Multimedia souhrn prezentací 1-3a

Data Compression

- Data compression involves data transformation (encoding) to reduce their size.

Motivation

- Saves memory needed for data storage.
- Saves communication time.

Advantages vs. Drawbacks

- **Advantages**:
  - Reduced storage requirements.
  - Faster data transmission.
- **Drawbacks**:
  - Slower access to the data.
  - Compressed data are harder to recover in case of hardware failure.
  - In the case of network communication failure, it is often easier to transmit the whole data block repeatedly.

Redundancy

- The basic compression principle is to remove redundancy in data.
- There are "objective" and "subjective" redundancies.
- Compression is based on repeating symbols, words, sentences, and context redundancy (e.g., English letter patterns).

Compression Types

- **Lossy / Lossless**:
  - Lossy: Original data may be decoded from the compressed version only approximately.
  - Lossless: Original data may be decoded exactly.
- **Streaming / Block**:
  - Streaming: Data flows continuously.
  - Block: Data is compressed in blocks.
- **Static / Adaptive / Something in Between**:
  - Static: Algorithm parameters do not change.
  - Adaptive: Parameters adapt to the current data, either globally or locally.

Terminology

- **Symbol**: Basic unit of work (e.g., bit, byte, pixel, record).
- **Word**: A finite sequence of symbols.
- **Message**: Input sequence of words to be encoded or output sequence.
- **Coder, Decoder (Compressor, Decompressor)**: Programs performing coding and decoding algorithms.
- The term **codec**: A codec must be capable of both coding and decoding.

Compression Efficacy

- Compression efficacy depends on input data.
- Textual data have different properties than audio streams, and understanding data characteristics improves compression efficiency.

Entropy

- Redundancy is linked to randomness and entropy.
- Entropy measures the amount of information in data.

Morse Alphabet:

- Morse code is a method of encoding text characters as sequences of dots and dashes. It's a form of non-reversible text compression that simplifies the representation of text by using a binary encoding.
- Example: In Morse code, the letter "A" is represented as ".-", "B" as "-...", and "C" as "-.-.".

Run-length Encoding (RLE):

- RLE is used for compressing data with long sequences of repeated symbols by representing them as a count of the symbol and the symbol itself.
- Example: If you have the string "aaabbbcccc," RLE would represent it as "3a3b4c."

Statistical Methods:

- These methods utilize knowledge about symbol or word frequency in the message to achieve compression. Two common techniques are Shannon-Fano and Huffman coding.
- **Shannon-Fano Coding (1949)**:
  - Symbols with higher probabilities get shorter codes.
  - Example: In Shannon-Fano coding, if the letter "E" is more frequent in the text, it will receive a shorter code compared to a less frequent letter like "Q."
- **Huffman Coding (1951)**:
  - Huffman coding assigns variable-length codes to symbols, with shorter codes for more frequent symbols.
  - Example: In Huffman coding, the letter "E" might be assigned the code "10," while the less frequent letter "Q" could have a code like "1101001."

Arithmetic Coding:

- Arithmetic coding encodes a message as a single real number in a specific range. It's highly efficient but requires precise implementation.
- Example: Given a probability distribution for characters in a message, you encode the message as a real number between 0 and 1. For example, if the message is "ABAB" and "A" has a higher probability, it might be encoded as 0.75.

Dictionary Methods:

- Dictionary-based methods utilize a dictionary of frequently occurring words or patterns to achieve compression. Examples include LZ77, LZ78, and LZW.
- **LZ77 (Sliding Window)**:

- LZ77 uses a sliding window approach to represent repeated sequences in data efficiently.
- It encodes repeated sequences as pairs, where the first number indicates how far back in the window to look, and the second number indicates the length of the sequence to copy.
- Example: In "abracadabra," "abra" might be represented as (4, 3), indicating "go back 4 characters, and copy the next 3 characters."

LZ78 (Lempel-Ziv 1978):

- LZ78 is a dictionary-based compression method that builds a dynamic dictionary as it processes input data.
- It encodes sequences by using a pair consisting of an index and the following character.
- The matched sequence and the following character are added to the dictionary as a new entry.
- Example: Input "ABABABABA" would be compressed as (0, 'A'), (0, 'B'), (2, 'A'), (3, 'B'), (5, 'A').

LZW (Lempel-Ziv-Welch):

- LZW expands on LZ78 by using a dynamic dictionary. It adds sequences to the dictionary as it processes the input.
- Example: In a dictionary with common words like "the," "quick," and "brown," it would assign shorter codes to these common words.

Context Methods:

- Context-based methods predict the next symbol based on the context of previous symbols. The prediction can lead to more efficient encoding.
- Example: In predictive text input on a smartphone, the system predicts the next word based on the previous words typed. This prediction reduces the number of keystrokes needed to type a sentence.

Block Methods (Burrows-Wheeler Transform and Blocksort):

- Block methods operate on fixed-length blocks of data. They don't directly compress data but rearrange it to improve the efficiency of subsequent compression techniques like RLE or Huffman coding.
- Example: In the Burrows-Wheeler Transform, the data is rearranged, and the last column of the sorted table becomes the transformed block. This rearrangement makes it easier to apply other compression methods efficiently.

Signals

What is a Signal

- Phenomena that progress in time are referred to as processes.
- When these phenomena are transformed into changes in a physical quantity that can be transmitted, analyzed, processed, or stored, they are called signals.
- Signals can be acoustic, optic, electric, and more.

Signal Types

- Continuous
- Discrete

Different Classification of Signals

- Deterministic
  - Periodic
  - Quasiperiodic
  - Transient
- Stochastic
  - Nonstationary
  - Stationary
    - Ergodic
    - Nonergodic

Periodic Signals

- A signal that repeats itself regularly is called periodic.
- Mathematically, $x(t) = x(t + nT)$, where T is the signal period.

Sinus

- The most basic natural periodic signal is a sinusoid.
- It describes the steady movement of a point on a circle viewed from the side.
- Characterized by parameters like period, frequency, amplitude, and phase.
- Mathematically: $x = A \sin(2\pi ft + \phi)$.

Fourier Series and Spectrum

- Any periodic signal can be decomposed into a sum of sine functions with different amplitudes, frequencies, and phase shifts.
- The Fourier series represents this decomposition.
- Spectrum: Assigns amplitude and phase to each frequency.
- Periodic signals have a discrete spectrum with components at integer multiples of the fundamental frequency.

Non-Periodic Signals

- Quasiperiodic Signal: Discrete spectrum, doesn't repeat itself.
- Transient Signal: Single-shot "movement."
- Stochastic (Random) Signal:
- Stationary: Constant mean value over time.
- Ergodic: Mean value of one realization is equal to the mean value of more realizations.
- Noise: Main representative of random signals.

Spectrum of Non-Periodic Signals

- As the period of a signal approaches infinity, its discrete spectrum becomes continuous.
- Fourier transform is used to represent the spectrum.
- Continuous spectrum is a complex function of frequency.

- Perfect reconstruction from spectrum is impossible if not all spectral components are available.

Signal Discretization

- Transformation from continuous to discrete form.
- Sampling: Taking "samples" at regular intervals in time.
- Quantization: Representing values with finite precision.
- Discretization leads to loss of information.
- Balancing between too few samples (signal degradation) and too many (data volume).

Influence of Sampling on the Signal

- Sampling theorem (Shannon, Nyquist, Kotělnikov): Sampling frequency should be at least twice the highest signal frequency (Nyquist frequency).
- Frequencies higher than Nyquist cause aliasing, leading to unwanted distortion.
- Antialiasing filters are used to ensure no high frequencies are present before sampling.

Influence of Quantization on the Signal

- Quantization adds noise to the original signal.
- Coarser quantization results in larger noise amplitude.
- Methods to optimize or limit quantization noise exist.

Discrete Fourier Transform (DFT)

- Transforms a series of N numbers to another series of N numbers.
- Frequency components are no longer continuous but quantized.
- DFT is widely used and often computed efficiently using the Fast Fourier Transform (FFT).

Short-Time Fourier Transform (STFT)

- Classic DFT transforms the entire input signal.
- For long signals, it's more useful to analyze the signal's spectrum over time.
- STFT computes the DFT for smaller time windows.

Two-Dimensional Case

- Signals in 2D, like images, can be analyzed using DFT or STFT.
- Base functions for DCT (used in JPEG) are 2D sinusoids.
- Convolution in the spatial domain corresponds to multiplication in the spectral domain and vice versa.

Transformation of the Signal from Discrete to Continuous Form

- Reconstruction of the analog signal.
- D/A conversion: Computation of signal values between samples.
- Ideal interpolation involves convolution with a sinc function, which is non-practical due to its infinite nature.

Sound and Hearing

Sound

- Sound is mechanical waves propagating through a compressible medium, such as air, gas, liquids, and solids.
- It is a longitudinal wave, where particle displacement is parallel to the direction of propagation.
- In air, normal barometric pressure is approximately 100,000 Pa.
- Sound is described using acoustic pressure, which is the tiny changes in air pressure caused by sound.

Levels

- Sound intensity is not perceived linearly; there is approximately a logarithmic relationship.
- Sound intensity is expressed in rational logarithmic form, such as sound pressure level (SPL) in decibels (dB).
- SPL is measured in reference to the acoustic pressure reference (p0), which is about 20 mPa.
- A ten-fold increase in vibration amplitude results in a 20 dB increase in SPL.

Sound and Signal

- Sound is a continuous analog 1D signal.
- It can be represented using spectrum or spectrogram, which shows short-time spectrum variations over time.

Human Auditory System

- The auditory system is complex, with information passing through various stages, including the outer ear, middle ear, inner ear (cochlea), auditory nerve, and higher centers.
- Cochlear filters in the inner ear transform sound into the frequency domain, and sound frequency is perceived logarithmically.
- Equal loudness contours demonstrate that different frequencies are perceived with varying loudness.
- Masking occurs when a strong stimulus suppresses the perception of a weak stimulus, especially if they have similar frequencies.

Human Auditory System

- The auditory system is complex, with information passing through various stages, including the outer ear, middle ear, inner ear (cochlea), auditory nerve, and higher centers.
- Cochlear filters in the inner ear transform sound into the frequency domain, and sound frequency is perceived logarithmically.
- Equal loudness contours demonstrate that different frequencies are perceived with varying loudness.

- Masking occurs when a strong stimulus suppresses the perception of a weak stimulus, especially if they have similar frequencies.

Sound Recording and Discretization

- Sound is continuous, and to acquire and process it, it's represented as a continuous electric signal.
- Audio signals may be recorded in an analog form or converted into digital data for storage.
- A/D (Analog to Digital) and D/A (Digital to Analog) converters are essential for digitization and reconstruction.

A/D and D/A Conversion

- A/D and D/A converters are essential for converting analog audio signals into digital and vice versa.

Digitization

- Sampling is the process of converting analog signals into digital form.
- It's essential to use a high-quality low-pass filter to prevent aliasing during sampling.
- Quantization introduces quantization noise, with the amount depending on the number of bits used.
- The Signal-to-Noise Ratio (S/N) is a measure of the conversion quality, with more bits leading to higher S/N ratios.

Sigma-Delta Converter

- Sigma-delta converters are commonly used in audio applications.
- They provide precision using only 1 bit but require high sampling frequencies.

D/A Conversion

- In D/A conversion, high-resolution samples are used to reconstruct the original analog signal.
- Sigma-delta converters are widely used in audio equipment, ensuring high-quality sound.

Formats

- **WAV (Waveform Audio File Format)**
  - Block structure.
  - Typically stores uncompressed PCM audio.
  - Offers flexibility in sampling frequencies, resolutions, and channels.
  - Supports additional data like triggers, regions, and markers.
- **AIFF (Audio Interchange File Format)**
  - Similar to WAV.
  - Often associated with MAC devices.
- **Proprietary Formats**
  - Used by audio processing software like Cubase and ProTools.
  - Designed for project-specific data structures.
  - Facilitates import and export of common formats.

- **Audio CD**
    - Stores uncompressed PCM data.
    - Standard: 44.1 kHz sampling frequency, 16-bit depth.
    - Capacity: around 74 minutes.
    - Self-correcting codes like CIRC for data integrity.
    - Uses phase-reverted laser beam to read pits.

Software

- Various audio editing software exists for different purposes, including multitrack recording, sound wave editing, and effects.

In General

- High-resolution and sampling frequency are important for professional audio.
- Care must be taken to avoid sharp changes at the selection boundaries, as they can create audible clicks and pops.

Cutting

- Audio editing involves working with a series of samples.
- Segments can be cut and copied to other parts of the audio.
- Care must be taken to avoid abrupt signal changes.
- Cross-fading is used to create smooth transitions.
- At the insertion point, you can overwrite, mix, or shift the existing signal to accommodate the new audio.

Amplitude Changes

- **Fade In / Fade Out**: Gradual increase or decrease in amplitude to avoid abrupt volume changes at the beginning or end of audio tracks.
- **Mute**: Setting samples to zero to create silence in specific parts of the audio.
- **Normalize**: Adjusting the entire audio block's amplitude so that the loudest sample matches a predetermined value, often -1 dB.
- **Noise Gate**: Silencing samples below a certain threshold to remove background noise or unwanted sounds during silent portions.
- **Dynamic Compression**: Adjusting sample amplitudes based on their current values to control the dynamic range, ensuring audio balance.

Effects

- **Amplitude Modulation (Tremolo):**
    - Tremolo modulates amplitude, creating a pulsating effect.
    - Adds dynamics and expressiveness to sound.
- **Chorus:**
    - Simulates multiple voices in unison.
    - Thickens and enriches sound with delayed copies.
- **Delay (Echo):**
    - Creates sound repetitions after a short delay.
    - Adds spaciousness and depth.

- **Flanger:**
  - Alters phase with a delayed signal.
  - Produces a sweeping, "wah-wah" effect.
- **Reverberation:**
  - Simulates acoustic spaces.
  - Adds depth and spatial dimension.
  - Parameters: decay time, pre-delay, room size.

Resampling and Requantization

- Resampling changes the sampling frequency.
- Requantization changes the bit resolution and may require dithering to reduce quantization noise.

Windows and Sound

- Standard sound recording and playback in Windows uses WDM drivers, which have latency and quality limitations.
- ASIO (Audio Stream Input/Output) offers low-latency, bit-perfect communication for professional audio applications.
- WaveRT drivers are introduced in newer Windows versions but may have limitations for professional use.

Audio Signal Compression

Motivation

- Human hearing capabilities:
  - Frequency range: Approximately 20 Hz to 20 kHz.
  - Dynamic range: Approximately 120 dB (amplitude ratio 1:10^6).
- Digital audio requirements:
  - Sampling frequency: At least 40 kHz.
  - Quantization: At least 20 bits.
  - Bitrate: At least 800 kbps in one channel.
  - 1 minute of stereo audio requires more than 10 MB.
  - High rates are not practical for network transmission and storage, hence the need for compression.

Compression

- **Lossless Compression**:
  - Common methods like LZW have limited compression ratios (approximately 1:1.2).
  - Algorithm FLAC (Free Lossless Audio Codec) achieves better compression (about 60% of WAV).
  - Principles:
    - Decorrelation of left and right channels.
    - Signal approximation using polynomial or linear predictive coding.
    - Residual encoding with a version of Huffman code.

- **Perceptual Compression** (Lossy Compression):
    - Achieves higher compression ratios while maintaining reasonable quality.
    - Perceptual coding is central.
    - Removal of non-perceivable signal parts.
    - Psychoacoustic models for estimation.
    - Quantization precision decreases in non-perceivable areas.
    - Signal is transformed to the frequency domain, where spectral components are quantized.
- **Frequency Transformations**:
    - Filterbanks.
    - MDCT (Modified Discrete Cosine Transform).
    - Signal is processed in overlapping blocks.
    - Weighting of samples with a window before the transform.
    - Combining spectral and temporal information.

Principle of a Perceptual Coder

- Transforms the signal to the frequency domain.
- Estimates a masked threshold based on spectral peaks and psychoacoustic models.
- Suggests lower bit resolution for masked spectral components.
- Injects quantization noise, which must stay below the masked threshold.
- Encodes the bitstream using Huffman code and writes it to a file.

Perceptual Coders

- Developed in the 1970s.
- The most widespread is MP3 (MPEG-1 Layer III or MPEG-2 Layer III).
- Other perceptual coders include Ogg Vorbis, Windows Media Audio, ATRAC, AAC, and more.

MP3 Format

- The MP3 format standard defines the data format but not the compression algorithm.
- Different encoder implementations can vary in quality.
- Structure:
    - Frequency transform: Filterbank and MDCT.
    - Psychoacoustic model.
    - Quantization and coding: Iterative loops.
    - Bitstream compression using Huffman code.
- Quality is set using a bitrate (e.g., 128 kbps) or other settings.
- Sampling frequency may be limited.
- Options for joint stereo, conversion to mono, and quality vs. speed trade-offs.

Problems of Perceptual Compression

- **Distortion**:
    - Loss of information leads to signal degradation.
    - Distortion in perceptual coders differs from harmonic distortion.

    - Block-oriented processing can lead to audible artifacts.
    - Good frequency resolution vs. temporal resolution trade-off.
    - Examples include pre-echo, birdies, and aliasing.
- **Pre-echo**:
    - Audible near impulse-like signals.
    - Caused by the averaging of the whole block.
    - Quantization noise can appear before the signal.
    - Common near drum beats and similar signals.
- **Birdies**:
    - Ringing or "sci-fi" sounds in the signal.
    - Caused by different quantization in subsequent blocks.
    - Quantization noise is not white noise but has tonal components.
    - Examples include birdies and their perception due to perceptual streaming.
- **Collapse of Stereo Image**:
    - Can occur with certain signals and compressions.
- **Aliasing**:
    - Appears if the compression reduces the sampling frequency without proper low-pass filtering.