Analysis-by-Synthesis Coding

Nimrod Peleg
Update: Jan 2004

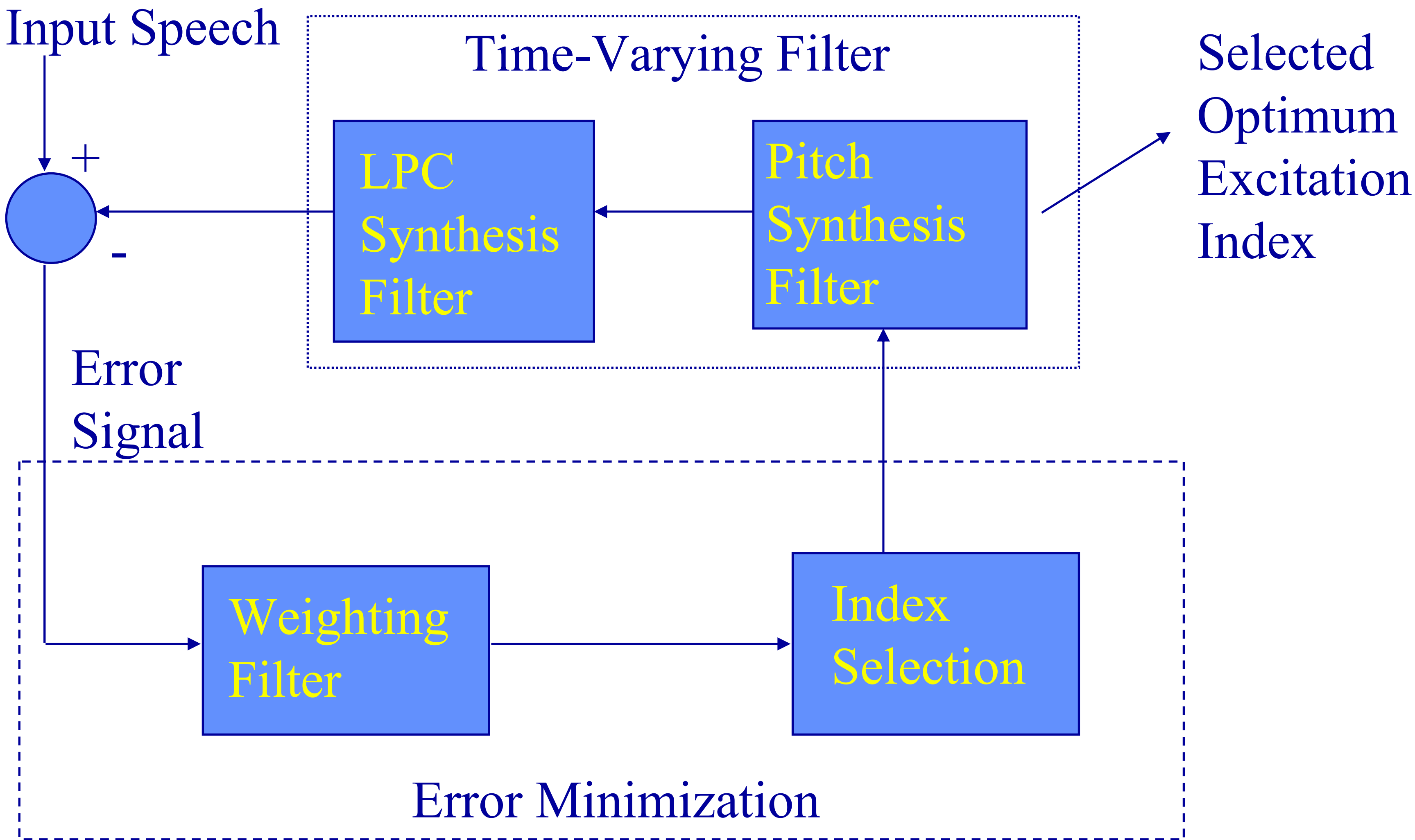
A-by-S Coding

- RELP, APC, ATC and SBC have sufficient quality at rates of 9.6-16Kbps
- Two main reasons:
  - coded speech is not analyzed to check efficiency (No distortion control)
  - Errors accumulated from previous frames are not taken into account in current frame (errors propagate since no "reset" in loop

Analysis-by-Synthesis LP Coders

- Analysis-by-Synthesis coders use close-loop for the excitation sequence determination
- An optimization process determines an excitation sequence which minimizes a measure of the difference between input and coded speech
- A weighting function is chosen to optimize for human ear
- A better quality achieved for 4.8-9.6Kbps

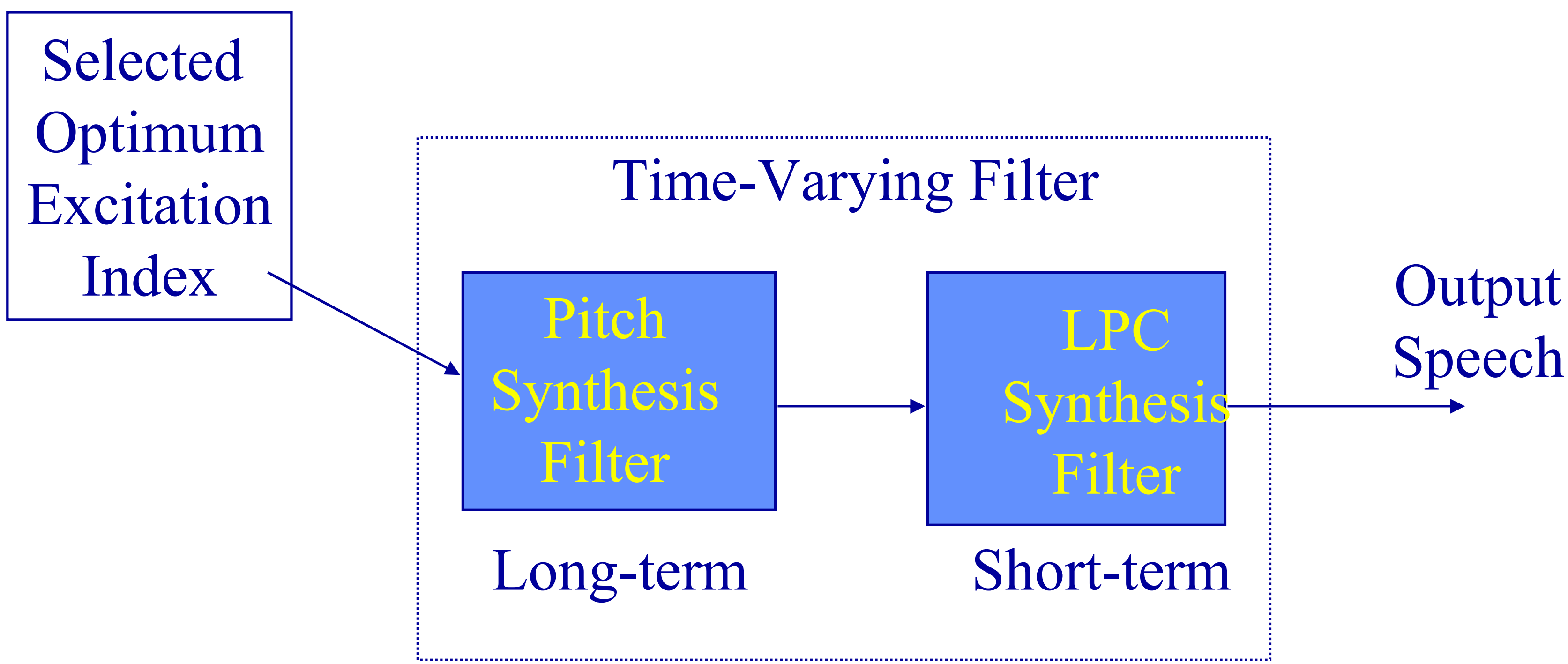
General A-by-S Encoder



A-by-S Operation

- Initialize filters (low-level random noise)
- Determine LPC coeff. on a frame of samples
- Divide the frame to subframes, for each:
  - calculate pitch parameters (delay & scaling)
  - Pitch and LPC filters are grouped together. The cascaded filter is used to determine the best secondary excitation, to minimize error between original and synthesized speech
- Final speech is generated by passing optimum secondary excitation through cascaded filter

General A-by-S Decoder



Short-Term Predictor

- Models the short-term correlation in the speech : Spectral envelope
- It is time varying to reflect changes in speech spectrum
- Adaptation rates: 20-30mSec
- Filter order: 8-12

Long-Term Predictor

- Models the long-term correlation in the speech: Pitch period
- Higher adaptation rates: 5-10mSec
- LTP can be omitted (as in MPLPC)
- Both LTP and STP have buffer memory of previous frame that provide smoothing effect to distortion (caused by block oriented analysis)

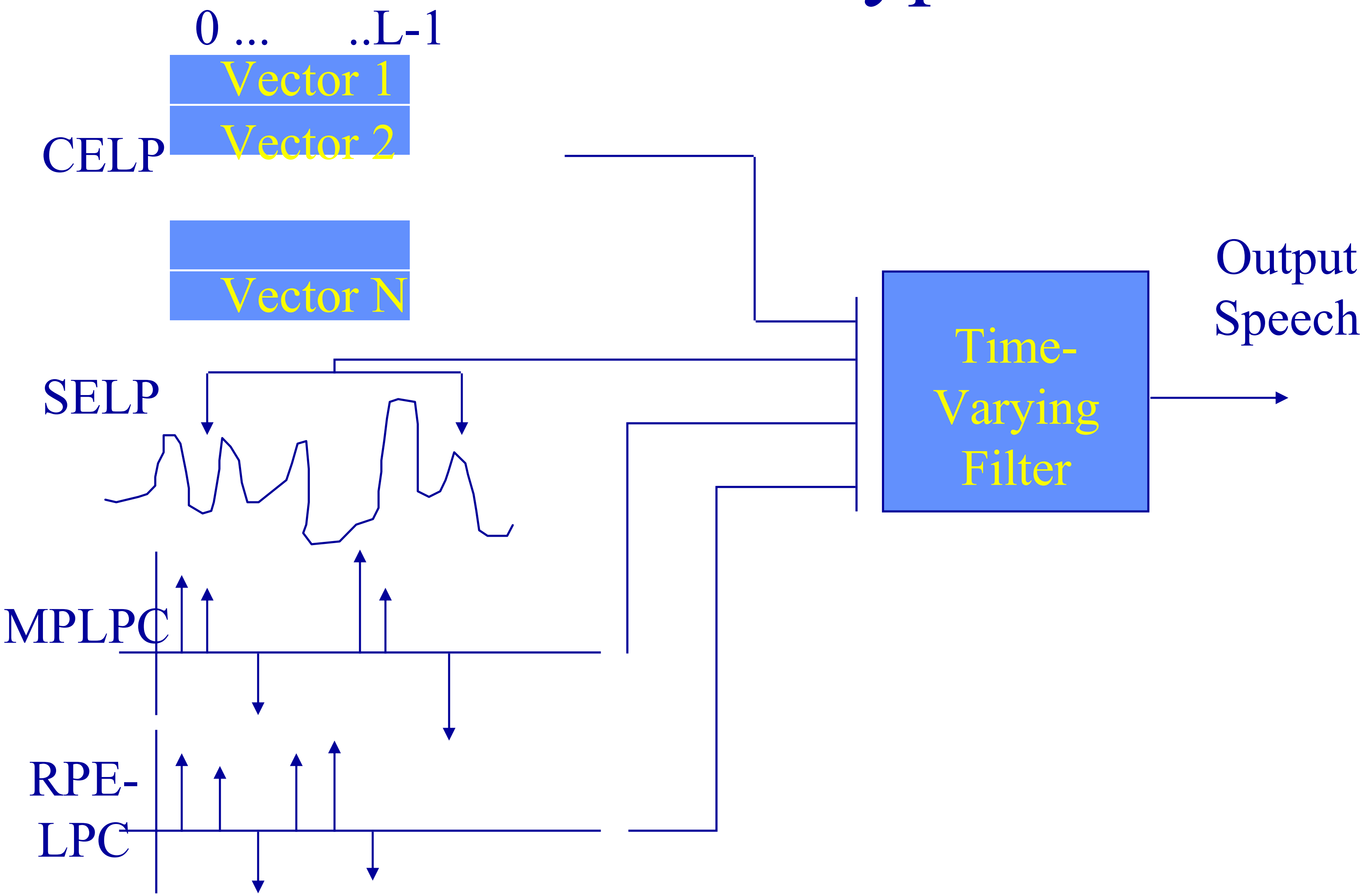
Error Minimization

- Minimizing perceptual error by selecting filter(s) parameters such that MSE between original and reconstructed signal is minimized
- This results in a flat spectrum of quantization error, (better results can be obtained by exploiting the auditory masking of human hearing)
- All the process done in transmitter only !

Excitation Generator

- Includes all information missing in the model parameters (residual signal)
- Several methods for it:
  - CELP (Code Excited LP)
  - SELP (Self Excited LP)
  - MPLPC (Multi-Pulse LP)
  - RPE-LPC (Regular-Pulse LPC)
  - MELP (Mixed Excitation LP)

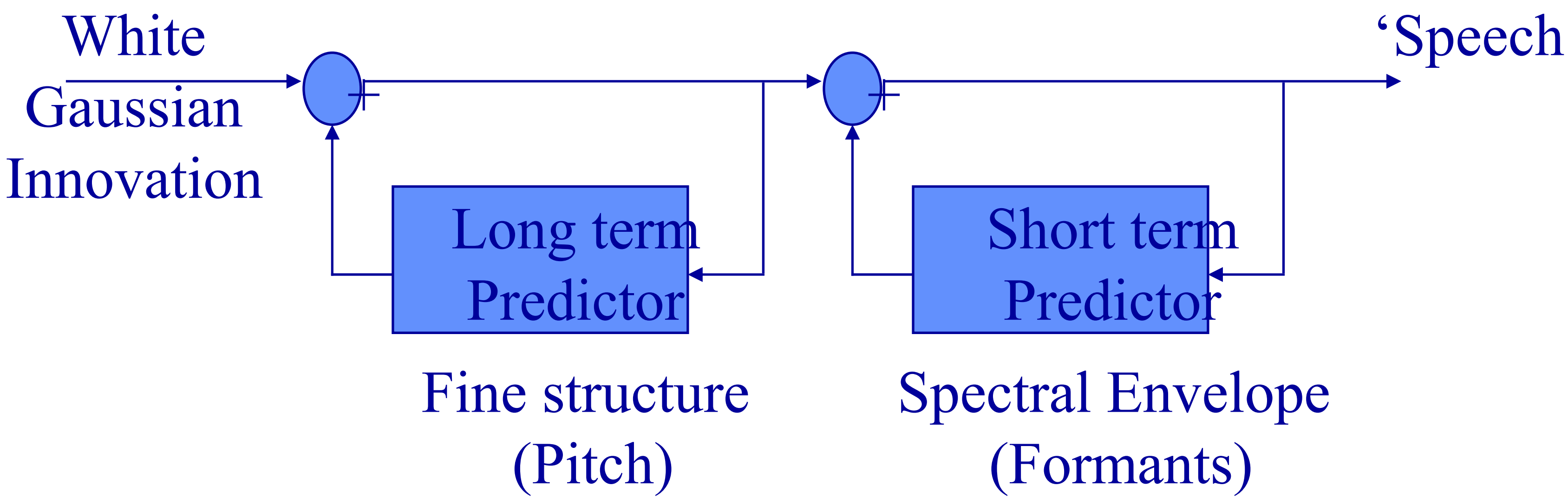
Excitation Types



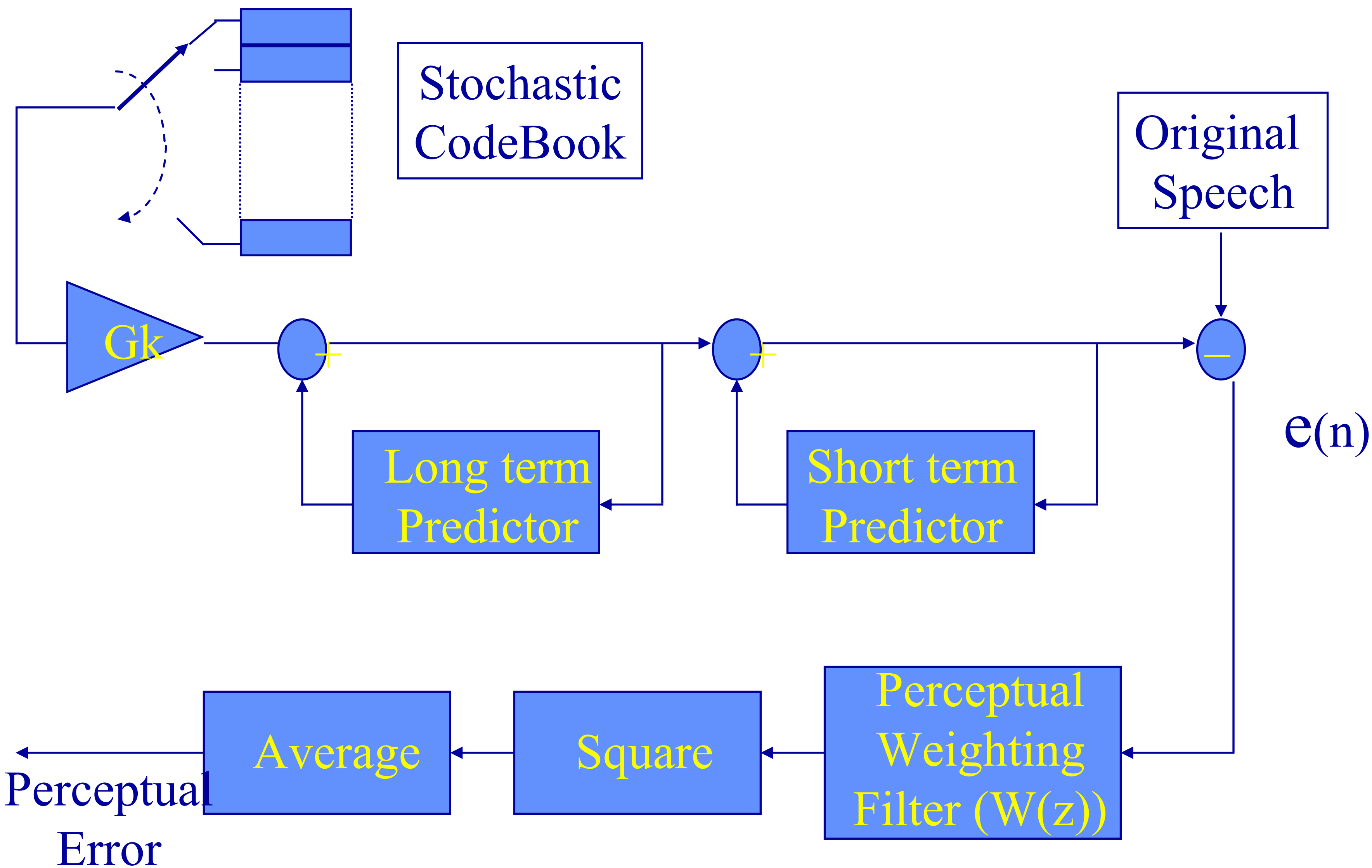
CELP

- Excitation vector chosen from a set of N possible stochastic vectors (+gain)
- The vector that produces lowest error is the desired excitation sequence
- Only the index needed to be sent to receiver
- Many schemes to codebook design, among them: unit variance white Gaussian random noise (Atal)

Speech Synthesis Model



CELP Basic Scheme

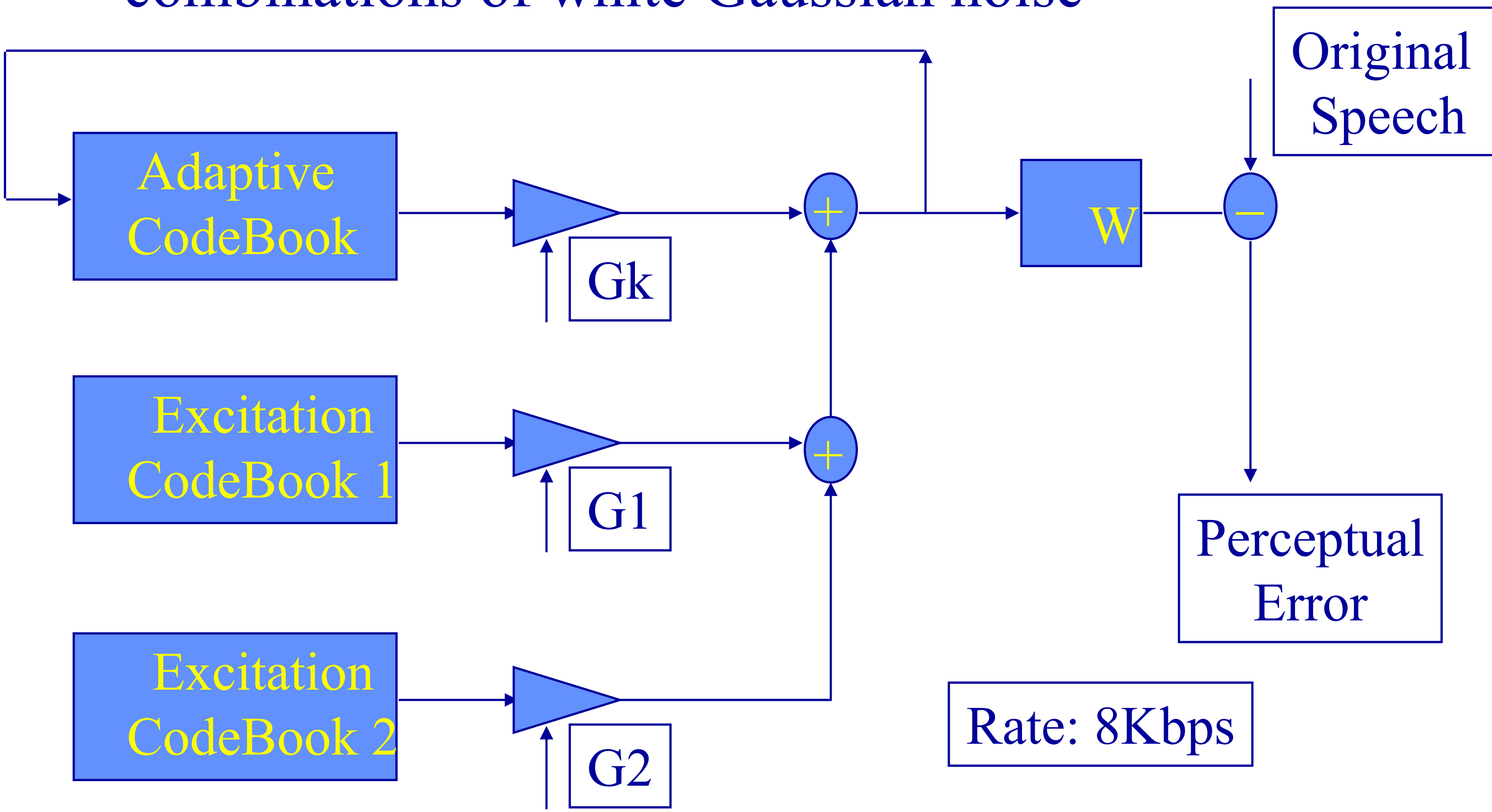


Optional Improvements

- Adaptive CodeBook (closed loop pitch prediction)
- Low-Delay (LD-CELP): Backward adaptive CELP Coder:
  - Linear prediction of order 50
  - Rate: 16Kbps (Toll quality)

Vector Sum Excitation LP (VSELP)

Orthogonal vectors create a basis for linear combinations of white Gaussian noise



Rate: 8Kbps

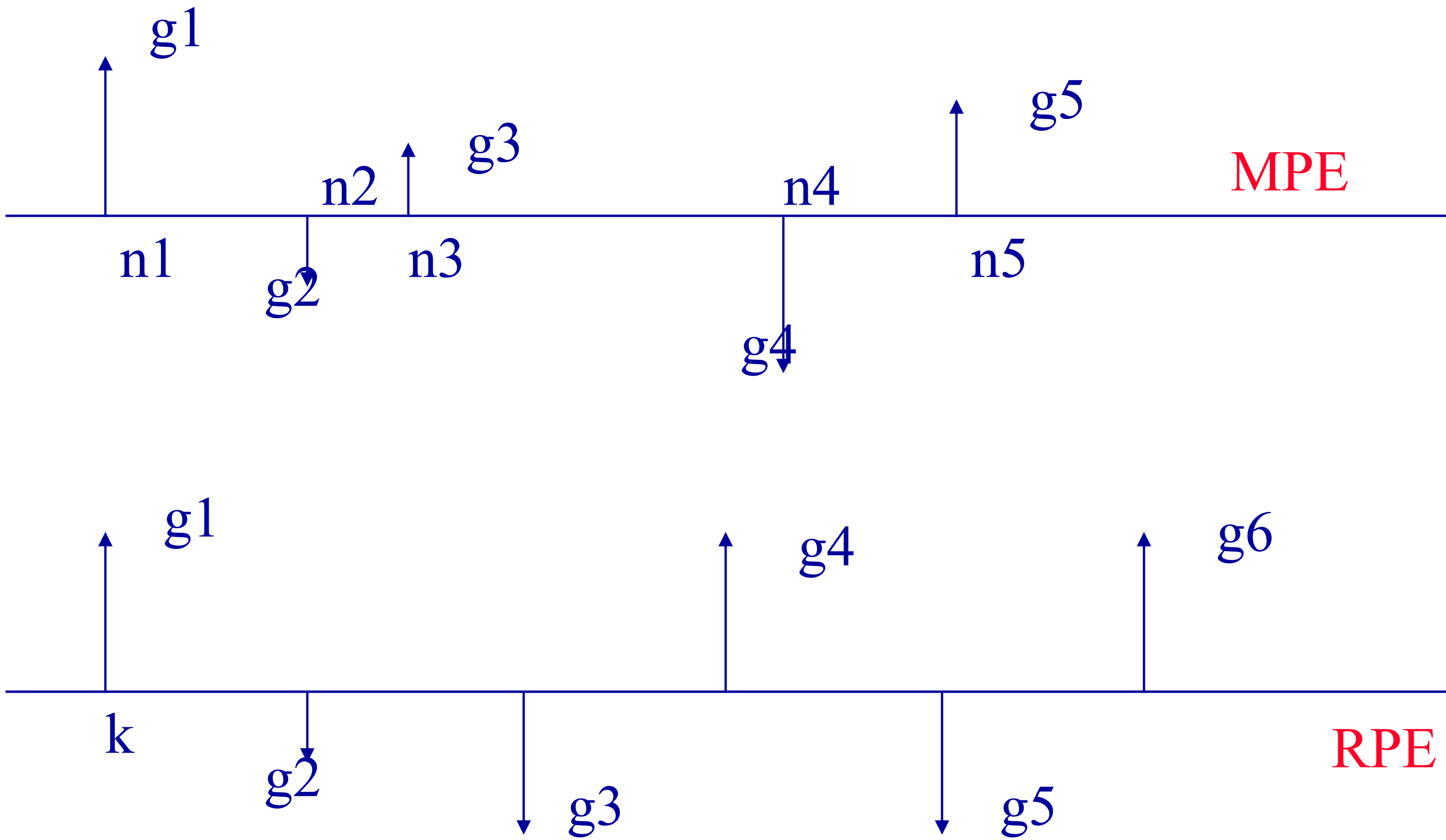
MPE-LPC

- The first A-by-S scheme (Multi Pulse Excitation)
- Excitation is specified by a small set of pulses with different amplitudes located at non-uniformly spaced intervals
- The only a-priory decision to make: The number of pulses required per block (usually about 1 per 10 samples, or 8 pulses for 10mSec)
- Locations and amplitudes should be sent

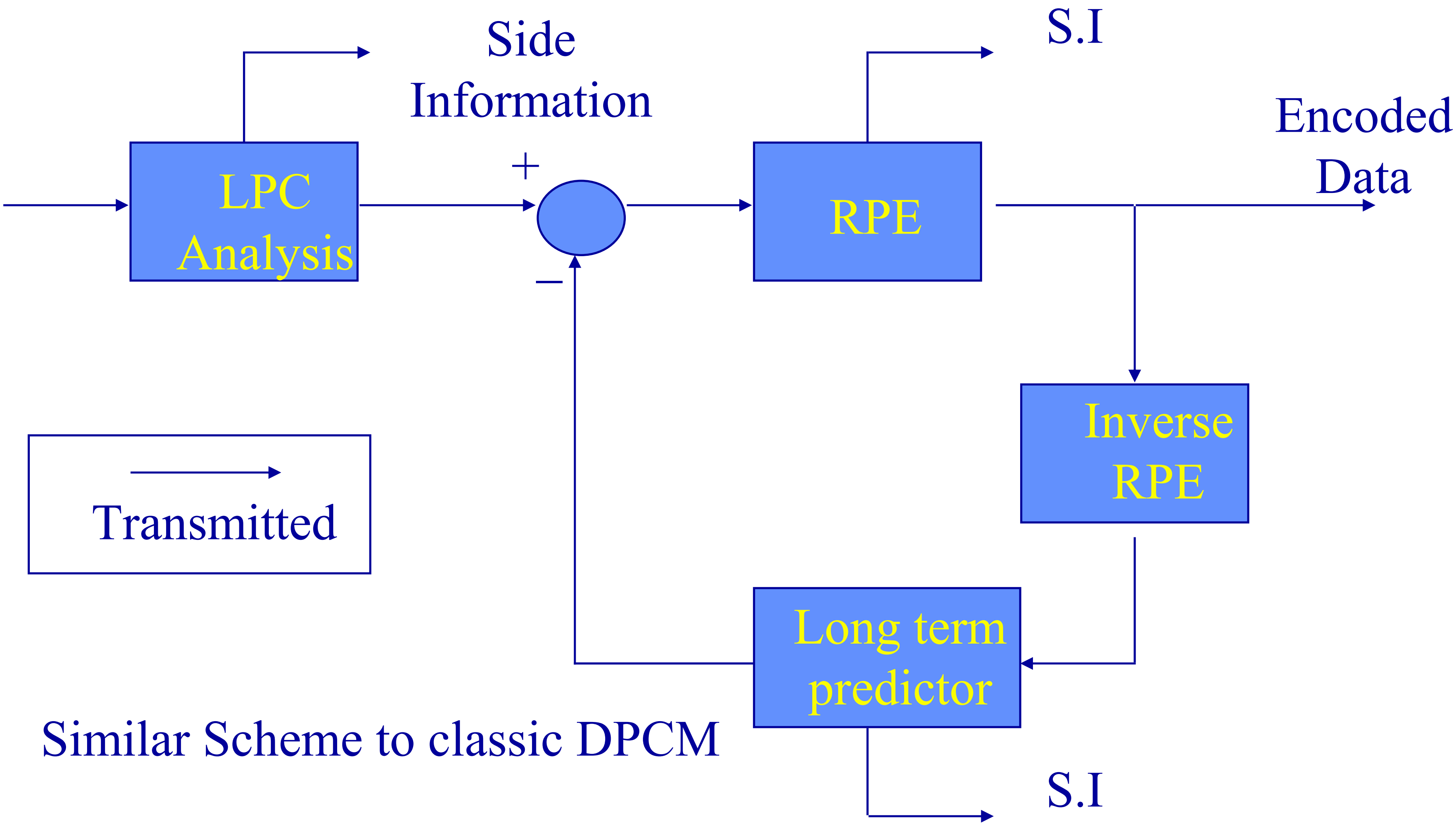
RPE-LPC

- regular Pulse Excitation LPC
- Pulses are equally spaced, and positions are specified by the first pulse position
- This prevents the need for position information
- Used in GSM

MPE Vs. RPE

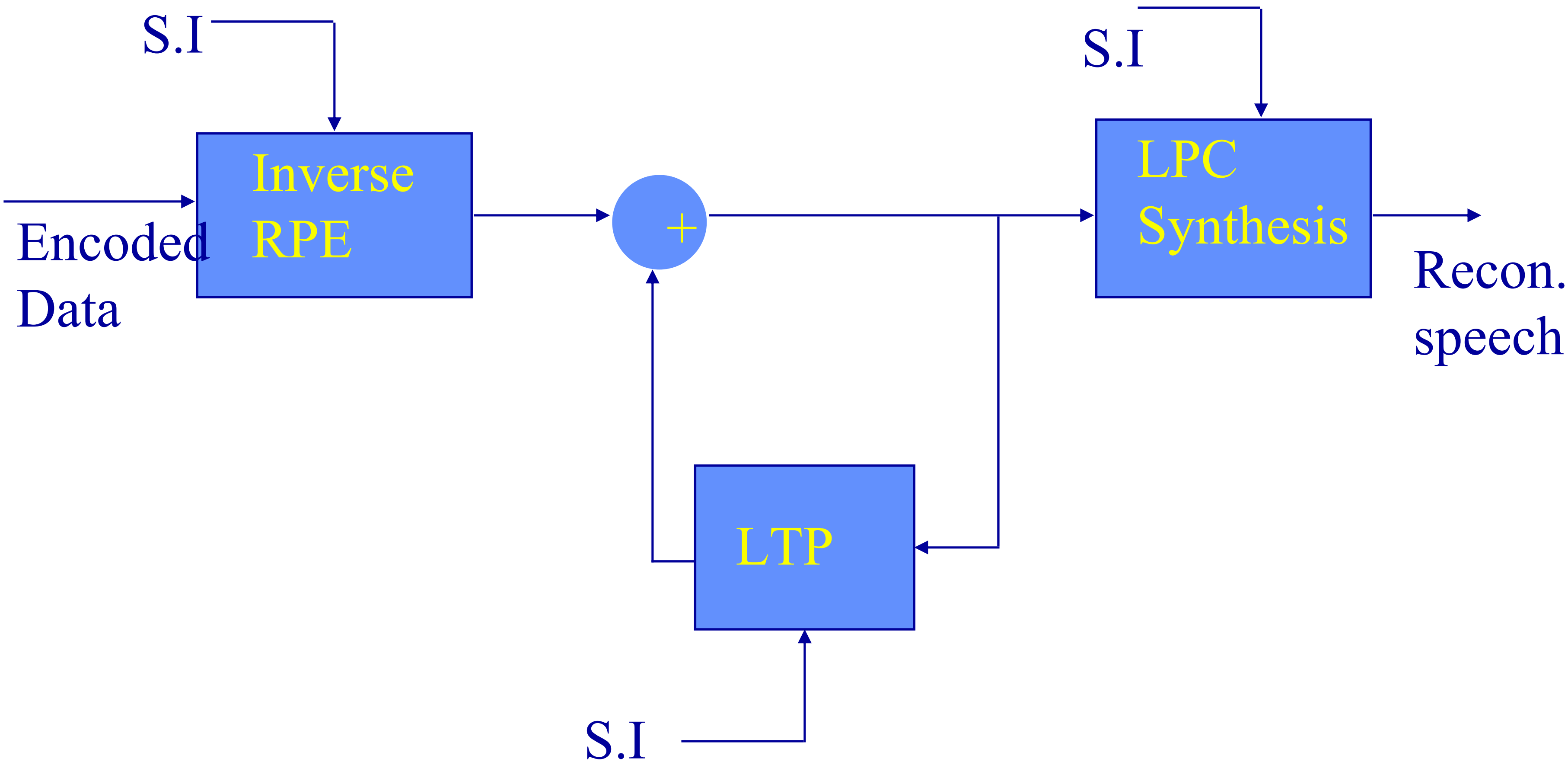


GSM Encoder



Similar Scheme to classic DPCM

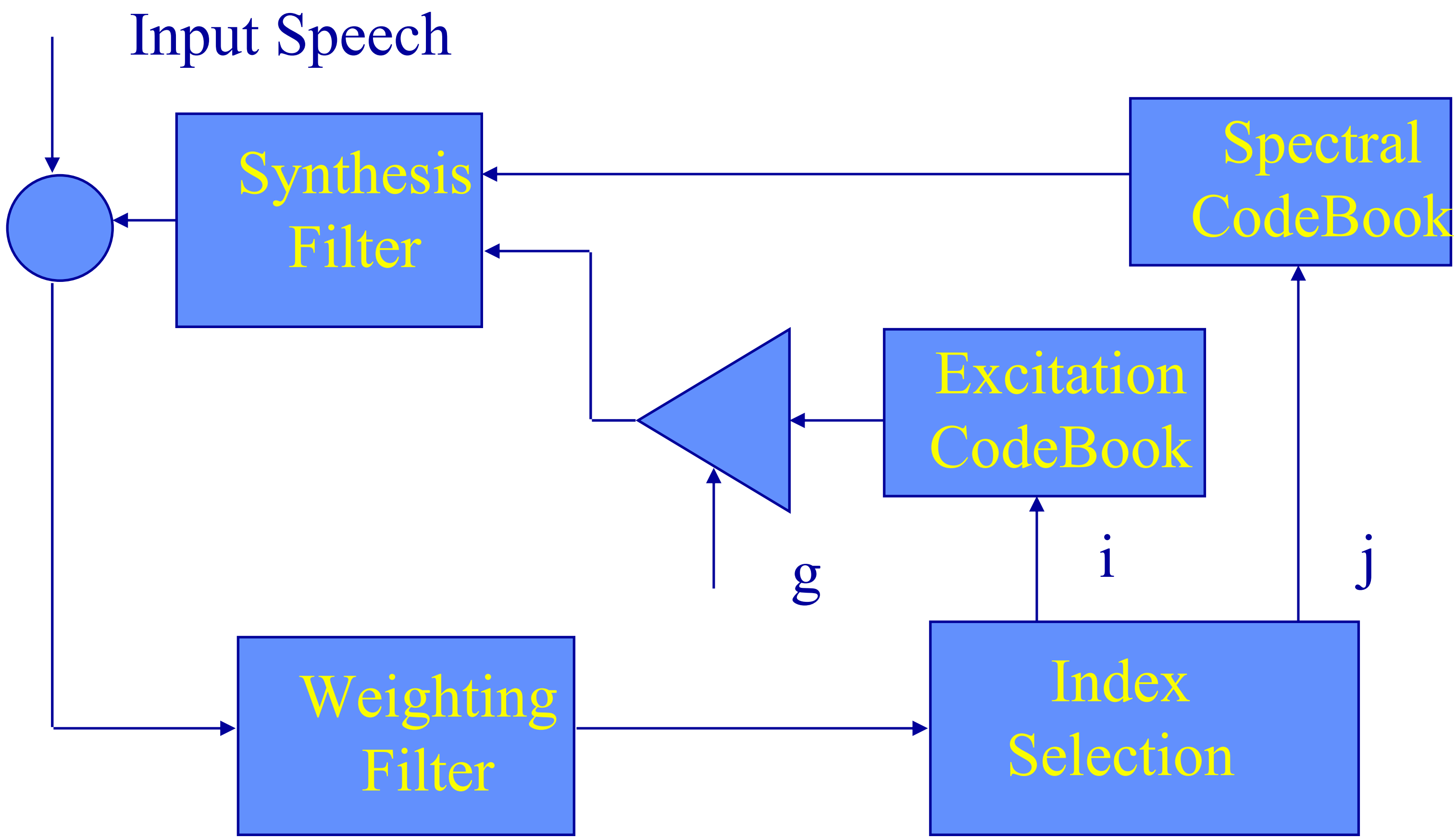
GSM Decoder


S.I
S.I
Inverse RPE
LPC Synthesis
+
Encoded Data
Recon. speech
LTP
S.I

Performance and Complexity (after Spanias,1994)

Technique	Rate(bps)	MOS	MIPS
PCM (G.711)	64K	4.3-4.4	0.01
ADPCM (G.726)	16K-40K	4.1-4.2	~2
LD-CELP (G.728)	16K	4.0-4.2	~19
RPE-LTP (GSM)	13k	3.47	6
VSELP (IS-54)	8k	3.45	13.5
CELP (DoD1016)	4.8k	3.2	16
LPC-10e (DoD1015)	2.4k	2.3	7
CS-ACELP (G.729/729a)	8k	4.2	17 / 10
CS-ACELP (G.723)	5.3k/6.3k	3.5/4.0	14.6 / 16
(Conjugate-Structure Algebraic CELP)			
MELP	2.4Kbps	2.9*	20

* Better in Noisy environment !


Input Speech
Synthesis Filter
Spectral CodeBook
Excitation CodeBook
g
i
j
Weighting Filter
Index Selection

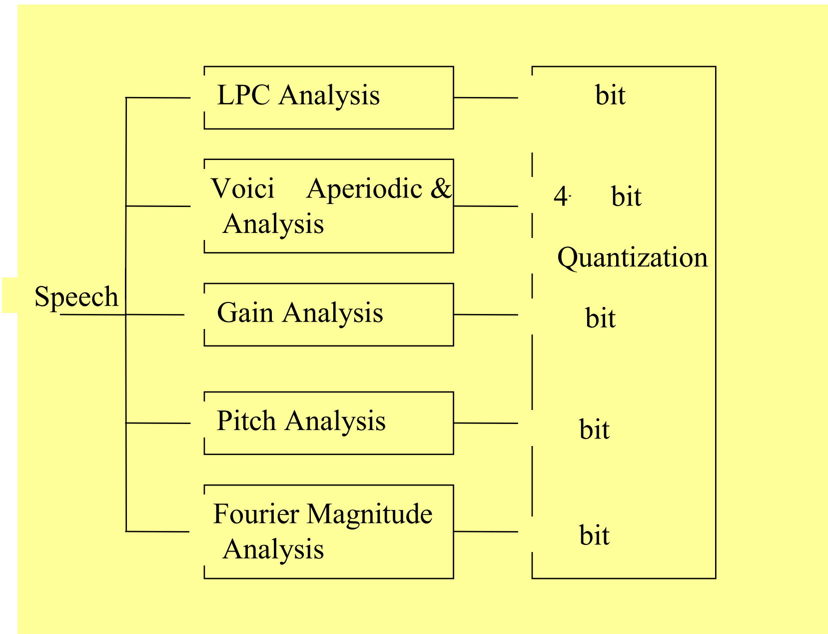
SELP

- Excitation signal is derived from the past history of the coded excitation function itself
- When the best vector is found, it is fed back with oldest samples discarded
- Its effectively a CELP coder with an adaptive codebook, containing no codebook...

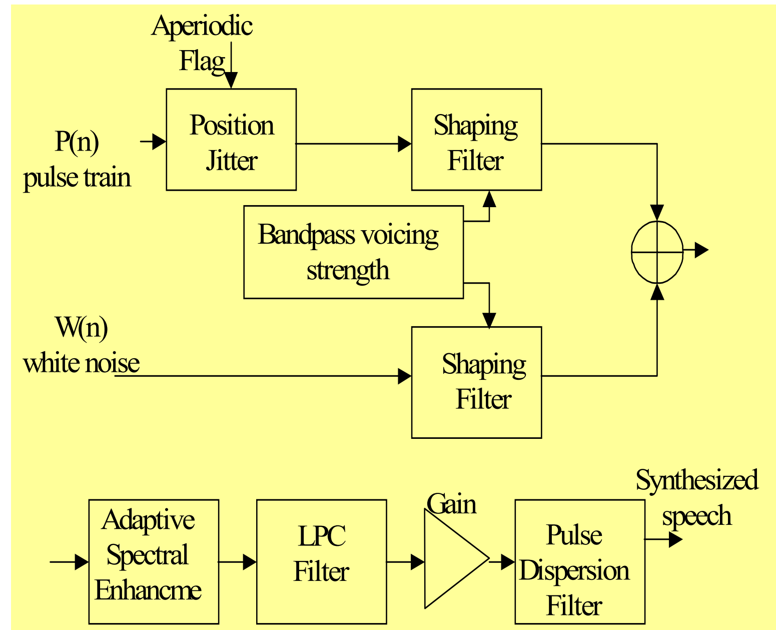
MELP Principles

- Mixed Excitation
- Aperiodic Pulses
- Adaptive Spectral Enhancement
- Pulse Dispersion
- Fourier Magnitudes

MELP Encoder/Decoder



Encoder



Decoder

MELP Coder Scheme