Predictive Modeling of Speech

Dr. Yogananda Isukapalli

PREDICTIVE MODELING OF SPEECH

Linear Predictive Coding (LPC):

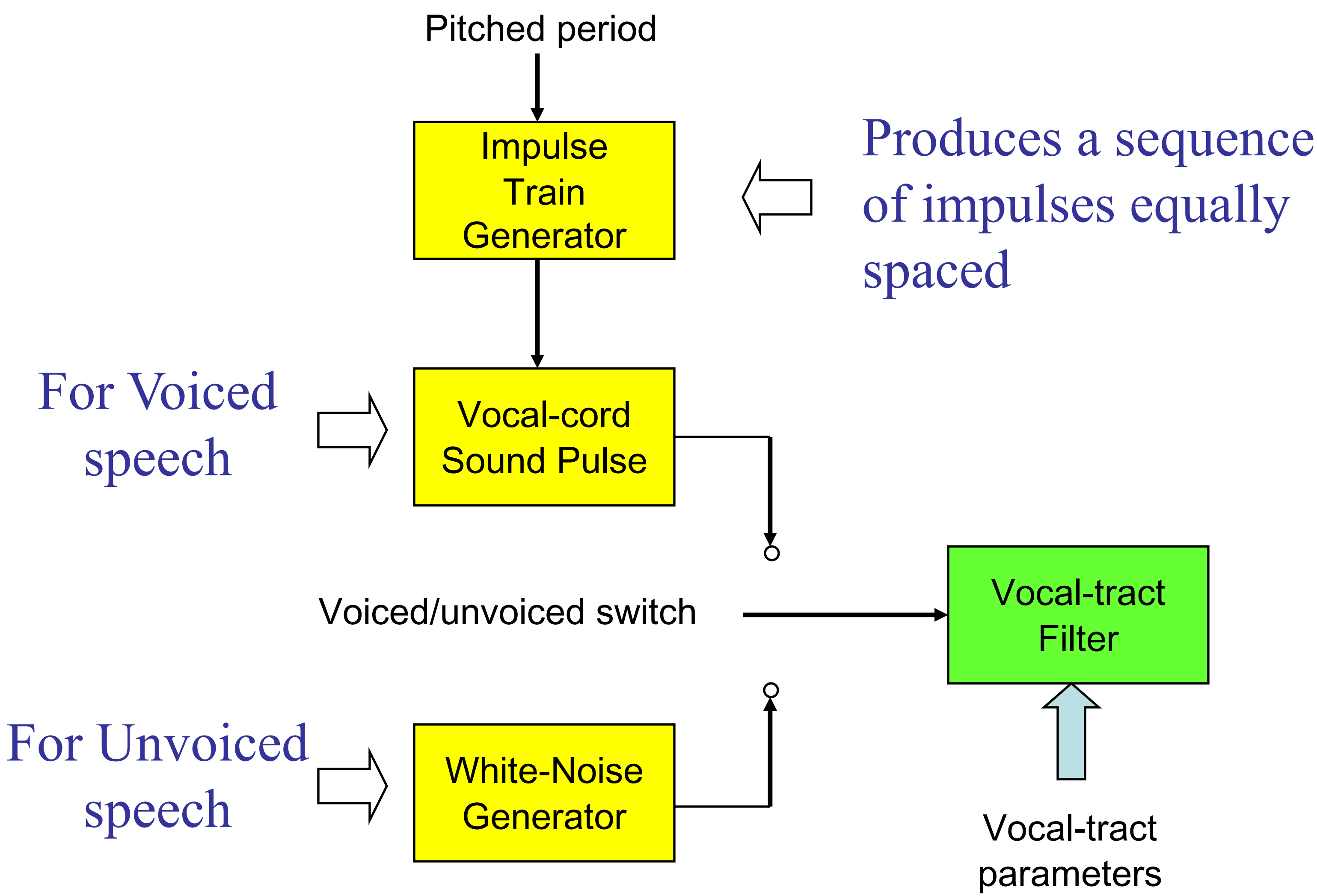
is defined as a digital method for encoding an analog signal in which a particular value is predicted by a **linear** function of the past values of the signal

- First proposed method for encoding Human Speech US DoD

- Human speech is produced in the vocal tract which can be approximated as a variable diameter tube.

- The **linear predictive coding** (LPC) model is based on a mathematical approximation of the vocal tract represented by a tube of a varying diameter

SPEECH PRODUCTION



Block Diagram of simplified model for the speech production process

SPEECH PRODUCTION

- The model assumes that the sound-generating mechanism is linearly separable from the intelligence-modulating vocal-tract filter.

- The precise form of the excitation depends on whether the Speech is voiced or unvoiced.

  - *Voiced speech* sound (such as /*i*/ in *e*ve) is generated from a quasi-periodic excitation of vocal-tract.
  - *Unvoiced speech* sound (such as /*f*/ in *f*ish) is generated from random sounds produced by turbulent airflow through a constriction along the vocal tract.

SPEECH PRODUCTION

Vocal-tract filter is represented by the all-pole transfer function,

$$H(z) = \frac{G}{1 + \sum_{k=1}^{M} a_k z^{-k}}$$

where

G: Gain factor

a_k: Real-value coefficient

The form of excitation applied to this filter is changed by switching between the voiced and unvoiced sounds.

SPEECH PRODUCTION

- ✓ Accurate modeling the short-term power spectral envelope plays an important role in the quality and intelligibility of the reconstructed **speech**.

- ✓ At low bit-rate **speech** coding, an **all-pole** filter is adopted to **model** the spectral information in **LPC** (linear predictive coding)-based coders.

- ✓ By minimizing the MSE between the actual speech samples and the predicted ones, an optimal set of weights for an all-pole (synthesis) digital filter can be determined

SPEECH PRODUCTION

Assuming an "all-pole" model, irrespective of speech signal envelope we are trying to estimate:

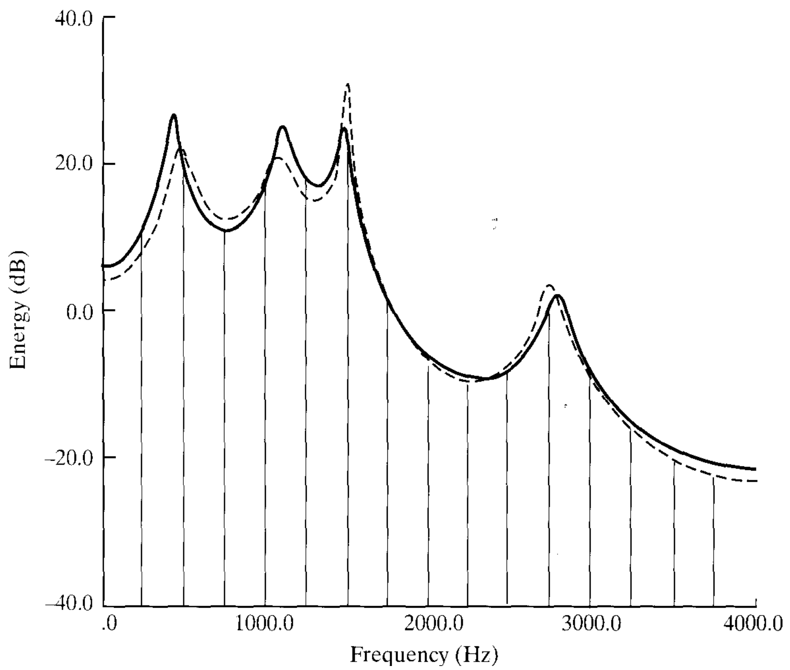
- If sound is unvoiced, usual LPC method gives "good estimate" of filter coefficients a_k

- If sound is voiced, usual LPC method gives a "biased estimate" of a_k with the estimate worsening as the pitch increases (error criterion being MSE)

Example: This example illustrates the limitations of standard all-pole modeling of periodic waveforms. Specifications are:

- All-pole filter order, $M = 12$
- Input Signal: periodic pulse sequence with N=32

SPEECH PRODUCTION



The solid line is the original 12-pole envelope, which goes through all the points.

The dashed line is the 12-pole linear predictive model for N=30 spectral lines

SPEECH PRODUCTION

El-jaroudi and Makhoul used "*Itakura-Saito distance measure*" as "**error criterion"** to overcome this problem to develop a new model called *"discrete all-pole model"*.

Itakura-Saito distance measure:

Let $u(n)$ be a real-valued, stationary stochastic process, it's Fourier Transform U_k is:

$$U_k = \sum_{k=0}^{N-1} \left(u_n e^{-jn\omega_k} \right) \qquad \mathrm{k = 0,1,2, \ldots N - 1}$$

SPEECH PRODUCTION

The auto-correlation function of $u(n)$ for lag m is:

$$r(m) = \frac{1}{N}\sum_{k=0}^{N-1}\left(S_k e^{jn\omega_k}\right) \qquad \text{k = 0,1,2,... N - 1}$$

where

$$S_k = \sum_{k=0}^{N-1}\left(r(m)e^{-jn\omega_k}\right) \qquad \text{k = 0,1,2,... N - 1}$$

Let vector **a** denote a set of *spectral parameters* as:

$$\mathrm{a} \;=\; [a_1, a_2, \ldots\ldots, a_M]^T$$

Note: See p.no.189-191 in textbook for proof

SPEECH PRODUCTION

Itakura-Saito distance measure is given by:

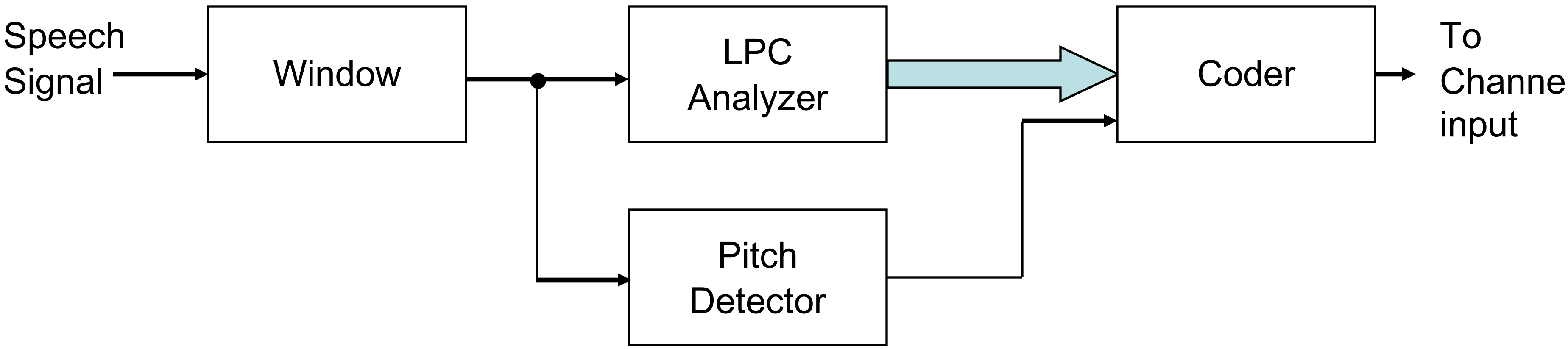
$$D_{IS}(a) = \sum_{k=0}^{N-1} \left(\frac{|U_k|^2}{S_k(a)} - \ln\left(\frac{|U_k|^2}{S_k(a)} \right) - 1 \right)$$

The time-reversed impulse response $h(-i)$ of the discrete frequency-sampled all-pole filter can be obtained from the relation of predictor coefficients to the auto-correlation as:

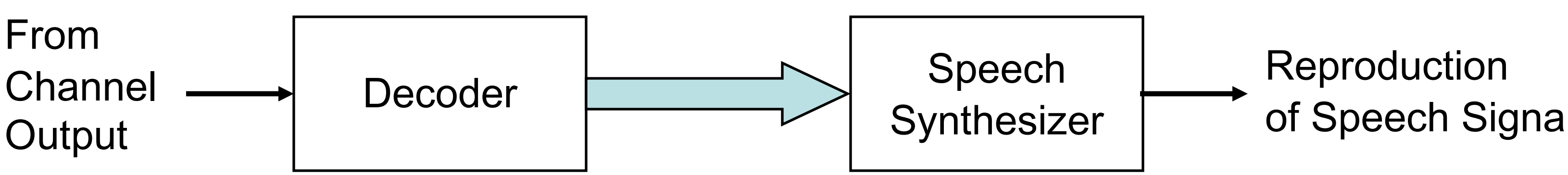
$$\hat{h}(-i) = \sum_{k=0}^{M} a_k r(i-k)$$

LPC Vocoder

LPC system for digital transmission and reception of speech signals over a communication channel



(a)



(b)

Block diagram of LPC Vocoder : a) Transmitter b) receiver

LPC Vocoder

Transmitter:

- Applies window (typically, 10 to 30 ms long) to the input
- Analyzes the input speech block by block
    - Performs Linear prediction
    - Pitch Detection
- Finally, following parameters are encoded:
    - Set of coefficients computed by LPC Analyzer
    - The Pitch Period
    - The Gain parameter
    - The voiced-unvoiced parameters

LPC Vocoder

Receiver:

Performs the inverse operations on the channel output:

- Decode incoming parameters
- Synthesize a speech signal from the parameters