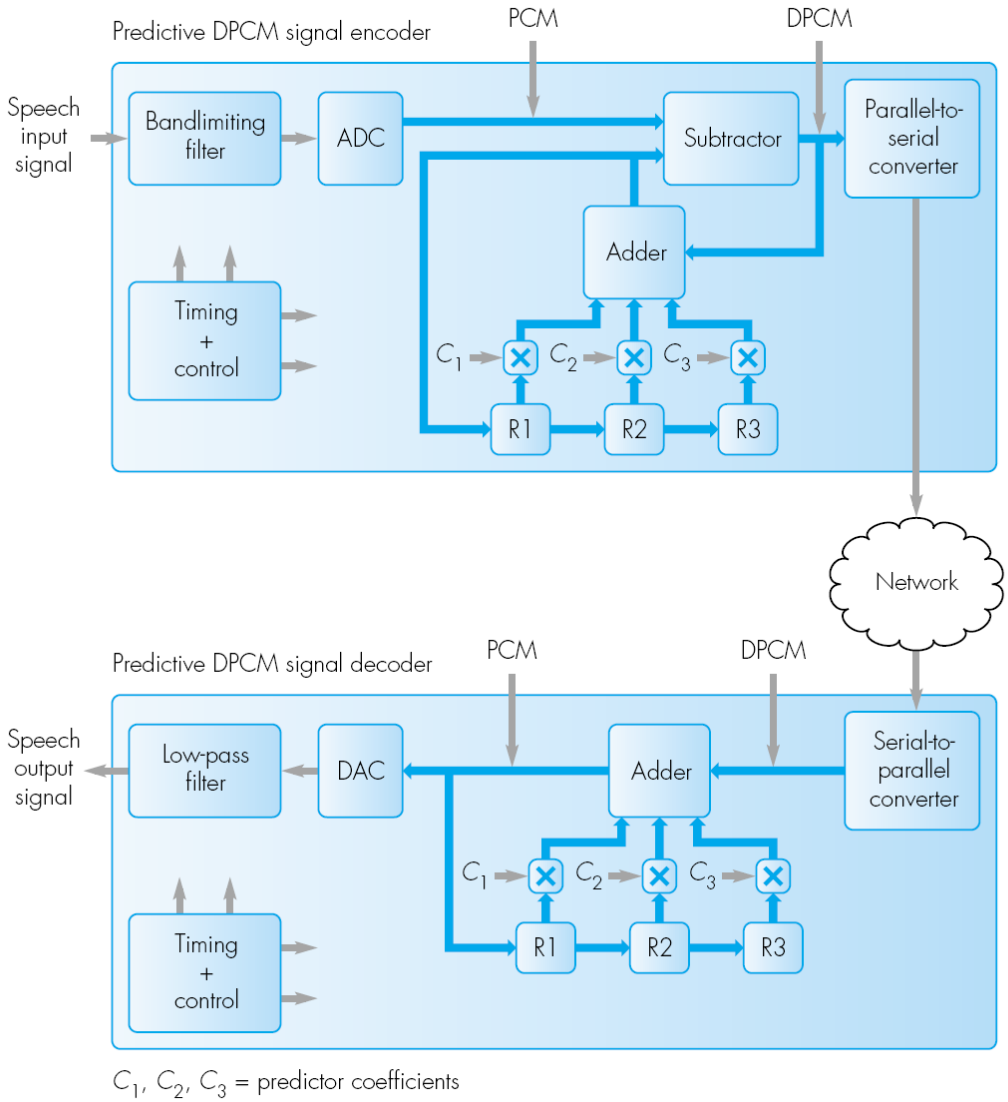


DPCM slope overload



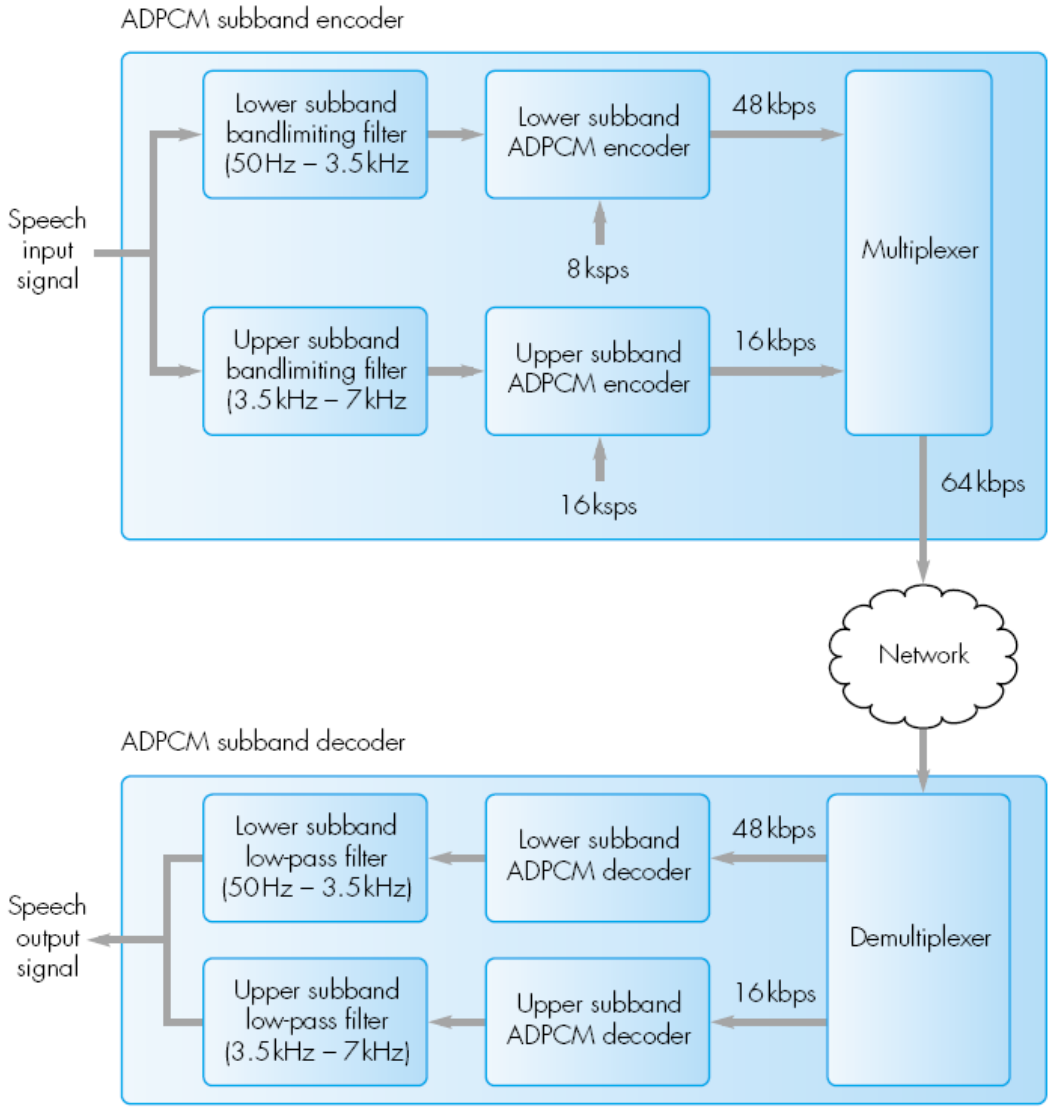
High frequencies differences needs a higher number of bits

Predictive DPCM



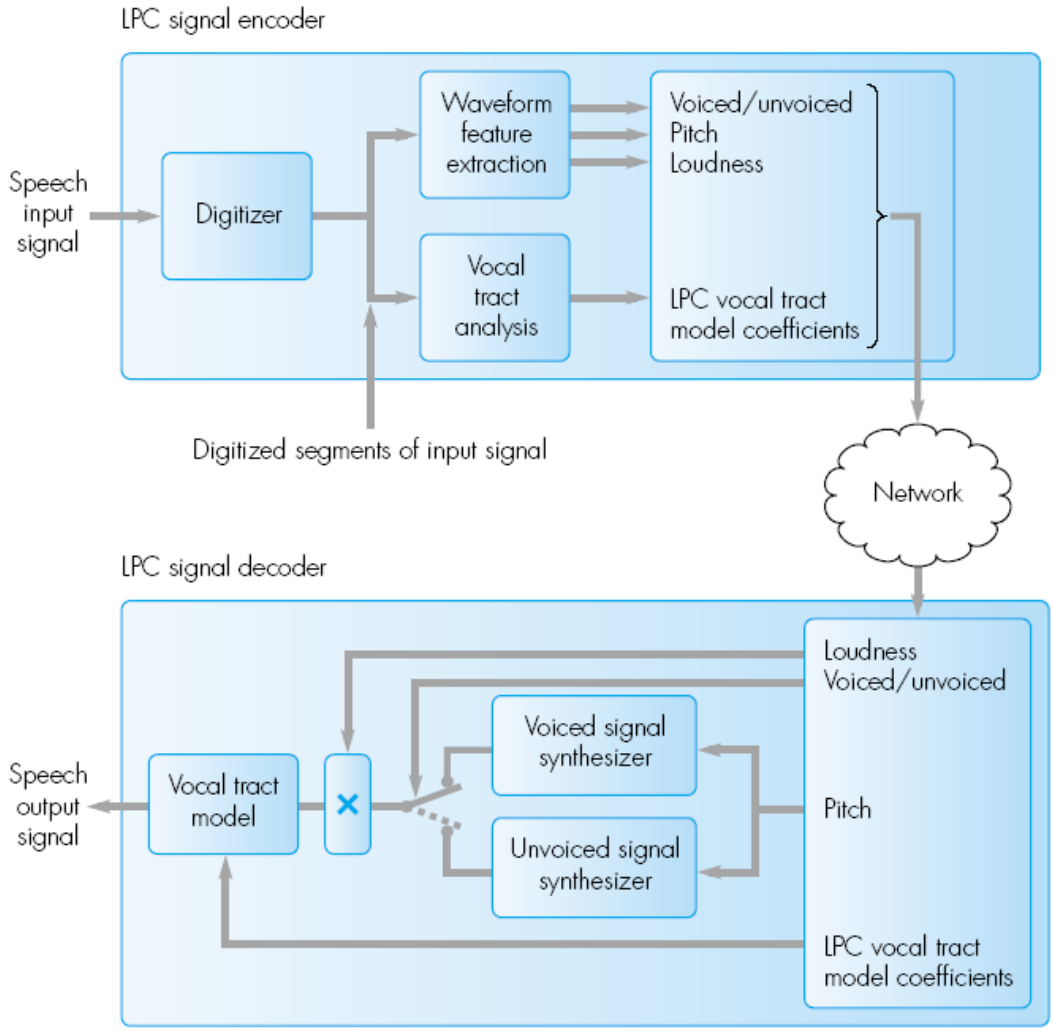
Prediction by using 3 registers and 3 coefficients

Adaptive DPCM



A filtering scheme is used

Linear Predictive Coding



LPC scheme

Code-excited LPC

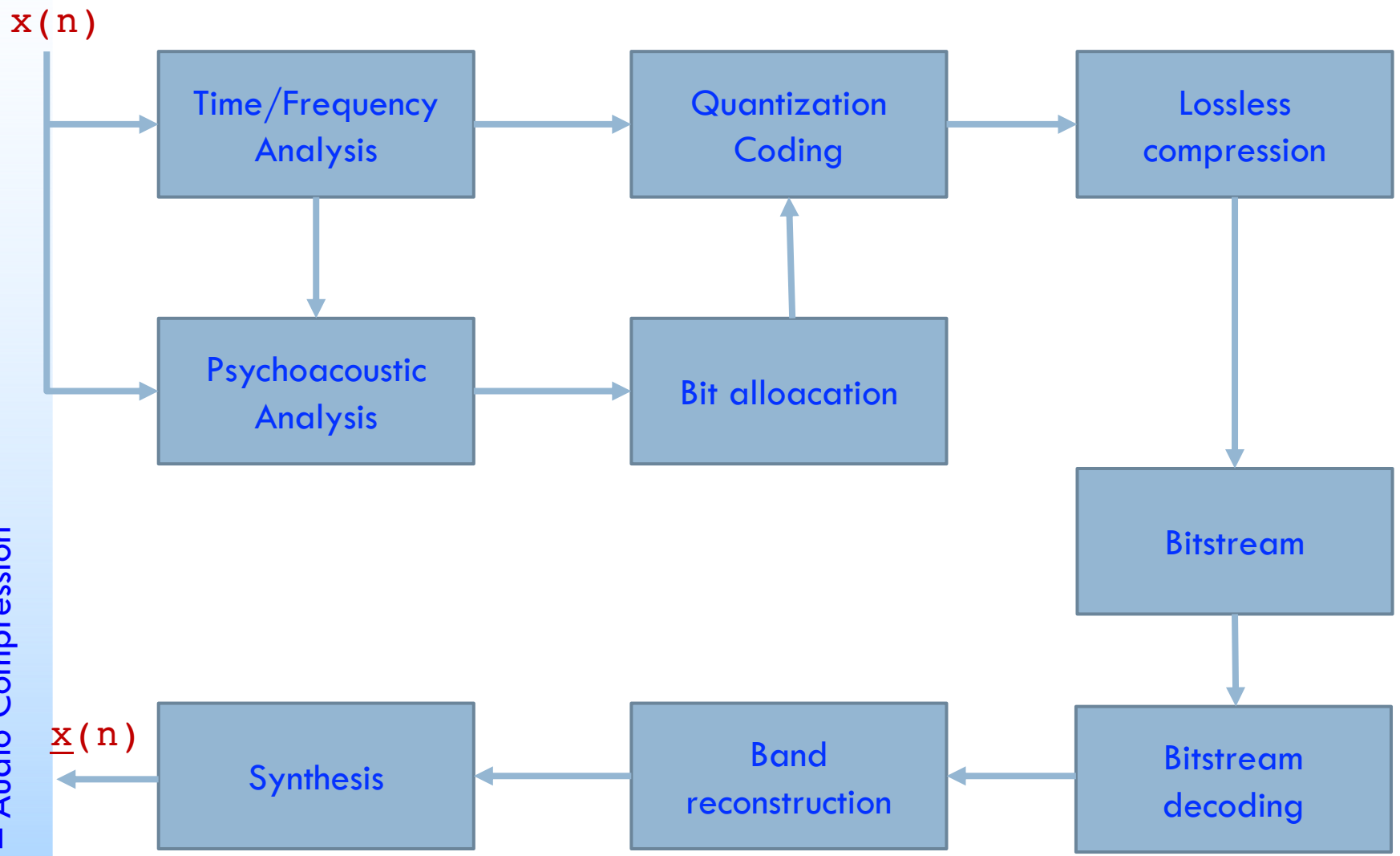
- CELP
  - Set of segments (templates)
    - named codebook
  - transmitted codeword
    - template with best matching with an input segment

Perceptual coding

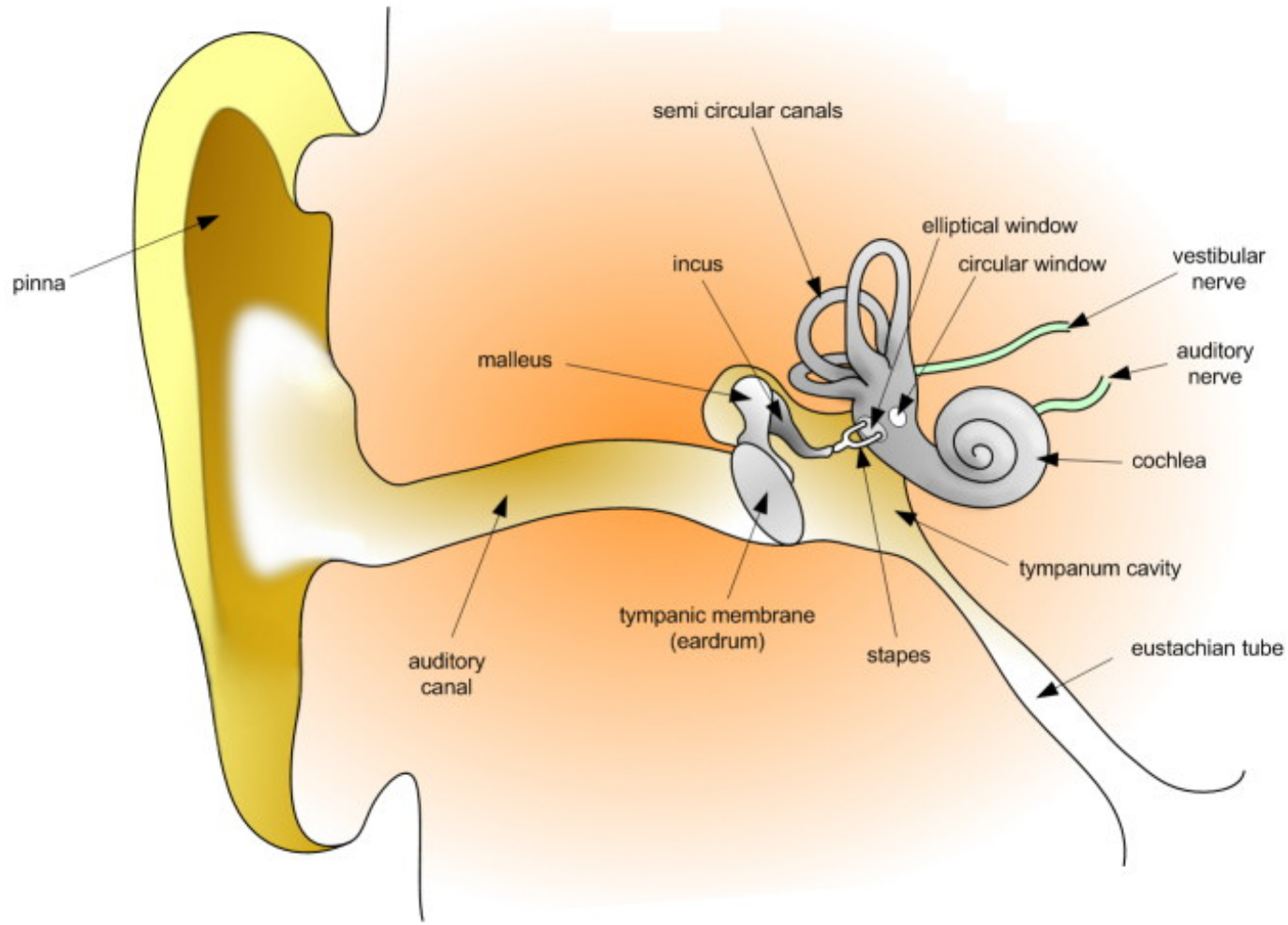
- psychoacoustic models
  - exploit the characteristics of the human ear
  - only the perceptual characteristics are transmitted

- Main aspects
  - frequency masking
  - temporal masking

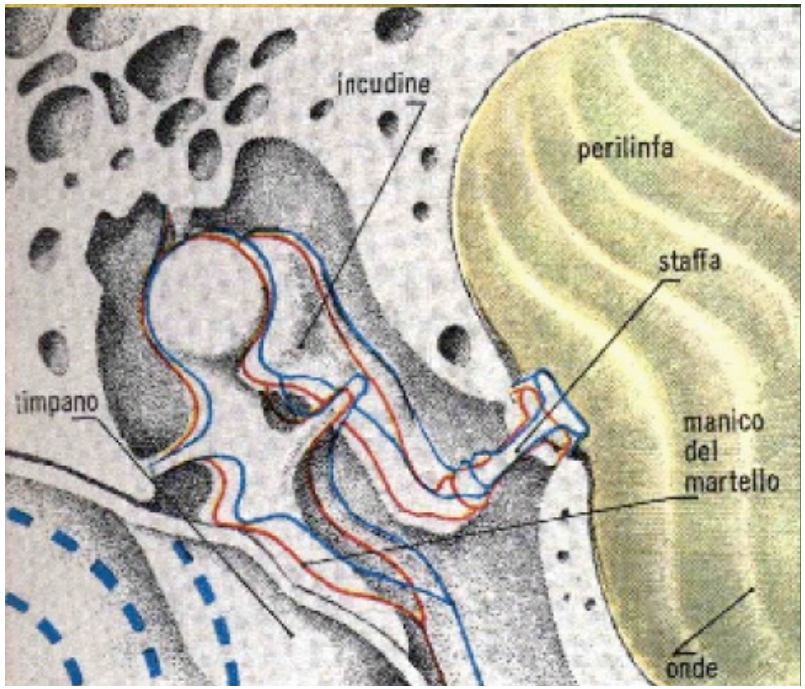
General scheme



Auditory perception

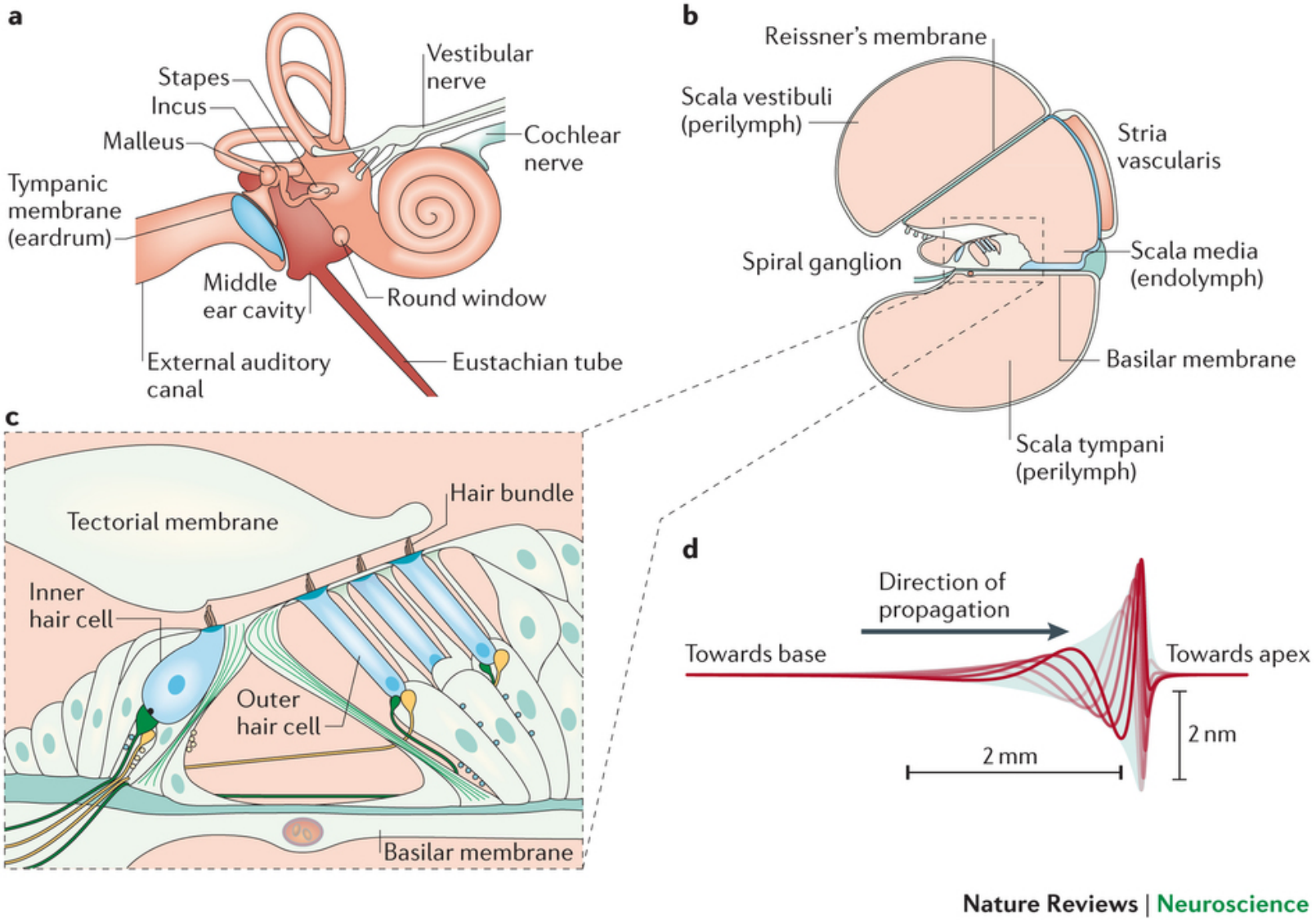


Middle ear

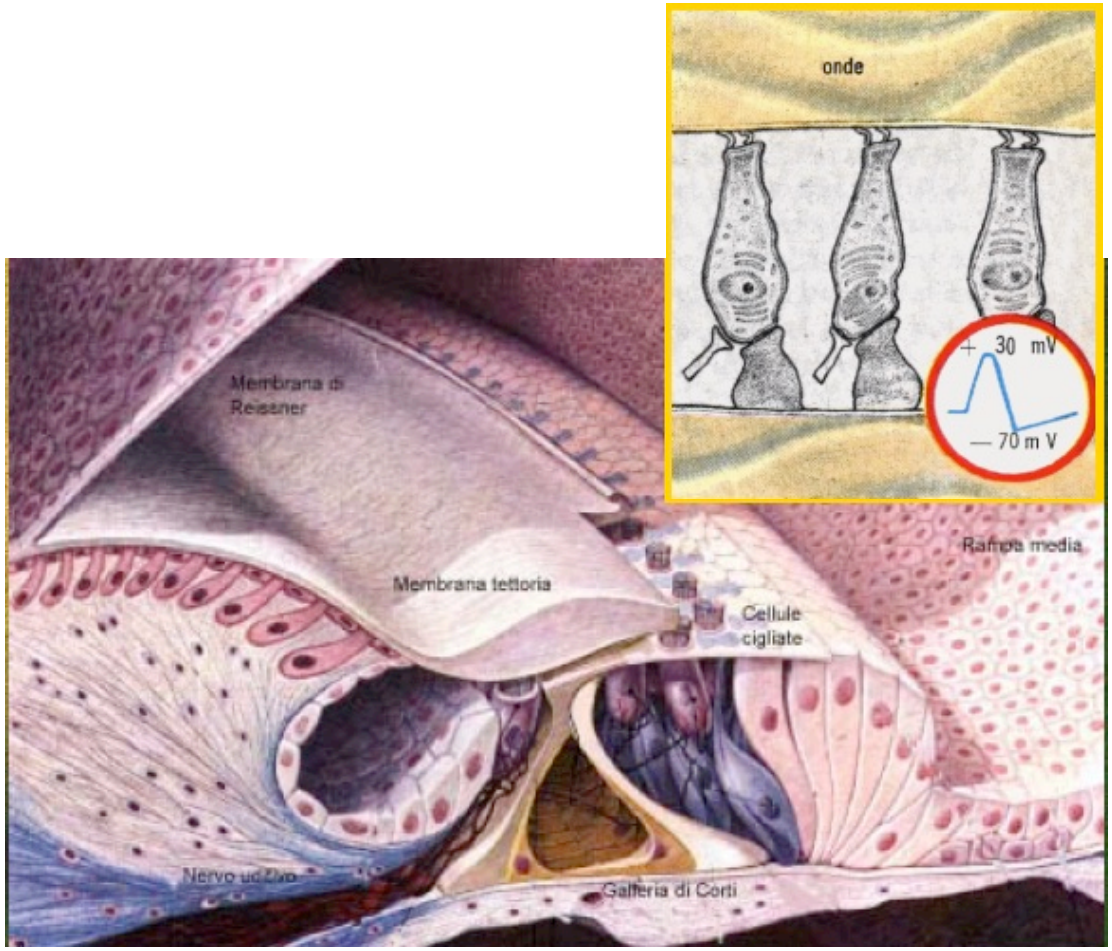


Vibration propagation

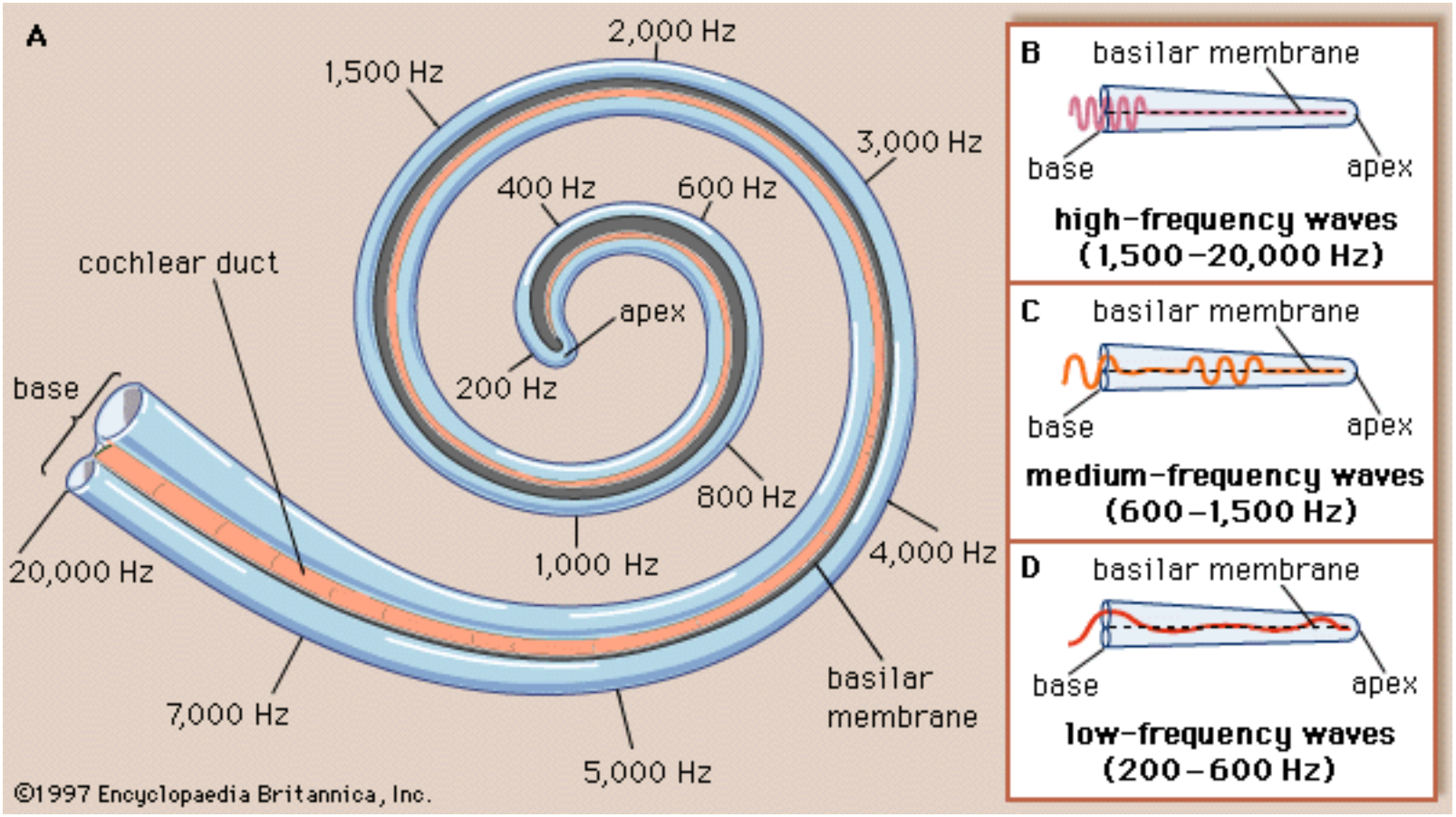
Cochlea


a
Stapes
Incus
Malleus
Tympanic membrane (eardrum)
External auditory canal
Middle ear cavity
Vestibular nerve
Cochlear nerve
Round window
Eustachian tube
b
Reissner's membrane
Scala vestibuli (perilymph)
Stria vascularis
Spiral ganglion
Scala media (endolymph)
Basilar membrane
Scala tympani (perilymph)
c
Tectorial membrane
Hair bundle
Inner hair cell
Outer hair cell
Basilar membrane
d
Direction of propagation
Towards base
Towards apex
2 nm
2 mm
Nature Reviews | Neuroscience

Corti organ

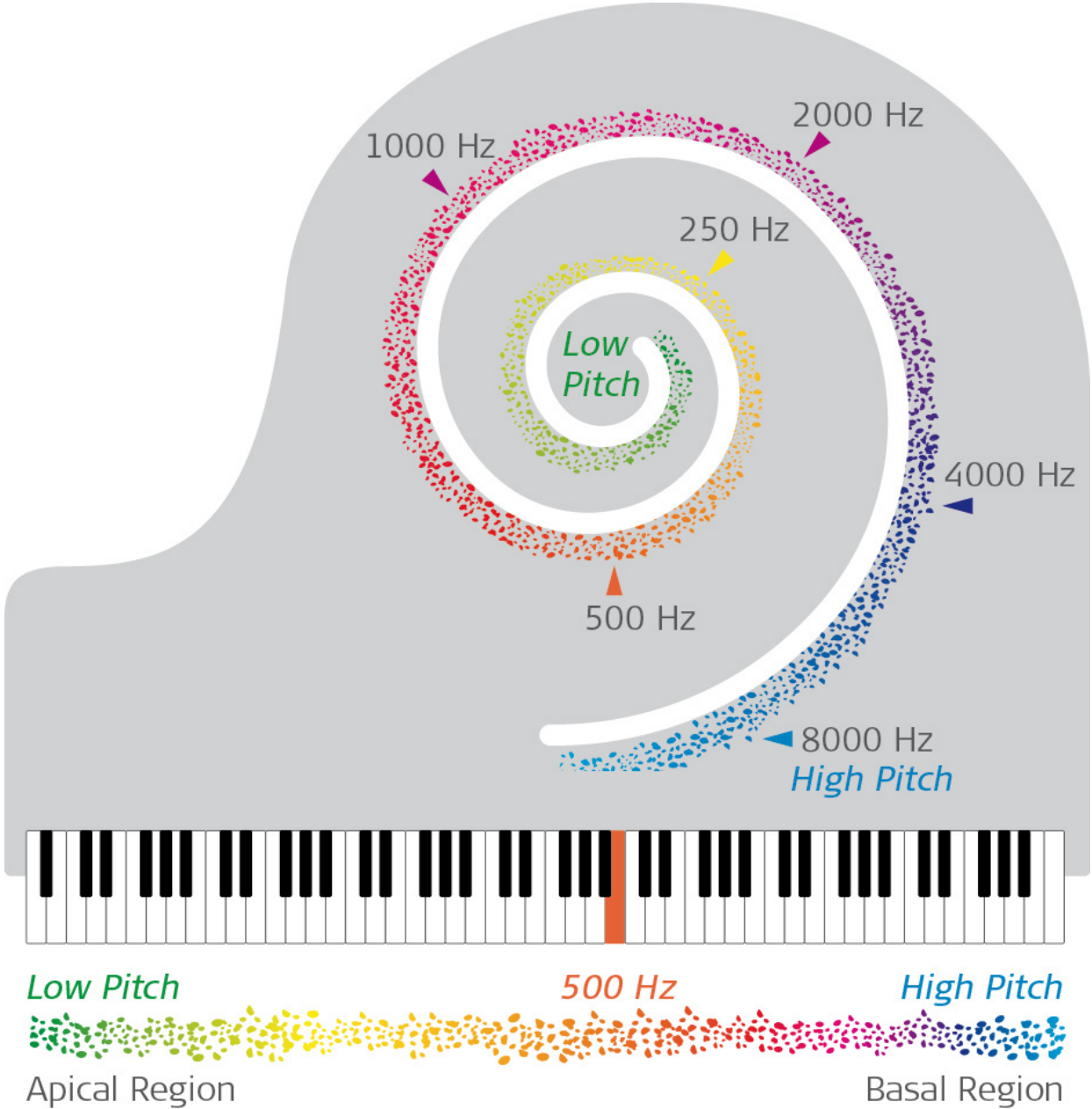

onde
+ 30 mV
— 70 m V
Membrana di Reissner
Membrana tettoria
Rampa media
Cellule cigliate
Nervo uditivo
Galleria di Corti

Basilar membrane



Bank of filters

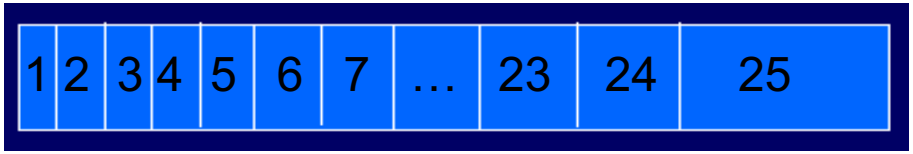
Basilar membrane



Herman
von Helmholtz

Basilar membrane

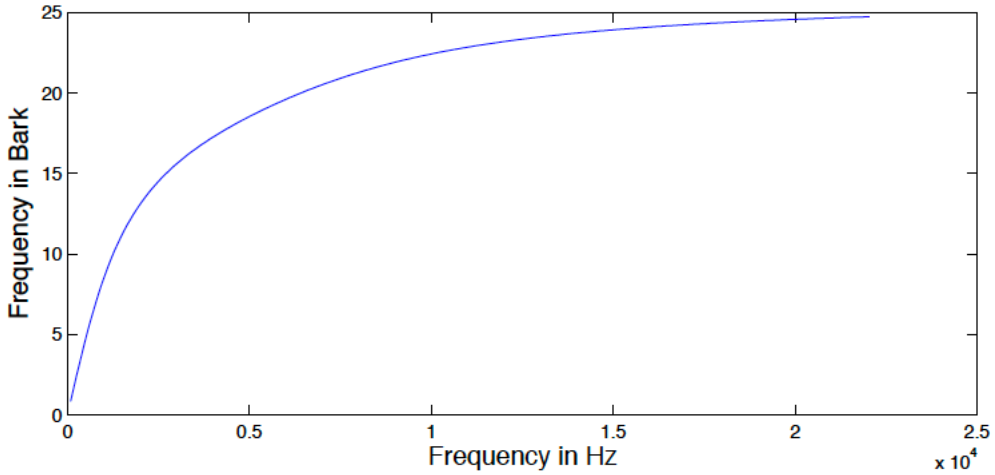
- Basilar membrane
  - 25 critical bands
  - non-linear behaviour (logartitmic)
  - pass-band bank filters



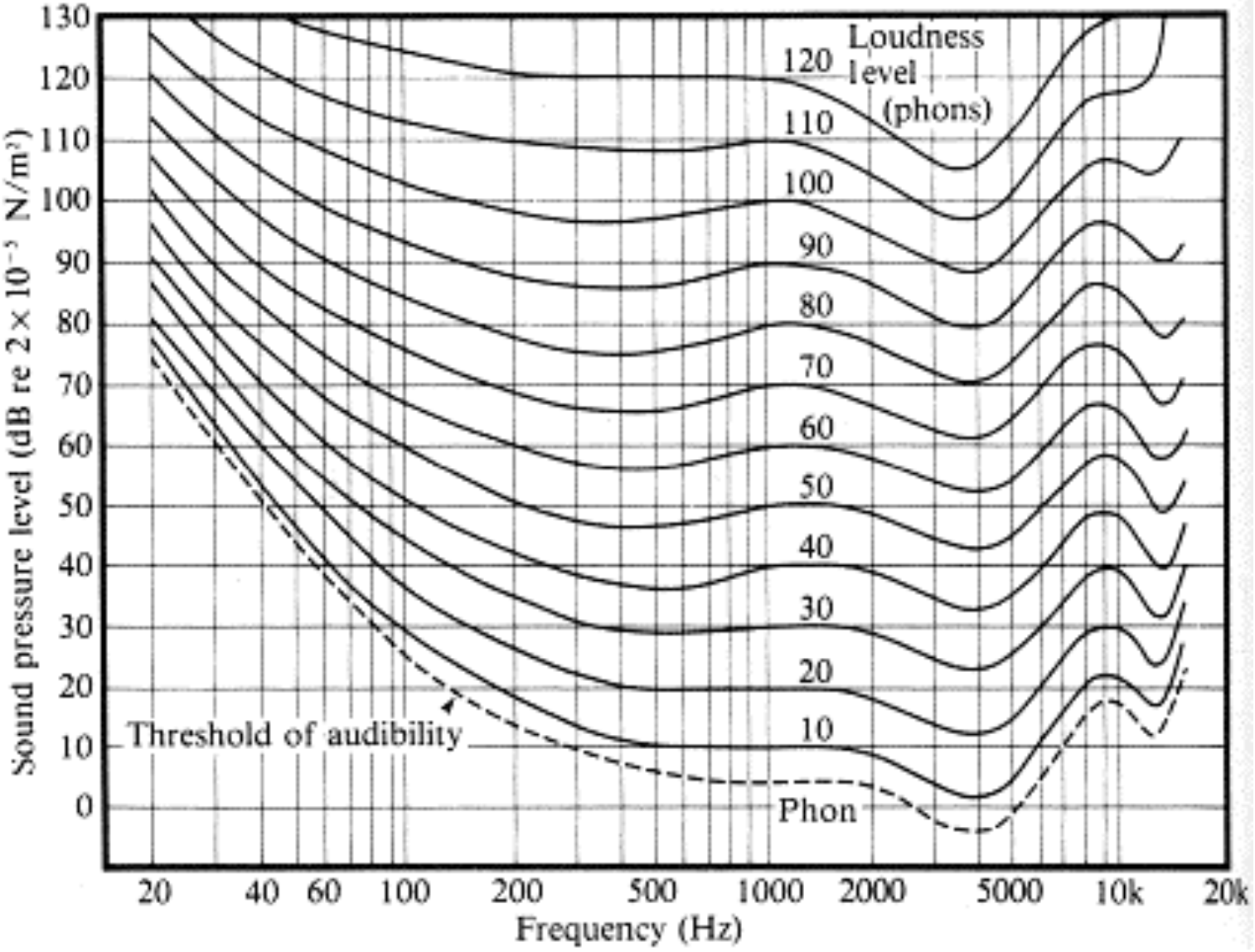
Non-linear basilar membrane critical bands

Bark scale

- Bark scale
  - $f < 500\ Hz$ $\quad f_{bark} = \frac{f}{100}$
  - $f \geq 500\ Hz$ $\quad f_{bark} = 9 + 4\log\frac{f}{100}$

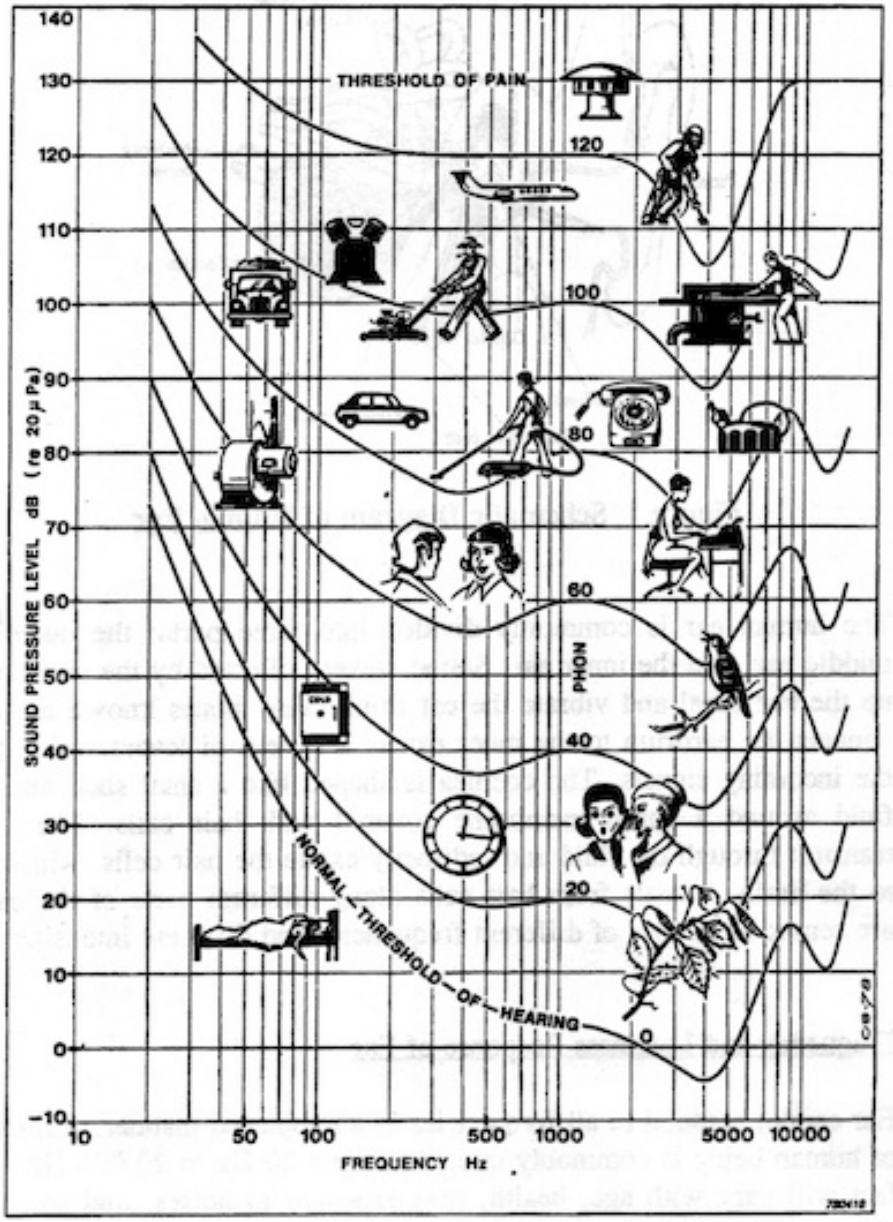


Fletcher – Munson Diagram



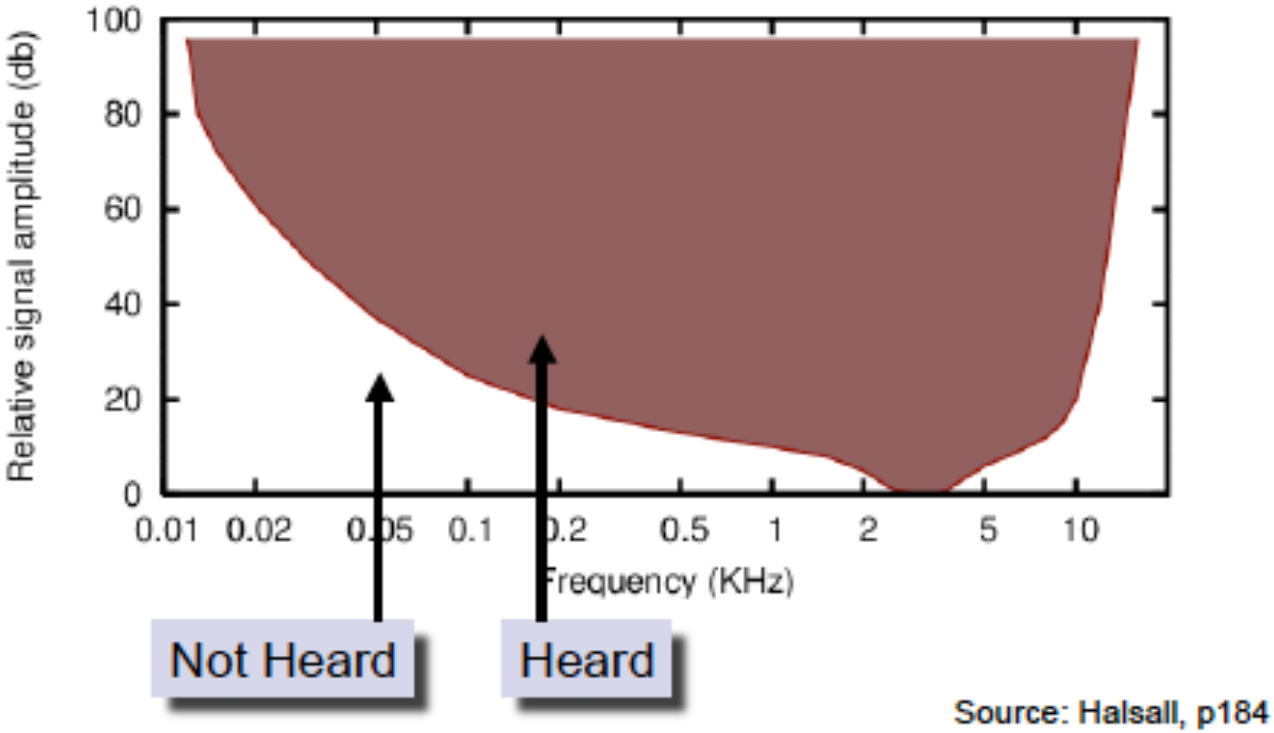
Equal-loudness contours for the human ear determined experimentally

Fletcher – Munson Diagram



Example of phones

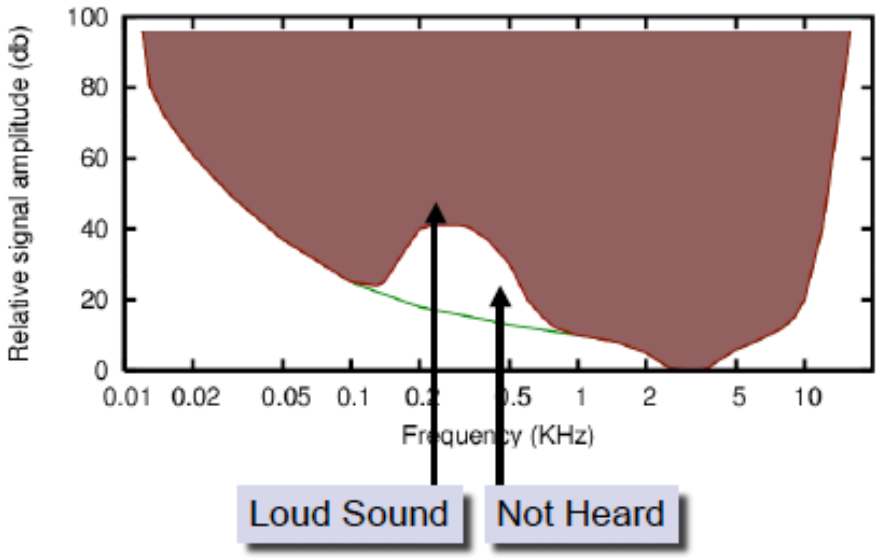
Hearing threshold



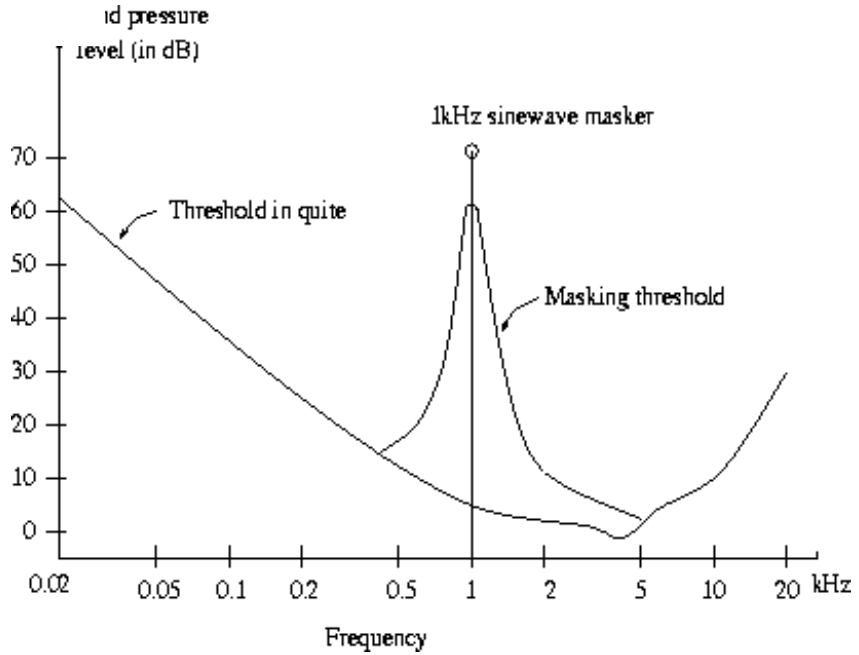
Source: Halsall, p184

Normal Treshold of Hearing

Simultaneous masking

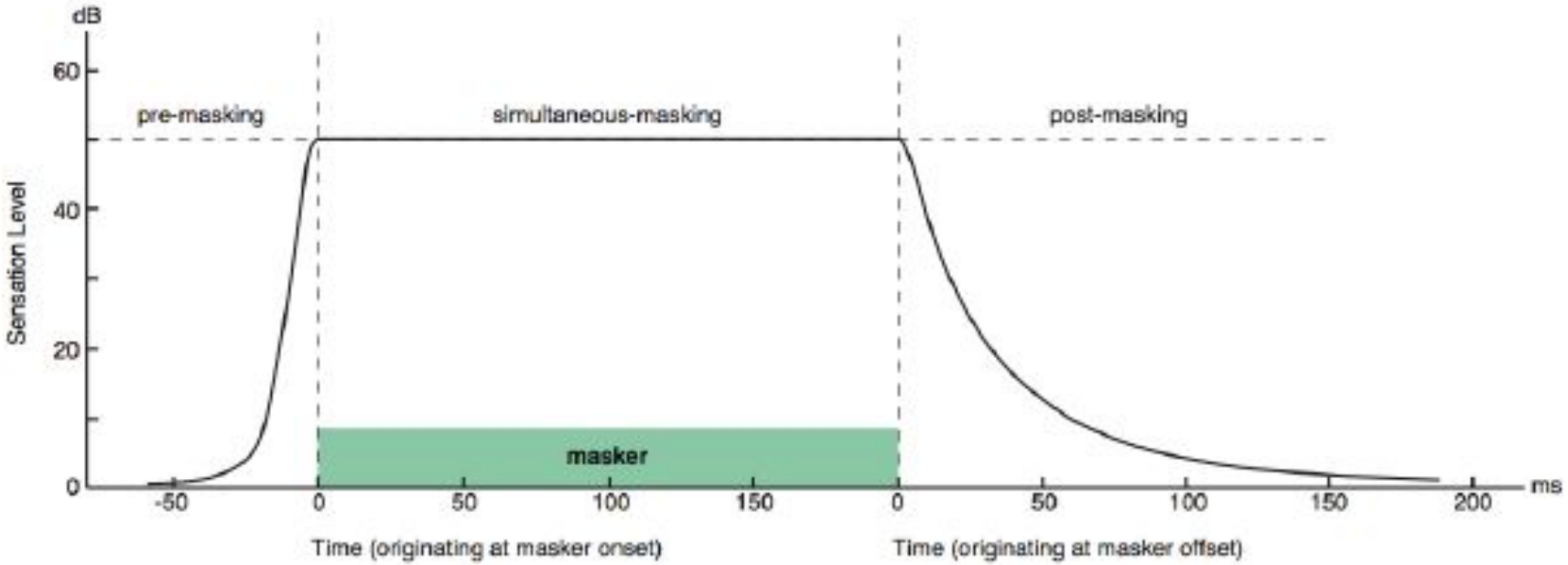


Masking threshold



Example of simultaneous masking

Temporal masking



Example of temporal masking

MPEG Audio Encoders

- Perceptual coding
  - unnecessary information eliminated
  - psychoacoustic model
  - masking mechanism
    - number of quantization bits

MPEG

- Moving Picture Experts Group (MPEG)
  - working group of authorities that was formed by ISO and IEC
  - standards for audio and video compression and transmission
  - established in 1988 by the initiative of
    - Hiroshi Yasuda (Nippon Telegraph and Telephone)
    - Leonardo Chiariglione
  - The first meeting was in May 1988 in Ottawa, Canada

MPEG standards

- MPEG-1 (1993)
  - *Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s*
  - ISO/IEC 11172
- MPEG-2 (1995)
  - *Generic coding of moving pictures and associated audio information*
  - ISO/IEC 13818
- MPEG-3
  - *standardizing scalable and multi-resolution compression*
  - intended for HDTV compression
  - was merged with MPEG-2

MPEG standards

- MPEG-4 (1998)
  - *Coding of audio-visual objects*
  - ISO/IEC 14496
- MPEG-7 (2002)
  - *Multimedia content description interface*
  - ISO/IEC 15938
- MPEG-21 (2001)
  - *Multimedia framework*
  - ISO/IEC 21000

MPEG standards

- MPEG-A (2007)
  - *Multimedia application format*
  - ISO/IEC 23000
  - e.g., MPEG music player application format, MPEG photo player application
- MPEG-B (2006)
  - *MPEG systems technologies*
  - ISO/IEC 23001
  - e.g., Binary MPEG format for XML, Fragment Request Units, Bitstream Syntax Description Language (BSDL)

MPEG standards

- MPEG-C (2006)
  - *MPEG video technologies*
  - ISO/IEC 23002
  - e.g., accuracy requirements for implementation of integer-output 8x8 inverse discrete cosine transform

- MPEG-D (2007)
  - *MPEG audio technologies*
  - ISO/IEC 23003
  - e.g., MPEG Surround, SAOC-Spatial Audio Object Coding and USAC-Unified Speech and Audio Coding

MPEG standards

- MPEG-E (2007)
  - *Multimedia Middleware*
  - ISO/IEC 23004
  - e.g., Architecture, Multimedia application programming interface (API), Component model

- MPEG-V (2011)
  - *Media context and control*
  - ISO/IEC 23005
  - e.g., Avatar characteristics, Sensor information, Architecture

MPEG standards

- MPEG-M (2010)
  - *MPEG eXtensible Middleware* (MXM)
  - ISO/IEC 23006
  - e.g., MXM architecture and technologies, API, MPEG extensible middleware (MXM) protocols
- MPEG-U (2010)
  - *Rich media user interfaces*
  - ISO/IEC 23007
  - e.g., Widgets

MPEG standards

- MPEG-H (2013)
  - *High Efficiency Coding and Media Delivery in Heterogeneous Environments*
  - ISO/IEC 23008
  - Part 1 – MPEG media transport; Part 2 – High Efficiency Video Coding; Part 3 – 3D Audio
- MPEG-DASH (2012)
  - *Information technology – Dynamic adaptive streaming over HTTP* (DASH)
  - ISO/IEC 23009
  - Media presentation description and segment formats

MPEG-1

- MPEG-1
  - standard for lossy compression of video and audio
  - designed to compress VHS-quality raw digital video and CD audio down to 1.5 Mbit/s (26:1 and 6:1 compression ratios respectively)
  - without excessive quality loss
    - video CDs
    - digital cable/satellite TV
    - digital audio broadcasting (DAB)

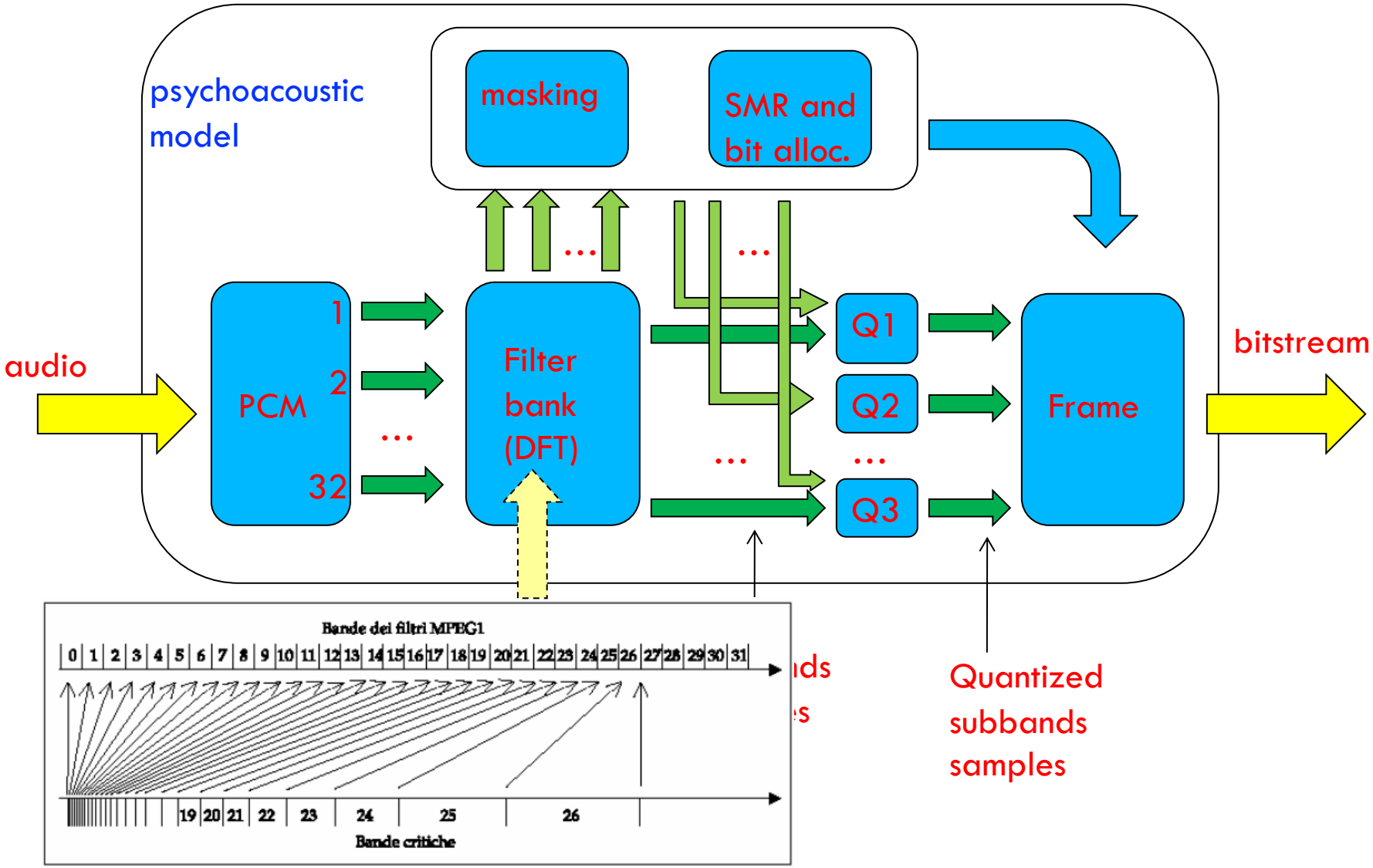
MPEG-1

- The standard consists of five Parts
  - ISO/IEC 11172-1 (1993)
    - System
  - ISO/IEC 11172-2 (1993)
    - Video
  - ISO/IEC 11172-3 (1993)
    - Audio
  - ISO/IEC 11172-4 (1995)
    - Compliance Testing
  - ISO/IEC TR 11172-5 (1998)
    - Software simulation

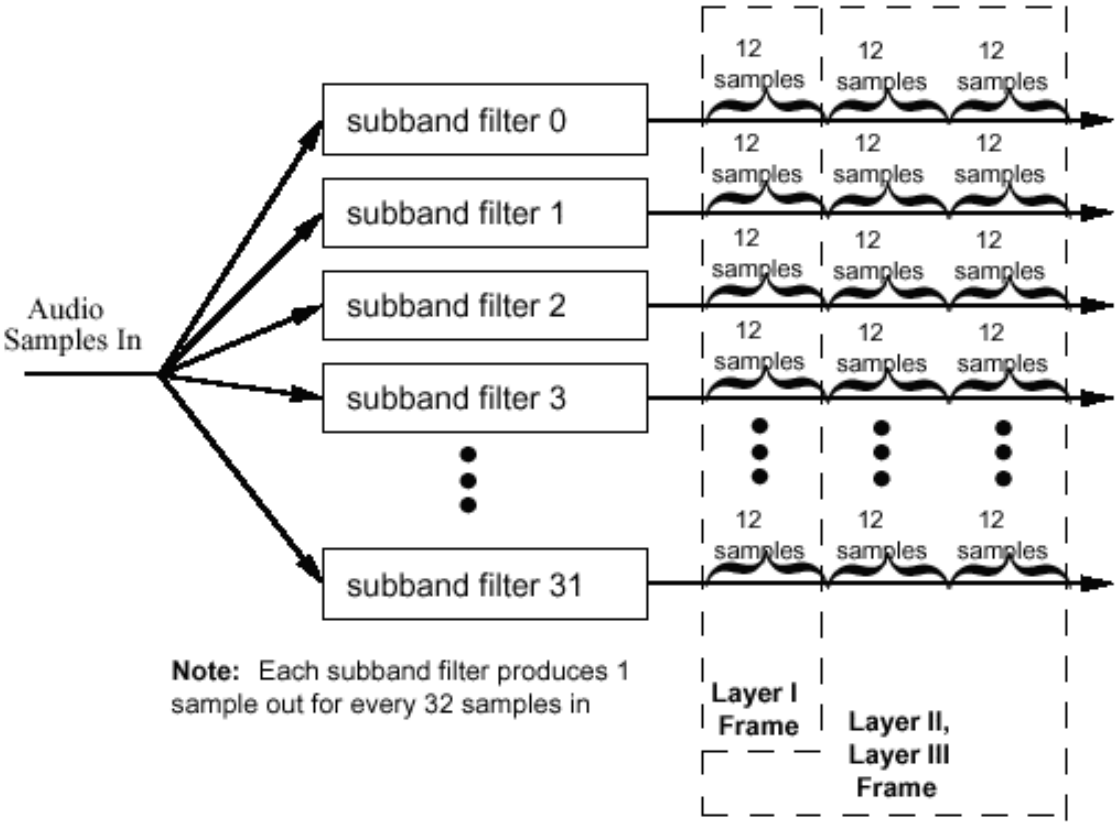
MPEG 1

- MPEG 1
  - Layer 1
    - Compressed bit rate: 32-448 Kbps
  - Layer II
    - Compressed bit rate: 32-192 Kbps
  - Layer III
    - Compressed bit rate: 64 Kbps
    - mp3

MPEG 1 – Layer 1



MPEG 1 – Layer 1



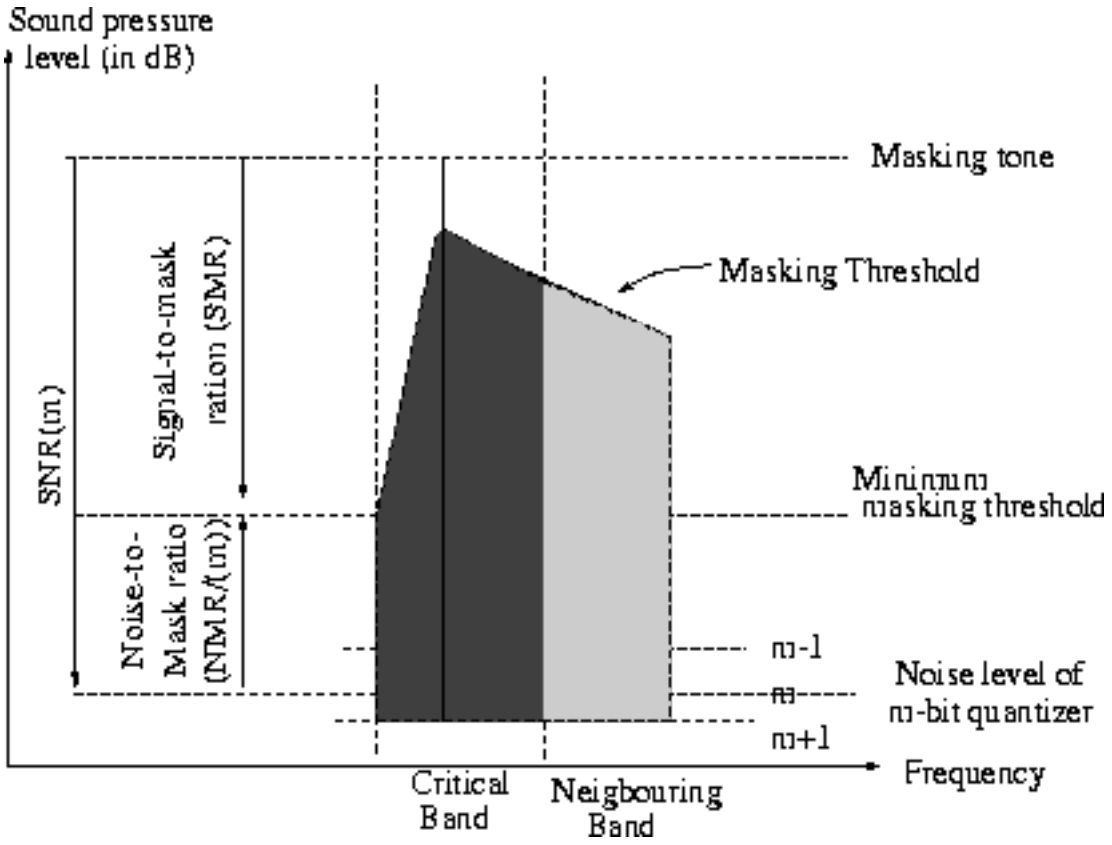
Definition of segments

Psychoacoustic model

- A 1024 points FFT is used

- Global masking thresholds

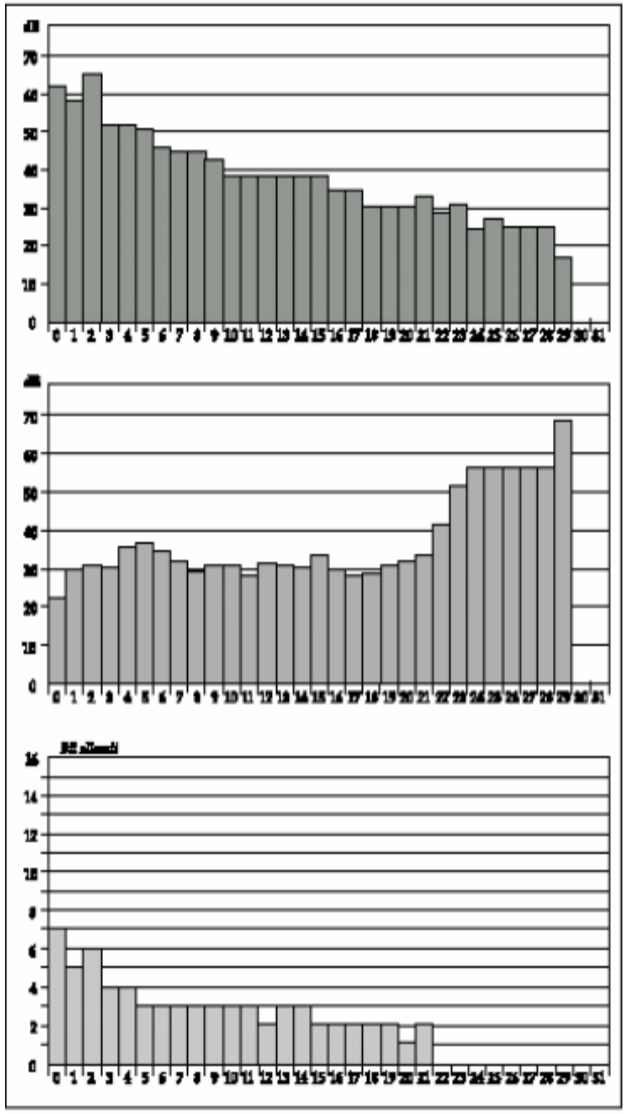
Global bit allocation



Bit allocation for each subband (NMR = SNR - SMR)

Then the subbands are placed in order of lowest to highest mask-to-noise ratio, and the lowest subband is allocated the smallest number of code bits and this process continues until no more code bits can be allocated

Global bit allocation

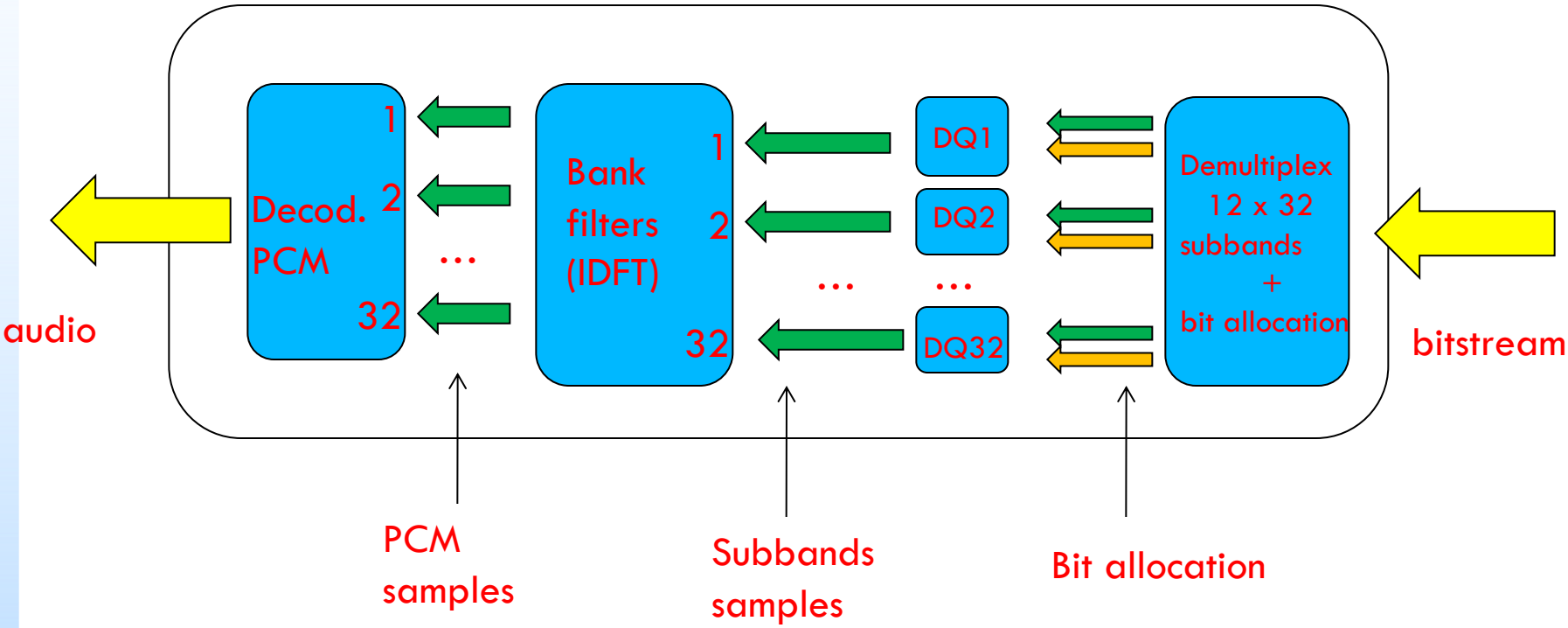


Amplitude spectrum

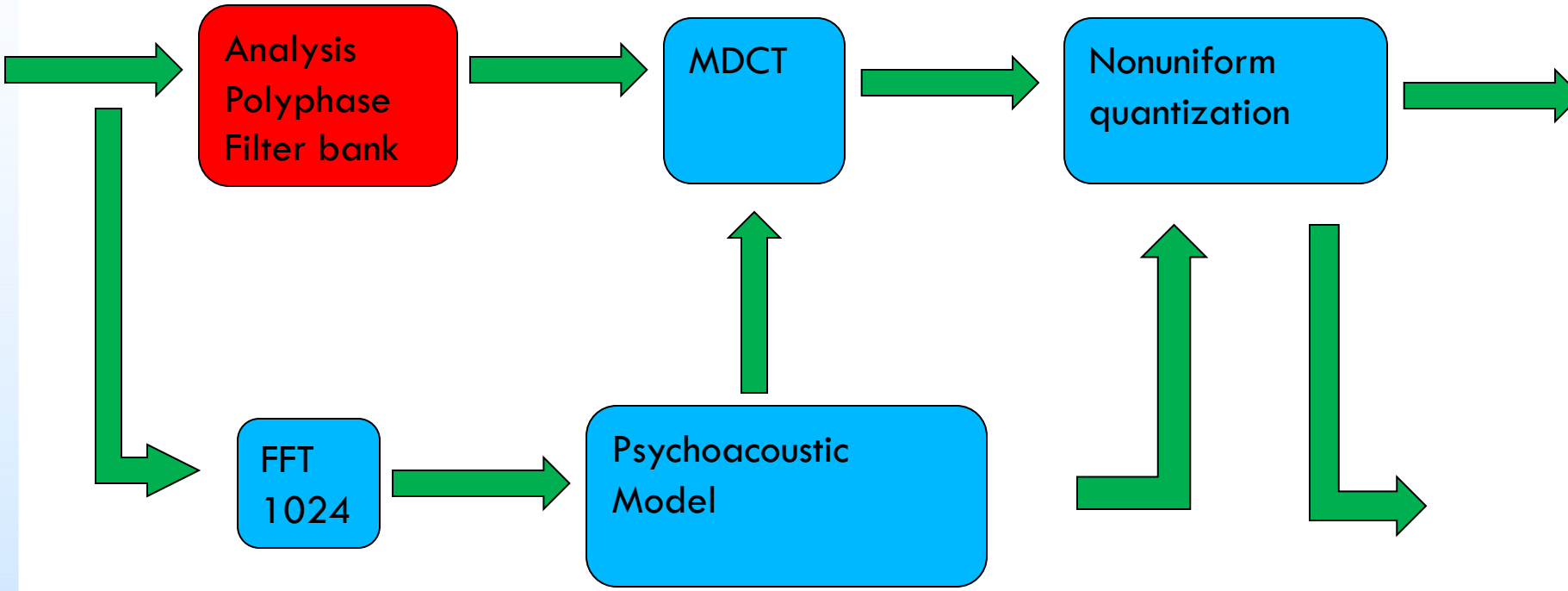
Global masking threshold

Subbands bit allocation

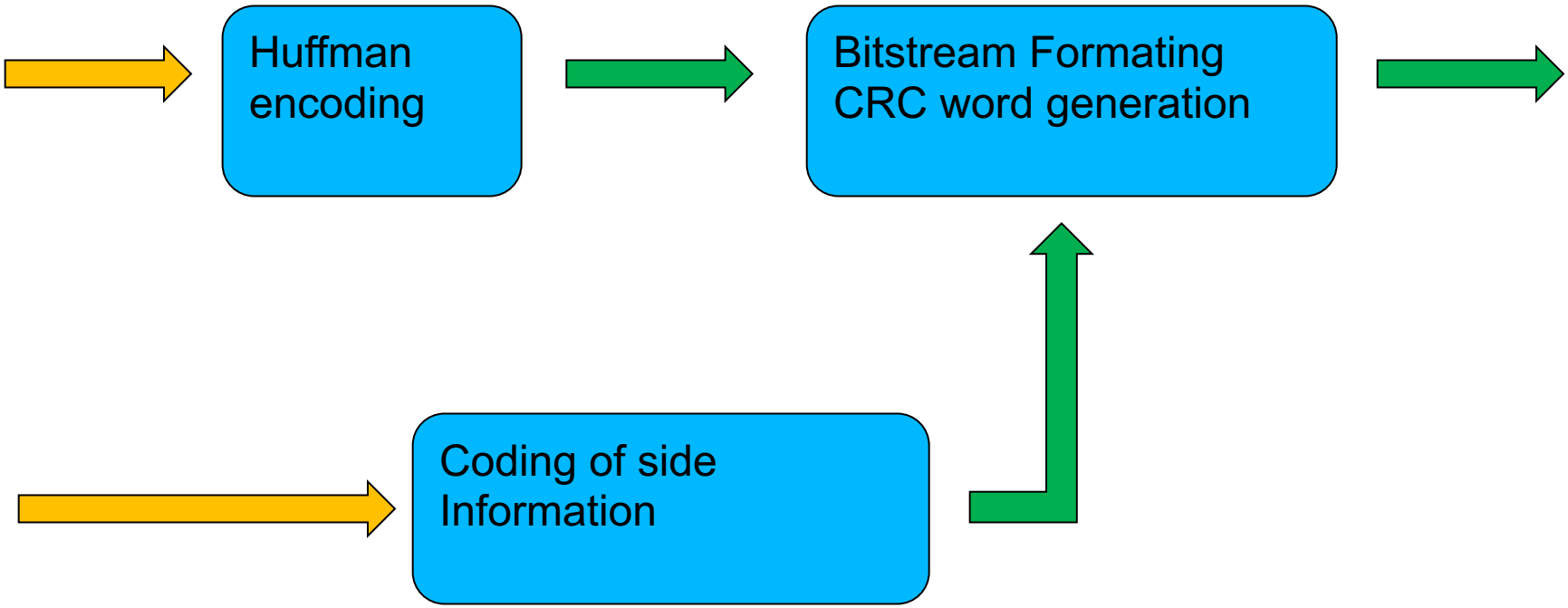
Decoder


audio
Decod. PCM
1
2
32
Bank filters (IDFT)
1
2
32
DQ1
DQ2
DQ32
Demultiplex 12 x 32 subbands + bit allocation
bitstream
PCM samples
Subbands samples
Bit allocation

MPEG - Layer 3

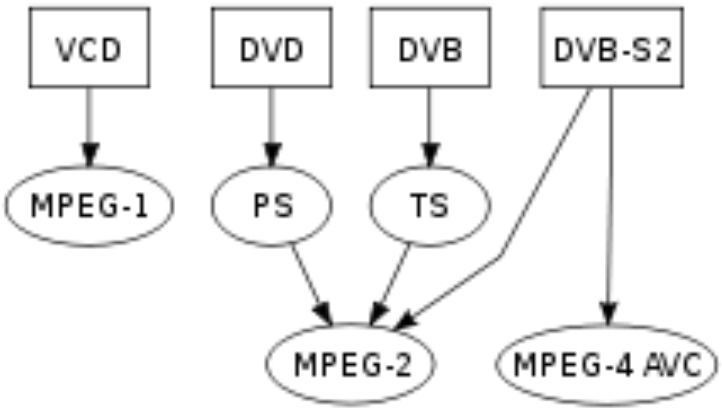

Analysis
Polyphase
Filter bank
MDCT
Nonuniform
quantization
FFT
1024
Psychoacoustic
Model

MPEG – Layer 3


Huffman encoding
Bitstream Formating CRC word generation
Coding of side Information

MPEG-2

- MPEG-2
  - generic coding of moving pictures and associated audio information
  - combination of lossy video compression and lossy audio data compression methods
  - storage and transmission of movies using currently available storage media and transmission bandwidth



MPEG-2 is used in Digital Video Broadcast and DVDs.
The MPEG transport stream, TS, and MPEG program stream, PS, are container formats

MPEG-2

- The standard consists of 9 Parts
  - ISO/IEC 13818-1 (2000)
    - Systems
  - ISO/IEC 13818-2 (2000)
    - Video
  - ISO/IEC 13818-3 (1998)
    - Audio
  - ISO/IEC 13818-4 (1998)
    - Conformance Testing
  - ISO/IEC 13818-1 (1997)
    - Software simulation

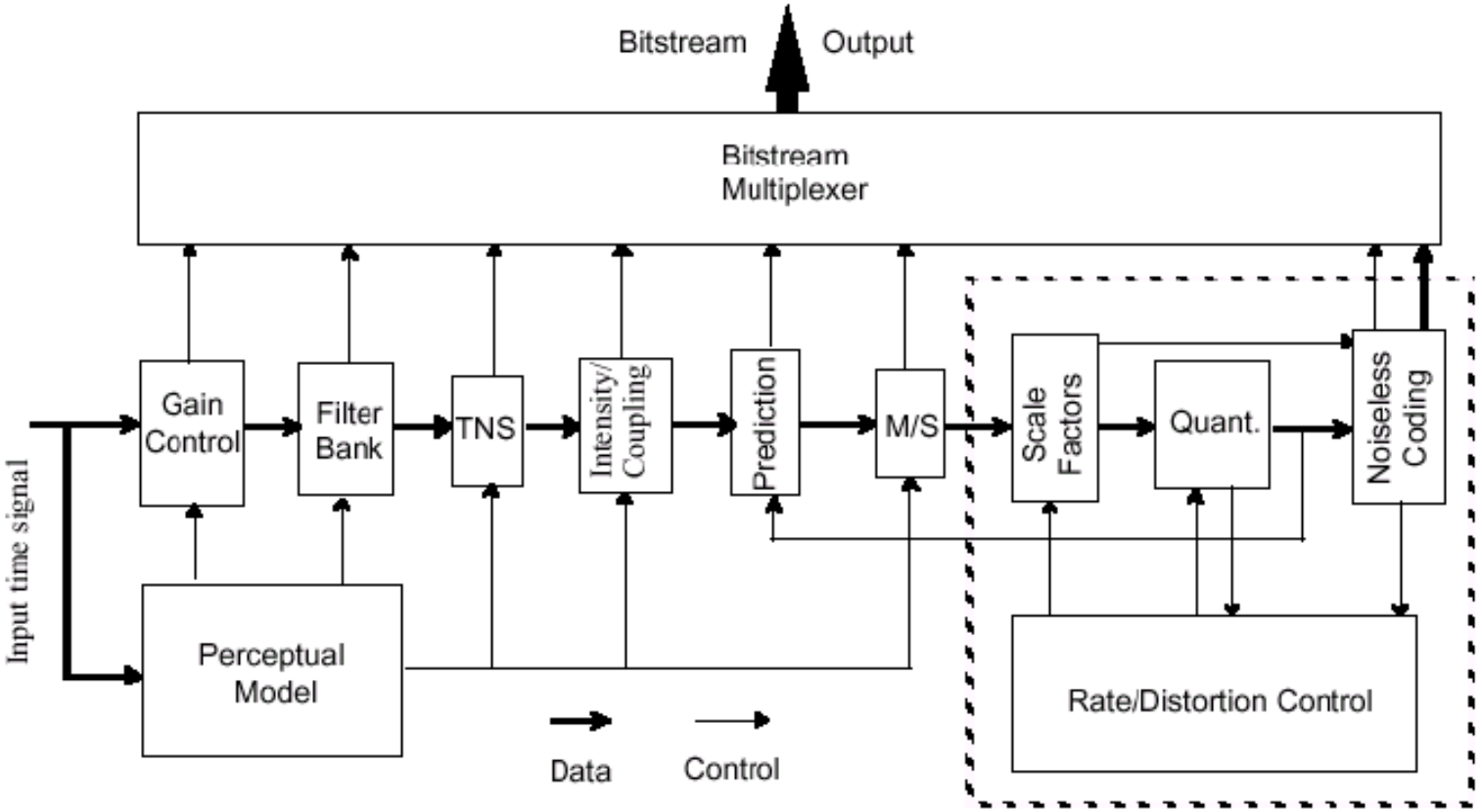
MPEG-2

- The standard consists of 9 Parts
  - ISO/IEC 13818-6 (1998)
    - Extensions for DSM-CC
  - ISO/IEC 13818-7 (1997)
    - Advanced Audio Coding (AAC
  - ISO/IEC 13818-8 (1996)
    - Extension for real time interface for systems decoders
  - ISO/IEC 13818-9 (1999)
    - Conformance extensions for Digital Storage Media Command and Control (DSM-CC)

MPEG 2 - AAC

- Adanced Audio Coding (AAC)
  - improvement for multichannel encoding
  - 48 channels
  - samplig frequency from 8 to 96 KHz for each channel

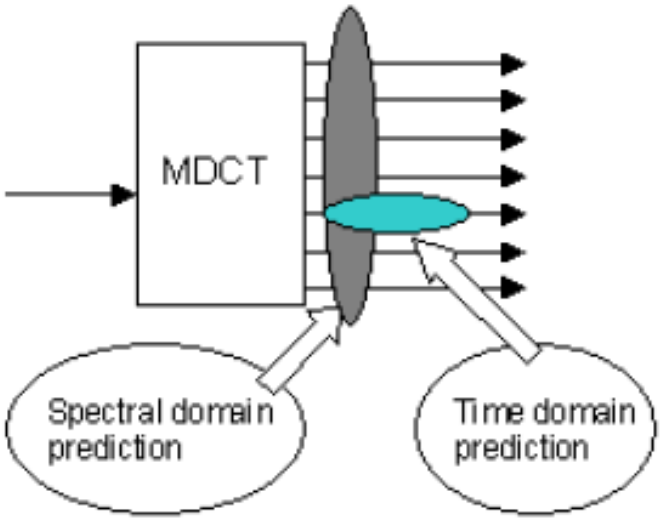
MPEG 2 - AAC



AAC encoding scheme

MPEG 2 - AAC

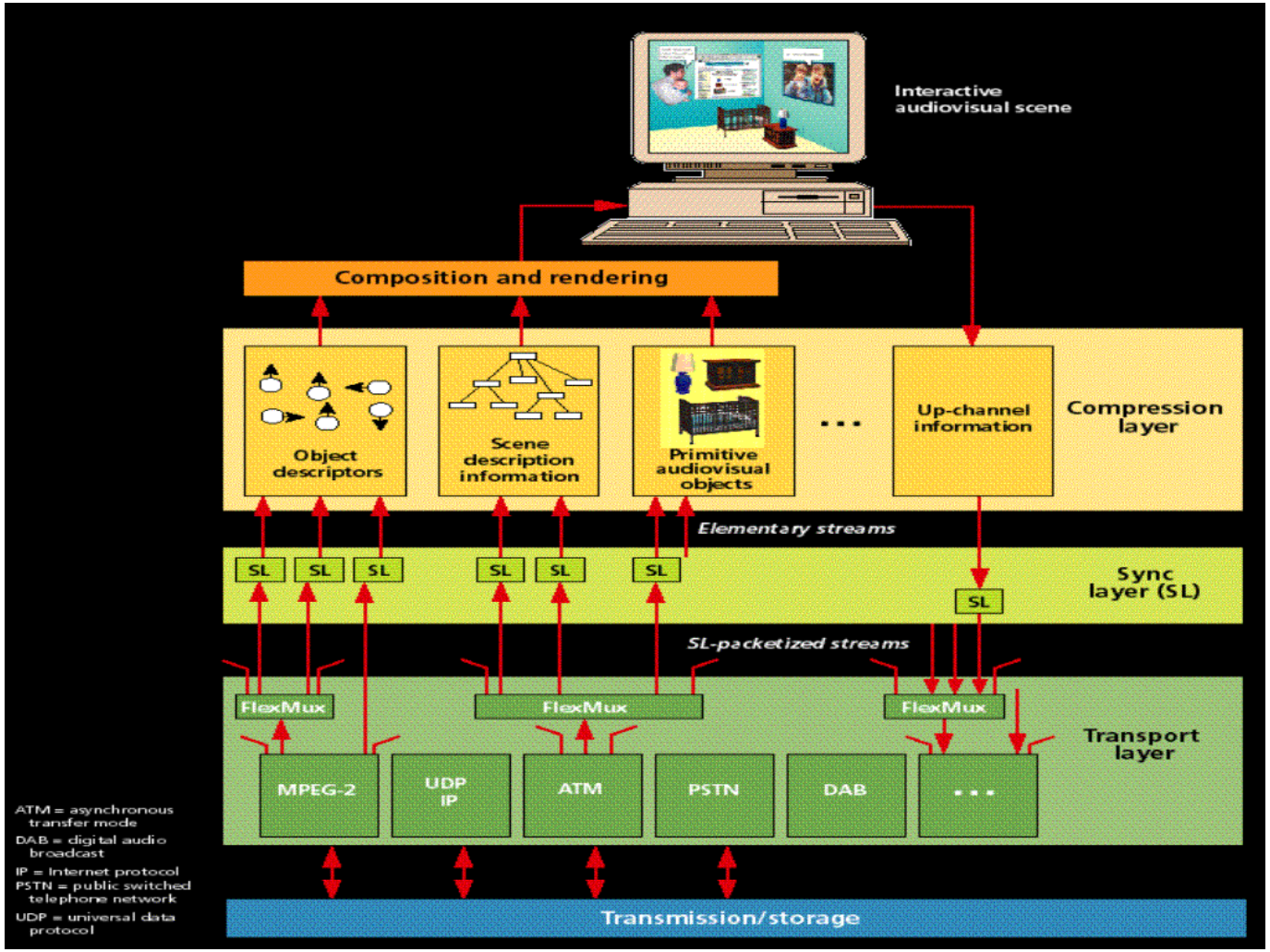
- Main concept – prediction
  - Prediction
  - Temporal Noise Shaping (TNS)



MPEG-4

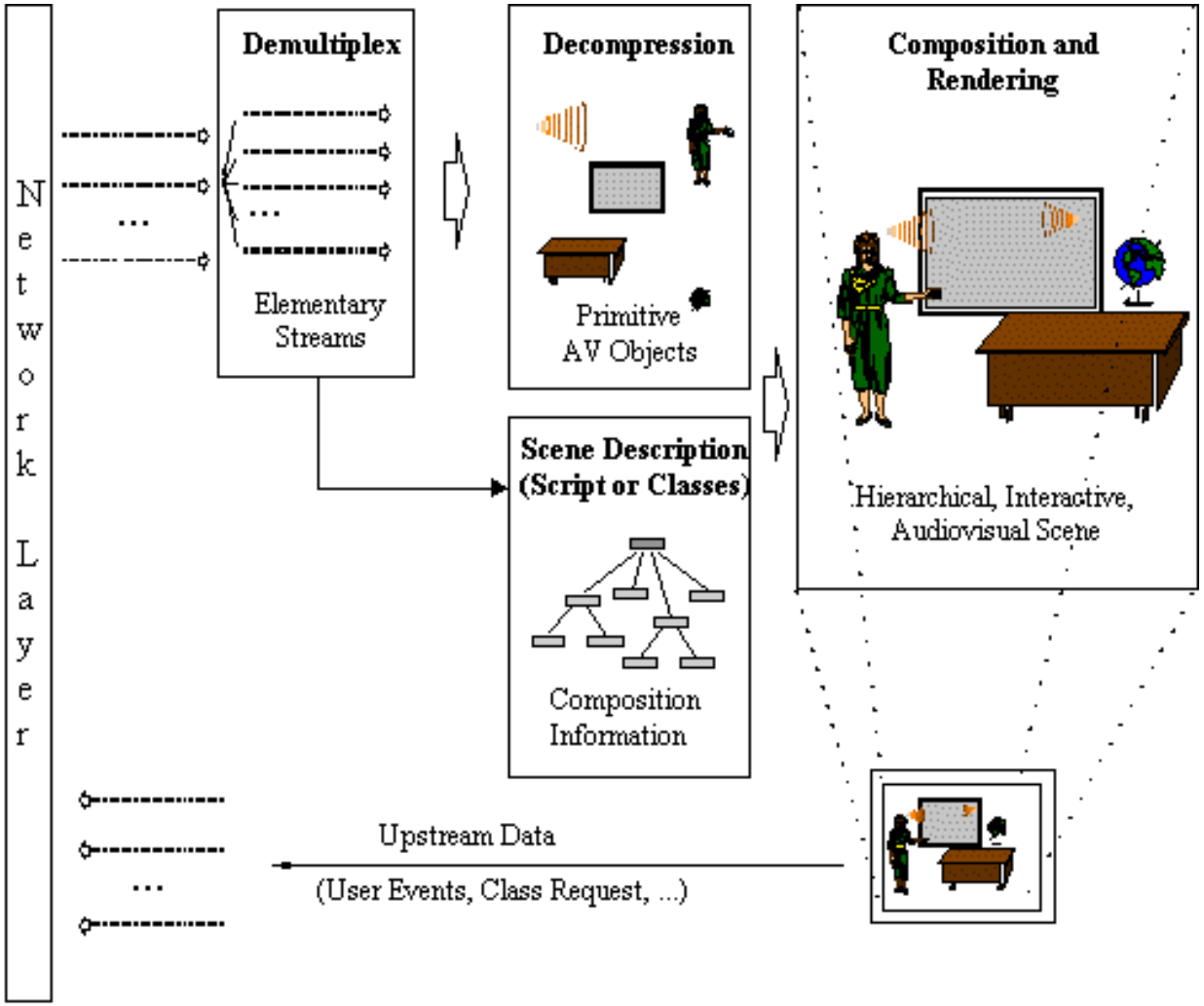
- Besides compression, it pays great attention to user interactivity
  - allows a larger number of users to create and communicate their multimedia presentations and applications on new infrastructures
    - Internet, mobile/wireless networks, …
  - adopt a new object-based coding approach
    - *media objects* are entities
    - media objects (audio and visual objects) can be either natural or synthetic
- bitrate covers a large range, between 5 kbps and 10 Mbps

MPEG-4



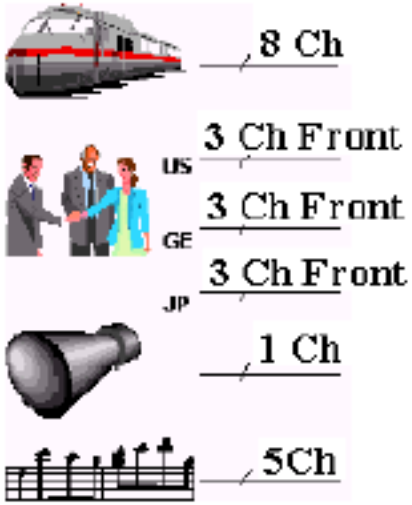
MPEG-4 components and layers

MPEG-4



Decoding, composition and rendering

MPEG 4



MPEG 4 considers each audio as an independent object

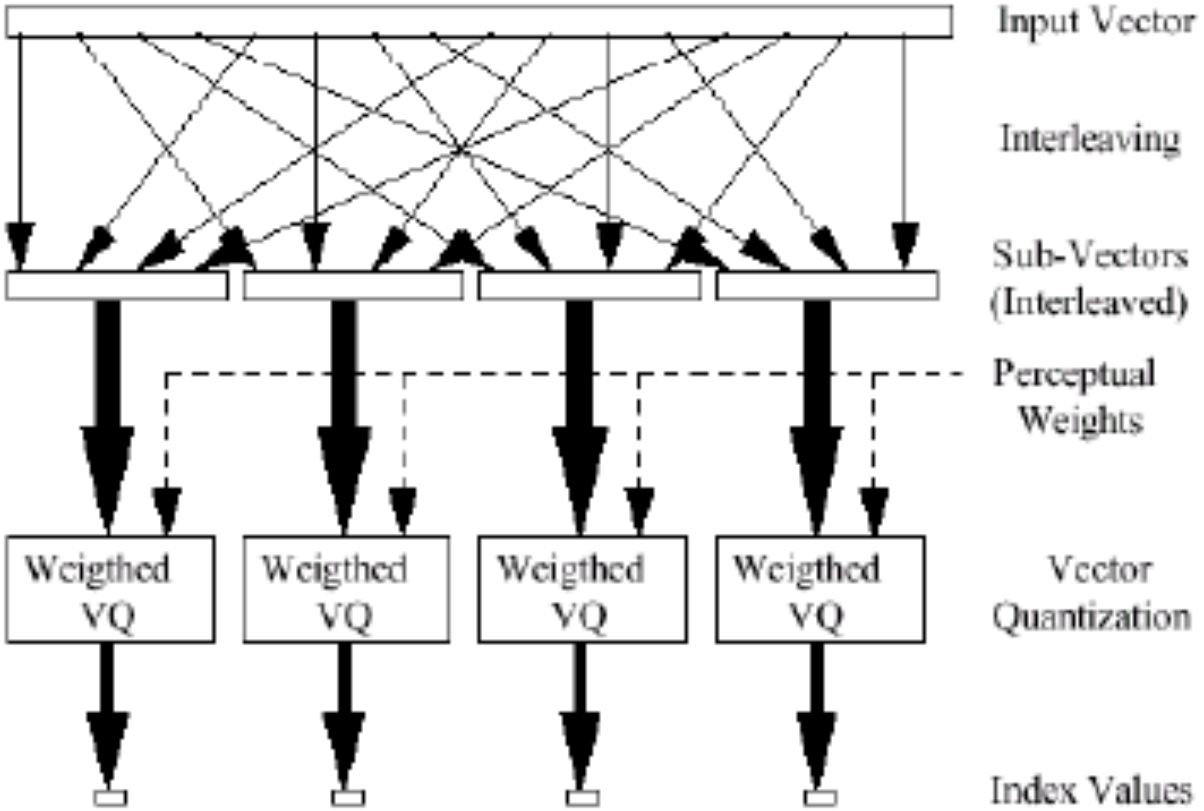
MPEG 4 - Speech Signal

- Speech signal
  - Synthesis Decoding Code Excited Linear Predictive (CELP)
    - Bitrate from 4 to 24 Kbit/s
  - Harmonic vector eXcitation Coding (HVXC)
    - Bitrate from 2 to 4 Kbit/s

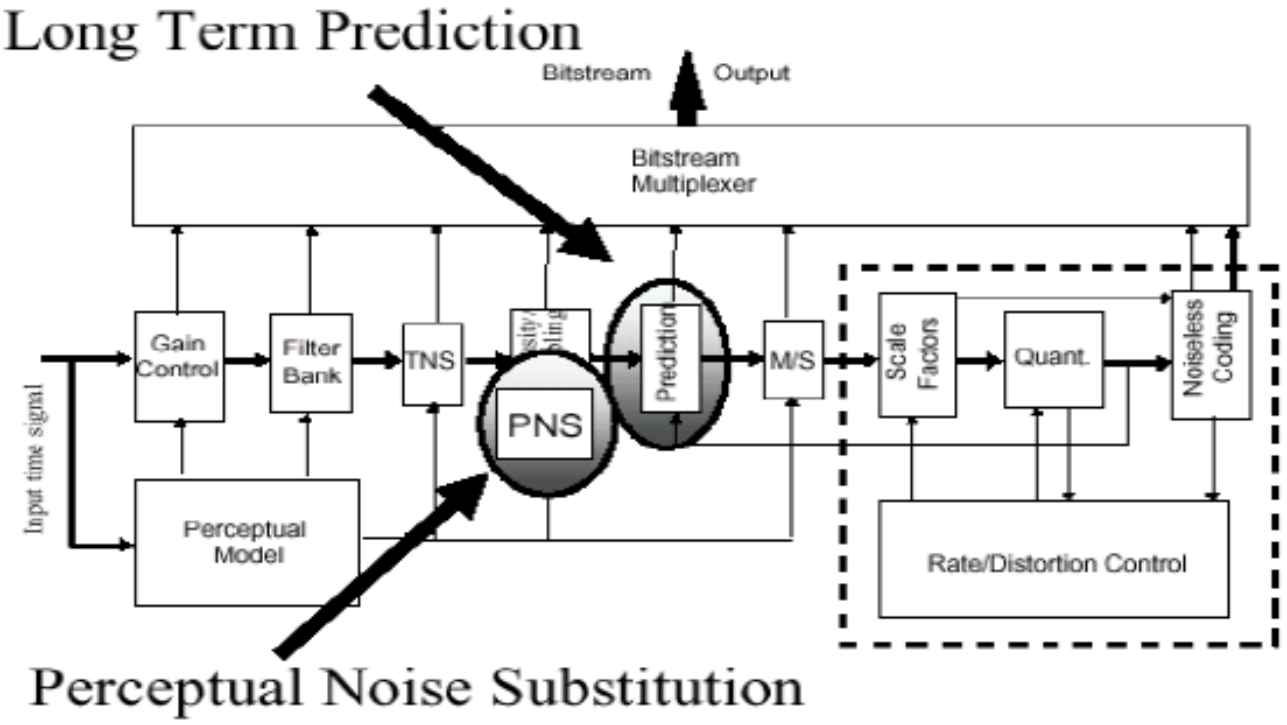
General Audio

- General Audio
  - Transform-domain Weighted Interleaved Vector Quanrization (TwinVQ)
    - less than 16 Kbit/s
  - AAC for greater bitrates

General Audio



TwinVQ scheme



TwinVQ scheme

Synthesized Speech

- Text to Speech
  - production of a sound voice from a text
  - Interface for compressed data

Synthesized Audio

- Structured Audio Orchestra Language (SAOL)
  - Set of musical instruments for reproducing
- Structured Audio Score Language (SASL)
  - what to produce

MPEG-7

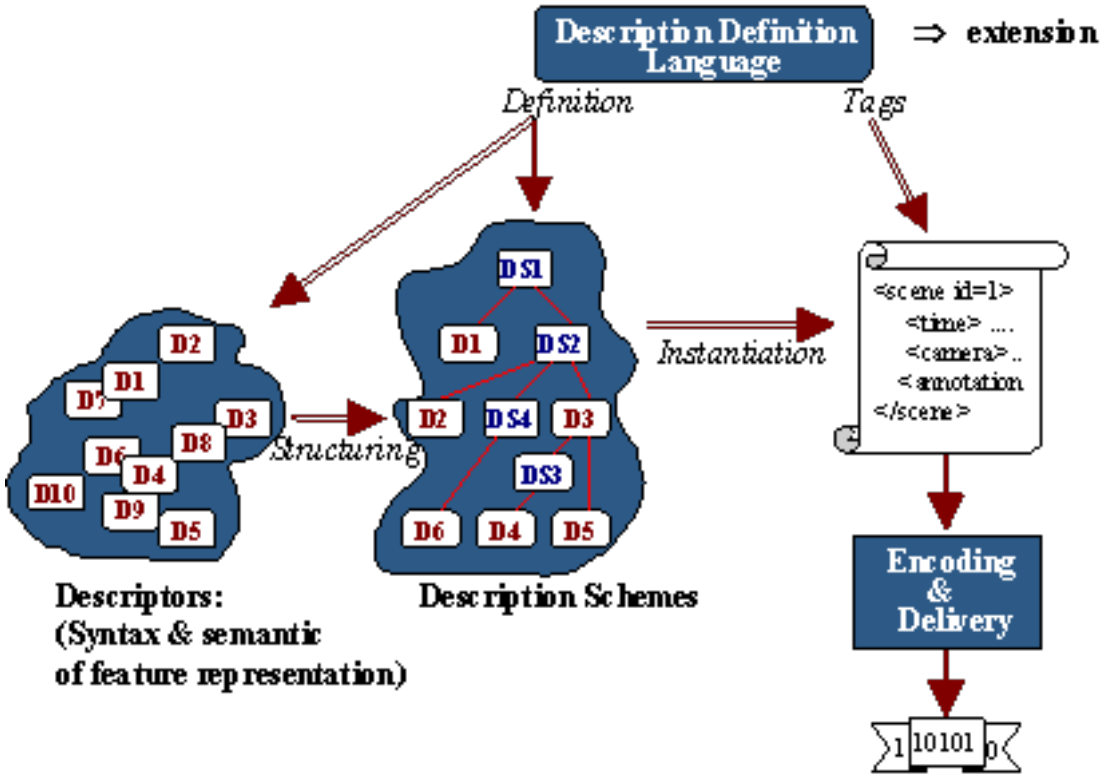
- Main objective
  - audiovisual content-based retrieval (or audiovisual object retrieval)
  - applications such as digital libraries
  - supports a variety of multimedia applications
    - pictures, graphics, 3D models, audio, speech, video, composition information, …

MPEG-7

- Descriptor (D)
  - low-level features
    - color,texture, shape, and motion
  - high-level features of semantic objects
    - events and abstract concepts
- Description Scheme (DS)
  - Specification of the structure and relationship between Ds and DSs
- Description Definition Language (DDL)
  - Syntactic rules to express and combine DSs and Ds

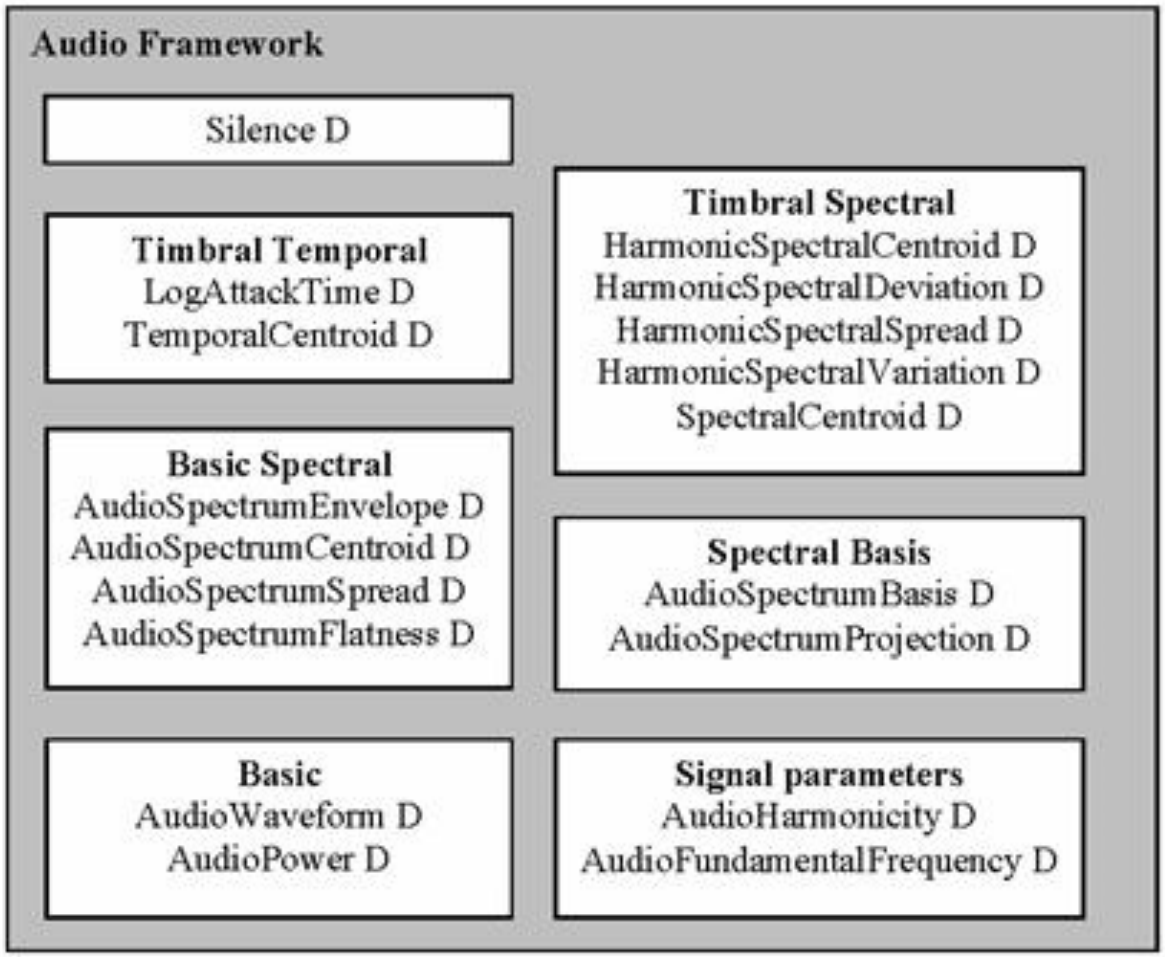
MPEG-7

Based on XML



Main parts of MPEG-7 standard

MPEG-7

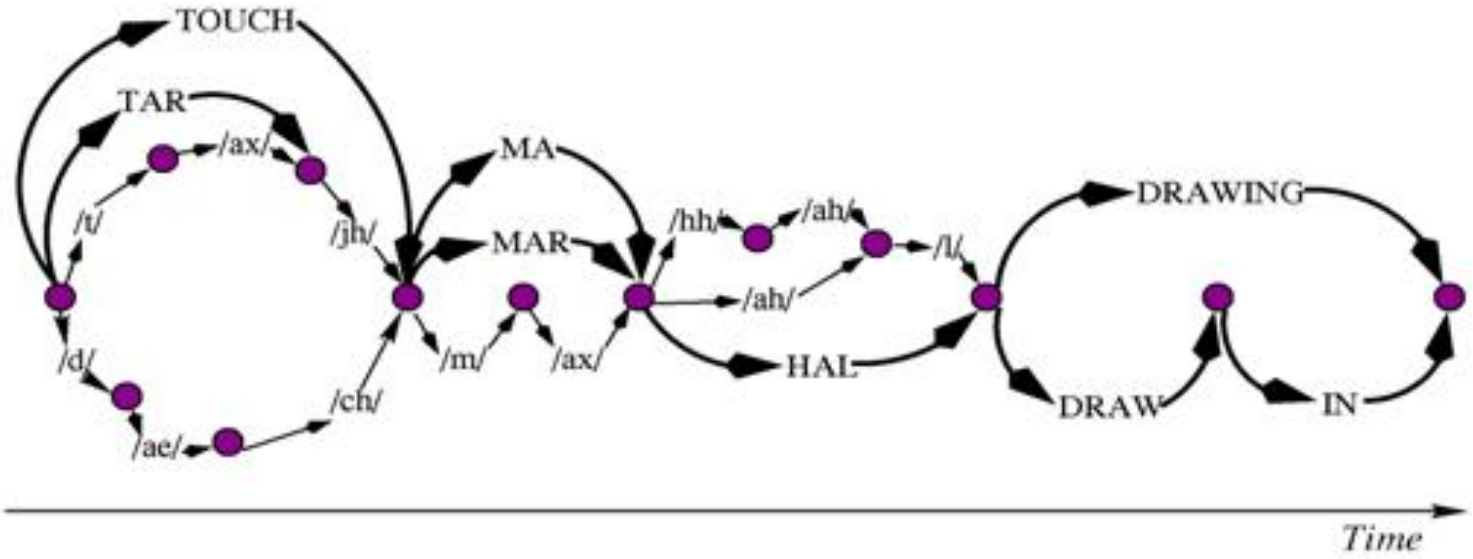


Audio DS

MPEG-7

- High level Audio DS
  - Audio Signature Description Scheme
  - Musical Instrument Timbre Description Tools
  - Melody Description Tools
  - General Sound Recognition and Indexing Description Tools
  - Spoken Content Description Tools

MPEG-7



Speech and DS

Dolby Digital

- Cinematographic encoding
  - Batman in 1992
- Home Theather (e.g., 5.1 Dolby System)
  - 5 channels
    - left, right, middle, left surround, right surround
  - Low Frequency Effects (LFE)
    - 1/10 sampling of the other channels

AC-3 Dolby Digital

- AC-3
  - perceptual digital audio coding technique that reduces the amount of data needed to produce high-quality sound
  - coding system designed specifically for multichannel digital audio