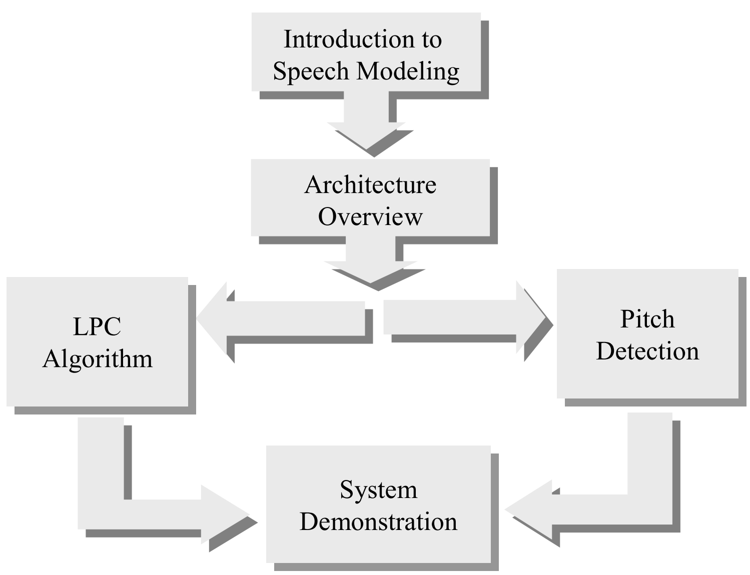
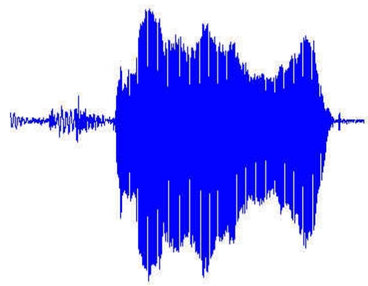
Implementation of Linear Predictive Coding (LPC) of Speech

213A class project
Spring 2000
Jean François Frigon and Vladislav Teplitsky

Outline

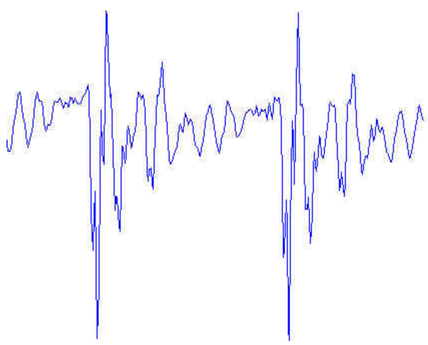


Speech Modeling – Non-stationary



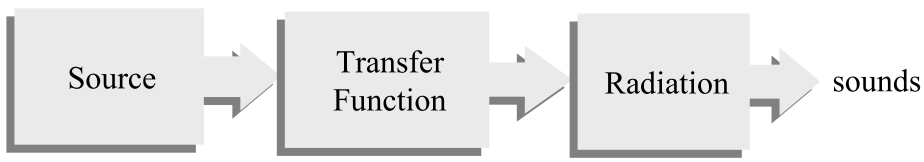
- Speech is a highly non-stationary signal
- Dynamically changes over time
- Changes occur very quickly

Speech Modeling – Frame Blocking



- Need to analyze the signal over many short segments, called frames
- Apply a short-duration (usually 20-30 msec) overlapping window (usually Hamming) to the speech signal in order to segment into frames
- A single frame of speech is stationary – perform analysis

Speech Modeling – LTI Model



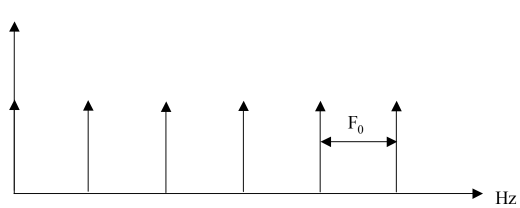
LTI Model is valid for

- moderately loud sounds
- short speech segments – frames (20 – 30 msec)

Implementation of LPC

Speech Modeling – Source (Voiced)

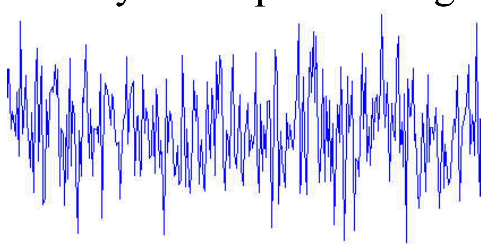
- Sounds are either **voiced** or **unvoiced**
- Voiced (e.g. all vowels) sounds are generated by vocal cords’ vibrations
- These vibrations are periodic in time, thus are approximated by an impulse train
- Spacing between impulses is the **pitch**, F_0



Implementation of LPC

Speech Modeling – Source (Unvoiced)

- Unvoiced sounds (e.g. /sh/, /s/, /p/) are generated without vocal cords' vibrations
- The excitation is modeled by a White Gaussian Noise source
- Unvoiced sounds have no pitch since they are excited by a non-periodic signal



Speech Modeling – Transfer Function

- Transfer function models the effects of the vocal tract on the source signal
- Transfer function is either all-pole (vowel model) or pole-and-zero (consonant model)
- Poles of the transfer function – resonances of the vocal tract - are called **formants**
- Human auditory system is much more sensitive to poles than to zeroes of the transfer function

Speech Modeling – Transfer Function

- We will consider only an all-pole transfer function of the form:

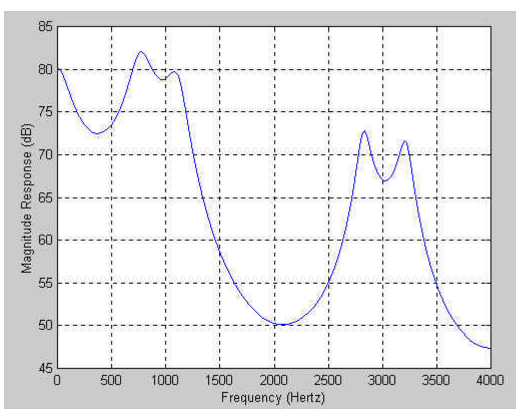
$$H(z) = \frac{G}{\prod_{i=1}^{p} (1 - a_i)(1 - a_i^*)}$$

- where G is the gain, p is the order (number of poles), and a_i is the pole.
- p ≈ 2×Bandwidth of signal (in kHz)+[2,3,4]
- e.g. BW=4 kHz, then

  p = 2×4 + [2,3,4] ∈ [10,11,12]

Speech Modeling – Transfer Function

- Example: a 10th order transfer function model:



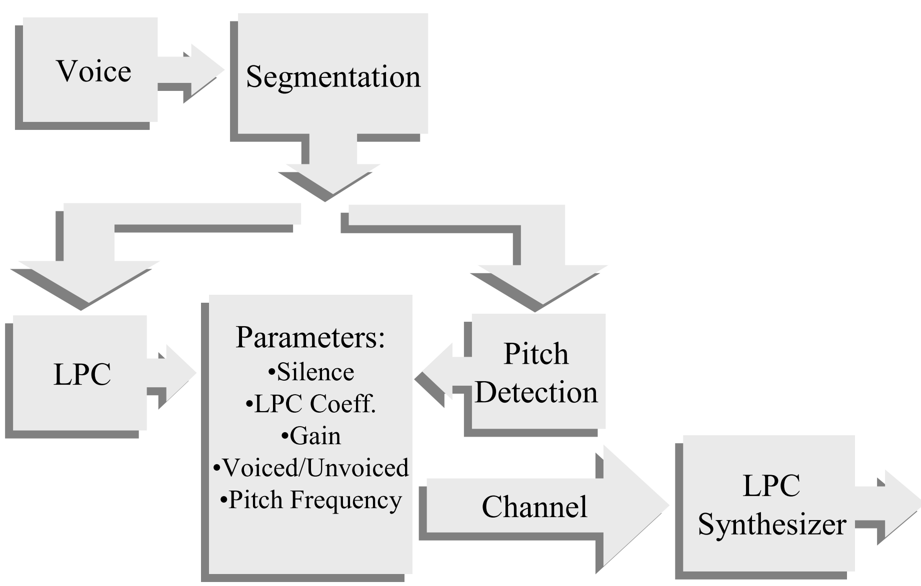
Speech Modeling – Radiation

- Models how sound is radiated by the lips
- Usually approximated by a digital differentiator:

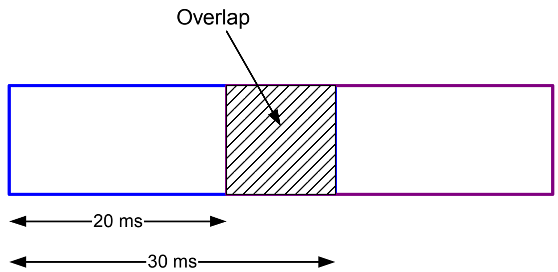
$$R(z) = 1 - z^{-1}$$

- Radiation is not important for classification of a sound
- Thus, we will omit it from our implementation

Architecture Overview



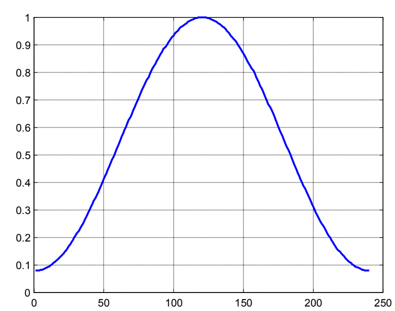
Voice Segmentation

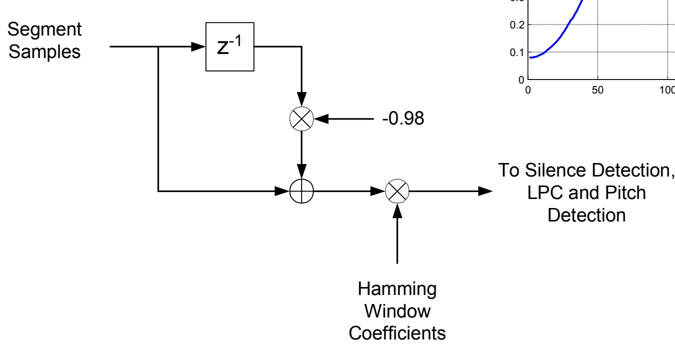


- 8000 samples/sec
- 20 ms step size (160 samples)
- 30 ms window (240 samples)
- Process 240 samples in 20 ms

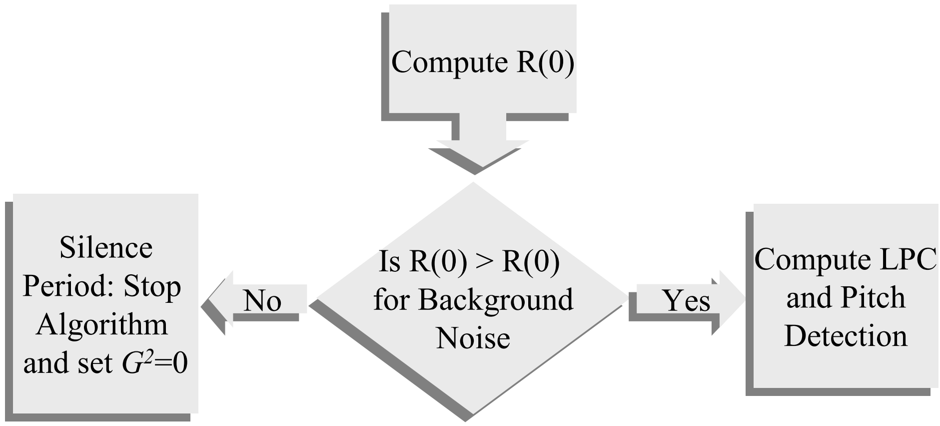
Voice Segmentation - Filtering and Windowing

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right)$$
$$n = 0 \ldots N-1$$





Voice Segmentation - Silence Detection



LPC - Motivation

Speech Difference Equation for a p^{th} order filter:

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + Gu(n)$$

Want to minimize the mean-squared prediction error:

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

For a single input impulse or stationary white noise, the obtained coefficients are identical to the a_k's

LPC - Autocorrelation (1)

If we assume that *s(n)* is zero outside the interval $0 \leq n \leq N-1$, we then need to solve the following set of linear equations:

$$\sum_{k=1}^{p} \alpha_k R(|i-k|) = R(i) \qquad 1 \leq i \leq p$$

Where:

$$R(k) = \sum_{m=0}^{N-1-k} s(m)s(m+k)$$

LPC - Autocorrelation (2)

In matrix form the set of linear equation can be expressed as:

$$\begin{bmatrix} R(0) & R(1) & R(2) & \dots & R(p-1) \\ R(1) & R(0) & R(1) & \dots & R(p-2) \\ R(2) & R(1) & R(0) & \dots & R(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ R(p-1) & R(p-2) & R(p-3) & \dots & R(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \dots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ R(3) \\ \dots \\ R(p) \end{bmatrix}$$

LPC - Levinson-Durbin Algorithm (1)

By exploiting

•Toeplitz structure of the matrix;

•Particular structure of the right-hand side of the linear system of equation

We can use the efficient Levinson-Durbin recursive procedure to solve this particular system of equations.

LPC - Levinson-Durbin Algorithm (2)

The Levinson-Durbin recursive procedure is given by:

$$\mathrm{E}^{(0)} = R(0)$$

for $1 \le i \le p$

$$k_i = \left[R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j) \right] \Big/ E^{(i-1)}$$

$$\alpha_i^{(i)} = k_i$$

for $1 \le j \le i-1$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}$$

$$E^{(i)} = \left(1 - k_i^2\right) E^{(i-1)}$$

The final solution is given by: $\alpha_j = \alpha_j^{(p)} \quad 1 \le j \le p$

LPC - Gain Coefficient

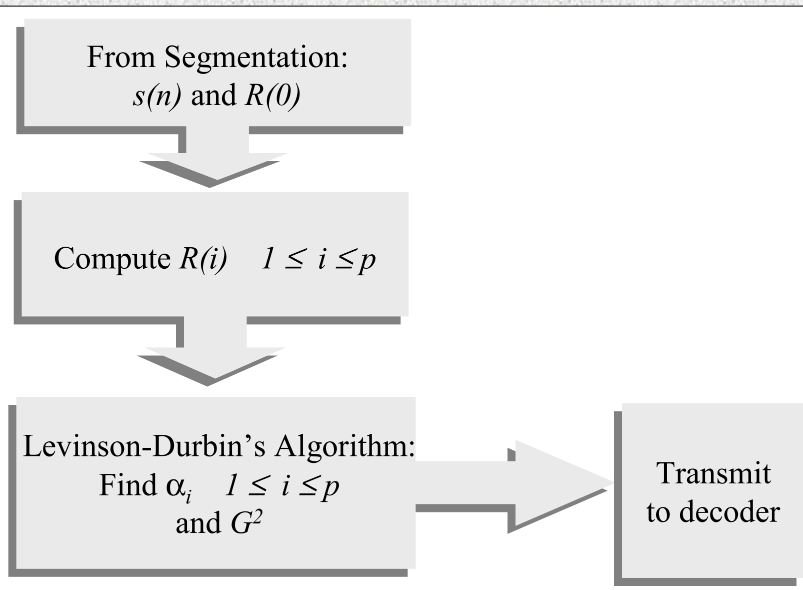
It can be shown that the gain coefficient is given by:

$$G^2 = R(0) - \sum_{k=1}^{p} \alpha_k R(k) = E_n$$

Where E_n is the minimum mean squared error prediction and is given by $E^{(p)}$ from Levinson-Durbin's Algorithm.

We will transmit G^2.

LPC Algorithm



Pitch Detection - Motivation

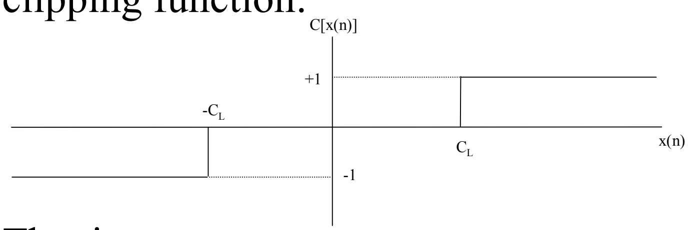
- Recall that source can be either a periodic impulse train spaced by F_0 or random noise
- Autocorrelation function of a speech frame:

$$R(k) = \sum_{m=0}^{N-k-1} x(m)x(m+k)$$

- If x(n) is periodic in N, then R(k) is also periodic in N
- Thus, we can compute R(k) and check if it's periodic

Pitch Detection – Motivation

- First we clip the frame using 3-level center clipping function:



- That is:

$$C[x(n)] = \begin{cases} +1 & \text{if} \quad x(n) > C_L \\ -1 & \text{if} \quad x(n) < -C_L \\ 0 & \text{otherwise} \end{cases}$$

Pitch Detection – Motivation

- Next we compute the modified autocorrelation function:

$$R_n(k) = \sum_{m=0}^{N-k-1} x(m)x(m+k)$$

- where $x(m)x(m+k)$ can have only 3 different values:

$$x(m)x(m+k) = \begin{cases} +1 & \text{if } x(m)=x(m+k) \\ -1 & \text{if } x(m)\neq x(m+k) \\ 0 & \text{if } x(m)=0 \text{ or } x(m+k)=0 \end{cases}$$

Pitch Detection – Motivation

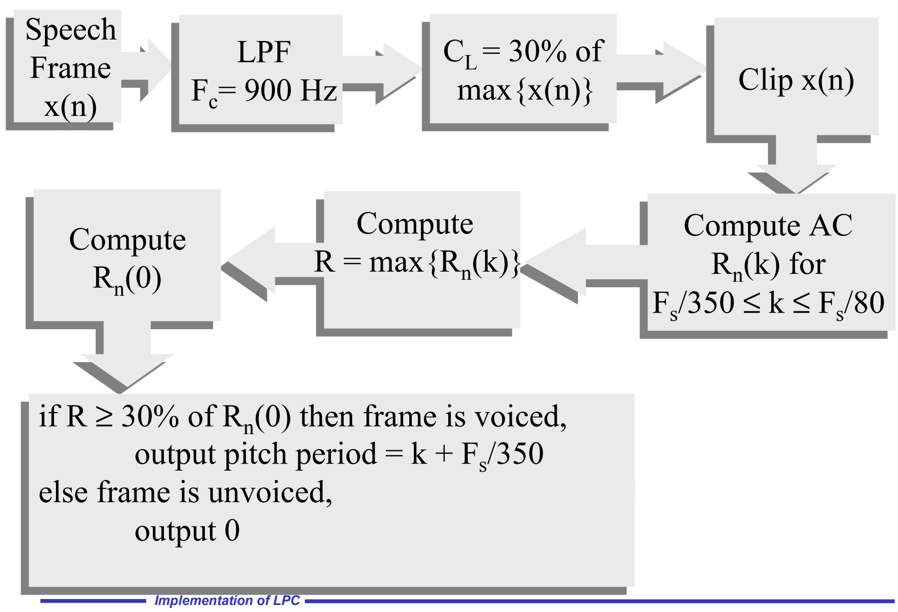
- We don't need to compute $R_n(k)$ for all values of k (i.e. $0 \le k \le N$)

	F_0 (Hz) min	F_0 (Hz) max
men	80	200
women	150	350

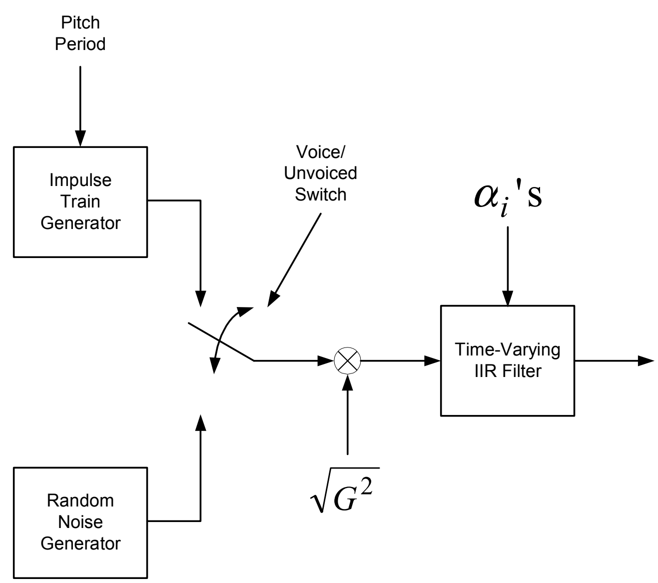
- Thus we only need to look in the range:

$$80 \text{ Hz} \le F_0 \le 350 \text{ Hz}$$

Pitch Detection – Algorithm



LPC Synthesizer



References

- L. R. Rabiner and R. W. Schafer. *Digital Processing of Speech Signals*. Prentice Hall, Englewood Cliffs, New Jersey, 1978.
- Douglas O'Shaugnessy. *Speech Communication Human and Machine.* Addison Wesley Books, 1978.
- M. M. Sondhi. *New Methods of Pitch Extraction.* IEEE Trans. Audio and Electroacoustics, Vol. AU-16, No. 2, pp. 262-266, June 1968.