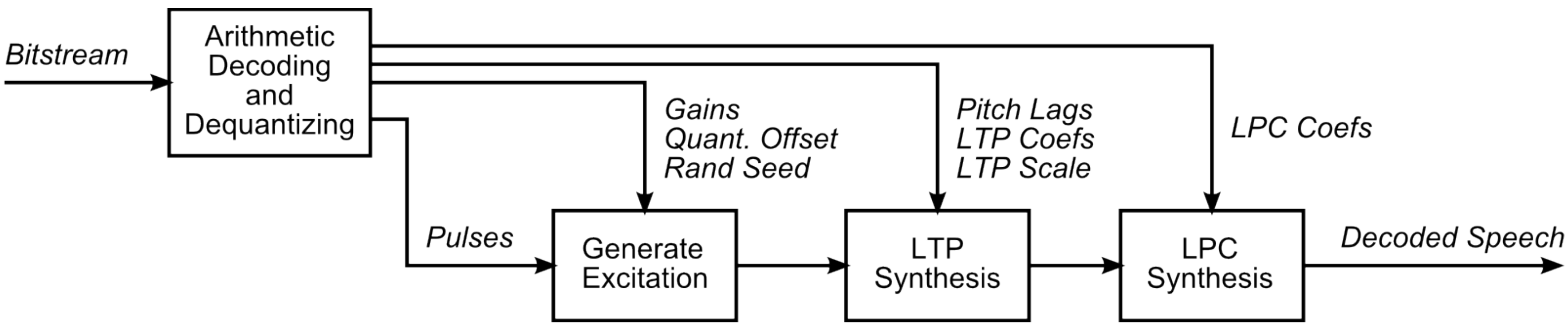
SILK Overview

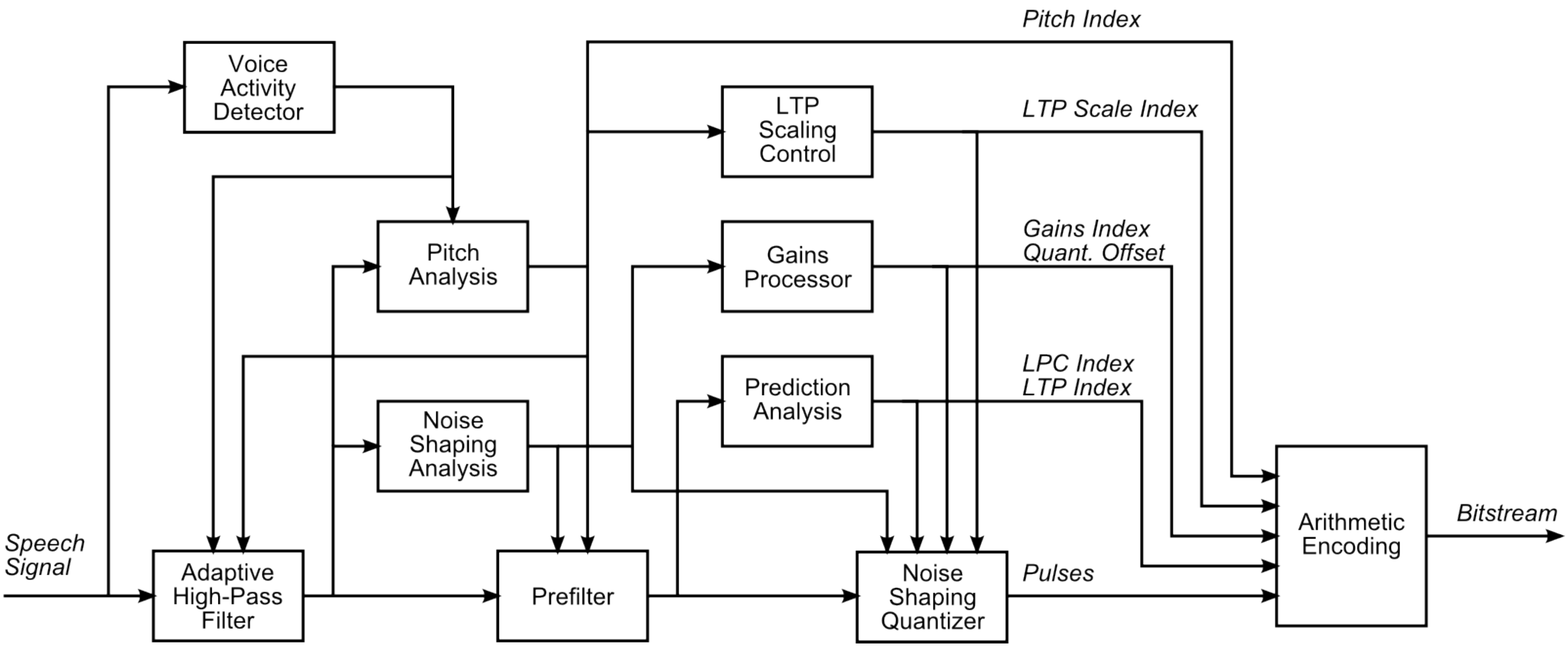
IETF codec WG, Nov 8, 2010

Koen Vos

Decoder


Bitstream
Arithmetic Decoding and Dequantizing
Gains
Quant. Offset
Rand Seed
Pitch Lags
LTP Coefs
LTP Scale
LPC Coefs
Pulses
Generate Excitation
LTP Synthesis
LPC Synthesis
Decoded Speech

Encoder



Pitch Index
Voice Activity Detector
LTP Scaling Control
LTP Scale Index
Pitch Analysis
Gains Processor
Gains Index
Quant. Offset
Prediction Analysis
LPC Index
LTP Index
Noise Shaping Analysis
Arithmetic Encoding
Bitstream
Speech Signal
Adaptive High-Pass Filter
Prefilter
Noise Shaping Quantizer
Pulses

Adaptive High-Pass Filter

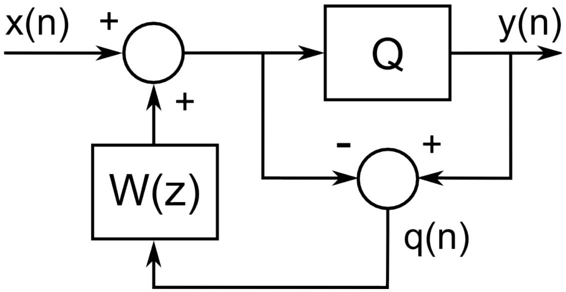
- 2^{nd} order IIR filter
- Cutoff frequency between 80 and 150 Hz
- Depends on:
  - Recent pitch lags: higher cutoff for high pitch frequencies
  - Noise levels: higher cutoff for noisy input

Prediction

- Short-term (LPC) + long-term (LTP)
- Re-estimate LPC given LTP coefficients
- Burg's method
- LPC coefficients coded as Line Spectral Frequencies (LSFs): multi-stage VQ with entropy coding of indices
- Interpolation of LFSs for first 10 ms

Two Noise Shaping Structures

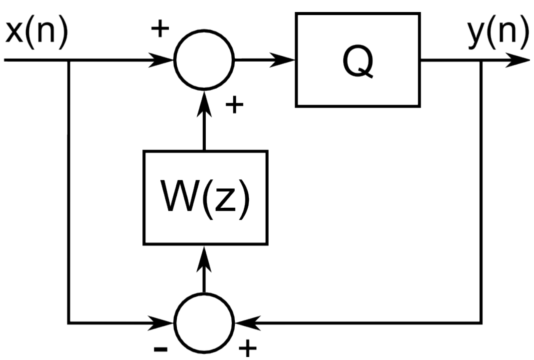
- Moving Average (most commonly used)



$$Y(z) = X(z) + (1 + W(z)) \cdot Q(z)$$

$$\text{with:} \quad W(z) = \sum_{n=1}^{D} w_n z^{-n}$$

- Autoregressive



$$Y(z) = X(z) + (Y(z) - X(z)) \cdot W(z) + Q(z)$$

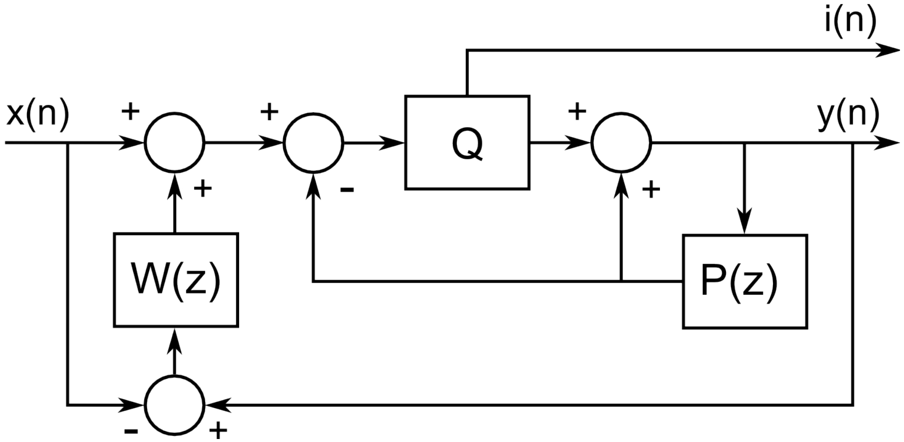
$$\Leftrightarrow$$

$$Y(z) = X(z) + \frac{Q(z)}{1 - W(z)}$$

$$\text{with:} \quad W(z) = \sum_{n=1}^{D} w_n z^{-n}$$

Note: for simplicity, quantization noise is treated as an independent, additive signal.

With Prediction

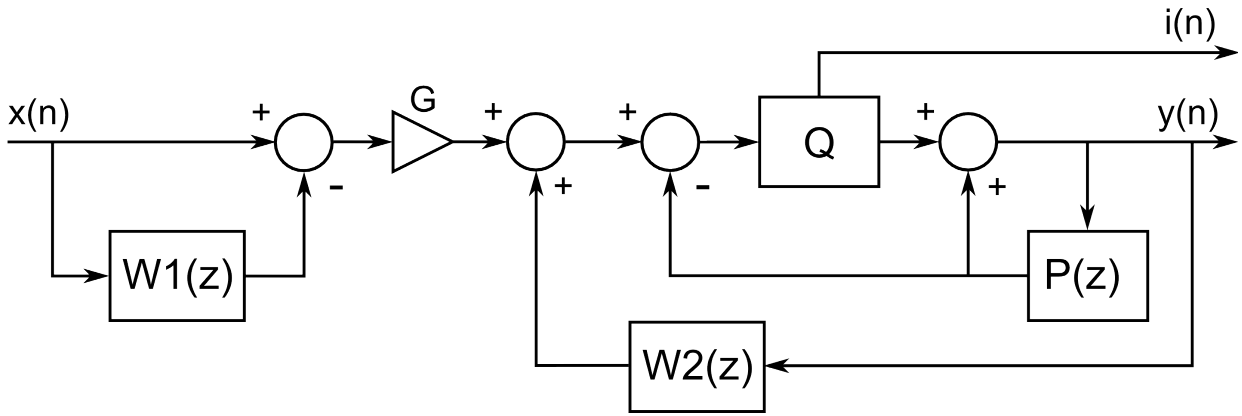


$$Y(z) = X(z) + \frac{Q(z)}{1 - W(z)}$$

- Predictor *P(z)* does LPC and LTP
- Noise shaping filter *W(z)* performs short-term and long-term noise shaping
- Setting W(z) = 0: closed-loop predictive quantizer adds white noise
- Setting *W(z) = P(z):* the quantizer becomes open-loop, P(z) determines noise
- In practice: something in between

Note: some high-rates assumptions were made.

Combined Prediction and Noise Shaping



$$Y(z) = G \cdot (1 - W_1(z)) \cdot X(z) + W_2(z) \cdot Y(z) + Q(z)$$

$$\Leftrightarrow$$

$$Y(z) = G \cdot \frac{1 - W_1(z)}{1 - W_2(z)} X(z) + \frac{1}{1 - W_2(z)} Q(z)$$

- Quantization noise is shaped by *W2(z)*
- Input is pre-filtered by (1-W1(z))/(1-W2(z)), and scaled by G`

Entropy Coding of Excitation

- Coded per 16-sample block
- First compute sum of absolute values
- Then recursively split the block in half, each time entropy coding the sum of absolute values in each half
- Signs of non-zero samples coded separately
- For large signal, LSBs are coded separately