Speech coding

- Requirement of speech coding : ⓐ high speech quality
  ⓑ low bit rate
  ⓒ low complexity
  ⓓ limited signal delay.

  Note : some contradicting requirements. we need trade-off.

- Speech coding algorithms can be mainly divided into three categories:
  ⓐ waveform coding
  ⓑ parametric coding
  ⓒ hybrid coding

Waveform coding : quantization of samples of speech signal itself or some intermediate signals, such as prediction error signal or subband signals.

Example: DPCM, ADPCM
quality: good, SNR = 30–35 dB
bit rate: $f_s = 8$ kHz, B = 32 kbits/sec

Parametric coders (vocoders): Do not code the waveform, but a set of model parameters. The model parameters are quantized and transmitted. Using the coded model parameters, speech is generated at the receiver.

Example: Vocoder
Application area: mainly for defence, military deployed in a no-mans land.
quality: synthetic speech
bit rate: B = 2.4 Kbits/sec.

Hybrid coding: ① A compromise between waveform coding and parametric coding.

② The parameters of a time-varying speech signal is quantized and send. Also the residual signal is quantized blockwise and send.

Example: ~~ALL~~ many telephone quality speech codec.

bit rate: 0.75 – 1.5 bits/sample.

quality: SNR ~ 10 dB

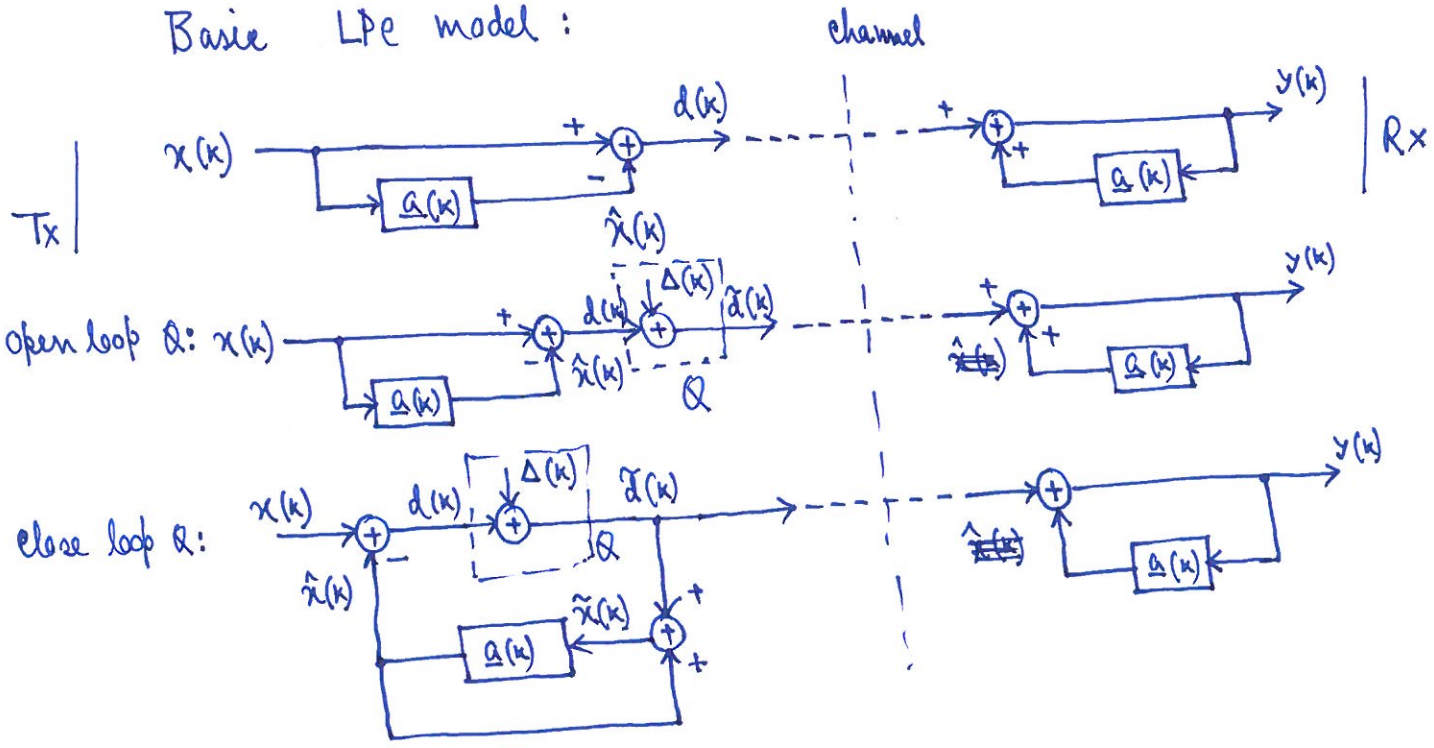
subjective quality: OK

NOTE: All the three approach uses linear prediction (LP).

Generic coding term: Linear Predictive Coding (LPC).

Quantization of the Residual Signal

Basic LPC model:



Quantization with open-loop prediction

Quantized residual:

$$\tilde{d}(k) = d(k) + \Delta(k) .$$

$$\tilde{D}(z) = D(z) + \Delta(z) .$$

Now, $\quad \tilde{D}(z) + A(z)\,Y(z) = Y(z)$

So, $\quad \dfrac{Y(z)}{\tilde{D}(z)} = \dfrac{1}{1-A(z)} .$

or, $\quad Y(z) = \dfrac{1}{1-A(z)}\,\tilde{D}(z) = \dfrac{1}{1-A(z)} \cdot D(z) + \dfrac{1}{1-A(z)} \cdot \Delta(z) .$

what is $D(z)$ & and $X(z)$ relation?

$$D(z) = X(z) - A(z)\,X(z)$$
$$= (1-A(z))\,X(z)$$

or, $\quad \dfrac{D(z)}{1-A(z)} = X(z)$

$$\therefore \quad Y(z) = X(z) + \frac{1}{1-A(z)}\,\Delta(z) .$$

or, $\quad y(k) = x(k) + r(k) \qquad$ where $r(k) = Z^{-1}\left(\dfrac{1}{1-A(z)}\,\Delta(z)\right) .$

Assume that the ~~noise~~ quantization noise $\Delta(k)$ is uniformly distributed as $\quad -\dfrac{\Delta d}{2} \le \Delta(k) \le +\dfrac{\Delta d}{2} \quad$ (in range)

with the power $\quad \sigma_\Delta^2 = \dfrac{(\Delta d)^2}{12}$

and a constant noise PSD

$$\phi_{\Delta\Delta}(e^{j\omega}) = \frac{(\Delta d)^2}{12} . \qquad \text{---- (i)}$$

Here, Δd denotes the quantizer's step size.

Note : ① The reconstruction error is $r(k)$.

② The $r(k)$ is a filtered version of $\Delta(k)$, the quantization noise.

③ The spectrum of the ~~quantization noise~~ reconstruction error $r(k)$ follows the spectral envelope of speech signal.

Why ? Because $\frac{1}{1-A(z)}$ is the synthesis filter that models the vocal tract.

④ This is called "noise shaping".

- Let us see the interplay between prediction gain, signal-to-reconstruction-error ratio, signal-to-quantization-noise ratio.

Signal-to-reconstruction error ratio = Signal-to-noise ratio for output signal. $\left(\frac{S}{N}\right)_y$.

$$\left(\frac{S}{N}\right)_y = \frac{\sigma_x^2}{\sigma_r^2} = \frac{\phi_{xx}(0)}{\varphi_{rr}(0)}$$

. we dont know $\varphi_{rr}(0)$.

Assumption : ⓐ Let us assume the LP provides a perfect job.

ⓑ The LP prediction error $d(k)$ is white.

ⓒ PSD: $\phi_{dd}(e^{j\Omega}) = \text{const} = \varphi_{dd}(0) = \sigma_d^2$ . . . . ii)

Now, $\phi_{dd}(e^{j\Omega}) = \phi_{xx}(e^{j\Omega}) \cdot |1 - A(e^{j\Omega})|^2$

or, $\frac{\varphi_{dd}(0)}{|1-A(e^{j\Omega})|^2} = \phi_{xx}(e^{j\Omega})$ [using ii)]

or, $\varphi_{dd}(0) \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{|1-A(e^{j\Omega})|^2} d\Omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_{xx}(e^{j\Omega}) d\Omega = \rho_{xx}(0)$
$= \sigma_x^2$.

$$\therefore \quad \frac{\sigma_x^2}{\sigma_d^2} = \frac{\varphi_{xx}(0)}{\varphi_{dd}(0)} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{|1 - A(e^{j\Omega})|^2}\, d\Omega = G_p \quad \ldots\ldots \text{(iii)}$$

Define $|H(e^{j\Omega})|^2 = \dfrac{1}{|1 - A(e^{j\Omega})|^2}$

Now, Signal-to-quantization-noise-ratio = Signal-to-noise ratio for prediction error $d(k)$

$$\left(\frac{S}{N}\right)_x = \frac{\sigma_d^2}{\sigma_\Delta^2} = \frac{\varphi_{dd}(0)}{\frac{(\Delta d)^2}{12}} \qquad \ldots \text{(iv)}$$

Now, $\phi_{rr}(e^{j\Omega}) = |H(e^{j\Omega})|^2 \cdot \phi_{\Delta\Delta}(e^{j\Omega})$

(because $\mathcal{Z}\{r(k)\} = \dfrac{1}{1-A(z)}\,\Delta(z)$

$R(z) = \dfrac{1}{1-A(z)} \cdot \Delta(z)$.

$$\therefore \quad \varphi_{rr}(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi_{rr}(e^{j\Omega})\, d\Omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} |H(e^{j\Omega})|^2\, \phi_{\Delta\Delta}(e^{j\Omega})\, d\Omega$$

$$= \frac{(\Delta d)^2}{12} \cdot \frac{1}{2\pi}\int_{-\pi}^{\pi} |H(e^{j\Omega})|^2\, d\Omega \qquad [\text{using (ii)}]$$

$$= \frac{(\Delta d)^2}{12} \cdot G_p \qquad [\text{using (iii)}]$$

$$\therefore \left(\frac{S}{N}\right)_y = \frac{\sigma_x^2}{\sigma_r^2} = \frac{\varphi_{xx}(0)}{\varphi_{rr}(0)} = \frac{G_p \cdot \varphi_{dd}(0)}{\frac{(\Delta d)^2}{12} \cdot G_p} = \frac{\varphi_{dd}(0)}{\frac{(\Delta d)^2}{12}} = \left(\frac{S}{N}\right)_x$$

(using (iii) and (iv)).

Main point: The ultimate quality is the $\left(\frac{S}{N}\right)_y$, which did not improve by open-loop prediction. But we got spectral shaping.

Closed loop Quantization

$$y(k) = \tilde{x}(k) = \hat{x}(k) + \tilde{d}(k)$$

$$= \hat{x}(k) + d(k) + \Delta(k)$$

$$= x(k) + \Delta(k) \ .$$

$$\uparrow$$

$$d(k) = x(k) - \hat{x}(k)$$

Note : (a) The reconstruction error is the quantization noise $\Delta(k)$.

(b) The reconstruction error is spectrally white.

$$\left(\frac{S}{N}\right)_y = \frac{\sigma_x^2}{\sigma_\Delta^2} = \frac{\sigma_x^2}{\sigma_d^2} \cdot \frac{\sigma_d^2}{\sigma_\Delta^2}$$

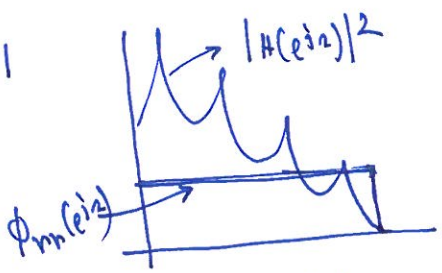
$$= G_p \cdot \left(\frac{S}{N}\right)_{\tilde{d}}$$

$$\therefore \ SNR_y = 10 \log_{10} G_p + 10 \log_{10}\left(\frac{S}{N}\right)_{\tilde{d}}$$

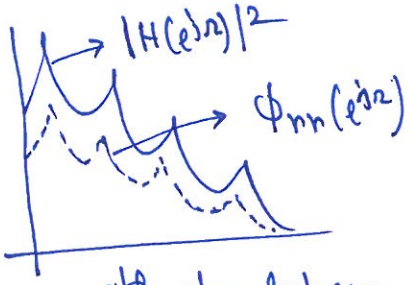
- Closed loop prediction (following 6-dB-per-bit rule), provides $\frac{10 \log_{10} G_p}{6}$ bit advantage than direct uniform quantization of speech signal (PCM).

- But, we did not get any spectral shaping.

- See Figure 8.10 of the reference book



closed loop case



open loop case

- Main points:

	closed loop	Open loop
spectral shaping	X	✓
SNR improvement	✓	X

But, we want both.

Q: Can we do in between? A scheme that combined closed loop and open loop.

Ans: Yes. It is possible by generalizing the closed loop structure and inserting a noise shaping filter.