Special Module on Media Processing and Communication

Dayalbagh Educational Institute
(DEI)
Dayalbagh Agra

Indian Institute of Technology Delhi
(IITD)
New Delhi

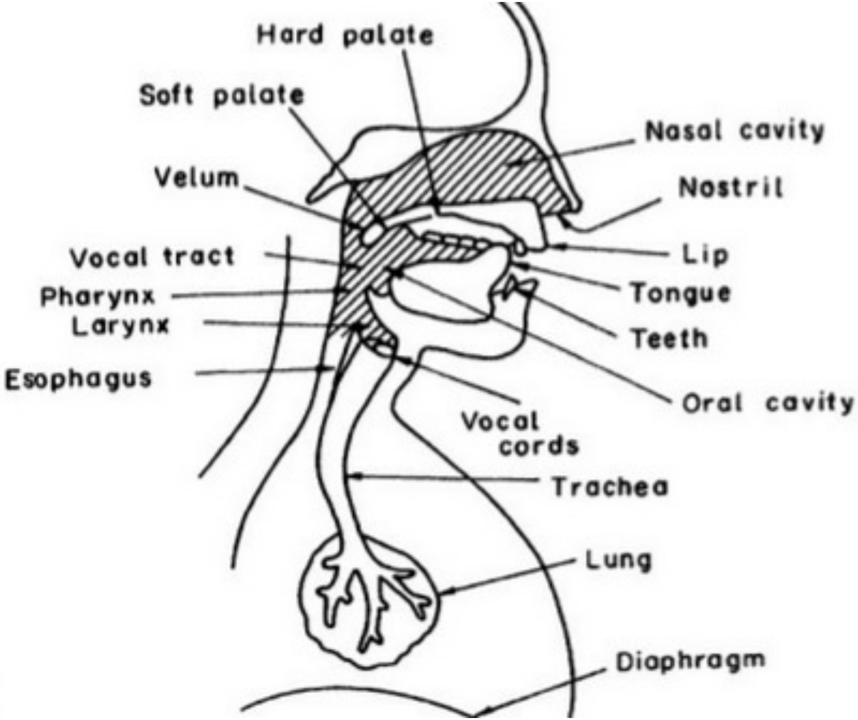
Speech Compression

Speech Production

Speech is produced by forcing air from the lungs through the vocal cords in to the vocal tract.

The sound is generated by vibrations. The pitch of the sound is controlled by varying the shape of the vocal track. The loudness is controlled by the amount of the air exhaled/inhaled.

The process is slow – sampling 20 ms

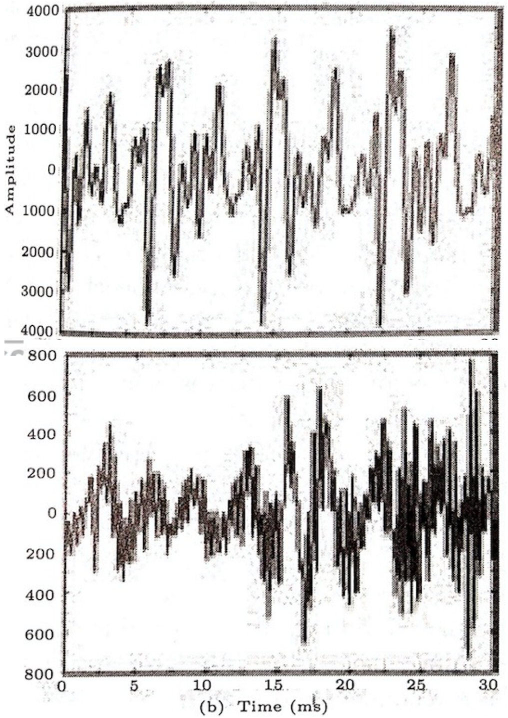


Speech Sounds

There can be Voiced sounds and Unvoiced sounds.

Voiced sounds are made while talking. The frequencies are restricted to 500 Hz to 2 KHz. There is periodicity.

Unvoiced sounds are heard but not part of speech. Samples are uncorrelated and random.



Voiced

Unvoiced

Speech Coding

Waveform Codecs

    Audio codecs (PCM, DPCM, ADPCM)

Source Codecs (Vocoders)

    Linear Predictive Coding (LPC)

Hybrid

    Code-excited Linear Predictive (CLP) Codec

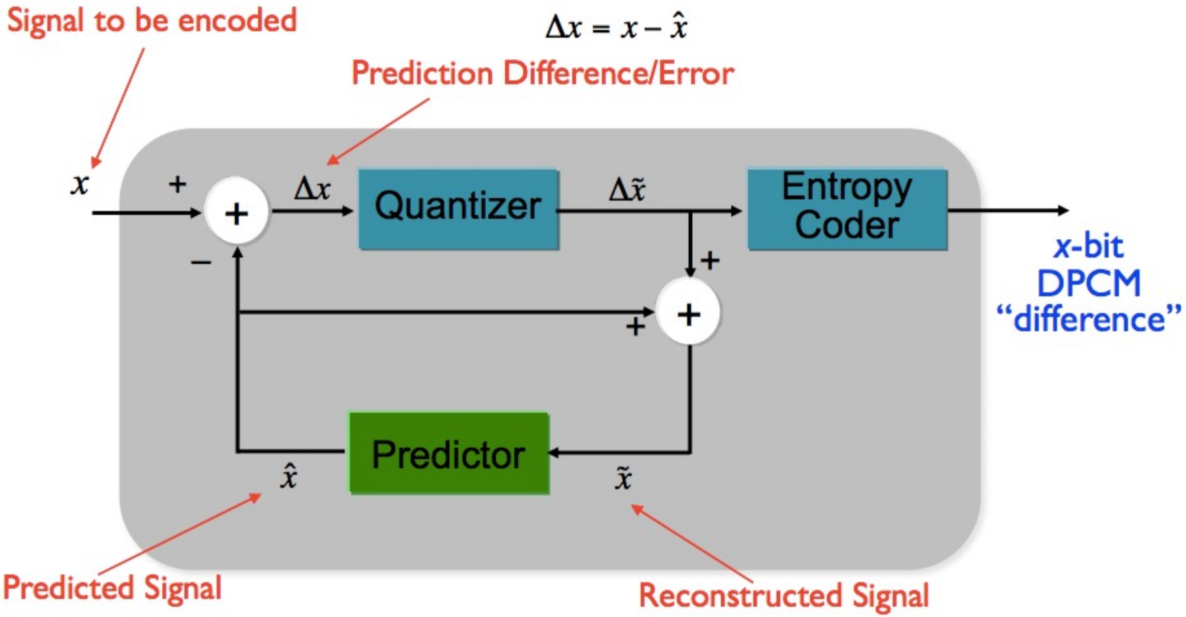
Waveform Coding

Similar to audio coding

- Pulse Code Modulation (PCM)
- Quantizers
  - μ-law and A-law companding methods
- Differential PCM (DPCM)
- Adaptive DPCM
- Subband Coding Algorithms (Psychoacoustic Model)

Differential Pulse Code Modulation(DPCM)

Revisit



$\hat{x}(n) = \tilde{x}(n-1)$ Simple difference

$\tilde{x} = \hat{x} + \Delta\tilde{x}$

$\hat{x}(n) = \sum_{k=1}^{N} h_k \tilde{x}(n-k)$ Linear predictor => minimizes quantization error

$\hat{x}(n) = 0$ PCM

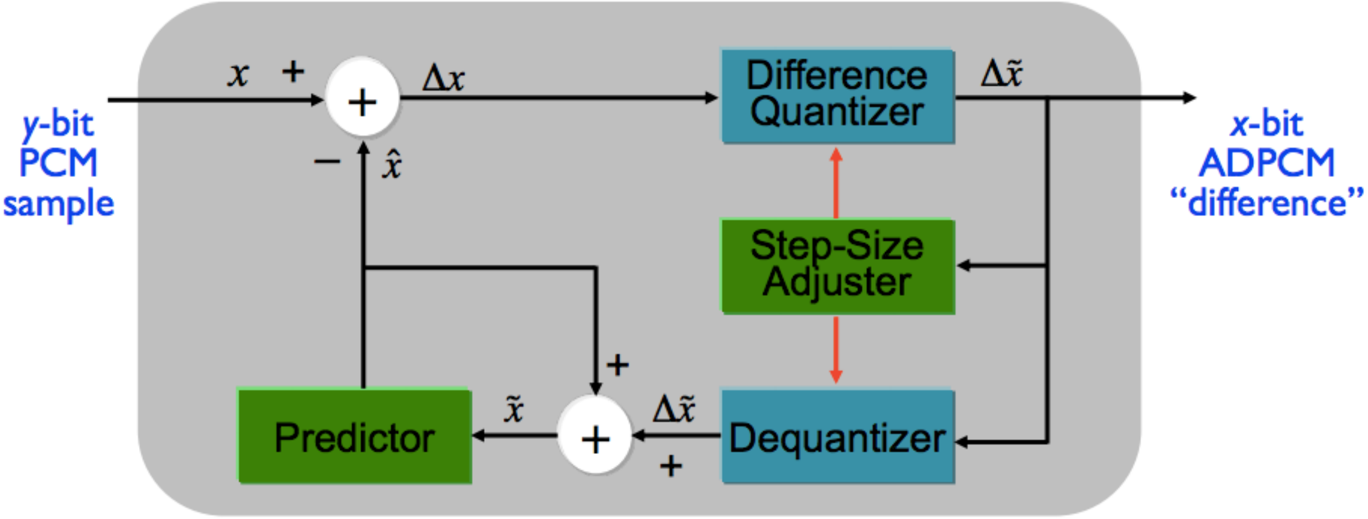
Source [2]

Adaptive DPCM

To ensure differences are always small…

- Adaptively change the step-size (quanta).
- (Adaptively) attempt to predict next sample value.

Revisit

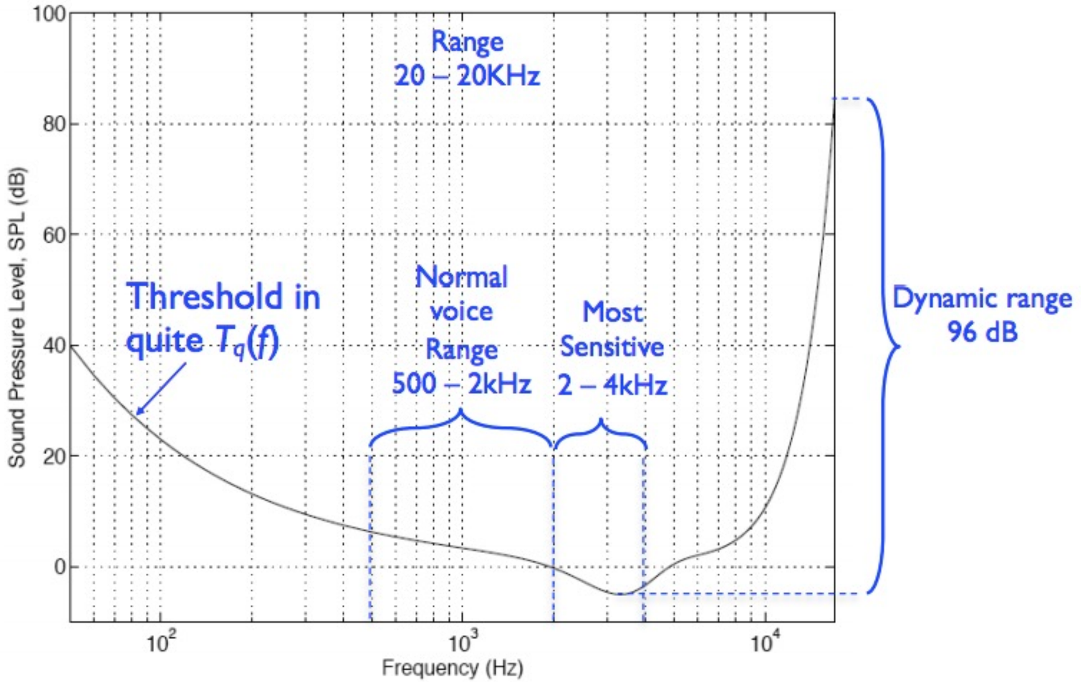


$$\tilde{x} = \hat{x} + \Delta\tilde{x}$$

Source [2]

Psychoacoustic Coding

Absolute threshold of h
in quite environment

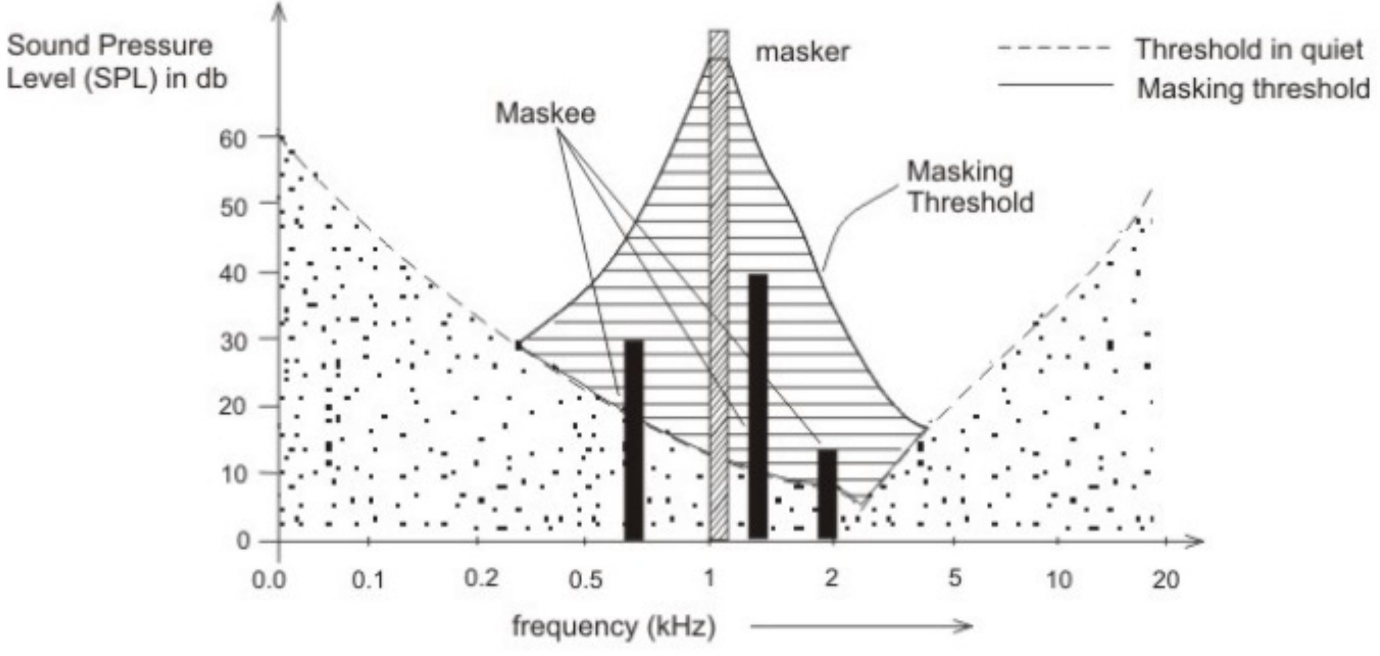


$$T_q(f) = 3.64(f/1000)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 10^{-3}(f/1000)^4$$

Psychoacoustic Coding

Simultaneous Masking

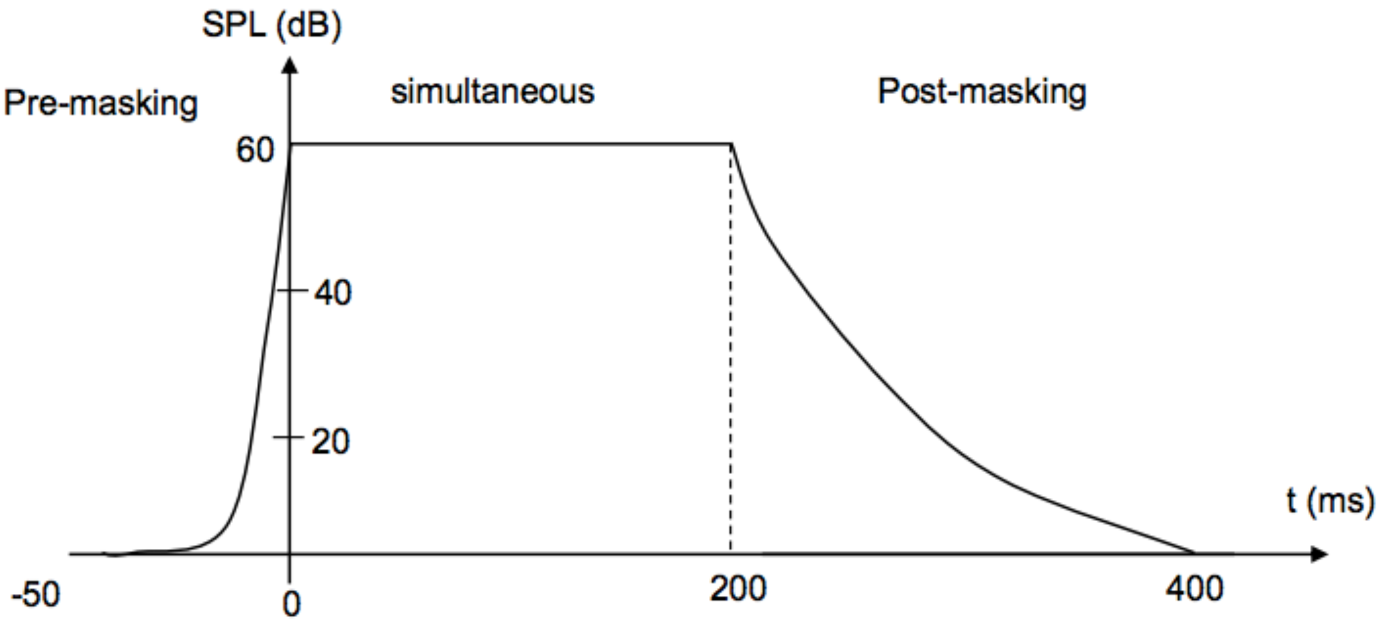
Revisit



Psychoacoustic Coding

TEMPORAL MASKING

Revisit



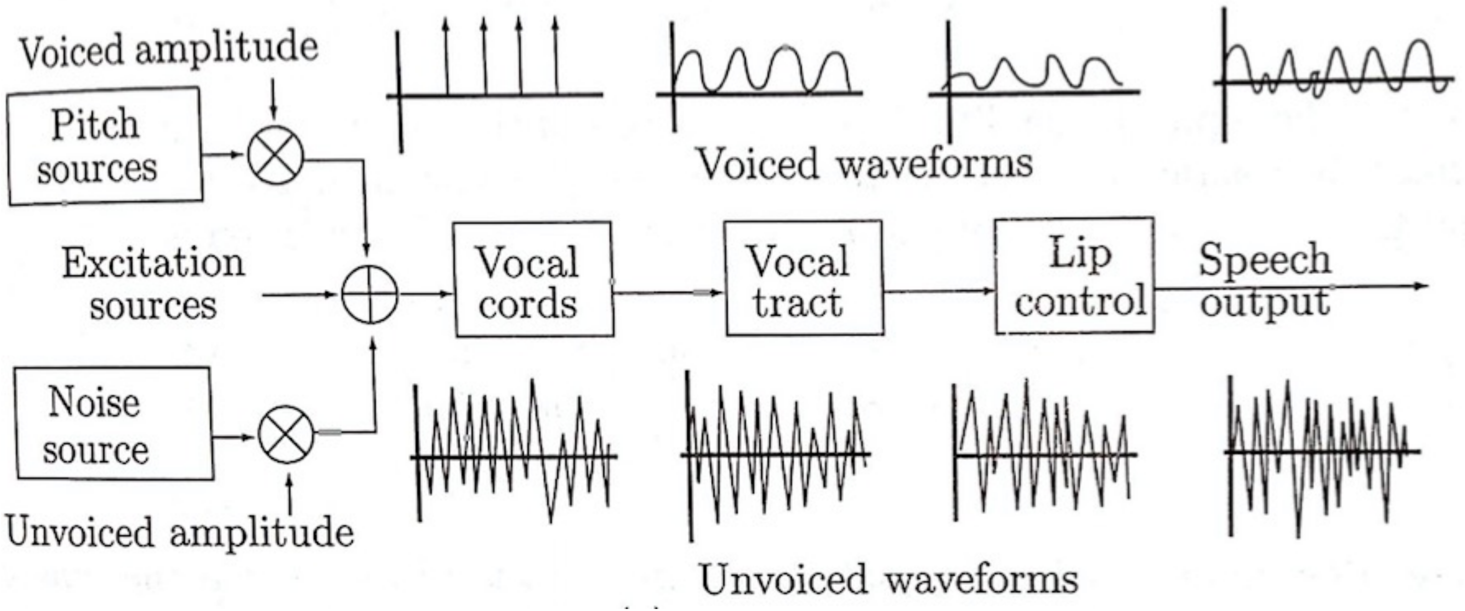
Source Coding

Uses a model for the source data that depend on certain parameters. The encoder computes (extract) parameters from input and encodes them.

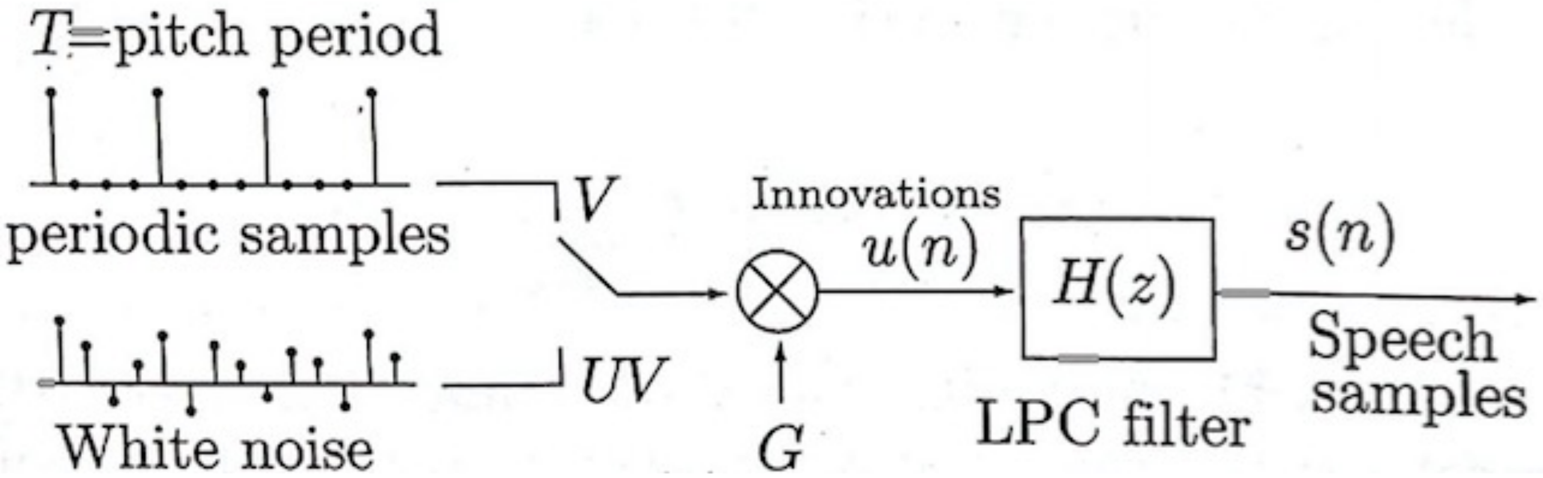
Decoder inputs the parameters and reconstructs the original speech using the model.

Also called vocoders (vocal coder). Linear Prediction Coder (LPC) provides a ***robust, reliable and accurate method*** for estimating the parameters of the linear system (the com combined vocal tract, glottal pulse, and radiation characteristic for voiced speech)

Linear Predictive Coding (LPC)



Linear Predictive Coding (LPC)



Linear Predictive Coding (LPC)

Differential Pulse Coding Modulation (DPCM) uses predictive modeling

$$\tilde{s}(n) = \sum_{k=1}^{p} \alpha_k s(n-k)$$

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k)$$

Obtained through minimization of e(n)

What if these related to the speech parameters !

Linear Predictive Coding (LPC)

Differential Pulse Coding Modulation (DPCM) uses predictive modeling

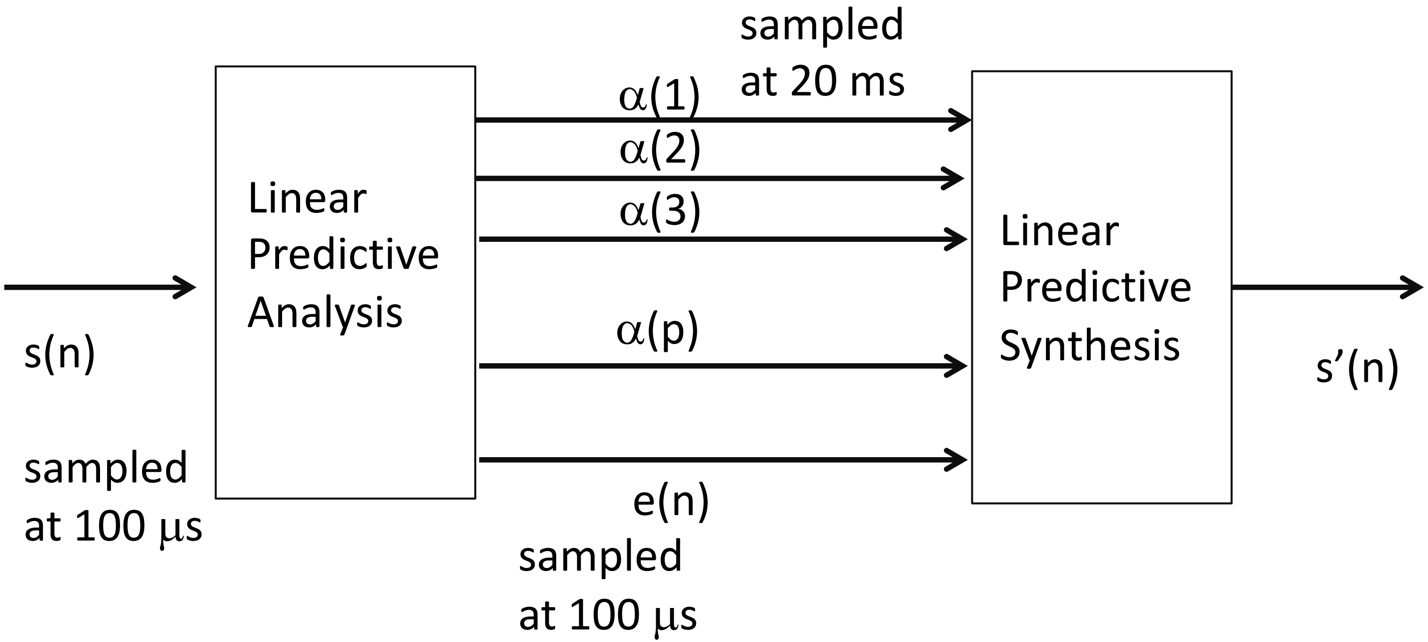
$$\tilde{s}(n) = \sum_{k=1}^{p} \alpha_k s(n-k)$$

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k)$$

Obtained through minimization of e(n)

What if these related to the speech parameters !

Linear Predictive Coding (LPC)



Linear Predictive Coding (LPC)

Given error of prediction in time domain

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k)$$

Taking Z-transform of both sides gives

$$E(z) = S(z) - \sum_{k=1}^{p} \alpha_k z^{-k} \; S(z)$$

$$= \left[1 - \sum_{k=1}^{p} \alpha_k z^{-k} \right] S(z)$$

$$\boxed{S(z) = \frac{1}{1 - \underbrace{\sum_{k=1}^{p} \alpha_k z^{-k}}_{P(z)}} E(z)}$$

Linear Predictive Coding (LPC)

Vocal Track Model

$$S(z) = \frac{1}{1 - \sum_{k=1}^{q} c_k z^{-k}} I(z)$$

Digital formant filter

a formant is the broad spectral maximum that results from an acoustic resonance of the human vocal tract

Generally an odd order filter

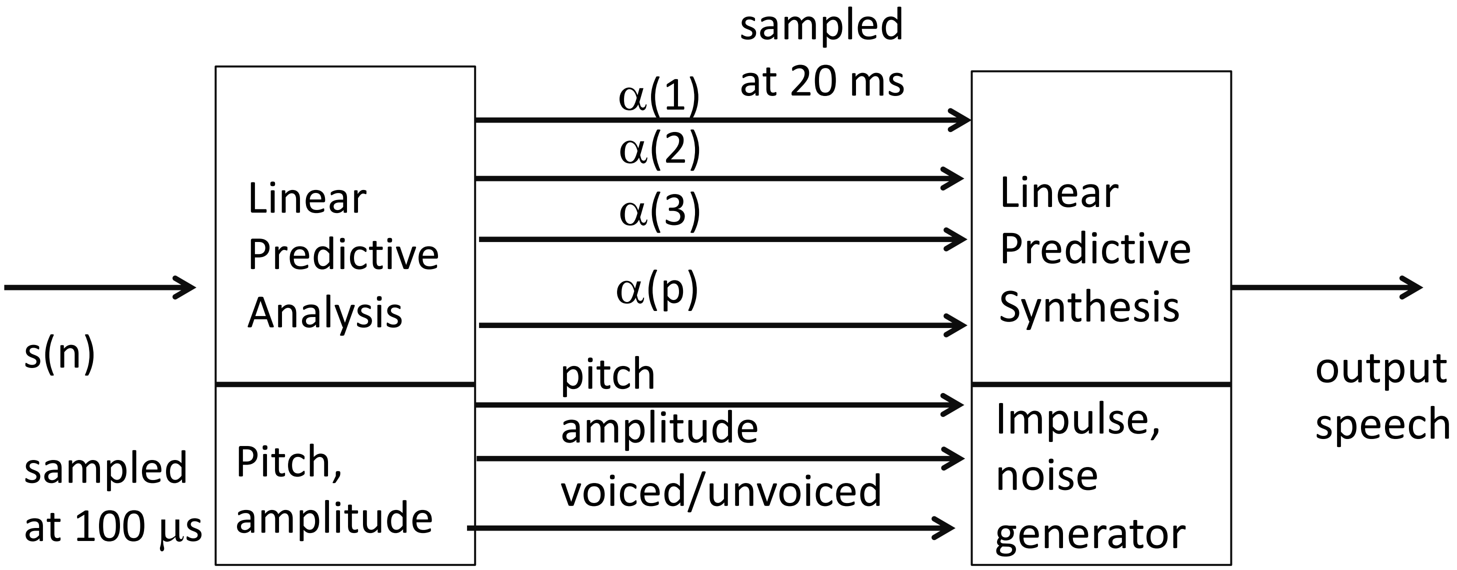
q=2n+1 where n is the number of formants

Input is the excitation impulse to the vocal track

Output is the S(z)

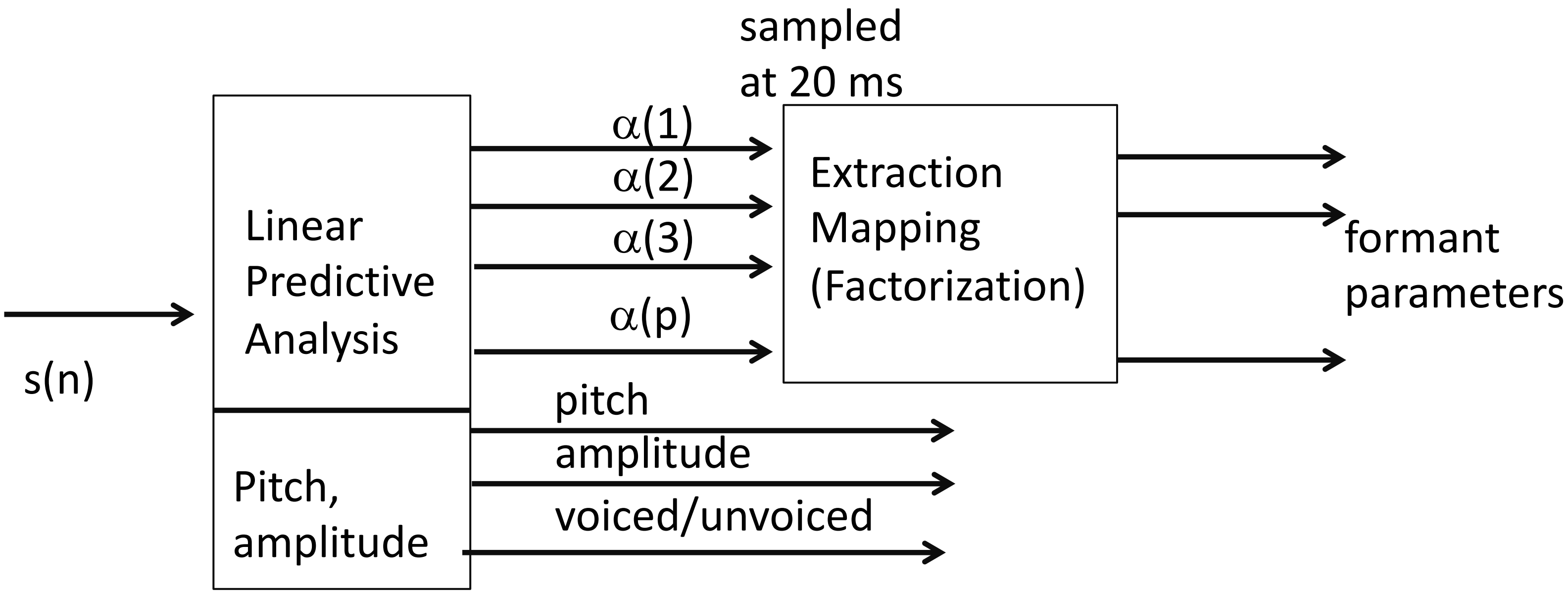
c_k give the position and bandwidths of the formant resonances

Linear Predictive Coding (LPC)



Very low bit speech coder

Linear Predictive Coding (LPC)



Linear Predictive Coding (LPC)

Computation of $\alpha(k)$ by minimization of the mean square error

$$E_{\hat{n}} = \sum_{m} e_{\hat{n}}^{2}(m) = \sum_{m} \left(s_{\hat{n}}(m) - \tilde{s}_{\hat{n}}(m)\right)^{2}$$

$$= \sum_{m} \left(s_{\hat{n}}(m) - \sum_{k=1}^{p} \alpha_k s_{\hat{n}}(m-k) \right)^{2}$$

References

1. NPTEL course on Multimedia Processing by Prof S Sengupta.
2. http://web.engr.oregonstate.edu/~benl/Courses/ece477.sp20/Lectures/
3. https://web.ece.ucsb.edu/Faculty/Rabiner/ece259/