3 Optimal and Adaptive Filtering

3.2: Linear Prediction

Optimal and Adaptive Filtering

3.2

1. **Wiener-Hopf filter**
    - Minimum Mean Square Error Estimation
    - The Wiener-Hopf solution
2. **Linear prediction**
    - The Wiener-Hopf filter as a predictor
    - Linear prediction for signal coding
3. **Adaptive filtering**
    - Steepest descend
    - Least Mean Square approach
4. **Applications of optimal and adaptive filtering**
    - …

Linear Prediction

1. **Introduction**
   - Modelling of a prediction problem
2. **The Wiener-Hopf filter as a predictor**
   - Problem specification
3. **Linear prediction for signal coding**
   - Coder/Decoder structure
   - Quantization of the prediction error
4. **Linear prediction coding of speech signals**
   - Speech signal characteristics
   - Short term and long term prediction
5. **Conclusions**

Introduction to signal prediction

Signal prediction: We estimate the value of a random signal at a given time instance ($x[n_0]$) , based on other time instance values (e.g.: $x[n_0 - 1], x[n_0 - 2], \cdots$) .

Design: We compare the current signal value ($x[n_0]$) with its estimation ($y[n_0]$)

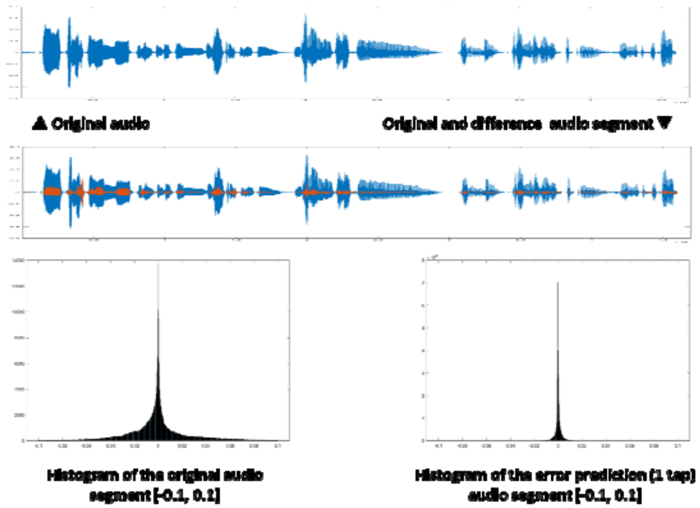
Use: The current signal value ($x[n_0]$) may not be available and we produce an estimation. If $x[n_0]$ is available, we produce the prediction error ($e[n_0]$)

The application assumes:

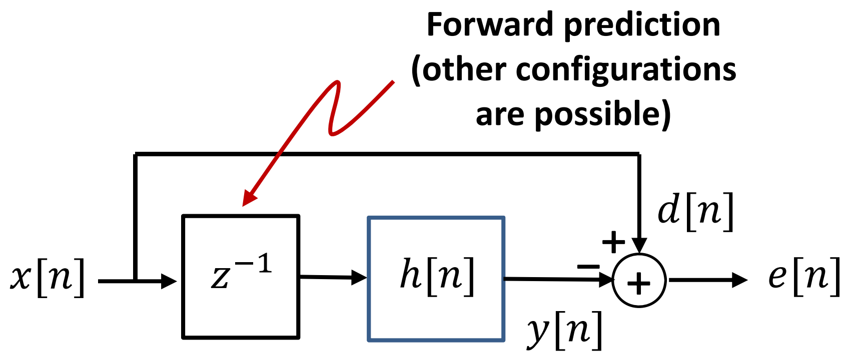
- Observations and reference belong to the same noisy process

Example of application:

- Speech coding and synthesis



The prediction error has a lower dynamic range and its quantization decreases the quantization noise power

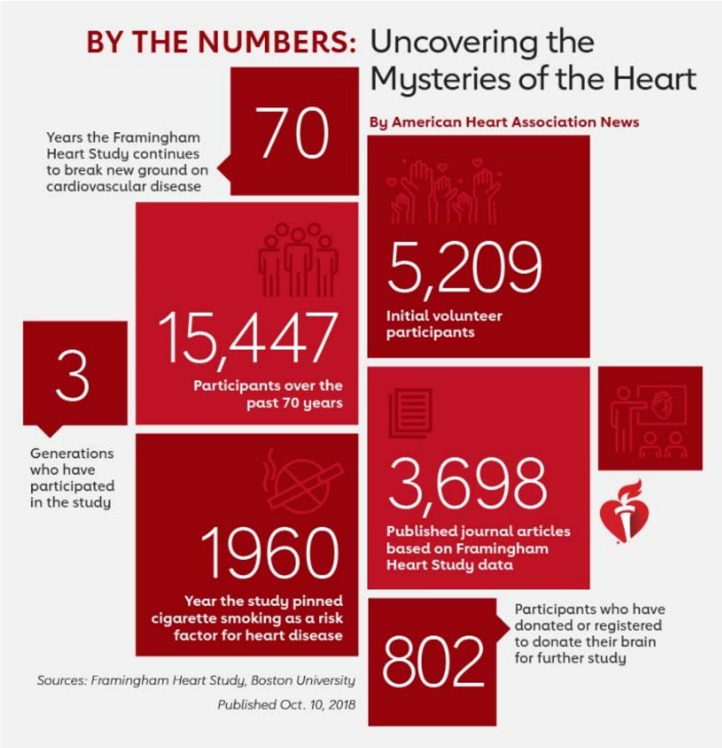


Introduction to signal prediction

3.2

However, these concepts can be extended to other scenarios and situations:

- **Framingham**: Cardiovascular disease study since 1948
- **Pasqual Maragall Foundation**: Alzheimer study since 2008



Major findings from the Framingham Heart Study, according to the researchers themselves:

1960s:
- Cigarette smoking increases risk of heart disease.
- Increased cholesterol and elevated blood pressure increase risk of heart disease.

1970s:
- Elevated blood pressure increases risk of stroke.
- Psychosocial factors affect risk of heart disease.

1980s:
- High levels of HDL cholesterol reduce risk of heart disease.

1990s:
- Having an enlarged left ventricle of the heart increases risk of stroke.

https://framinghamheartstudy.org/

https://fpmaragall.org/en/

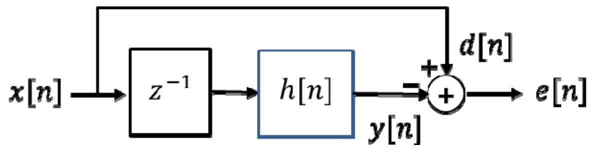
Different prediction configurations

In the context of **linear prediction**, we can define three possible scenarios:

- **Forward prediction**: The current sample is estimated using only previous samples:
  - Forecasting a given parameter value
- **Backward prediction**: The current sample is estimated using only future samples:
  - "Remembering" a given value. Implies delay.
- **Linear smoothing** (or interpolation): The current sample is estimated combining past and future samples:
  - Recovering a damaged signal

Note: Commonly, in signal processing applications, what it is important is **the ability to obtain a good estimation** of a sample, pretending that is known, rather than forecasting it:

- Coding applications



The delay denotes that previous samples are used; that is, we perform a **forward prediction**

Table Tennis frames #40, #41 and #42

Linear Prediction

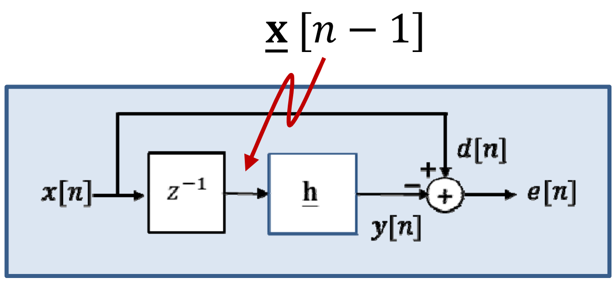
3.2

1. Introduction
   - Modelling of a prediction problem
2. **The Wiener-Hopf filter as a predictor**
   - Problem specification
3. **Linear prediction for signal coding**
   - Coder/Decoder structure
   - Quantization of the prediction error
4. **Linear prediction coding of speech signals**
   - Speech signal characteristics
   - Short term and long term prediction
5. **Conclusions**

The Wiener-Hopf filter as a predictor

Let us analyze the FIR Wiener-Hopf filter in the context of **forward prediction**. Let us assume that we want to predict a given stationary process ($s[n]$). In that case:

- **Reference**: $d[n] = s[n]$
- **Observations**: $\underline{\mathbf{x}}[n] = \underline{\mathbf{s}}[n-1]$



Previous samples of the observation signal have been **buffered** to be used in the prediction

With this scenario, the Wiener-Hopf solution implies:

$$\boxed{\underline{\mathbf{h}}_{opt} = \underline{\underline{\mathbf{R}}}_x^{-1}\underline{\mathbf{r}}_{xd}}$$

$$\underline{\mathbf{r}}_{xd} = E\{\underline{\mathbf{s}}[n-1]s[n]\} = \begin{bmatrix} E\{s[n-1]\,s[n]\} \\ E\{s[n-2]\,s[n]\} \\ \cdots \\ E\{s[n-N]\,s[n]\} \end{bmatrix} = \begin{bmatrix} r_s[-1] \\ r_s[-2] \\ \cdots \\ r_s[-N] \end{bmatrix} = \underline{\mathbf{r}}_s[-1]$$

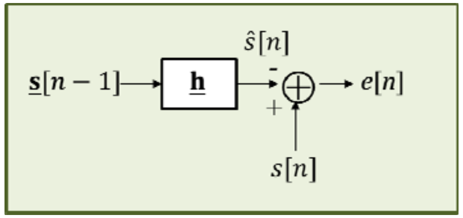
◀ **Cross-correlation vector**

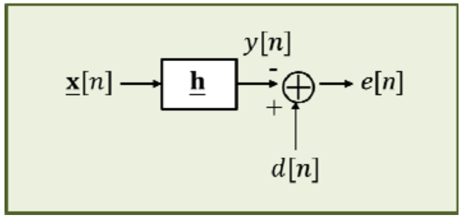
$$\underline{\underline{\mathbf{R}}}_x = E\{\underline{\mathbf{s}}[n-1]\,\underline{\mathbf{s}}^T[n-1]\} = \begin{bmatrix} r_s[0] & r_s[1] & & r_s[N-1] \\ r_s[-1] & r_s[0] & & r_s[N-2] \\ \cdots & \cdots & \cdots & \cdots \\ r_s[-N+1] & r_s[-N+2] & & r_s[0] \end{bmatrix} = \underline{\underline{\mathbf{R}}}_s$$

◀ **Correlation matrix**

The Wiener-Hopf filter as a predictor

3.2





Relation between variables in both schemes:

$d[n] \Rightarrow s[n]$:	reference signal	$\Rightarrow$	current sample
$\underline{\mathbf{x}}[n] \Rightarrow \underline{\mathbf{s}}[n-1]$:	N data samples	$\Rightarrow$	N previous samples
$\underline{\mathbf{h}} \Rightarrow \underline{\mathbf{h}}$:	filter (N taps)	$\Rightarrow$	predictor filter (N taps)
$y[n] \Rightarrow \hat{s}[n]$:	filtered signal	$\Rightarrow$	current predicted sample
$e[n] \Rightarrow e[n]$:	prediction error	$\Rightarrow$	prediction error

- When the optimal filter is used:
  - Error is “orthogonal” to data: $$E\{\underline{\mathbf{s}}[n-1]e[n]\} = \underline{0}$$
  - The power of the error is lower than the power of the reference signal: $$E\{s^2[n]\} \geq E\{e^2[n]\}$$
  - The minimum error power is $$\varepsilon = r_s[0] - \underline{\mathbf{h}}^T{}_{opt}\,\underline{\mathbf{r}}_s$$
- The expression for the optimal filter is $$\underline{\underline{\mathbf{R}}}_s \cdot \underline{\mathbf{h}}_{opt} = \underline{\mathbf{r}}_s$$
- The power of the error, for any filter ($\underline{\mathbf{h}}$) is $$E\{e^2[n]\} = \varepsilon + (\underline{\mathbf{h}} - \underline{\mathbf{h}}_{opt})^T \underline{\underline{\mathbf{R}}}_s (\underline{\mathbf{h}} - \underline{\mathbf{h}}_{opt})$$

Linear Prediction

3.2

1. **Introduction**
   - Modelling of a prediction problem
2. **The Wiener-Hopf filter as a predictor**
   - Problem specification
3. **Linear prediction for signal coding**
   - Coder/Decoder structure
   - Quantization of the prediction error
4. **Linear prediction coding of speech signals**
   - Speech signal characteristics
   - Short term and long term prediction
5. **Conclusions**

Linear Prediction Coding (LPC)

3.2

- Assuming **stationarity**, the Wiener-Hopf filter minimizes the MSE between the process and its estimation (MMSE: **minimum error power**):
  - Signals are processed by (close to) stationary segments: **frames** (in speech coding, typically 20 ms)

- The power of the error is lower than the power of the reference signal. That allows defining a **coding gain** (G_c):

$$\sigma_s^2 = E\{s^2[n]\} \geq E\{e^2[n]\} = \sigma_e^2 \quad \Rightarrow \quad \boxed{G_c = \frac{\sigma_s^2}{\sigma_e^2}}$$

- Given a filter different from the optimal one (e.g.: the **quantized filter** ($\underline{\mathbf{h}}_q$)), the obtained error power and actual coding gain can be computed:

$$\boxed{E\{e^2[n]\} = \varepsilon + (\underline{\mathbf{h}}_q - \underline{\mathbf{h}}_{opt})^T \underline{\underline{\mathbf{R}}}_s (\underline{\mathbf{h}}_q - \underline{\mathbf{h}}_{opt})}$$

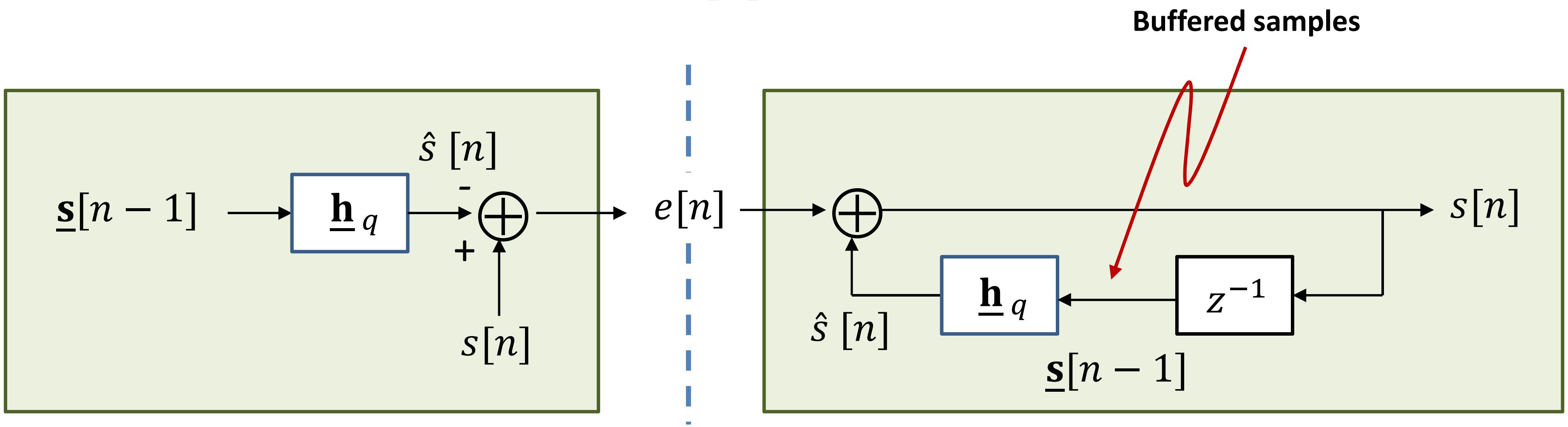
LPC: Coder / Decoder structure

3.2

- For each frame (assumed to be a stationary signal), the decoder receives **the filter** that has been used for predicting the signal and **the prediction error**.

$s[n]$: current sample
$\underline{\mathbf{s}}[n-1]$: N previous samples
$\underline{\mathbf{h}}$: predictor filter (N taps)
$\hat{s}[n]$: current predicted sample
$e[n]$: prediction error

- ❑ Assuming **amplitude-discrete signals** ($s[n]$, $\hat{s}[n]$, $e[n] \in \mathbb{Z}$), the receiver can reconstruct the original signal ($s[n]$) without loss.



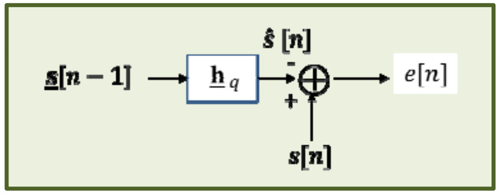
- **Internal variables are kept** when starting processing a new frame.

LPC: Coder / Decoder structure

3.2

CODER

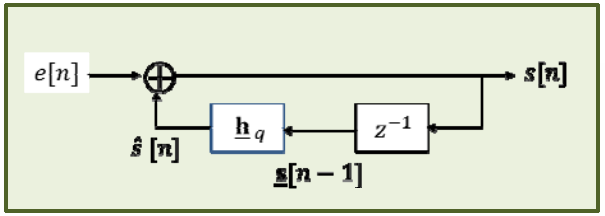
n=0



n=1

DECODER

n=0



n=1

Linear Prediction

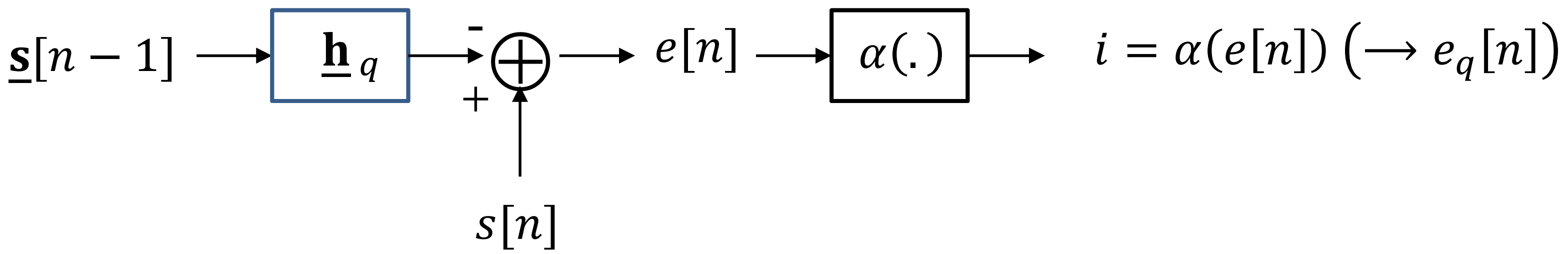
3.2

1. **Introduction**
   - Modelling of a prediction problem
2. **The Wiener-Hopf filter as a predictor**
   - Problem specification
3. **Linear prediction for signal coding**
   - Coder/Decoder structure
   - Quantization of the prediction error
4. **Linear prediction coding of speech signals**
   - Speech signal characteristics
   - Short term and long term prediction
5. **Conclusions**

Quantization of the prediction error

3.2

- So far, we have concentrated on the **quantization** of values that come directly from a **signal** (voice, audio, image) or are **model coefficients** (filter coefficients).
- Should the same strategy be used in the case of **quantizing prediction error samples**?
- The **coding scheme** in that case can be the following:



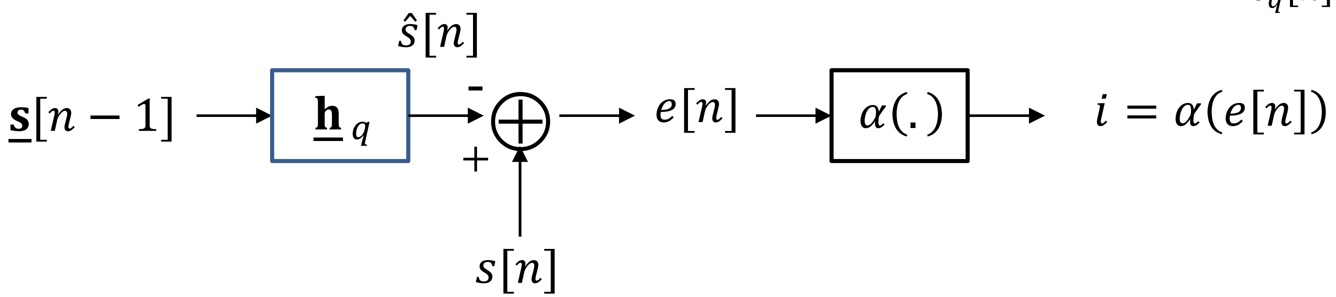
In this situation, how does the decoder work?

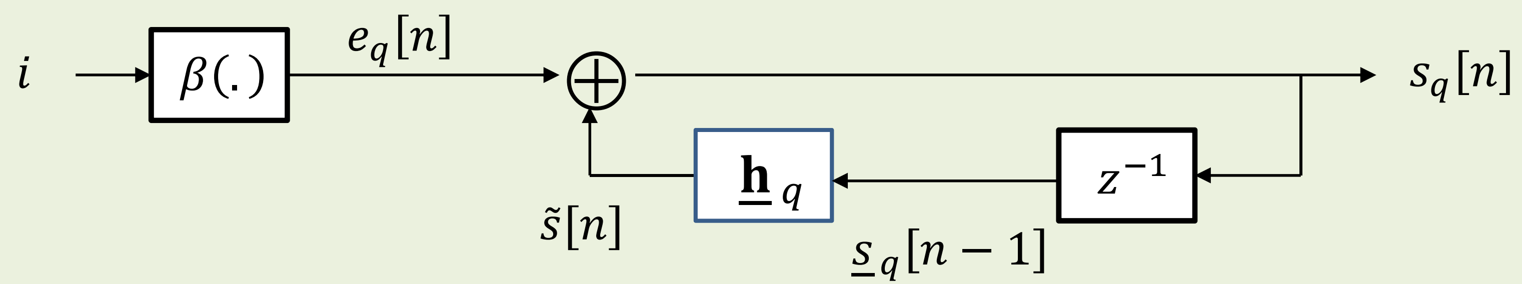
Quantization of the prediction error

3.2

Predictive coding/decoding systems: $e[n] = e_q[n] + \varepsilon_q[n]$

$e[n]$: prediction error
$e_q[n]$: quantized error
$\varepsilon_q[n]$: quantization error



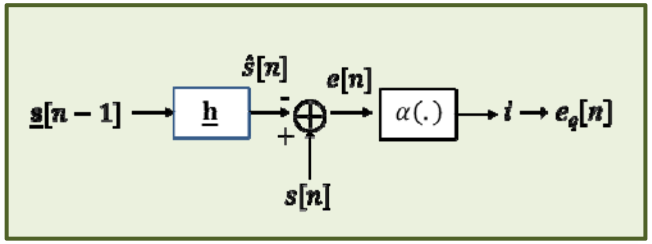


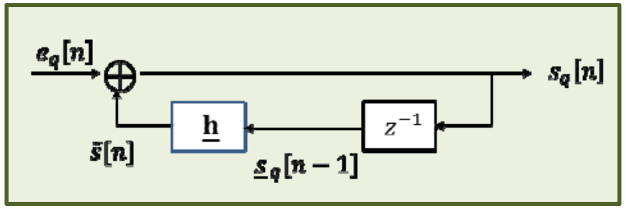
- For simplicity, let us assume that the exact values of the N filter coefficients are available at the receiver side ($\underline{\mathbf{h}}$) and so are the first N samples of the signal ($\underline{\mathbf{s}}\,[N-1]$).

Quantization of the prediction error

3.2

CODER $n = N$

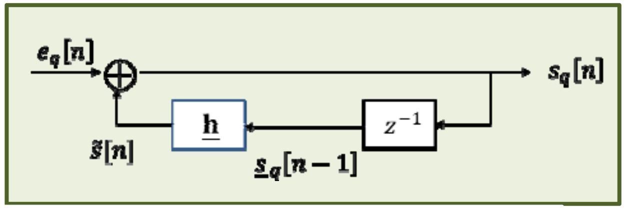




DECODER $n = N$

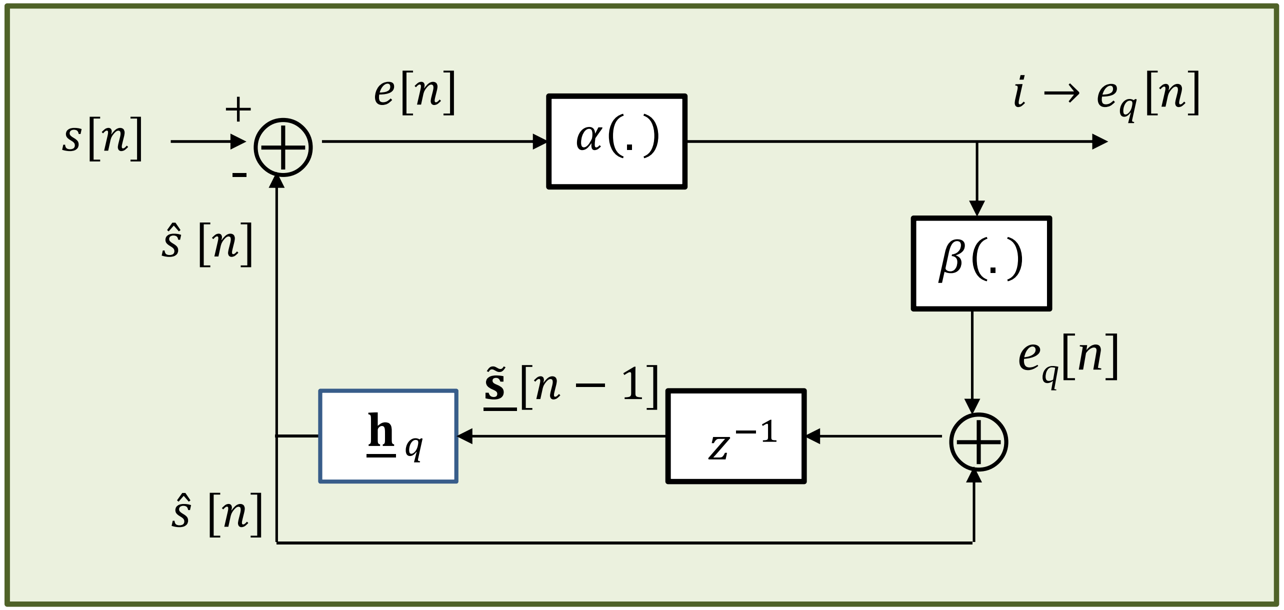
Quantization of the prediction error

CODER $n = N+1$



Encoder with an embedded decoder

At the encoder, the current sample is predicted using the previous samples as they are available at the decoder side; that is, that have been computed taking into account **the prediction error**.



In a predictive coder, the encoder and the decoder have to work with the same samples (**decoded samples**), to control the **propagation of the quantization error**.

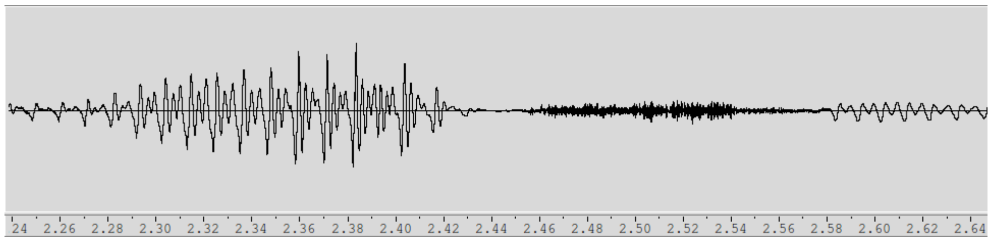
Linear Prediction

1. **Introduction**
   - Modelling of a prediction problem
2. **The Wiener-Hopf filter as a predictor**
   - Problem specification
3. **Linear prediction for signal coding**
   - Coder/Decoder structure
   - Quantization of the prediction error
4. **Linear prediction coding of speech signals**
   - Speech signal characteristics
   - Short term and long term prediction
5. **Conclusions**

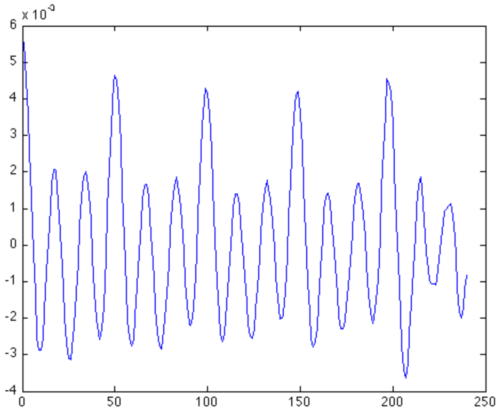
Temporal redundancy in speech signals

3.2

In speech signals, there is usually a **high temporal correlation** (similarity) **between consecutive (or close) samples** that can be appreciated in the signal itself, its (estimated) autocorrelation or its (estimated) spectral density.

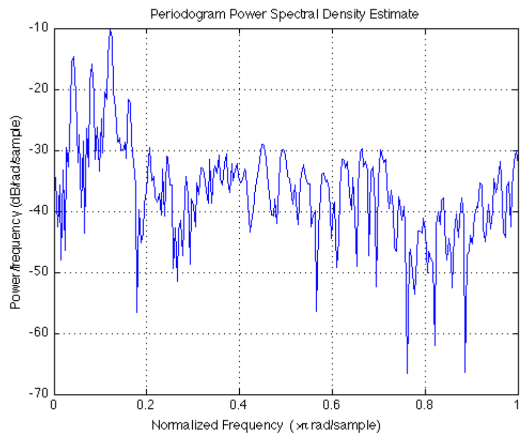


Speech signal: Voiced and Unvoiced parts



◄ **Autocorrelation of a frame of voiced signal**

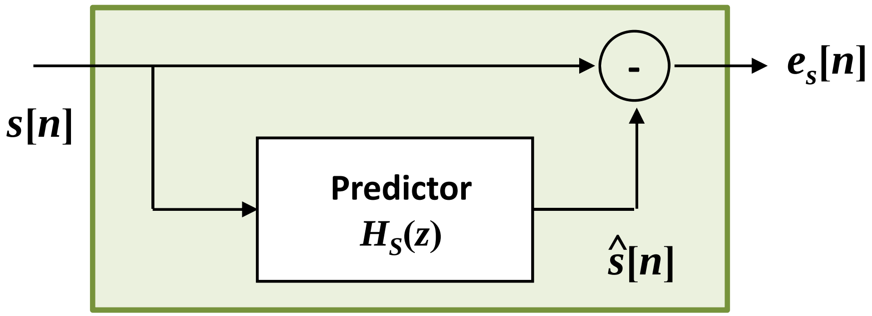
Spectral density of a frame of voiced signal ►



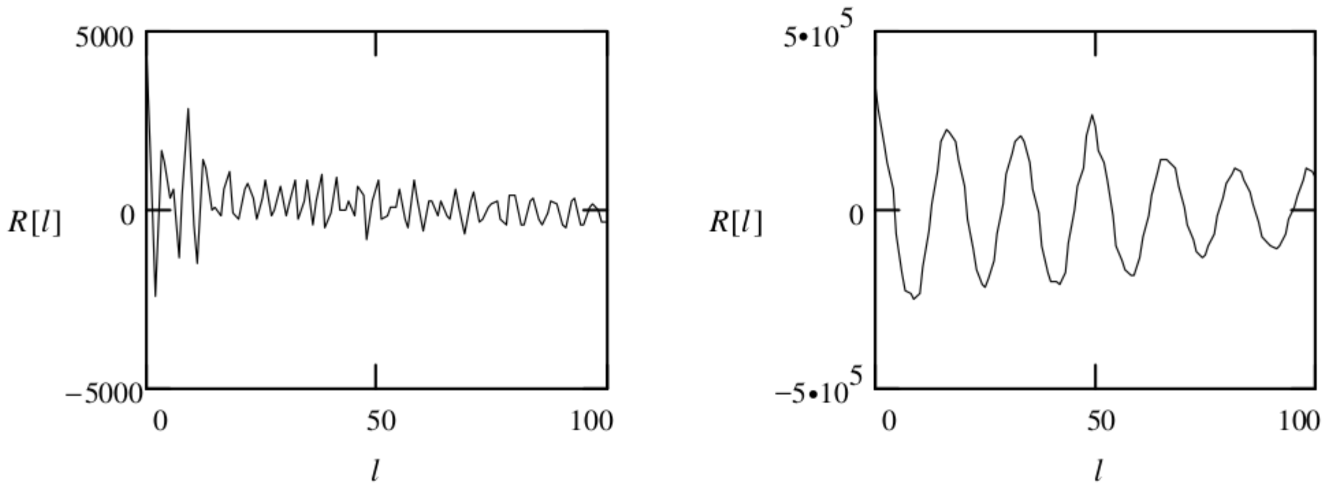
Temporal redundancy in speech signals

3.2

Linear prediction leads to **higher prediction gains** in the case of voiced signals than unvoiced signals.

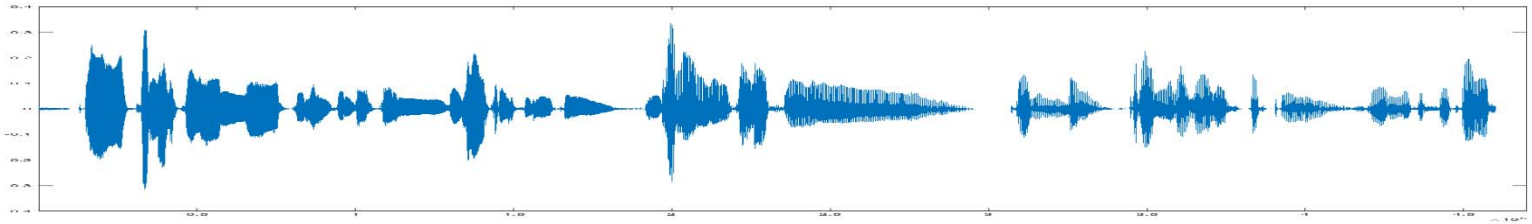


$$H_S(z) = \sum_{i=1}^{N} a_i z^{-i}$$



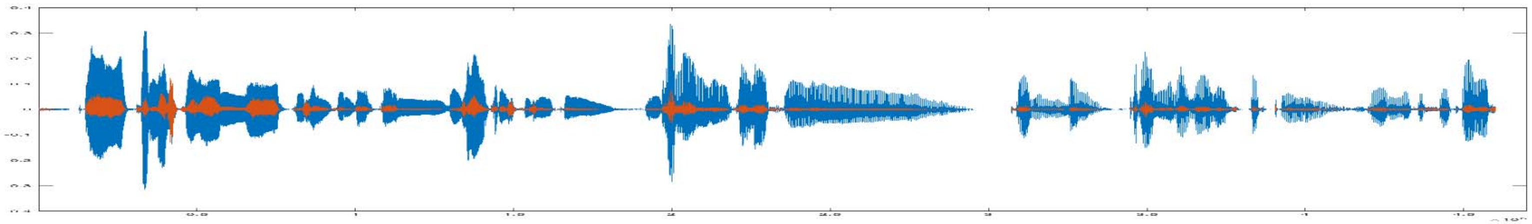
Correlation of an unvoiced and a voiced signal

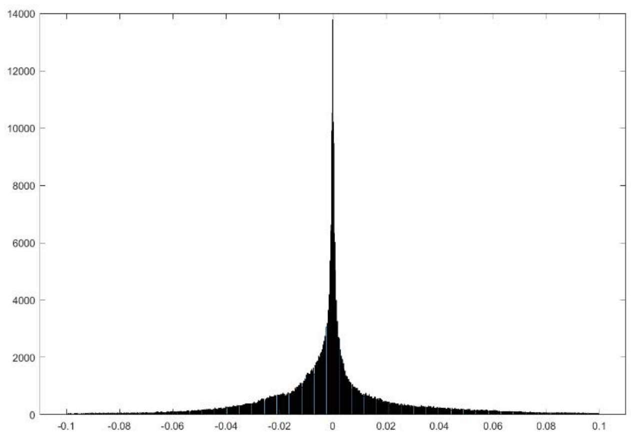
Temporal redundancy in speech signals



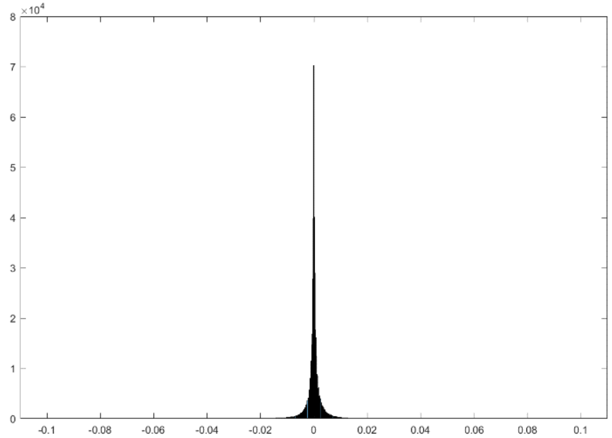
▲ Original audio

Original and difference audio segment ▼





Histogram of the original audio segment [-0.1, 0.1]



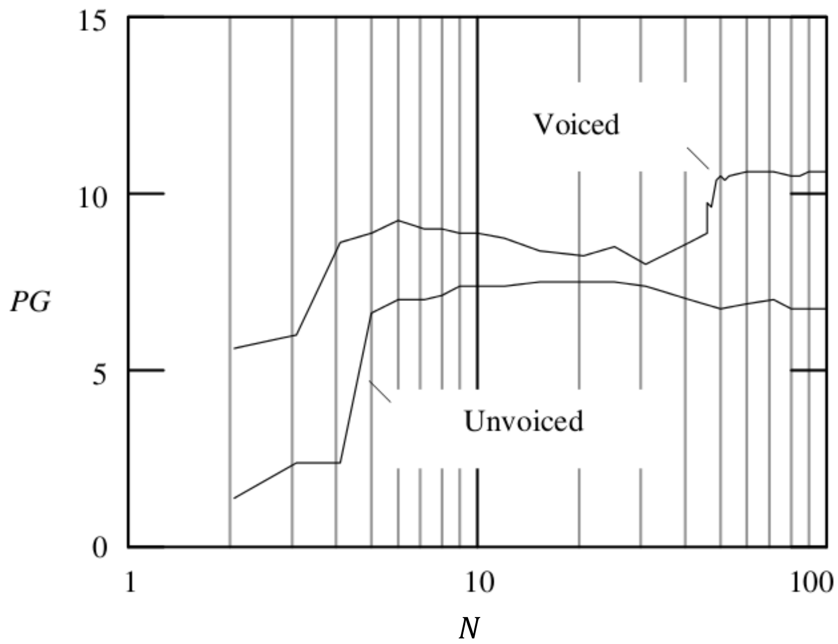
Histogram of the error prediction (1 tap) audio segment [-0.1, 0.1]

Temporal redundancy in speech signals

3.2

A large increase in the performance comes from the fact of including in the predictor **nearby samples** ($N = 8 - 10$) as well as samples that are near to **one pitch period apart** (T).

- However, samples in the **middle of this range** do not substantially improve the prediction gain.



◂ **Prediction gain (PG) when increasing the predictor order (N) for two given realizations (voiced/unvoiced)**

Linear Prediction

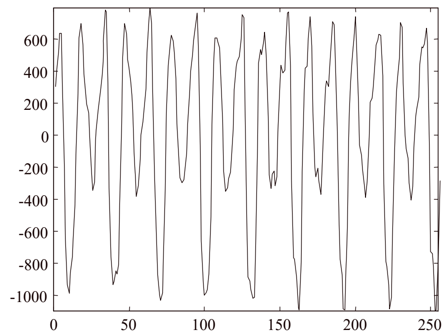
3.2

1. **Introduction**
   - Modelling of a prediction problem
2. **The Wiener-Hopf filter as a predictor**
   - Problem specification
3. **Linear prediction for signal coding**
   - Coder/Decoder structure
   - Quantization of the prediction error
4. **Linear prediction coding of speech signals**
   - Speech signal characteristics
   - Short term and long term prediction
5. **Conclusions**

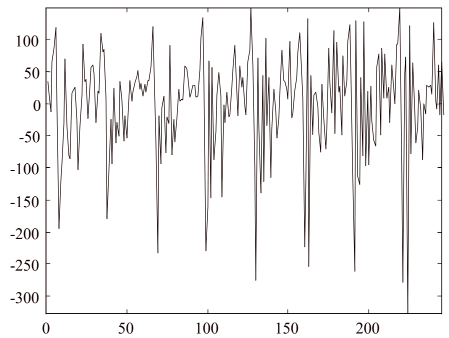
The prediction error

When using a filter predictor that takes into account the closest samples to the current one (**short term prediction**), the information about the periodicity of the signal is not exploited.

- In voiced speech signals, the (short term) prediction error presents a periodicity at one period pitch distance



Voiced speech signal



Short term prediction error

In order to improve the prediction, a **long term predictor** is concatenated to the previous short term predictor:

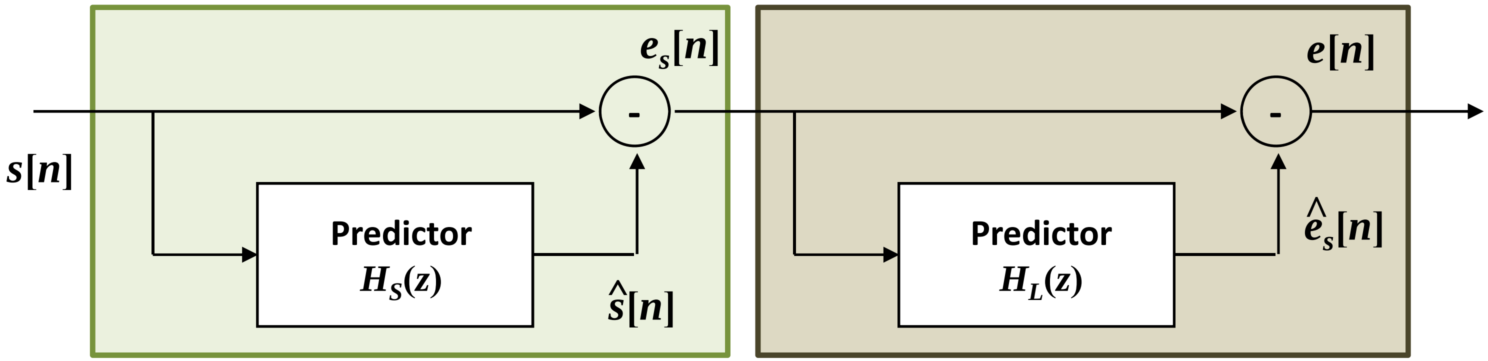
- The power of the error decreases and the error spectrum is closer to a white one.

Short term and long term prediction

3.2

The redundancy present in voiced signals due to their close-to-periodical nature can be exploited using a **long term predictor** in cascade with the previous short term one.

- The **intermediate samples** are not used in the prediction

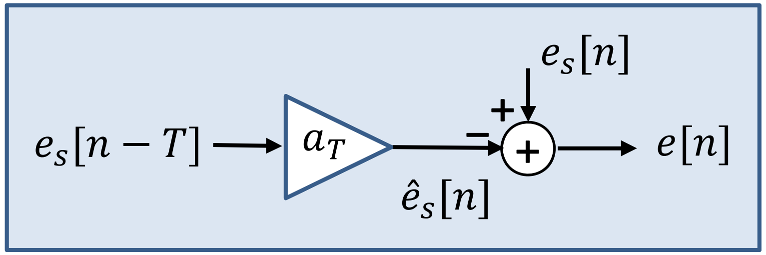


$$H_S(z) = \sum_{i=1}^{N} a_i z^{-i} \qquad H_L(z) = \sum_{i=T-1}^{T+1} a_i z^{-i}$$

- The long term predictor usually has **1 – 3 coefficients**.

Long term prediction: Pitch estimation

When **only one sample is used**, the algorithm to find the parameters of the long term predictor (T, a_T) implies a (simple) Wiener-Hopf like optimization.



$$e[n] = e_s[n] - \hat{e}_s[n] = e_s[n] - a_T e_s[n-T]$$

- Obtain the value of the single-tap long term predictor parameters (T, a_T) as a minimization of the MSE over the available data (one frame of speech signal or, actually, one fourth of a frame)

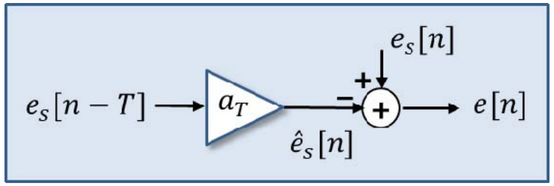
$$J(a_T, T) = \sum_n e^2[n] = \sum_n (e_s[n] - \hat{e}_s[n])^2 = \sum_n (e_s[n] - a_T e_s[n-T])^2$$

$$\frac{\partial}{\partial a_T} J(a_T, T) = -2 \sum_n (e_s[n] - a_T e_s[n-T]) e_s[n-T] = 0$$

Long term prediction: Pitch estimation

3.2

$$\frac{\partial}{\partial a_T} J(a_T, T) = \sum_n (e_s[n] - a_T e_s[n-T]) e_s[n-T] = 0$$



$$\sum_n e_s[n] e_s[n-T] - a_T \sum_n e_s^2[n-T] = 0$$

$$a_{T_{opt}} = \frac{\sum_n e_s[n] e_s[n-T]}{\sum_n e_s^2[n-T]}$$

- The previous expression implies estimating two autocorrelation values
- The optimal value of a_T depends on the pitch value (T), and there is no close form to obtain the optimum T value.
- An **exhaustive search** of the pitch period (T) value leading to the minimum sum of squared error provides with the best pair of parameters

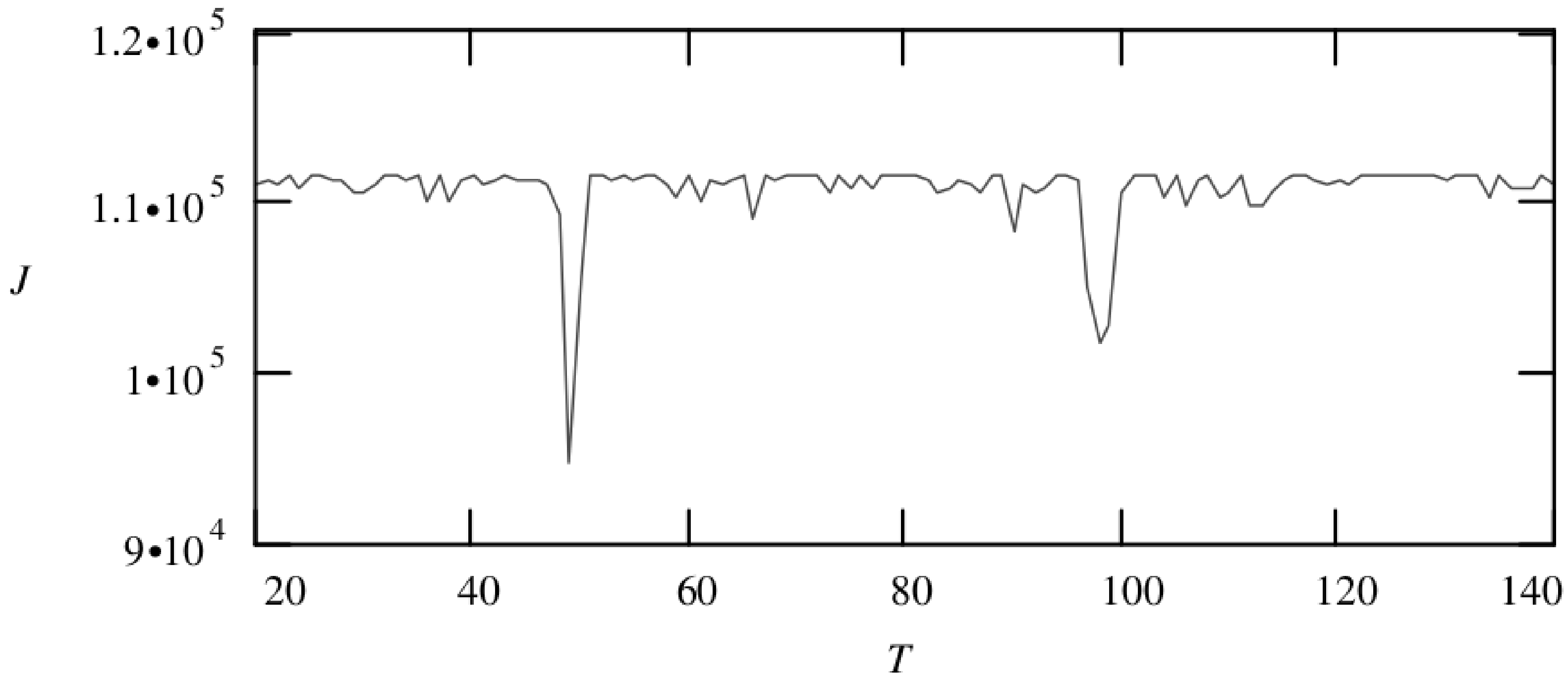
$$J(a_T, T) = \sum_n (e_s[n] - a_T e_s[n-T])^2$$

$$J(a_T, T) = \sum_n e_s^2[n] - \frac{(\sum_n e_s[n] e_s[n-T])^2}{\sum_n e_s^2[n-T]}$$

Long term prediction: Pitch estimation

An example of the evolution of the sum of squared error (J) as a function of the pitch period (T):

- **Typical values** for the pitch period (T) are $20 \leq T \leq 140$

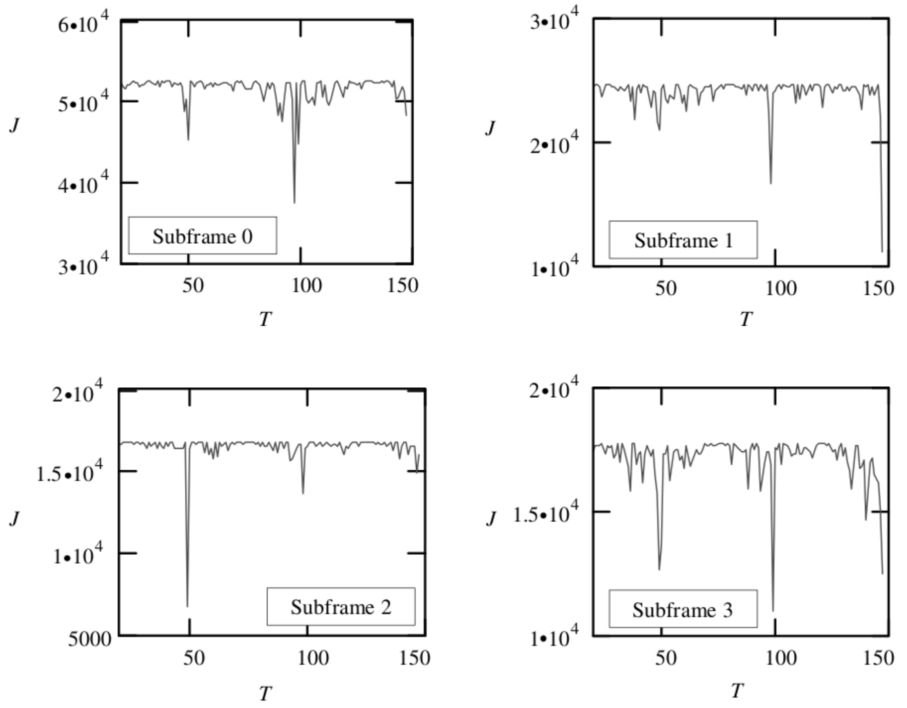


Wai C. Chu, *Speech Coding Algorithms*, John Wiley & Sons 2003.

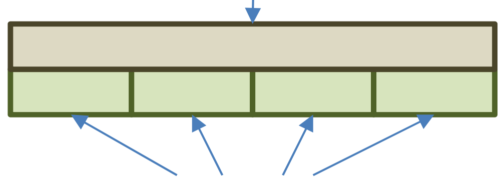
Long term prediction: Pitch estimation

3.2

- The pitch period **varies much faster** than the coefficients of the filter. Therefore, the computation of the pitch period over a whole frame does not provide a good estimation.
- Frames are subdivided into **sub-frames** and the pitch period is independently estimated in each sub-frame.



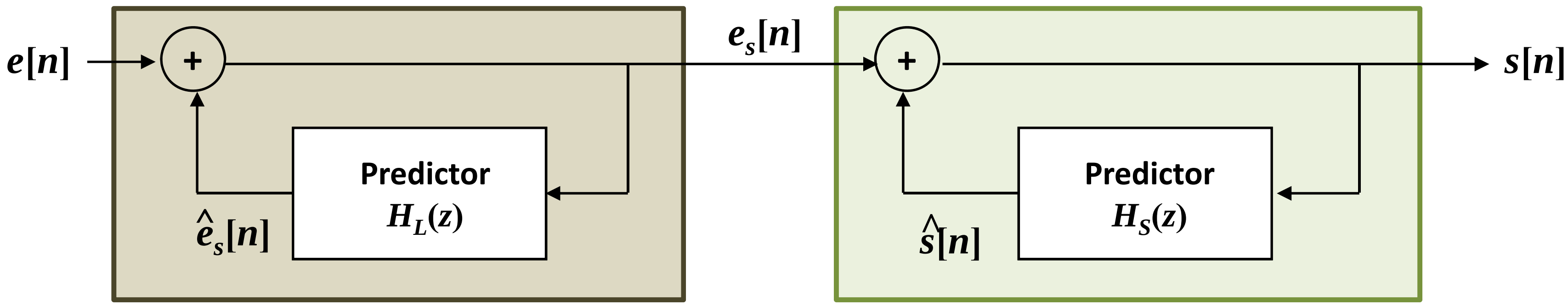
Result of estimating the pitch period in four consecutive sub-frames.



Wai C. Chu, *Speech Coding Algorithms*, John Wiley & Sons 2003.

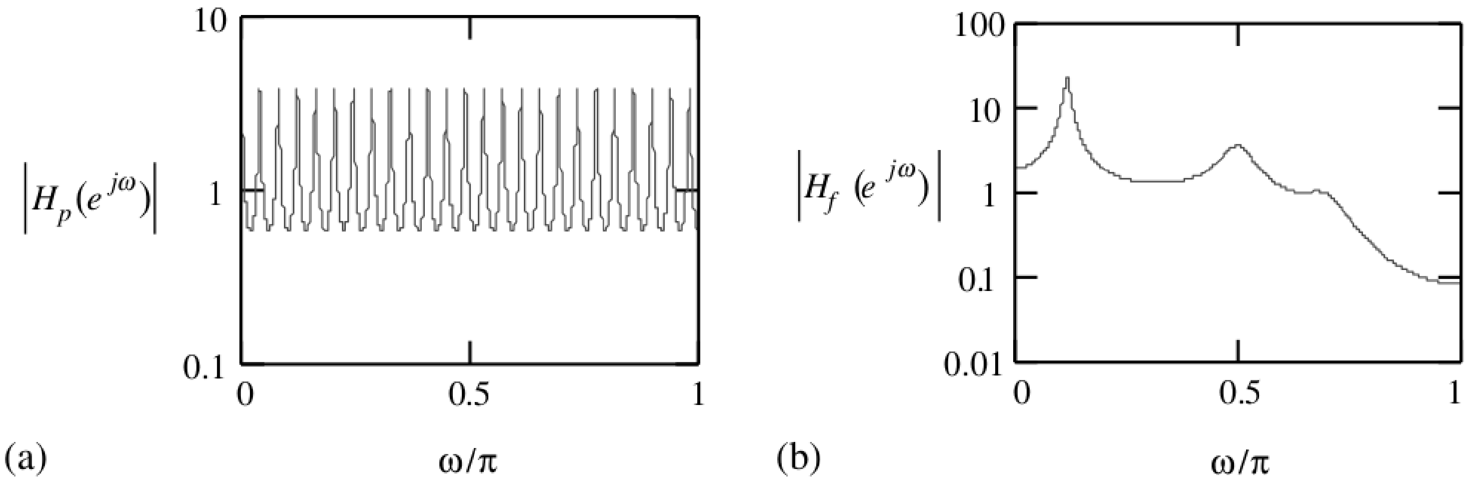
Analysis of the decoder side

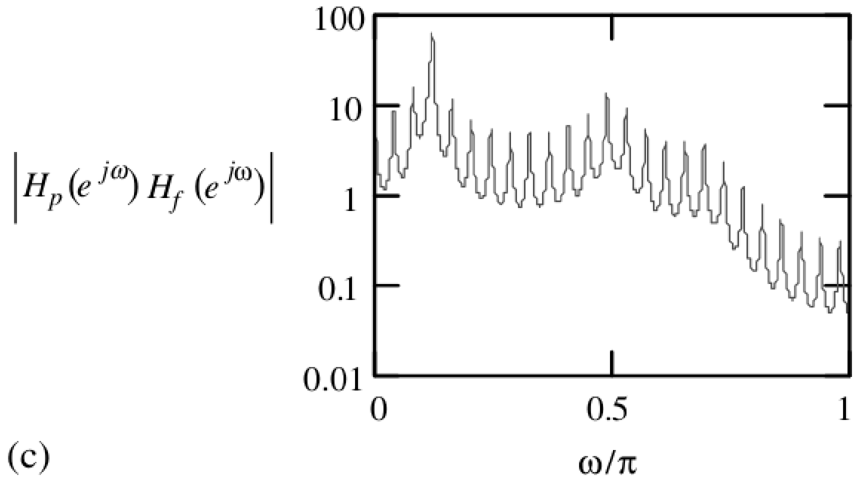
- The synthesis of the speech signal at the receiver side implies the implementation of the **inverse filters**.
- Quantization of the **short term and long term filter coefficients** has to ensure **filter stability** and **transparent quantization**.



$$H_L(z) = \frac{1}{1 - \sum_{i=T-1}^{T+1} a_i z^{-i}} \qquad H_s(z) = \frac{1}{1 - \sum_{i=1}^{N} a_i z^{-i}}$$

Analysis of the decoder side





Combination of a short term and a long term pair of synthesis filters. The result is a cascade connection of a pitch synthesis filter and a formant synthesis filter that reproduces the power spectrum of a voiced signal.

Wai C. Chu, *Speech Coding Algorithms*, John Wiley & Sons 2003.

The GSM 6-10 speech coding standard

GSM 6.10 Full-Rate

13 Kbit/s Regular-Pulse Excitation

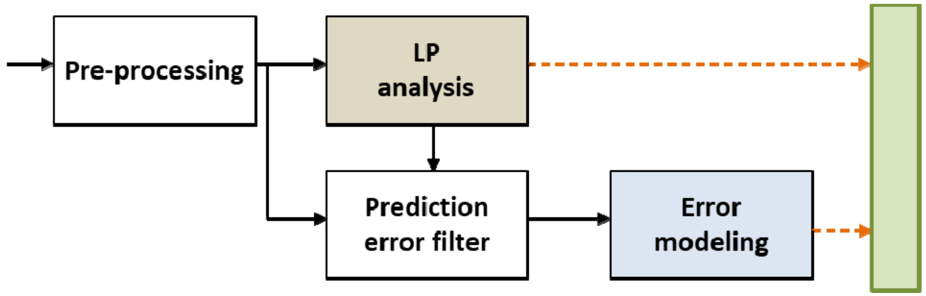
Published (1991) as an ETSI standard (ETS 300 961)

- **Sampling frequency**: 8 Khz
- **Bitrate**: 13 Kbit/s
- **Frame length**: 20 ms (160 samples)
- **Typical algorithmic delay**: 25 ms
- **PSQM testing under ideal conditions**: MOS of 4.1
- **Main use**: First digital speech coding standard used in the GSM digital mobile phone system. Still widely used around the world.
- **Other characteristics**:
  - Vocal tract modeled using a **8th order** (short-term) predictor.
  - Excitation estimated on sub-frames of **5 ms**

The GSM 6-10 speech coding standard

Pre-processing:

- Apply pre-emphasis filters on 20 ms frames



Short term prediction:

- Compute **short-term 8-order LPC**.
  - Transform reflection coefficients in log-area (**LAR**) before quantization
  - ➢ **Shorter** than current ones: Lower quality
- Quantize using uniform quantizers, with **specific range for each coefficient**.
  - #Bits per frame = 6 + 6 + 5 + 5 + 4 + 4 + 3 + 3 = **36 bits**
- Generate 4-sets of linearly interpolated LAR to be applied in different parts of the frame.
  - Better adaptation to **non-stationary signals**

The GSM 6-10 speech coding standard

Long term prediction: 5 ms sub-frame

- Compute **delay** (**pitch**) using autocorrelation method.
  - Pitch range: 40 - 120 samples $\rightarrow$ **7 bits**.
- Compute and quantize **prediction gain**: **2 bits**.

Quantization of residual signal (excitation): 5ms sub-frames

- **Regular-Pulse Excitation:** 40 samples **decimated** into 3 down-sampled subsequences of 14, 13 and 13 samples.
  - The first one is split into 2 ones of 13 samples each.
- **4 possible** decimated **sequences**:
  - Select the one with **higher energy** (**2 bits**)
- **APCM quantization** of the 13 values:
  - Maximum value: **6-bits log-quantizer**.
  - Normalized samples: **3 bits uniform** mid-rise quantizer.

The GSM 6-10 speech coding standard

3.2

Computing the bit rate:

- #Bits per frame for the **short-term 8-order LPC**:
  - 6 + 6 + 5 + 5 + 4 + 4 + 3 + 3 = **36 bits**
- #Bits per sub-frame for the **Pitch range**:
  - 40 - 120 samples $\rightarrow$ **7 bits**.
- #Bits per sub-frame for the **Prediction gain**:
  - **2 bits**.
- #Bits per sub-frame for the **Residual signal**:
  - 4 possible decimated sequences $\rightarrow$ **2 bits**.
- #Bits per sub-frame for the **APCM quantization**:
  - Maximum value $\rightarrow$ **6 bits.**
  - Normalized 13 samples $\rightarrow$ **3 bits**.

Total amount per frame: 36 + 4*(7+2) + 4*(2 + 6 + 3*13) = 260 bits

Total Bitrate (frame length: 20 ms): 260 bits/frame * 50 frame/s = 13Kbits/s

Linear Prediction