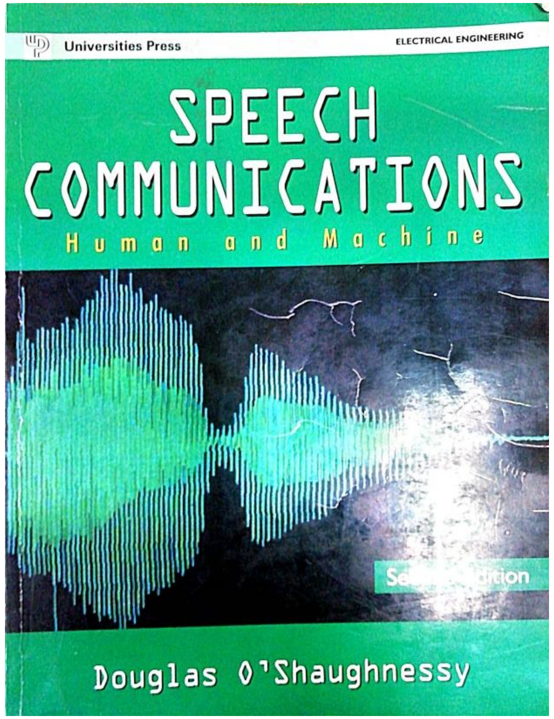
Speech Coding

Objective

- Efficient secure storage
- Efficient transmission
- Applications
  - Digital voice transmission over analog channels
  - Wireless transmission (Mobile communications)



Evaluation of Coders

- Bit rate
- Cost of transmission and storage
- Complexity and speed
- Output speech quality

Speech Quality

- Broadcast, 50 – 7000 Hz, 32-64 Kbps
- Toll/wireline, 200 – 3200 Hz, 8-64 Kbps, SNR > 30 dB
- Communications, high intelligibility and noticeable distortion (4 Kbps)
- Synthetic, substantial degradation and 90% intelligibility

Classes of Coders

- Waveform coder (sample by sample)
  - PCM
- Time domain waveform coder (periodicity and slowly varying intensity)
  - LP coders, DPCM, DM, ADPCM
- Spectral domain waveform coder (spectral properties)
- Source coder / Vocoder (speech production)

Distortion Measures

- Objective measures (SNR, SEGSNR)
- Subjective measures (MOS, DMOS)
  - Intelligibility
  - Naturalness
  - Speaker characteristics

Speech Redundancies

- Speech production
  - Changes in the VT shape and speech spectrum are relatively slow compared to sampling frequency
  - Glottal vibration is relatively slow
  - Periodicity (successive segments are nearly identical)
  - Most of the speech spectral energy is at low frequencies
  - Periodic vs noisy excitation

- Auditory characteristics
  - Phase information
  - Masking (weighting to different frequencies)

Time Domain Waveform Coding

- Slow variation of speech energy
- Periodicity
- Dominant low frequency energy

Adaptive PCM

variable step size and gain

size of the window

APCM

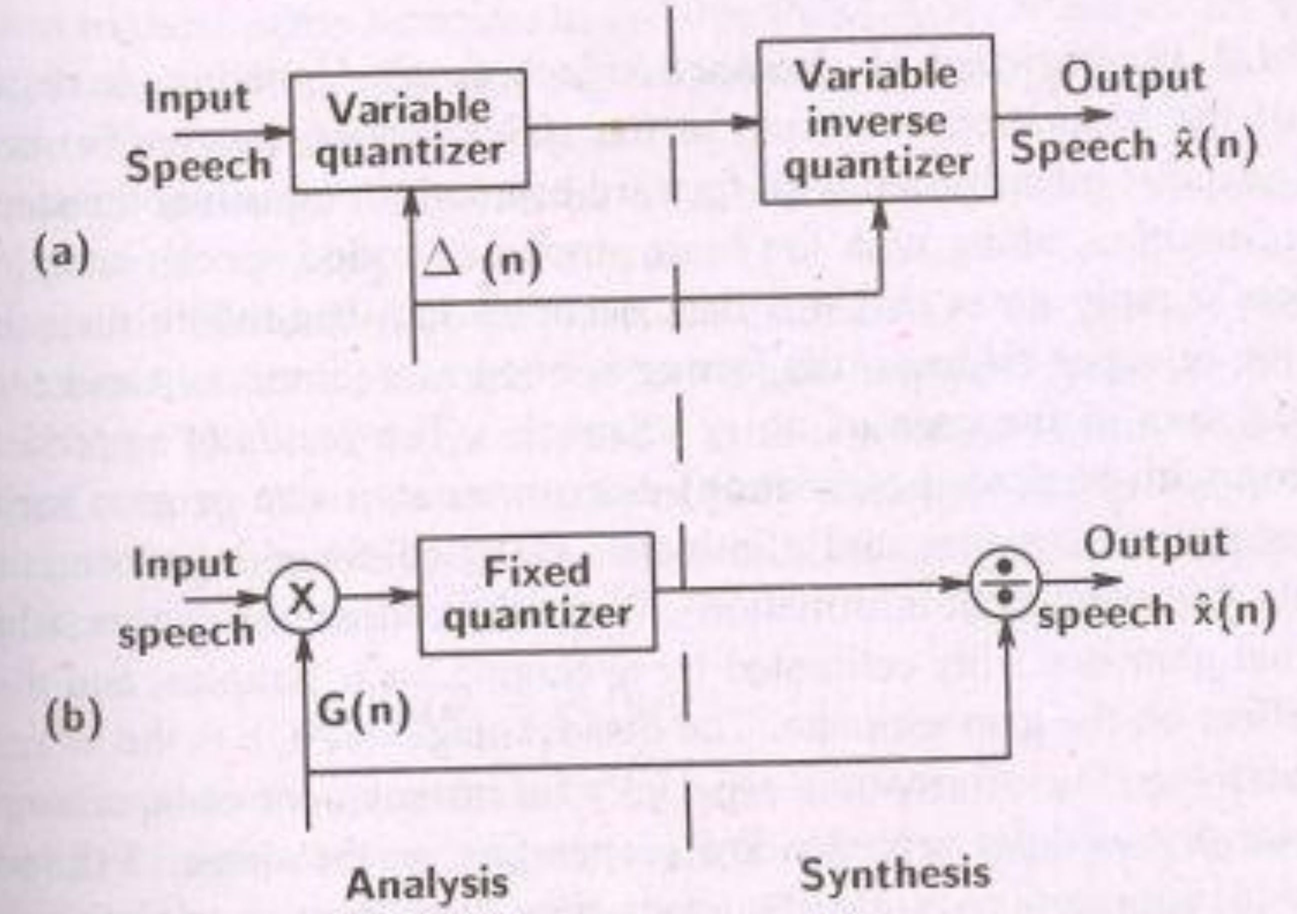


Figure 7.13 Block diagram representation of adaptive quantization: (a) variable–step size version, (b) variable–gain version.

Exploiting Properties of Spectral Envelope

- Differential PCM
  - Short term correlation (1-3 ms), spectral envelope
  - Long term correlation (pitch periodicity), harmonic structure

- Exploiting periodicity of voiced speech
  - Pitch adaptive predictive coders

- Exploiting the auditory limitation
  - Noise shaping

DPCM

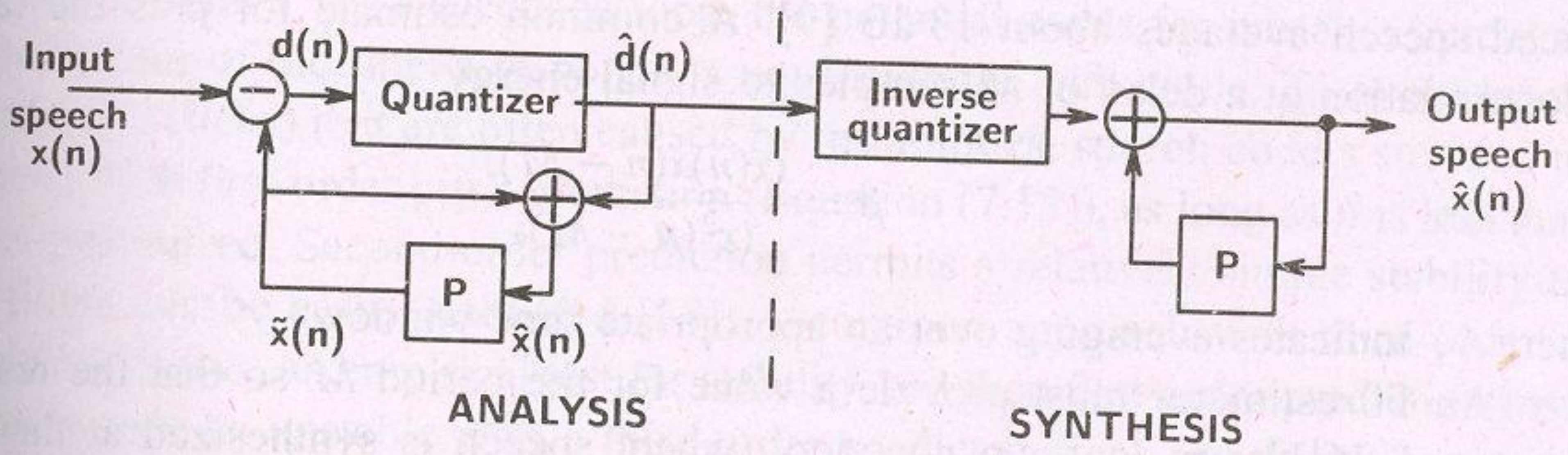


Figure 7.17 General differential PCM: coder on the left, decoder on the right. P is a linear predictor, and the inverse quantizer simply converts transmitted codes back into a single $\hat{d}(n)$ value.

DPCM (Cont..)

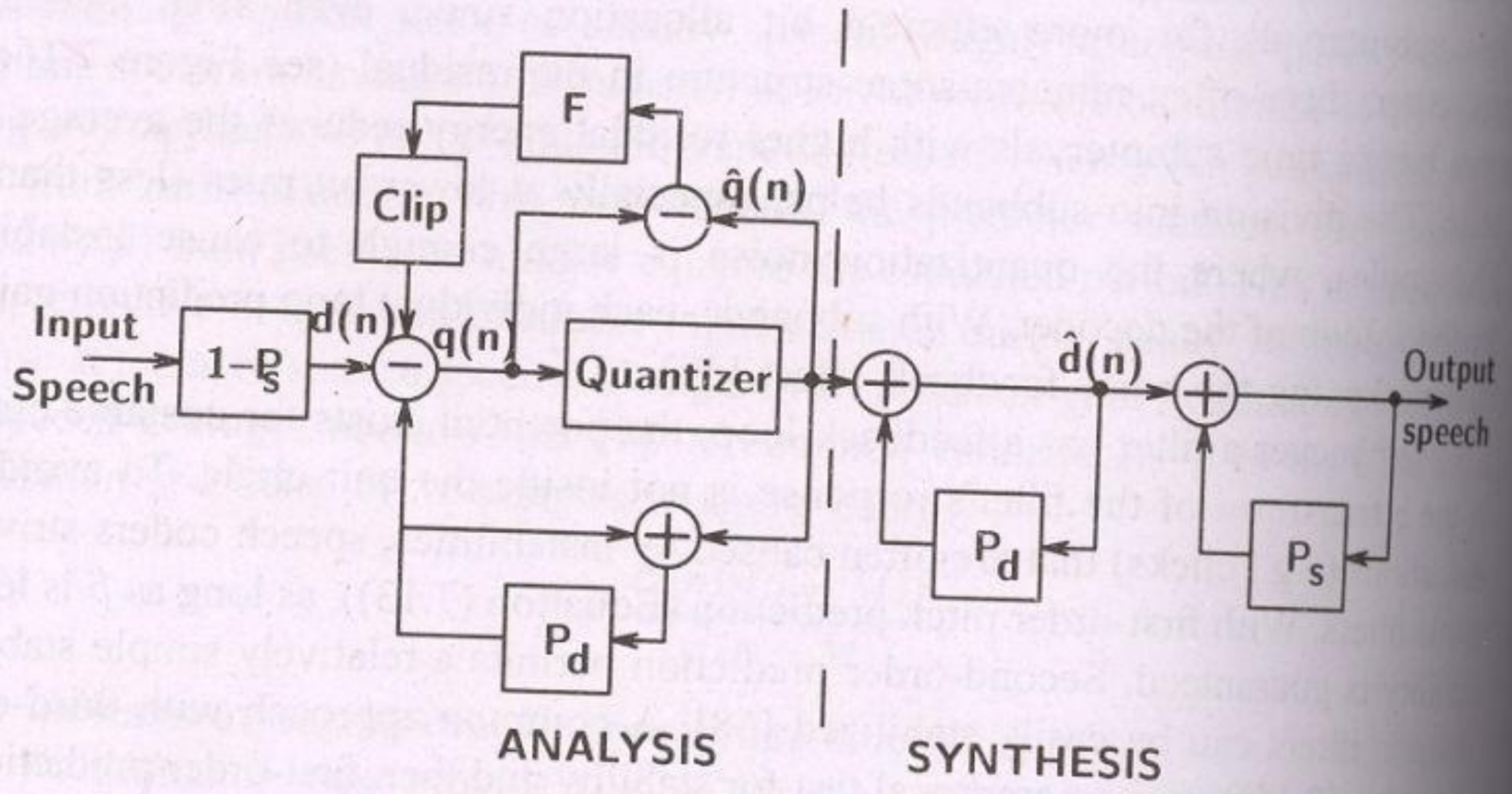


Figure 7.18 APC system with two stages of prediction: a predictor, P_s, based on the short-time spectral envelope and another, P_d, based on pitch structure. The filter F shapes the quantization noise to have less perceptual effect on the output speech. (After Atal and Schroeder [61] © IEEE.)

Linear Predictive Coding

- DPCM, LDM, ADM, ADPCM
- LP coding vs LP analysis by synthesis
- Different excitation models
  - Simple and mixed excitation model
  - Residual excited LP vocoder
  - Multipulse excited LP vocoder

Delta Modulation

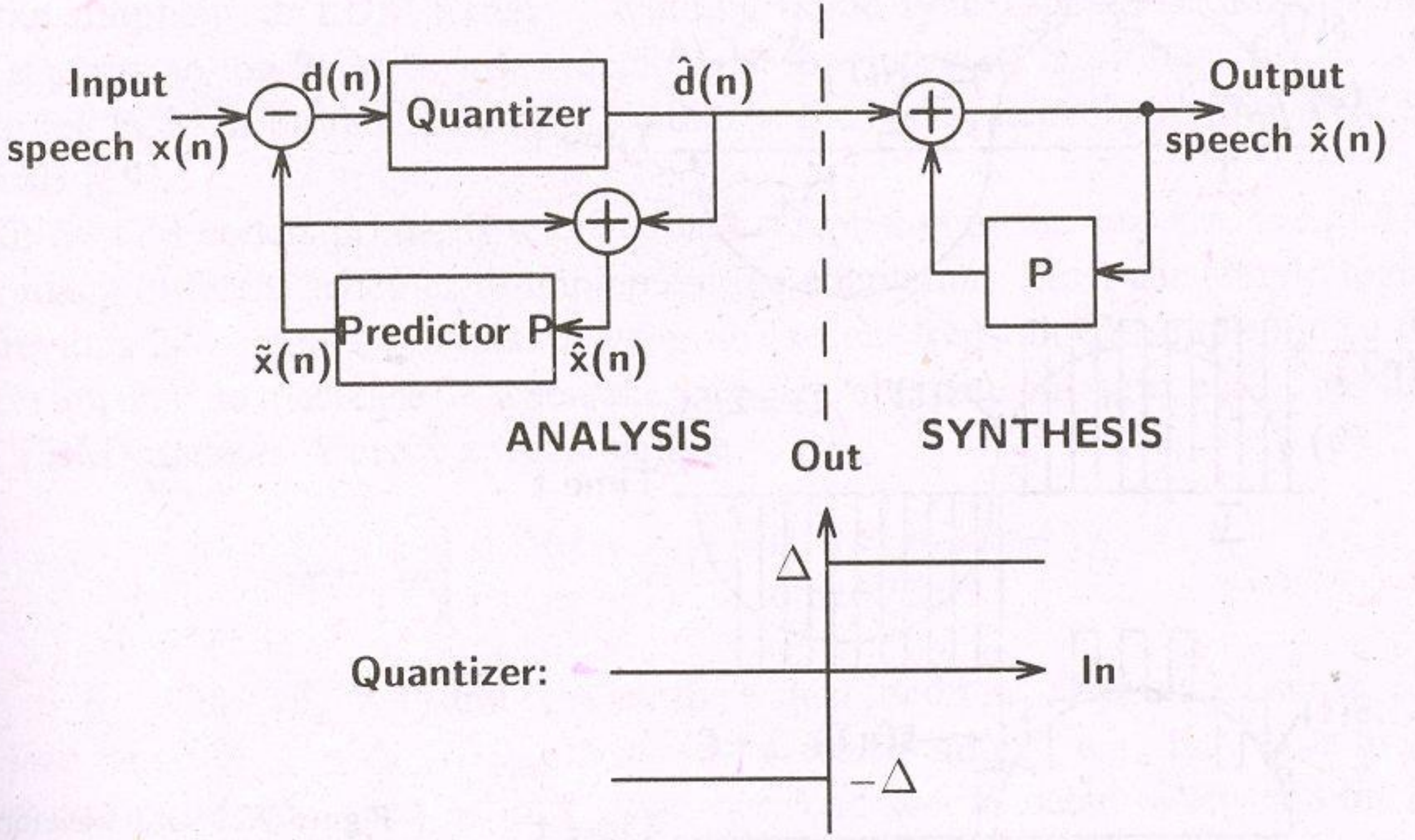


Figure 7.21 Delta modulation system.

SNR for Different Coding Schemes

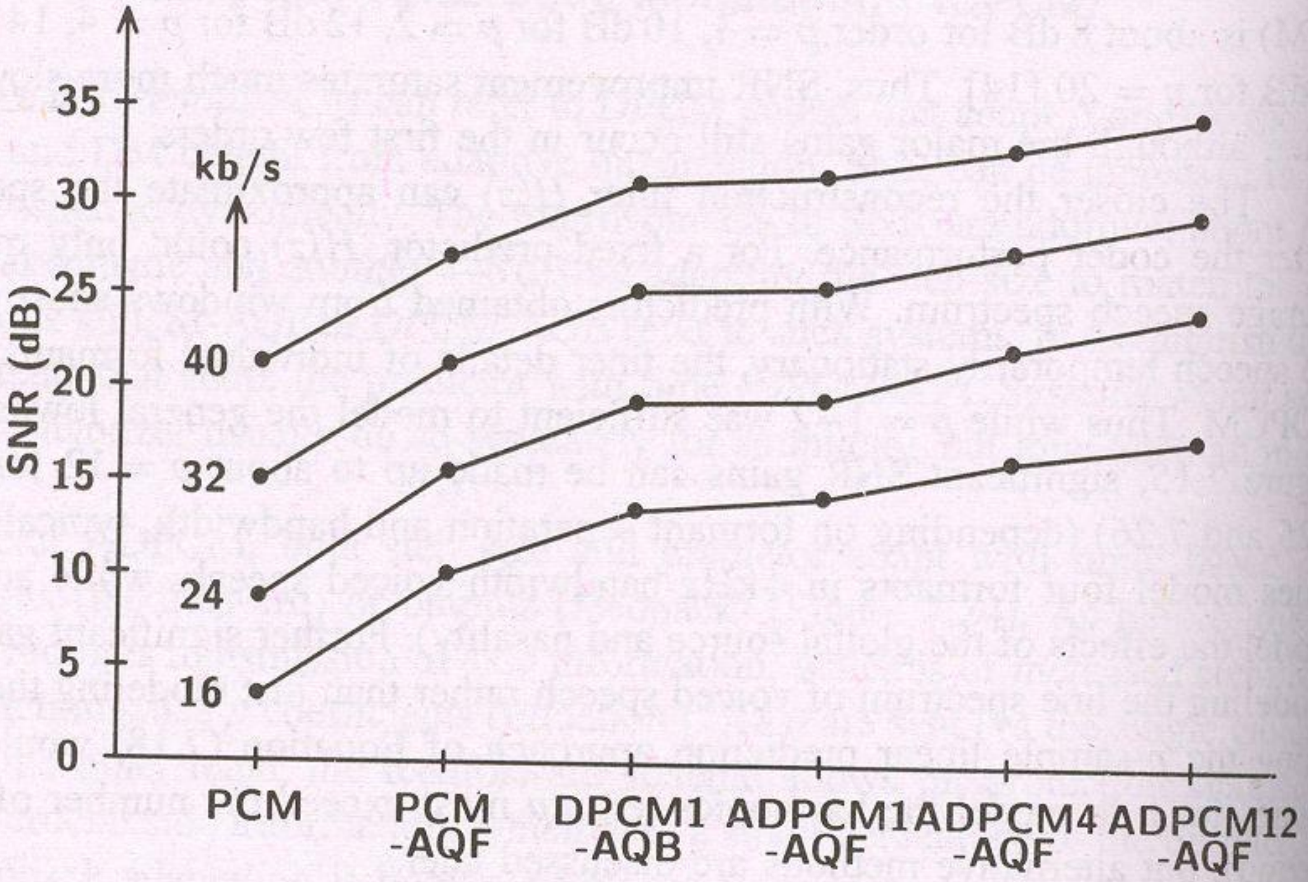


Figure 7.26 SNR for quantization with 16 kb/s ($B = 2$) up to 40 kb/s ($B = 5$). DPCM*i* represents DPCM with an *i*th-order predictor. (Reprinted with permission from *The Bell System Technical Journal* [76] © 1975, AT&T.)

RELP Vocoder

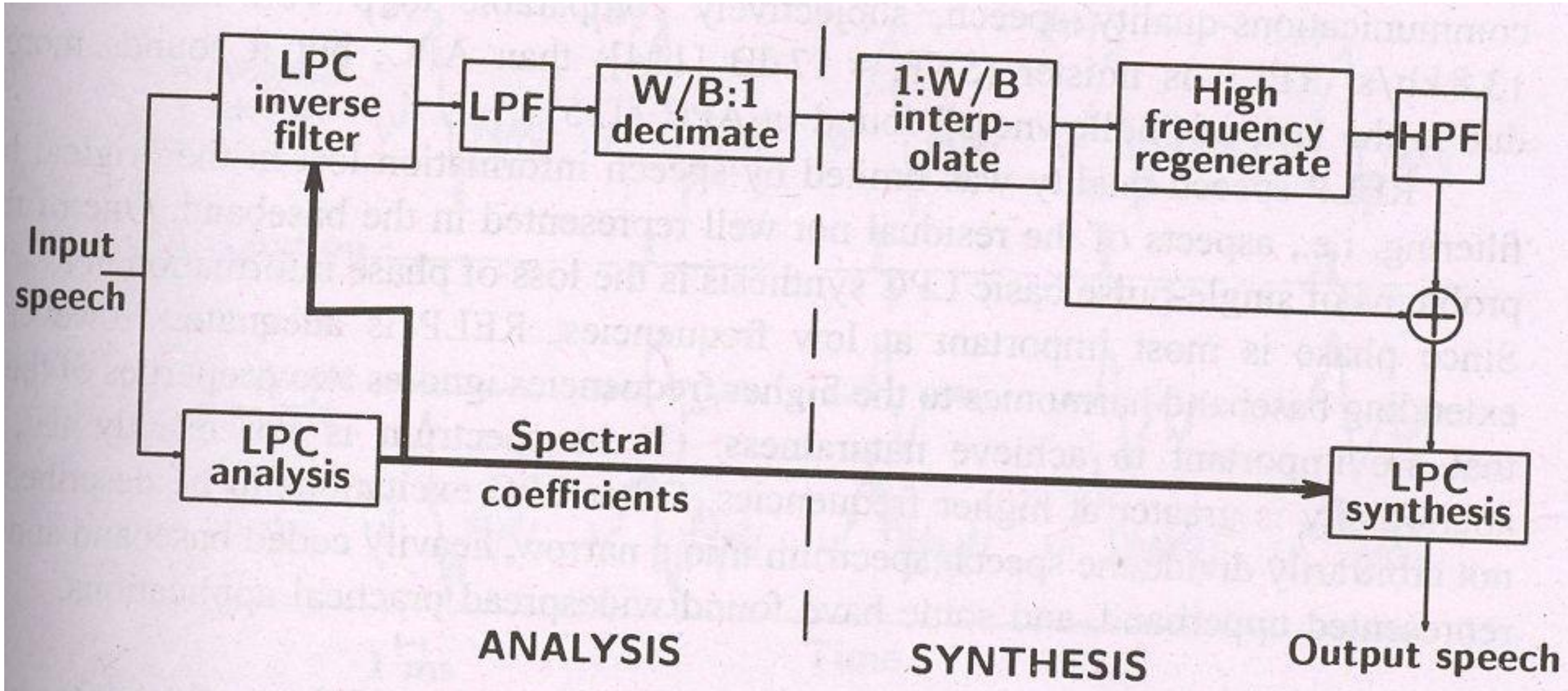


Figure 7.27 Block diagram of a residual-excited linear predictive (RELP) vocoder. LPF and HPF represent lowpass and highpass filters, respectively. *W* and *B* are the bandwidths of the original speech and the decimated residual, respectively. (After Viswanathan *et al.* [111] © IEEE.)

Analysis by Synthesis Coder

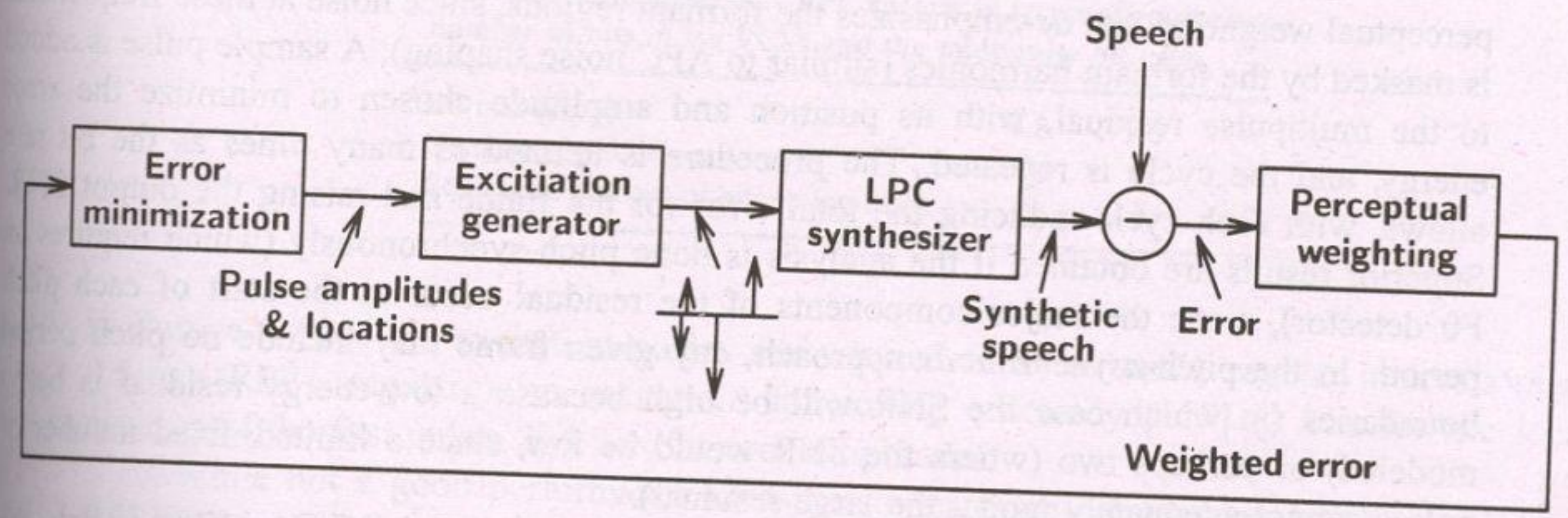


Figure 7.28 An analysis-by-synthesis procedure for determining locations and amplitudes of pulses in MLPC. (After Atal and Remde [116] © IEEE.)

Multi-Pulse Excitation

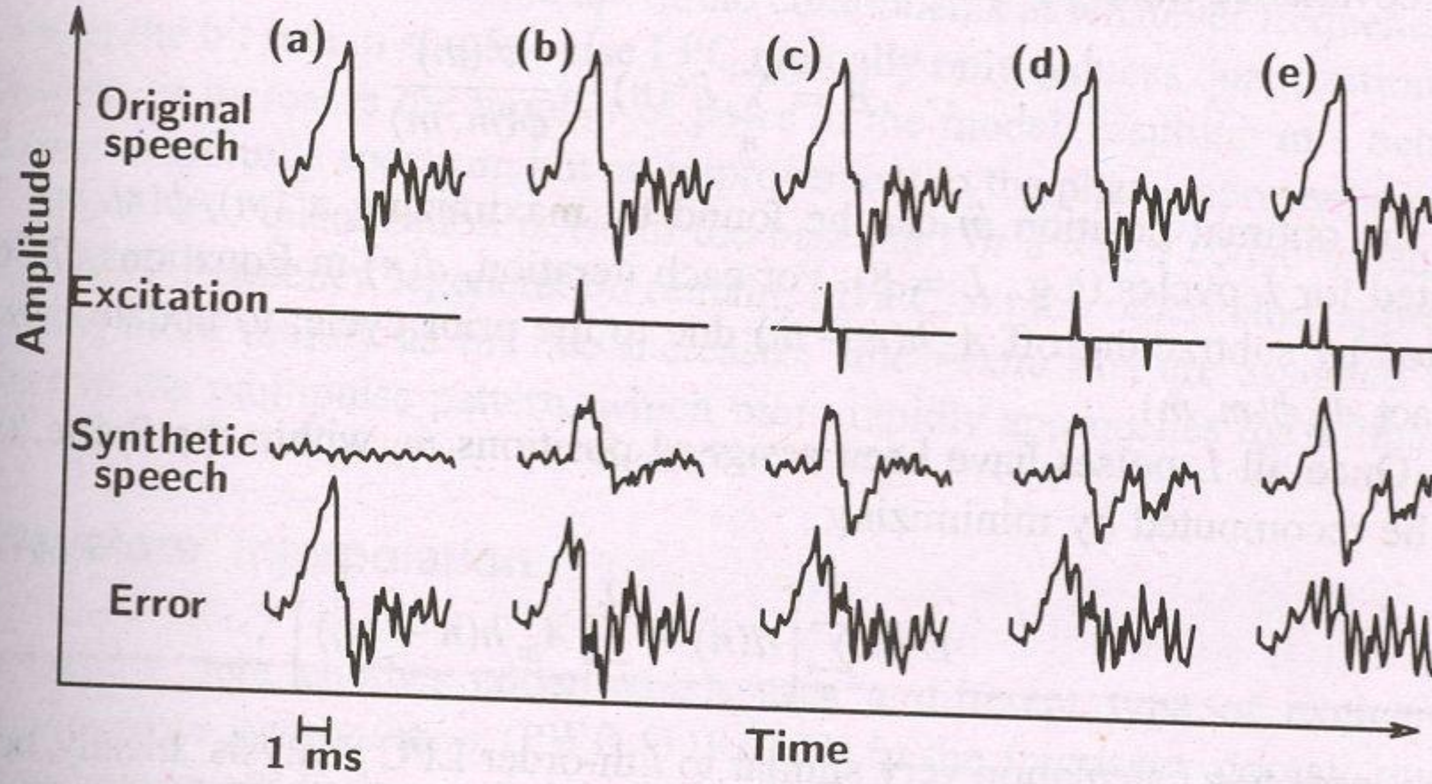


Figure 7.29 Five iterations in the determination of the excitation pulses. In (a), the output is due to the effects of pulses from the previous excitation frame. As each new pulse is added in (b)–(d), the error energy decreases, and the synthetic speech more closely resembles the input. (After Atal and Remde [116] © IEEE.)

Filter-Bank Coder

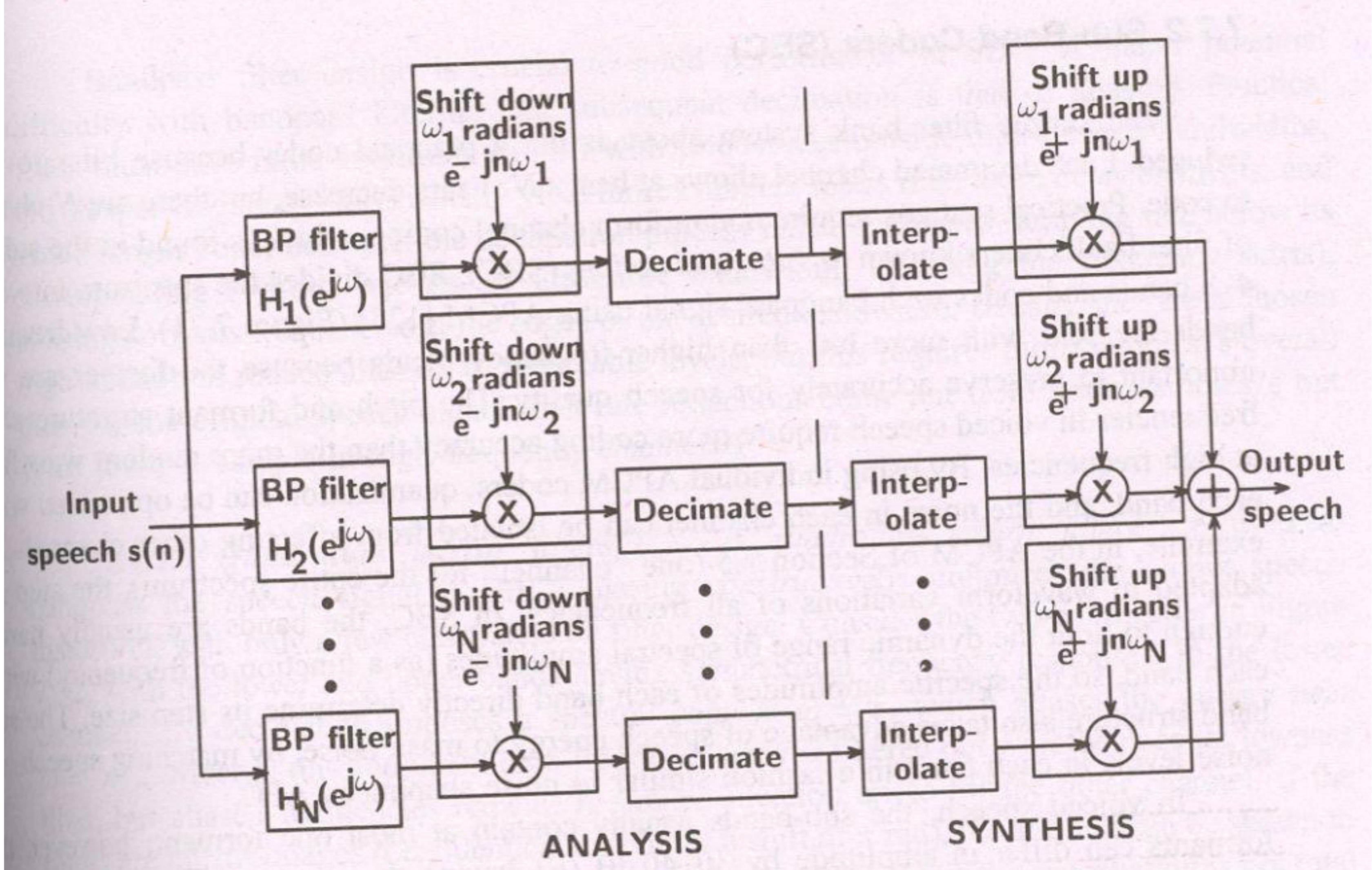


Figure 7.30 Analysis and synthesis operations for a filter bank coder.

Sub-Band Coder

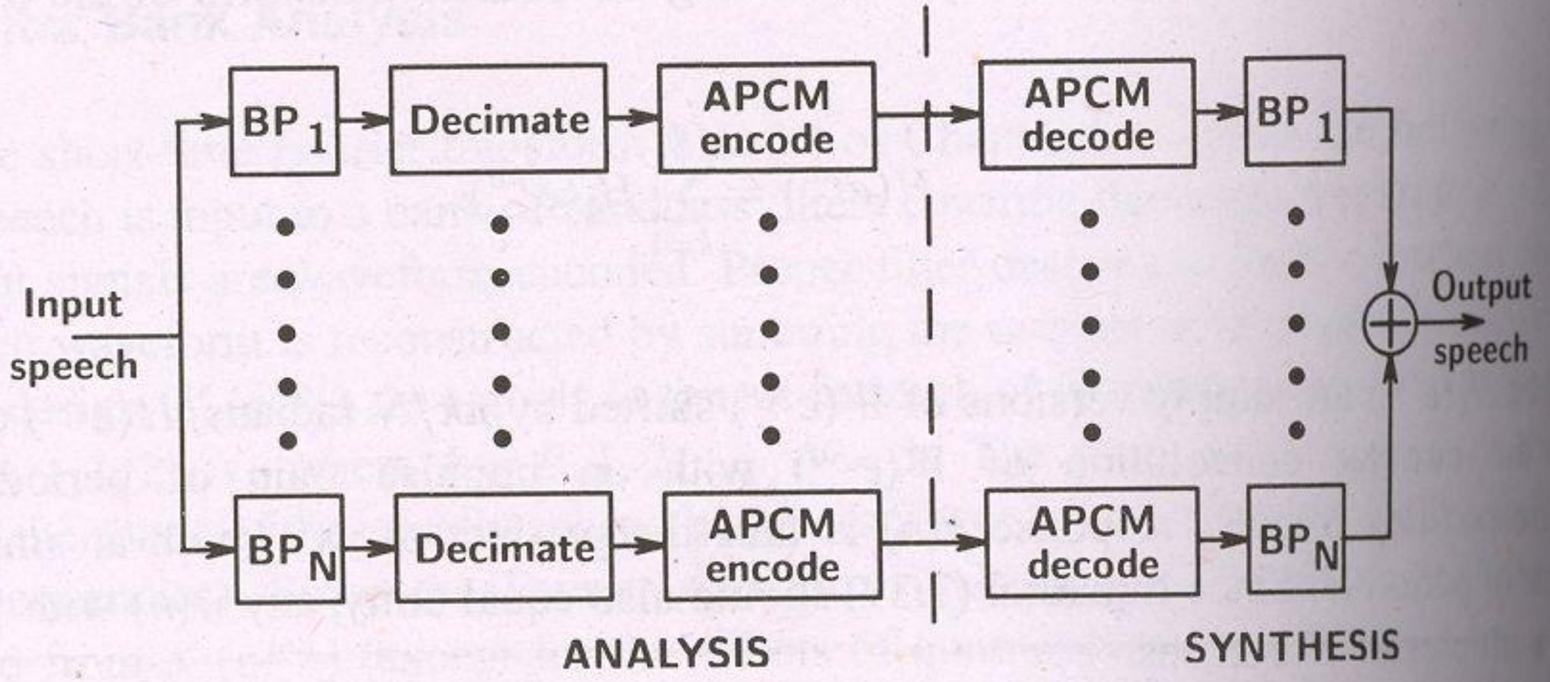


Figure 7.31 Sub-band coder with N bands. The decimators have output sampling rates of $2f_i$, where f_i is the high-frequency cutoff for bandpass filter BP_i. In many systems, the decimation rate can be as low as twice the bandwidth of BP_i. The decoder in each channel includes the interpolator corresponding to the channel's decimator. (After Flanagan *et al.* [9] © IEEE.)

Adaptive Transform Coder

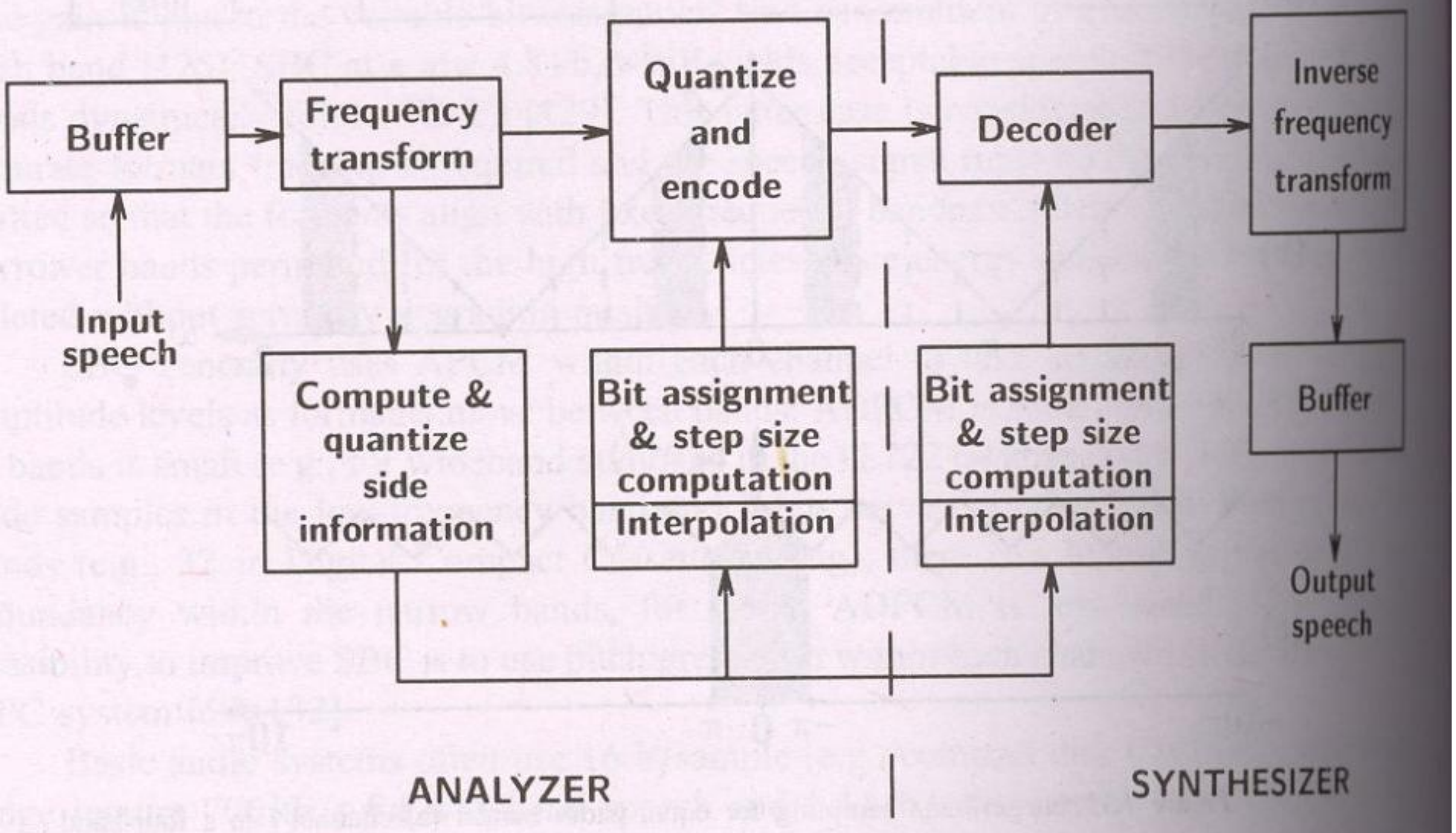


Figure 7.34 Adaptive transform coding. (After Tribolet and Crochiere [137] © IEEE.)

Harmonic Coder

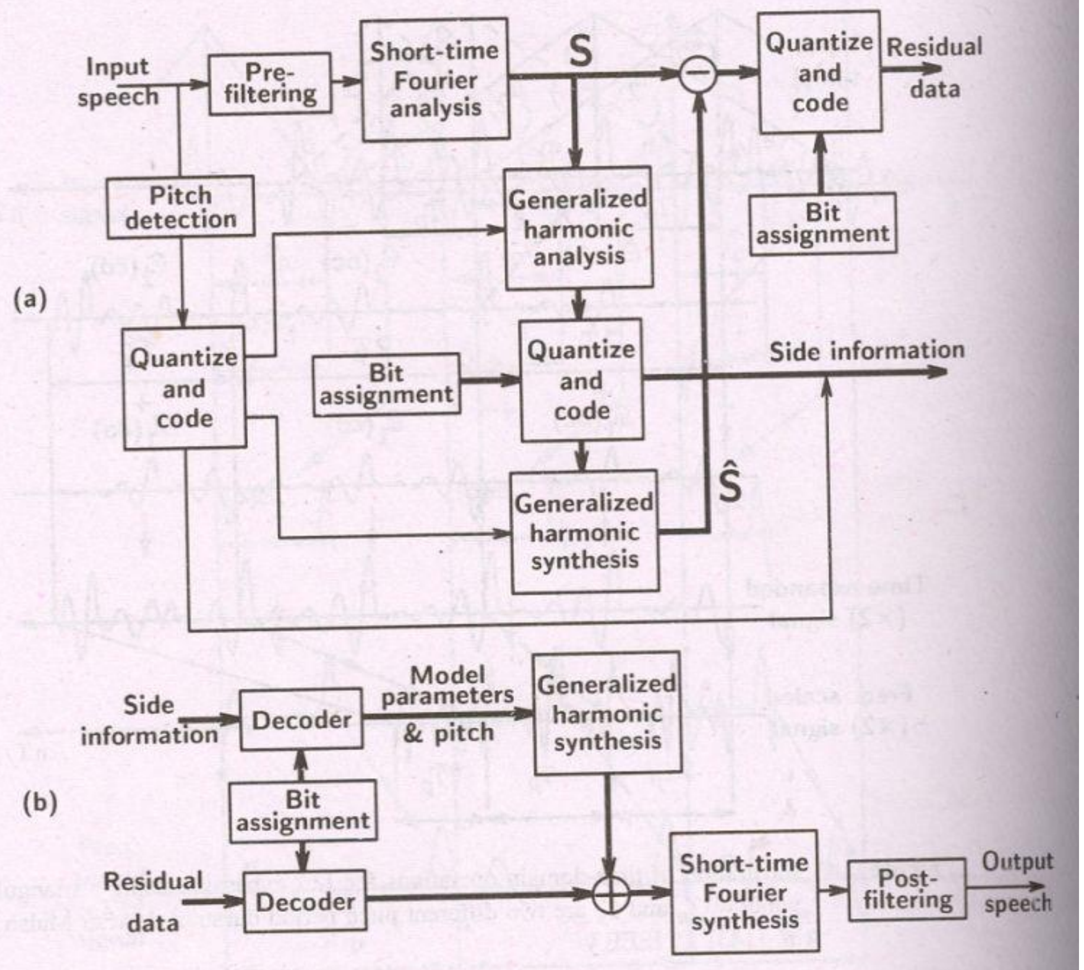


Figure 7.44 Block diagram of a harmonic coder: (a) transmitter, (b) receiver. (After Almeida and Tribolet [150] © IEEE.)

Harmonic Coder with Linear Prediction

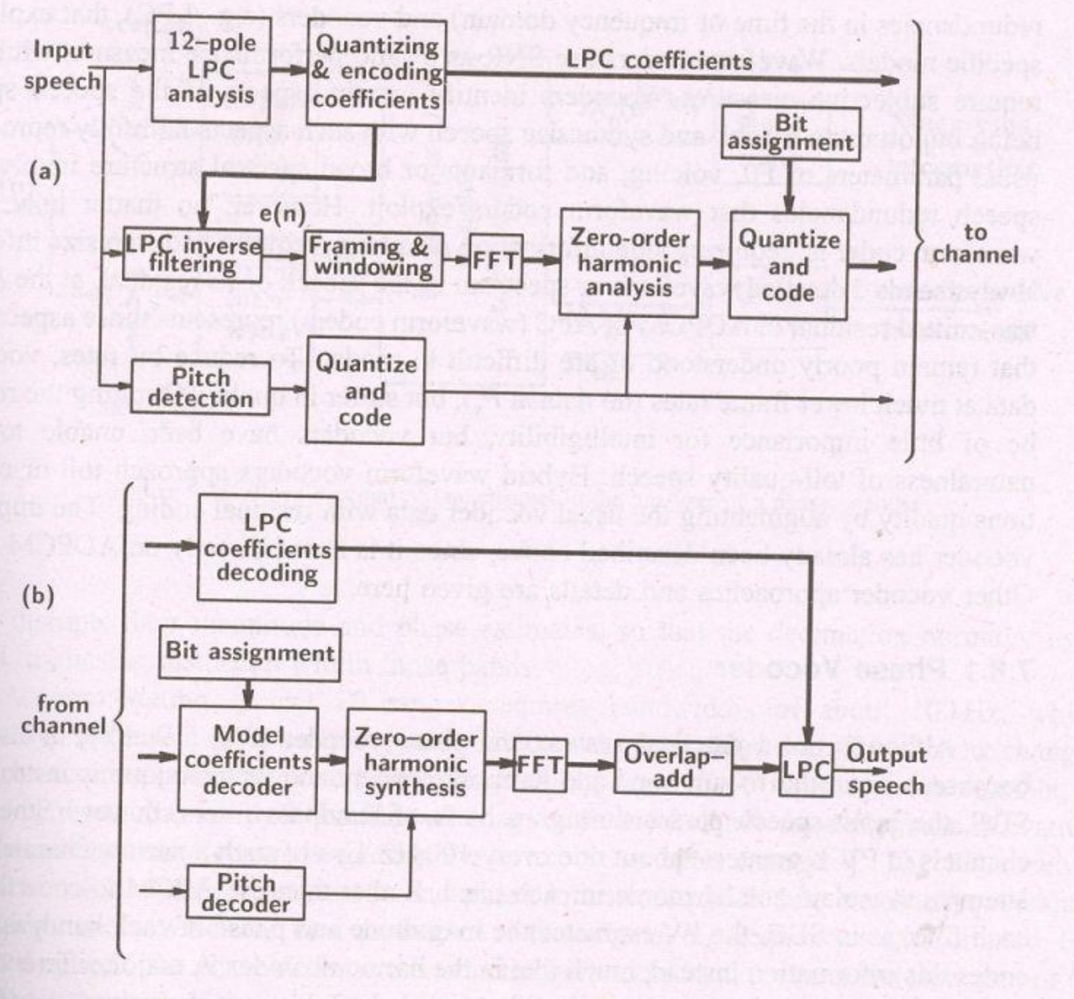


Figure 7.45 Block diagram of a 4.8 kb/s harmonic coder using linear prediction: (a) transmitter, (b) receiver. (After Almeida and Tribolet [150] © IEEE.)

Phase Vocoder

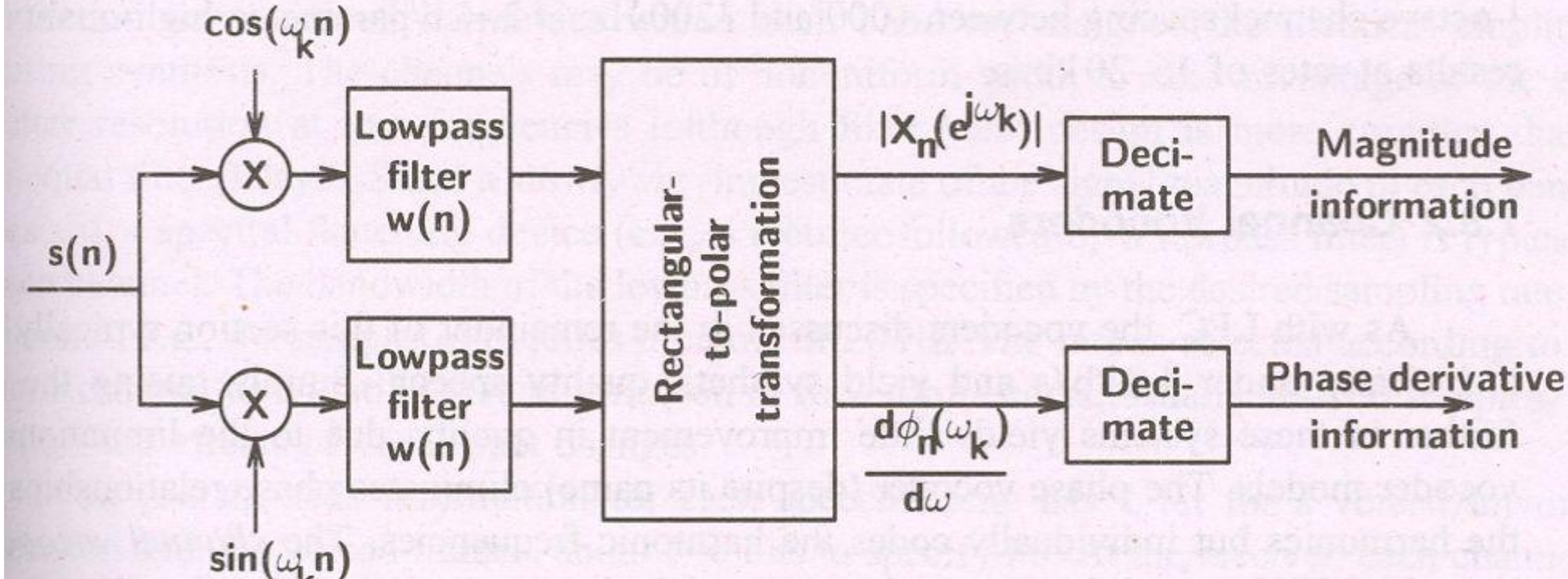
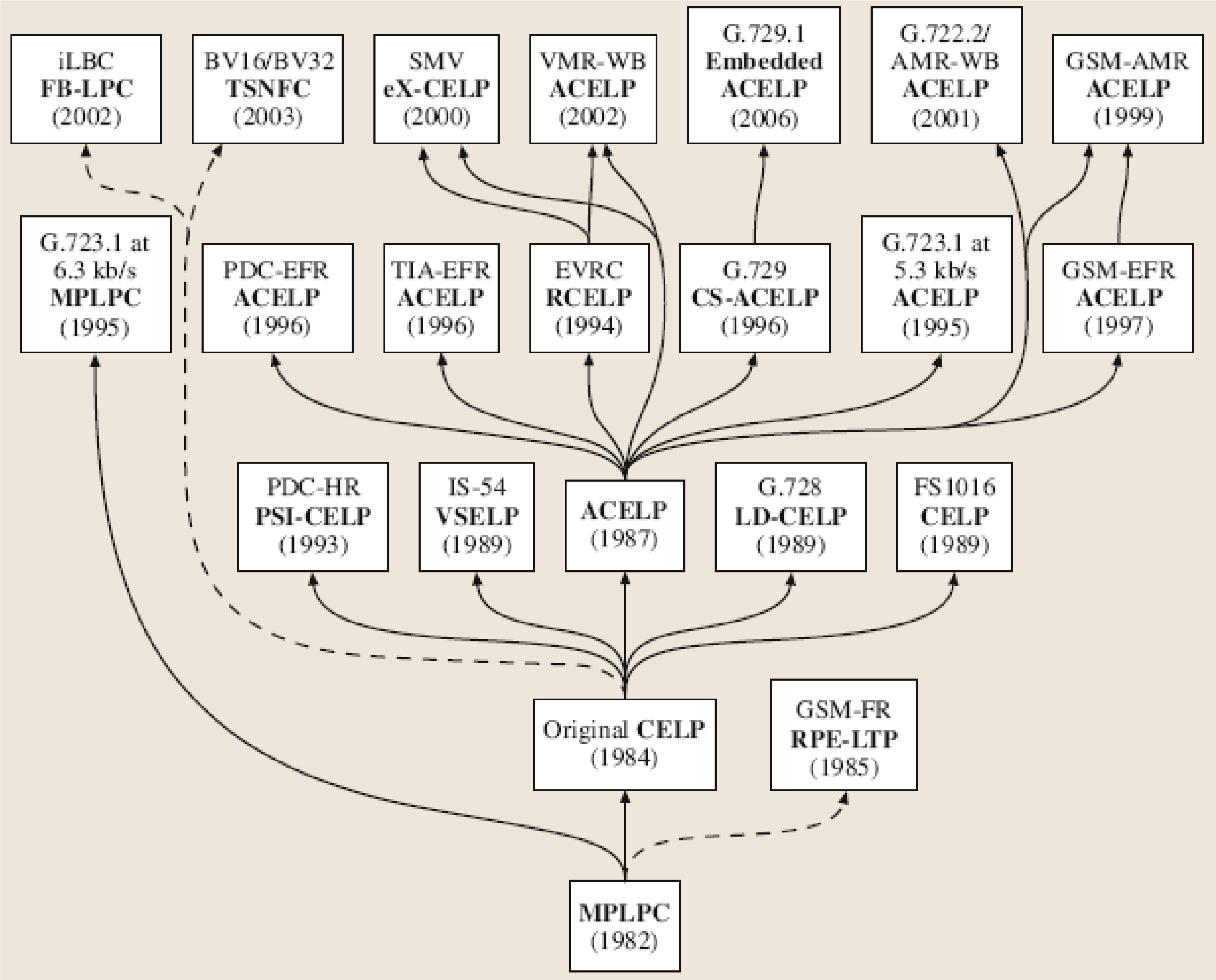
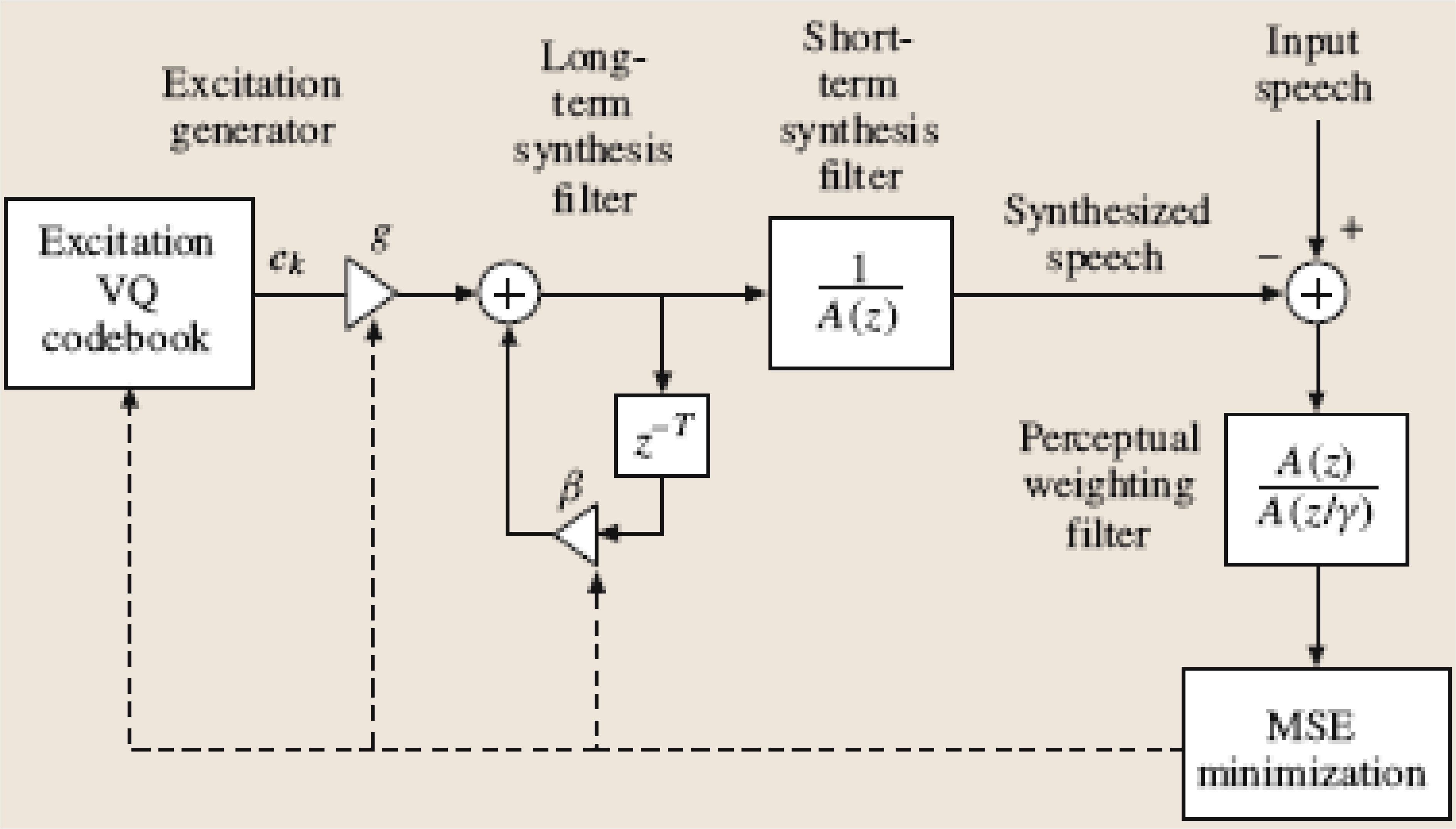


Figure 7.46 Block diagram of one channel in the analyzer of a phase vocoder.

Family of Analysis by Synthesis Coders



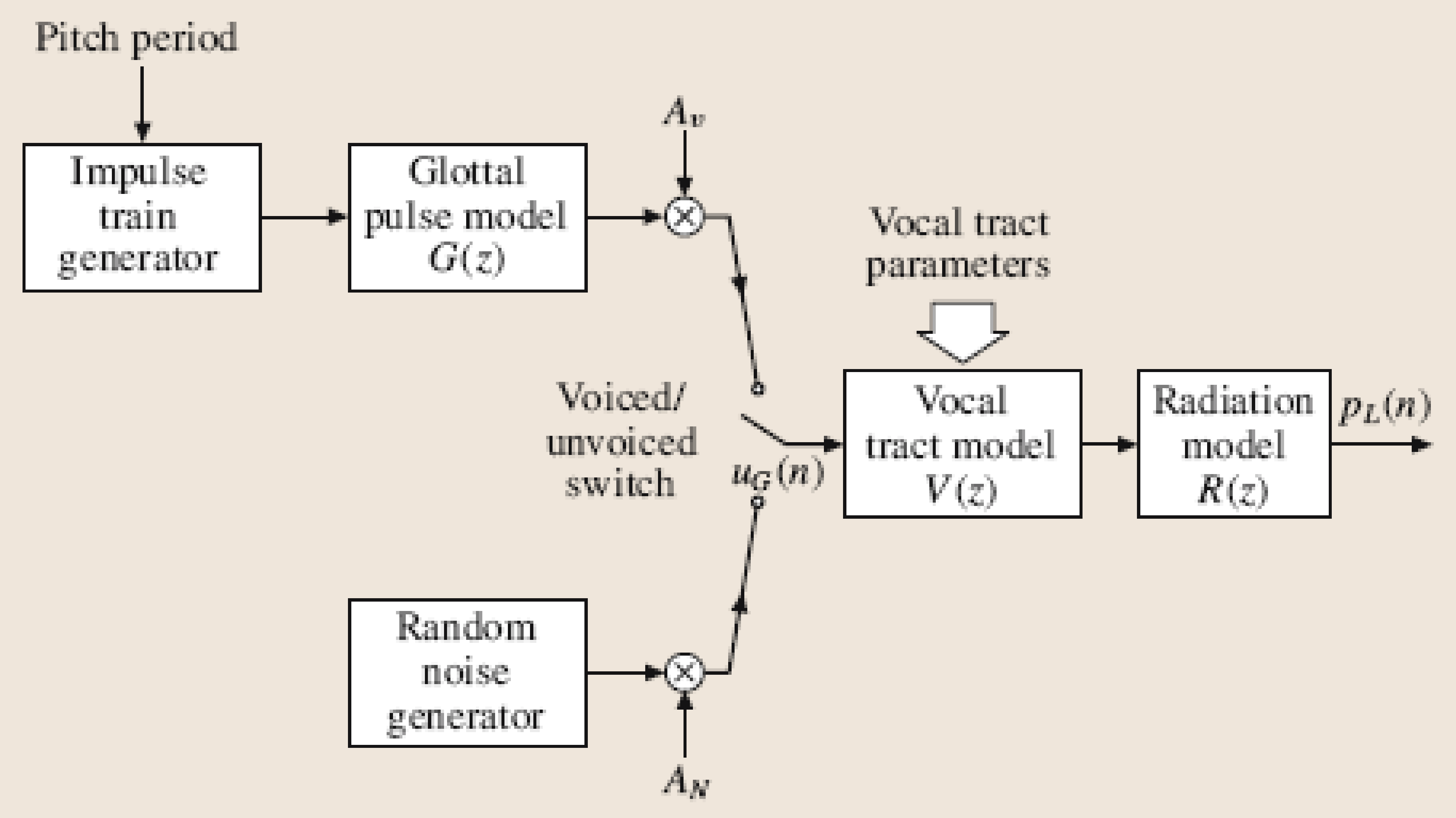
CELP Coder



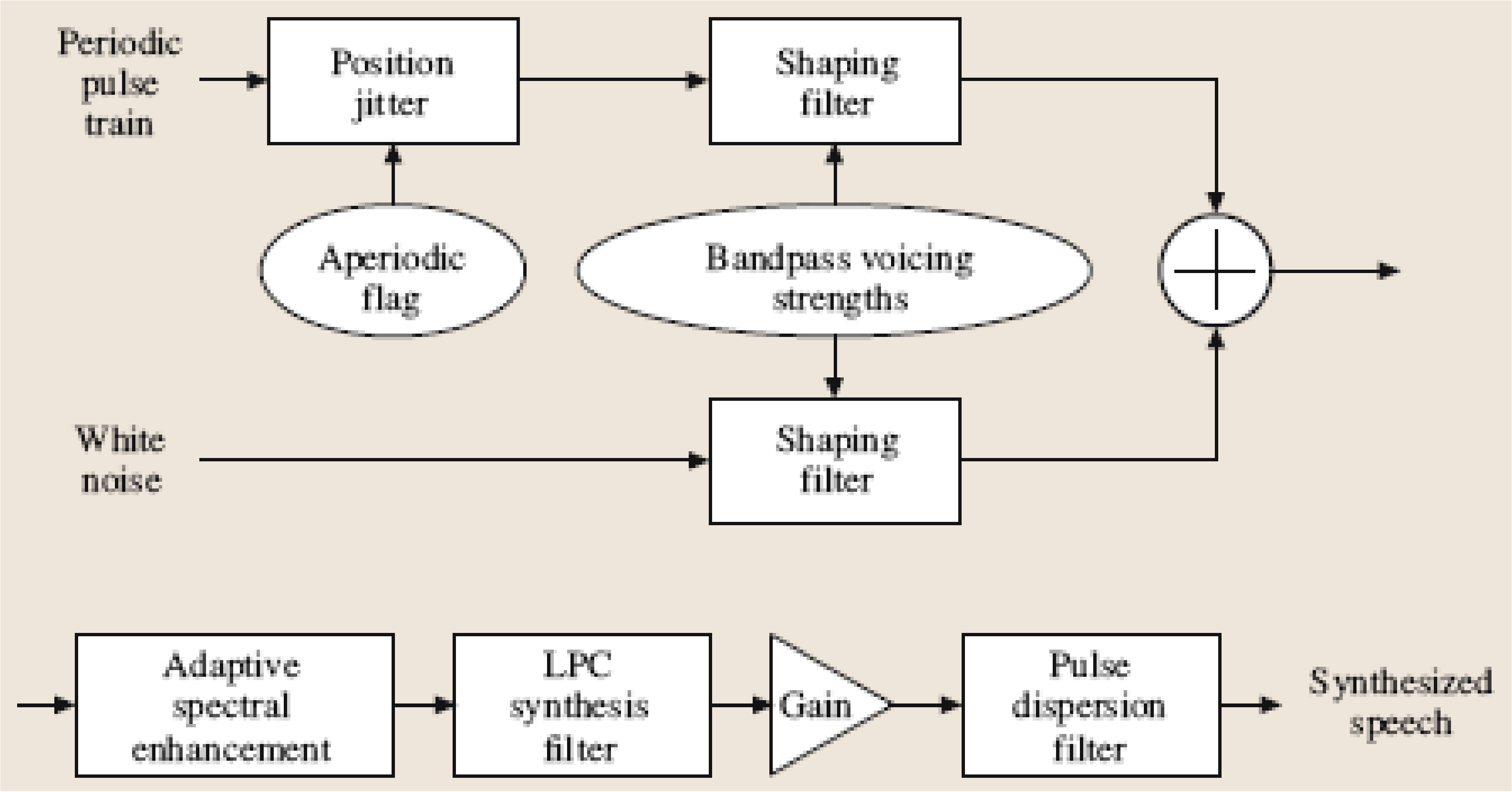
Excitation generator
Long-term synthesis filter
Short-term synthesis filter
Input speech
Excitation VQ codebook
c_k
g
$\frac{1}{A(z)}$
Synthesized speech
+
−
z^{-T}
β
Perceptual weighting filter
$\frac{A(z)}{A(z/\gamma)}$
MSE minimization

Low-Bit Rate Coders

Linear Model of Speech Production



Mixed Excitation Linear Prediction (MELP) Synthesizer



MELP Coder Bit Allocation (2.4 Kbps)

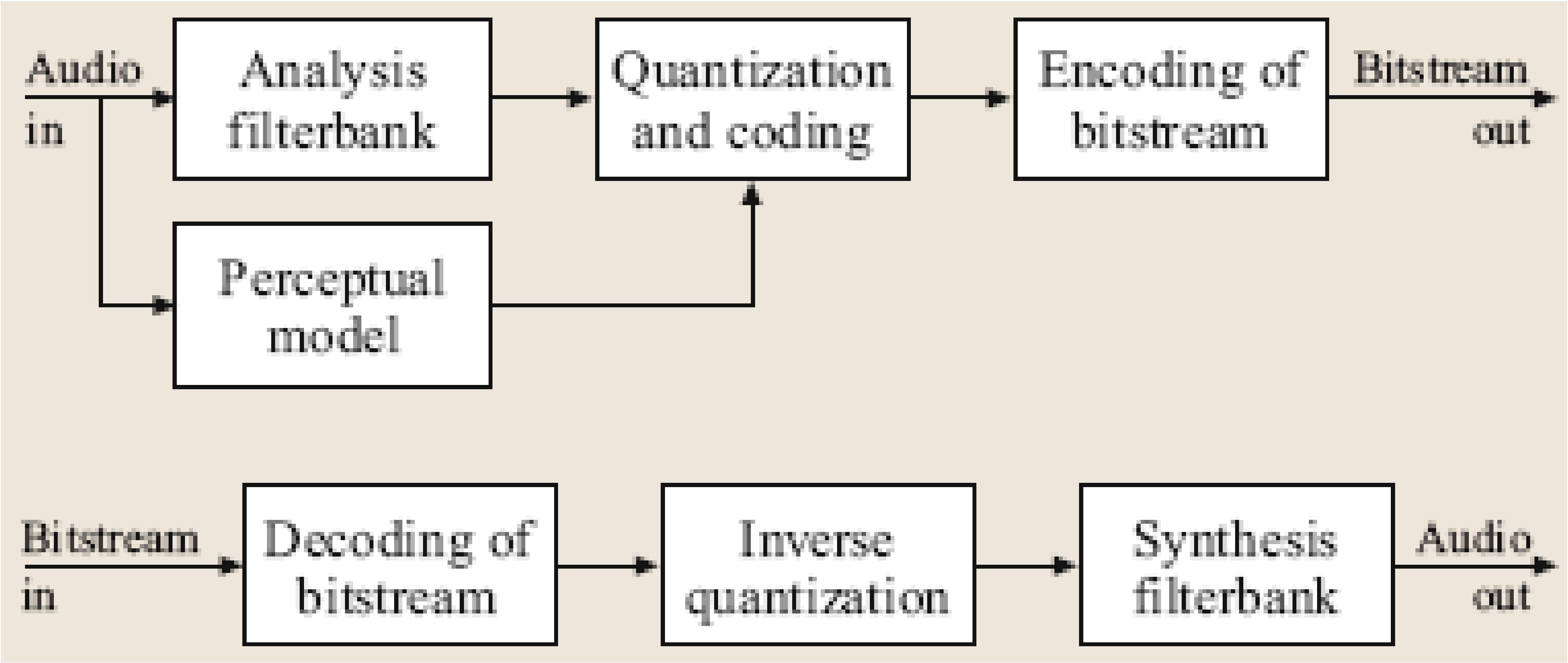
Parameters	Voiced	Unvoiced
LSF	25	25
Fourier magnitudes	8	–
Gain (2 per frame)	8	8
Pitch and overall voicing	7	7
Bandpass voicing	4	–
Aperiodic flag	1	–
Error protection	–	13
Sync bit	1	1
Total bits/22.5 ms	54	54

Low bit rate speech coding standards

- MIL-STD 3005 (MELP 2.4 Kbps)
- NATO STANAG 4591 (MELP 2.4/1.2/0.6 Kbps)

Perceptual audio coding

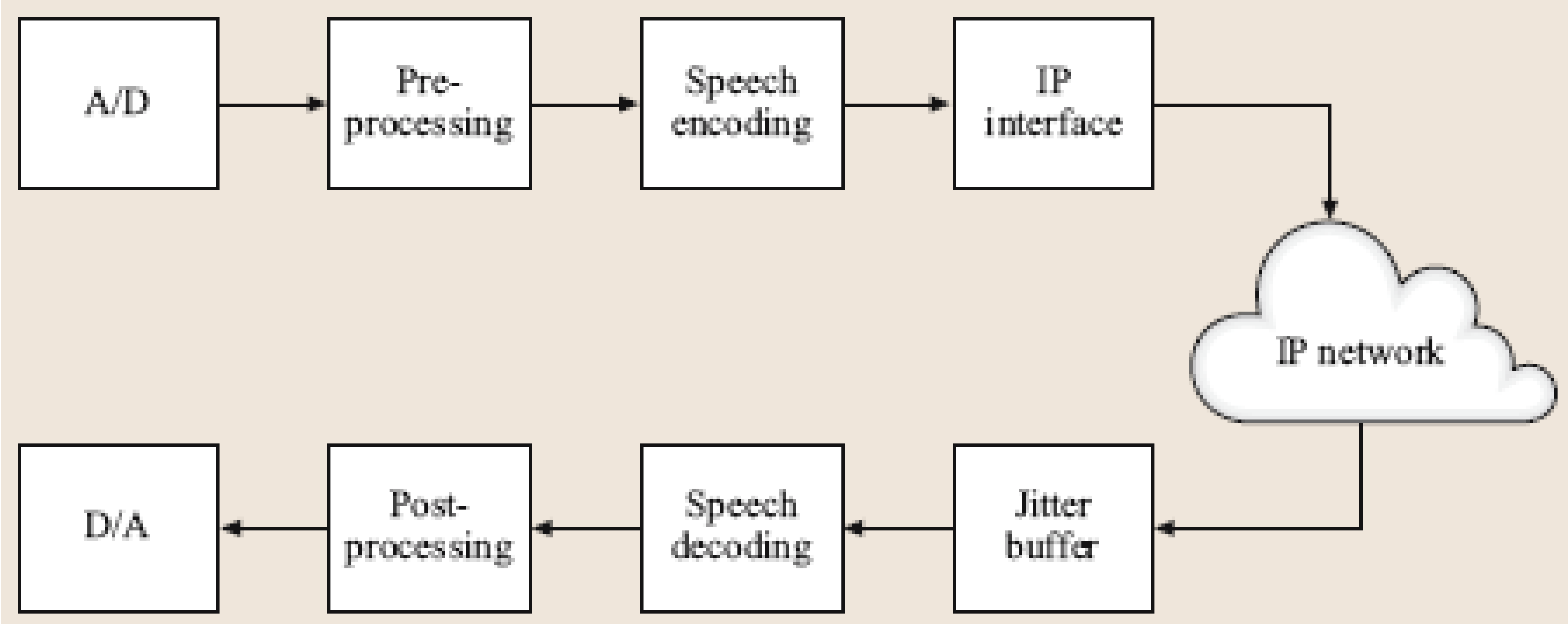
Perceptual Audio Coder



Standard Audio Coders

- MPEG-1
- MPEG-2
- MPEG-2 Advanced Audio Coding
- MPEG-4 Advanced Audio Coding

VoIP: speech coding for packet networks



VoIP System

- Echo cancellation
- Speech codec
- Jitter buffer
- Packet loss recovery

- Packet Loss Concealment
  - ✓ Nonparametric concealment
  - ✓ Parametric concealment

Thank You