Lecture Notes for EE290T1

Avideh Zakhor

11/20/99

Audio Compression

Audio	vs	Speech
Noise generated by anything.		Limited to 4KHz. Max.

TABLE 8.1 Common Bandwidths and Sampling Rates for Speech and Audio

Input	Frequency Range (Hz)	Sampling Rate (1000 samples/second)
Telephone speech	200–3400	8
Wideband speech	50–7000	16
Wideband audio	20–20,000	44.1 or 48

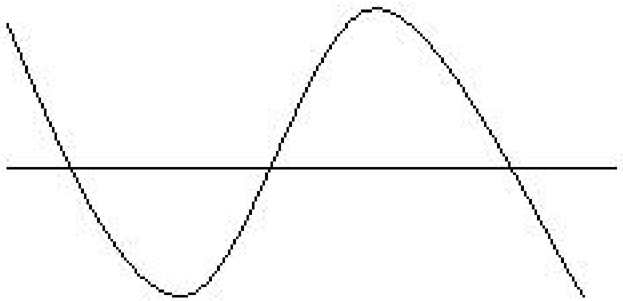
TABLE 8.2 Comparison of Commercially Available Audio Coding Systems

	Bit Rate	Quality	Complexity	Main Applications	Available Since
MPEG-1 Layer 1	32–448 kbps total	transparent @ 192 kbps/ channel, as per (ISO 1991c)	low encoder/ decoder	DCC	1991
MPEG-1 Layer 2	32–384 kbps total	transparent @ 128 kbps/ channel, as per (ITU 1994)	low decoder	DAB, CD-I DVD	1991
MP3 ← MPEG-1 Layer 3	32–320 kbps total	transparent @ 96 kbps/ channel, as per (ITU 1994)	low decoder	ISDN, satellite radio systems, Internet audio	1993
Dolby AC-2	128–192 kbps/ channel	transparent @ 128 kbps/ channel, as per (ITU 1994)	low encoder/ decoder	point to point, cable	1989
Dolby AC-3	32–640 kbps	transparent @ 384 kbps/5.1 channel, as per (ITU 1995)	low decoder	point to multipoint, HDTV, cable, SD-DVD	1991
Sony ATRAC	140 kbps/ channel		low encoder/ decoder	MD	1992
AT&T PAC			low decoder		
MPEG-AAC (NBC)	64 kbps/ channel	transparent			1997

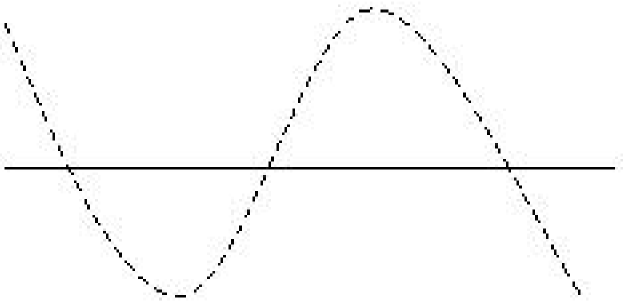
Adapted from Brandenburg and Bosi (1995).

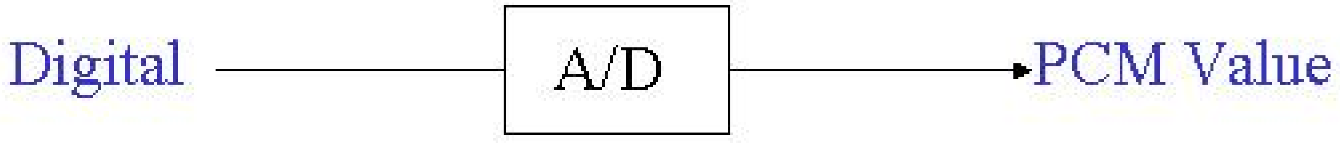
Audio Compression

PCM Analog



Digitize





Uniform Quantization $\longrightarrow$ Each extra bit gives an additional 6dB in SNR

A peak into the dB scale:

• 0dB	weakest audible sound pressure level
• 25dB	minimum noise level in typical recording studio
• 35dB	noise level in quiet home
• 120dB	loudest noise level before discomfort

Data Rates: 44.1 kHz sampling rate and 16bits/sample
sterio ==> 2 channels ==> 1.4Mbits/s

Log-PCM:

Quantization

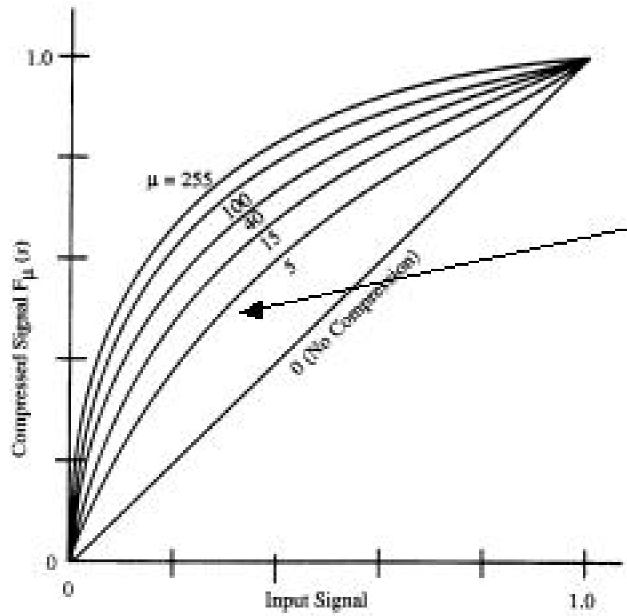


FIGURE 4.4 μ-Law Companding Characteristic

Boost low amplitude signal

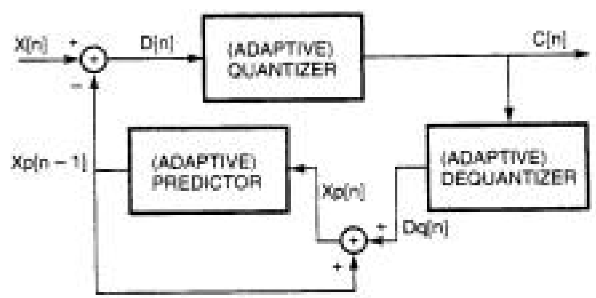
$$F\mu(s) = \frac{\ln(1+\mu|s|)}{\ln(1+\mu)}\,\mathrm{sgn}(s)$$

: rescale to give priority to low amplitudes

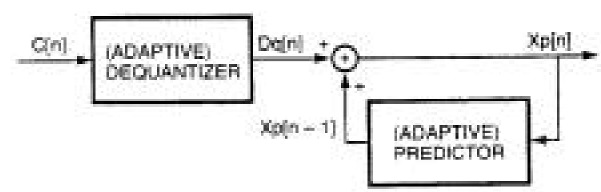
Uniform Quantizer: Output SNR decreases linearly with decreasing input signal power. Use log-PCM to compensate (scheme invented by telephone companies).

Adaptive Differential PCM:

- Adaptive version of DPCM
- Make the quantization or prediction adaptive



(a) ADPCM Encoder



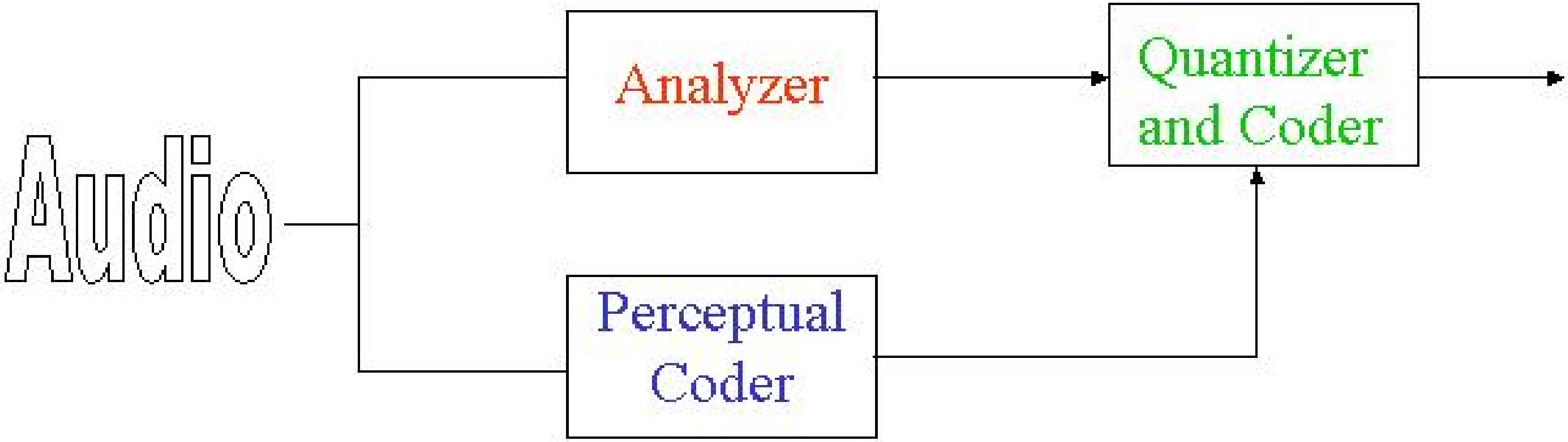
(b) ADPCM Decoder

MPEG Audio Compression

Basic Idea: Masking and Perceptual Coding

"Threshold in quiet" - - → Threshold level above which sound will be audible in a quiet environment. This threshold is a function of frequency.

Masking Threshold: Sound pressure level below which one signal (maskee) will not be heard, given the presence of another (masker).



The Perceptual Coder is used by the Quantizer/Coder to remove data that will be masked.

MPEG 1: (layer 1) 32 equally spaced subbands of 750Hz each. Each has 511 taps.

(layer 2) 1024-FFT instead of 512-FFT

(layer 3) MP3 - after the analysis FB, MDCT further subdivides each subband ==> bit allocation is better

MPEG2: some additional functionalities ==> designed for theatre systems.