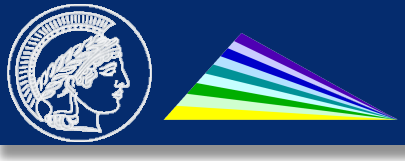


Compression of Audio and Video

Paul Trilsbeek



Codecs and File Formats

- Important to distinguish between codec and file format: codec = COder + DECoder, describes the algorithm with which media (for example images or video) are coded and then again decoded
- Examples of codecs:
  - Audio: mp3, ogg vorbis, Linear PCM
  - Video: MPEG1/2, H.264, MJPEG2000
- File format specifies how resulting bytes are packaged, structured. Some “container” formats can contain audio and video encoded with various codecs.
- Examples of container formats:
  - Audio: AIFF, WAV, WMA
  - Video: MOV, WMV, AVI, MP4, MPEG2 PS

DOBES media formats

- MPI/DOBES Standards: end 2001 discussion about compression formats in DOBES - results: uncompressed is best and strongly recommended, however not always practically possible
  - For images: TIFF, JPEG or PNG
  - For video: MPEG2 and MPEG4/H. 264(MJPEG2000 lossless for specific collections) as archiving format, MPEG1/MPEG4 as derived formats
  - For audio:16 bit linear PCM 44.1 or 48 kHz WAV file (higher bit rates and sampling frequencies need to be converted)

Compression of audio

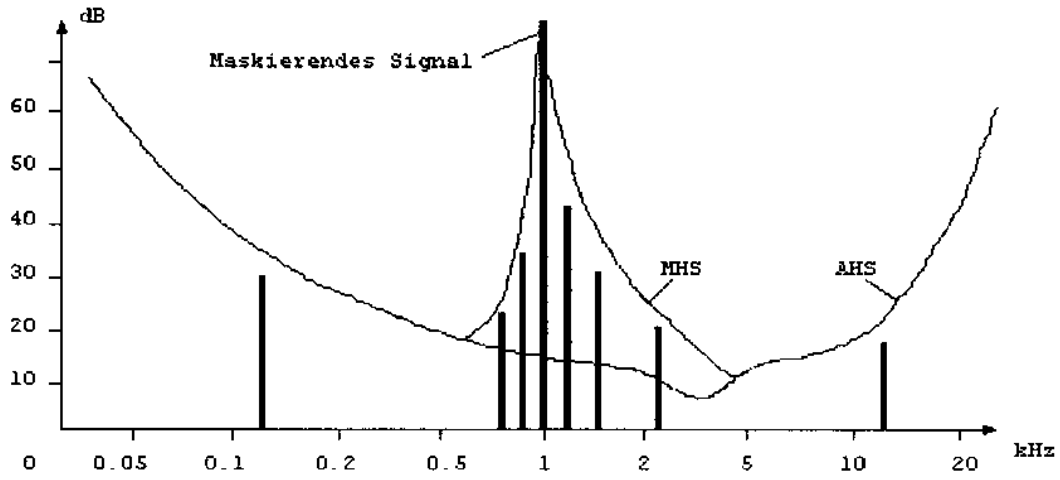
•„perceptual coding“: data-reduction modelled according to human perception, takes advantage of our perceptual inadequacies, throws away information which is not processed by our senses

• audio: phenomenon of masking, certain information cannot be processed anymore because it is simply masked by other information:

  • simultaneous masking: if frequency components are too close we perceive them as one sensation

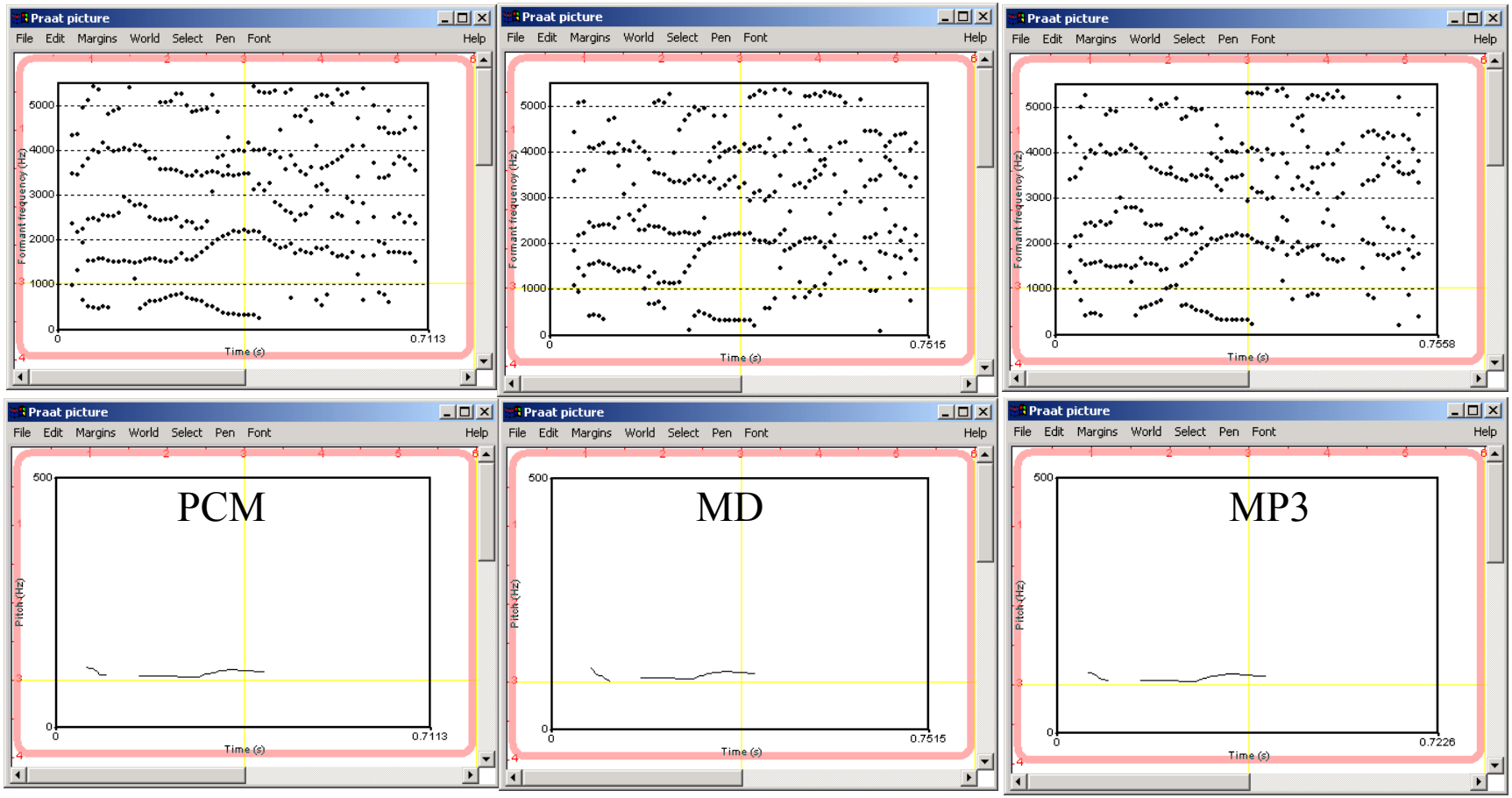
  • spectral masking: our hearing perception is divided into so-called critical bands, within a critical band a weaker tone is masked by a stronger one

  • temporal masking: a tone has a pre- and post masking effect over time



tones below frequency-dependent threshold are not perceived

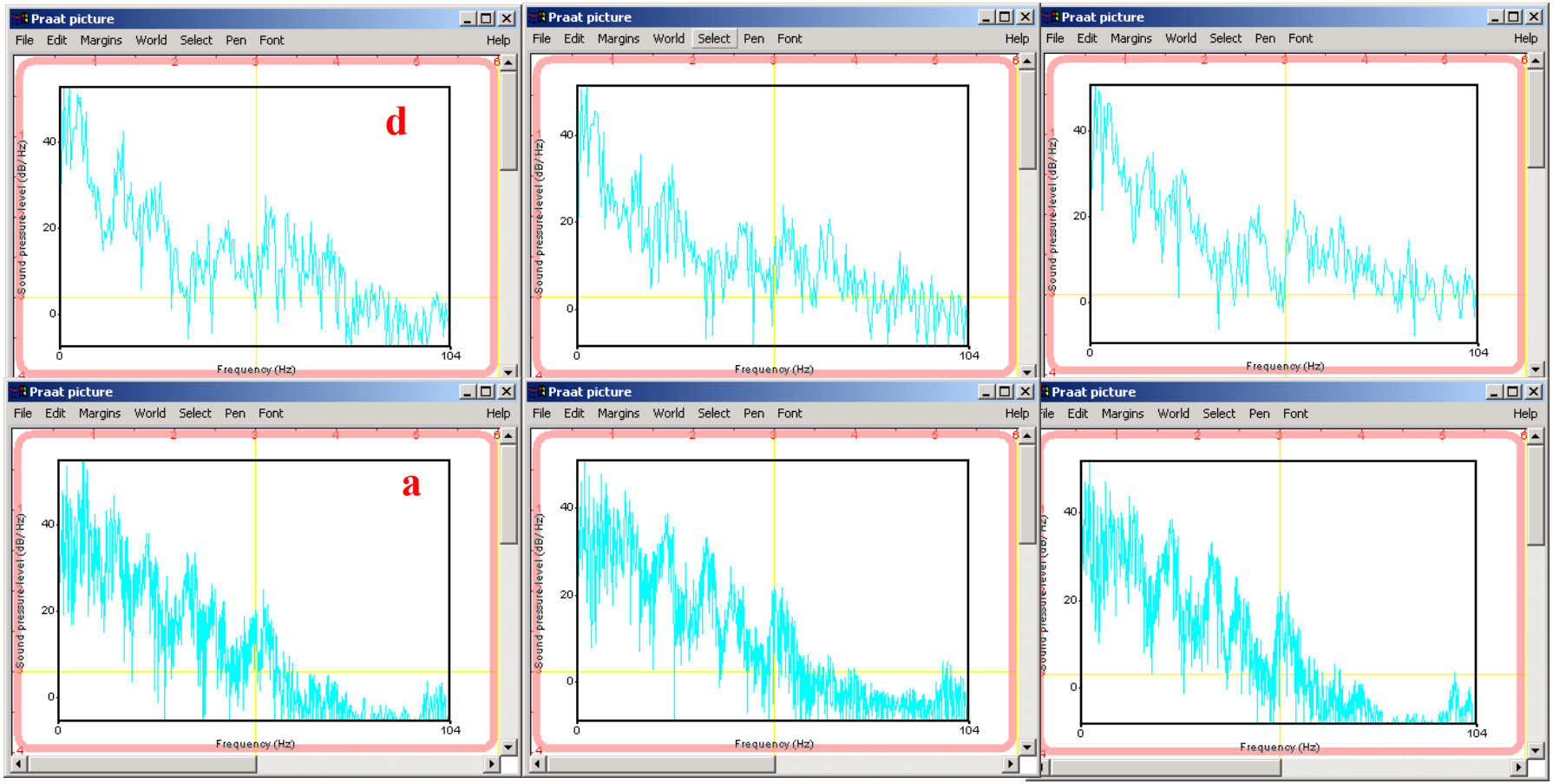
Audio Compression



- formant tracking no problem
- relevant formants discovered
- problems in the high frequency range where formant concept problematic
- tracks smoother for PCM

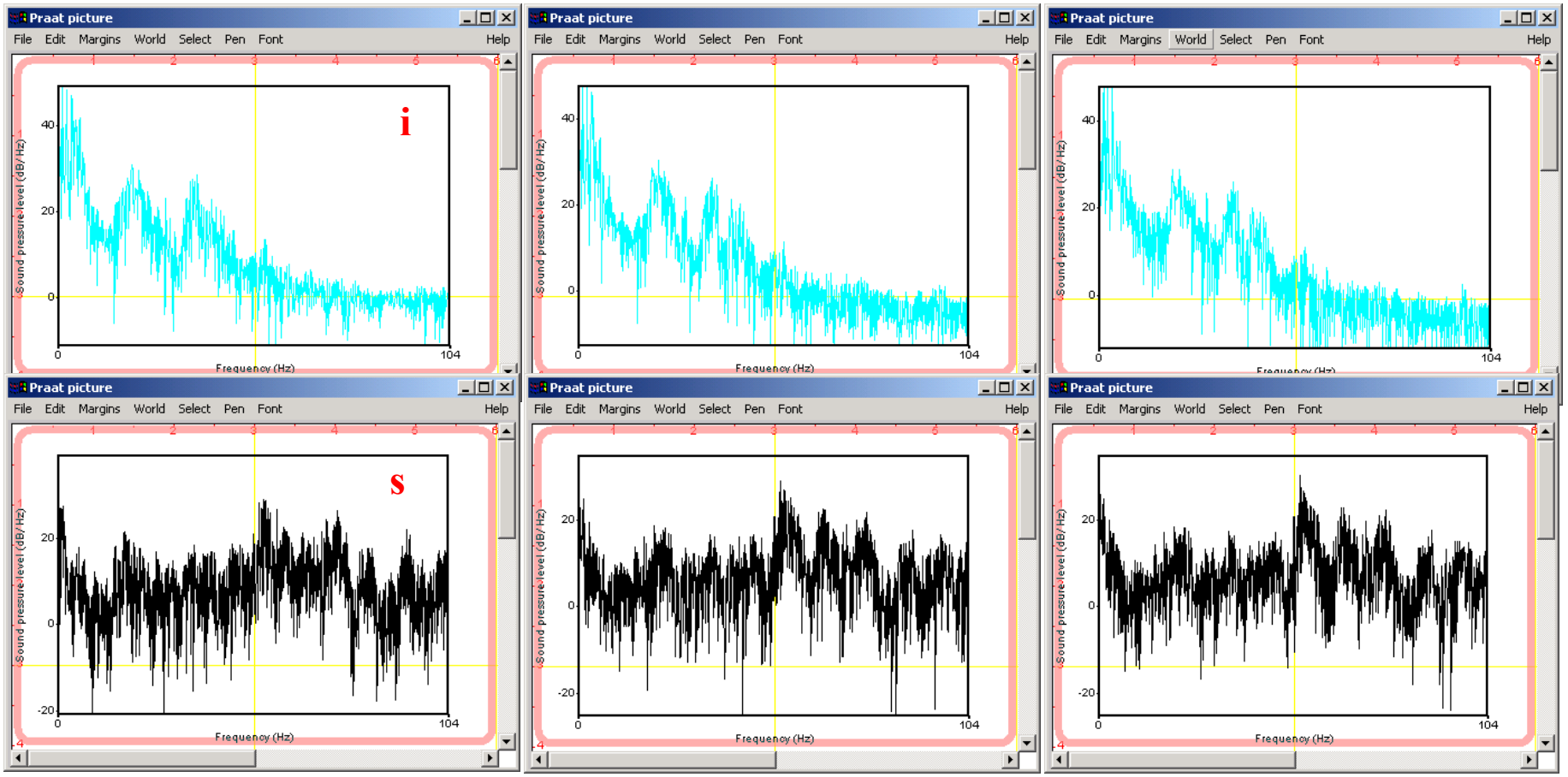
- pitch contours no problem
- done in time domain

Audio Compression



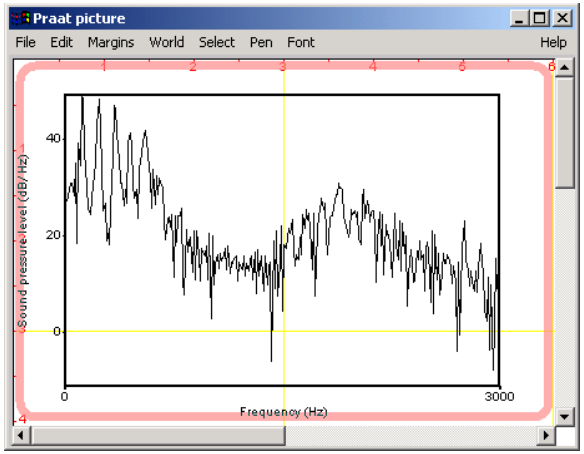
- spectral envelopes are preserved, especially resonance positions
- differences especially in the high f region (broad critical bands)

Audio Compression

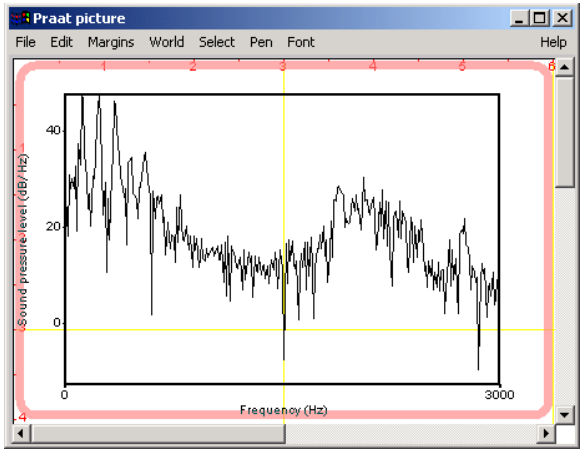


- some filtering aspects between the high freqency resonances
- recognizable differences for the s due to broad critical bands in high f region
- difference is not perceivable by listening

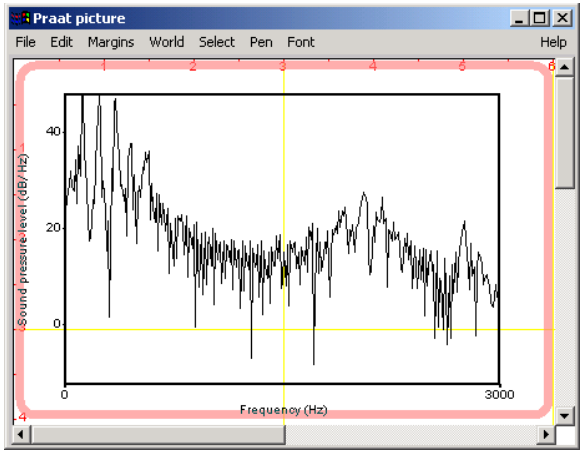
Audio Compression



PCM

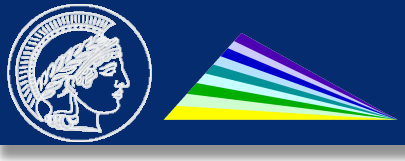


MD



MP3

- from high resolution spectra for the “i” we can see good representation of the harmonics due to small critical bands (150 Hz)
- for the high f resonances harmonic structure is slightly distorted obviously not relevant for perception



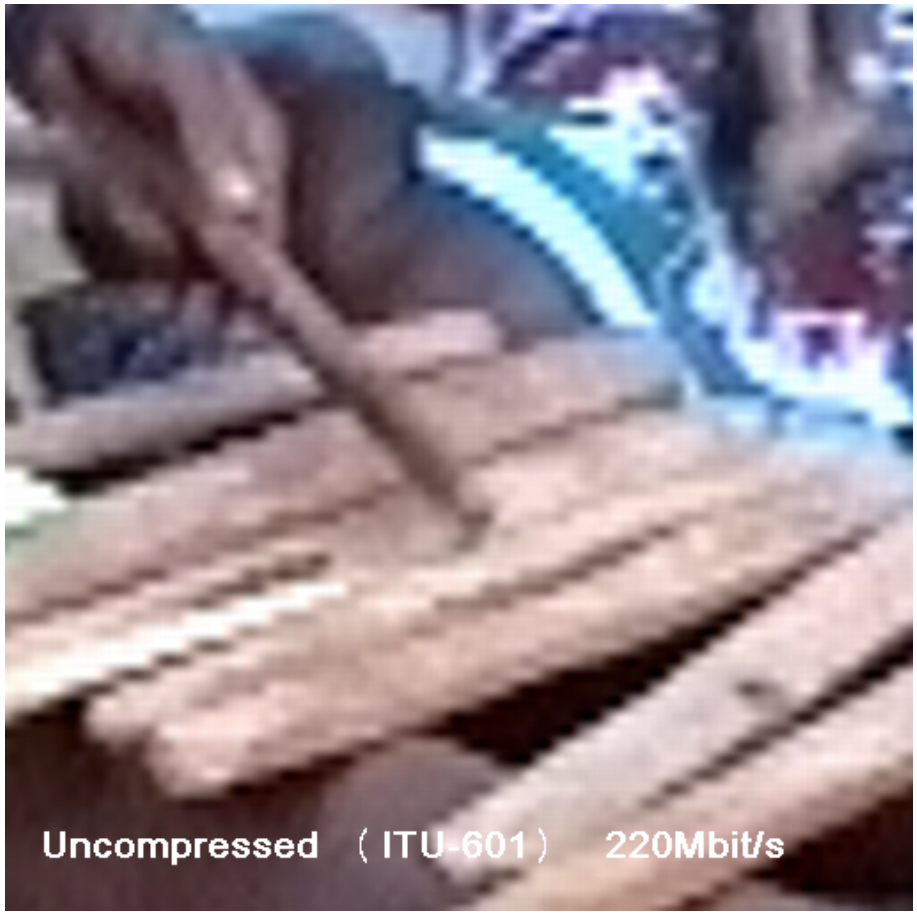
Audio Compression

- results Wittenburg
  - almost no visible difference in Pitch, Spectrum, basic formants except at high frequencies
- results van Son
  - only small consequences even for production parameters
  - change of microphone has larger consequences
- results Campbell
  - no difference for hearing tests
- well known is
  - time filtering after hard sounds

Compression of Video

Intraframe coding (spatial):	**Interframe coding (temporal):**
- compression within one frame	- compression between frames
- “perceptual coding”: known limitations of the human eye are exploited: small color changes are perceived less accurately than small changes in brightness - data reduction through: 1. averaging together groups of pixels (color); 2. throwing away high freq. Information which is not perceivable anymore; 3. quantization (high more than low, color information more than brightness)	- only the differences between frames are encoded as often a large number of the pixels will be the same on a series of frames (e.g. a background doesn‘t change) - vectors describe how the pixels move from frame to frame (motion compensation and estimation), then only the vectors are encoded – this is a reduction step
- editing easier (DV)	- editing computationally intensive (MPEG2/MPEG4)

Compression artifacts





Images courtesy of Dietrich Schüller and Franz Pavuza, Phonogrammarchiv Vienna

Compression artifacts





Images courtesy of Dietrich Schüller and Franz Pavuza, Phonogrammarchiv Vienna

Uncompressed Video Data Rates

	SDTV (Pal)	HDTV	Digital Cinema (2K, 4K)	35mm film
pixels	720Hx576V	1920x1080	1920x1080 2048x1080 4096x2160	~ 4096x2160
data-rate (Mbit/s)	165,9	> 995,3	10.872 (4K)	

Record compressed or not?

- For audio-only recordings no excuse not to record uncompressed: flash recorders record uncompressed, memory cheap enough
- For Video: uncompressed video recording not possible with consumer/prosumer equipment. Be aware of possible artifacts.
- Audio with video: if possible record uncompressed, if not feasible it’s not the end of the world. Microphone quality has more effect on sound quality then applying compression (with high bit rate)

Compression for long-term archiving

- Converting from one lossy compressed format to another introduces artifacts
- Archives therefore store data in as uncompressed form as possible
- Film industry and film and video archives starting to use lossless compressed video in MJPEG2000, MPI does this for special material