EI IV MULTIMEDIA TECHNOLOGY

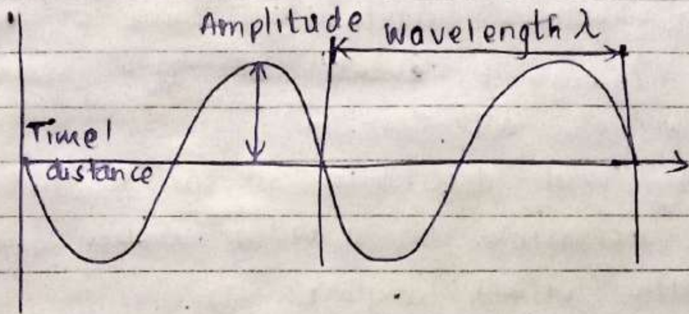
Page No. 1

Unit 3 Audio and Video Processing

~~[6181] 196~~ [6004]-568

Q.1 a) What is the sound wave? What are the uses of audio in computer applications? Explain with suitable example

Ans



Ans. A sound wave is a mechanical wave that travels through a medium like air, water, or solids as vibrations. It is a longitudinal wave, meaning the particles in the medium vibrate parallel to the direction of the wave travels, creating areas of compression (high pressure) and refraction (low pressure). Sound waves are characterized by properties like frequency, (pitch), amplitude (loudness), and speed (how fast it travels through a specific medium.

Audio is fundamental to modern applications, enabling numerous functionalities and user experiences. Here are some key uses of audio in computer applications:

- Communication : Audio allows for real-time conversation and interaction.
  e.g. Voice calls and video ~~calls~~ conferencing using applications like Zoom or Microsoft Teams, where microphones capture voice as audio input and speakers output the voice of others.
- Entertainment and Media: The consumption of music, movies, and games relies heavily on audio.
  e.g. Streaming music on platforms like Spotify or Apple Music, where vast libraries of digital audio files are played back through a computer's

sound system.

- User Interfaces and feedback Audio cues can enhance user interaction by providing non-intrusive feedback
  e.g System alert sounds in operating systems (like the Windows notification chime or macOS alert sound) signal events to the user without needing visual attention.
- Accessibility and Assistance. Audio technology makes computers accessible to a wider audience, especially those with visual impairments.
  e.g. Screen reader software such as JAWS or NVDA, which uses synthesized speech (text-to-speech) to read out screen content and help users navigate interfaces.
- Digital Audio Workstations (DAWs) These are applications specifically designed for recording, editing, and producing audio professionally.
  e.g. Creating multi-track music using software like Ableton Live or Logic Pro, where musicians manipulate waveform data to compose, mix, and master songs.
- Speech Recognition and Virtual Assistants: Computers process language for command and control.
  e.g. Interacting with virtual assistants like Google Assistant or Amazon Alexa on a computer, where the application interprets the user's spoken audio command to perform tasks.

Q1b) List the various audio formats supported by the internet Explain PCM and AAC audio file formats

Ans The internet supports various audio formats, which fall into three main categories unconpressed, lossless compressed, and lossy compressed. The most widely supported formats across web browsers include MP3, AAC, Ogg Vorbis, and WAV

Common Internet Audio Formats

- Uncompressed formats These offer the highest quality as they retain all original audio data, but result in very large file sizes.
- WAV (Waveform Audio File Format): The standard uncompressed format for ~~w~~ Windows systems and the default for CDs, it is widely supported by browsers.
- AIFF (Audio Interchange File Format): Apple's equivalent to WAV, also uncompressed PCM-based audio.
- Lossless Compressed formats: These reduce file size without any quality loss by using compression algorithms that allow the original data to be perfectly reconstructed.
- FLAC (Free Lossless Audio Codec): A popular, open-source format favoured by audiophiles for high-resolution audio archiving.
- ALAC (Apple Lossless Audio Codec): Apple's alternative to FLAC, compatible with Apple devices
- Lossless Compressed Formats: These achieve the smallest file sizes by permanently discarding some inaudible or less noticeable audio information. This is the most common type for streaming and general web use.
- MP3 (MPEG-1 Audio Layer III): The most common audio format universally supported accross almost all devices and browsers.
- AAC (Advanced Audio Coding): Designed as the successor to MP3, it offers better sound quality at the same or lower bitrates and is ~~w~~ widely used by platforms like Apple Music and YouTube.

• Ogg Vorbis: A free, open-source alternative to ~~AtP~~ MP3 and AAC, often used by Spotify for streaming.
• Opus: A modern, low-latency format ideal for real-time internet communication like voice chats and is also used for general audio streaming

PCM (Pulse Code Modulation)

PCM is not a file format in itself but underlying method to convert analog audio ~~sounds~~ signals into a digital format. It forms the basis for uncompressed audio files like WAV and AIFF.

The process involves three steps:

1. Sampling: The amplitude of the continuous analog signal is measured at regular, uniform intervals (e.g. 44,100 times per second for CD quality)
2. Quantization: Each ~~samp~~ sampled value is rounded to the nearest value within a predetermined set of digital steps. The number of steps is determined by the bit depth (e.g., 16-bit for a CD)
3. Encoding: The quantized values are converted into binary data (1s and 0s) that a computer can store and process.

• Characteristics: PCM captures the most faithful reproduction of the original analog sound possible in the digital domain, with no data compression or quality loss. The drawback is very large file sizes, making it less practical for general internet streaming but ideal for professional audio production and archiving.

AAC (Advanced Audio Coding)

AAC is a lossy compression audio coding format designed to be more efficient than MP3.

AAC uses a complex psychoacoustic model to identify and discard the parts of the audio spectrum least critical to human hearing. This significantly reduces the file size while aiming to maintain high perceived audio quality. Different specifications, such as Low Complexity (AAC LC) and High Efficiency (HE-AAC), cater to various uses and bitrates.

- Characteristics: It generally provides better sound quality than an MP3 file at the same bitrate. Its efficiency makes it the standard for streaming services, portable devices, and digital ~~rado~~ radio. AAC files are often found within M4A or MP4 container formats.

Q.2 a) What is the purpose of video editing? Explain various stages of video editing considering sample application.

Ans Video editing serves the primary purpose of meticulously crafting raw footage into a cohesive, impactful narrative or message. It involves selecting, arranging, and modifying clips to convey emotion, guide the viewer's attention, and meet specific communication goals, such as entertaining an audience, educating viewers, or promoting a product.

The process transforms disparate, unorganized shots into a polished final product.

Stages of Video Editing

Video editing is typically broken down into three main phases: ingest / logging, the editing process itself (rough cut, fine cut, final cut), and output / delivery.

1. Ingest and Logging: Transferring footage from cameras / storage devices into the editing software and organizing it. This includes reviewing clips, adding metadata (labels, descriptions), and selecting 'selects' (best takes).

2 Assembly / Rough Cut The initial phase of arranging chosen clips into a basic sequence that follows the script or story structure. The focus is on the narrative ~~and~~ flow and sequence, not yet on perfect timing or effects.

3 Fine Cut: The editor refines the timing, rhythm, and transitions. Edits become more precise, and elements like sound design, visual effects (VFX), color correction, and music are integrated to enhance the emotional impact and visual quality.

4 Final Cut and Output: The client or director approves the final cut. The finished video is then exported (rendered) into the required format and compression settings for its final destination (e.g., YouTube, broadcast TV, social media)

Sample Application: A 30-Second Travel Advertisement

Consider the process of creating a 30-second video advertisement promoting a new travel destination.

1 Ingest & Logging: The editor imports hours of footage - scenic views, people enjoying activities, close-ups of local food - into the editing software. They label clips as "Beach Sunset," "Hiking Trail," "Zip-lining Action," and mark the best 10 seconds of each.

2 Assembly / Rough Cut: The editor arranges the "selects" in the general order of the planned commercial: an enticing opening shot ~~at~~ of the beach, followed by active shots to build excitement, and a final shot of the destination logo.

3 Fine Cut

. Timing: The editor tightens cuts to align with a dynamic background music track, ensuring quick transitions to maintain energy.

Sound: A voiceover artist's track is added and balanced with the music and ambient sound effects

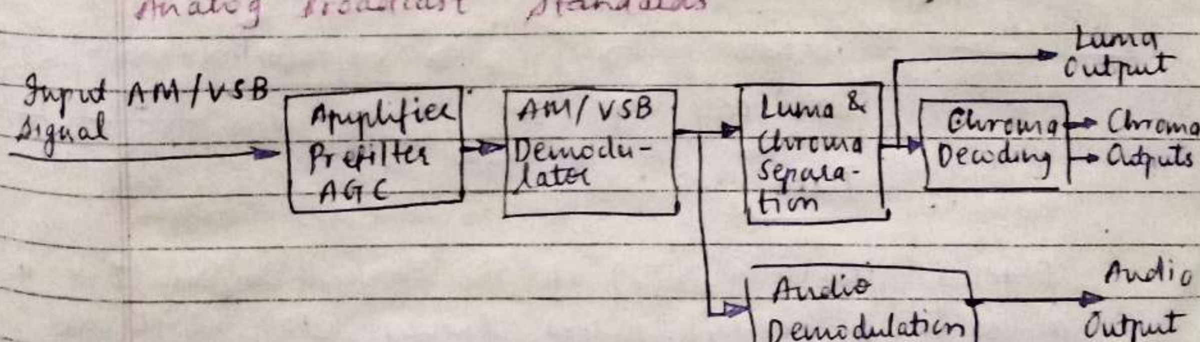
· Visuals: Color grading is applied to make the blues of the water more vibrant and the sunsets warmer. A call-to-action graphic is created and placed at the end.

4 Final Cut & Output: The director approves the ad. The editor exports the file as a high-resolution, compressed version for web streaming on YouTube.

Q.2.b) Draw and explain video broadcast standards

Ans.: Video broadcast standards are the technical rules that ensure consistent transmission and reception of television signals, covering aspects like resolution, frame rate, and color encoding The most common traditional analog standards were NTSC, PAL, and SECAM, which have largely been replaced by digital standards such as those from the ATSC and DVB communities

Analog Broadcast Standards



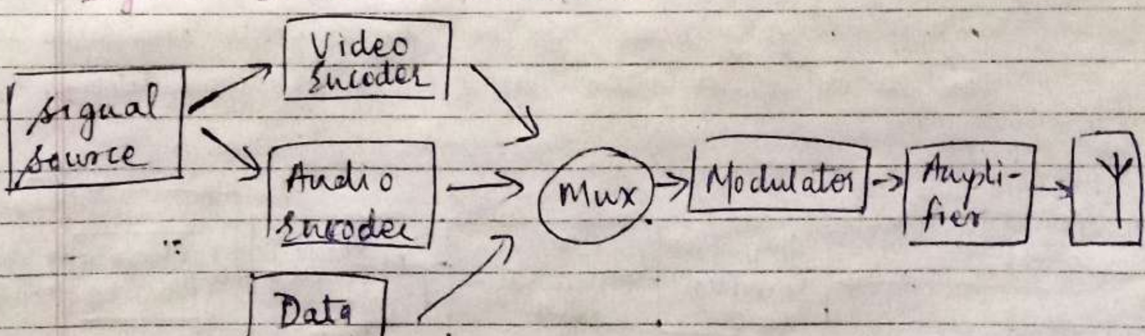
Analog television represents brightness, color, and sound as continuous signals (amplitude, phase, and frequency modulation). The quality of the picture degrades with a weak signal, becoming "snowy" or subject to interference.

The three main standards worldwide were:

· NTSC (National Television System Committee): Primarily used in North America, Japan, and South Korea. It defined a system with 525 horizontal scan lines and a refresh rate of 60 fields (30 frames) per second using interlaced scanning (drawing odd lines, then even lines)

- PAL (Phase Alternating Line): used in Europe, Australia, and many other parts of the world. It increased the resolution to 625 horizontal lines but slowed the frame rate to 50 fields (25 frames) per second.
- SECAM (Sequential Color and Memory) Used in France and parts of Eastern Europe. It also used a 625-line, 50 Hz system but differed in its color encoding technology, transmitting color information one line at a time.
- These standards are generally not compatible with each other due to difference in line counts, frame rates, and color encoding methods.

Digital Broadcast Standards



Digital television (DTV) systems transmit information as a series of on/off pulses (binary data), which are compressed and encoded. This method allows for efficient bandwidth usage, enabling higher resolutions and multiple ~~channels~~ channels within the same frequency space an analog signal used.

Modern digital standards include:

- ATSC (Advanced Television Systems Committee) Adopted in the US, Canada, Mexico, and South Korea, replacing NTSC. It includes support for various formats, including high-definition television (HDTV) resolutions like 720p and 1080i/p, and uses MPEG-2 or H.264 video encoding.

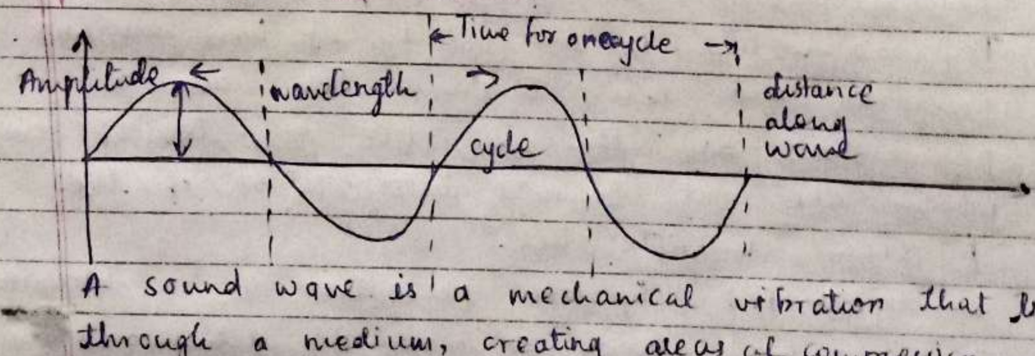
- DVB (Digital Video Broadcasting) A widely adopted suite of standards (e.g., DVB-T for terrestrial, DVB-C for cable) used across Europe and many other regions. It supports video encoding standards such as MPEG-2 and MPEG-4 (H.264) and advanced audio coding.
- ISDB (Integrated Services Digital Broadcasting) Developed and used in ~~po~~ Japan and parts of South America.

Digital standards maintain high quality until the signal drops below a certain threshold (the "digital cliff"), at which point ~~recepto~~ reception fails entirely, unlike the gradual degradation of analog signals.

[6181]-196

Q1 a) What is a sound wave? Explain the characteristics of sound waves. What are psychoacoustic parameters of sound?

Ans.



A sound wave is a mechanical vibration that travels through a medium, creating areas of compression and rarefaction. Its main characteristics are amplitude, which determines loudness, frequency, which determines pitch; and wavelength, the distance of one complete wave cycle. Psychoacoustic parameters describe how humans perceive sound, including properties like loudness, pitch, ~~tibre~~ timbre, and duration.

Characteristics of Sound Waves

- Amplitude The maximum displacement of a particle from its resting position: It determines the sound's loudness; a larger amplitude means a louder sound.
- Frequency The number of sound wave cycles that pass a fixed point per second. It is measured in Hertz (Hz) and determines the pitch of the sound. Higher frequency means a higher pitch.
- Wavelength The distance between two consecutive compressions or rarefactions in a sound wave
- Time Period The time it takes for one complete cycle of the sound wave to pass a fixed point.
- Velocity (or Speed) The speed at which the sound wave travels through the medium.
- Timbre The quality of a musical note or sound that distinguishes it from others with the same pitch and loudness. It is determined by the combination of a fundamental frequency and its harmonics.

Psychoacoustic parameters of sound

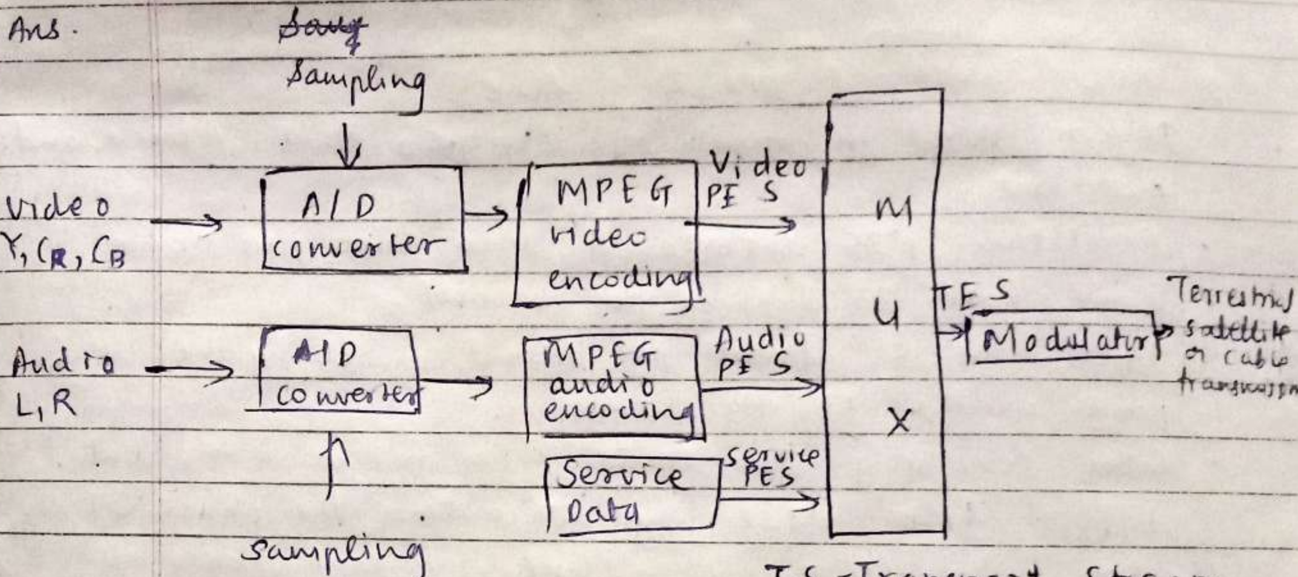
Psychoacoustics is the study of how humans perceive and interpret sound. Psychoacoustic parameters are the subjective properties of sound, which are our brain's interpretation of the physical characteristics.

- Loudness The subjective perception of the intensity of a sound, which is primarily determined by its amplitude.
- Pitch The subjective perception of how "high" or "low" a sound is, which is determined by its frequency.
- Timbre The quality or "color" of a sound that distinguishes it from other sounds of the same pitch and loudness.
- Duration The length of time a sound is heard.

• Other parameters: Other less common parameters include sharpness and roughness, which relate to the subjective perception of a sound's spectral characteristics.

Q.5.b) Explain any three video transmission standards in detail.

Ans.



TS - Transport Stream
PES - Packetized Elementary Stream

Three detailed video compression standards are NTSC, PAL, and H.264. NTSC and PAL are analog broadcast standards that differ in their frame rate, lines per frame and color encoding, with NTSC at 30 frames per second within 525 lines (used in North America) and PAL at 25 frames per second with 625 lines (used in Europe). H.264 is a digital standard that uses efficient compression techniques like intra-frame and inter-frame coding, with a 4×4 transform and variable block sizes, to significantly reduce bandwidth and storage requirements for digital video like streaming and DVDs.

1. NTSC (National Television System Committee)
• Region: Primarily used in North America and Japan.
• Resolution: 525 interlaced scan lines per frame.

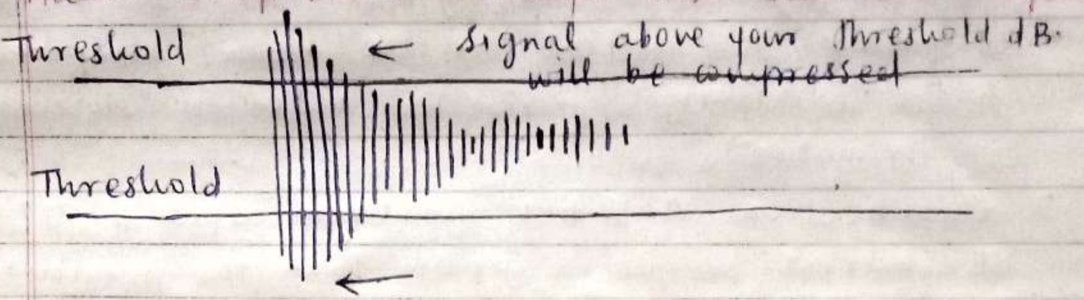
- Frame Rate: 30 frames per second (or 29.97 fps).
- Field Rate: 60 fields per second, matching the local power grid's 60 Hz frequency.
- Color Encoding: Transmits color information along with the brightness (luminance) signal. It uses a color subcarrier frequency of 3.579545 MHz.

2. PAL (Phase Alternating Line)
- Region: Used in most of Europe, Asia, Africa, and Australia.
- Resolution: 625 interlaced scan lines per frame.
- Frame Rate: 25 frames per second.
- Field Rate: 50 fields per second, matching the local power grid's 50 Hz frequency.
- Color Encoding: Transmits color information, but differs from NTSC by alternating the phase of the color information on a line-by-line basis to correct for phase errors.

3. H.264 (AVC - Advanced Video Coding)
- Type: A modern digital video compression standard.
- Function: Reduces the size of video files while maintaining high quality through advanced compression algorithms.
- Key features:
  - Uses a combination of intra-frame and inter-frame coding.
  - Employs a 4x4 transform and variable block sizes for motion compensation.
  - Utilizes quarter-pixel motion estimation and an ~~inapp~~ in-loop adaptive deblocking filter for improved efficiency and picture quality.
- Applications: Widely adopted for applications like high-definition television (HDTV), DVDs, video streaming services, and video surveillance.

Q 2a) What is Audio compression? Explain the DM Audio compression technique with suitable example.

Ans



Audio compression is a signal processing technique used to reduce the amount of data required to represent audio signals, enabling more efficient storage and transmission. This can involve either:

- Data compression (lossy or lossless), which reduces file size by removing redundant and perceptually irrelevant information.
- Dynamic range compression, a mixing tool that reduces the difference between the loudest and quietest parts of an audio signal to create a more consistent volume level.

Data Modulation (DM) is a method of data compression that focuses on simplicity and low bandwidth usage.

Delta Modulation (DM) Audio Compression Technique

Delta Modulation is a simple analog-to-digital conversion (ADC) and digital-to-analog signal conversion technique that uses a single bit per sample to represent only the change (delta) in the audio signal's amplitude ~~to~~ from the previous ~~example~~. sample.

How it Works.

Instead of transmitting the absolute value of each sample, DM transmits a binary signal (a '1' or '0') indicating whether the current sample's amplitude is greater than or less than the previous approximated sample.

1. Sampling: The analog signal is sampled at a high rate (higher than the Nyquist rate).
2. Comparison: The current sample value is compared to the previously approximated value (a "staircase approximation").
3. Quantization: A 1-bit quantizer outputs a '1' if the current sample is greater than the previous approximation, and a '0' if it is less than or equal to it.
4. Transmission: Only this 1-bit stream is transmitted.
5. Reconstruction: At the receiver, an integrator accumulates the bits (adding or subtracting a fixed step size, Δ) to reconstruct a staircase approximation of the original signal, which is then smoothed with a low-pass filter.

Example

Consider an analog signal that needs to be converted to a digital stream using Delta Modulation, with an initial reconstructed value of 0 and a fixed step size (Δ) of 1 unit

Sample Time	I/P Signal Value	Prev. Approx. Value	Difference (I/P - Prev.)	Transmitted bit	New Approx. value
1	0.5	0	+0.5 (+ve)	1	0+1 = 1
2	1.2	1	+0.2 (+ve)	1	1+1 = 2
3	1.8	2	-0.2 (-ve)	0	2-1 = 1
4	2.5	1	+1.5 (+ve)	1	1+1 = 2
5	2.8	2	+0.8 (+ve)	1	2+1 = 3
6	2.1	3	-0.4 (-ve)	0	3-1 = 2

The transmitted digital bitstream for this sequence is 1, 1, 0, 1, 1, 0. This minimal data rate makes it efficient for low-bandwidth applications like basic speech transmission, where quality is less critical than simple implementation and timely delivery.

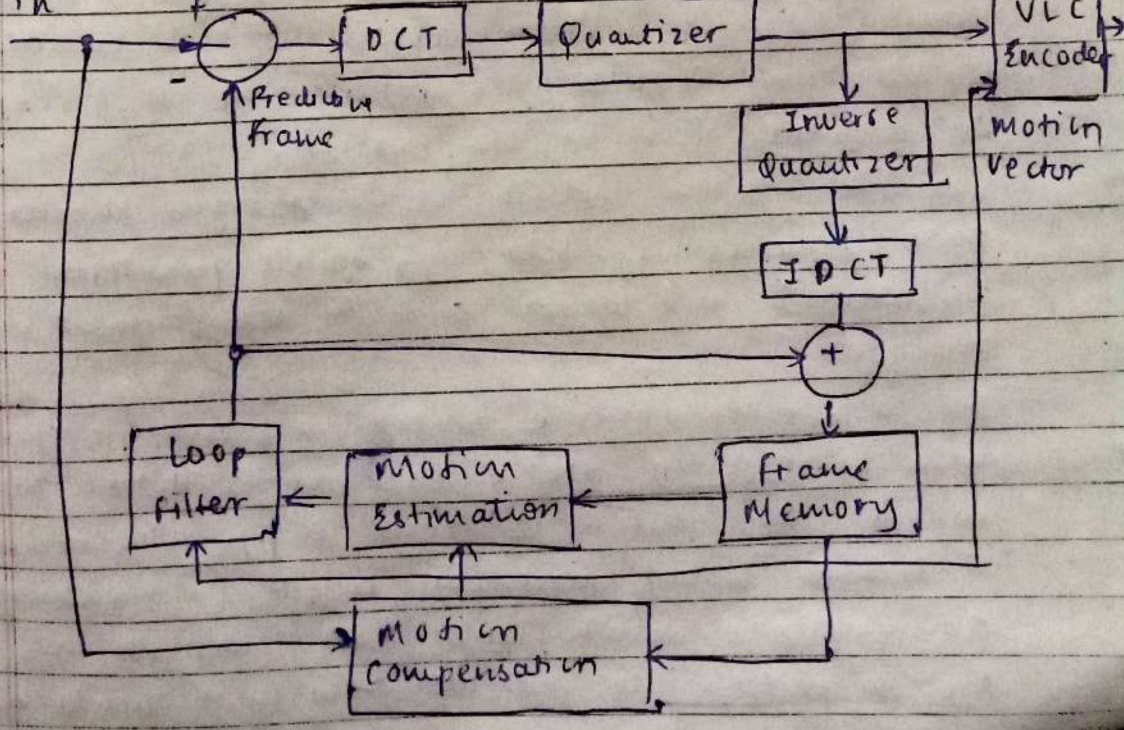
Limitations:

- Slope Overload: If the input signal changes too rapidly for the fixed step size to track, distortion occurs.
- Granular Noise: If the step size is too large when the signal is relatively flat, the staircase approximation "jumps" around the true value creating noise.

Adaptive Delta Modulation (ADM) addresses these issues by dynamically adjusting the step size based on the signal's characteristics.

Q.2.b) Explain H-261 and H-263 Video Compression techniques with suitable examples.

Ans



H.261 and H.263 are video compression standards developed by the ITU-T, primarily for videoconferencing and video telephony. They both ~~other~~ employ techniques to reduce spacial and temporal redundancy in video sequences.

H.261:

H.261 was the first international digital video compression standard, published in 1990, and designed for videoconferencing over ISDN (Integrated Services Digital Network) lines at bit rates ranging from 64 kbps to 2 Mbps.

Key Techniques

- Discrete Cosine Transform (DCT): Divides frame into 8x8 blocks and transforms spacial information into frequency coefficients, allowing for quantization and discarding of less significant data.
- Motion Compensation: For inter-coded frames (P-frames) it predicts the current frame based on a previous reference frame (I-frame) by estimating motion vectors for 16x16 macroblocks. Only the difference (residual) and motion vectors are then encoded.
- Quantization: Reduces the precision of DCT coefficients to achieve further compression.
- Variable Length Coding (VLC): Assigns shorter codes to more frequent symbols (e.g., small quantized coefficients) and longer codes to less frequent ones.

Example

In a videoconference where a person is speaking with a relatively static background, H.261 would encode the initial I-frame fully. Subsequent P-frames would primarily encode the movement of the speaker's mouth and facial expressions, using motion vectors to indicate how these parts have shifted from the previous frame, alongside the residual data for changes not captured by motion

H.263

H.263, introduced in 1996, aimed to improve upon H.261, particularly for very low bit-rate applications, such as video telephony over conventional telephone lines (POTS) at less than 64 kbps. It introduced several enhancements for better compression efficiency.

Key Enhancements over H.261

Half pixel Motion Estimation: Allows motion vectors to point to half-pixel positions, leading to more precise motion compensation and smaller residuals.

Bidirectional Predicted (B) frames: In addition to I and P frames, H.263 introduced B-frames, which can use both past and future frames for prediction, further improving temporal compression.

Advanced Prediction Modes: Includes options like 4 motion vectors per macroblock (one for each 8x8 block within a macroblock), and Overlapped Block Motion Compensation (OBMC) for smoother image transitions.

Expanded Picture Formats: Supported a wider range of resolutions beyond CIF and QCIF.

Example

In a low-bitrate video call over a modem, H.263's half pixel motion estimation could more accurately track ~~subtle~~ subtle head movements, reducing the amount of residual data needed. The use of B-frames could further compress sequences with ~~more~~ more complex motion, like a hand gesture, by predicting its position from both the proceeding and following frames in the sequence.

Q.1 a) What is a sound wave? Explain the characteristics of sound waves. Explain AIFF and VOC audio file formats

Ans. AIFF and VOC are both audio file formats, but they differ significantly in their origin, use, and features. AIFF (Audio Interchange File Format), developed by Apple, is a high-quality, uncompressed format with good metadata support, similar to the WAV format. In contrast, VOC (Creative Voice), developed by Creative Labs for its Sound Blaster cards, was an early format popular in the late 1980s and early ~~190~~ 1990s, primarily for PC sound and music.

AIFF (Audio Interchange File Format)

- Origin: Developed by Apple in 1988, based on the Interchange File Format (IFF).
- Quality: Typically uncompressed, resulting in high-fidelity audio similar to a CD, though a compressed version (AIFF-C) also exists.
- Features: Supports advanced metadata, including album artwork, song titles, and looping points, making it suitable for professional and editing use. It can also store MIDI data.
- File size: Larger file sizes due to the uncompressed nature, which can lead to longer download times and greater storage needs.
- Compatibility: Originally for Macs but now widely compatible across different platforms.

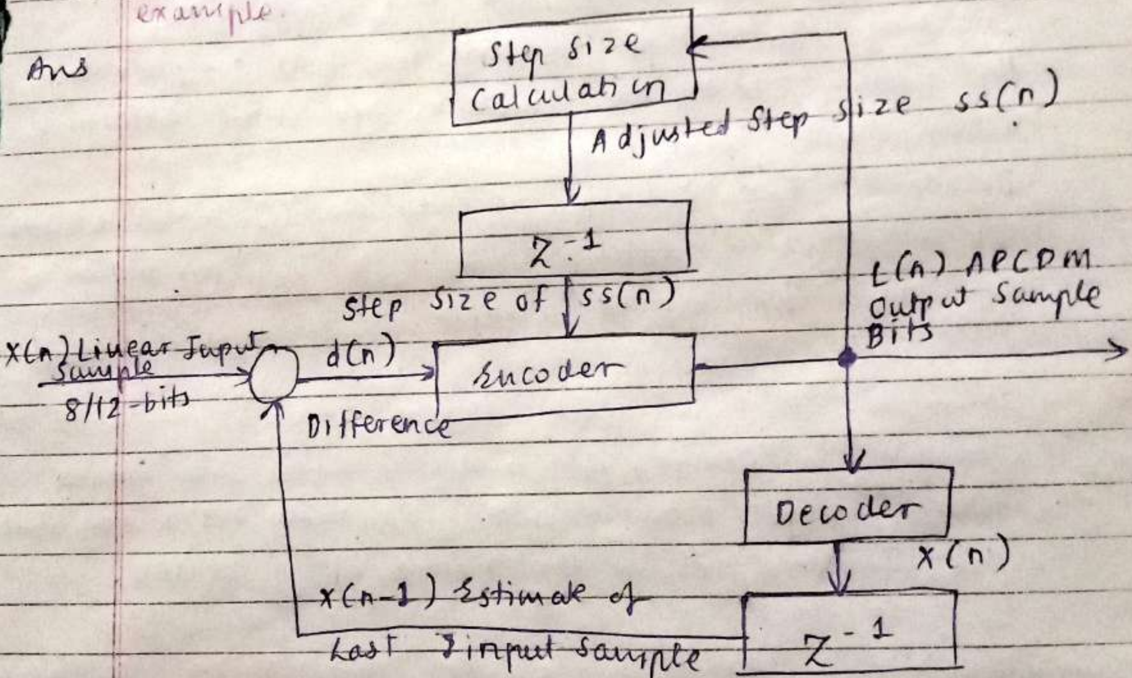
VOC (Creative Voice)

- Origin: Developed by Creative Technology for its Sound Blaster sound cards.
- Quality: A format for storing digital audio data.
- Features: Designed primarily for storing sound samples and music files on PCs

- File Size: Generally smaller than uncompressed formats like AIFF, as it was developed for a time when storage was more limited.
- Compatibility: Largely obsolete, but still can be encountered in older PC-based audio projects and recordings.

Q1 b) What is Audio compression? Explain the ADPCM Audio compression ~~tchr~~ technique with suitable example.

Ans



Audio Compression reduces digital audio data to make it more efficient for storage and transmission, and Adaptive Differential Pulse Code Modulation (APCDM) is a specific technique that does this by encoding the difference between predicted and actual audio samples. It works by using a prediction function to estimate the next sample, and then only quantizing the difference (error) between that prediction and the actual value, which is then transmitted & instead of the full sample.

How APCDM Works

- Prediction: A predictor analyzes the previous samples to predict the value of the value of the next sample.
- Quantization: The system calculates the difference (or "delta") between the predicted sample and the actual sample.
- Adaptation: The quantizer adapts to the size of this difference. If the difference is large, it uses more bits; if it is small, it uses fewer bits.
- Encoding: Instead of sending the full sample value, the system sends the smaller, quantized difference value.
- Decoding: The decoder receives the quantized difference. It adds this difference to its own prediction to reconstruct the original sample.

Example. A simplified ADPCM scenario

Imagine a series of audio samples with values that change slowly and predictably. Let's say that the samples are in the range of 0 to 100.

Sample Number	Actual Sample Value	Predicted Value	Difference
1	50	N/A	N/A
2	55	50	5
3	58	55	3
4	65	58	7

1. For Sample 1: The value 50 is sent as is, as there is no previous sample to predict from.
2. For sample 2: The decoder predicts the next sample will be 50. The difference is 55 - 5 = 50. The system sends the value 5 instead of 55.
3. For sample 3. The decoder predicts the next sample

will be 55 (using the previous actual value) The difference is 58 - 55 = 3 The system sends the value 3 instead of 58.

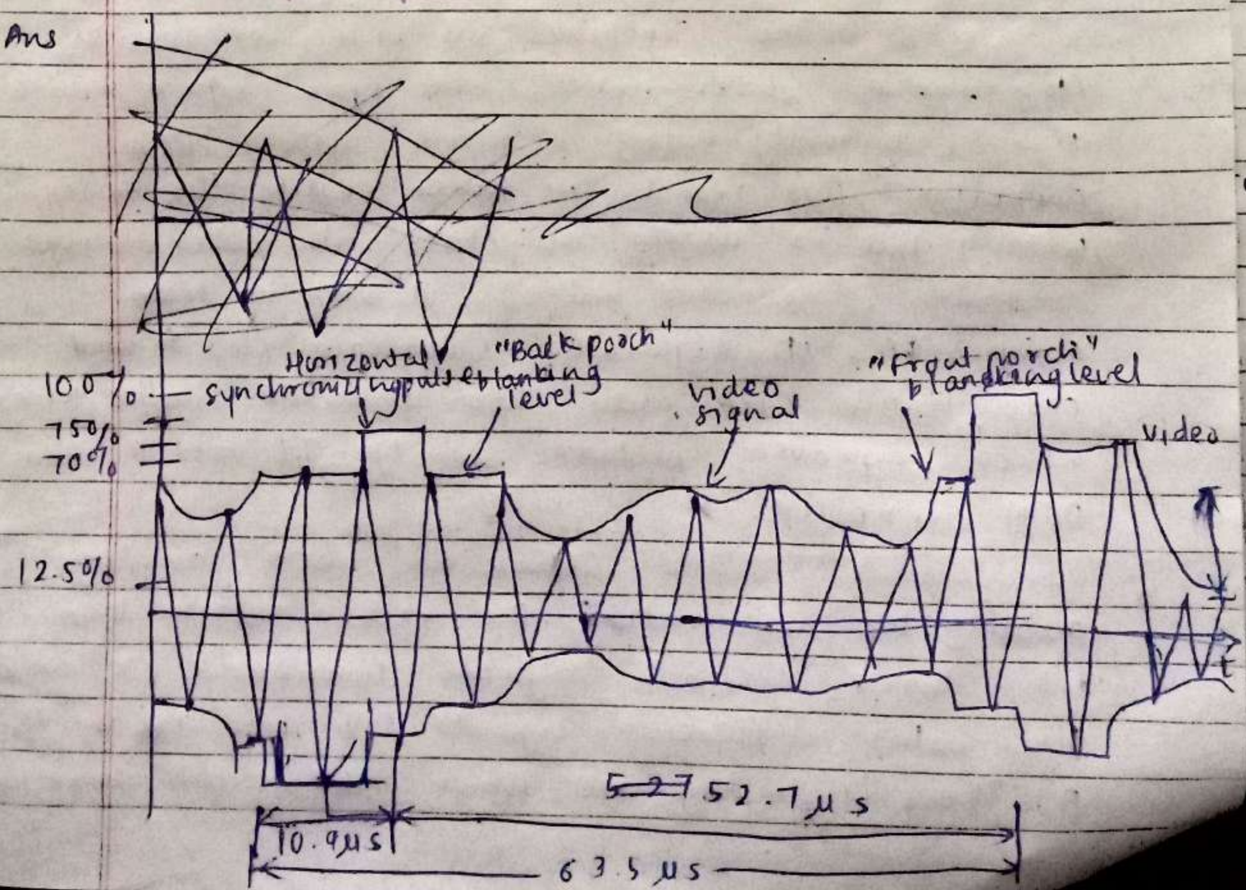
4. For sample 4: The decoder predicts the next sample will be 58. The difference is 65 - 58 = 7 The system sends the value 7 instead of 65

Compression ratio and bit reduction

By sending only the smaller "difference" values, which are often much smaller than the original sample values, the system significantly reduces the amount of data ~~stored~~ that needs to be transmitted or stored. The amount of data saved depends on the signal's characteristics and how well the predictor can estimate the next sample. In many cases, this method can achieve a high compression ratio, such as converting a 16-bit sample down to a 4-bit value.

Q.2.a) What is video signal? Explain the video signals format with a figure. Explain AVI and MOV video file formats.

Ans



A video signal is an electronic representation of visual information that allows moving images to be transmitted from a source (like a camera or computer) to a display device (like a TV or monitor). These signals incorporate luminance (brightness), chrominance (color), and synchronization pulses to ensure proper display alignment. Video signals are broadly categorized into analog and digital formats.

Video Signal Formats

Video signals can be analog, digital, or a hybrid, each with distinct methods for transmitting data. The format determines the quality, resolution, and compatibility with various devices

Analog Video Signals

Analog video signals use a continuous, time-varying electrical signal to represent images. Quality can degrade during transmission or copying due to noise interference.

- Composite Video Combines all video components (luminance, chrominance, and sync) into a single signal, typically using a single yellow RCA connector. This results in lower quality due to signal mixing but is widely compatible with older devices
- S-Video (Separated Video): A step up from composite, it separates luminance (Y) and chrominance (C) onto two separate signal lines, which improves picture clarity. It uses a mini-DIN connector.
- Component Video Offers the best analog quality by separating the video information into three distinct signals: luminance (Y) and two color difference signals (Pb and Pr, or R, G, B). These typically use three RCA or BNC connectors

- VGA (Video Graphics Array): An analog interface commonly used for computer monitors, transmitting Red, Green, Blue, horizontal sync, and vertical sync signals over a 15-pin connector.

Digital video signals.

Digital video signals use encoded binary data (0s and 1s) to represent moving images. They do not lose quality when copied and support much higher resolutions and frame rates than analog formats

- DVI (Digital Visual Interface): Designed for transferring digital video from a computer to a monitor. It can support digital-only (DVI-D), analog-only (DVI-A), or integrated (DVI-I) signals.
- HDMI (High-Definition Multimedia Interface): The current standard for home entertainment, it carries uncompressed digital video and multi-channel digital audio over a single cable.
- DisplayPort (DP): A high-performance interface designed to replace DVI and VGA, using packetized data transmission (like Ethernet) for high resolutions and multi-stream capabilities.

AVI and MOV Video File Formats

AVI and MOV are container formats, which define how audio, video, and other data (like subtitles) are stored in a single file, independent of the actual compression codec used.

AVI (Audio Video Interleave)

- Developer: Microsoft (1992)
- Primary Use: Archiving, general Windows compatibility
- Compression: Generally less efficient, results in larger files
- Quality: High quality (especially with less compression)

- Compatibility: Wide compatibility across many platforms/players
- Editing: Less editing friendly (simpler structure)

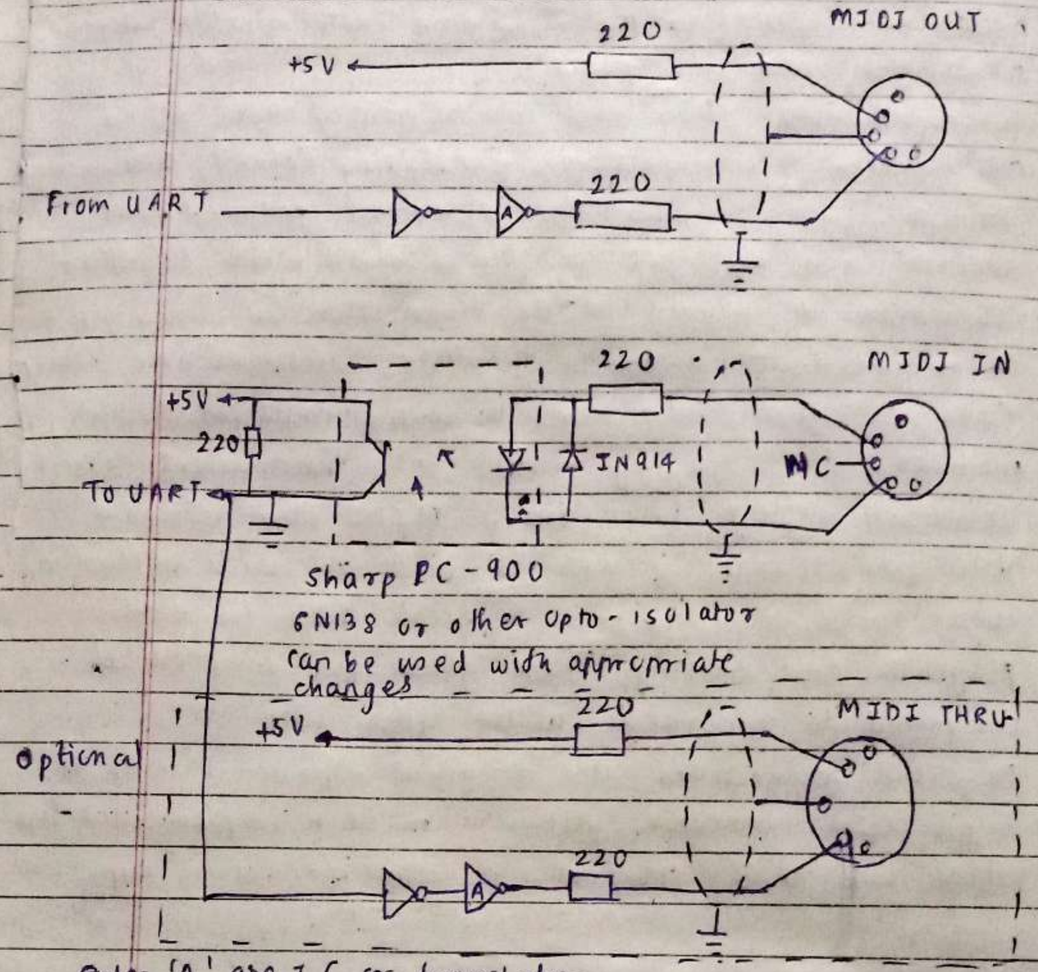
A solid, widely supported older format known for its high-quality output when minimal compression is used. However, the large file sizes and less flexible structure (e.g., it doesn't officially support metadata well) make it less ideal for web streaming compared to modern formats like MP4.

MOV (QuickTime File Format).

- Developer: Apple Inc. (1991)
- Primary Use: Professional editing, Apple ecosystem
- Compression: More efficient compression, smaller file sizes
- Quality: High quality (maintains quality with better compression)
- Compatibility: Best on macOS/iOS; requires third-party players on some non-Apple devices
- Editing: Excellent for editing (supports various tracks/metadata)

Developed for Apple's QuickTime, it is highly flexible and can manage multiple tracks of audio, video, and text simultaneously, making it a standard for professional video editing. It offers good compression efficiency and is optimized for the Apple ecosystem, though it is still widely supported by many other players, such as VLC media player.

* MIDI



Gates 'A' are IC or translator

The MIDI (Musical Instrument Digital Interface) standard is a technical protocol and digital interface that allows a wide variety of electronic musical instruments, computers, and related audio devices to communicate and exchange performance data. It does not transmit actual audio signals but rather digital instructions about how music is created, such as a note's pitch, timing, and velocity.

Core Concepts

- Messages, Not Sound: MIDI data is a sequence of commands (messages) that tell a receiving device which note to play, when to start and stop it, how loud it should be (velocity), and other expressive details like pitch blends or modulation. The actual sound produced depends on the receiving instrument's internal sound generator or "patch")

- Interoperability: Developed in the early 1980s by a collaboration of manufacturers (including Roland, Sequential Circuits, and Yamaha), the standard ensured that devices from different companies could work together.
- Channels: A single MIDI connection can carry up to 16 independent channels of data, allowing one device (like a keyboard controller) to control multiple sound modules, each playing a different instrument sound.
- Standard MIDI Files (SMF): MIDI also defines a file format (usually with a .mid or .midi extension) that stores timestamped musical data for later playback and editing. These files are very small compared to recorded audio files like MP3s.
- Physical Connections: The original standard uses a 5-pin DIN connector, although modern implementations often use USB, Ethernet, or other connection types.

Key Message Types

MIDI messages fall into several categories.

- Channel Voice Messages Transmit real-time performance data over specific channels, including:
- Note On/Off: Indicates when a note begins and ~~ed~~ ends, and with what velocity.
- Control Change (CC): Used for adjusting various instrument parameters like volume, panning, modulation, and ~~pedal~~ pedal use.
- Program Change: Changes the instrument sound (patch) being used on a specific channel.
- Pitch Band: Controls the continuous upward or downward bending of a note's pitch.
- System Messages Apply to all devices in the system.
- System Exclusive (SysEx): Manufacturer-specific

messages that allow in-depth control over proprietary features, beyond the standard MIDI commands

- System Real-Time: Messages for synchronization, such as timing clocks, start, stop, and continue commands.

Evolution to MIDI 2.0

The original MIDI 1.0 standard remained largely unchanged for decades, but the MIDI 2.0 standard was introduced in 2020 as an extension. MIDI 2.0 is a two-way (bi-directional) communication protocol, unlike the one-way MIDI 1.0, which allows devices to auto-configure and exchange information. It also provides higher resolution for controllers and velocity, better timing, and expanded expressive capabilities.

* WMA

WMA, which stands for Windows Media Audio, is a digital audio format developed by Microsoft for encoding and compressing audio files. It uses the .wma file extension and can offer various compression levels, from lossy compression (which reduces file size by sacrificing some audio quality) to lossless (which preserves full quality). The format is based on Microsoft's Advanced Systems Format (ASF) container and supports digital rights management (DRM) for copy protection.

Key Features of WMA

- Developed by Microsoft: WMA was created by Microsoft as an alternative to other audio formats like MP3.
- Compression: WMA is a compression format, meaning it reduces file sizes. Different WMA codecs are available to control trade-off between file size and audio quality

- Lossy: The original WMA codec is a lossy format, similar to MP3, that reduces file size
- Lossless: A WMA Lossless codec exists that compresses the audio without losing any quality
- WMA Voice: A version designed specifically for voice content, using very low bitrates
- Digital Rights Management (DRM): A key feature of WMA is its built-in support for DRM, which allows for the control of how audio files can be copied or played.
- Container format: WMA files are often stored within the Advanced Systems Format (ASF) container
- Compatibility: While it is natively supported on Windows, it ~~is~~ is compatible with many portable media players, although support on other operating systems like Android, is limited compared to more popular formats like MP3, notes Mastering The Mix.

* OGG.

OGG is an open-source, free and copyright-free container format that stores digital media, most commonly audio compressed with the Vorbis codec. It was developed by the Xiph.Org Foundation and is known for efficient compression that balances file size and audio quality.

Key Features of OGG:

- Container Format: OGG is a container, meaning it can hold different types of data streams, such as audio (Vorbis, Opus, FLAC) and video (Theora), along with metadata.
- Open and free: It is free to use for both commercial and non-commercial purposes, which has made it a popular choice for developers and content creators.

- Efficient compression: It uses codecs to compress audio and video, resulting in smaller file sizes while maintaining good quality, sometimes better than MP3 at similar bitrates.
- Streaming capability: The format was designed with streaming in mind, making it suitable for internet radio and other online content.
- Common file extensions: Common extensions include .ogg for audio, but others like .ogv (video) and .oga are also used.

OGG vs MP3

- Quality: At certain bitrates, OGG Vorbis can be considered to have better quality than MP3, especially at lower bitrates.
- Compatibility: MP3 is still more widely compatible with hardware devices than OGG, though OGG support has become more common in software.
- Licensing: OGG is royalty-free, while the MP3 format has associated ~~licensing~~ licensing fees.

* MP3

MPEG-1 Audio Layer 3 (MP3) is a popular digital audio format that uses lossy data compression to significantly reduce file size while preserving a high level of audio quality. Developed in the late 1980s and released as a standard in 1993, it became a dominant format for music storage, sharing, and playback across a wide range of devices due to its smaller file size compared to older formats like WAV

Key Features of MP3

- Compression: It removes audio data that is less perceptible to the human ear to make files smaller.

23/7/24

~~load local data inpath ('/student/ ... /hivedirectory');~~
~~create external table student (id int, name string, marks float)~~

~~row format delimited~~
~~fields separated by ','~~
~~location '/Hivedirectory';~~

- File size: MP3 files are a fraction of the size of uncompressed audio files, making them easy and to store and transfer.
- Compatibility: It is widely supported on computers, smartphones, portable media players, and various software programs.
- ID3 Tags: MP3 was the first audio file format to widely support ID3 tags, which store metadata like the song title, artist, and album.
- Bitrate: The quality of an MP3 is determined by its bitrate; higher bitrates (like 320 kbps) offer better quality than lower ones (like 120 kbps)

How it works

- An MP3 file is created through a process of encoding audio data using an encoder.
- This compressed audio data is then stored in an MP3 file, which can be played back on any device or software that supports the format.
- when playing the file, a decoder converts the compressed data back into an audio signal that can be heard through speakers or headphones.

* MPEG

MPEG audio compression is a lossy compression technique that works by exploiting the limitations of human hearing, such as masking, to remove or reduce information that is inaudible or imperceptible to the listener. Key stages include psychoacoustic modeling, which determines which sounds can be discarded, and applying an algorithm like Huffman or arithmetic coding to efficiently represent the remaining data in a smaller bitstream. The standard defines several layers (like Layers I, II, and III for MP3) offering different trade-offs between complexity, efficiency, and compression ratio.

Psychoacoustic Modeling

- Frequency masking: A loud sound can make a quieter sound in a close frequency range inaudible. MPEG removes or reduces data for the masked sound.
- Temporal masking: A loud sound can also make quieter sounds that occur immediately before (pre-masking) or after (post-masking) it inaudible.
- Threshold of hearing: This sets a lower limit on how quiet a sound can be and still be heard, regardless of masking.

Compression Process

1. Partitioning: The audio signal is divided into smaller blocks, and a filter ~~bta~~ bank is applied to separate the signal into different frequency sub-bands.
2. Psychoacoustic analysis: A psychoacoustic model analyzes each sub-band to determine the masking threshold and the "noise-masking threshold," which is the minimum level of data required to be kept for it to remain audible.
3. Bit allocation: Based on the psychoacoustic

model, bits are allocated to each sub-band. More bits are allocated to sub-bands that are more critical to the overall sound quality, while fewer are used for those that can be more heavily compressed.

4 Quantization: The data for each sub-band is quantized. The amount of quantization depends on the bit allocation, which, in turn, is determined by the psycho acoustic model. sub-bands with less importance are quantized more aggressively.

- Entropy coding: The quantized data is then encoded using an efficient entropy coding method, such as Huffman coding, to further reduce its size.

MPEG Layers and bit rates

- Layer I: The simplest, used in formats like MP1 and with fewer bits, but less efficient compression.
- Layer II: More efficient than Layer I, commonly used in formats like MP2.
- Layer III: The most efficient layer, used in MP3 format, which can achieve significant compression ratios of up to 1:10 while remaining perceptually lossless.
- MPEG-2 AAC: An advanced audio coding standard that is more efficient than MP3 and is used in formats like AAC.

* EDTV

EDTV, or Enhance Definition Television, is a digital video transmission standard that offers better picture quality than SDTV (Standard Definition Television) but not as good as HDTV (High Definition Television). It is primarily

17-09-2024 Thursday

U-III Communication Protocols and IoT Challenges

IEEE 802.11

sy sy sy A S E

defined by its 480p progressive scan resolution (or 576p in PAL regions), which provides a sharper, clearer image with reduced motion artifacts compared to the interlaced SDTV signals (480i)

Key characteristics

- Resolution: 480p (in NTSC regions) or 576p (in PAL regions).
- Scan type: Progressive scan, meaning each line of the image is drawn one after the other, resulting in a more stable picture than the interlaced scan of SDTV (480i/576i).
- Aspect ratio: Traditionally 4:3 (more square)
- Upconversion: EDTV sets often have the ability to "upconvert" lower-resolution 480i signals, such as from DVDs or older broadcasts, to 480p to display them with improved clarity.
- Audio: Some EDTV standards include enhancements like Dolby Digital surround sound.

* CCIR - Consultative Committee for Int'l Radio

CCIR video standards are a set of international analog and early digital video transmission specifications, with common analog examples like Systems B/G (625 lines, 50 Hz) and M (525 lines, 60 Hz). Digital standards like CCIR 601 were developed for broadcast quality and featured formats such as a 625-line interlaced scan with a 165 Mbps data rate. Other standards include System D (625 lines, with PAL/SECAM) and the Japanese Clear-Vision system, which was an enhanced version of System M.

Key Analog CCIR Standards

- System B: Used for VHF bands, with a 7 MHz channel bandwidth.

system G used for UHF channels, typically with system B and PAL color.
- System M Used in the Americas and parts of Asia, with a 525-line, 60 Hz frame rate.
- System D: A 625-line system used in Eastern Europe and China, often paired with PAL and SECAM.
- System L: Used in France, with a 625-line, 25 Hz frame rate and a 8 MHz channel bandwidth.

Key Digital CCIR Standards
- CCIR 601 (ITU-R BT. 601): A standard for digital component video intended for studio use, defining a resolution of 720×485 for NTSC systems and 720×576 for PAL systems.
- CIF (Common Intermediate Format): A lower-resolution standard for digital video, considered a temporary standard for a time.
- QCIF (Quarter-CIF): A lower-resolution format that is a quarter of the resolution of CIF.

Characteristics
- Analog: These systems used a variety of frame rates (like 25 or 29.97 fps), line counts (e.g., 525 or 625), and bandwidths (e.g., 6 or 7 MHz) to transmit video and audio signals.
- Digital: Digital standards aim to improve on analog systems by providing higher quality, standardized data rates, and resolutions. CCIR 601 specified a data rate of about 165 Mbps for NTSC systems, for example.
- Color systems: The analog standards sometimes didn't specify a color system, but were often paired with other standards like NTSC or PAL.

* CIF

CIF, or Common Intermediate Format, is a video transmission standard with a resolution of 352 x 288 pixels and a 4:3 aspect ratio, designed to act as a compromise between NTSC and PAL broadcast standards. It was created for early digital video applications like video conferencing to ensure consistent and compatible video processing across different systems and is now widely used in areas such as security surveillance and telemedicine.

Features

- Resolution: 352 x 288 pixels
- Aspect Ratio: 4:3
- Purpose: To create a common format for video conferencing systems to communicate with each other regardless of their regional broadcast standard (NTSC or PAL).
- Frame Rate: Based on NTSC's frame rate of 30000/1001 (approximately 29.97) frames per second.
- Color Encoding: Uses YCbCr with 4:2:0 color sampling.
- Variations:
  - QCIF (Quarter-CIF): 176 x 144 pixels
  - SQCIF (Sub Quarter-(CIF): 128 x 96 pixels
  - 4CIF: 704 x 576 pixels
  - 2CIF: 720 x 240 pixels.

* SIF

SIF, or Source Input Format, is a video standard developed by the MPEG commiltee for transmitting and storing digital video, particularly for applications like DVDs and older high-definition technologies. It defines specific resolutions, such as 352x240 for NTSC and 352x288 for PAL, and a frame rate of 29.97

fps for NTSC and 25 fps for PAL. SIF is derived from the ITU-R BT 601 standard and is a component of the MPEG-1 video format

Characteristics

- Source: Developed by the MPEG committee as an input format for video encoding.
- Purpose: To enable the storage and transmission of digital video for various applications.
- Color space: Uses the ~~YC~~ YCbCr colour coordinate system, where Y is luminance and Cb and Cr are color difference signals
- Relationship to other standards: Derived from the ITU-R BT. 601 standard and is the typical input format for MPEG-1 video.

* HDTV

The main HDTV transmission standards are the three video formats: 720p (1280 x 720 pixels, progressive scan); 1080i (1920 x 1080 pixels, interlaced scan), and 1080 p (1920 x 1080 pixels, progressive scan). These are part of the broader television standards, which use a widescreen 16:9 aspect ratio and support multiple audio channels, unlike older standard definition (SDTV) systems. The ITU-R BT. 709 recommendation is a key standard defining parameters like color space and quantization for HDTV.

Transmission Standards and Formats

- 720p: A progressive scan format with a resolution of 1280 x 720 pixels.
- 1080i: An interlaced scan format with a resolution of 1920 x 1080 pixels.
- 1080p: A progressive scan format with a resolution of 1920 x 1080 pixels, offering

the clearest and smoothest image for fast-moving content.

- ITU-R BT.709: A standard that defines the digital encoding of video signals, including color information, quantization, and the conversion from RGB to luminance and color difference signals.
- 16:9 aspect ratio: The standard widescreen format for HDTV, which is wider than the 4:3 aspec ratio used in standard definition television.
- Digital audio: HDTV supports multiple audio channels, a significant improvement over the two-channel audio of SDTV.

* How it works
- HDTV relies on digital broadcasting, which can be transmitted via cable, satellite, or over-the-air (terrestrial).
- Analog transmission is not capable of carrying the high-definition signal due to capacity limitations.
- The digital signal is compressed to reduce the required bandwidth for transmission.
- The receiver decodes the digital signal and reconstructs the image based on the chosen format (e.g., 720p, 1080i, or 1080p).

* RM FLV 3GP

RM, FLV, and 3GP are all video file formats, but they have different origins and uses. RM (RealMedia) was created by RealNetworks for streaming, FLV (Flash Video) was used by Adobe Flash Player for web video, and 3GP (3GPP) was designed specifically for older mobile phones with limited storage and bandwidth.

RM (RealMedia)

- Originally used for streaming video and audio over the internet
- Often associated with the RealPlayer software.

FLV (Flash video)

- A container format used by Adobe Flash player for delivering video on websites.
- Was very common for online video content before the widespread adoption of HTML5 video

3GP (3GPP)

- Designed for mobile phones, particularly for the 3G network.
- Files have low resolutions to save bandwidth and storage, making them ideal for mobile devices.
- Often used for Multimedia Messaging service (MMS) and Mobile TV broadcasts.

Which one to use?

- ~~For general us~~ If you're working with older mobile content or need a file that is optimized for low bandwidth, 3GP can be useful.
- FLV and RM are largely considered. legacy formats, especially as Flash is no longer supported.

* MPEG Video Compression

MPEG video compression is a lossy compression standard developed by the Moving Picture Experts Group that reduces video file sizes by exploiting temporal and spacial redundancy. It achieves this through two main techniques: motion compensation (predicting future frames from past ones) and Discrete Cosine Transform (DCT) (compressing the remaining differences). This method allows for high-quality video at smaller file sizes, making it suitabe for applications from streaming to professional editing.

How it works:

- Motion Compensation: Exploits temporal redundancy by finding and encoding the differences between frames, rather than re-encoding the entire frame.
- Frame Types: Uses different types of frames to achieve efficiency:
  - I-frames (Intra-coded): Keyframes that are compressed independently, like a still image.
  - P-frames (Predicted): Frames that are predicted from previous frames.
  - B-frames (Bi-directionally predicted): Frames that are predicted from both previous and future frames, providing the highest compression.
- Discrete Cosine Transform (DCT): Used to reduce spacial redundancy within each frame by transforming the pixel data into frequency coefficients.
- Quantization: Follows the DCT and is used to reduce the amount of information by lowering the precision of the frequency coefficients, especially the high-frequency ones, which are less perceptible to the human eye.
- Chroma Subsampling: Reduces the amount of color Information, as the human eye is less sensitive to changes in color than brightness.