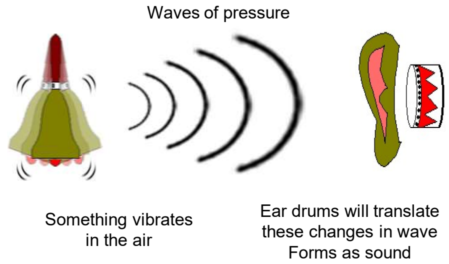
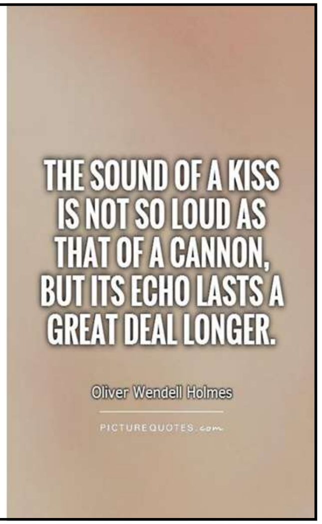
CHAPTER 5
FUNDAMENTAL OF AUDIO

IT4440
MULTIMEDIA
AND GAMES

WHY WE NEED SOUND ?



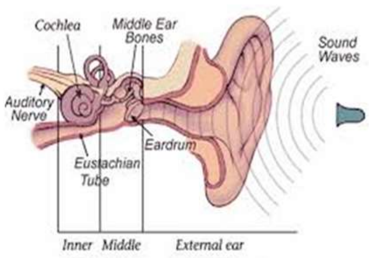


WHAT IS SOUND?

Sound is a type of energy which produces the sensation of hearing.

Sound comprises the spoken word, voices, music and even noise.

It is a complex relationship involving:
- a **vibrating object** (sound source)
- a **transmission medium** (usually air)
- a **receiver** (ear) and;
- a **preceptor** (brain).



Anatomy of the Ear

ELEMENTS OF SOUND PERCEPTION

A distinct use of the term *sound* from its use in physics is that in physiology and psychology, where the term refers to the subject of *perception* by the brain

There are six experimentally separable ways in which sound waves are analysed.

Pitch how "low" or "high"

Duration how "long" or "short"

Loudness how "loud" or "soft"

Timbre the quality of different sounds

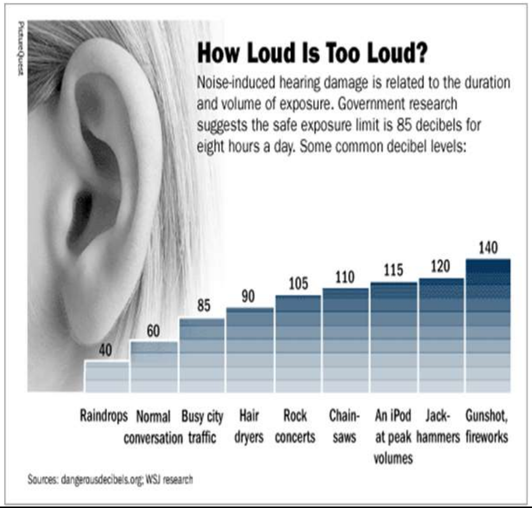
sonic texture the number of sound sources and the interaction between them

spatial location the cognitive placement of a sound in an environmental context

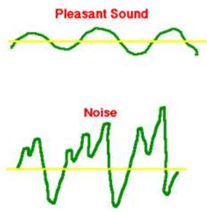
POWER OF SOUND

Sound pressure is measured in
→ **dB (decibel)**

Sound waves are known as **waveforms.**



SOUND

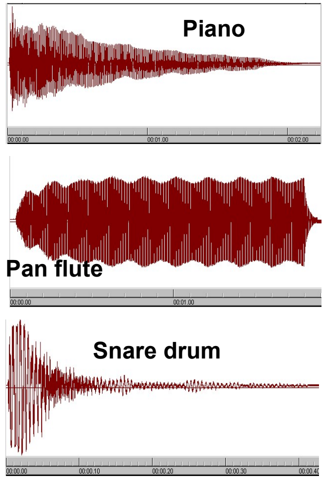


A pleasant sound has a regular wave pattern. The pattern is repeated over and over.

Noise is a term often used to refer to an unwanted sound

the waves of noise are irregular.

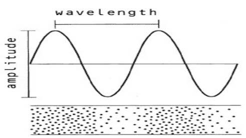
They do not have a repeated pattern.

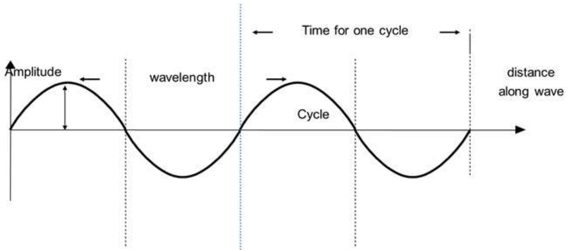


CHARACTERISTIC OF SOUND WAVES

Sound is described in terms of two characteristics:

- **Frequency** (or pitch)
- **Amplitude** (or loudness)





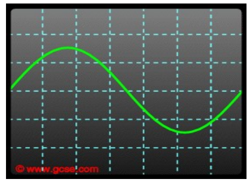
FREQUENCY

Frequency is a measure of how many vibrations occur in one second. This is measured in *Hertz* (abbreviation Hz) and directly corresponds to the ***pitch*** of a sound.

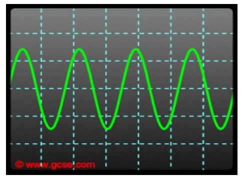
- The more frequent vibration occurs the higher the pitch of the sound.

Optimally, people can hear from **20 Hz to 20,000** Hz (20 kHz)

- Sounds below 20 Hz are infrasonic
- sounds above 20 kHz are ultrasonic.



Low pitch

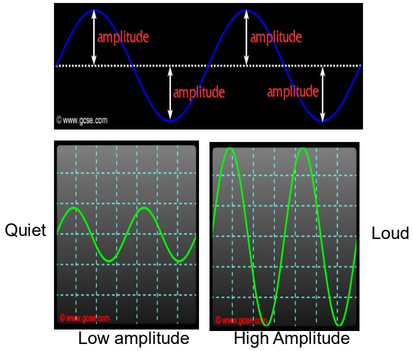


High pitch

AMPLITUDE

Amplitude is the *maximum displacement* of a wave from an equilibrium position.

- The louder a sound, the more energy it has and large amplitude.
- The amplitude relates to how loud a sound is.



ANALOGUE TO DIGITAL AUDIO

Analogue audio

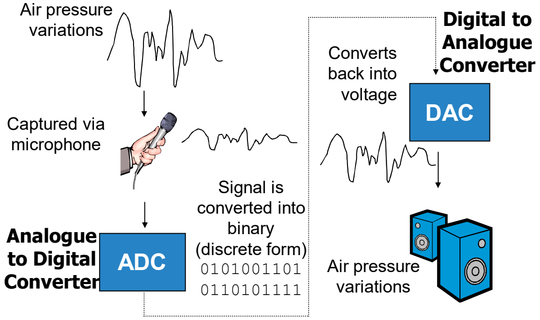
- non digital tape or audio tape recording of sound.
- an electronic signal that carries its information of sound as continuous fluctuating voltage value.

Digitizing the process of converting an analog signal to a digital one.

- A sound is recorded by making a **measurement of the amplitude** of the sound at regular intervals which are defined by the **"sample rate".**
  - The act of taking the measurement is often called "sampling"
  - aeach measurement is called a "sample point".

DIGITAL AUDIO SYSTEM

Digital audio data is the representation of sound, stored in the **form of samples point**.

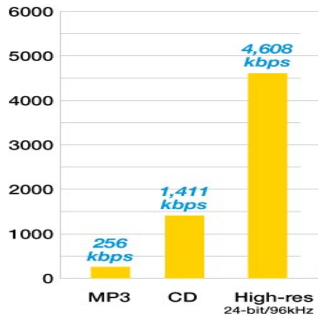


AUDIO QUALITY

Sampling frequency 44.1KHz
Quantization bits 16 (65536 intervals)

Quality factors for digital audio file (Quality of digital recording) depends on

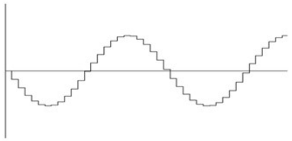
- Sampling Rate: the number of samples point taken per second (Hz).
- Sample Size (resolution) the number of bits used to record the value of a sample in a digitized signal.



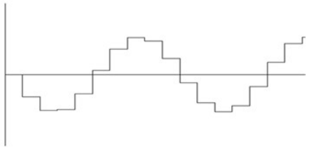
EXAMPLE

The three sampling frequencies most often used in multimedia are 44.1 **kHz, 22.05 kHz and 11.025 kHz.**

- The higher the sampling rate, the more the measurements are taken (better quality).
- The lower the sampling rate, the lesser the measurements are taken (low quality).



High Sampling Rate



Low Sampling Rate

Other than that, it also depends on:

- The **quality of original** audio source.
- The **quality of capture device** & supporting hardware.
- The characteristics used for capture.
- The capability of the playback environment.

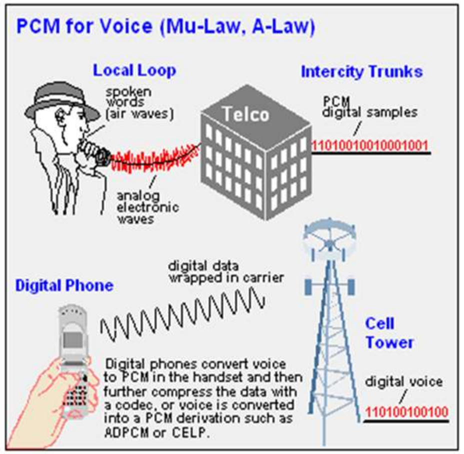
PCM

WHAT IS PCM ?

PCM - pulse-code modulation: The industry standard method for converting one analog voice signal into a digital signal

There are five basic steps to PCM: Sampling, Quantization, Companding, Encoding, Framing

It is the standard form of digital audio in computers, Compact Discs, digital telephony and other digital audio applications.



HISTORY

PCM was invented by P. M. Rainey of Western Electric in 1926

and later improved by British engineer Alec Reeves in 1937 while working for International Telephone and Telegraph in France.

He filed for a French patent in 1938, and his U.S. patent was granted in 1943

Low Noise

Long Distance

Data Storage

Encoded Signal

PULSE MODULATION

The transmission of analog data or speech which is in continuous form is known as pulse modulation.

At some certain levels or points, the wave formation can be seen in a pulse modulation system.

In this synchronizing, pulses are sent with the information related to the signal at different time samples.

There are two main categories in which pulse modulation can be divided.

Analog pulse modulation types
PAM(Pulse Amplitude Modulation)
PTM(Pulse Time Modulation)

Digital pulse modulation types
PCM(Pulse Code Modulation)
PDM(Pulse Delta Modulation)

HOW PCM WORKS ?

Sampling

Quantizing

Companding

Encoding

Framing

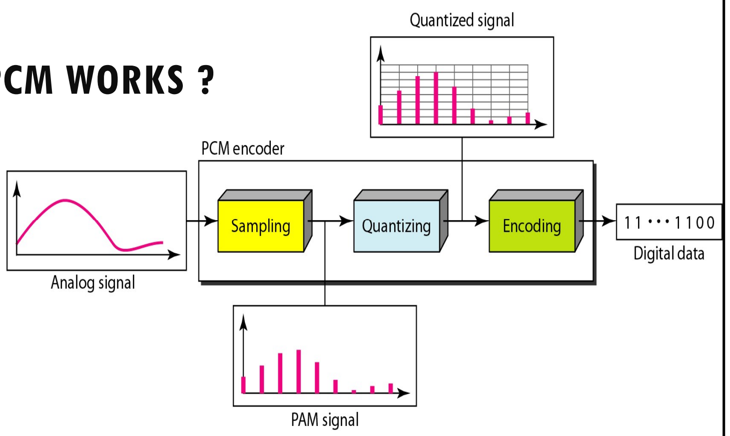


Figure 01 : Block Diagram of Pulse Code Modulation

SAMPLING

Analog signal is sampled every T_S secs.

T_s is referred to as the sampling interval.

$f_s = 1/T_s$ is called the sampling rate or sampling frequency.

There are 3 sampling methods:

- Ideal Natural Flattop

The process is referred to as pulse amplitude modulation PAM and the outcome is a signal with analog (non integer) values

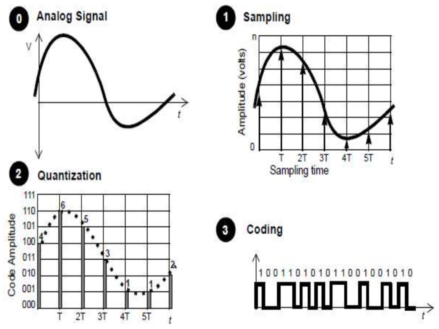
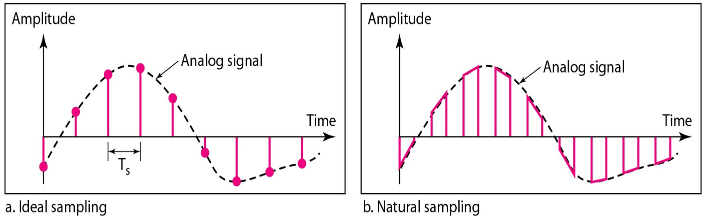


Figure 1.2 The three steps of digitalization of a signal: sampling of the signal, quantization of the amplitude, and binary encoding.



There are 3 sampling methods:

- Ideal - an impulse at each sampling instant
- Natural - a pulse of short width with varying amplitude
- Flattop - sample and hold, like natural but with single amplitude value

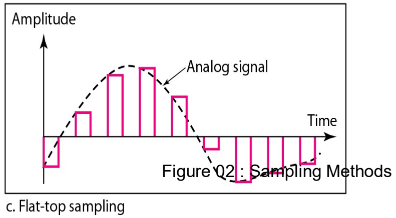


Figure 02 : Sampling Methods

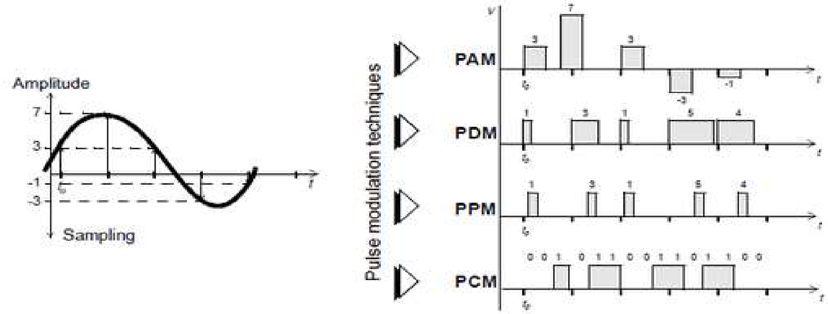
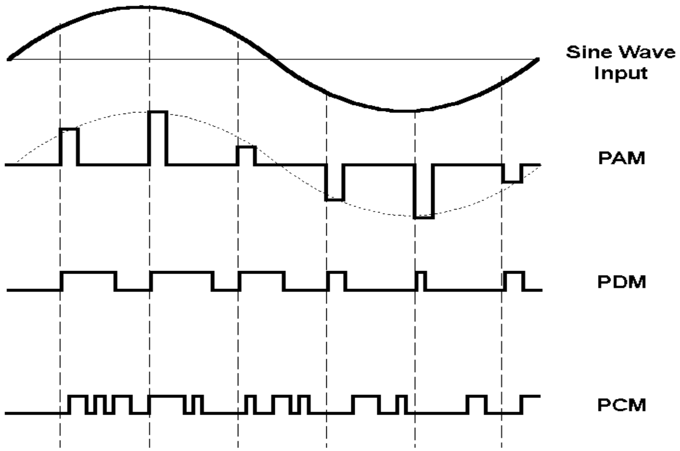


Figure 1.1 Pulse code modulation (PCM) was the technology selected to digitalize the voice in telephone networks. Other pulse techniques are pulse amplitude modulation (PAM), pulse duration modulation (PDM), and pulse position modulation (PPM).



Pulse Modulation Techniques

SAMPLING THEORY (NYQUIST)

- If input signal has maximum frequency (bandwidth) f, sampling frequency must be at least 2f
- With a low-pass filter to interpolate between samples, the input signal can be fully reconstructed

Example : highest frequency of telephone voice channel is 3.4 kHz

Sampling rate ≥ 2 x 3.4

≥ 6.8 kHz

Hence sample rate is 8 kHz

QUANTIZATION

The samples are divided into many discrete levels.

Then each sample is numbered according to their corresponding level.

Quantizing a signal will result in some distortion, called quantizing noise, it is greater for low-amplitude signals.

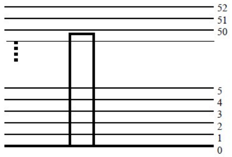
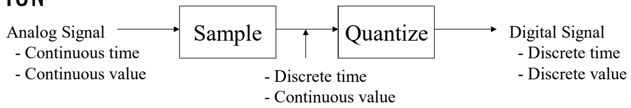
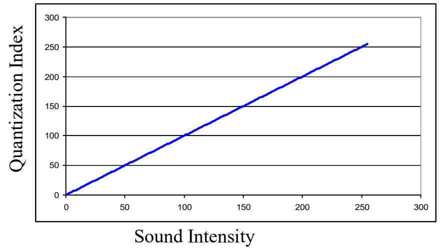


Figure 03 : Linear Quantization

LINEAR QUANTIZATION



It is also the second step required by A/D conversion.



PERCEPTUAL QUANTIZATION (U-LAW)

Quantization is a non linear transformation which maps elements from a continuous set to a finite set.

Want intensity values logarithmically mapped over N quantization units

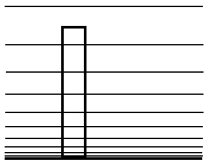
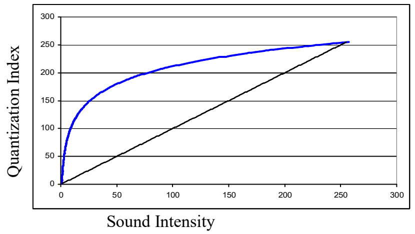


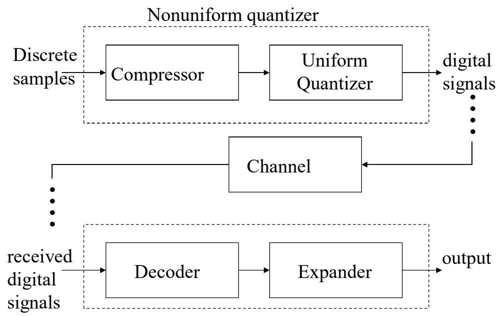
Figure 04 : Non Linear Quantization



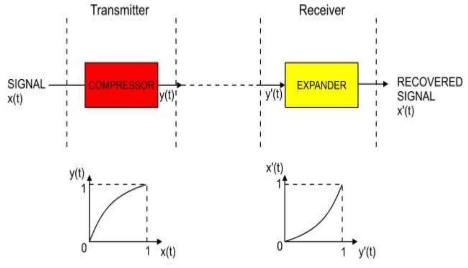
"COMPANDING."

Companding is the process of compressing and then expanding

- The higher amplitude analog signals are compressed prior to transmission and then expanded
- The signal is compressed at the transmitter to divide low-amplitude signals into more levels.
- Companding is used to reduce quantizing noise.



Companding Process



Quantizing a signal will result in some distortion because we do not code the exact voltage of the input signal.

This distortion is called quantizing noise, it is greater for low-amplitude signals.

Companding is used to reduce quantizing noise.

The signal is compressed at the transmitter to divide low-amplitude signals into more levels.

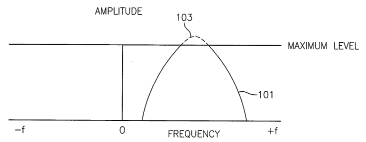
When the signal is decoded at the receiving system, it is expanded by reversing the compression process.

WHAT IS M LAW ?

The μ-law algorithm is a companding algorithm, primarily used in 8-bit PCM in North America and Japan

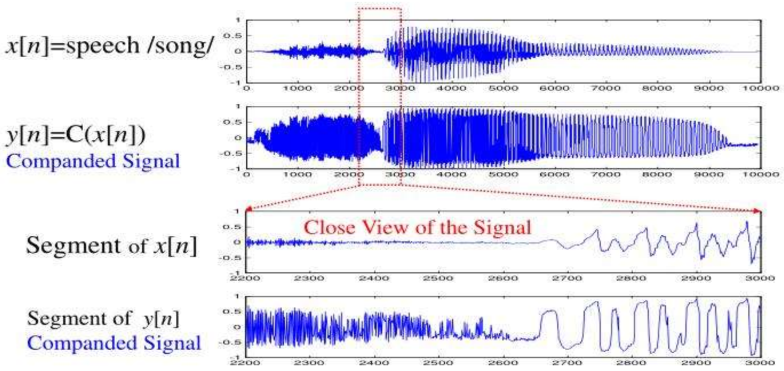
It's based on the observation that many signals are statistically more likely to be near a low signal level than a high signal level.

Most quality codecs use μ -law encoded samples.



30

Example: μ-law Companding



WHAT IS A LAW ?

A-Law is the standard codec algorithm for p modulation (PCM) from the ITU-T (the Teleco Standardization Sector of the International Telecommunications Union.

A-Law compression is extremely similar t compression

Standard companding algorithm used in digital communications systems (telephor part) to optimize the dynamic range of a (generally a voice) for digitizing, i.e.,

PCM – Companding Rules

- μ-law compressor
  - $y = \ln(1+\mu x)/\ln(1+\mu)$
  - μ is degree of compression
  - Used in North America telephone systems
  - Soft voices actually get amplified while stro suppressed
- A-law compressor
  - $y = Ax/(1 + \ln A)$ for $0 \le x \le 1/A$
  - $y = (1 + \ln Ax)/(1+ \ln A)$ for $1/A \le x \le 1$
  - Used in Europe and followers

A LAW AND M LAW

- Both methods compress 2:1 ratio.

  μ -Law has a larger dynamic range compared to A-law

  μ -Law has worse distortion with small signals compared to A-law

  μ -Law is used in North-America and Japan while A-law is commonly used in Europe

  A-law takes precedence over μ -law with international calls

μ-law and A-law

Widely used compression algorithms

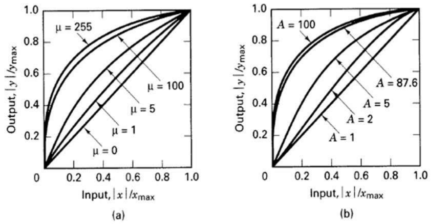


Figure 2.20 Compression characteristics. (a) μ-law characteristic. (b) A-law characteristic.

ENCODING

After compression and quantization of the input signal, it will be one of the 256 discrete signal levels that can be assigned an 8-bit code.

The process of assigning an 8-bit code to represent the signal level is known as encoding.

We represent levels in binary format.

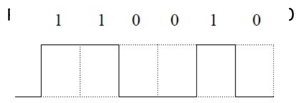


Figure 05 : Representation of sample signal

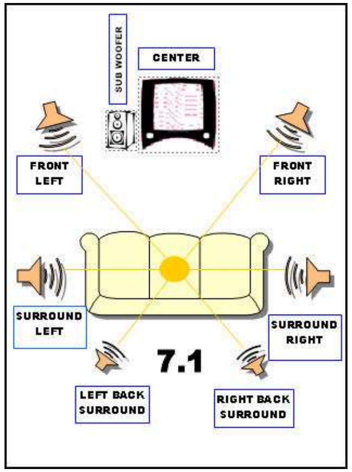
FRAMING

A frame is simply a set of samples with the same timestamp, one for each channel.

E.g. if your audio is mono, 1 frame is 1 sample.

If your audio is stereo, 1 frame is 2 samples (left + right).

When we say 44100 kHz we mean 44100 samples *per channel* per second, i.e. 44100 frames per second.



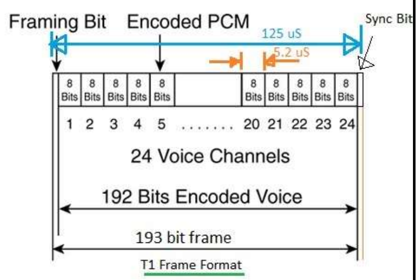
FRAMING

The encoded 8-bit signal is time division multiplexed with 23 other 8-bit signals to generate 192 bits for 24 signals.

A single framing bit is added to these 192 bits to make a 193-bit frame.

Framing BitsFollow an established pattern of 1s and 0s for 12 frames (1,0,0,0,1,1,0,1,1,1,0,0) and then repeat in each of the succeeding 12 frames.

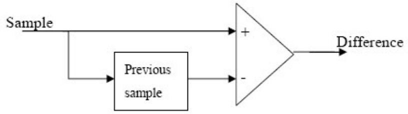
This pattern is used by the receiving terminal to stay synchronized with the received frames



DIFFERENTIAL PULSE CODE MODULATION (DPCM)

What if we look at sample differences, not the samples themselves?

- $d_t = x_t - x_{t-1}$
- Differences tend to be smaller
  - Use 4 bits instead of 12, maybe?



Differential pulse-code modulation (**DPCM**) is a signal encoder that uses the baseline of pulse-code modulation (PCM) but adds some functionalities based on the prediction of the samples of the signal

DIFFERENTIAL PCM

Differential pulse-code modulation is a signal encoder that uses the baseline of pulse-code modulation but adds some functionalities based on the prediction of the samples of the signal.

For the signals which does not change rapidly from one sample to next sample.

By knowing the past behavior of a signal up to a certain point in time, it is possible to make some inference about the future values

DM **DELTA MODULATION**

A delta modulation (DM or Δ-modulation) is an analog-to-digital and digital-to-analog signal conversion technique used for transmission of voice information where quality is not of primary importance.

DM is the simplest form of differential pulse-code modulation (DPCM) where the difference between successive samples are encoded into n-bit data streams

PCM	Delta modulation
PCM is a method used to digitally represent sampled analog signals.	Digital to analog convert
Encoder-Decoder	Modulator
Signal needs encode-decode both sides	Signal can modulate-demodulate
Large process	Easy process
Costly	Cheap

ADAPTIVE DIFFERENTIAL PULSE CODE MODULATION (ADPCM)

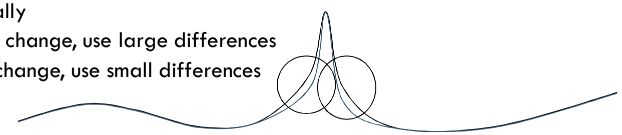
(ADPCM) is a method used to convert analog signals to binary signals.

The technique converts the analog signals by taking frequent samples of the sound and representing the value of the sampled modulation in binary form.
The technique is a variation of the digitized method known as pulse code modulation.

Adaptive similar to DPCM, but adjusts the width of the quantization steps

Encode difference in 4 bits, but vary the mapping of bits to difference dynamically

- If rapid change, use large differences
- If slow change, use small differences

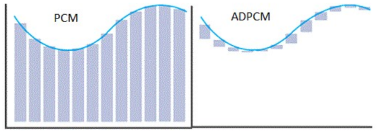


ADAPTIVE DIFFERENTIAL PCM

ADPCM, commonly termed as a form of compression, is a more efficient way of storing waveforms than 16-bit or 8-bit PCM.

ADPCM stores the value differences between two adjacent PCM samples and makes some assumptions that allow data reduction.

Because of these assumptions low frequencies are properly reproduced



ADAPTIVE DIFFERENTIAL PCM

An ADPCM algorithm is used to map a series of 8 bit µ-law (or a-law) PCM samples into a series of 4 bit ADPCM samples. In this way, the capacity of the line is doubled.

Some ADPCM techniques are used in Voice over IP communications.

ADPCM was also used by Interactive Multimedia Association for development of legacy audio codec known as ADPCM DVI, IMA ADPCM or DVI4

Techniques to compress general digital audio signals include µ -law and adaptive differential pulse code modulation.

These approaches apply low-complexity, low-compression, and medium audio quality algorithms to audio signals.

APPLICATIONS

Audio Compression Format	Algorithm	Bit Rate	Bits per sample	Implementations
G.711	Companding A-law ; µ-law ; PCM	64 kbit/s	8 bit	FFmpeg, Ekiga, Asterisk (PBX)
G.711.1	A-law ; µ-law	64, 80, 96 kbit/s	16 bit	VoIP
G.721	ADPCM	32 kbit/s	13 bit	
G.722	ADPCM	64 kbit/s	14 bit	QuickTime, RealPlayer

ADVANTAGES & DISADVANTAGES OF USING AUDIO

Sound adds life to any multimedia application and plays important role in effective marketing presentations.

Advantages

- Ensure important information is noticed.
- Add interest.
- Can communicate more directly than other media.

Disadvantages

- Easily overused.
- Requires special equipment for quality production.
- Not as memorable as visual media.