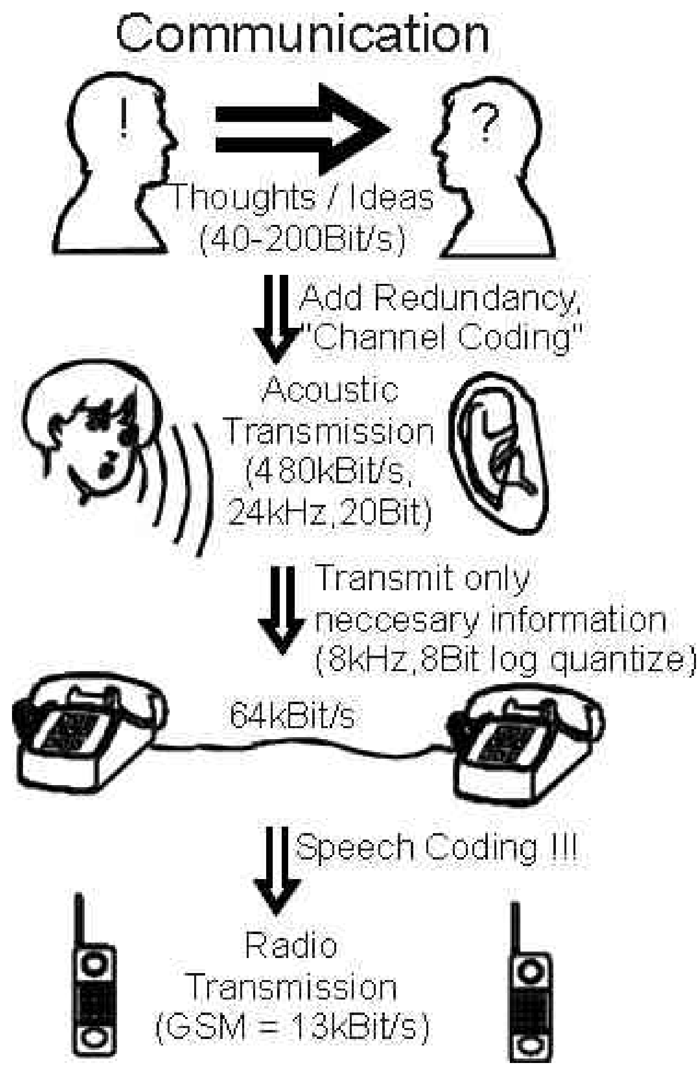
Introduction to Speech Coding



Nimrod Peleg
Update: Oct. 2009

Goals and Tradeoffs

Reduce bitrate while preserving needed quality

Tradeoffs:

- Quality (*Broadcast*, *Toll*, *Communication*, *Synthetic*)
- Bit Rate
- Complexity
- Robustness
- Delay

Speech Coding Types

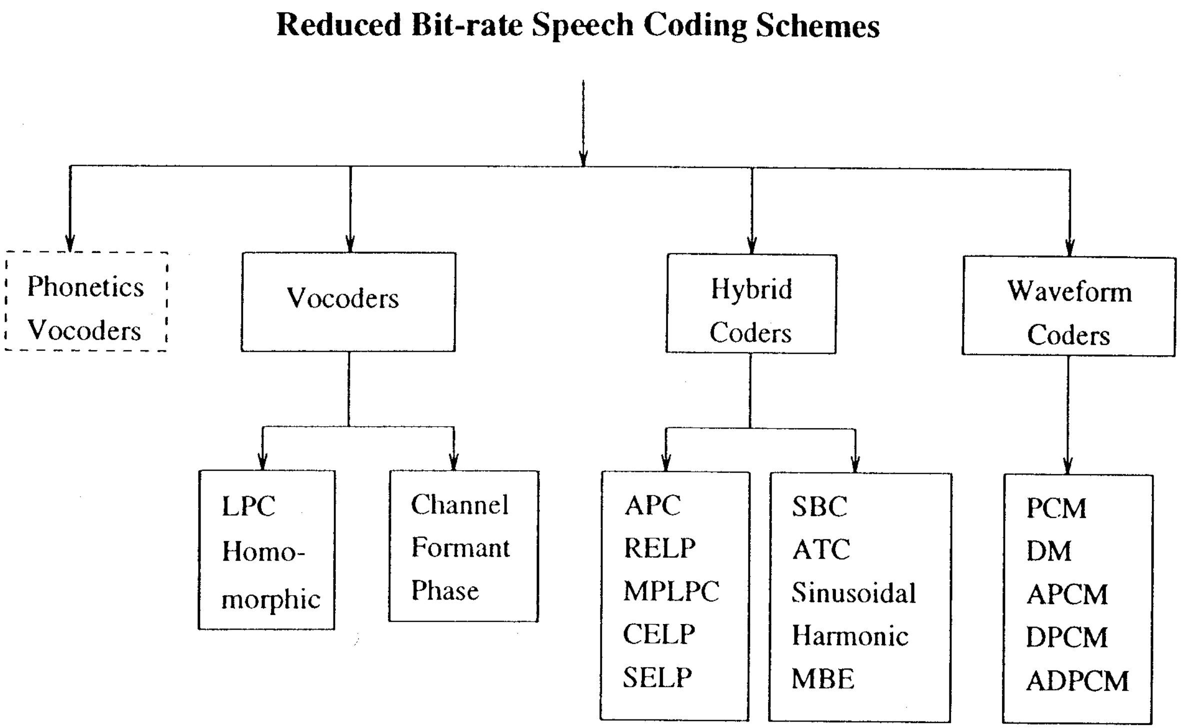


Figure 5.1 Classification of speech coding schemes

Digital Speech Coding Standards

The main schemes analyze the signal, remove
The redundancies and efficiently code the
non-redundant parts.

"Classical" telephony speech codecs:

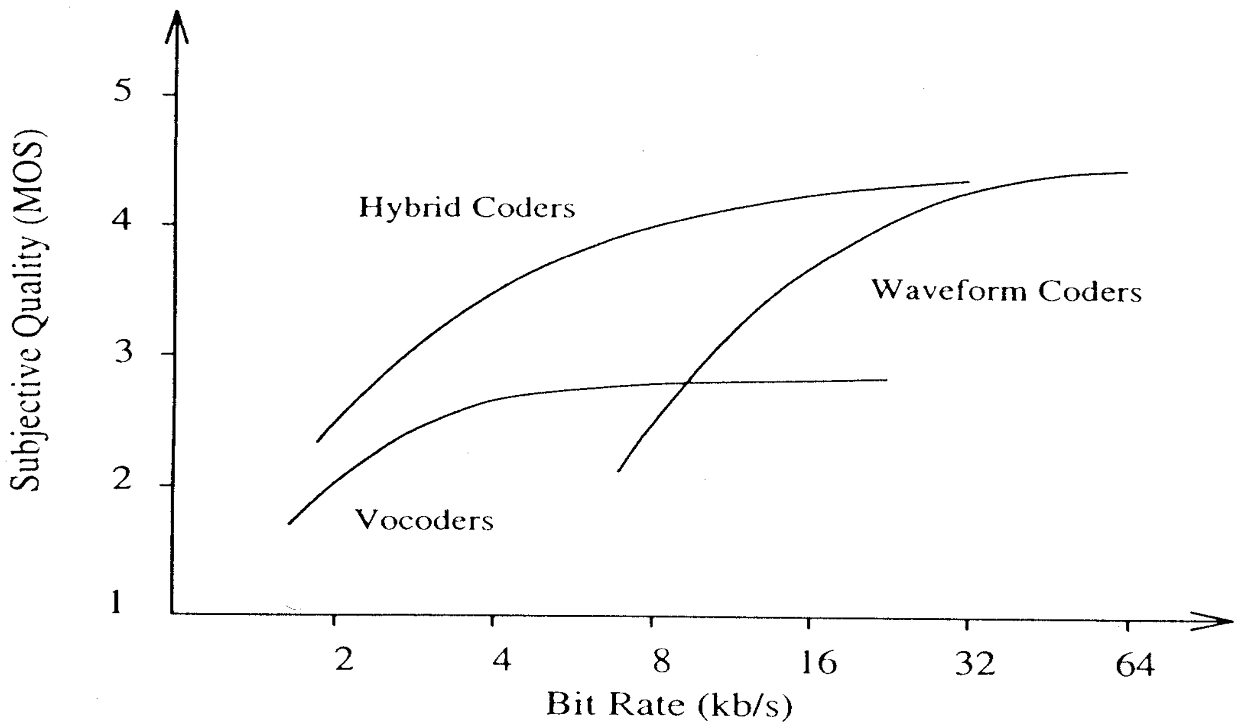
Coder	Application	Bitrate (Kbps)	Year
PCM	PSTN (1st Gen.)	64	1972
ADPCM	PSTN (2nd Gen.)	32	1984

Modern Speech Coding Standards

Coder	Application	Bitrate(Kbps)	Year
LD-CELP	PSTN	16	1992
APC	INMARSAT	16	1985
RPE-LTP	GSM	13	1991
VSELP	North Am. DMR	8	1992
MELP	Communication (US)	2.4/1.2	1997/2000
ACELP	Video Conf., Internet	5.3-6.3	1995
CELP	US Federal	4.8	1991
LPC-10	US Federal	2.4	1977

Quality Comparison

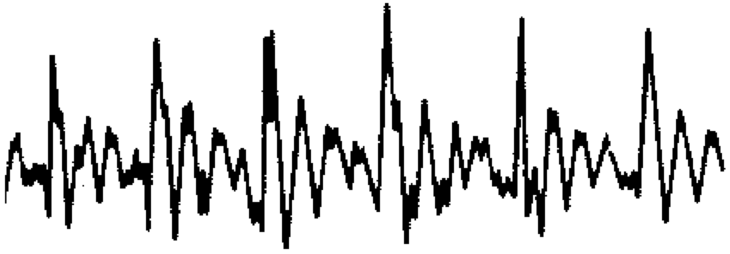
Mean Opinion Score (**MOS**):
Bad (1), Poor (2), Fair (3), Good (4), Excellent (5)



Quality comparison of speech coding schemes

Waveform Coding

- Preserves the general shape of the signal, and contains very little speech specific information
- Operates on a sample-by-sample basis
- Simple to implement in h/w and s/w
- Low delay
- Basic log-PCM 64Kbps is quality reference
- Very popular (many standards) and remains so due to replacements costs...



Waveform coding: PCM

Uses uniform Quantization

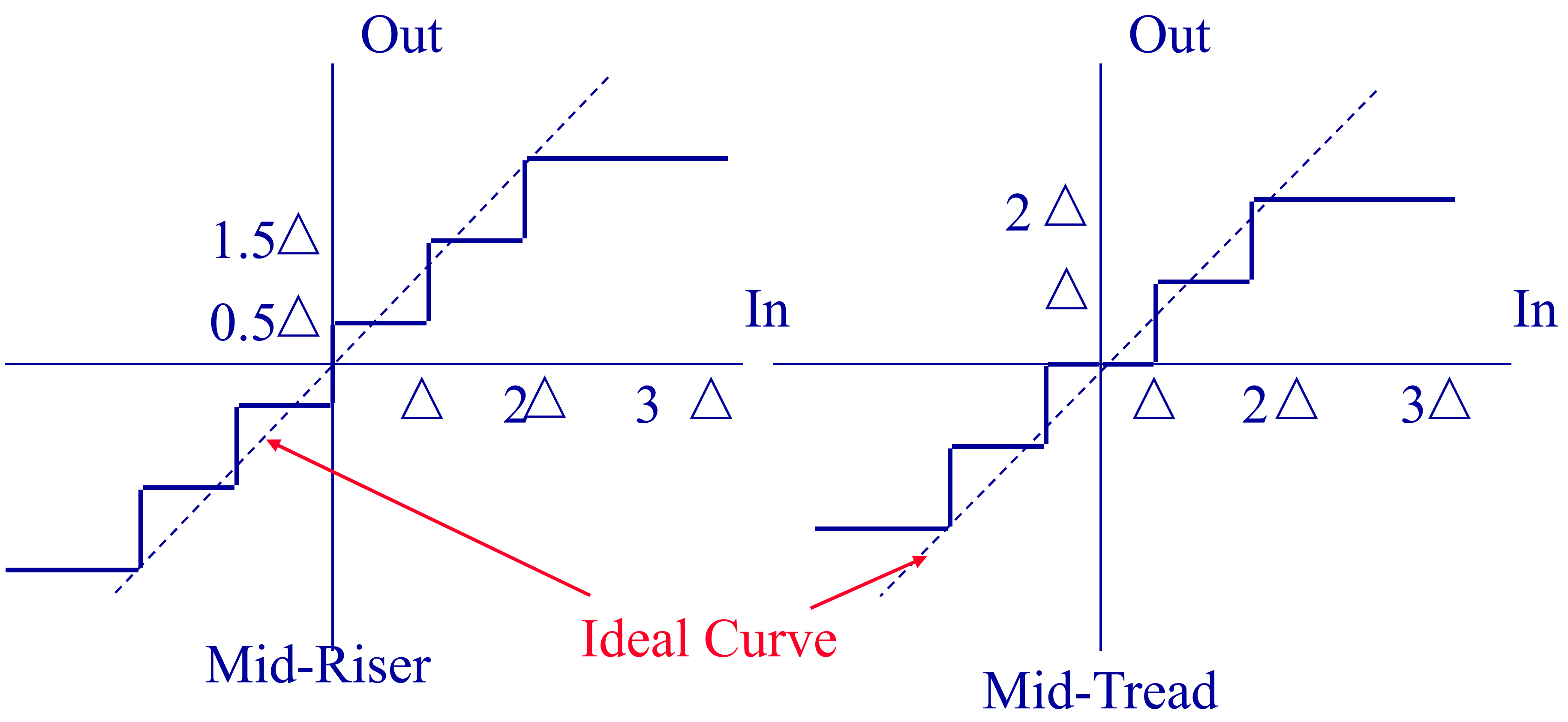
- Memoryless process, quantizes amplitudes by rounding each sample to a set of discrete values
- **No signal redundancy taken into account**

$$SNR = 10\log\frac{\sigma_x^2}{\sigma_q^2} = 6.02B - K \quad \text{(dB)}$$

B: Quantization Resolution

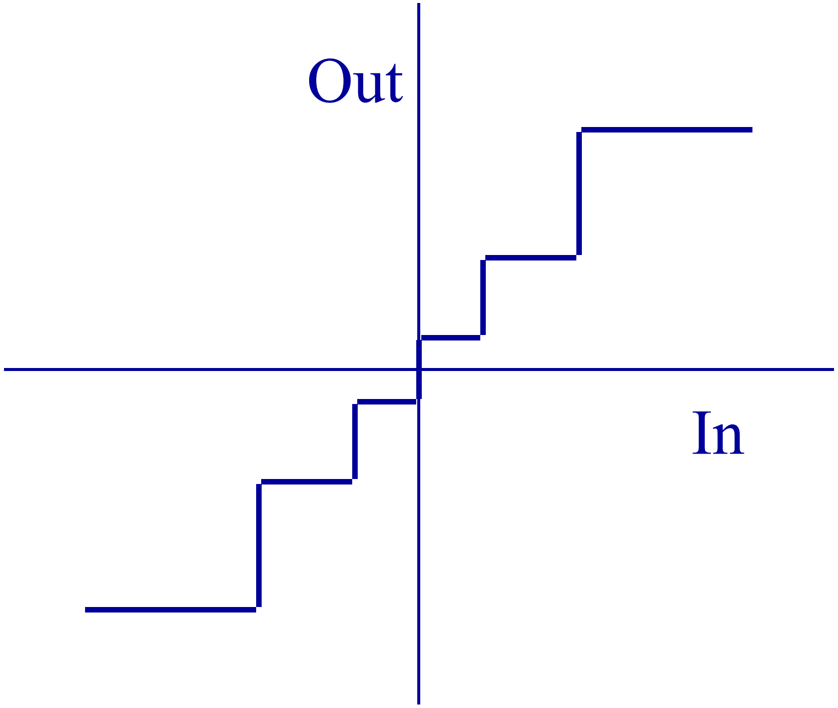
K is step size dependent (~ 5 - 8 dB)

Reminder: Uniform Quantizers



Two Common Uniform Quantizers

Reminder: Non-uniform Quantizer



If the distribution of input signal is not uniform - there is no reason why the quantizer should - so , non-uniform Quantizers are used

Optimal Quantizer design: by Max - Lloyd techniques, widely used for speech coding

In Matlab, Communications Toolbox

- **lloyds** :
- Optimize quantization parameters using the Lloyd algorithm
- Syntax
  - [partition,codebook] = lloyds(training_set,initcodebook)
- Description:

[partition,codebook] = lloyds(training_set,initcodebook)

optimizes the scalar quantization parameters partition and codebook for the training data in the vector training_set. initcodebook, a vector of length at least 2, is the initial guess of the codebook values. The output codebook is a vector of the same length as initcodebook. The output partition is a vector whose length is one less than the length of codebook.

[1] Lloyd, S. P., "Least Squares Quantization in PCM," IEEE Transactions on Information Theory, Vol IT-28, March, 1982, pp. 129-137.

[2] Max, J., "Quantizing for Minimum Distortion," IRE Transactions on Information Theory, Vol. IT-6, March, 1960, pp. 7-12

Logarithmic Quantizers

- Low complexity alternative to achieve good performance for signal with wide dynamic range
- Consists of: logarithmic transformation, than uniform quantization and expander that reconstructs the original signal dynamic range by inverse mapping function
- Two companding schemes exists: A-low,μ-low: for both, quantization noise close to uniform quantizer, but do not change much with changing signal variance (7bit~12bit uniform)

Logarithmic Quantizer (Cont'd)

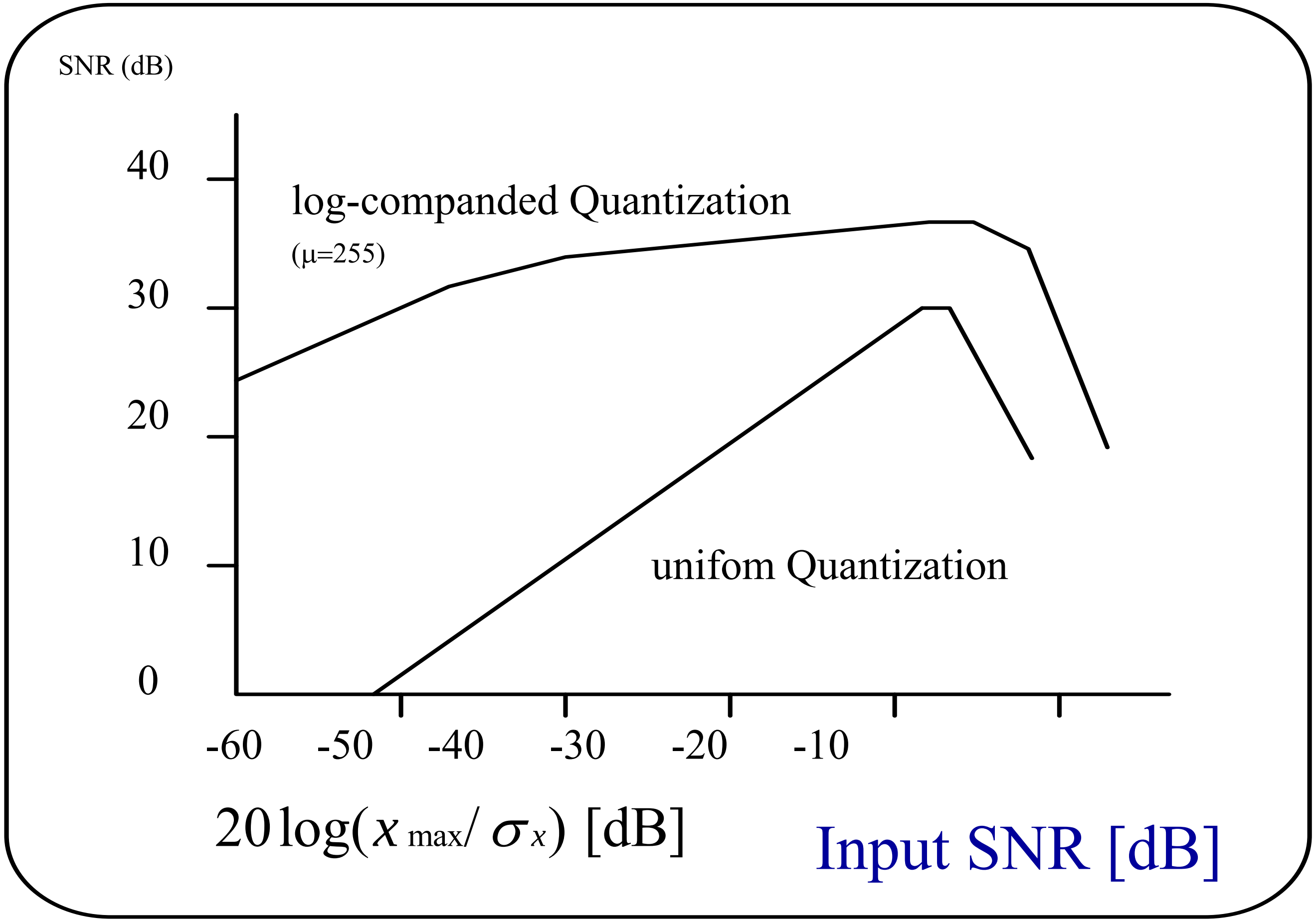
- SNR ~ $3L^2 / k*X^2max$

L: Number of levels k:Const.

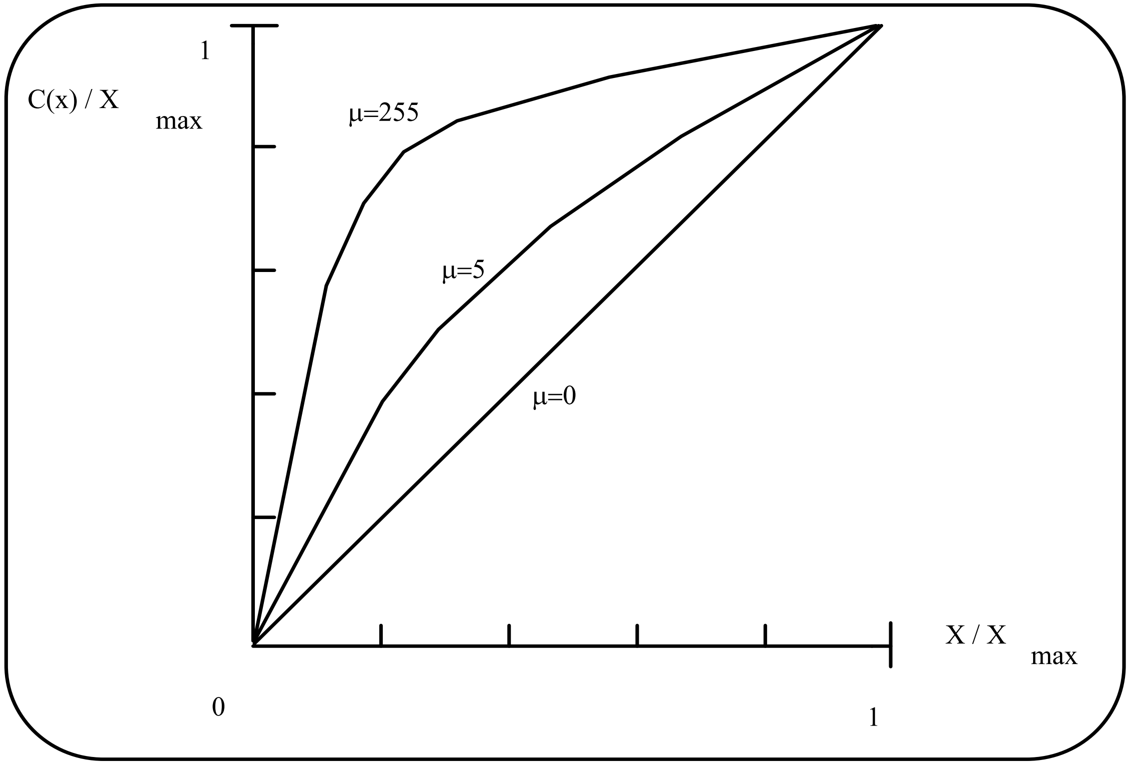
PCM Vs. µ-low Quantization

SNR Vs. Variance in 8-bit uniform and logarithmic quantizer

Output SNR [dB]



Input SNR [dB]



μ-low
“compander”

$$C(x) = x_{\max} \frac{\ln(1+\mu(x(/x\max}{\ln(1+\mu)} \operatorname{sgn}(x)$$

For small values: linear behavior

The keyword: Prediction

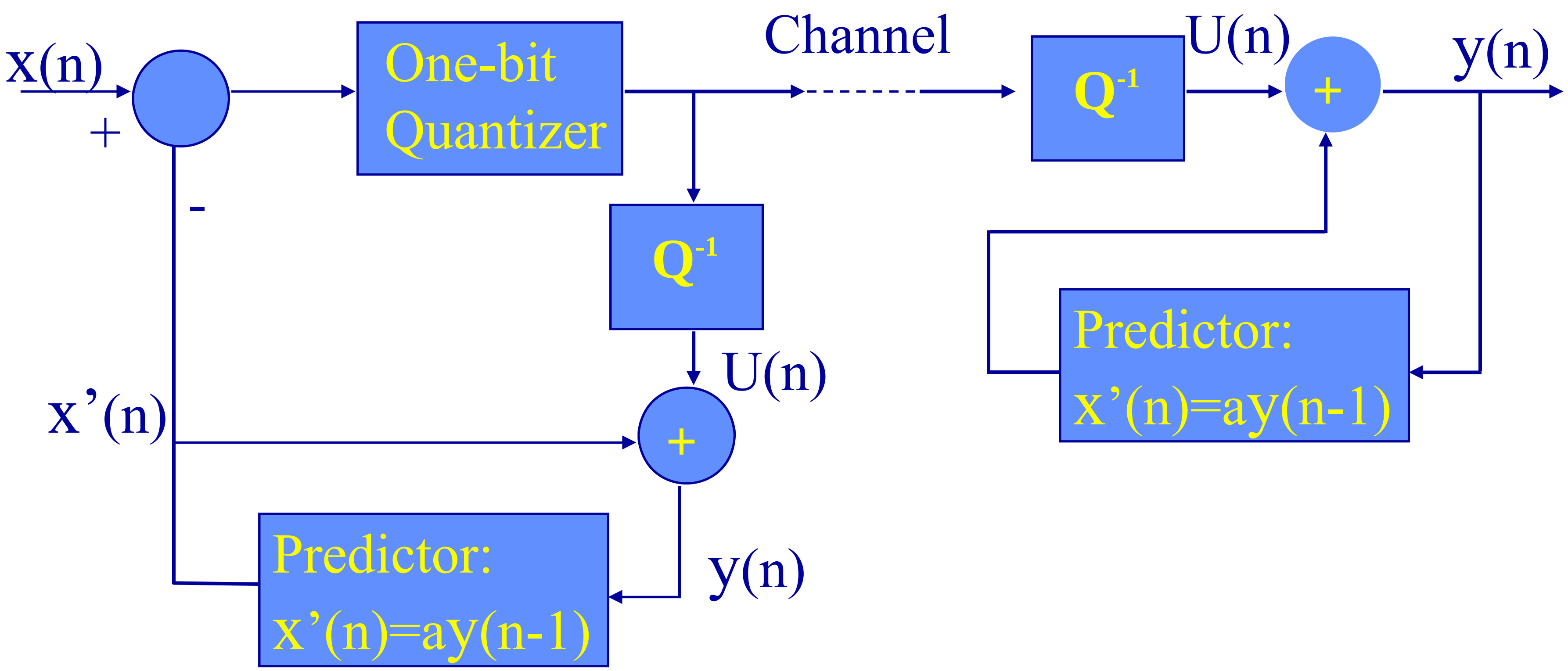
"קשה להתנבא, במיוחד לגבי העתיד"

אני נביא של מה שהיה / יהודה עמיחי

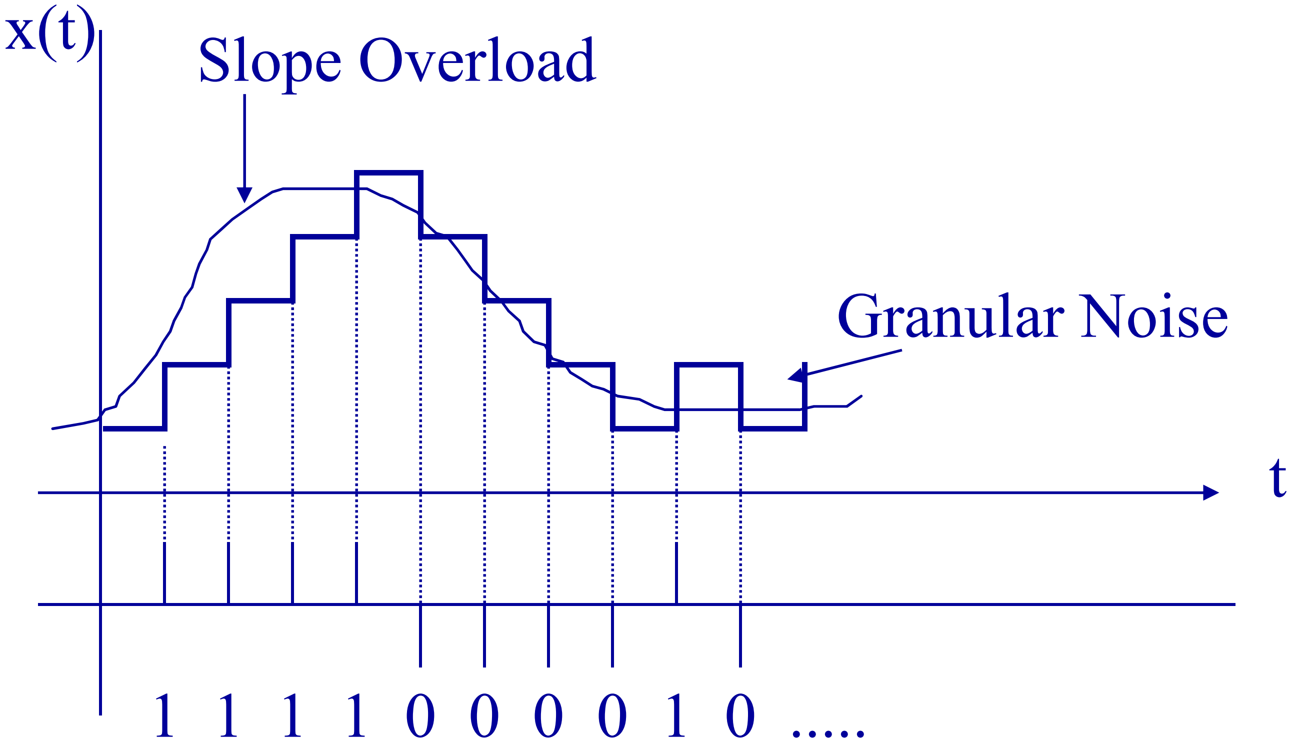
אֲנִי חוֹשֵׁב שֶׁהַחַיִּים הֵם חֲזָרוֹת לִקְרַאת
הַהַצָּגָה הָאֲמִתִּית. וּבַחֲזָרוֹת אֶפְשָׁר עוֹד לְהַכְנִיס
שִׁנּוּיִּים, לִמְחֹק מִשְׁפָּט וּלְהוֹסִיף דִּבּוּר, לְהַחֲלִיף
שַׂחֲקָנִים וּבַמָּאִים וְאוּלַמּוֹת עַד לַהַצָּגָה הָאֲמִתִּית
שֶׁבָּהּ שׁוּב אֵין מְשַׁנִּים, וְלֹא מְשַׁנֶּה שֶׁאֵין מְשַׁנֶּה
כִּי הַהַצָּגָה מוּרֶדֶת מִיָּד אַחַר הַפַּעַם הָרִאשׁוֹנָה.

Waveform coding: DM

- Delta Modulation:
  - Predictive coding that uses first order prediction and one-bit adaptive quantizer



Delta Modulation (Cont'd)



Delta Modulation (Cont'd)

- For reasonable speech quality, over-sampling at rates of 16-50KHz needed
- DM outperforms PCM at rates of 50Kbps
- Simple for implementation
- Many DM schemes developed, including second order predictors, and adaptive quantization step
- Used for high end audio : your CD works like this !

CVSD: Continuously Variable Slope Delta-modulation

- Based on adaptive delta modulation.
- uses a predictor and basically output a ‘1’ if the slope is increasing faster than the predictor and output a ‘0’ if it is decreasing slower than the predictor.

- The step size represented by the bit can vary and if the algorithm detects slope overload (i.e. the feedback signal is not keeping up with the input signal) the step size can be increased.

- In CVSD the step size is increased by looking at the last 3 or 4 samples and increasing the step size if they are all 1’s or 0’s.

CVSD Con't

Change Delta according to last 3-4 bits:

delta(n)=b*delta(n-1) + a(n)*delta0

$b=1-eps^2$ eps -->0

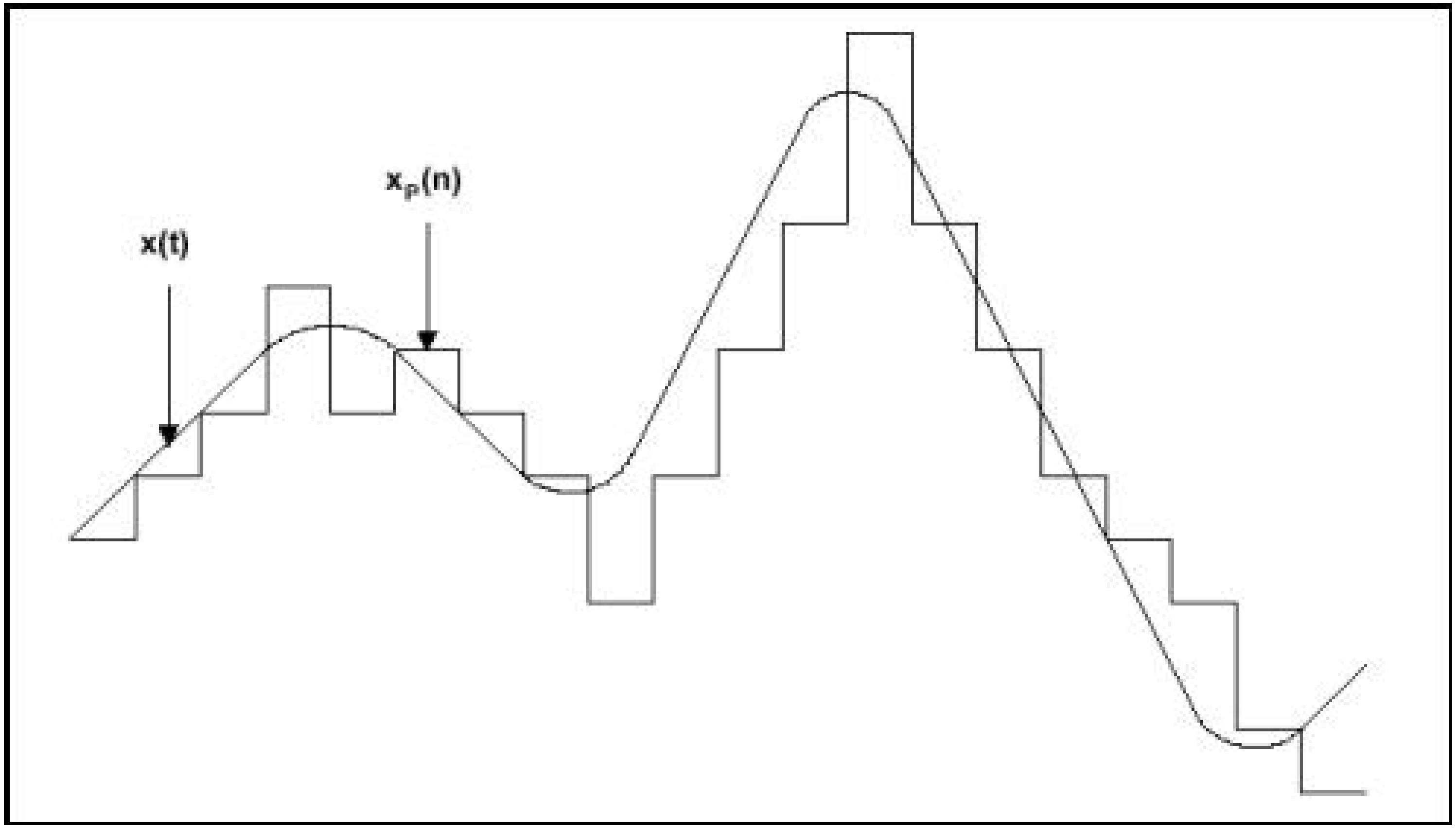
a(n) is '1' when last bits are '1,1,1' or '-1,-1,-1'

which indicates monotonous increasing/decreasing signal.

- If 'a' is 0 for a long time, delta becomes $delta_{min}$

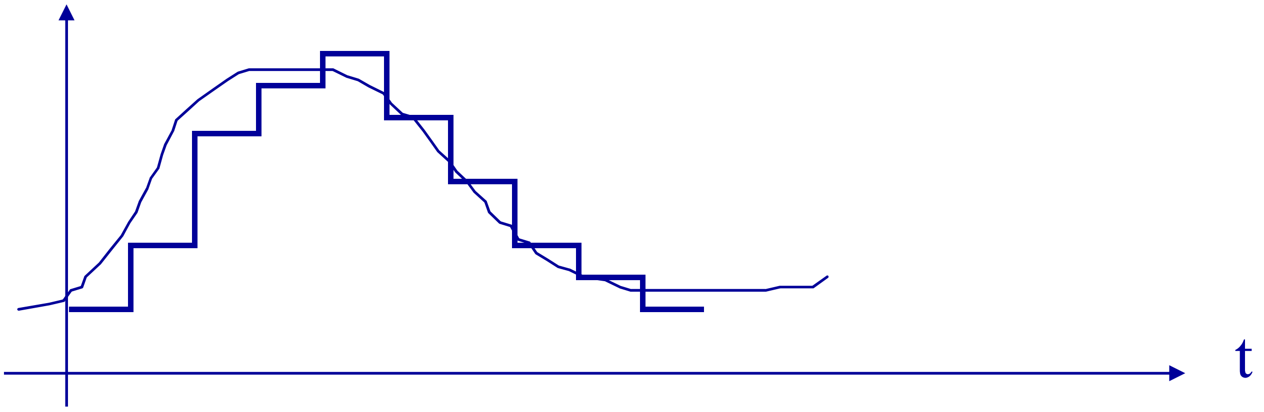
- Four-bit companding seems most effective for data rates greater than 32kbps while three-bit companding is more appropriate for data rates less than 32kbps.

CVSD Quantization



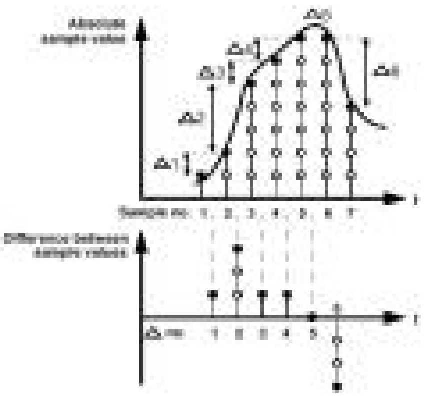
Adaptive PCM

- APCM: PCM with adaptive step size
- In Feed-Forward system, the step size is transmitted as side information
- In Feedback system, step size is estimated from past coded (and reconstructed) samples

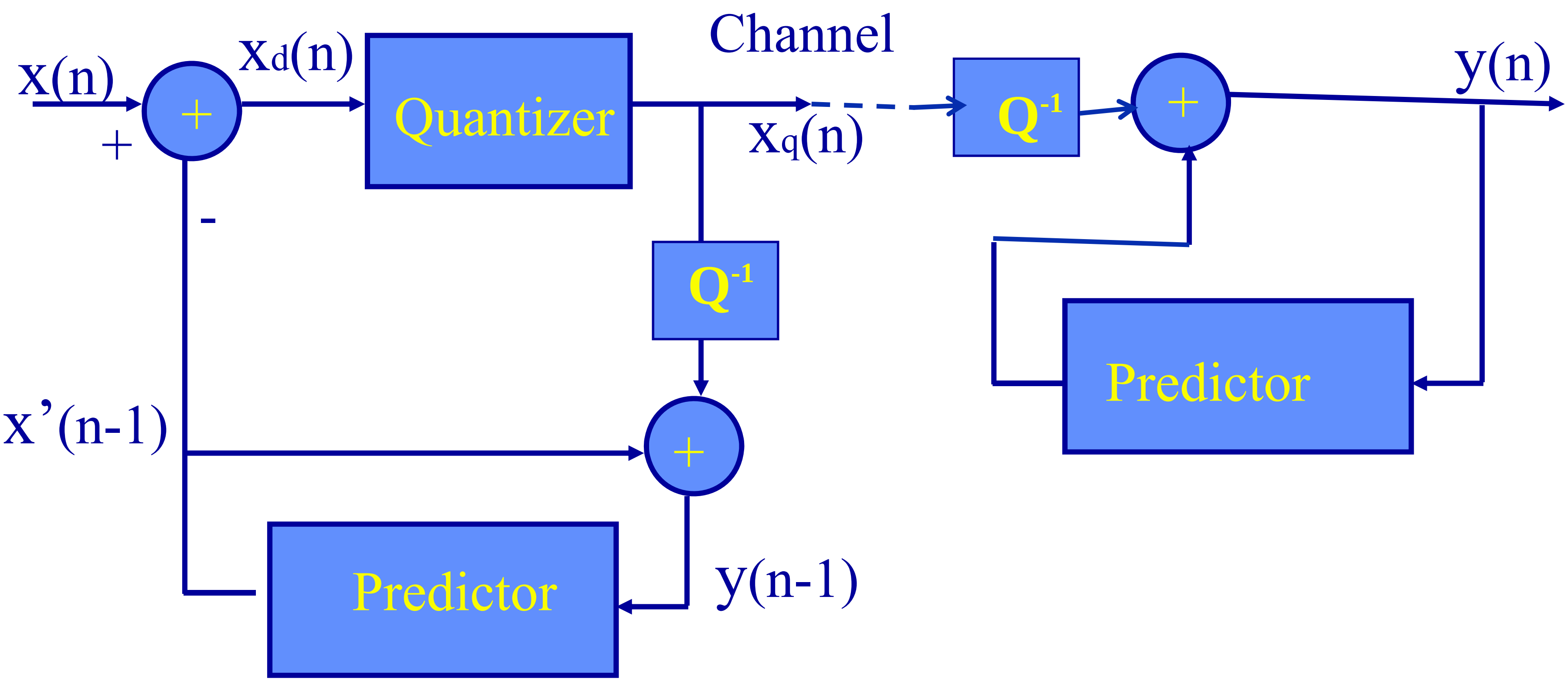


DPCM

- Predictive coding that uses short-term fixed predictor and fixed quantizer
- More efficient, utilizes redundancy by exploiting the correlation between adjacent samples.



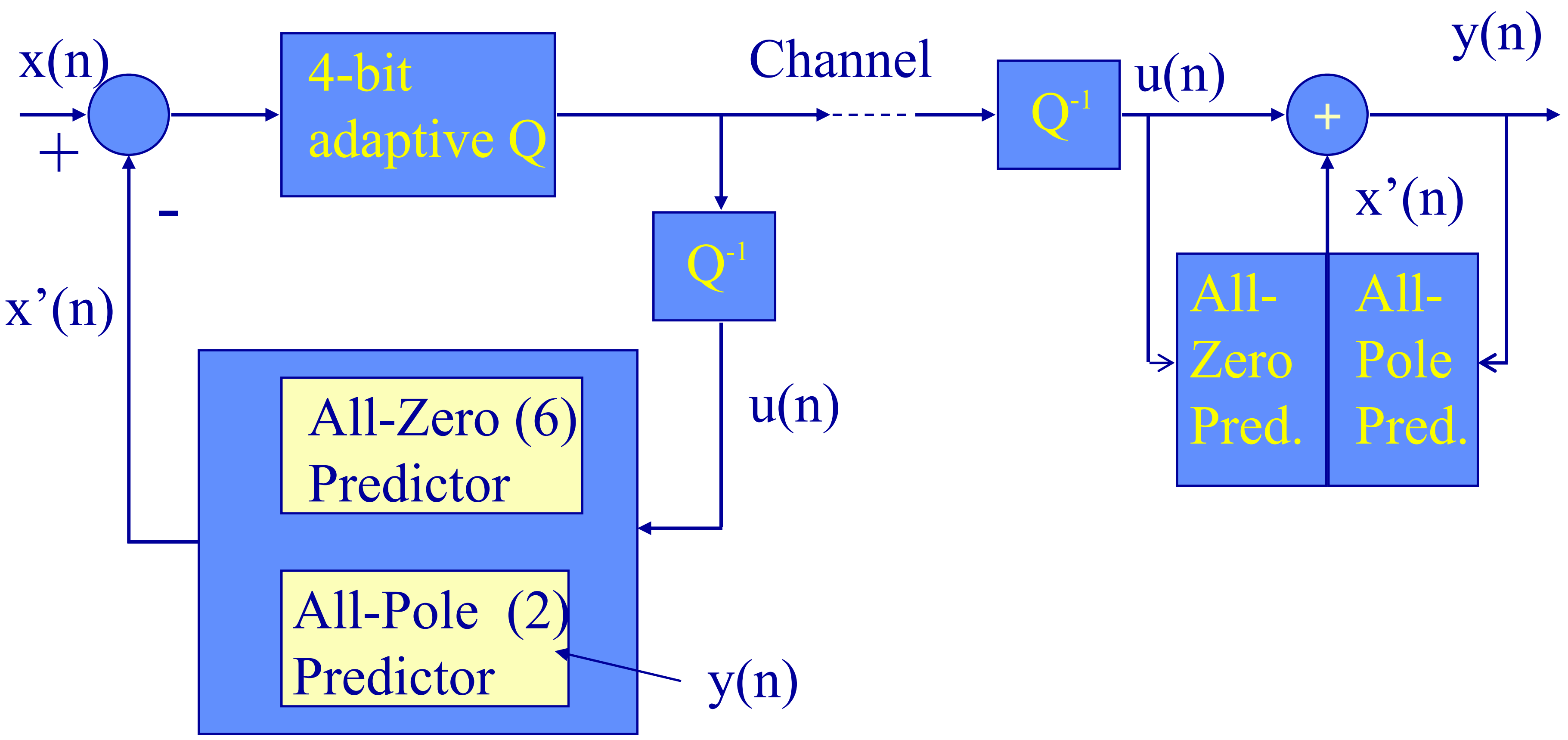
DPCM Scheme (Forward prediction)



Waveform coding: ADPCM

- Involving predictor adaptation, Quantizer adaptation into the DPCM scheme
- CCITT ADPCM at 32Kbps achieves quality close to PCM at 64Kbps (toll quality)
- The mentioned scheme uses short-term backward adaptive predictor and an adaptive 4-bit quantizer
- MOS: ~4.0 , Delay: one sample
- Quality degrades quickly: far from toll at 16Kbps

Simplified ADPCM Scheme



ADPCM Full System

- Use both:
  - predict the next sample based on the previous ones and code the error
  - adapt the quantizer
- 2:1 compression (4 bits per sample)

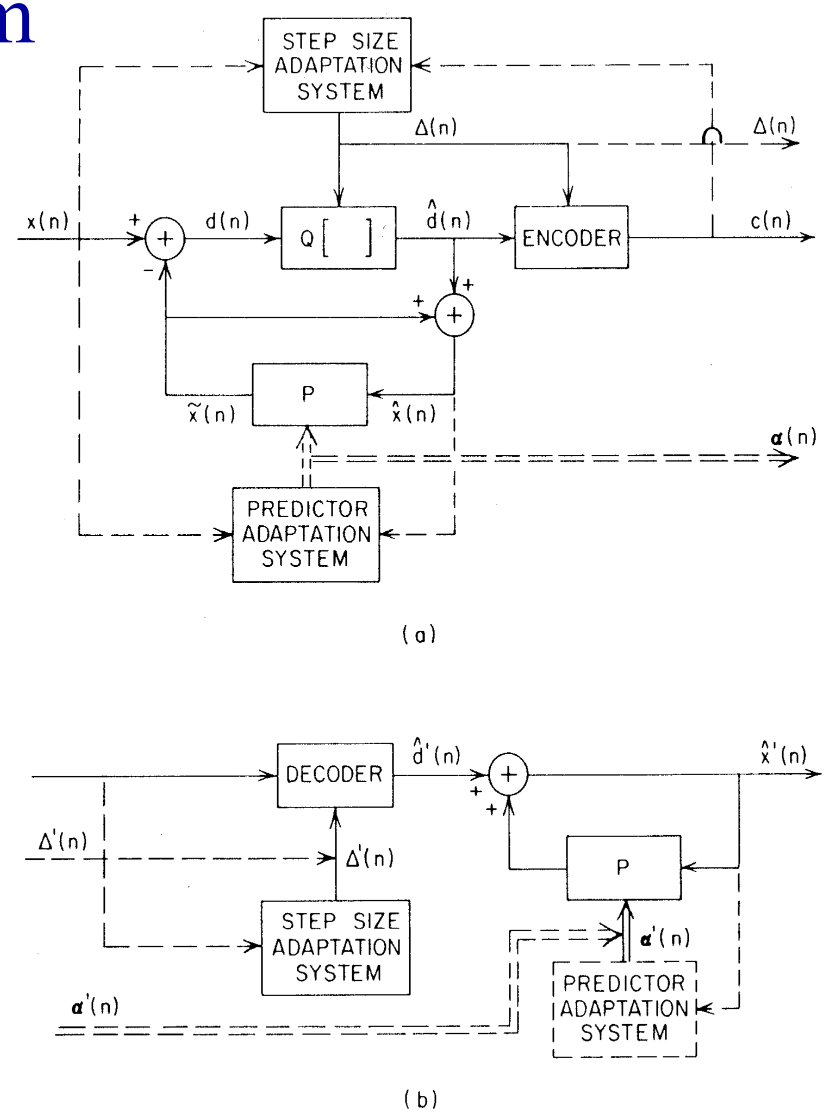
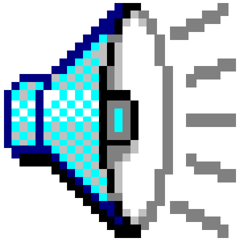


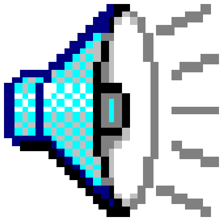
Fig. 5.41 ADPCM system with both adaptive quantization and adaptive prediction; (a) coder; (b) decoder.

PCM example



This is the original speech signal sampled at 8000 samples/second and u-law quantized at 8 bits/sample (64Kbps). Approximately 4 seconds of speech.

ADPCM example

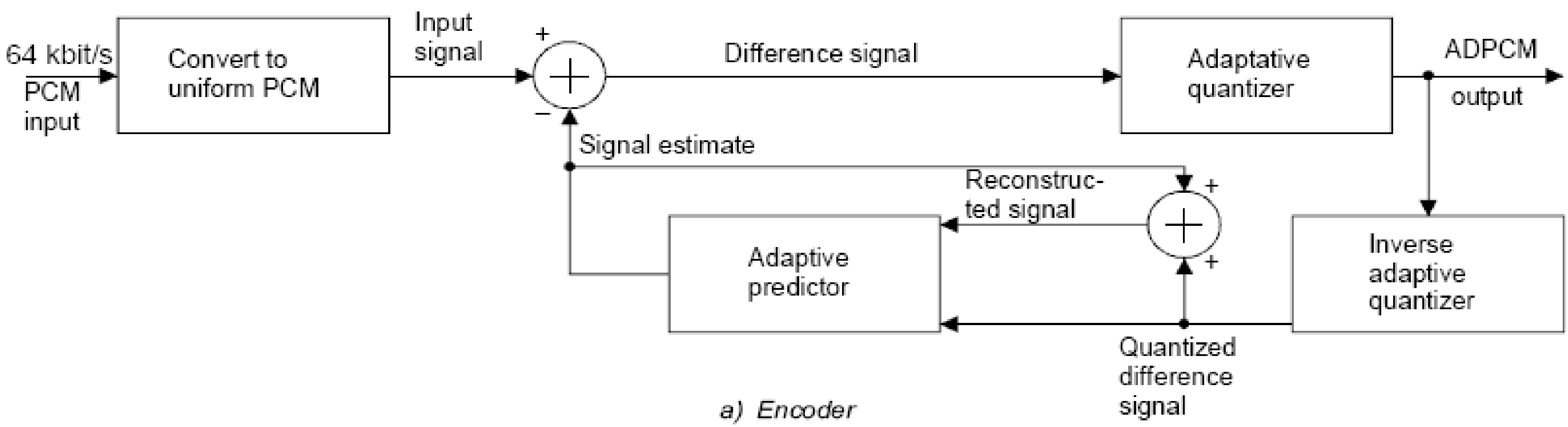


This is speech compressed using the Adaptive Differential Pulse Coded Modulation (ADPCM) scheme.
The bit rate is 4 bits/sample (compression ratio of 2:1): 32Kbps

ITU G.726 Recommendation

- G.726 is an ADPCM speech codec standard covering the transmission of voice at rates of 16, 24, 32, and 40 Kbit/s
- Often referred to by the bit size of a sample: 2-3-4-5 bits respectively
- Created in 1990.
- It is one of the standard codecs used in trunk phone systems.

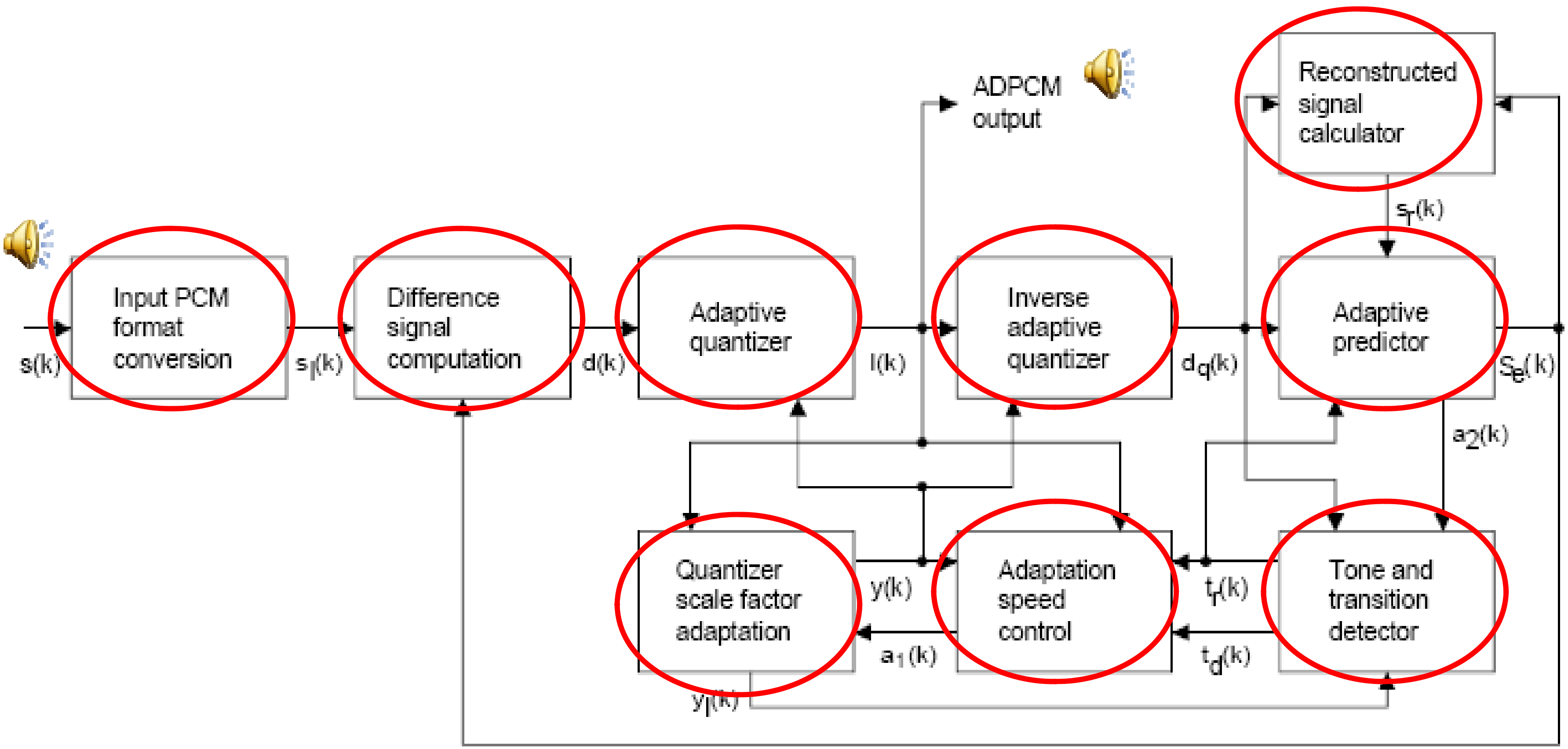
G.726 Encoder Signal Flow



Non-linear quantization on a difference signal

(between the original signal and it's estimation)

G.726 - Encoder block schematic



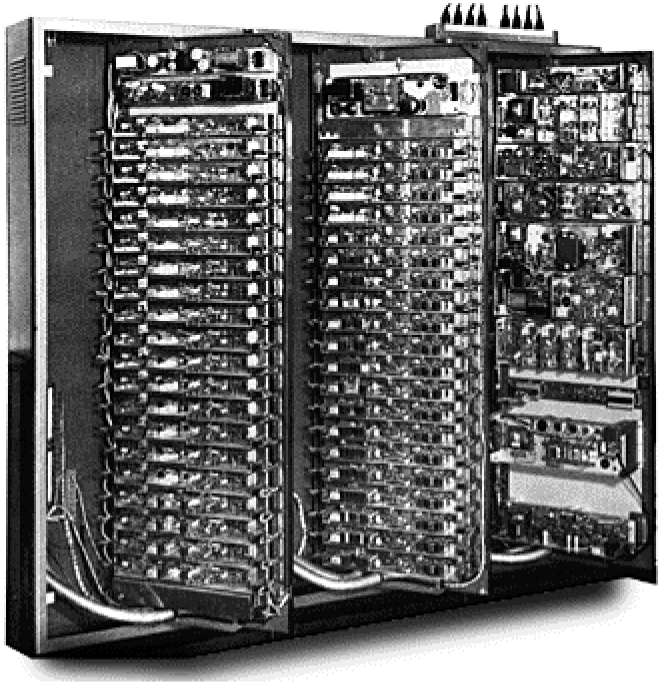
G.726 – Encoder Blocks

- **Input PCM format conversion** – Converts the input signal from A-law / μ-law to a uniform PCM signal.
- **Difference signal computation** – calculates the difference signal from the uniform PCM signal and the estimated signal.
- **Adaptive Quantizer** – quantizes the logarithmic representation of the difference signal (Non-uniform !)
- **Inverse Adaptive Quantizer** – A quantized version of the difference signal.

Encoder Blocks Cont’d

- **Quantizer Scale Factor Adaptation** – Computes the scaling factor for the quantizer according to the fluctuations (variance) of the difference signal.
- **Adaptation Speed Control** - A weight factor in the calculation of the scale factor.
- **Adaptive Predictor & Reconstructed Signal Calculator** – Computes the estimated signal from the quantizer difference signal.
- **Tone and Transition Detector** - Improves system’s performances for modem signals.

Vocoder:
Voice Coder

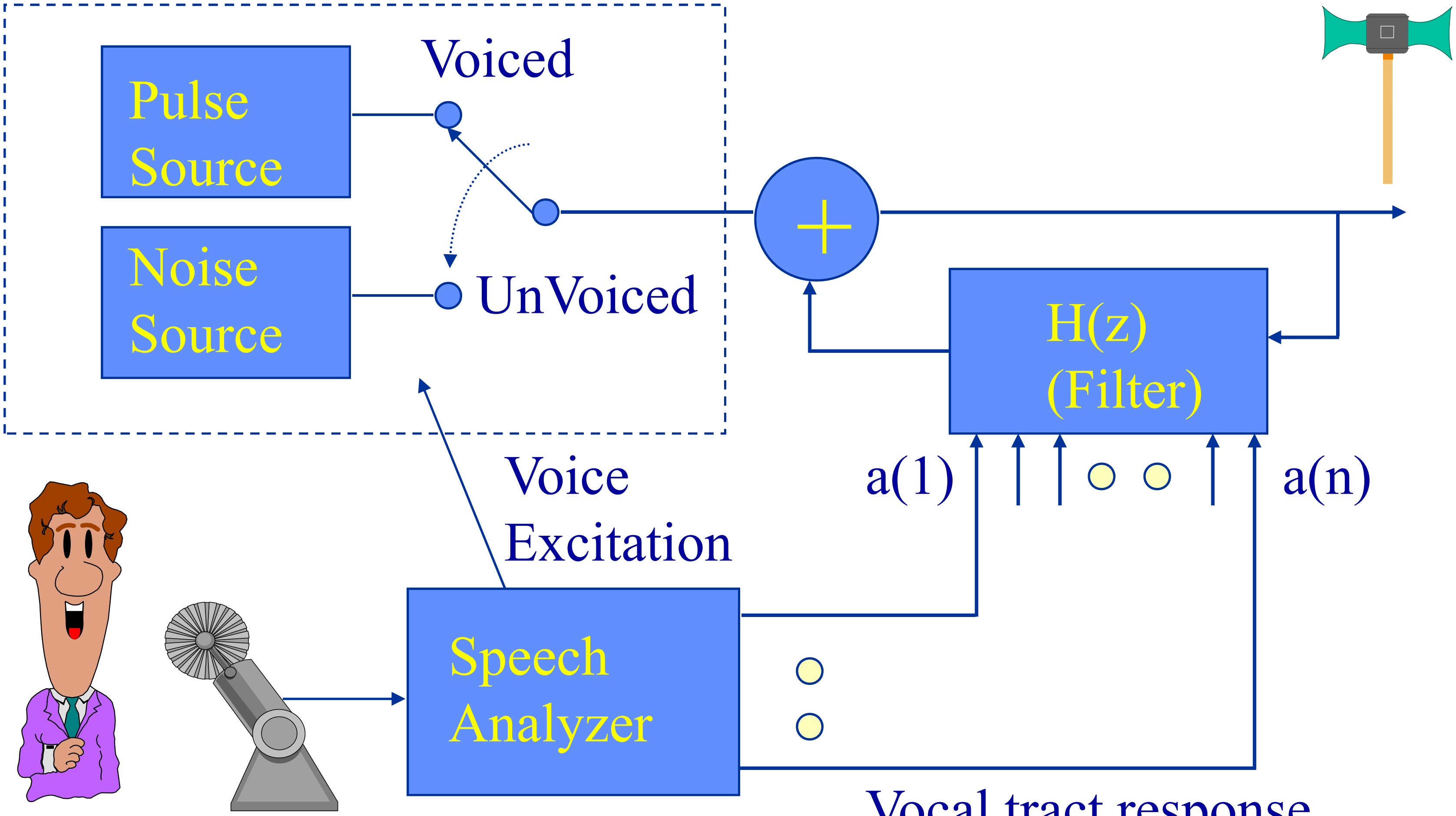


Siemens Synthesiser, 1959

VoCoders (Voice Coders)

- Very speech specific: the analyzer extracts a set of parameters representing speech production model
- BitRate: about 4.8Kbps, and very popular 2.4Kbps LPC-10
- SNR measure is useless
- Formant Vocoder: synthesize speech by a set of bandpass filters
- Both used also for speech synthesizers

LPC VoCoder

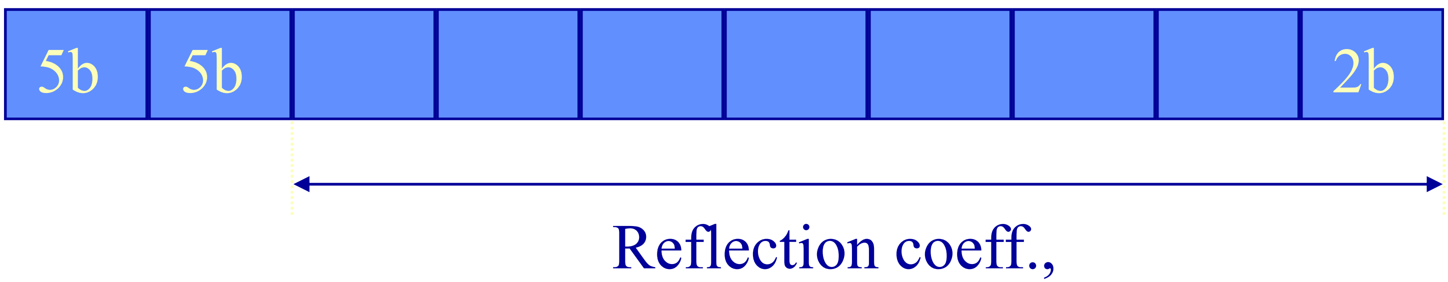

Pulse
Source
Noise
Source
Voiced
UnVoiced
+
H(z)
(Filter)
a(1)
a(n)
Voice
Excitation
Speech
Analyzer
Vocal tract response

LPC Vocoders features

- Employ source filter model
- For each frame needs to calculate: LPC, V/UV decision , power, and Pitch (if voiced)
- Most bits used for LP coefficients
- Commonly used representations:
  - LP: if there are no quantization problems
  - Reflection: robust but not efficient
  - Line Spectral Pairs: most efficient
- VQ saves about 50% of bits (~25 per frame)

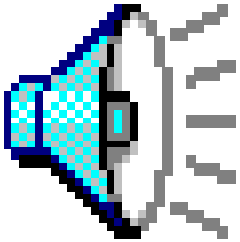
Example: LPC-10

- Sample rate: 8KHz, 180 sam./frame, 44.4 fps
- LP analysis of order 10



- Pitch and V/UV decision: 7b
- Gain: 5b
- Total: 54 bpf, 2400bps, VERY “buzzy”, very bad for background noise (mobile phone...)

LPC-10 Example



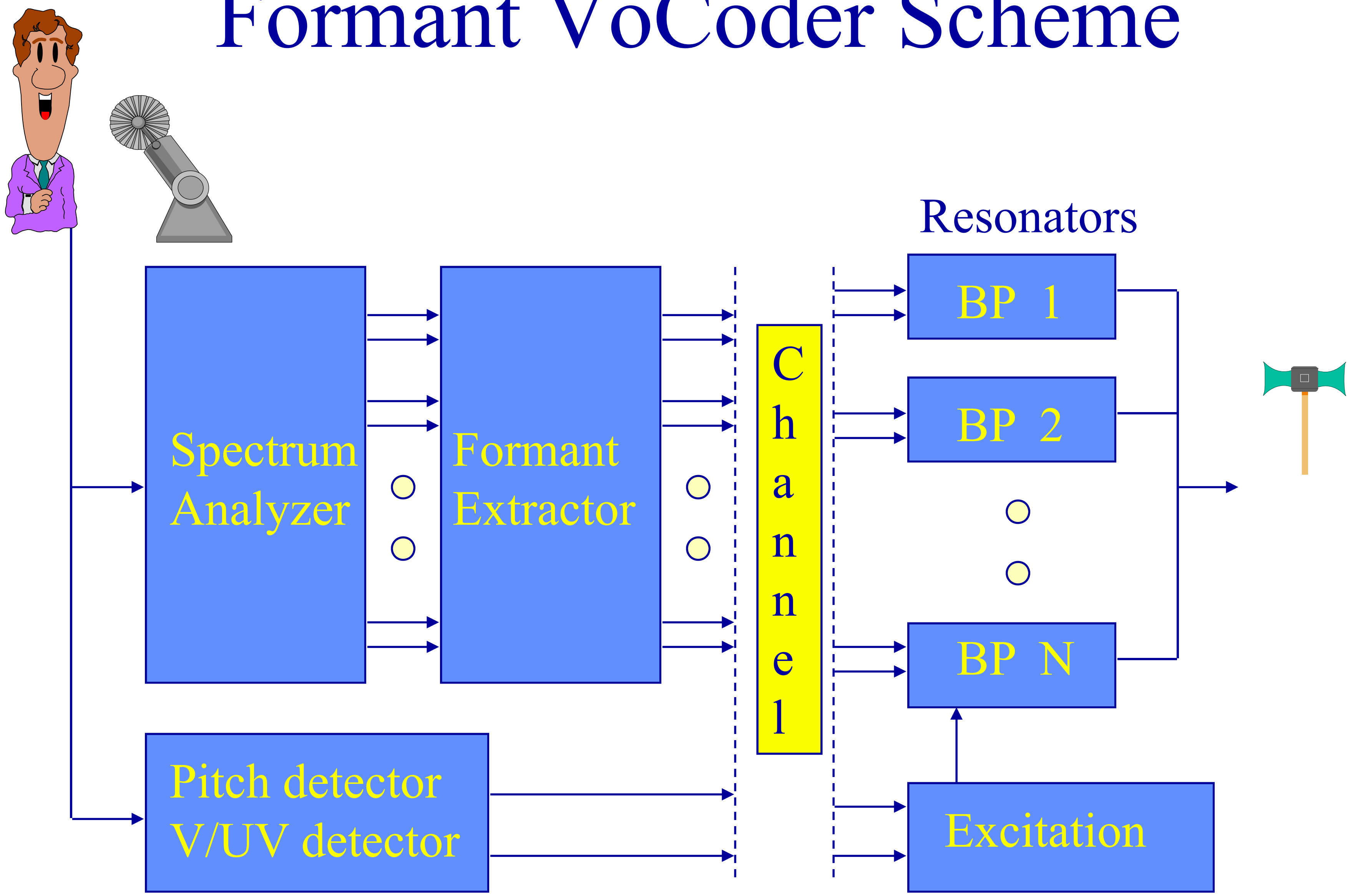
This is speech compressed using the Linear Predictive Coding (LPC10) scheme.

The bit rate is 0.3 bits/sample (compression ratio of 26.6:1): 2400bps

Formant VoCoder

- Formant parameters are transmitted
  - band-width and center frequency)
- Better compression ration
- Difficult to implement: performance is dictated by accurate formant location
- With some human intervention capable of producing very high quality speech.

Formant VoCoder Scheme



SPEECH CODERS

Year of Introduction	Bit Rates kbits/s	Description	MOS
1972	64	PCM (for PSTN)	4.4
1976	2.4	LPC-10 (Fed. Std. 1015)	2.7
1984	32	G.721 ADPCM (for PSTN)	4.1
1987	24	G.723 ADPCM	4.0
1990	16	G.726 ADPCM	3.9
1990	4.15	Inmarsat (Satellite)	≈3.2
1991	13	GSM (European Cellular)	3.6
1991	4.8	CELP (Fed. Std. 1016)	3.2
1992	16	G.728 (Low delay-CELP)	4.0
1992	8	VSELP (US Cellular)	3.5
1993	1-8	QSELP (US CDMA)	≈3.4
1993	6.8	VSELP (Japanese Cellular)	≈3.3
1995	8	G.729 (new toll-quality)	≈4.2
1995	6.3	G.723.1 (in H.323 & H.324)	3.98
1995	5-6	Half-Rate GSM	≈3.4
1996	2.4	New low rate Fed. Std.	≈3.3

Table 5.2 Digital speech coding standards

Rate (kb/s)	Application	Type of Coder	Year of Operation
64	PSTN (1st Generation)	Pulse Code Modulation (PCM)	1972
32	PSTN (2nd Generation)	Adaptive Differential PCM (ADPCM)	1984
16	PSTN (3rd Generation)	Low Delay Code Excited Linear Predictive Coding (LD-CELP)	1992
16	INMARSAT Standard B (Maritime)	Adaptive Predictive Coding (APC)	1985
13	Pan-European Digital Mobile Radio (DMR) Cellular System (GSM)	Regular-Pulse Excitation Long-Term Prediction (RPE-LTP)	1991
9.6	Skyphone (Aeronautical)	Multi-Pulse Linear Predictive Coding (MPLPC)	1990
8	North American DMR (Mobile)	Vector Sum Excited Linear Predictive Coding (VSELP)	1992
6.7	Japanese DMR (Mobile)	VSELP	1993
5 to 6	Half-Rate GSM (Mobile)	CELP?	>1993
6.4	INMARSAT Standard M (Land-Mobile)	Multi-Band Excitation (MBE)	1993
4.8	U.S. Government Federal Standard	CELP	1991
4.8	NASA MSAT-X (Mobile Satellite)	Vector Adaptive Predictive Coding (VAPC)	1991
2.4	U.S. Government Federal Standard	Linear Predictive Coding (LPC-10)	1977

Various Standard Codecs

Modern Speech Codec: MELP