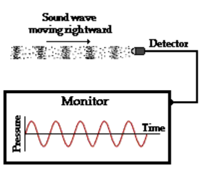
Multimedia Software Systems
CS4551

Audio Compression

What is Sound?

- Physics Introduction
  - Sound is a waveform like light.
  - It involves molecules of air being compressed and expanded under the action of some physical device.
  - Without air, there is no sound.
- A speaker in an audio system
  - Vibrates back and forth and produces a longitudinal pressure wave that we perceive as sound.
- Recording instruments convert the pressure wave to an electrical waveform signal, which is then sampled and quantized to get a digital signal.

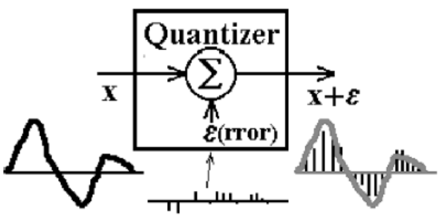


What is Sound? (2)

- High **frequency** sound correspond to high **pitch** (degree of highness and lowness) sound.
- **Tone** is a quality of a sound. i.e. different musical instrument produces different sounds despite that they play same frequencies.
- Amount energy of sound which spans a given radius is **intensity** of sound.

Sound - Sampling and Quantization

- From an analog signal (continuous measurement of pressure wave, eg. Continuous-valued voltages produced by microphones) to a digital signal



- Sampling – digitization in time dimension (Nyquist Theorem)
- Quantization - digitization in the amplitude dimension
  - Quantization introduces error! – Listen to 16, 12, 8, 4 bit music and see the difference.

Sound - Sampling and Quantization (2)

- Digital audio is represented as a one- dimensional set of samples.
- Digital audio is represented in the form of channels.
  - The number of channels describes whether the audio signal is mono, stereo, or even surround sound.
  - Each channel consists of a sequence of samples and can be described by a sampling rate and quantization bits.

Sound - Sampling and Quantization (3)

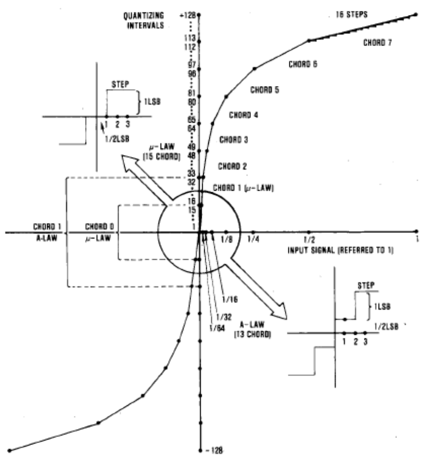
- Telephone-quality speech
  - Sampling rate = 8KHz ,
  - Quantization = 16bits/sample
  - Bit rate is = 8K x 16 = 128 Kbps
- CDs (stereo channels)
  - Sampling rate =44.1KHz
  - Quantization = 16bits/sample
  - Bit rate is 2 x 16 x 44100 =1.4 Mbps!
  - CD Storage = 10.5 Megabytes/minute
  - CD can hold on 70 minutes of audio
- Surround Sound Systems with 5 channels.
  - Dolby AC-3 used in many cinema uses 5.1 channels. ".1" represents subwoofer.

Non-Linear Quantization

- Assume the speech signal is normalized to the range [-1, 1]. If we examine a typical speech signal and its histogram, we shall see that we rarely use the extreme values +1 and -1.
  – In a linear quantization scheme, we assign as many reconstruction levels for larger amplitudes as for smaller amplitudes, which are more probable to occur.

Non-Linear Quantization

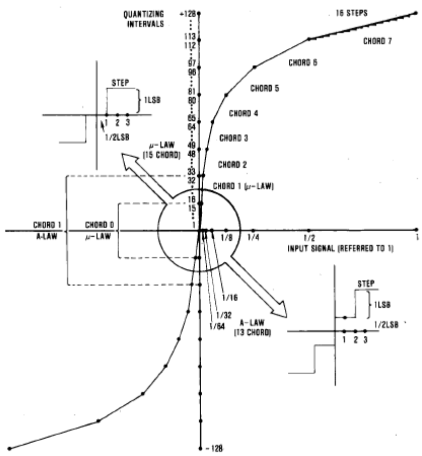
- A-law, µ-law
- Companding Law - (COMPression - expANDING) schemes
- Both are used in telephone networks.
- They expand small values and compress large values.
- When a signal goes through a compander, small amplitudes are mapped into a larger interval and larger amplitudes are mapped into a smaller interval.



Non-Linear Quantization

- More quantization levels are used for the values that originated from small amplitudes.

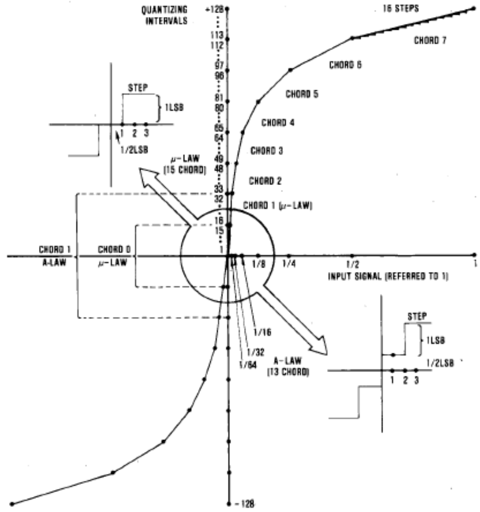
- This scheme is equivalent to applying non-uniform quantization to the original signal, where smaller quantization levels are used for smaller values and larger quantization levels are used for larger values.



μ-law/A-Law Coding

- Unlike linear quantization, the logarithmic step spacing represents low-amplitude audio samples with greater accuracy than it does higher amplitude samples.
- A-law is used in European telephone network.
- μ-law is used in America and Japan.
- The International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) Recommendation G.711 codifies the A-law and μ-law encoding scheme.

µ-law/A-Law Coding



It compensates for fact that quantization error much more audible around 0 amplitude.

It results in a SNR that is superior to that obtained by linear encoding for a given number of bits.

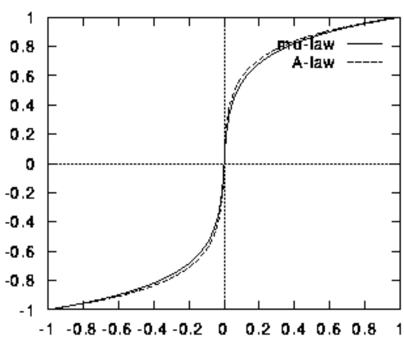
µ-law/A-Law Coding

$$F(x) = \frac{\operatorname{sgn}(x) * \ln(1+\mu|x|)}{\ln(1+\mu)} \qquad 0 \le |x| \le 1$$

u Law Equation

$$F(x) = \begin{cases} \dfrac{A*|x|}{1+\ln(A)} & 0 \le |x| < \dfrac{1}{A} \\[2ex] \dfrac{\operatorname{sgn}(x)*(1+\ln(A|x|))}{1+\ln(A)} & \dfrac{1}{A} \le |x| \le 1 \end{cases}$$

A-Law Equation



A Comparison to the Visual Domain

- Sound is a 1D signal with amplitude at time *t*
- Then, should it be simple to compress sound compared to a 2D image signal and 3D video signals?
- Consider human perception factors - human auditory system is more sensitive to quality degradation than the visual system. As a result, humans are more prone to compressed audio errors than compressed image and video errors.
- Compression ratios attained for video and images are greater than those attained for audio.

Audio Compression

- We need to take advantage of redundancy/correlation in the signal by statistically studying the signal –but just that is not enough!
- The amount of redundancy that can be removed all through out is very little and hence all the coding methods for audio generally give a lower compression ratio than images or video.
- Any other techniques?

Types of Audio Compression Techniques

- Audio Compression techniques can be broadly classified into different categories depending on how sound is “understood”.
- Sound is a Waveform
  - Use Statistical Distribution /etc.
  - Not a good idea in general by itself
- Sound is Perceived (Perception-Based)
  - Psycho acoustically motivated
  - Need to understand the human auditory system
- Sound is Produced

Sound as Waveform

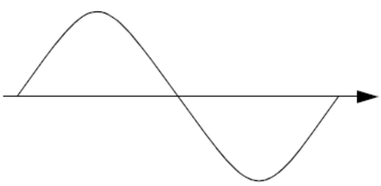
- Waveform Compression Techniques
  - Uses variants of PCM techniques.
  - PCM (Pulse Code Modulation)
  - DPCM (Differential PCM)
  - DM (Delta Modulation), Adaptive DM
  - ADPCM (Adaptive DPCM)

PCM (Pulse Code Modulation)

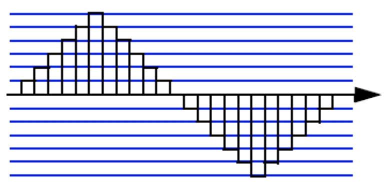
- PCM is a formal term for the sampling and quantization. It involves sampling rate and quantizer (uniform or non-uniform).
  - Speech (8KHz, 16bits/sample)
  - CDs music - (stereo channels, 44.1KHz, 16bits/sample)

PCM (Pulse Code Modulation)

each sample *f(n)* ⟶ $\boxed{\text{Quantizer}}$ ⟶ quantized sample *f(n)* in *k* bits



analog signal

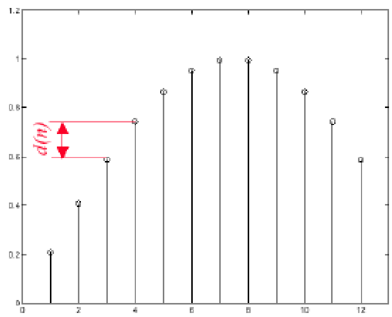


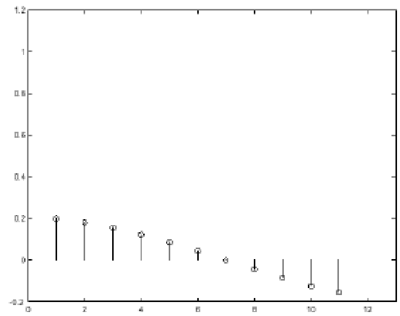
digital signal

DPCM (Differential PCM)

- Predictive coding :
  - *Let (n-1)*th sample be $f(n-1)$.
  - In general, use $f(n-1)$ as the *predicted* value for $f(n)$.
- Differential PCM Encoding *(DPCM):*
  - Don't quantize and transmit sample $f(n)$ directly
  - Compute the residual $e(n) = f(n) - f(n-1)$. Quantize $e(n)$ and transmit $\underline{e}(n)$ (let's say that the quantized $e(n)$ is $\underline{e}(n)$)
- We can show that SNR $_{DPCM}$ > SNR $_{PCM}$

DPCM





DPCM - Detail

f_n : input signal

$\tilde{f}_n$: reconstructed signal

$\hat{f}_n$: predicted signal

defined as a function of previous reconstructed values

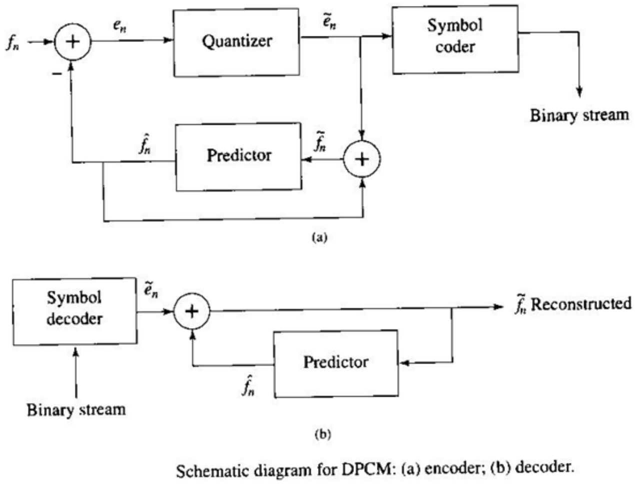
$e_n = f_n - \hat{f}_n$

$\tilde{e}_n = Q[e_n]$

Transmit $\tilde{e}_n$.

The decoder reconstructs the signal using $\tilde{f}_n = \hat{f}_n + \tilde{e}_n$.

Closed-Loop DPCM



Schematic diagram for DPCM: (a) encoder; (b) decoder.

Closed-Loop DPCM

- Example
  - Consider the sequence of 130 150 140 200 230.

$$\tilde{e}_n = Q[e_n] = 16 * \text{trunc}\left[(255 + e_n)/16\right] - 256 + 8$$
$$\tilde{f}_n = \hat{f}_n + \tilde{e}_n$$

f^$_n$ = trunc[(f~$_{n-1}$ + f~$_{n-2}$)/2] for prediction

- Assume that the first value will be transmitted without loss. What is the encoded values using the following prediction and quantization scheme? What is the decoded result?

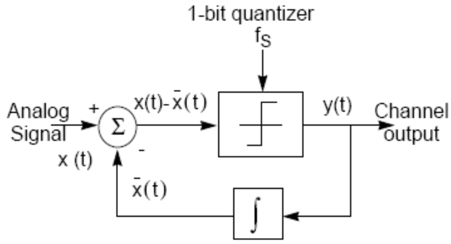
$$\begin{aligned} \hat{f} &= \boxed{130}, \quad 130, \quad 142, \quad 144, \quad 167 \\ e &= \boxed{0}, \quad 20, \quad -2, \quad 56, \quad 63 \\ \tilde{e} &= \boxed{0}, \quad 24, \quad -8, \quad 56, \quad 56 \\ \tilde{f} &= \boxed{130}, \quad 154, \quad 134, \quad 200, \quad 223 \end{aligned}$$

DPCM quantizer reconstruction levels

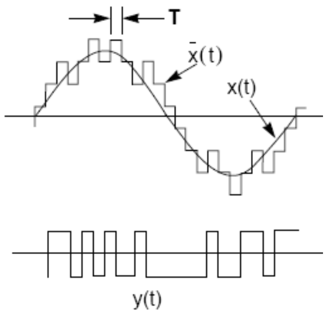
e_n in range	Quantized to value
−255 .. −240	−248
−239 .. −224	−232
⋮	⋮
−31 .. −16	−24
−15 .. 0	−8
1 .. 16	8
17 .. 32	24
⋮	⋮
225 .. 240	232
241 .. 255	248

DM (Delta Modulation)

- Delta Modulation
  - Like DPCM but only encodes differences using a single bit suggesting a delta increase or a delta decrease
  - Good for signals that don't change rapidly



(a) Modulation



DM

f_n : input signal

$\tilde{f}_n$: reconstructed signal

$\hat{f}_n = \tilde{f}_{n-1}$

$e_n = f_n - \hat{f}_n$

$$\tilde{e}_n = \begin{cases} +delta \text{ if } e_n > 0, \text{ where } delta \text{ is a constant} \\ -delta \text{ otherwise} \end{cases}$$

Transmit $\tilde{e}_n$.

The decoder reconstructs the signal using $\tilde{f}_n = \hat{f}_n + \tilde{e}_n$.

DM (Delta Modulation)

- Consider the samples 10, 11, 13, and 15. Suppose delta = 4.
  - Delta encoder uses 1-bit encoder for encoding differences using a fixed delta value.
  - Let's use *1 for delta increase* and *0 for delta decrease*, then the encoder generates 10, **1, 0, 1.**
  - The decoder reconstructs 10, 14, 10, 14.

$$\begin{array}{llll} \hat{f}_2 = 10, & e_2 = 11 - 10 = 1, & \tilde{e}_2 = 4, & \tilde{f}_2 = 10 + 4 = 14 \\ \hat{f}_3 = 14, & e_3 = 13 - 14 = -1, & \tilde{e}_3 = -4, & \tilde{f}_3 = 14 - 4 = 10 \\ \hat{f}_4 = 10, & e_4 = 15 - 10 = 5, & \tilde{e}_4 = 4, & \tilde{f}_4 = 10 + 4 = 14 \end{array}$$

Adaptive DM

- Adaptive DM
  - If the slope of the actual curve is high, the staircase approximation cannot keep up. Adaptive DM changes step size *delta* adaptively in response to the signal's current properties.
  - One way to change the *delta* size adaptively :
    - The encoder considers the previous N bits of output ($N = 3$ or $N = 4$ are very common) to determine adjustments to the *delta* size.
    - If the previous N bits are all 1s or 0s, the step size is doubled.
    - Otherwise, the step size is halved.
    - The step size is adjusted for every input sample processed.

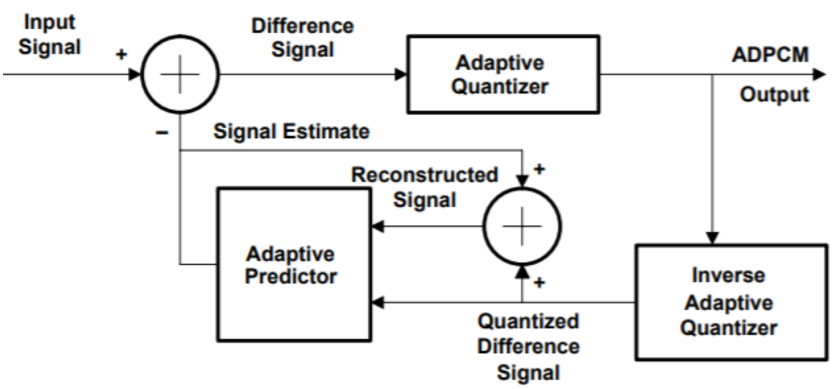
ADPCM (Adaptive DPCM)

- Adaptive Differential Pulse Code Modulation
  - Sophisticated version of DPCM.
  - Adapts predictor to signal characteristics
  - Also adapts width of quantization steps to signal characteristics
  - Better quality than DPCM with same storage requirements

ADPCM (Adaptive DPCM)

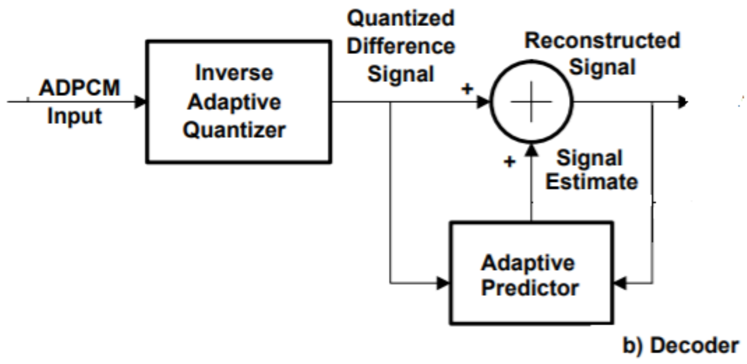
- Like DPCM, it codes the differences between the quantized audio signals using only a small number of specific bits which adaptively vary by signal.
- For example, G.726 ADPCM encodes difference in 4 bits, but vary the mapping of bits to difference dynamically.

ADPCM (Adaptive DPCM) - Encoder



a) Encoder

ADPCM (Adaptive DPCM) - Decoder



ADPCM (Adaptive DPCM)

- Adaptive Quantization
  - If rapid changing signal that produce difference with large fluctuations, use large differences.
  - If slow changing signal that produce difference signals with small fluctuations, use small differences.

- Adaptive Prediction
  - Generally, change the coefficient *a* of the prediction

ADPCM (Adaptive DPCM)

- Prediction is usually done based on M Previous Values (previously reconstructed quantized values)

$$\hat{f}_n = \sum_{i=1}^{M} a_i \tilde{f}_{n-i}$$

- ADPCM adaptively change a_i values that minimizes

$$\min \sum_{n=1}^{N} (f_n - \hat{f}_n)^2$$

- Simply solve

$$\min \sum_{n=1}^{N} (f_n - \sum_{i=1}^{M} a_i f_{n-i})^2$$

Audio Coding : Main Standards

- ITU **Speech** Coding Standards
  - ITU G.711
  - ITU G.722
  - ITU G.726
  - ITU G.728
  - ITU G.729

ITU – G.7xx

- ITU G.711
  - Designed for telephone bandwidth speech signal (3Khz)
  - Does direct sample-by-sample non-uniform quantization (PCM with u-law/A-law encoding scheme.
  - Encoder creates a 64 kbps bitstream for a signal sampled at 8 kHz.
  - Provides the lowest delay possible (1 sample) and the lowest complexity.
  - High-rate and no recovery mechanism, used as the default coder for ISDN video telephony
- ITU G.722
  - Designed for 7-Khz bandwidth voice or music
  - Operating at 48, 56 and 64 kbps for sample audio data at a rate of 16 kHz.
  - Divides signal in two bands (high-pass and low-pass), which are then encoded with different modalities. Two sub-band ADPCM.
  - G.722 is preferred over G.711 PCM for teleconference-type applications. Music quality is not perfectly transparent.

ITU – G.7xx

- ITU G.726
  - ADPCM speech codec standard covering the transmission of voice at rates of 16, 24, 32, and 40 kbps for a signal sampled at 8 kHz.

    c.f ADPCM implementation on TI DSP
- ITU G.728
  - Speech coding operating at 16 kbps for low-bit rate (64-128 Kb/s) ISDN video telephony
  - Hybrid between the lower bit-rate model-based coders (G.729) and ADPCM coders
- ITU G.729
  - Coding of speech at 8 kbps using model-based coders that use special models of production(synthesis) of speech