

Speech Coding Standards

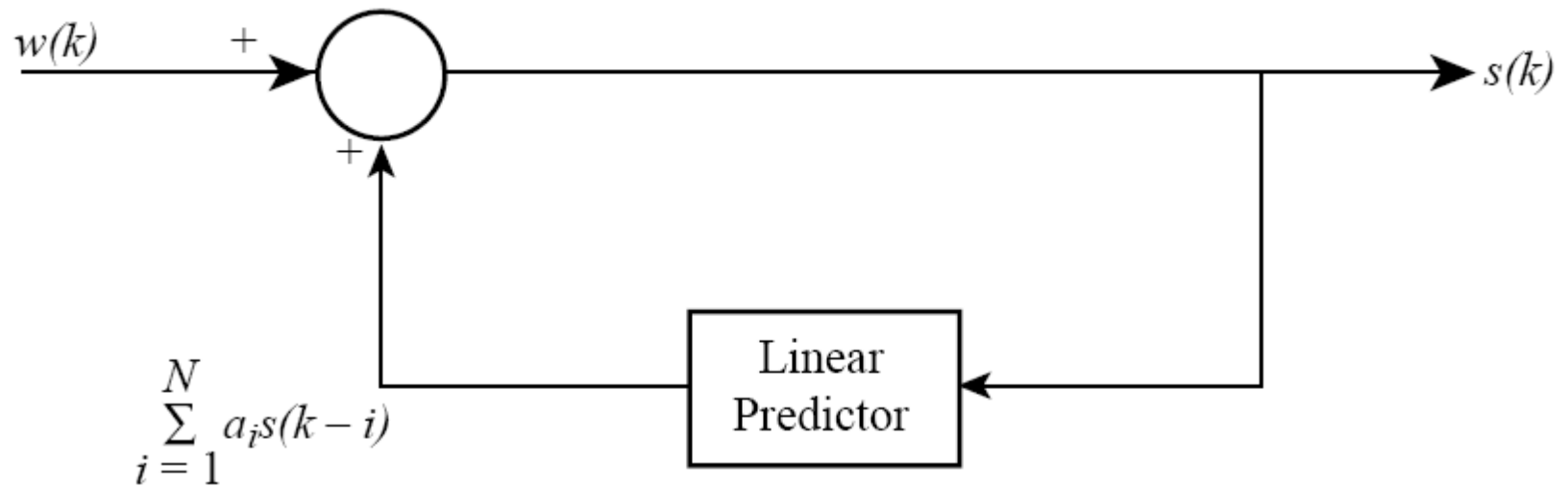
Jerry D. Gibson

ECE 241

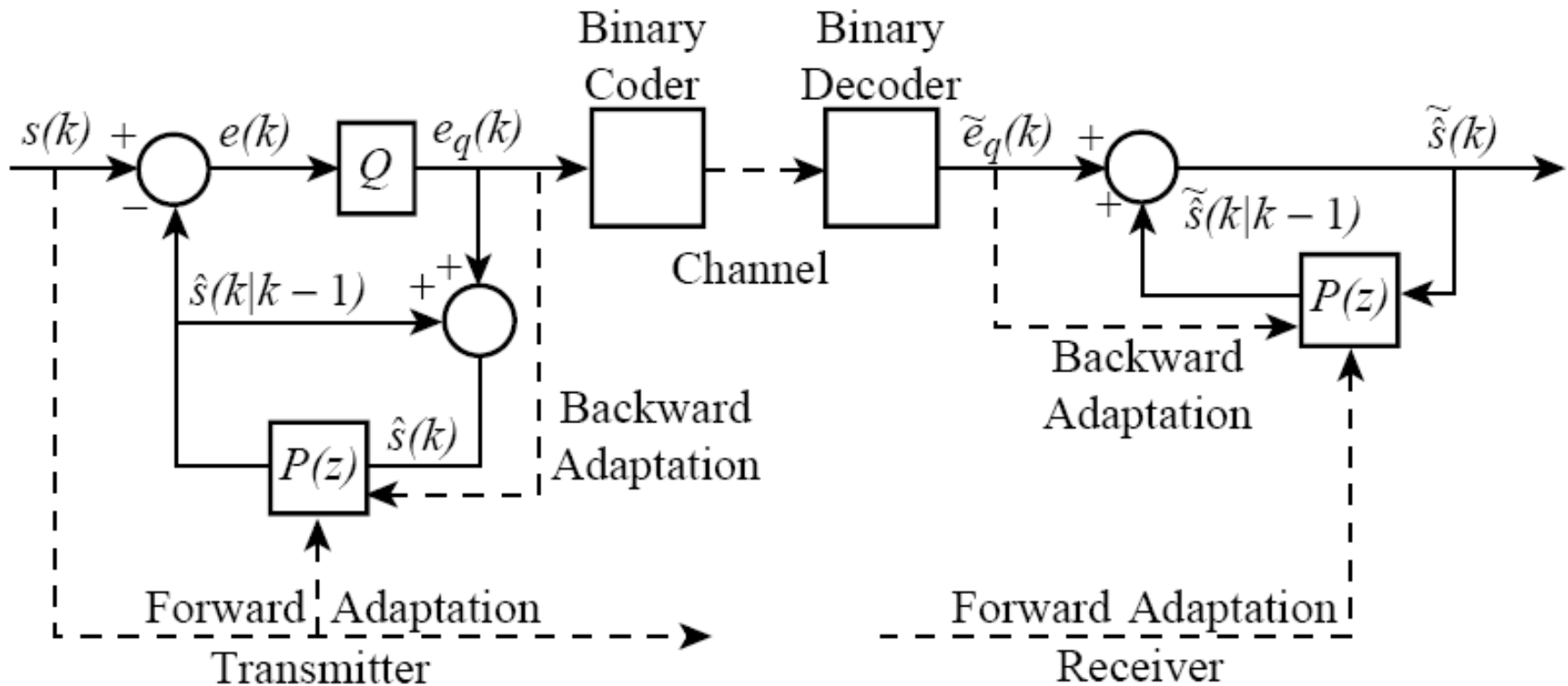
Speech Coding Standards

Year of Introduction	Bit Rates (kbps)	Description	MOS
1972	64	PCM (for PSTN)	4.4
1976	2.4	LPC-10 (U.S. Federal Standard 1015)	2.7
1984	32	G.721 ADPCM (for PSTN)	4.1
1990	4.15	INMARSAT (satellite)	≈3.2
1991	13	GSM (European cellular)	3.6
1991	4.8	CELP (U.S. Federal Standard 1016)	3.2
1992	16	G.728 (low-delay CELP)	4.0
1992	8	VSELP (NA cellular)	3.5
1993	1–8	QCELP (NA 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 and H.324)	3.98
1995	5–6	Half-rate GSM	≈3.4
1996	2.4	New low-rate U.S. Federal Standard	≈3.3

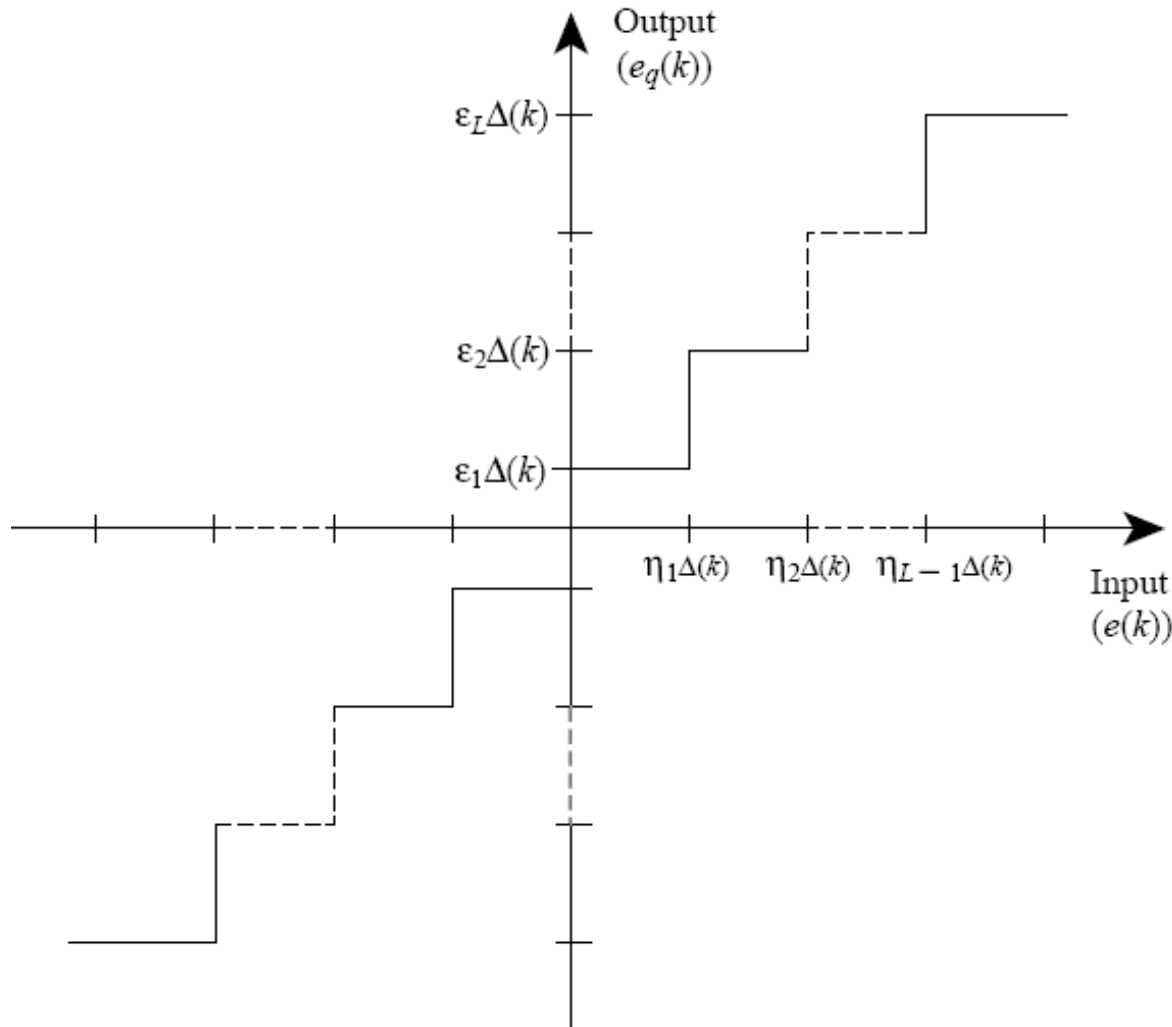
Linear Prediction Model



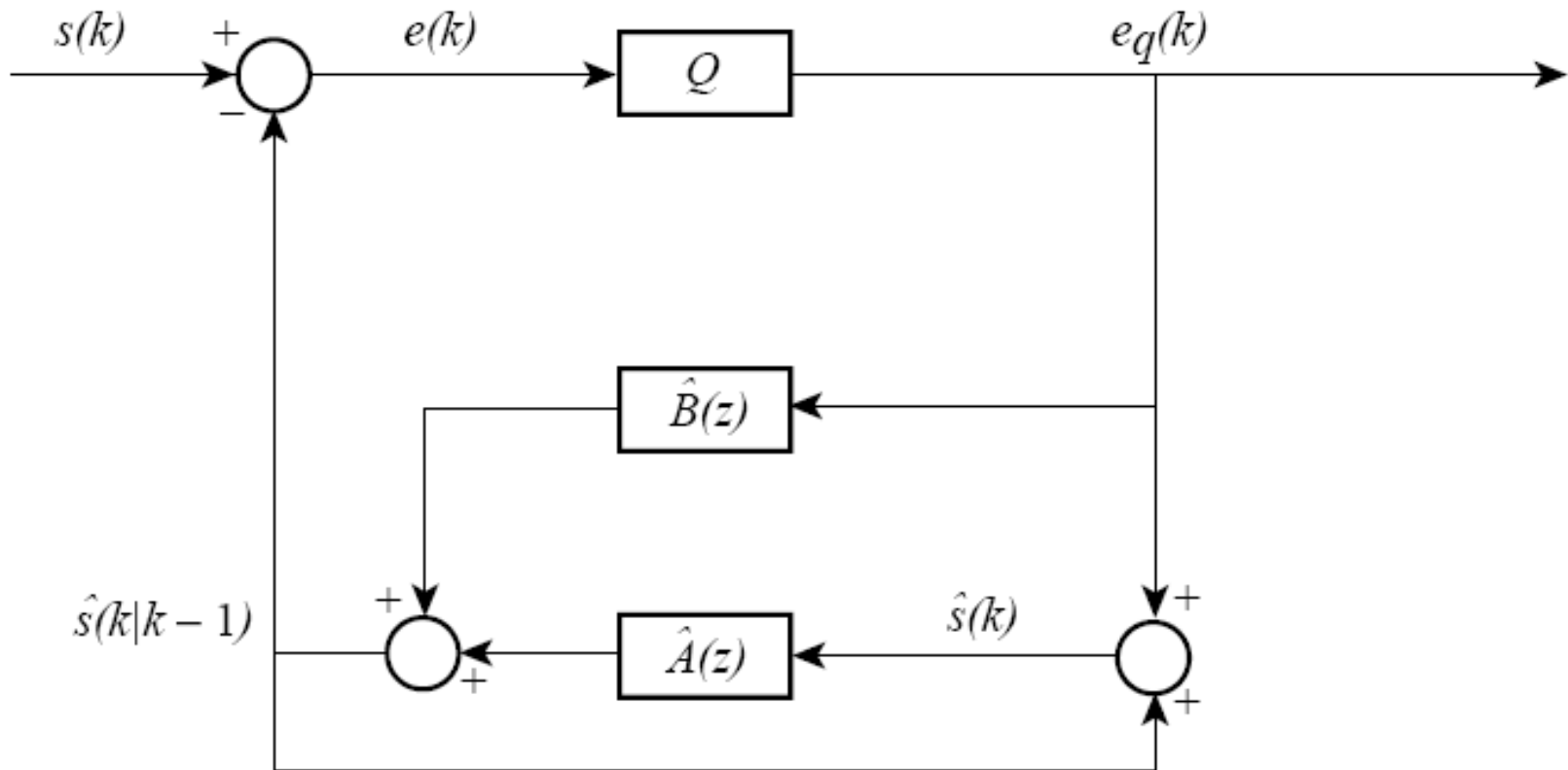
Differential Pulse Code Modulation



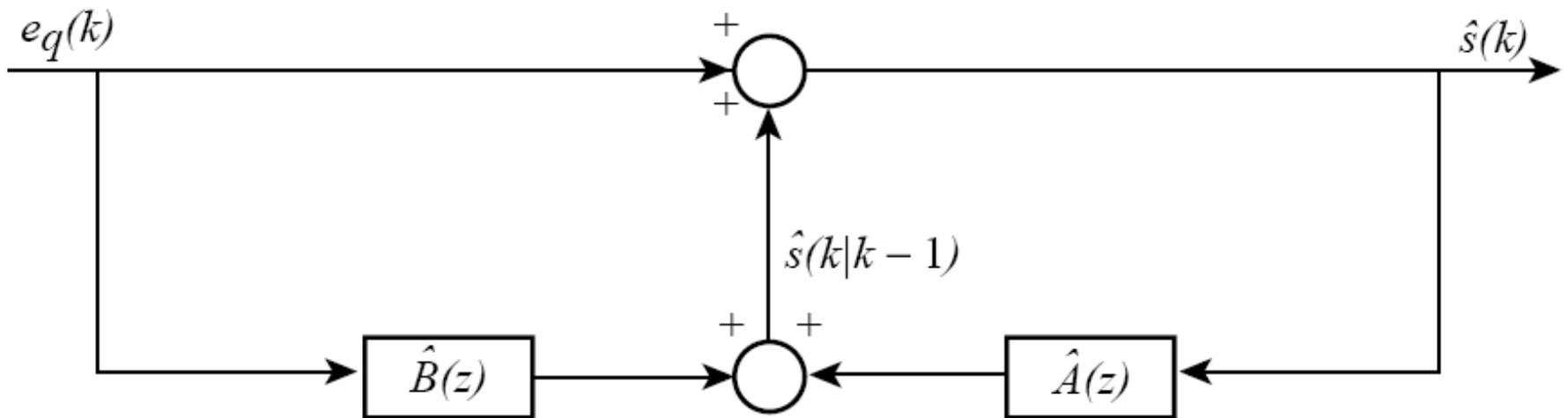
Symmetric Quantizer with Adaptive Step Size



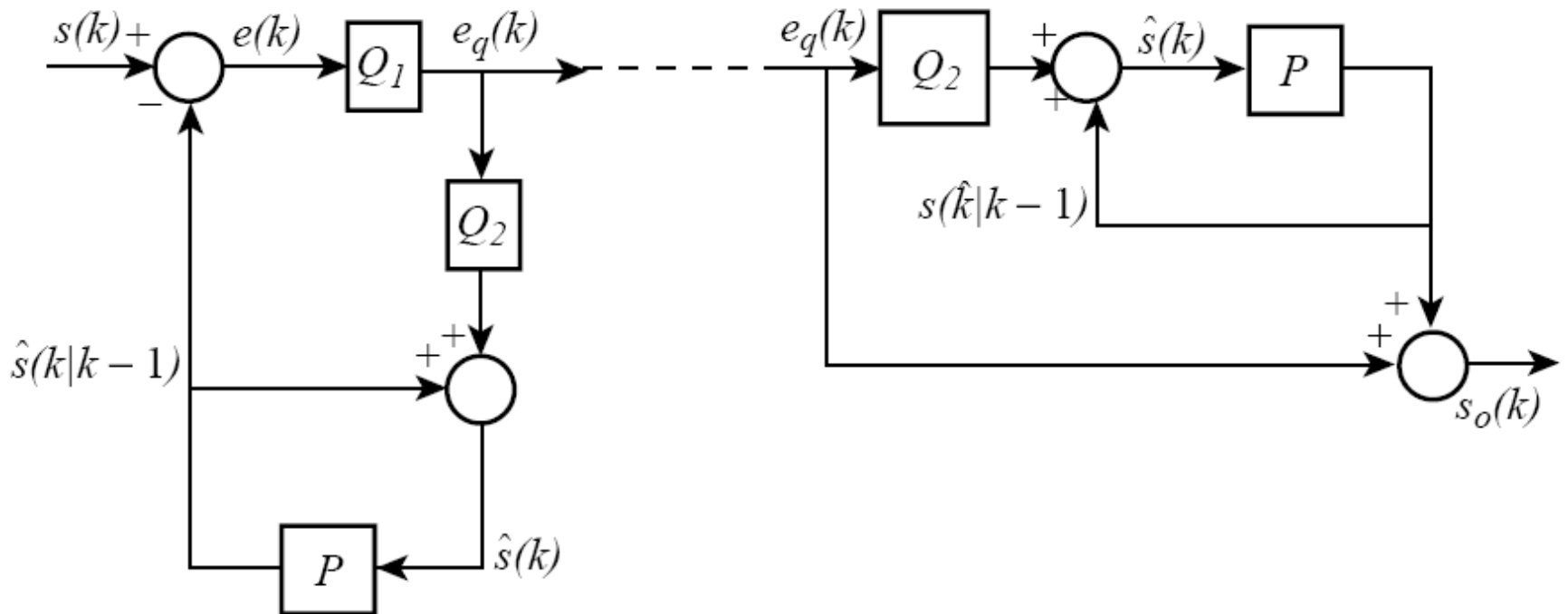
Differential Encoder with Pole-Zero Predictor



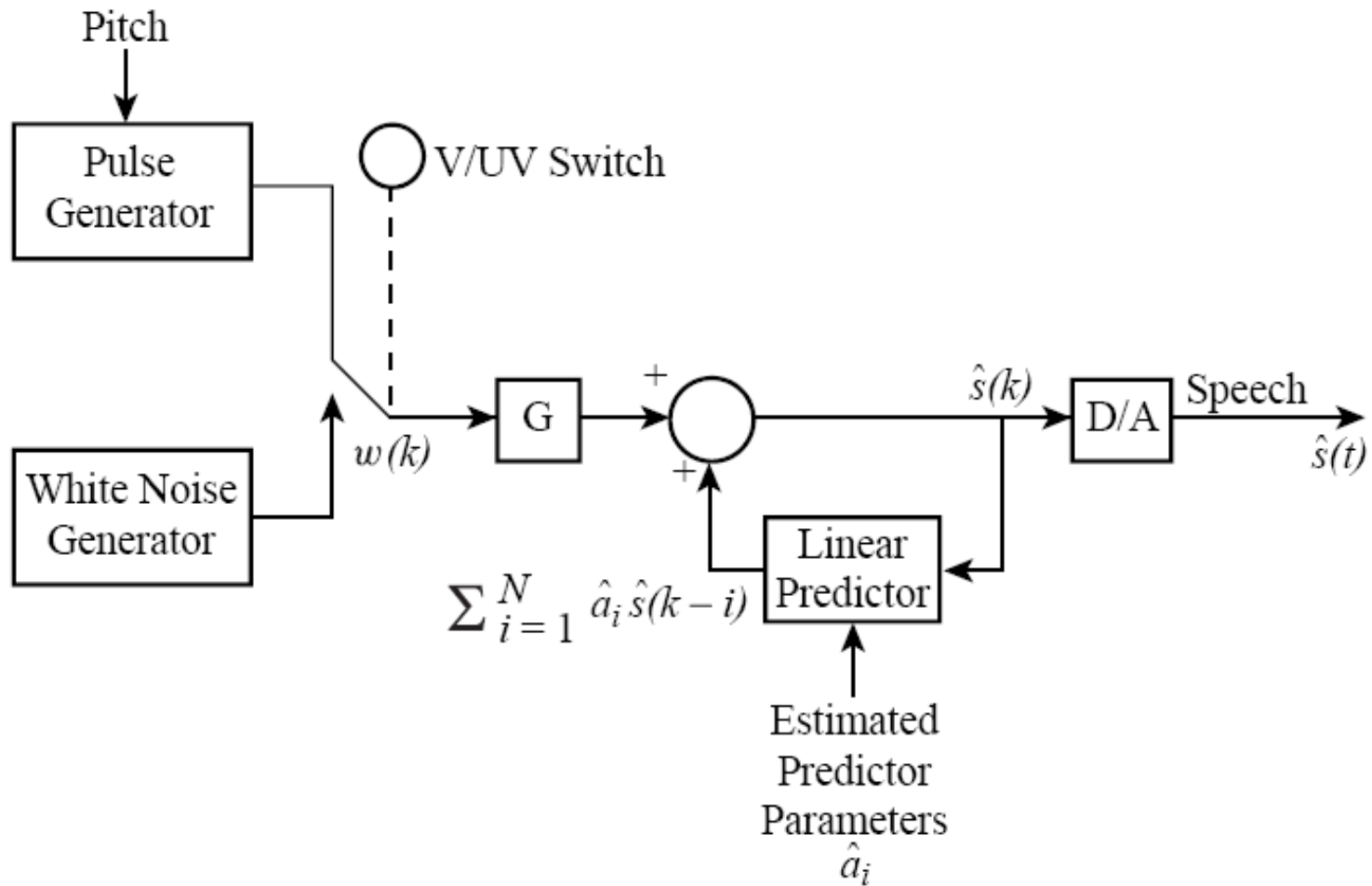
Decoder with Pole-Zero Predictor



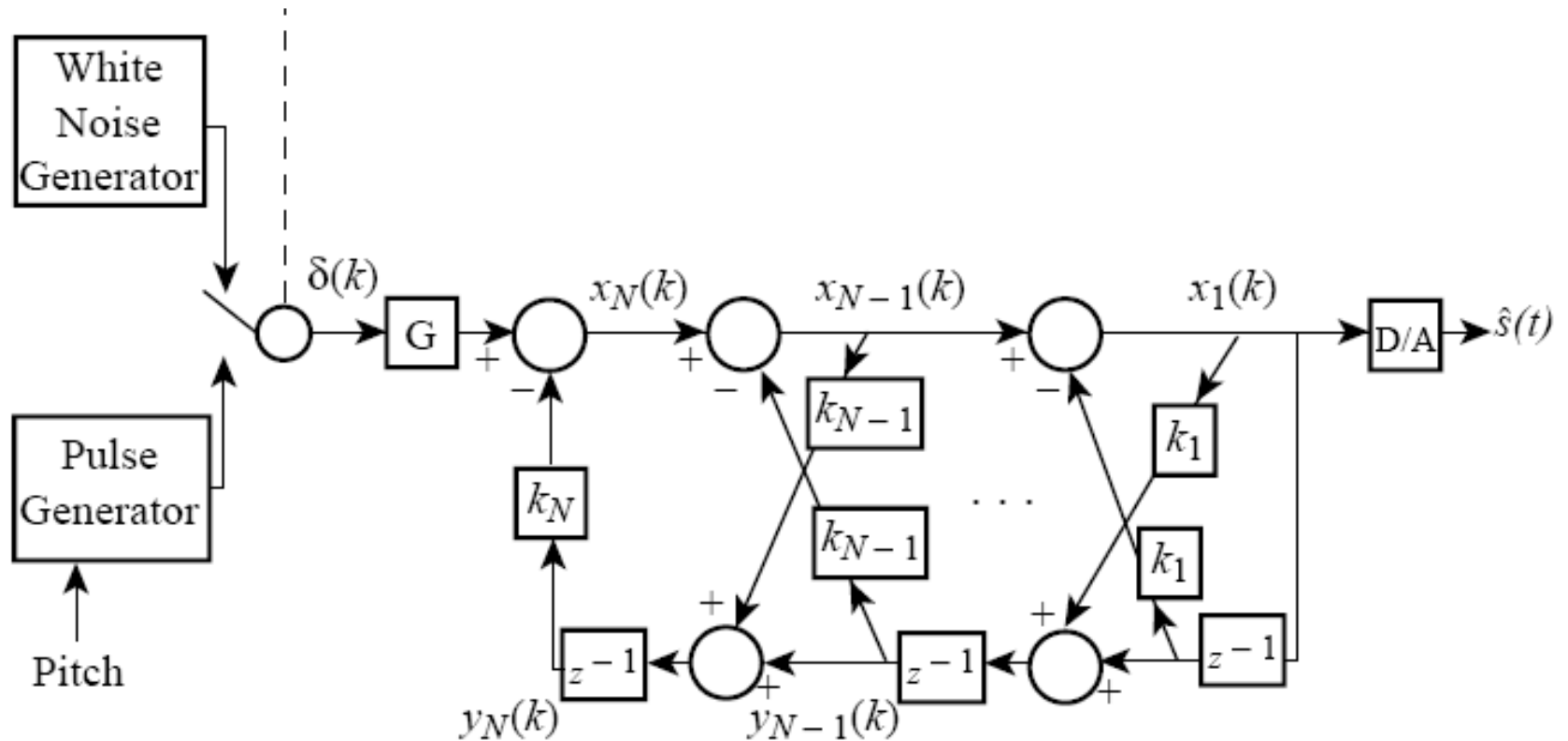
Embedded DPCM



Linear Predictive Coder



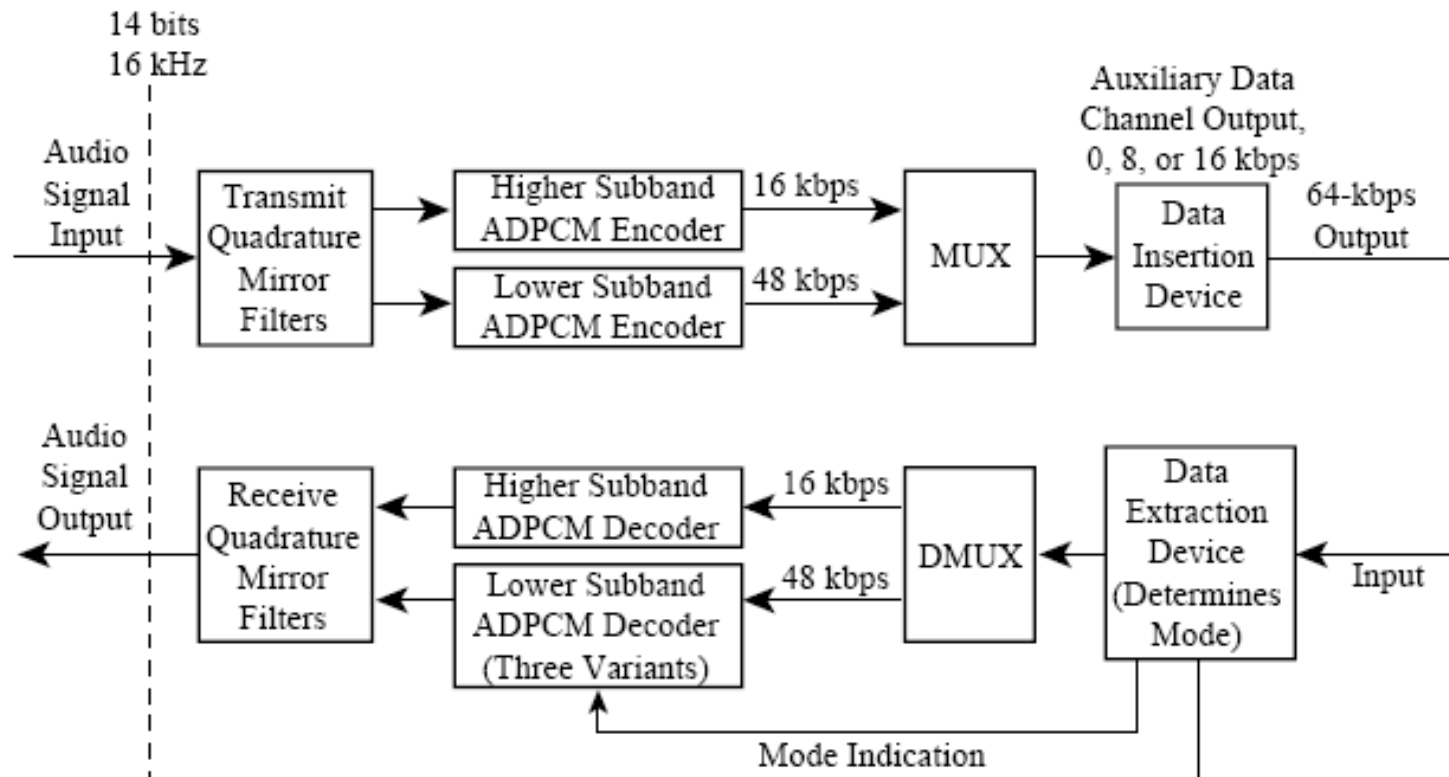
LPC Lattice Structure



Common Audio Bandwidths and Sampling Rates

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

Block Diagram of G.722



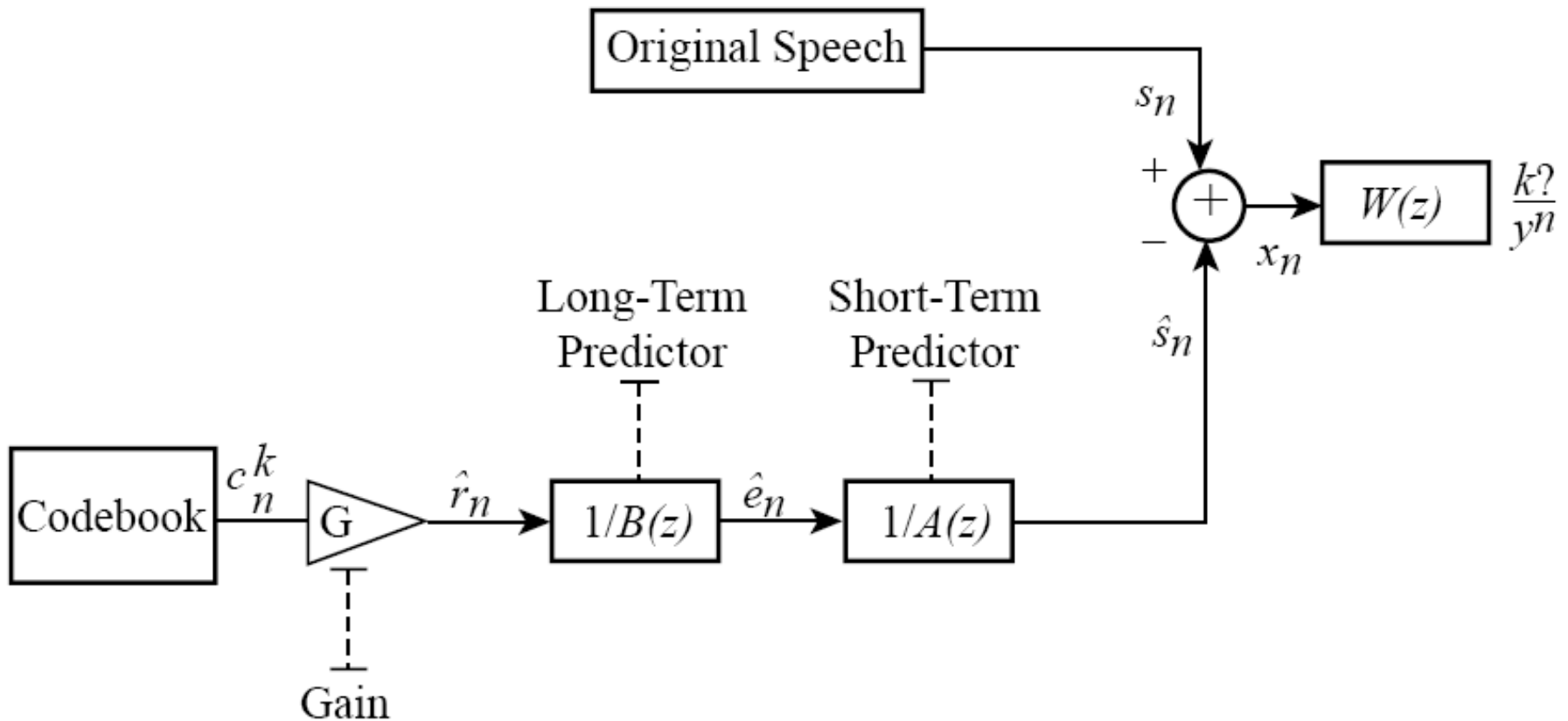
MOS Values for G.722

Bit Rate	Male	Female	Mean
48 kbps	3.7	3.7	3.7
56 kbps	4.3	3.7	4.0
64 kbps	4.0	4.1	4.1

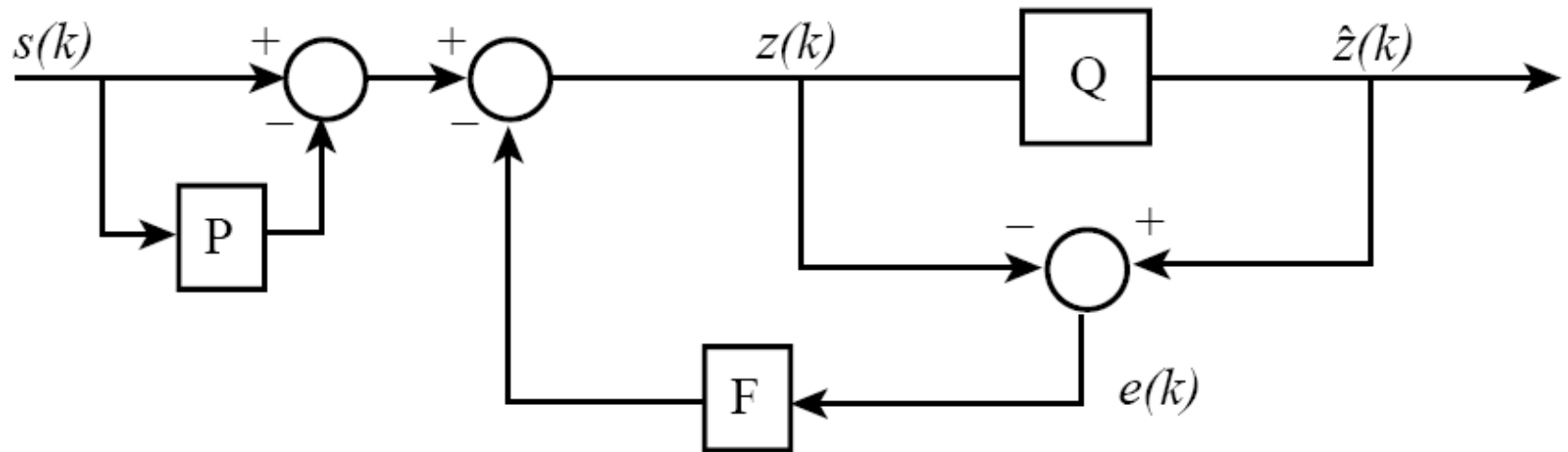
G.722 MOS versus BER

CCITT Wideband Speech Coder	BER=0	BER=10 ⁻⁴	BER=10 ⁻³
Source signal	4.3		
64 kbps	4.3	3.8	3.0
48 kbps	3.8	3.6	3.0

Codebook Search in CELP



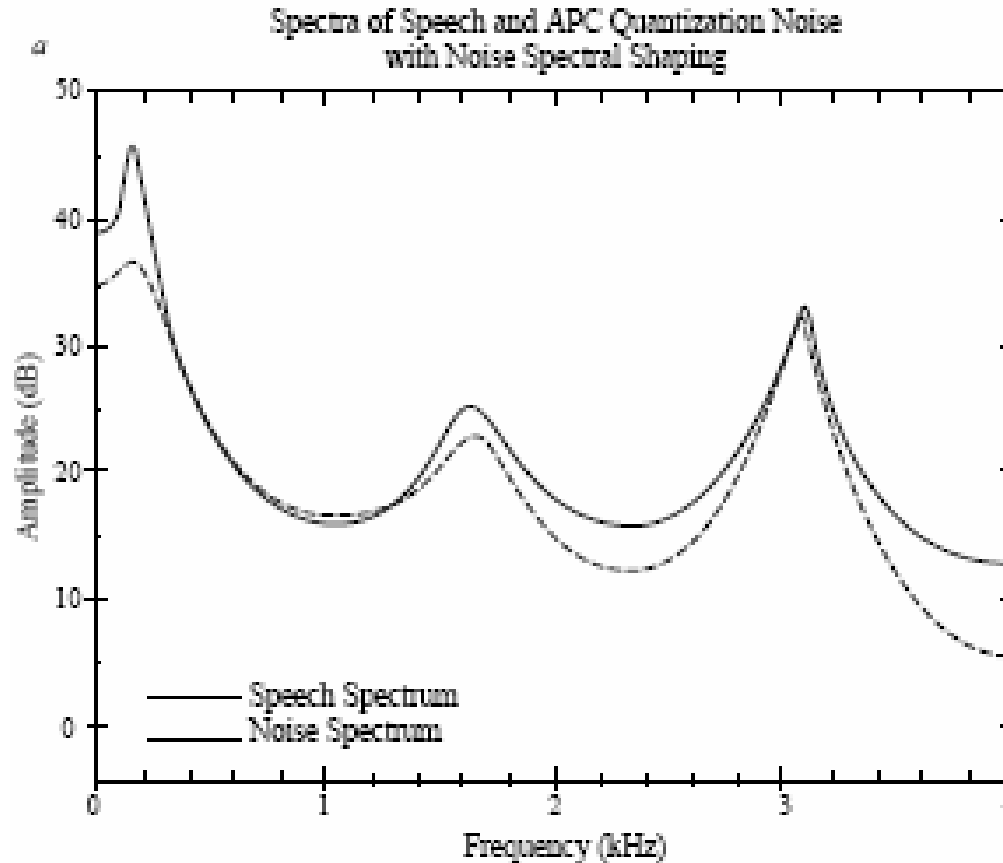
Noise Spectral Shaping



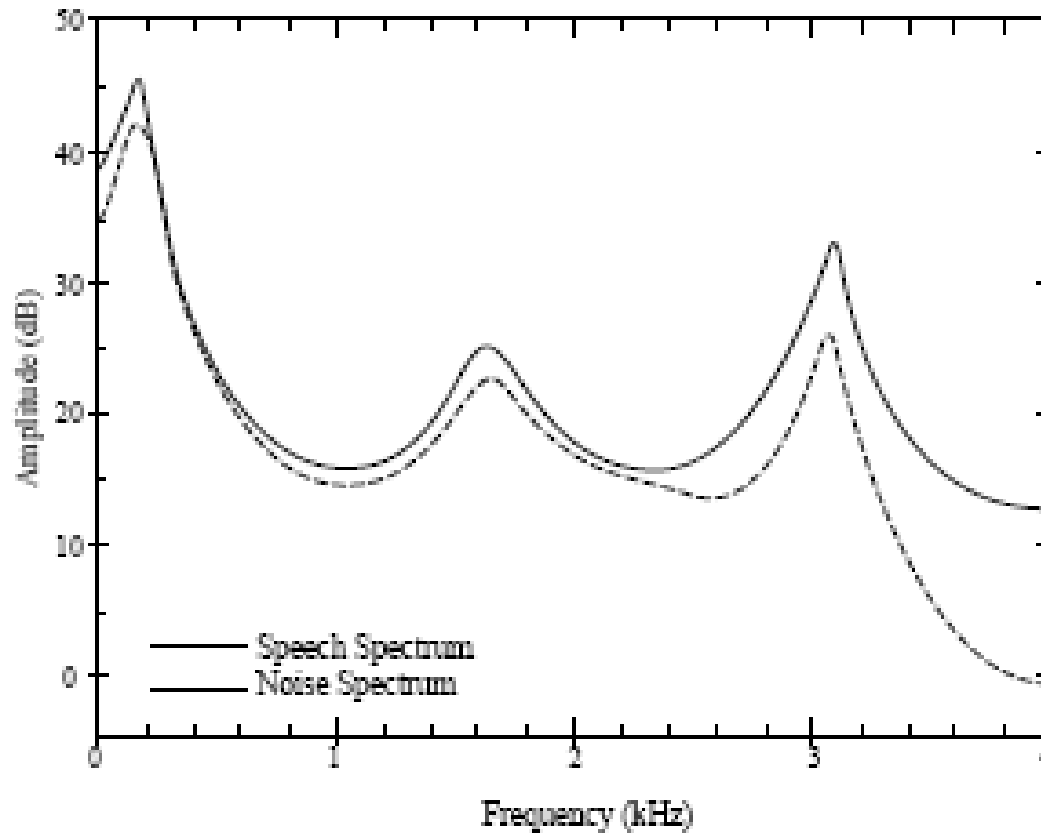
Noise Shaping Filter

$$F(z) = P(\alpha^{-1}z) = \sum_{i=1}^N \alpha^i a_i z^{-i}$$

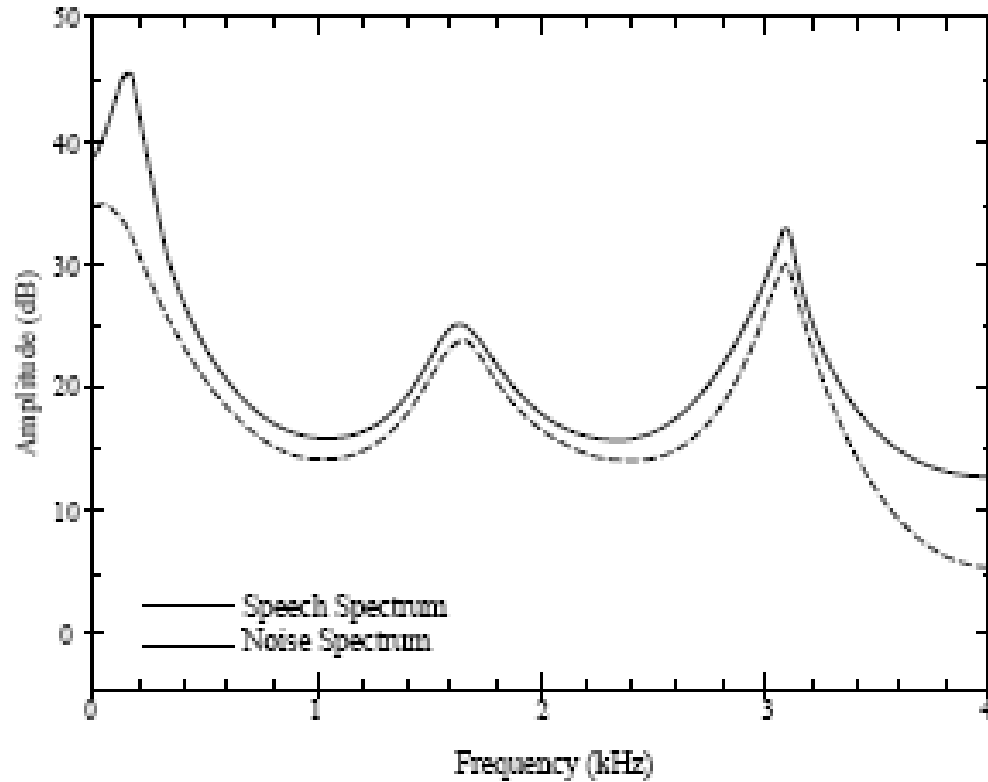
No Noise Spectral Shaping



Maximum Noise Spectral Shaping



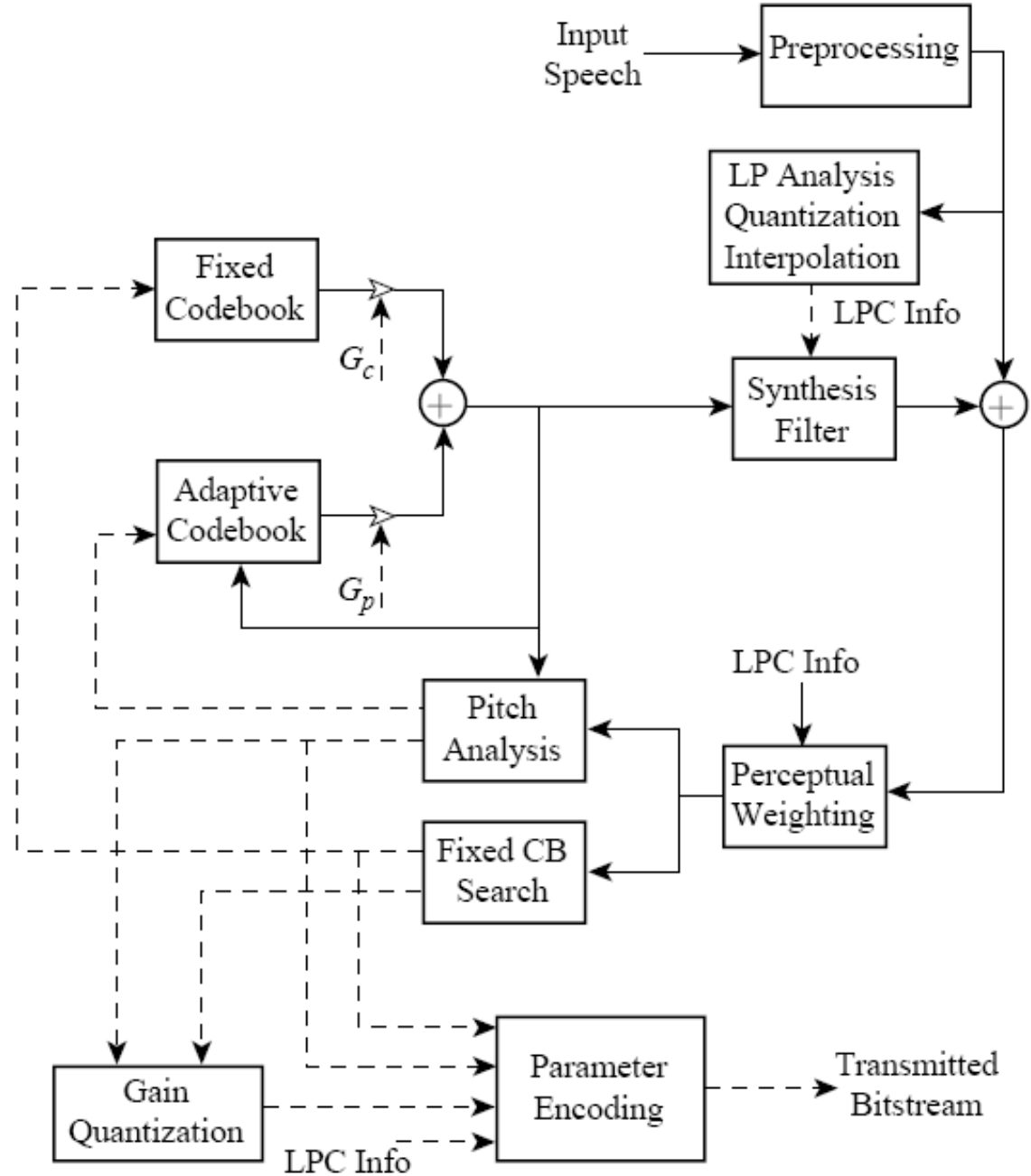
Perceptually Best Spectral Shaping



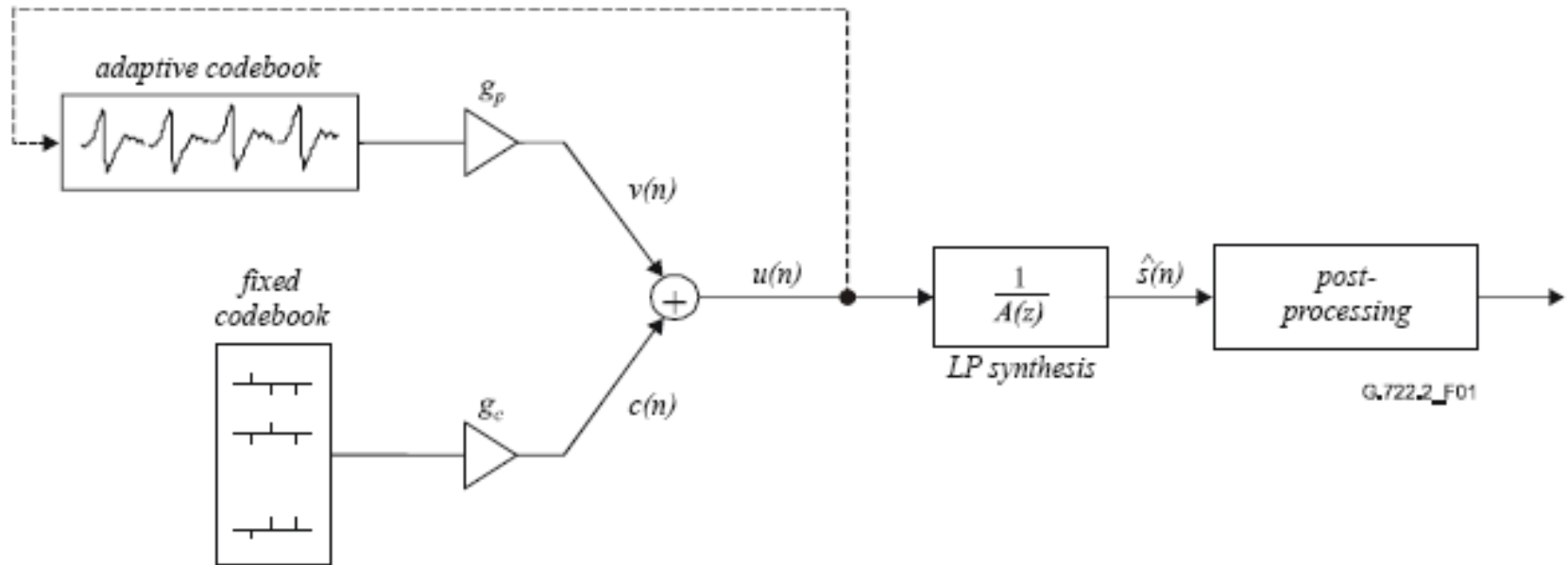
Perceptual Weighting

$$H(z) = W(z) = \frac{1 - \sum_{i=1}^N \beta^i a_i z^{-i}}{1 - \sum_{i=1}^N \alpha^i a_i z^{-i}}$$

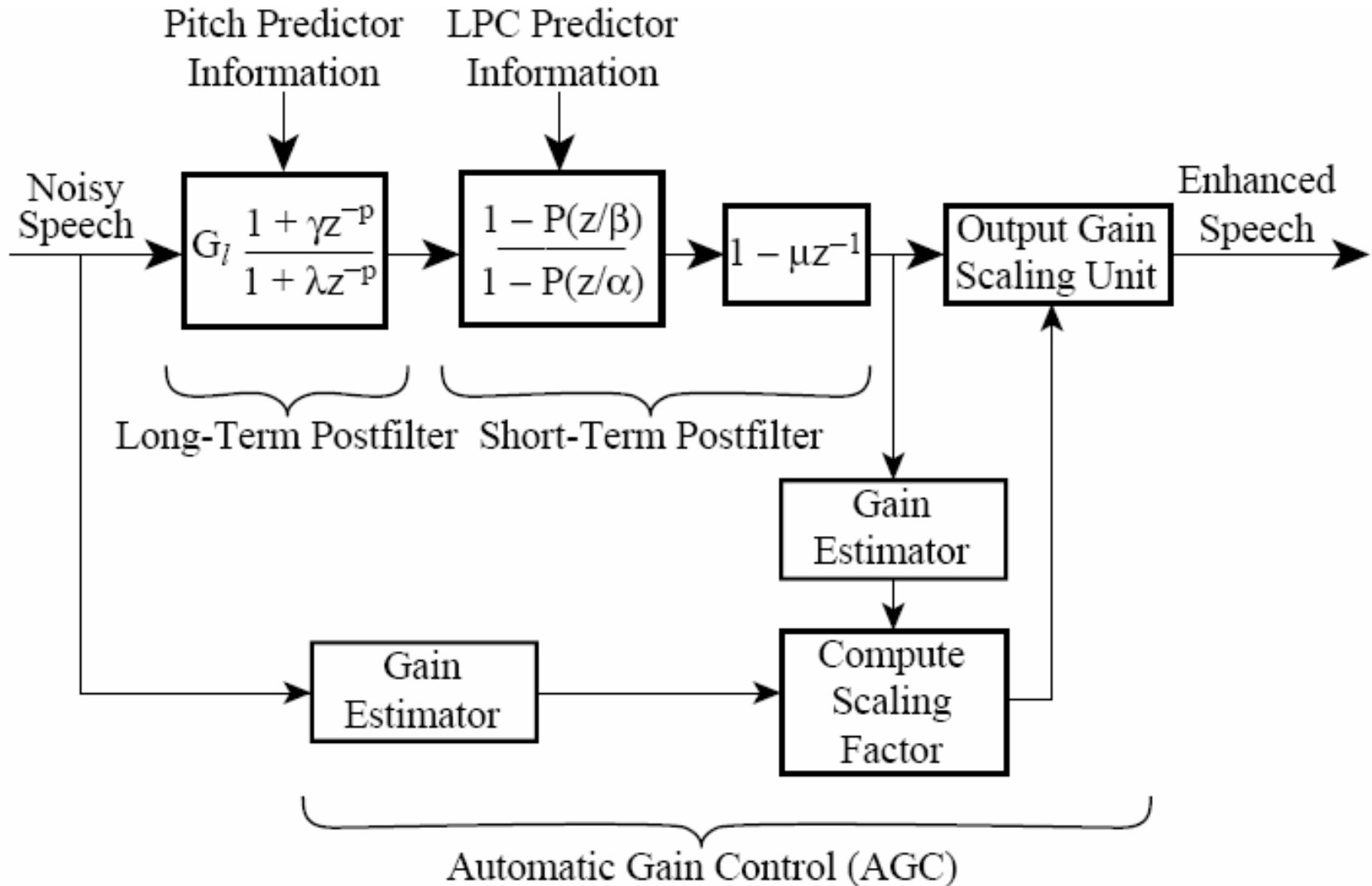
CELP Analysis



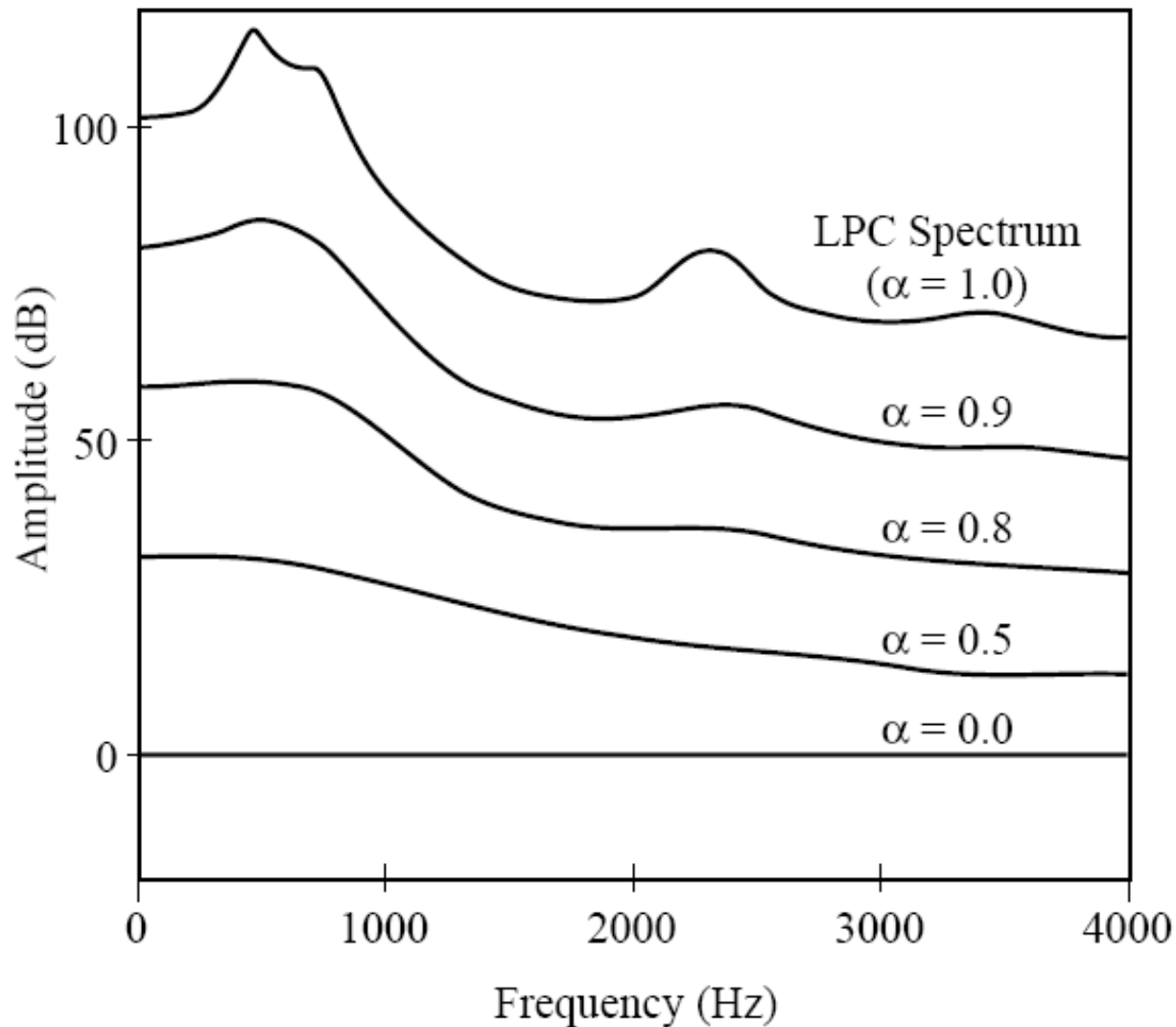
CELP Synthesis Model



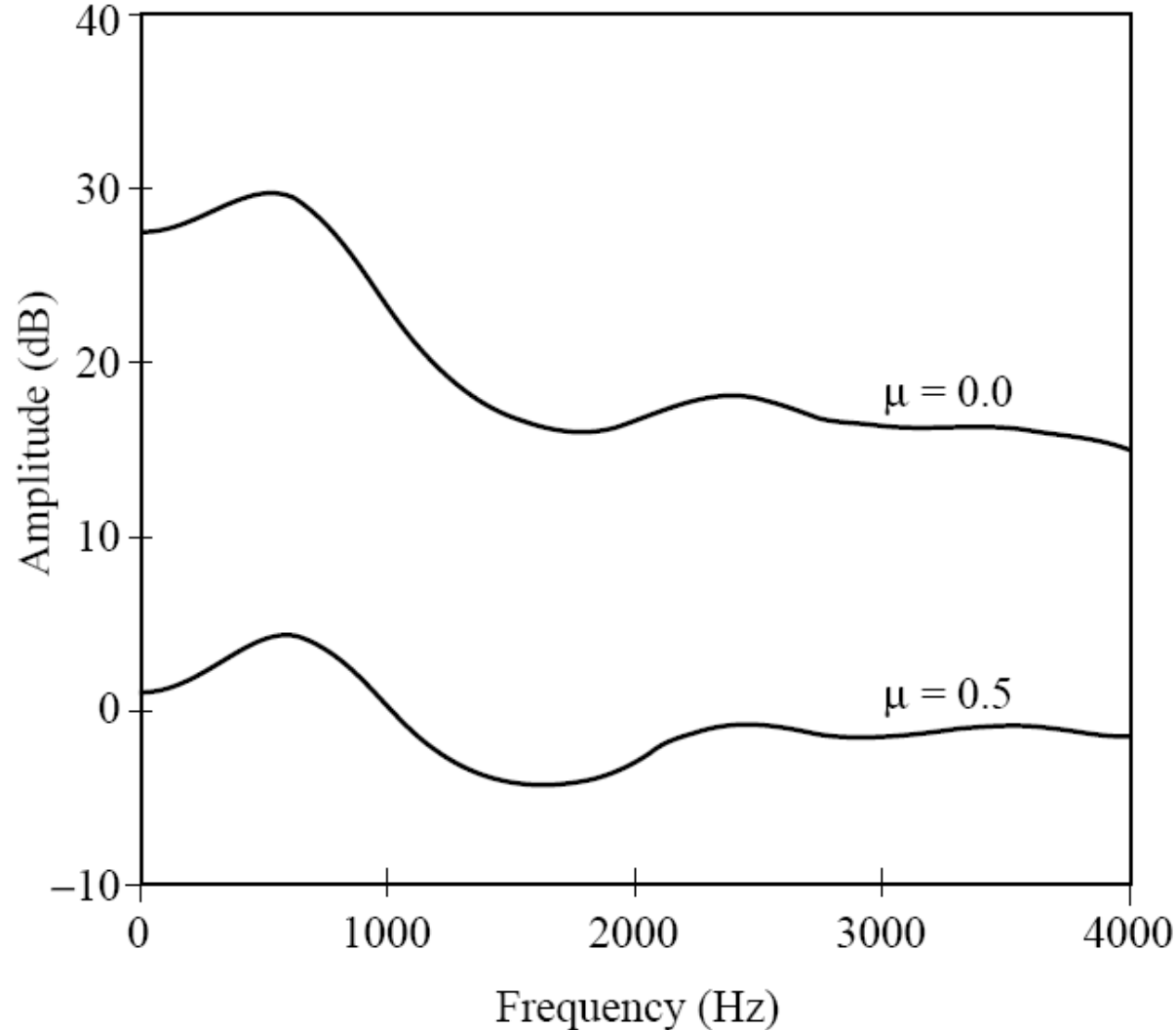
General Form of a Postfilter



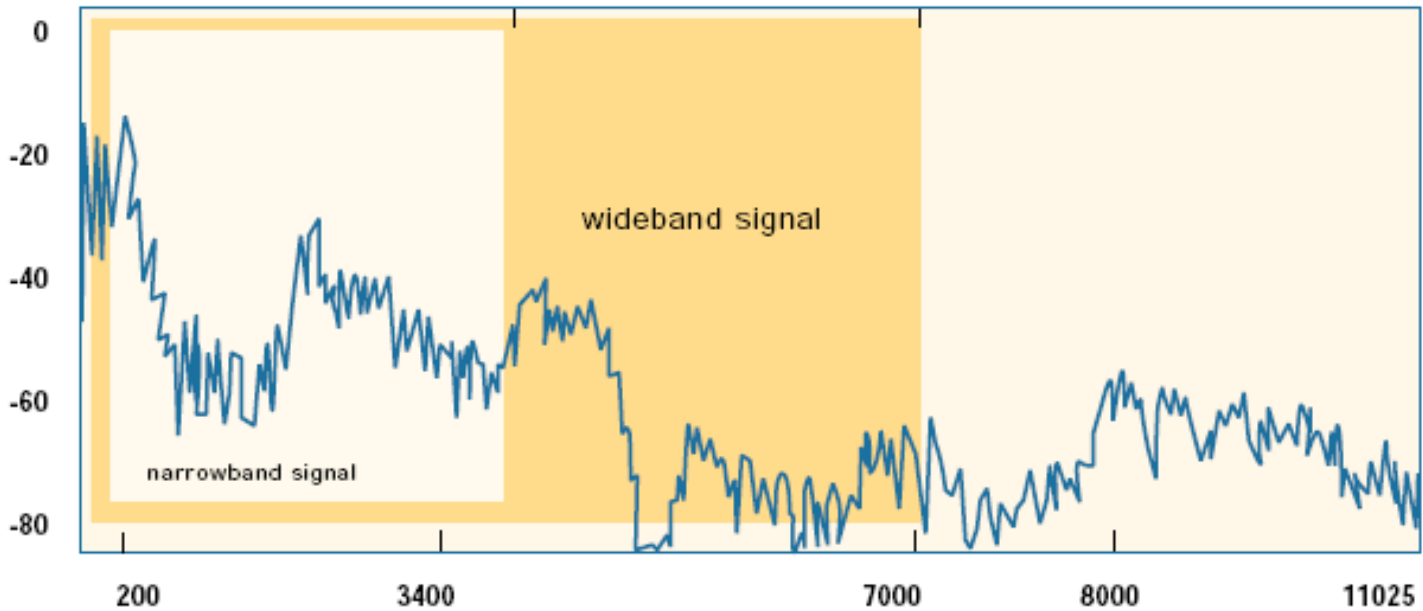
All Pole Postfilter Responses



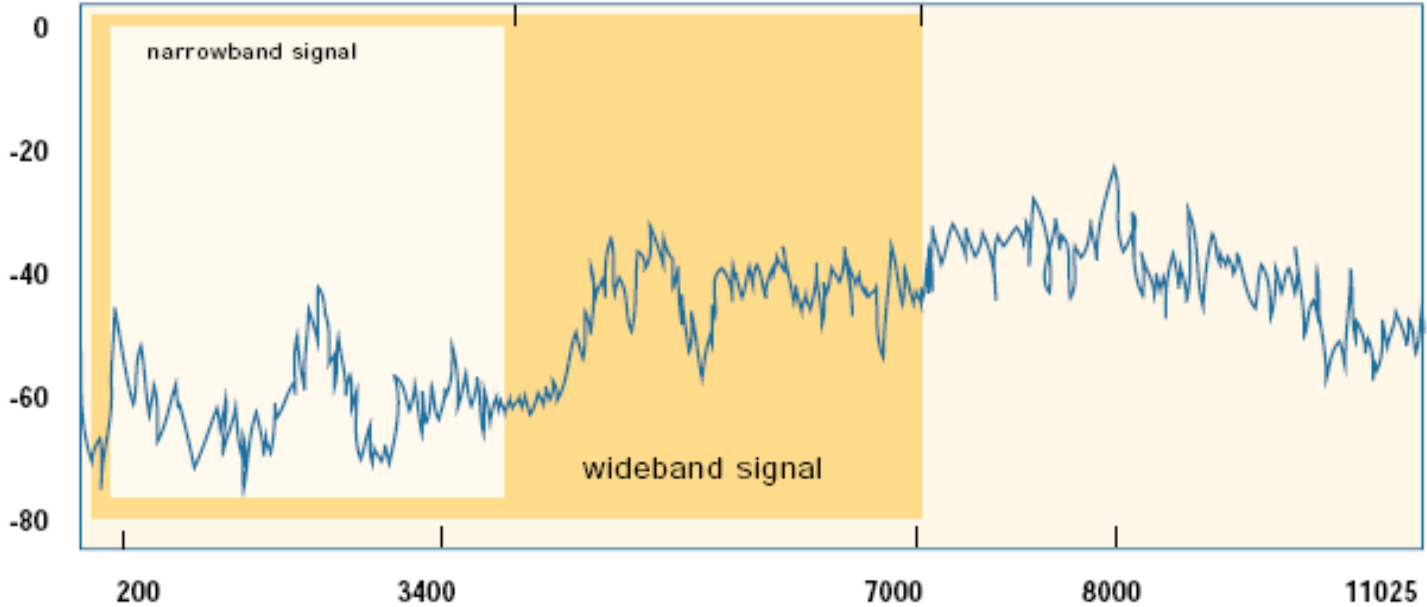
Pole-Zero Postfilter with Spectral Tilt Correction



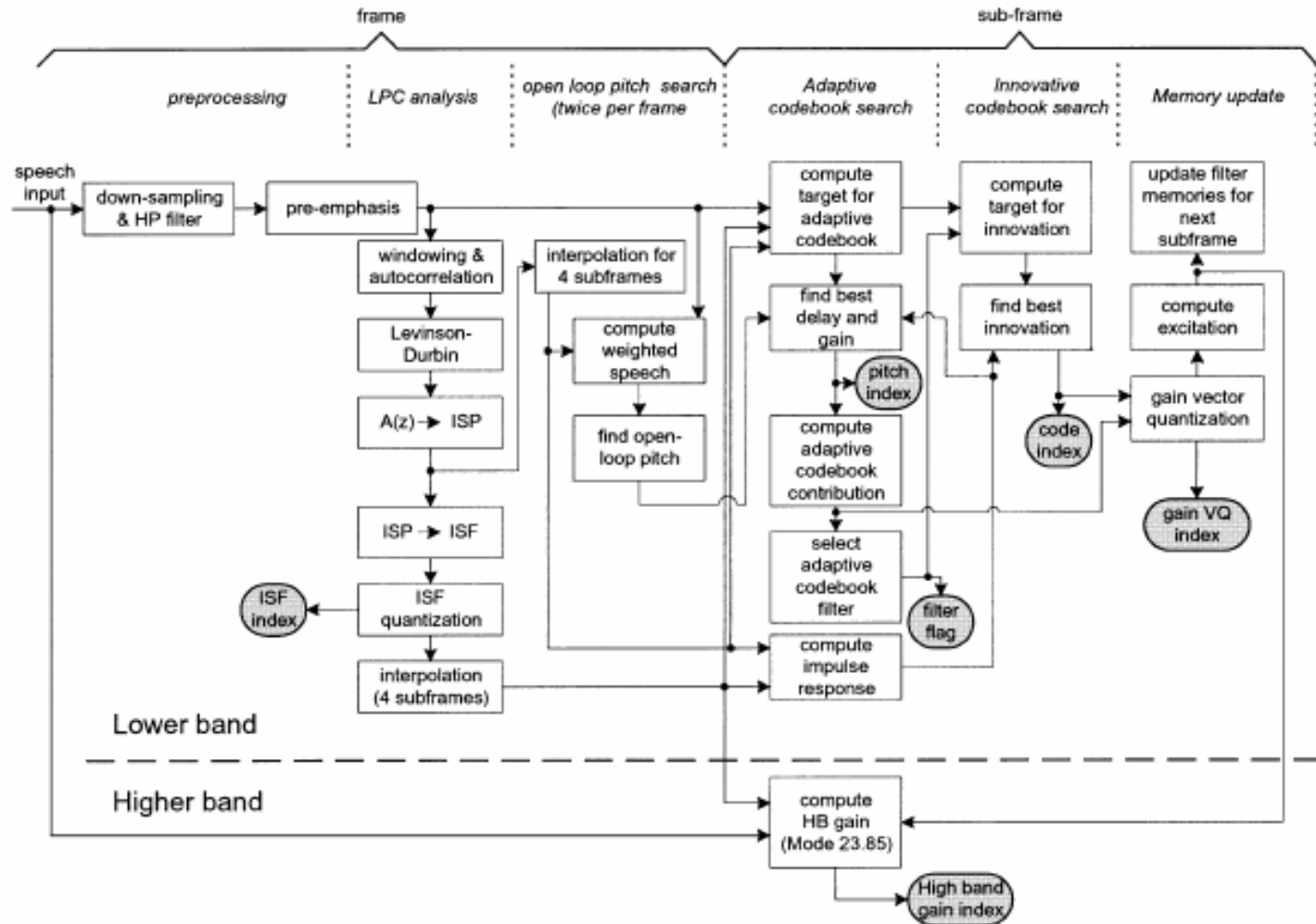
Voiced Speech



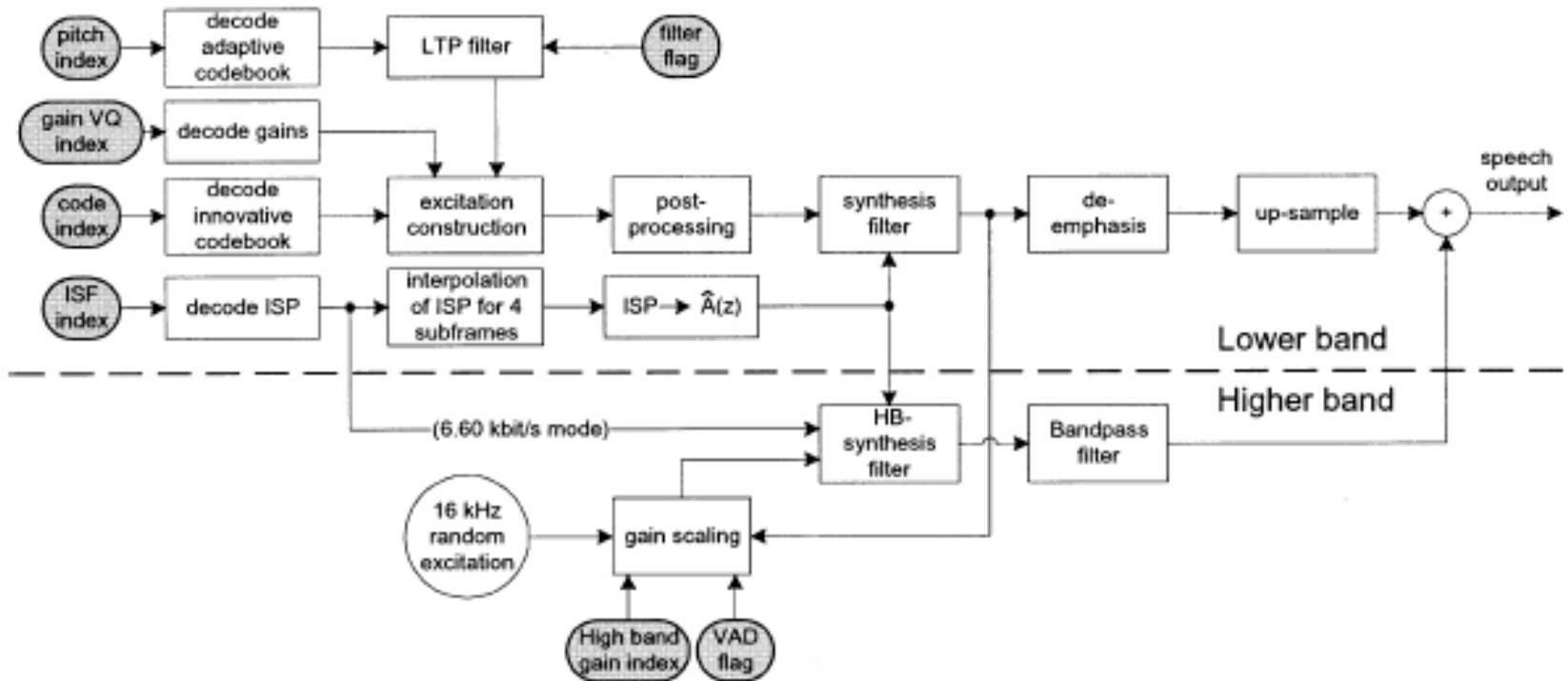
Unvoiced Speech



AMR Wideband Encoder



AMR Wideband Decoder



Bit Allocation for the AMR-WB Modes

PARAMETER	CODEC MODE [kb/s]								
	6.60	8.85	12.65	14.25	15.85	18.25	19.85	23.05	23.85
<i>VAD flag</i>	1	1	1	1	1	1	1	1	1
<i>LTP filtering flag</i>	0	0	4	4	4	4	4	4	4
<i>ISP</i>	36	46	46	46	46	46	46	46	46
<i>Pitch delay</i>	23	26	30	30	30	30	30	30	30
<i>Algebraic code</i>	48	80	144	176	208	256	288	352	352
<i>Gains</i>	24	24	28	28	28	28	28	28	28
<i>High-band energy</i>	0	0	0	0	0	0	0	0	16
<i>Total per frame</i>	132	177	253	285	317	365	397	461	477

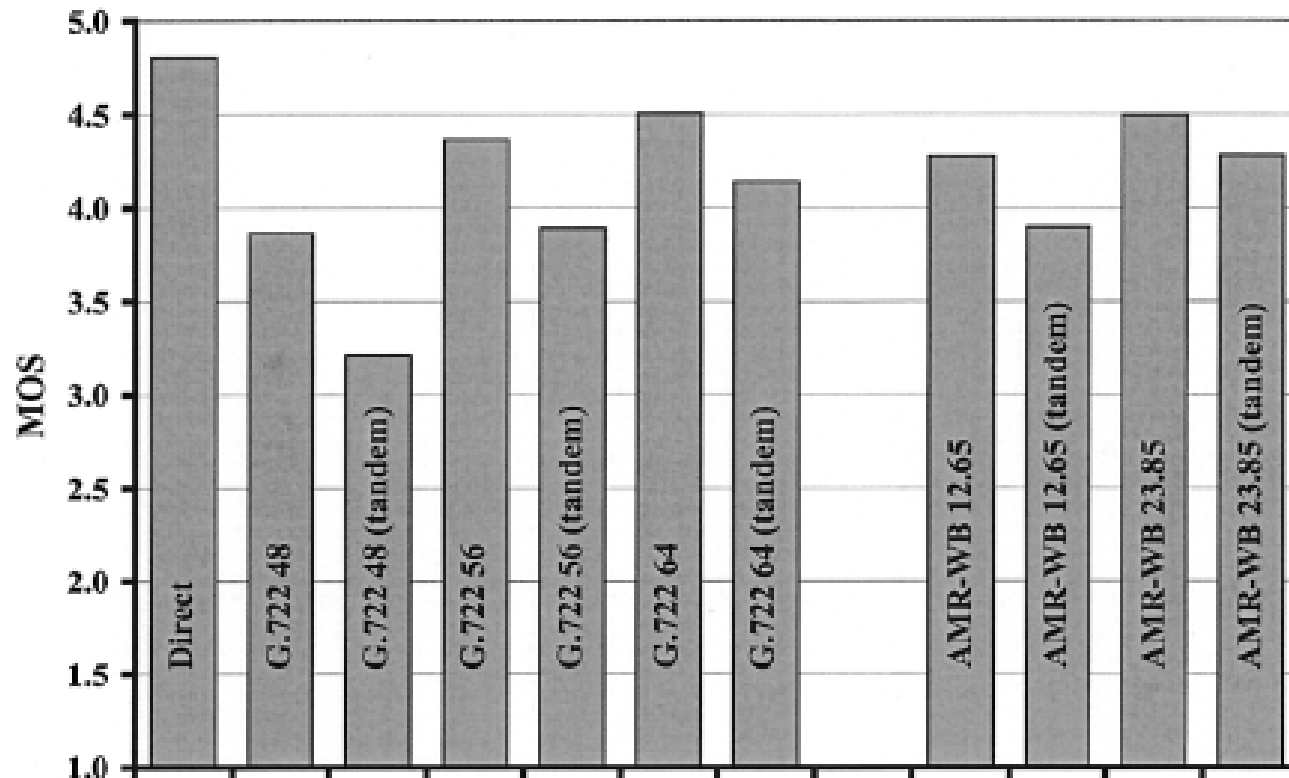
Bit Allocation for a 20 msec Frame

Mode	Parameter	1st subframe	2nd subframe	3rd subframe	4th subframe	Total per frame
12.65 kbit/s	VAD-flag					1
	ISP					46
	LTP-filtering	1	1	1	1	4
	Pitch delay	9	6	9	6	30
	Algebraic code	36	36	36	36	144
	Gains	7	7	7	7	28
	Total					253

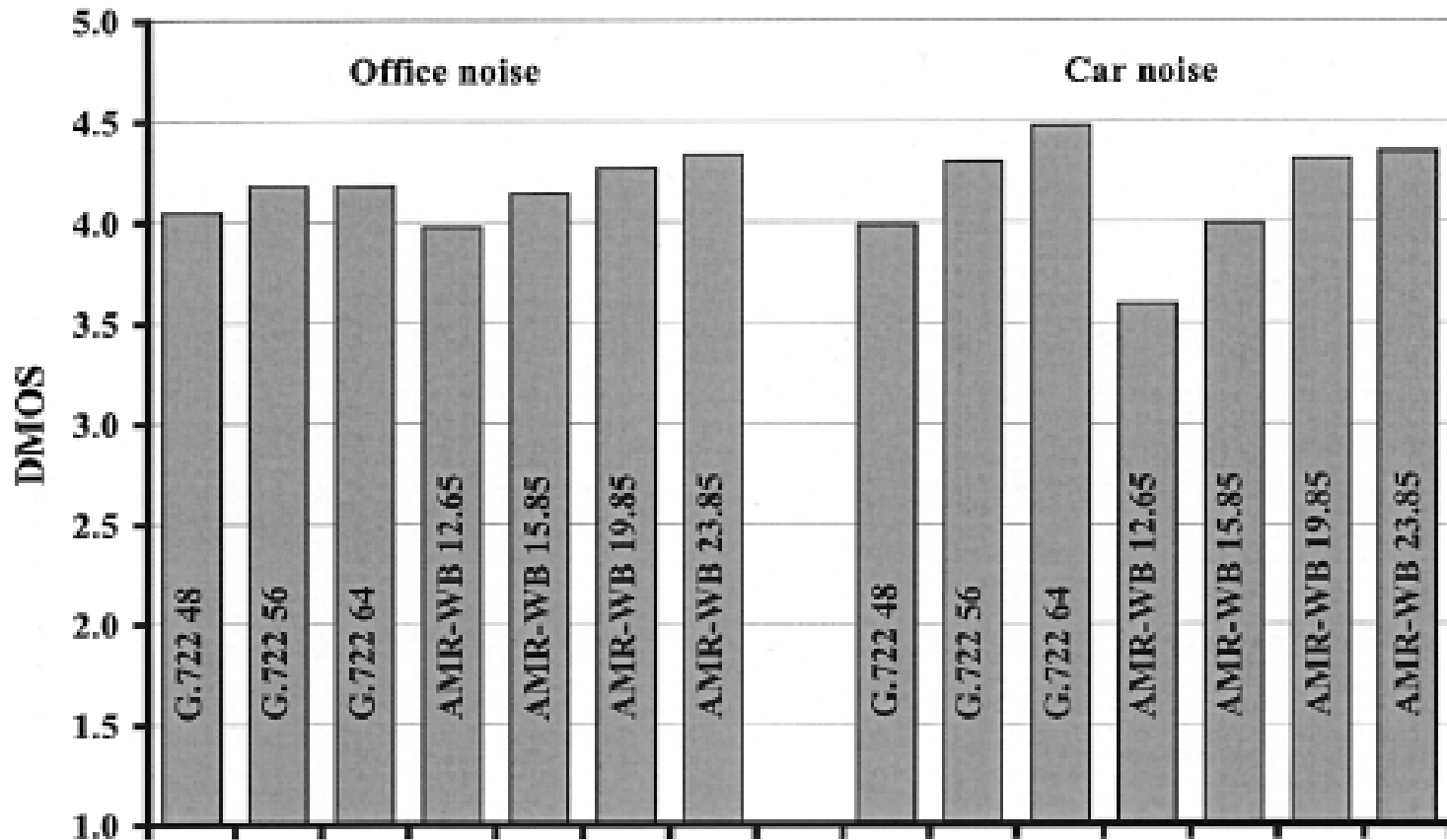
Fixed Codebook Pulse Positions

Track	Valid Pulse Positions in Subframe
T_0	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
T_1	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
T_2	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
T_3	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

AMR-WB Performance for Clean Speech (French)



AMR-WB Performance with Background Noise (English)



AMR-WB Performance with Music Inputs

