Digital audio & video, part I

- Introduction
- Digital audio
  + Psycho acoustics
  + Digital presentation of sound
- Digital images
  + JPEG

Introduction

- Multimedia uses often sound, images, and videos from natural sources
- First, the source has to be converted into signal
  + microphone, camera, video camera
- Next, analog signals are converted into digital
  + sampling, A/D-transformation
- Often, amount of the information is also reduced
  + compression

Introduction (cont.)

- Compression method can be *lossless* or *lossy*
- Compressed information is easier to store and transfer
- Compressed information has to be decompressed before use
- Digital to analog conversion has also to be done
- After this, the signal can be played (or shown) to the user

Digital audio application areas

- Computer generated sound
- Sound storage and processing
- Digital communications
- Answering service
- Speech synthesis
- Speech recognition
- Computerized call center
- Presentation of data as sound (Sonification)

Psycho acoustics

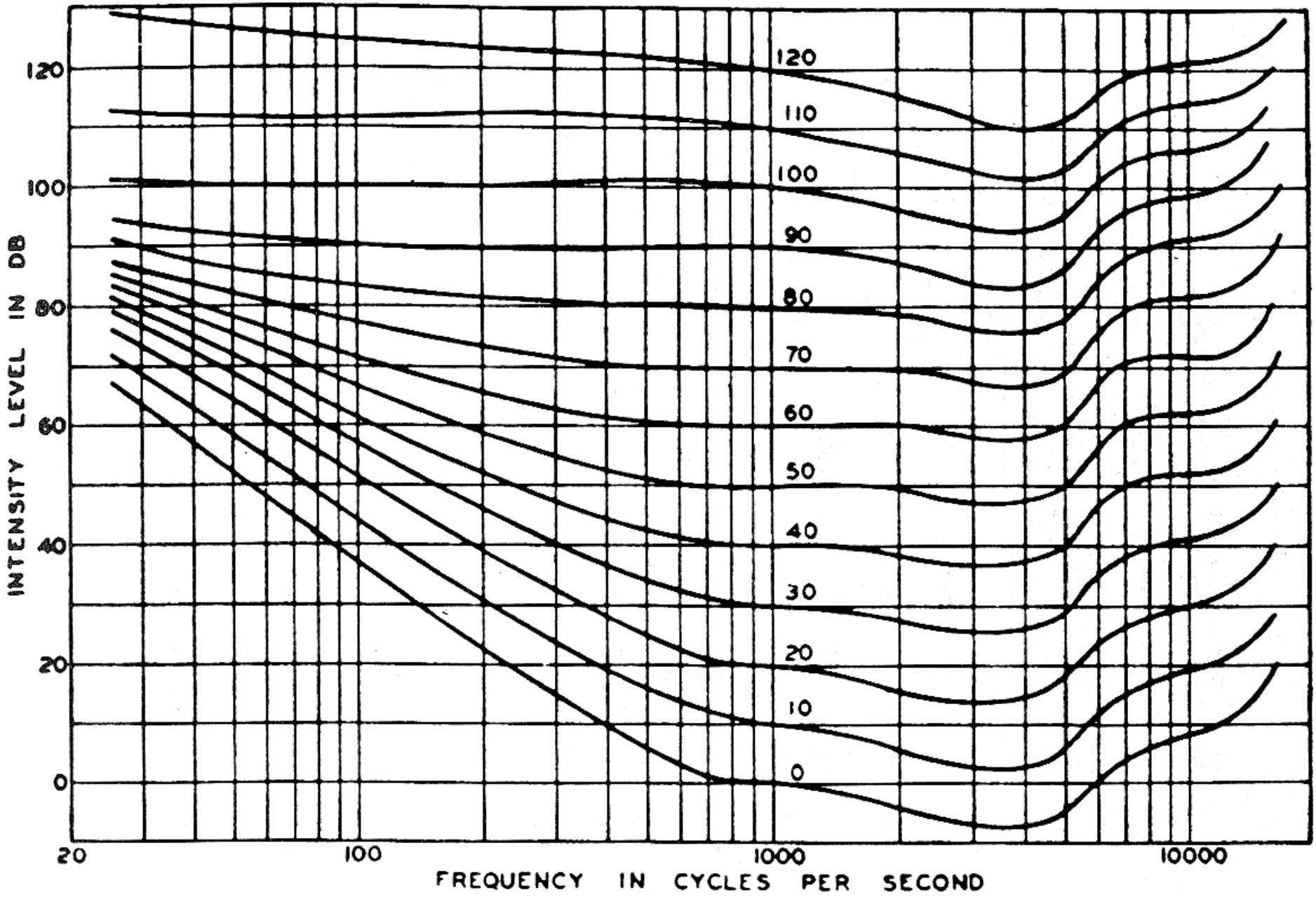
- Frequency band
- Dynamic range
- Frequency properties
- Time effect
- Masking
- Phase
- Binaural hearing and localization

Frequency band

- Humans can hear frequencies 20 Hz - 20 kHz
- Older people have much narrower range
- The frequency that we hear can be different than the physical frequency

Dynamic range

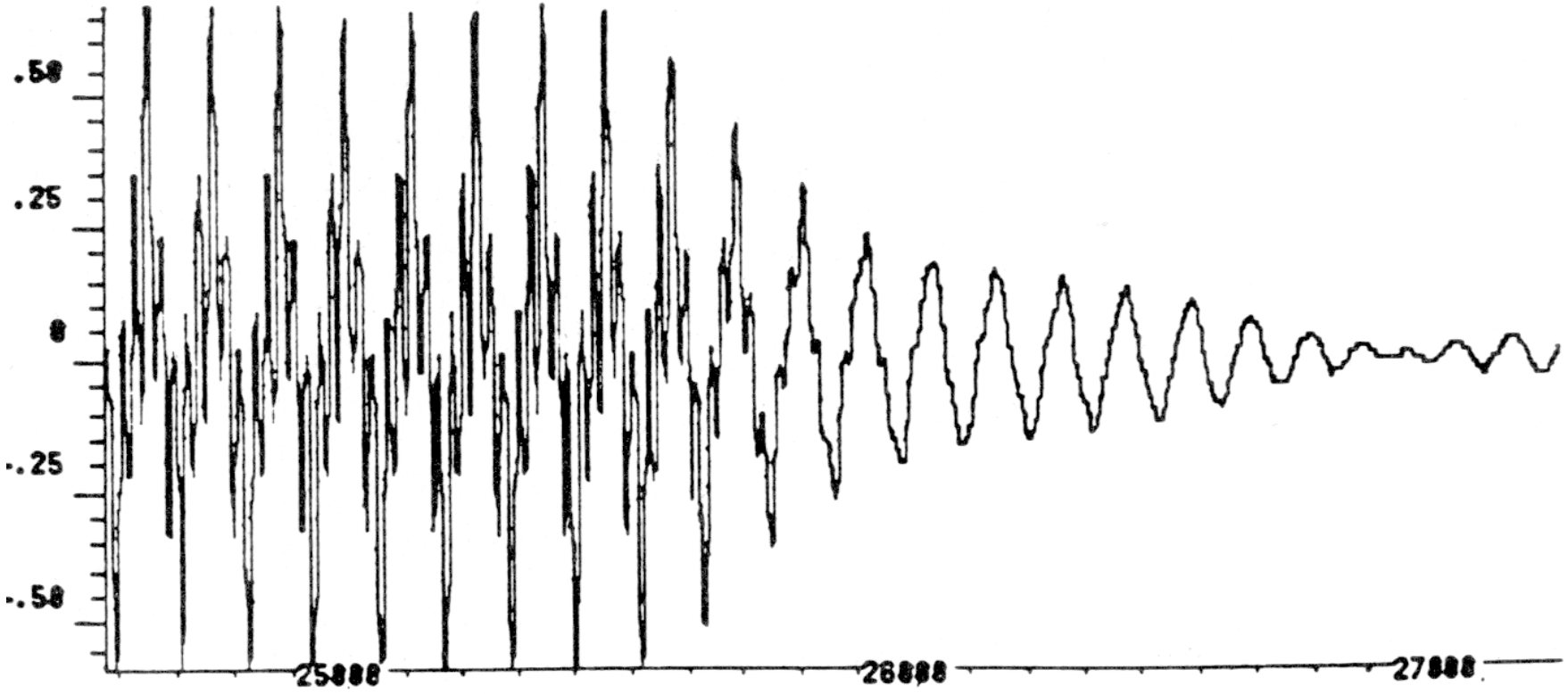
- At pain level, amplitude can be 1,000,000 times the sound at the threshold of audibility
- The measurement unit is dB = 20 $\log_{10}(A/B)$
- The threshold is 0 dB and pain level 100 - 120 dB
- Hearing is sense, which cannot be directly measured
  + the pitch of a sound changes according to amplitude
  + the loudness depends on frequency


120
110
100
90
80
70
60
50
40
30
20
10
0
120
100
80
60
40
20
0
INTENSITY LEVEL IN DB
20
100
1000
10000
FREQUENCY IN CYCLES PER SECOND

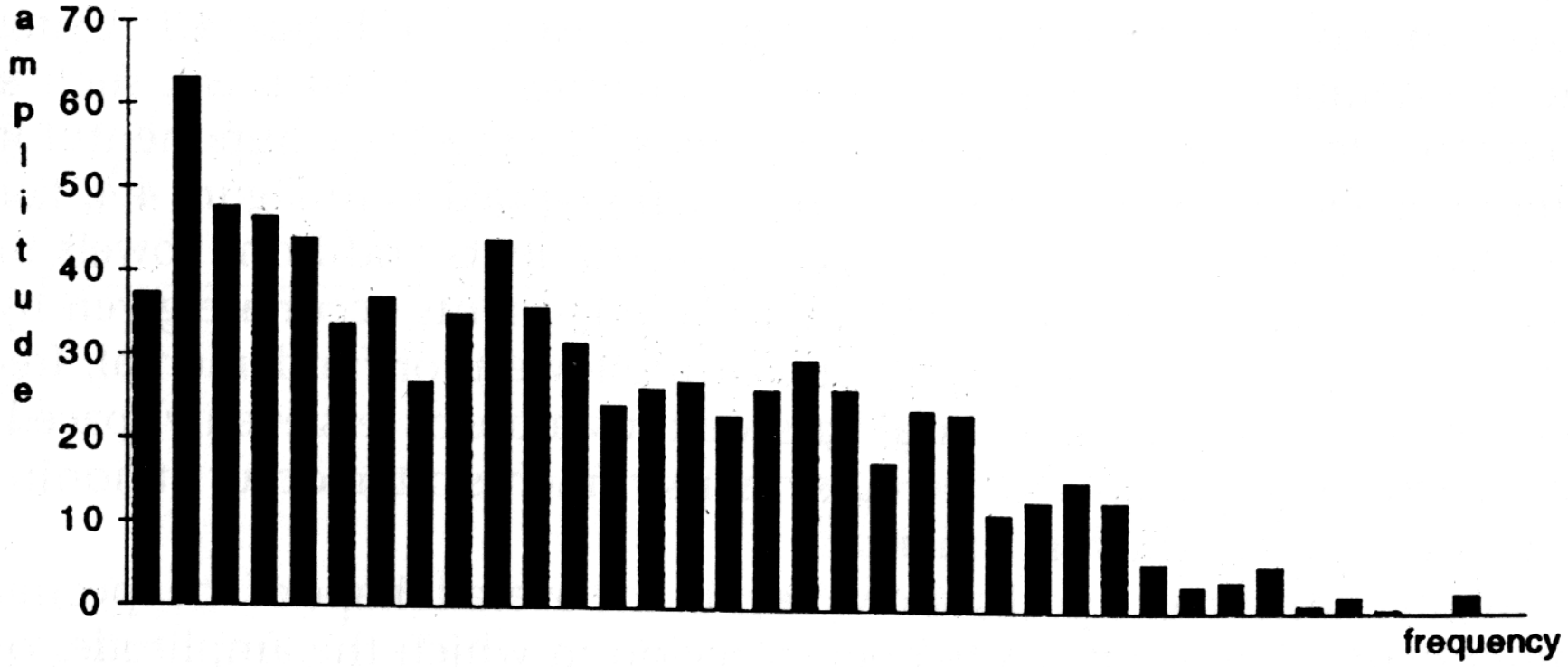
Frequency properties

- Natural sound are sums of many frequencies
- The frequencies can be calculated with Fourier analysis
- Natural sounds contain typically harmonic frequencies of the base frequency
- Ear is sensitive to valleys and hills of the spectrum
- Distinctive spots are called formants
- E.g., vocals distinction is based on formants

Clarinet sound



Frequency domain



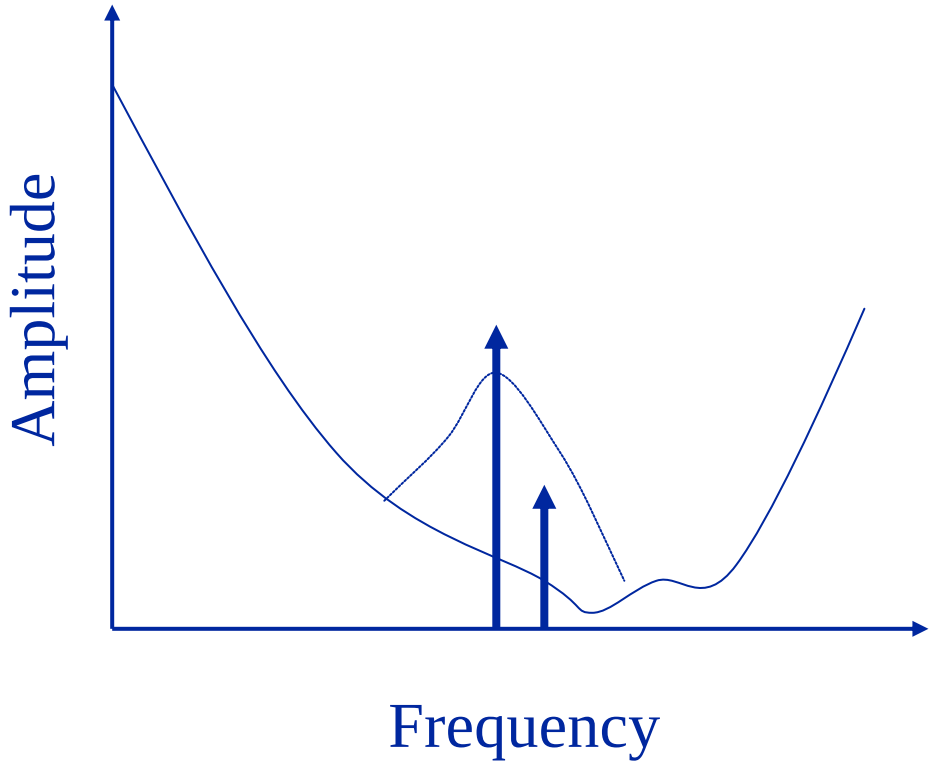
Time effect

- Sounds of instruments have three parts:
  + Attack, steady-state, and decay
- In simple synthesis, sound is generated with frequency components and their loudness is changed at different stages
- In real, the frequency components of the spectrum change constantly
- Hearing is especially sensitive in attack phase

Masking

- Sounds can mask each other partly or fully
- They can also change each other
- A certain frequency sound changes the threshold of audibility around larger area
- The sound have to be critical distance away so that they are heard separately
- Critical limit grows has frequency get higher

Masking (cont.)



Phase

- Same frequency sounds can have different phases
- A phase of 180 degrees cancels the sound
- There is evidence that, humans can hear phase difference

Binaural hearing and localization

- Humans can determine the location of sound
  + loudness, phase difference, frequency
- Skull, ear lopes, and hearing organs filter sound
- In addition, reflections have strong effect
- Sound sources should be placed in the same location and visual information

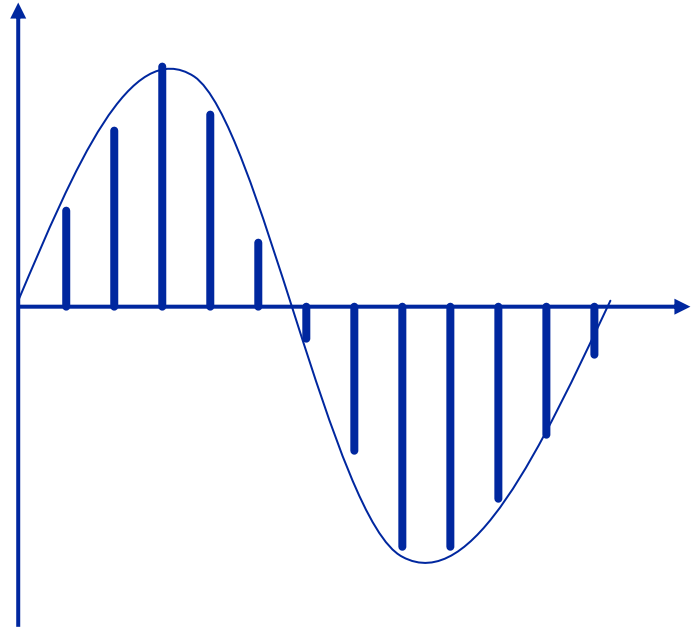
Digital presentation of sound

- Coding in time domains
- Transformations
- Linear prediction
- Parametric coding
- Digital transfer of audio

Coding in time domain

- Samples are taken at sample frequency
- The sample frequency has to be at least twice the maximum frequency (so called Nyquist frequency)
- Common sample frequencies are 8, 44.1, and 48 kHz
- The value of amplitude at sampling moment is coded as numeric value

Pulse Code Modulation (PCM)



Coding in time domain (cont.)

- Sampling causes quantization error
- Each bit improves the signal to noise ration

  + $20 \log_{10} 2 = 6$ dB
- Often 16 bits are used

  + $16 * 6$ dB $= 96$ dB
- The human dynamic hearing range is more (about 120 dB)

Coding in time dimension (cont.)

- Transformations between analog and digital signals are done with A/D and D/A converters
- In addition, filtering is required
  + Anti-alias and Reconstruction filters
- In high quality systems, they can also be error source
- This problem can be solved with oversampling
- The computer can also cause over hearing

Other coding methods

- In delta modulation (DPCM), only the difference between successive samples is coded
- In adaptive delta modulation (ADPCM), the step size can change

Other coding methods (cont.)

- In speech A law is often used (A = 87,6)

$$y ? \begin{cases} ? \dfrac{Ax}{1 ? \ln A}, & (0 ? x ? \dfrac{1}{A}) \\ ? \dfrac{1 ? \ln(Ax)}{1 ? \ln A}, & (\dfrac{1}{A} ? x ? 1) \end{cases}$$

- also ? law is common (? = 255)

$$y ? \frac{\ln(1 ? ?x)}{\ln(1 ? ?)}, (0 ? x ? 1)$$

Transformations

- Transformations can be used to present the content in different domain
- The goal is to make the signal transfer more efficient and robust

Fourier transformation

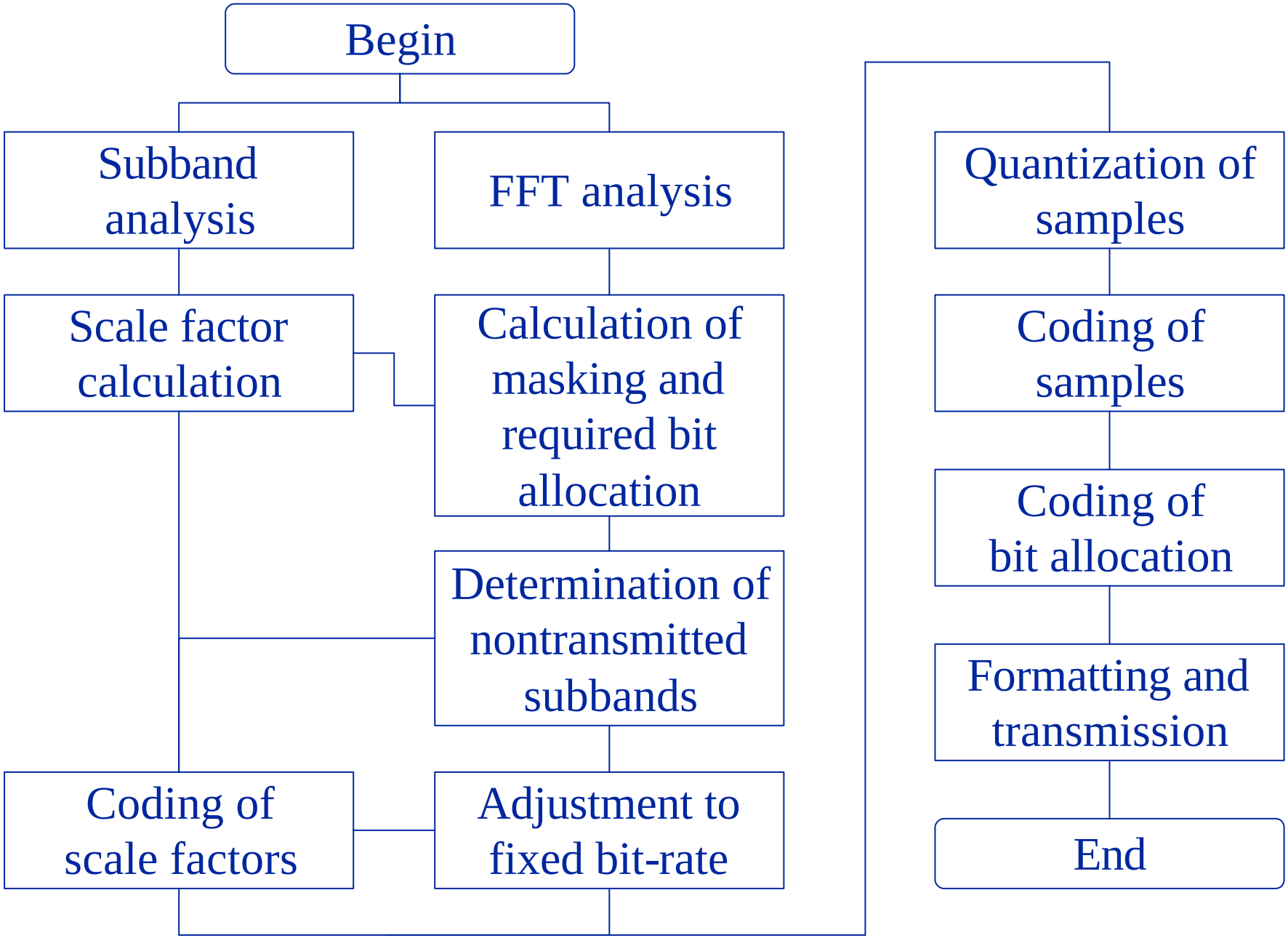
- Fourier coefficients represent the signal accurately in frequency domain
- Static signals can be present presented exactly with Fourier coefficients
- Discrete Fourier transformation has to be used with dynamic signals
- Coefficients are usually calculated with the Fast Fourier Transformation (FFT) algorithm

Frequency bands

- Masking effect can also be used in coding
- Signal is first divided into frequency bands, which are then coded separately (Subband Coding)
- E.g., Mini Disc -records (Sony), DCC cassettes (Philips) and MP3
- The methods has also been used with speech coding and recognition

MPEG audio

- MPEG audio uses Subband coding
- Signal is divided into 32 bands (Layer 1)
- The division is done to groups of 384 samples
- FFT transformation is used to find band with pure sine signals and noise
- Only interesting channels are coded
- The bit allocation per channel varies
- Layer I over 128 kbps / channel


Begin
Subband analysis
FFT analysis
Scale factor calculation
Calculation of masking and required bit allocation
Determination of nontransmitted subbands
Coding of scale factors
Adjustment to fixed bit-rate
Quantization of samples
Coding of samples
Coding of bit allocation
Formatting and transmission
End

MPEG audio (cont.)

- Layer 2
- about 128 kbps / channel
- 1152 samples per group
- 3 scaling factors
- 36 frequency bands

- Layer 3 (MP3)
- about 64 kbps / channel
- filter bank
- Huffmann coding
- bits used for coding can vary

Code-excited linear prediction (CELP)

- Speech can also be coded with synthesis
- Decoder receives control data and feeds it into synthesis filter
- Synthesis filter “mimics” speech using linear prediction
  + simple control signal produces thus speech
- Encoded tries different control signals and chooses the one that produces sound closest to the input signal

Digital audio transfer

- Transfer in digital format is preferable
- Professionals require better quality than CD
- Several sampling frequencies and resolutions
- Special converters exist for frequency transformations
- Several formats exist

Transfer formats

- Audio Engineering Society / European Broadcast Union (AES/EBU)
- PCM format
- 16-20 bits / sample
- stereo
- includes clock
- status (sampling frequency, etc.)

- Multichannel Audio Digital Interface (MADI)
- 56 channels
- 24 bits / sample
- in addition, several manufacturer specific formats

JPEG

- Objectives
- Architectures
- DCT coding and quantization
- Statistical coding
- Lossless coding
- Efficiency

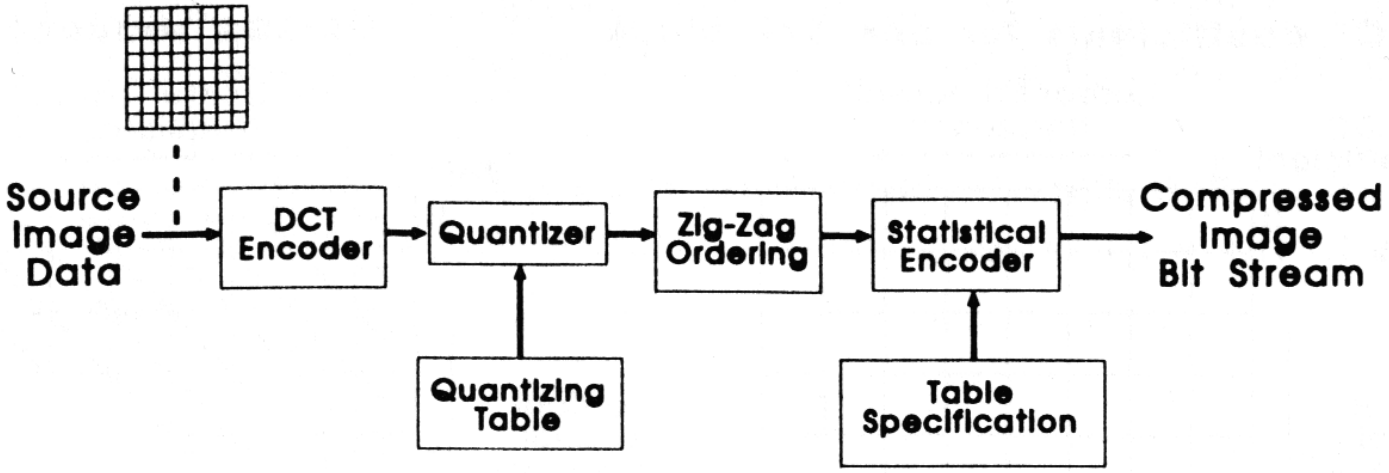
Objectives

- Compression rate / image quality can be selected
- Works with all kinds of images
- Both software and hardware implementation
- Four different modes
  - + sequential coding (original order)
  - + progressive coding (multiphase coding)
  - + lossless coding (perfect copy)
  - + hierarchical coding (many resolutions)

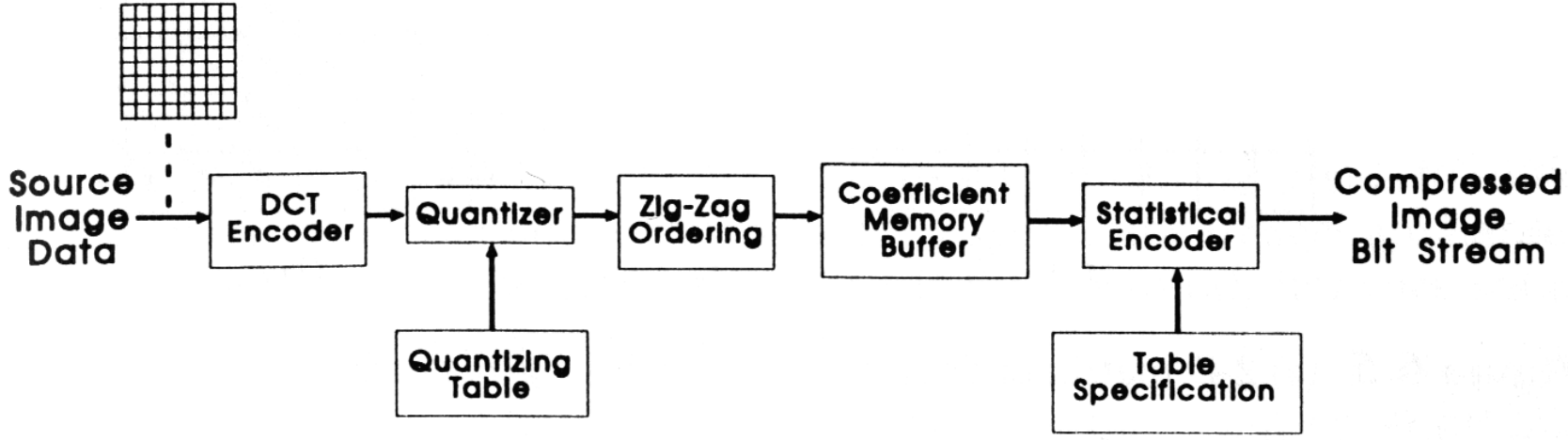
JPEG Architectures

- Lossy modes use DCT for 8 x 8 pixel blocks
- Sequential mode outputs the DCT-coefficients block by block
- Progressive mode outputs the DCT-coefficient in groups
- Hierarchical mode encodes several resolutions at the same time

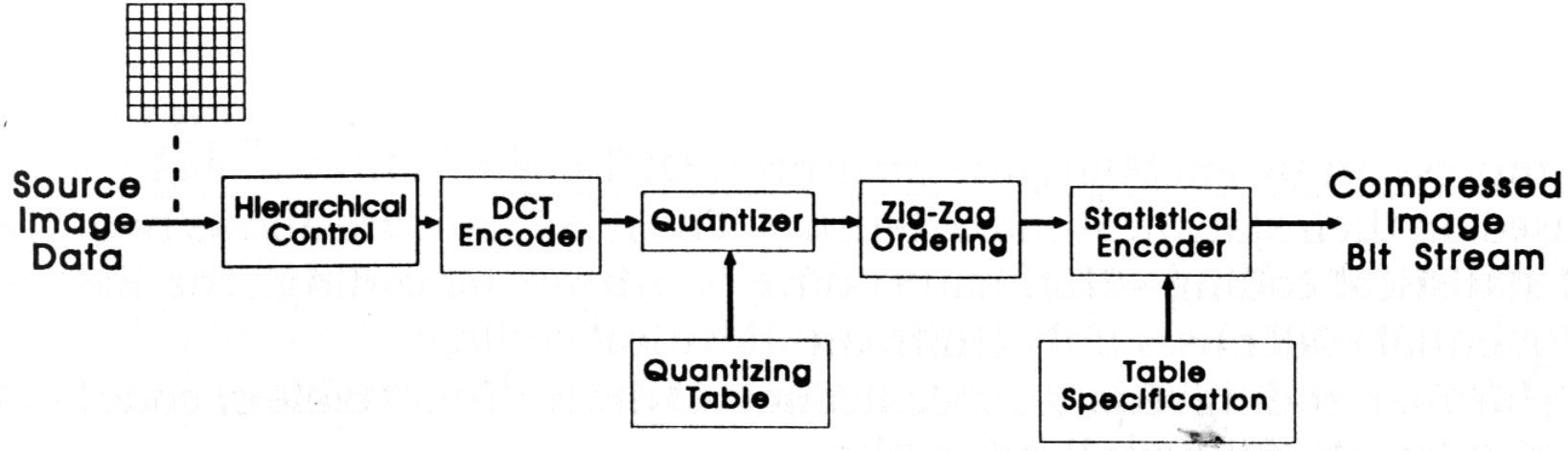
Sequential JPEG



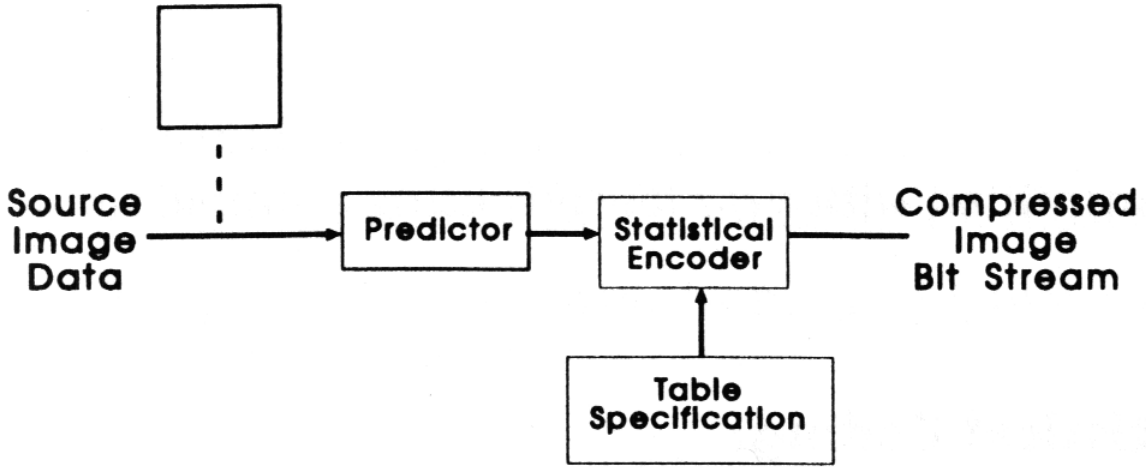
Progressive JPEG



Hierarchical JPEG



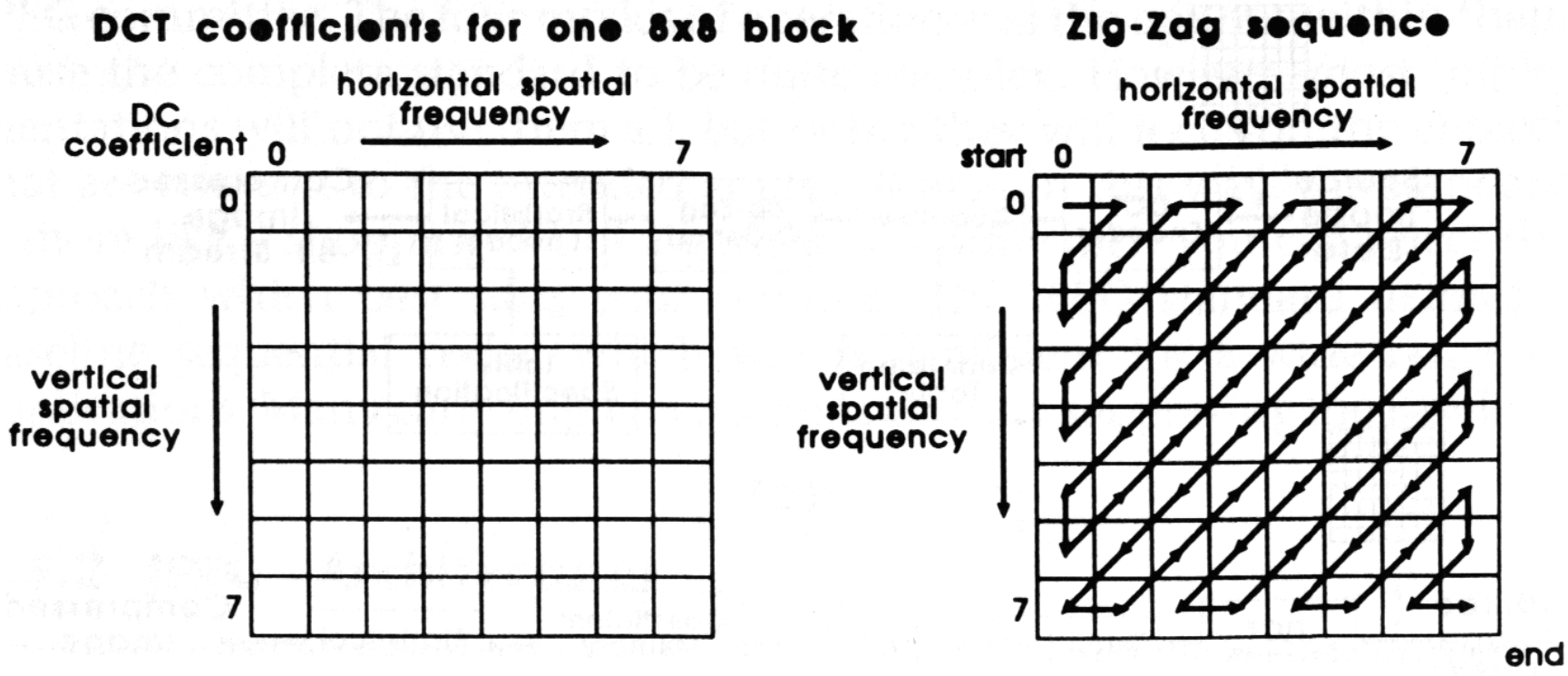
Lossless JPEG



DCT and Quantization

- The DCT-coefficients can be represented as a matrix
- The quantization is done according to a quantization table
- The coefficients are put in Zig-Zag order
- This places the zero coefficients in the end of the run
- Finally Run-Length coding eliminates the zeros

DCT coding


DCT coefficients for one 8x8 block
DC coefficient
horizontal spatial frequency
0
7
0
7
vertical spatial frequency
Zig-Zag sequence
horizontal spatial frequency
start
0
7
0
7
vertical spatial frequency
end

Statistical Coding

- Uses either Huffman or arithmetic coding
- Huffman coding requires a separate table
- Arithmetic coding does not require a table, but need more computation
- In addition, the compression ratio of arithmetic coding is 5 - 10 % better

Lossless Encoding

- Lossless encoding utilizes prediction
- Seven different alternatives
  + how many and which pixels are used
- Predictive encoding can reach compression ratio of 2:1

Efficiency

- 0,25 - 0,5 bpp: reasonable - good quality
- 0,5 - 0,75 bpp: good - very good quality
- 0,75 - 1,5 bpp: very good quality
- 1,5 - 2,00 bpp: same as original