Digital Speech and Audio Coding

Lecture 3

Short-Time Fourier Transform, Systems, Multi-rate Signal Processing

Mathew Magimai Doss and Petr Motlicek

Idiap Research Institute, Martigny

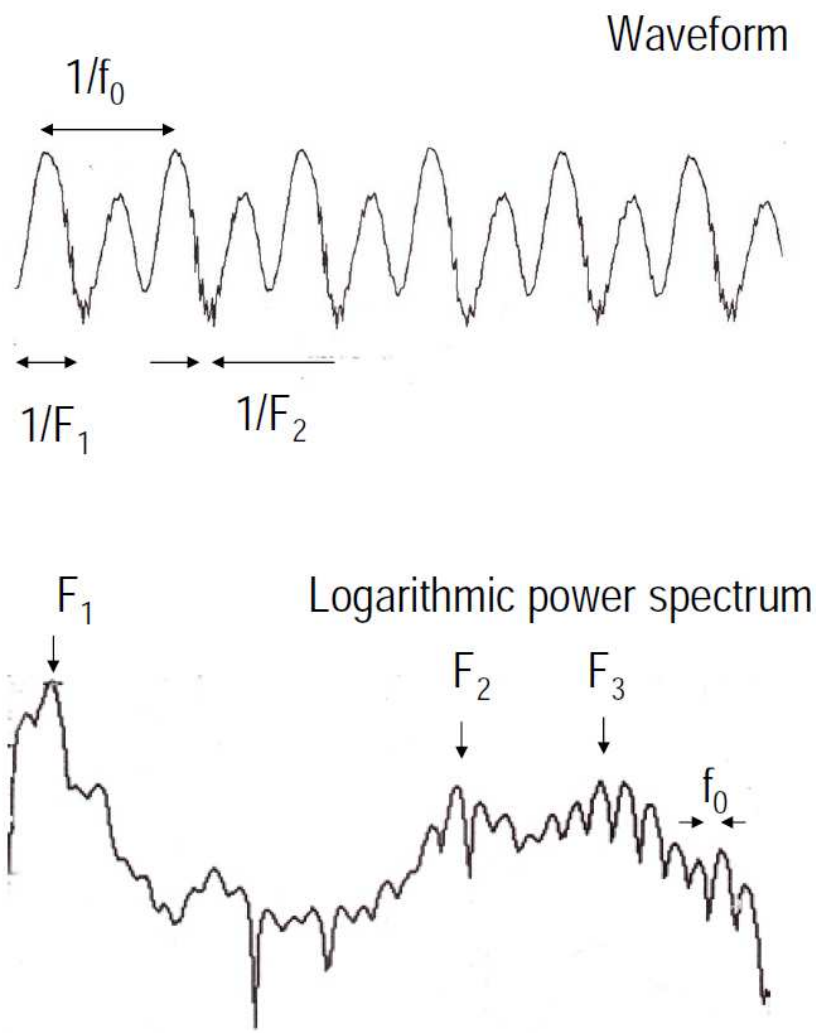
http://www.idiap.ch/

Ecole Polytechnique Fédérale de Lausanne, Switzerland

Previous lecture

- Signals
- Frequency Analysis
- Sampling . . . Nyquist theorem
- quantization - linear - logarithmic, scalar - vector

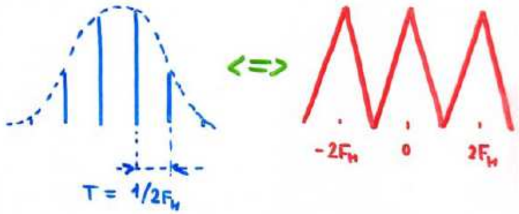
Time-frequency plot for speech



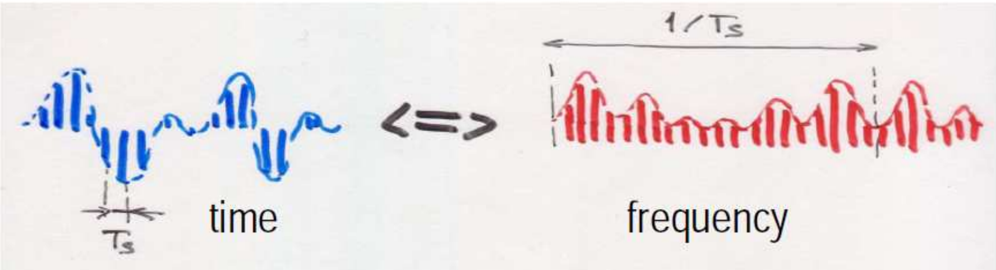
Periodicity

Periodicity in one domain implies discrete values in another domain

Sampling in one domain implies periodicity in another domain



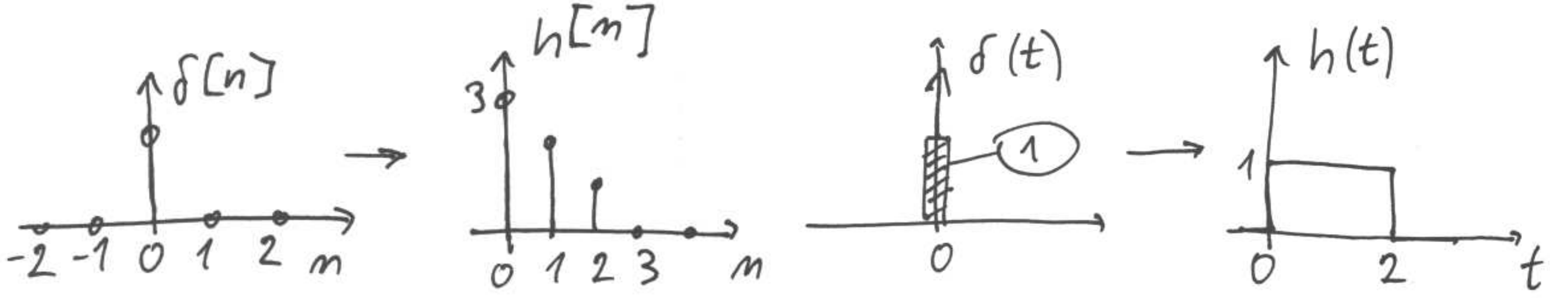
Windowed (periodic) discrete signal has always discrete and periodic spectrum



Discrete-time systems

LTI (Linear Time Invariant systems):

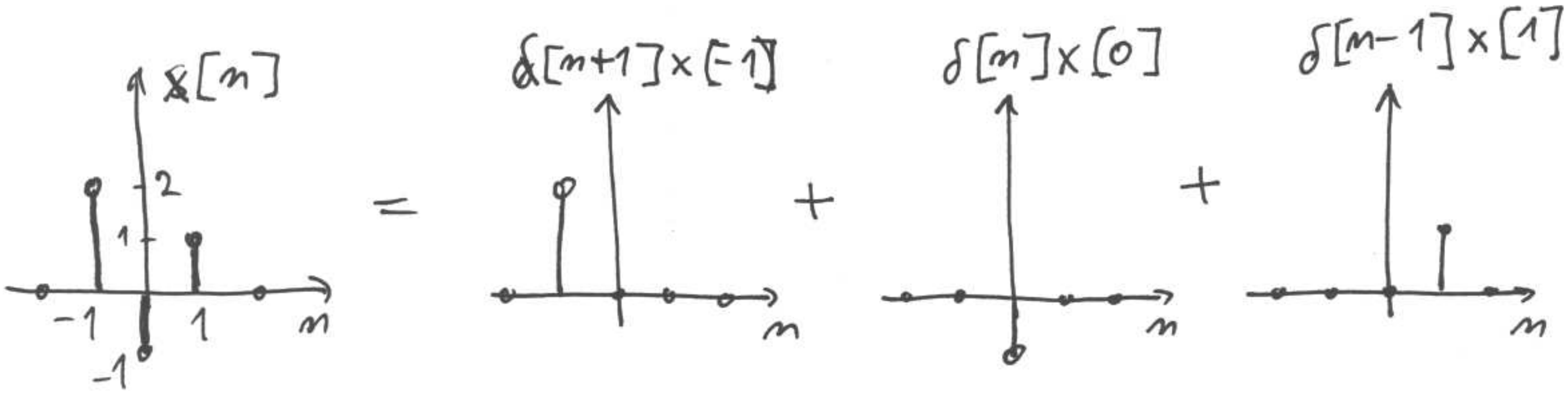
- Most important characteristics: impulse response - how systems react to an unit impulse
- However, we are more interested how the system reacts to any kind of signal $x[n]$. We can decompose $x[n]$ to set of unit impulses:



Discrete-time systems

Decomposition of signal into discrete unit impulses:

$$x[n] = \sum_{k=-\infty}^{+\infty} x[k]\delta[n-k]$$



Reaction of the system on the shifted unit impulse: $\delta[n-k]$ can be denoted to as $h_k[n]$. If the system is an LTI system, then all $h_n[k]$ are similar and/but shifted in time: $h_k[n] = h[n-k]$. Each shifted impulse will start its $h_k[n]$.

Discrete-time systems

We then just sum all together:

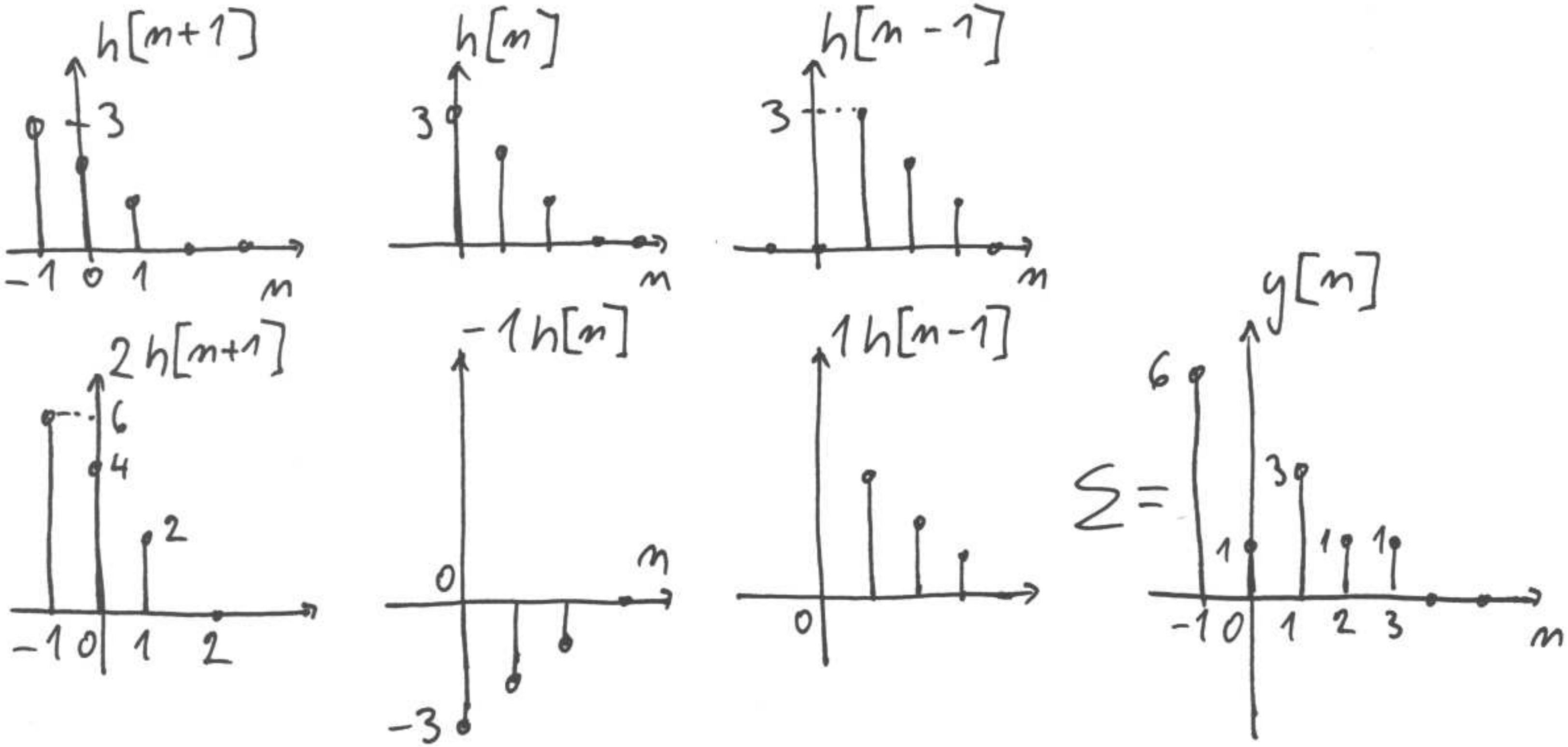
$$y[n] = \sum_{k=-\infty}^{+\infty} x[k]h[n-k]$$

which is often written as

$$y[n] = x[n] \star h[n] \iff \text{convolution}$$

Discrete-time systems

Example for $h[n]$ a $x[n]$:



Frequency characteristic of systems

Similarly to an unit impulse, the input signal will be the complex exponential:

$$x[n] = e^{j\omega_1 n},$$

with normalized angular frequency ω_1

$$y[n] = h[n] \star x[n] = \sum_{k=0}^{\infty} h[k] x[n-k] = \sum_{k=0}^{\infty} h[k] e^{j\omega_1 (n-k)} = e^{j\omega_1 n} \sum_{k=0}^{\infty} h[k] e^{-j\omega_1 k}.$$

The output signal contains also the input part multiplied by:

$$H(e^{j\omega_1}) = \sum_{k=0}^{\infty} h[k] e^{-j\omega_1 k}$$

and we can write:

$$y[n] = x[n] H(e^{j\omega_1})$$

Frequency characteristic of systems

For arbitrary frequency, we get (complex) frequency charasteristics:

$$H(e^{j\omega}) = \sum_{k=0}^{\infty} h[k]e^{-j\omega k}$$

We can notice that the frequency characteristics is a **DTFT-projection** of impulse response:

$$h[n] \longrightarrow H(e^{j\omega})$$

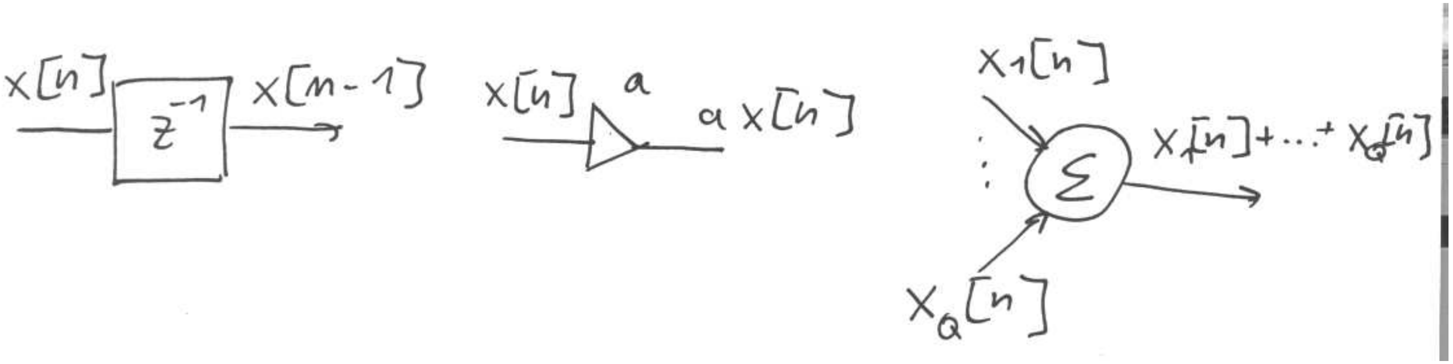
Properties:

- periodicity of spectra (also impulse response is a discrete signal!) – we should correctly mark $H(e^{j\omega})$ as $\tilde{H}(e^{j\omega})$:
- symmetry: $H(e^{j\omega}) = H^{\star}(e^{-j\omega})$

Fundamental blocks of systems:

input: $x[n]$, output $y[n]$, where n is a pointer to the sample (discrete time).

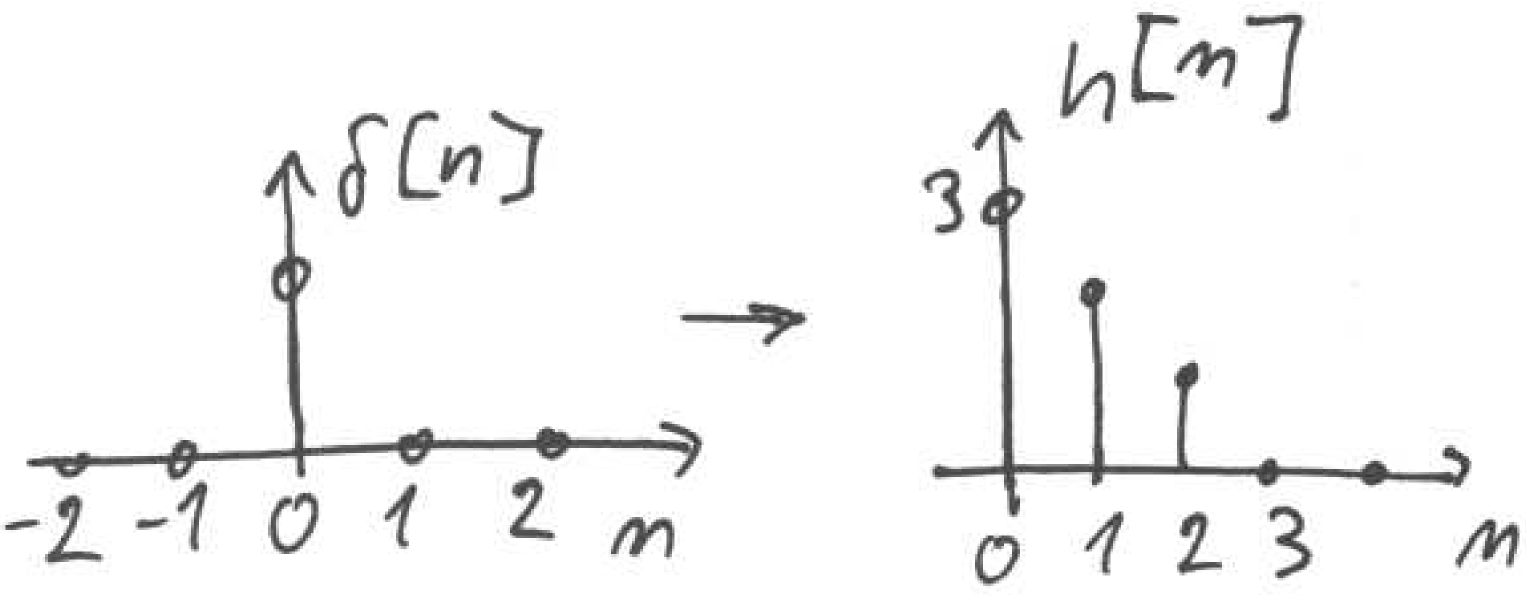
Fundamental blocks:



Fundamental blocks of systems:

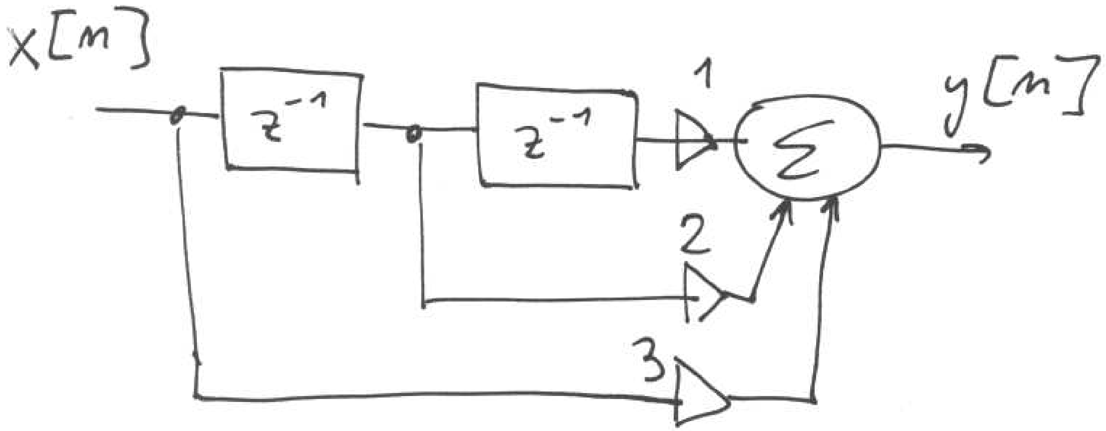
- Delay - will hold a sample for one sampling period and then it will be returned.
- multiplication - multiplies a sample by some coefficient.
- addition - ...

An example :



Fundamental blocks of systems:

can be built up:

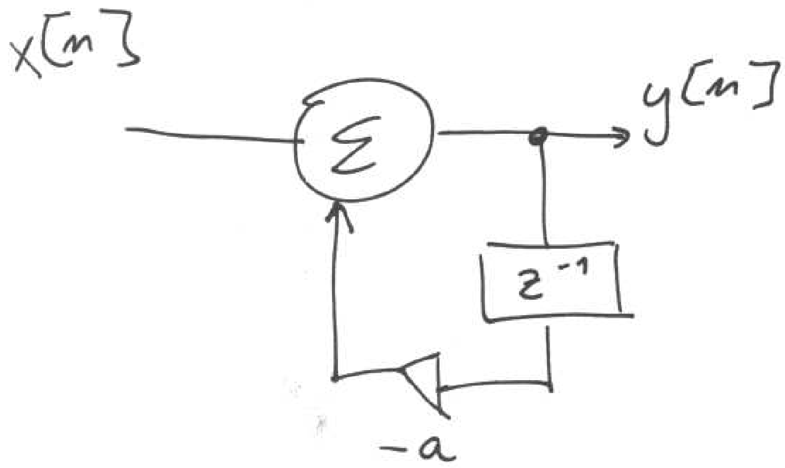


To prove it $\Longleftrightarrow$ check an impulse response.

Non-recursive and recursive systems

If the filter works only with an actual and delayed samples of an input signal – It's impulse response is **finite - finite impulse response** – **FIR** - non-recursive filters.

In case of **recursive** filters, we take into account also delayed samples of the output, for example:



Non-recursive and recursive systems

This filter has an impulse response:

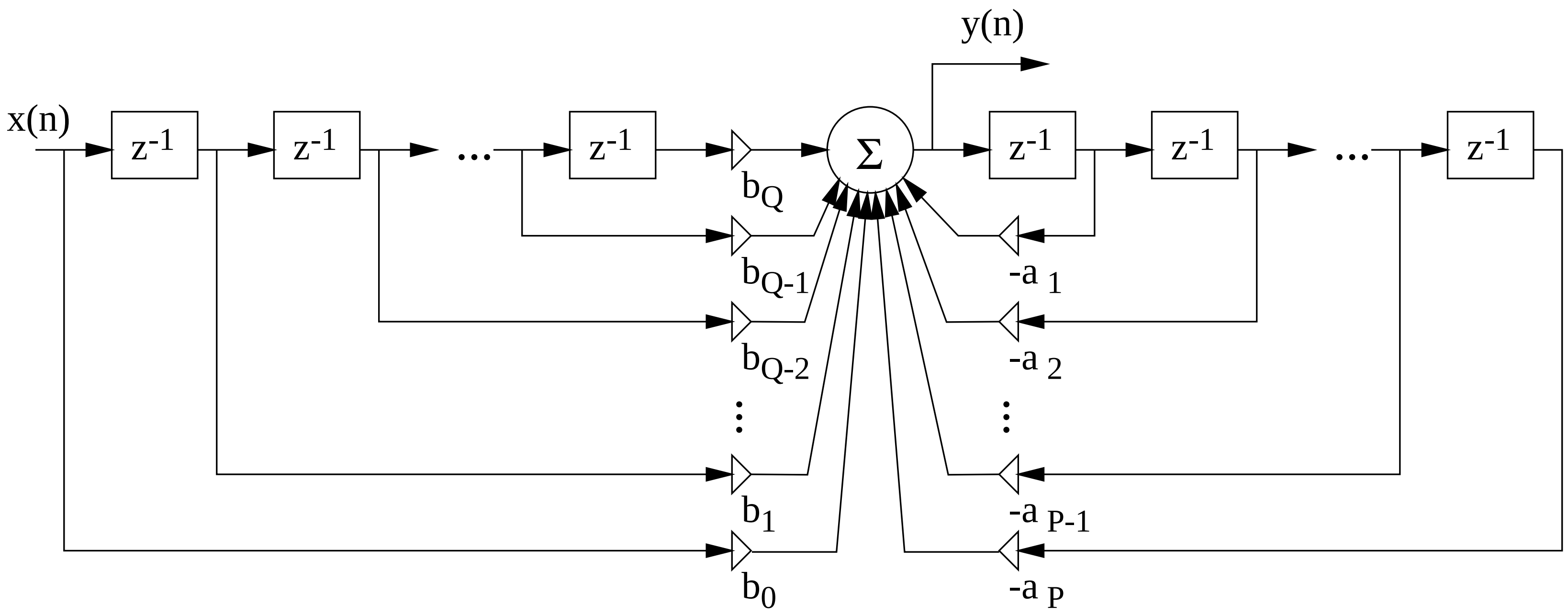
$$h[n] = \begin{cases} 0 & \text{for} \quad n < 0 \\ 1 \quad -a \quad (-a)^2 \quad (-a)^3 \ldots & \text{for} \quad n = 0, 1, 2, 3, \ldots \end{cases}$$

thus:

$$h[n] = \begin{cases} 0 & \text{for} \quad n < 0 \\ (-a)^n & \text{for} \quad n \geq 0 \end{cases}$$

The impulse response is **infinite - infinite impulse response** – **IIR**.

General recursive system



Output can be written by differencial equation:

$$y[n] = \sum_{k=0}^{Q} b_k x[n-k] - \sum_{k=1}^{P} a_k y[n-k]$$

z-TRANSFORM

will help us, similarly to Laplace transform in continuous domain, to describe discrete signals and systems using complex variable z.
z-transform is defined:

$$X(z) = \sum_{n=-\infty}^{\infty} x[n] z^{-n},$$

where z is complex variable. Let's mark it:

$$x[n] \longrightarrow X(z)$$

inverse transform

$$X(z) \longrightarrow x[n]$$

3 properties:

- **Linearity**:

$$x_1[n] \longrightarrow X_1(z)$$

$$x_2[n] \longrightarrow X_2(z)$$

$$ax_1[n] + bx_2[n] \longrightarrow aX_1(z) + bX_2(z)$$

- **Delay of signals**:

$$x[n] \longrightarrow X(z)$$

$$x[n-k] \longrightarrow \sum_{n=-\infty}^{\infty} x[n-k]z^{-n} =$$

$$= \sum_{n=-\infty}^{\infty} x[n]z^{-n-k} = z^{-k} \sum_{n=-\infty}^{\infty} x[n]z^{-n} = z^{-k}X(z)$$

$$x[n-1] \longrightarrow z^{-1}X(z)$$

1 sample delay:

$$\longrightarrow \boxed{z^{-1}} \longrightarrow$$

- **Relation to DTFT**: Fourier transform with discrete time:

$$\tilde{X}(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n}$$

it is quite similar to ZT, if z will be $e^{j\omega}$:

$$\tilde{X}(e^{j\omega}) = X(z)|_{z=e^{j\omega}},$$

Transformation function of recursive system

For system:

we will define:

$$H(z) = \frac{Y(z)}{X(z)}$$

$$y[n] = \sum_{k=0}^{Q} b_k x[n-k] - \sum_{k=1}^{P} a_k y[n-k] \longrightarrow Y(z) = \sum_{k=0}^{Q} b_k X(z) z^{-k} - \sum_{k=1}^{P} a_k Y(z) z^{-k}$$

$$Y(z) + \sum_{k=1}^{P} a_k Y(z) z^{-k} = \sum_{k=0}^{Q} b_k X(z) z^{-k}$$

and we get:

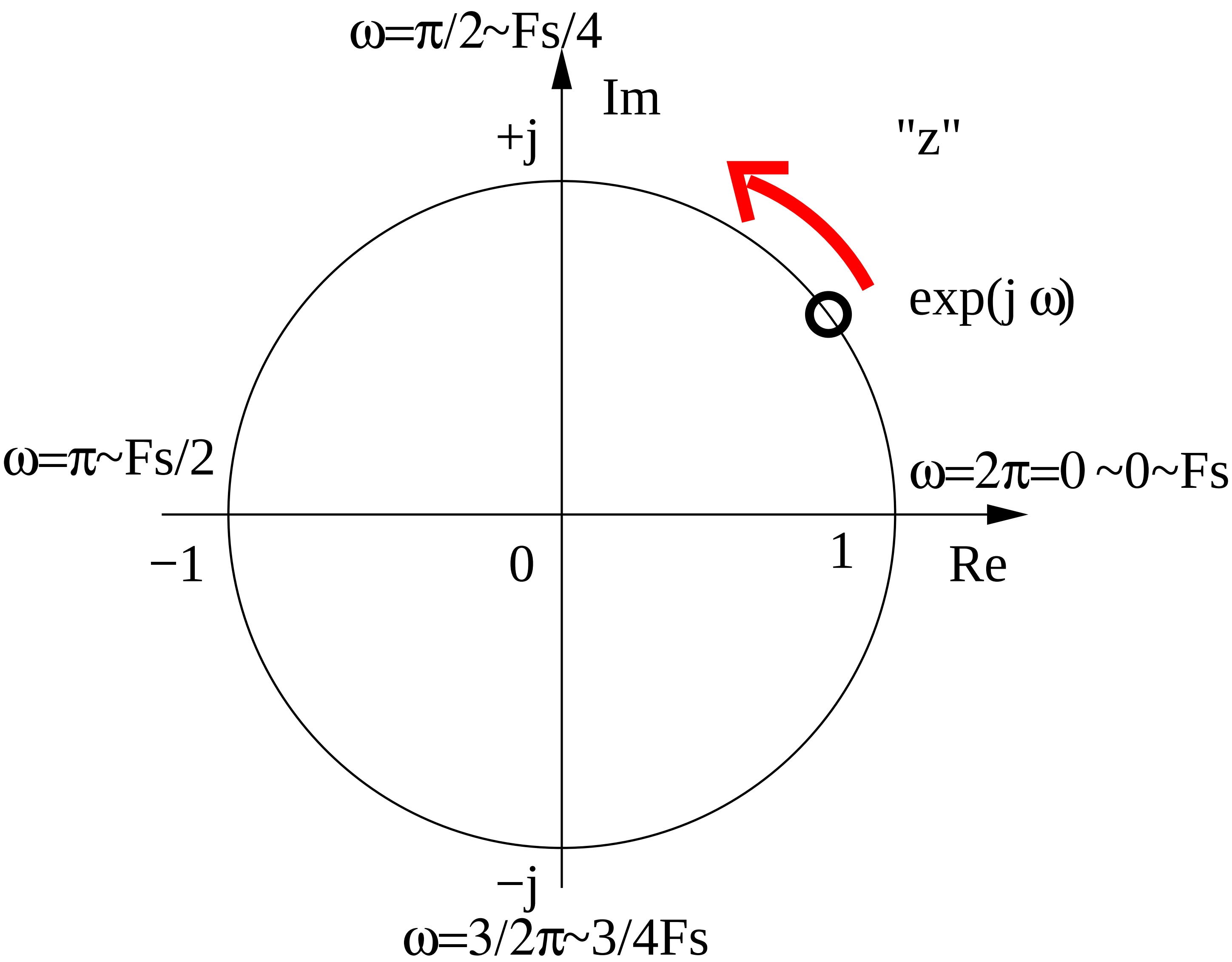
$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^{Q} b_k z^{-k}}{1 + \sum_{k=1}^{P} a_k z^{-k}} = \frac{B(z)}{A(z)},$$

Frequency characteristics of filters

We just replace z by $e^{j\omega}$ and we keep ω varying in interval we are interested – e.g. from 0 to π (half of sampling frequency):

$$H(e^{j\omega}) = H(z)|_{z=e^{j\omega}} = \frac{\sum_{k=0}^{Q} b_k e^{-j\omega k}}{1 + \sum_{k=1}^{P} a_k e^{-j\omega k}}$$

Frequency characteristics of filters



Frequency characteristic from nulls and poles

$H(e^{j\omega})$:

$$H(e^{j\omega}) = b_0 z^{-(Q-P)} \frac{\prod_{k=1}^{Q}(z - n_k)}{\prod_{k=1}^{P}(z - p_k),} |_{z=e^{j\omega}} = b_0 e^{j\omega(P-Q)} \frac{\prod_{k=1}^{Q}(e^{j\omega} - n_k)}{\prod_{k=1}^{P}(e^{j\omega} - p_k),}$$

Example - non-recursive filter:

$$y[n] = x[n] + 0.5x[n-1]$$

- Impulse response?
- Transformation function (coef. a, b)?
- Frequency characteristics?
- by hand – freq. charfacteristics using nulls and poles.

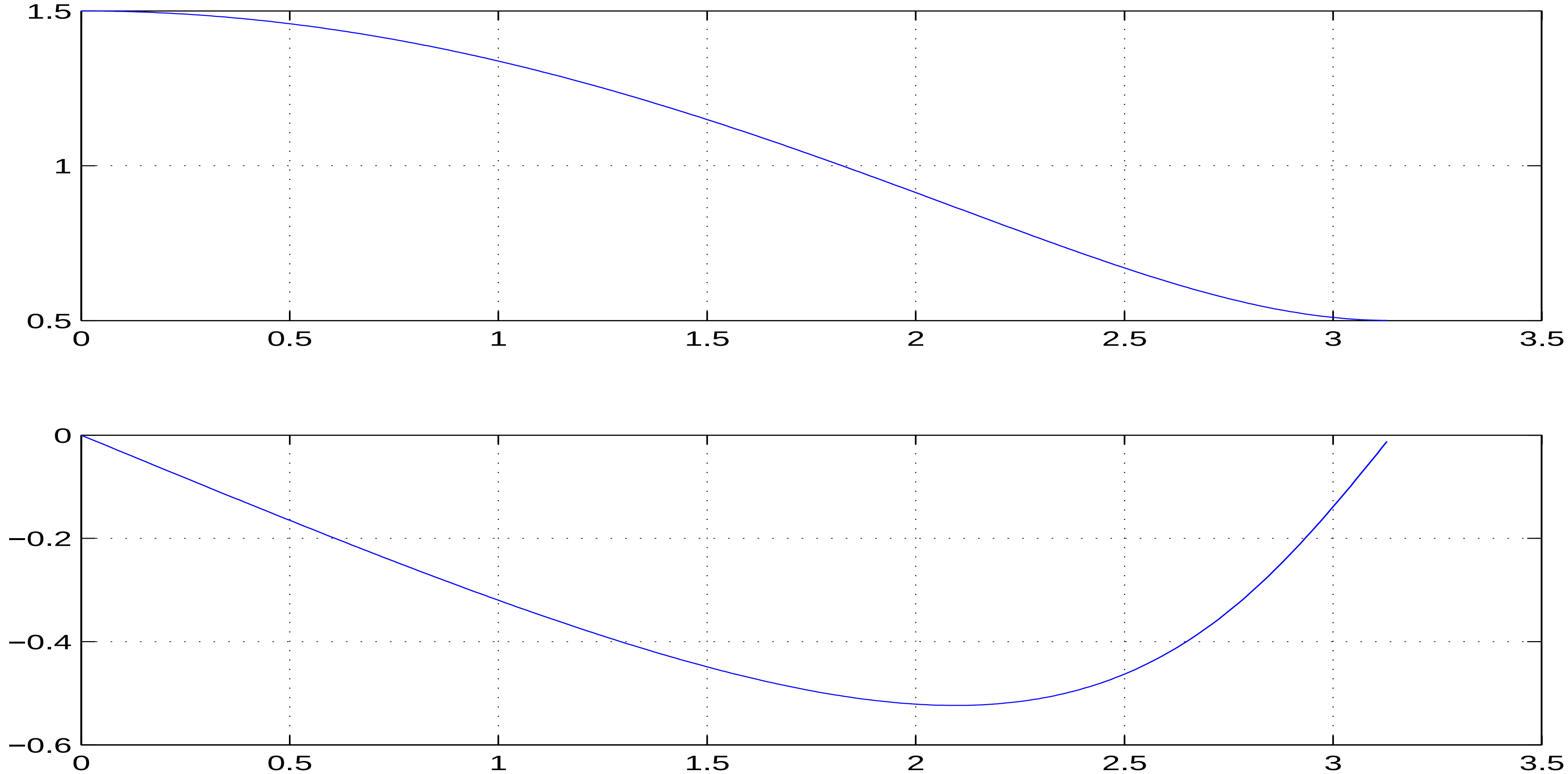
Solution:

- $h[n] = 1,\ 0.5$ for $n = 0, 1$, null elsewhere.
- $Y(z) = X(z) + 0.5X(z)z^{-1} \qquad Y(z) = X(z)[1 + 0.5z^{-1}]$
  $H(z) = 1 + 0.5z^{-1} = \frac{1+0.5z^{-1}}{1}$, thus $b_0 = 1$, $b_1 = 0.5$, $a_0 = 1$.

Example - non-recursive filter:

- $z = e^{j\omega}$, let's call:

```
H=freqz([1 0.5],[1],256); om=(0:255)/256 * pi;
subplot(211); plot(om,abs(H)); grid
subplot(212); plot(om,angle(H)); grid
```



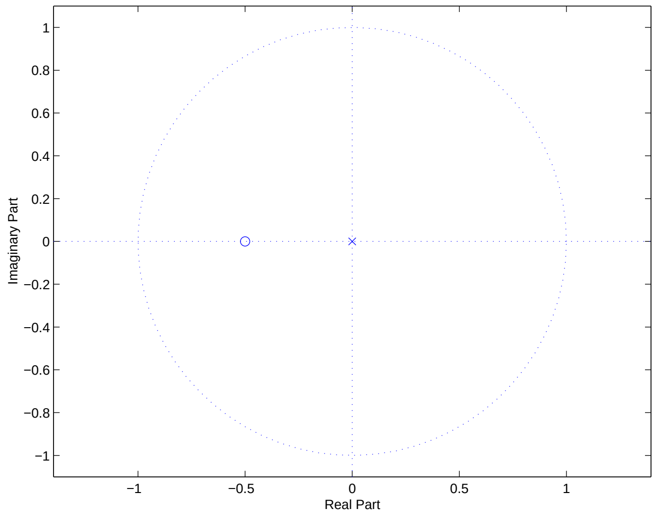
Example - non-recursive filter:

$\Rightarrow$ Low pass filter:

- Nulls and poles: $H(z) = \frac{1+0.5z^{-1}}{1} = \frac{z(1+0.5z^{-1})}{z} = \frac{z+0.5}{z}$

  Numerator is equal to zero for $z = -0.5$, thus, filter will have 1 null $n_1 = -0.5$.

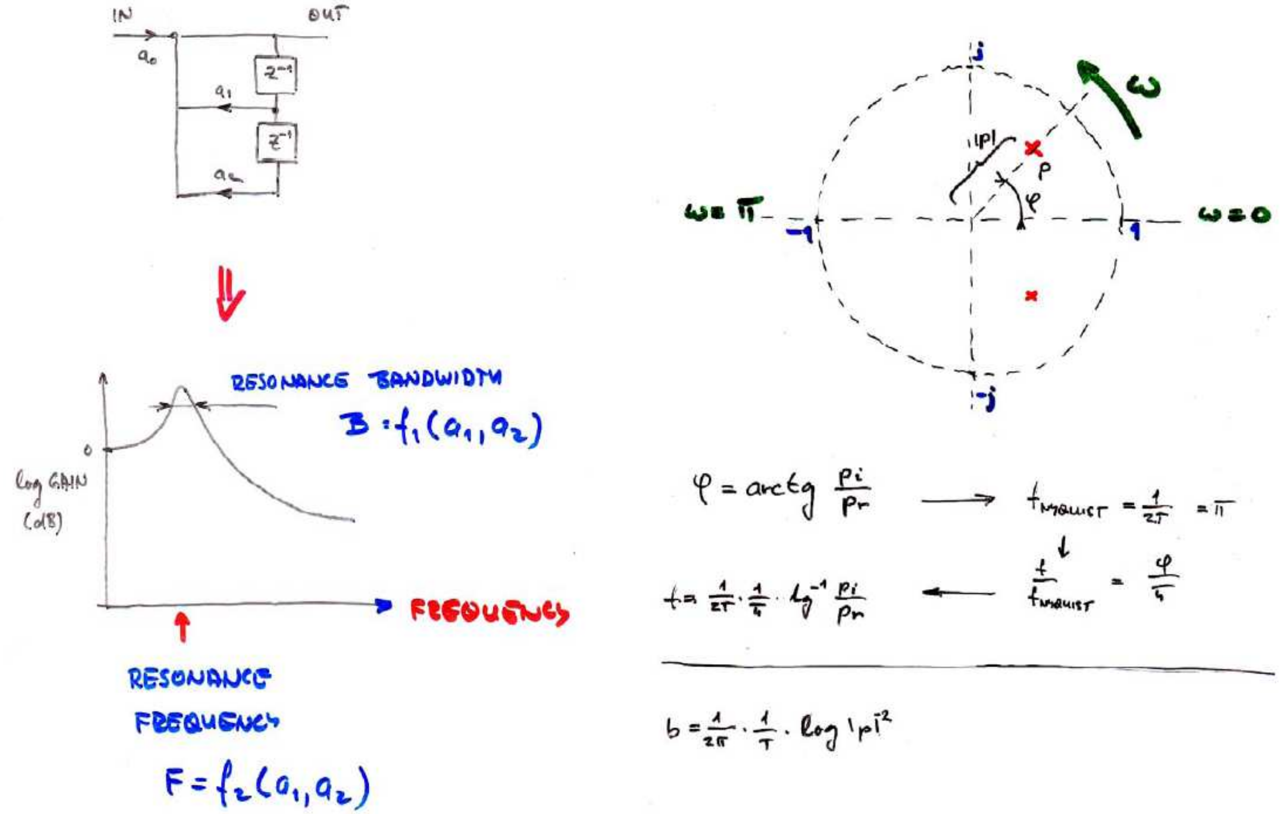
  Denominator will be zero for $z = 0$, thus one pole: $p_1 = 0$



- Freq. characteristics using nulls and poles:

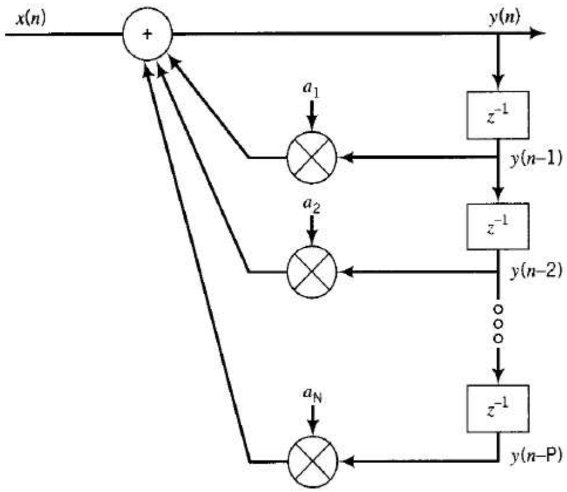
  $H(z) = \frac{z-(-0.5)}{z-0}$ $\qquad$ $H(e^{j\omega}) = \frac{e^{j\omega}-(-0.5)}{e^{j\omega}-0}$

Digital resonator



All-pole model

All-pole (autoregressive) model



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

a) ROOTS OF DENOMINATOR

$$1 - \sum_{k=1}^{M} a_k \cdot z^{-k} = \prod_{k=1}^{M} (1 - p_k \cdot z^{-1})$$

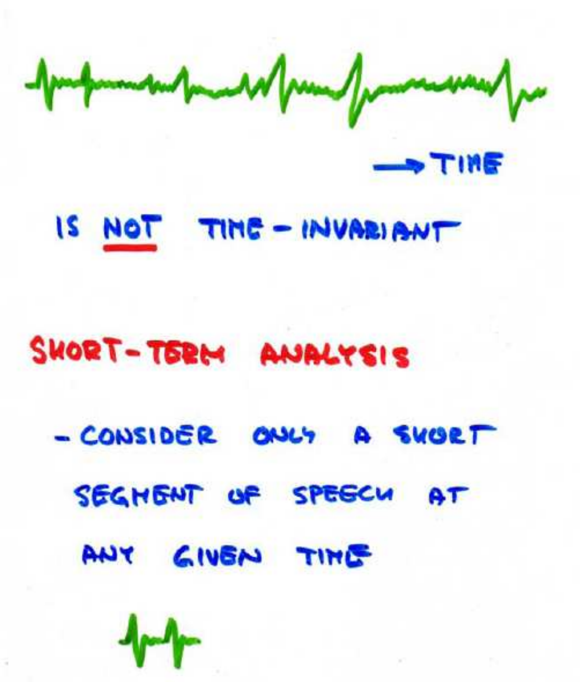
$$p_k = p_r + j p_i$$

FREQUENCY $f_k = \frac{f_{SAMPLE}}{2\pi} \cdot \tan^{-1} \frac{p_i}{p_r}$

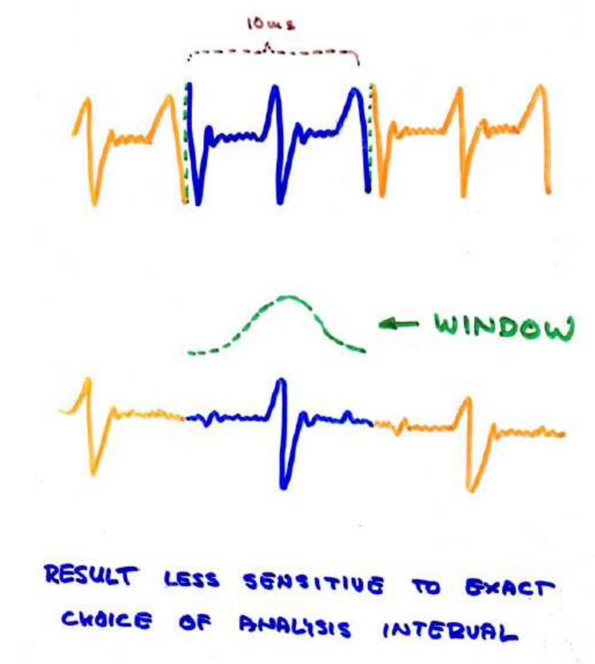
BANDWIDTH $b_k = -\frac{f_{SAMPLE}}{2\pi} \cdot \ln (p_r^2 + p_i)^2$

Short-Time Analysis

Speech signal



Non-stationary turns into periodic



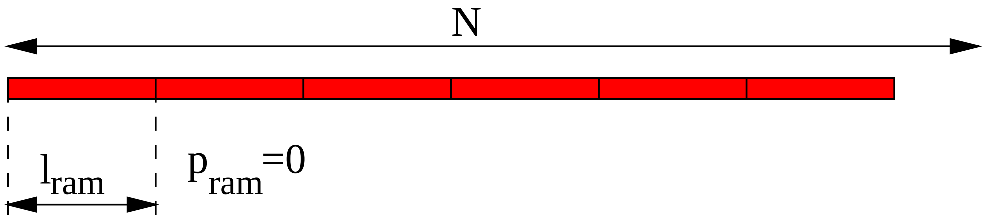
Short-Time Analysis

Frames:

- Signal should be stationary. BUT – is not
- Division into shorter segments
- New parameters:
  - Length of segments L: short enough to ensure stationarity; long enough to well estimate parameters of the segment.
  - Overlap O: small - fast computing but parameters can change dramataically from one segment to the other; efficiency due to bit-rates!!
  - frame-shift: $S = L - O$

Short-Time Analysis

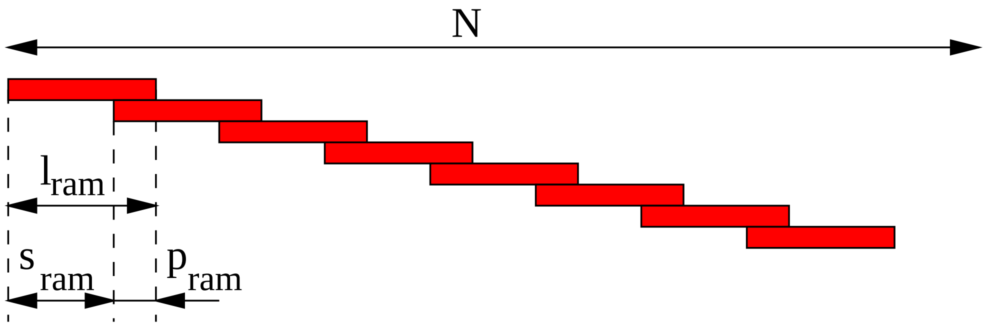
No overlapping $O = 0$:



$$N_{ram} = \left\lfloor \frac{N}{l_{ram}} \right\rfloor$$

$\ldots \lfloor \cdot \rfloor$ means flooring.

Short-Time Analysis



$$N_{ram} = 1 + \left\lfloor \frac{N - l_{ram}}{s_{ram}} \right\rfloor$$

...if the signal is at least one frame long.

Short-Time Analysis

Selection of the signal by a window:

- Rectangular window:

$$w[n] = \begin{cases} 1, & \text{for} \quad 0 \leq n \leq l_{ram} - 1 \\ 0, & \text{elsewhere} \end{cases}$$

- Hamming window - attenuation of sides:

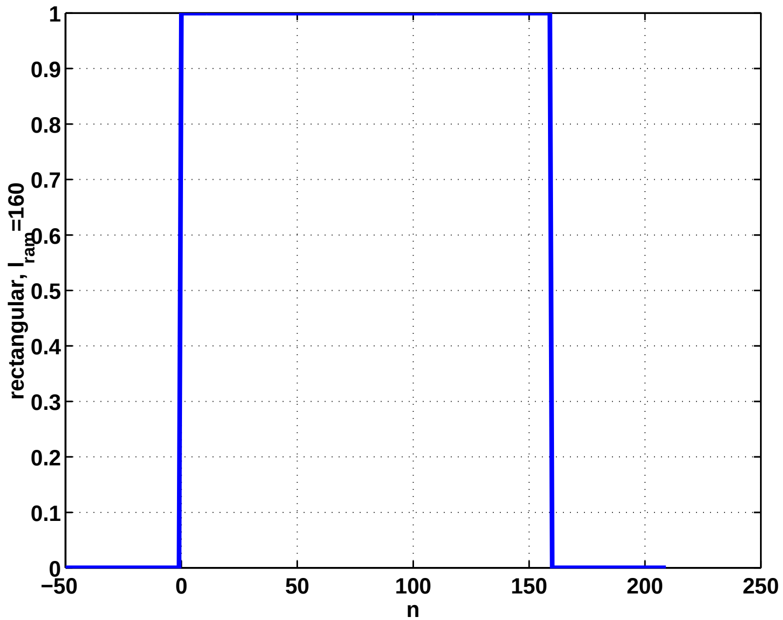
$$w[n] = \begin{cases} 0.54 - 0.46 \cos \dfrac{2\pi n}{l_{ram} - 1} & \text{for} \quad 0 \leq n \leq l_{ram} - 1 \\ 0 & \text{elsewhere} \end{cases}$$

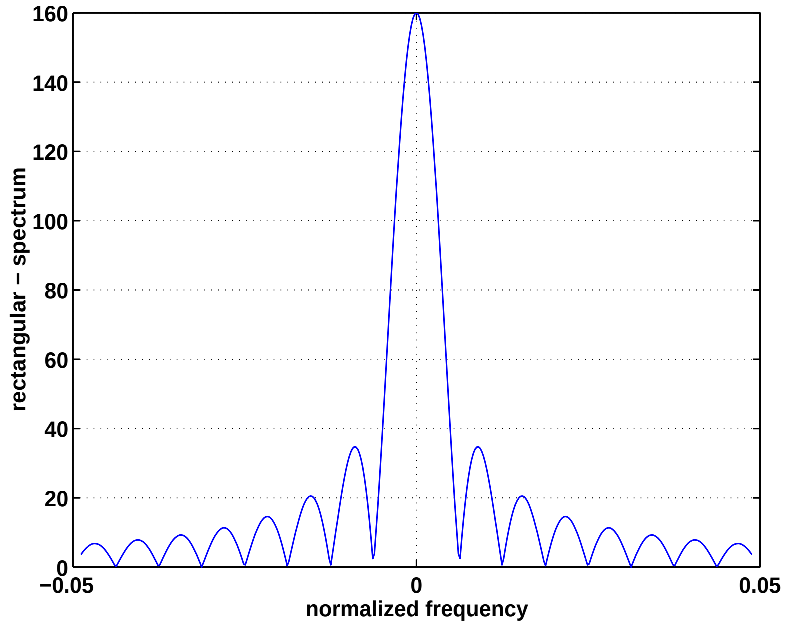
Short-Time Analysis

What will happen with spectrum of the signal chopped by the window: Multiplication of the signal by a window in time domain $\Longleftrightarrow$ convolution of the both spectra

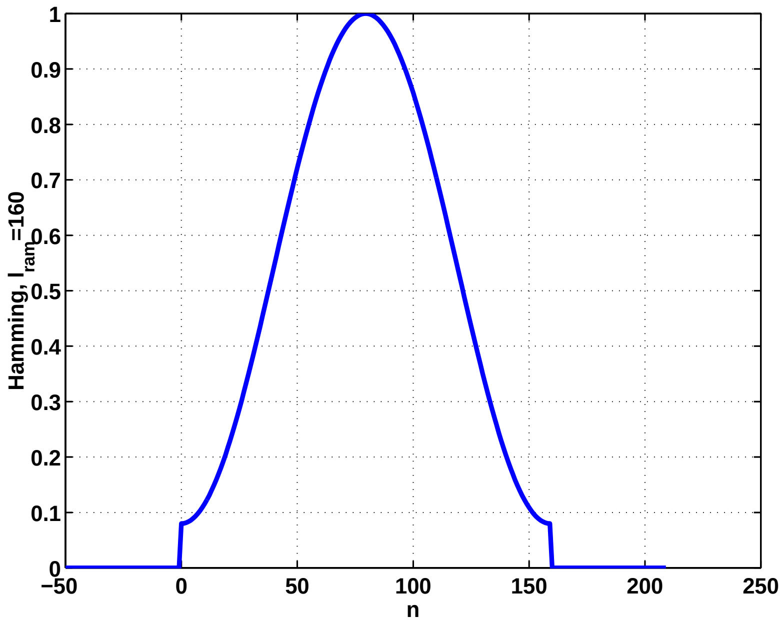
$$X(f) = S(f) \star W(f)$$

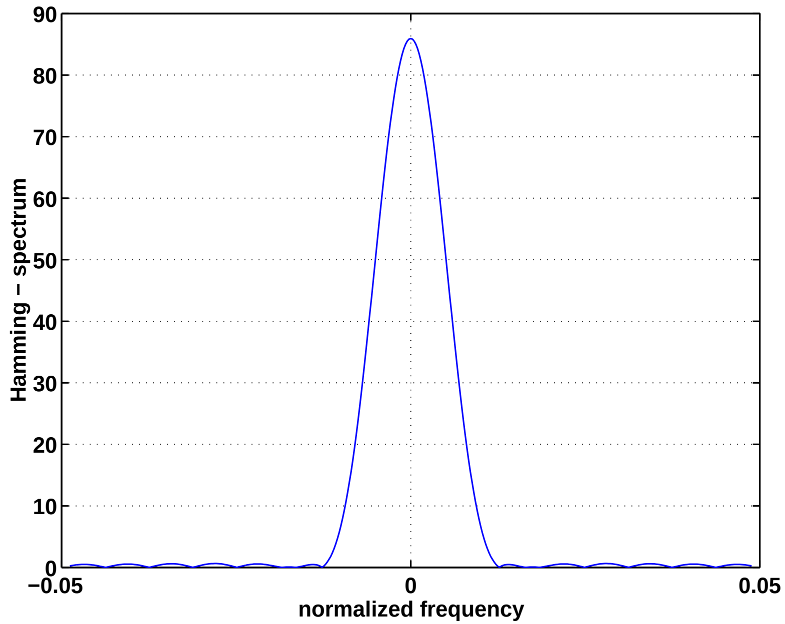
Comparison of Rectangular and Hamming window:





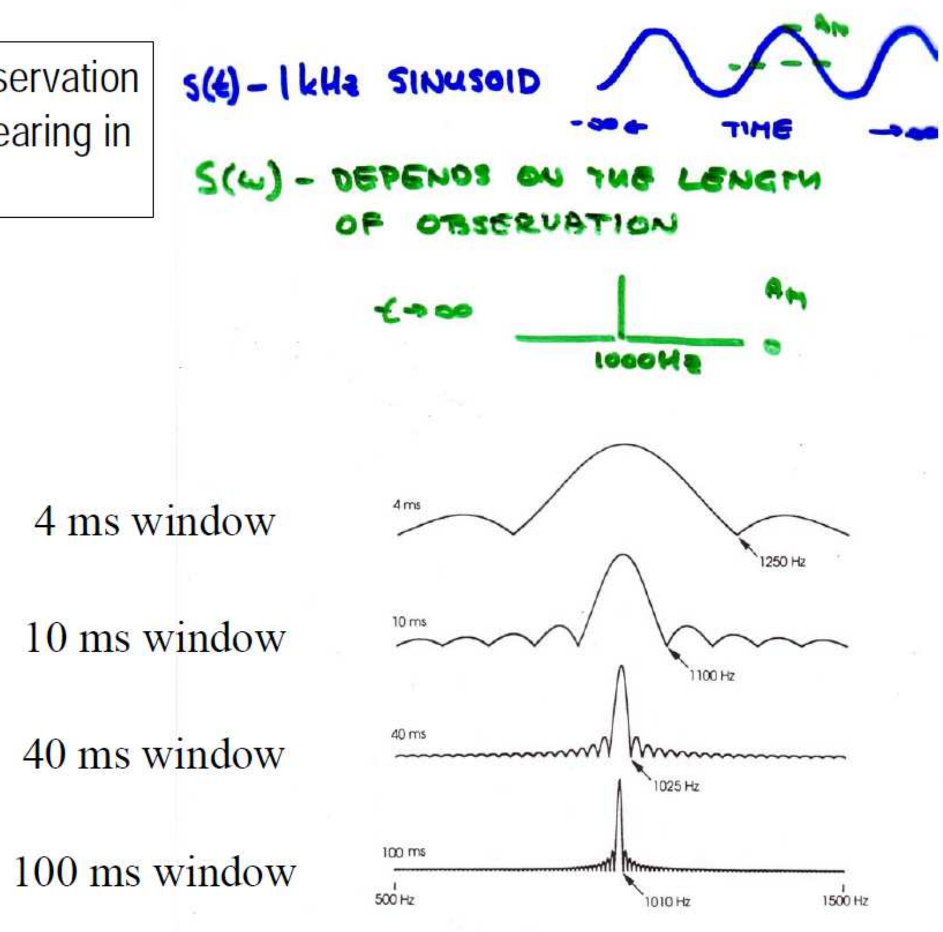
Short-Time Analysis





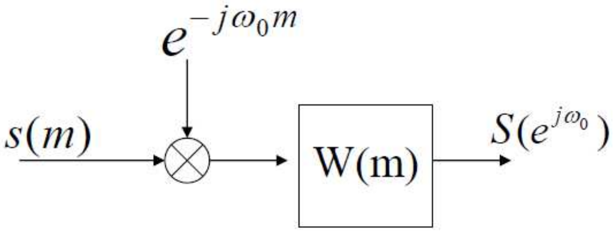
Short-Time Analysis

Finite length of observation always means smearing in frequency



Short-Time fourier Transform

$$S_n(e^{j\omega}) = \sum_{m=-\infty}^{\infty} s(m)e^{-jm\omega} \cdot w(n-m)$$

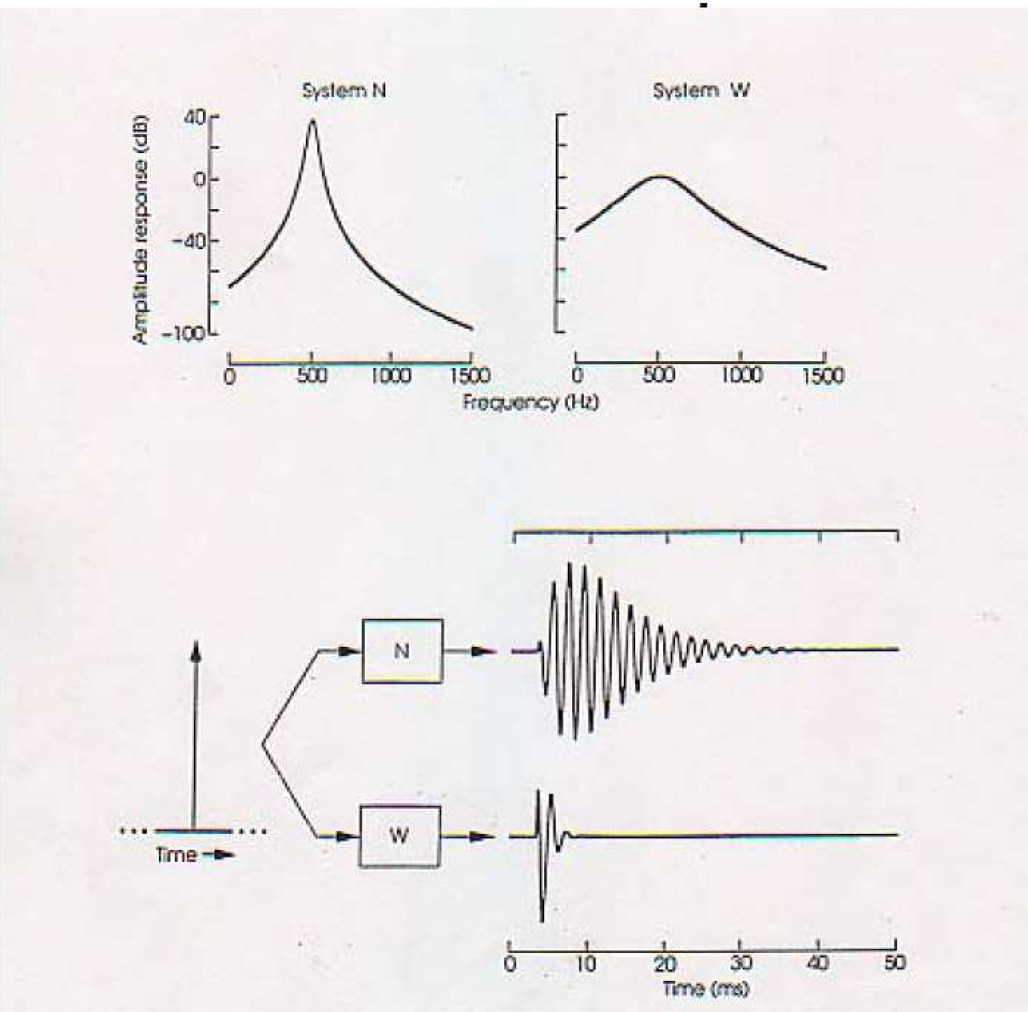


if ω is fixed (at a particular frequency ω_0), the equation above represents convolution of two terms

$$s(m) \cdot e^{-j\omega_0 m} * w(m)$$

The convolution represents linear filtering by a band - pass filter with center frequency ω_0 and the filter shape given by frequency response $W(\omega)$ of the window w(m)

Uncertainty principle



Sharper frequency response implies longer impulse response

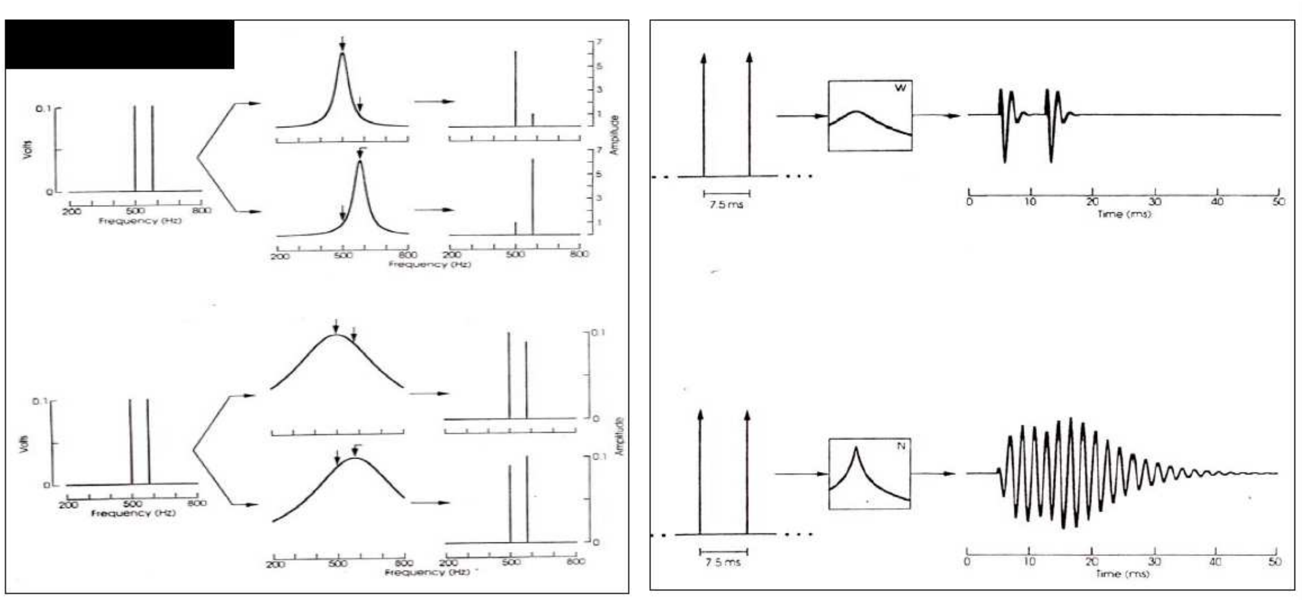
- better frequency resolution means worse temporal resolution
- you cannot know accurately both the frequency of the stimulus and its location (uncertainty principle)

Uncertainty principle

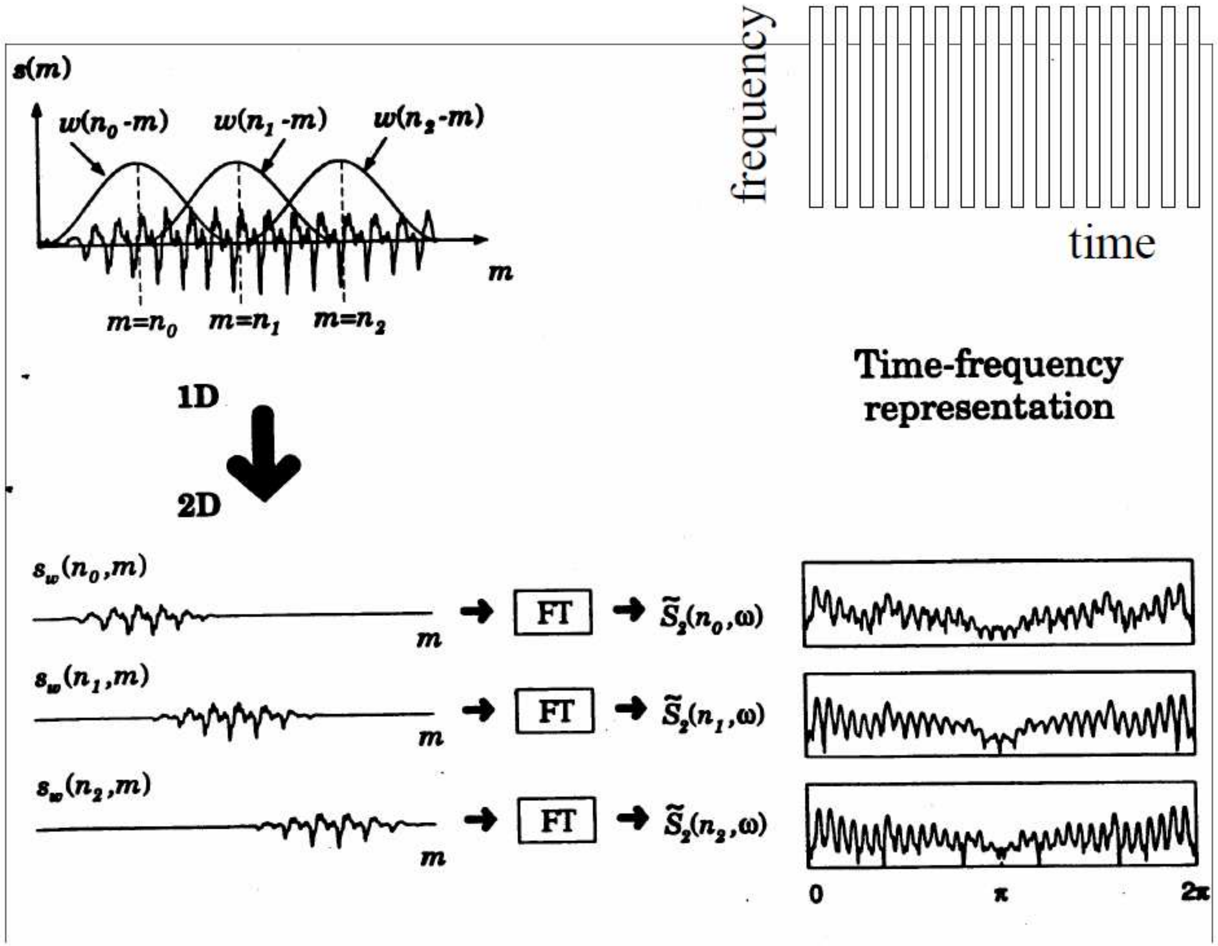
Originally from nuclear physics [Heisenberg], also applies to signal processing [Gabor]:

Product of time and frequency resolution is constant

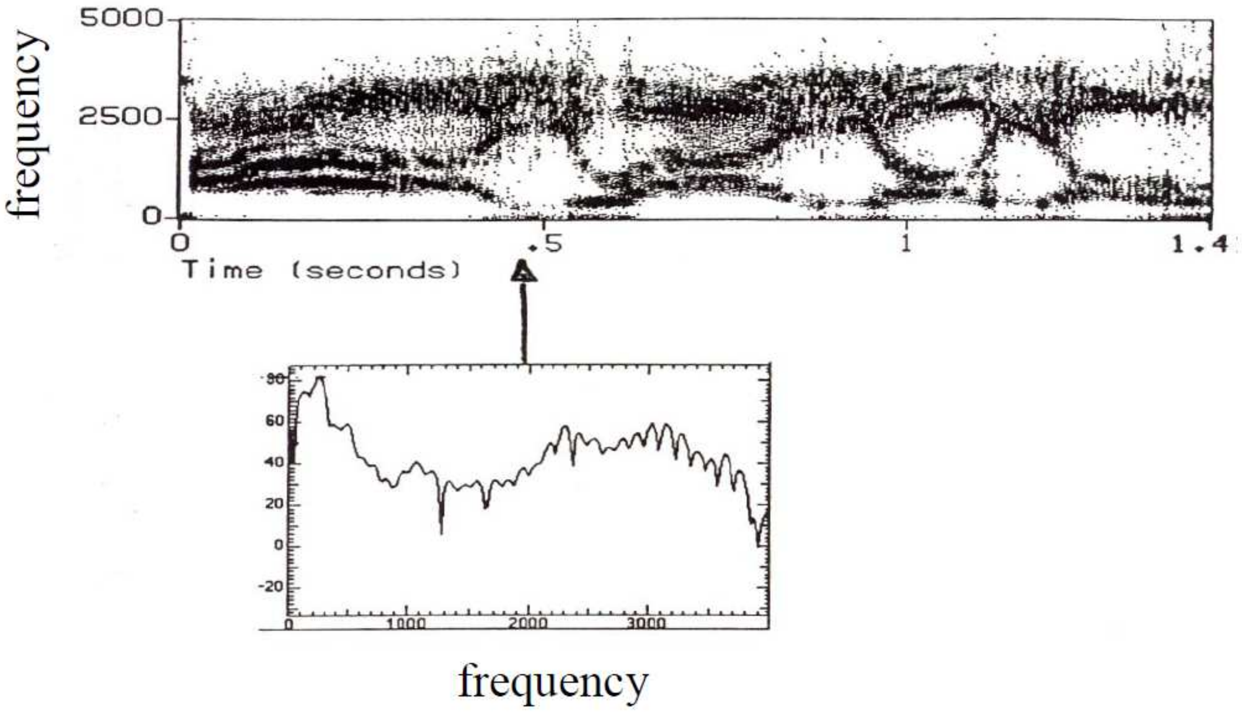
(*High frequency resolution means low temporal resolution and vice-versa*)



Spectrogram



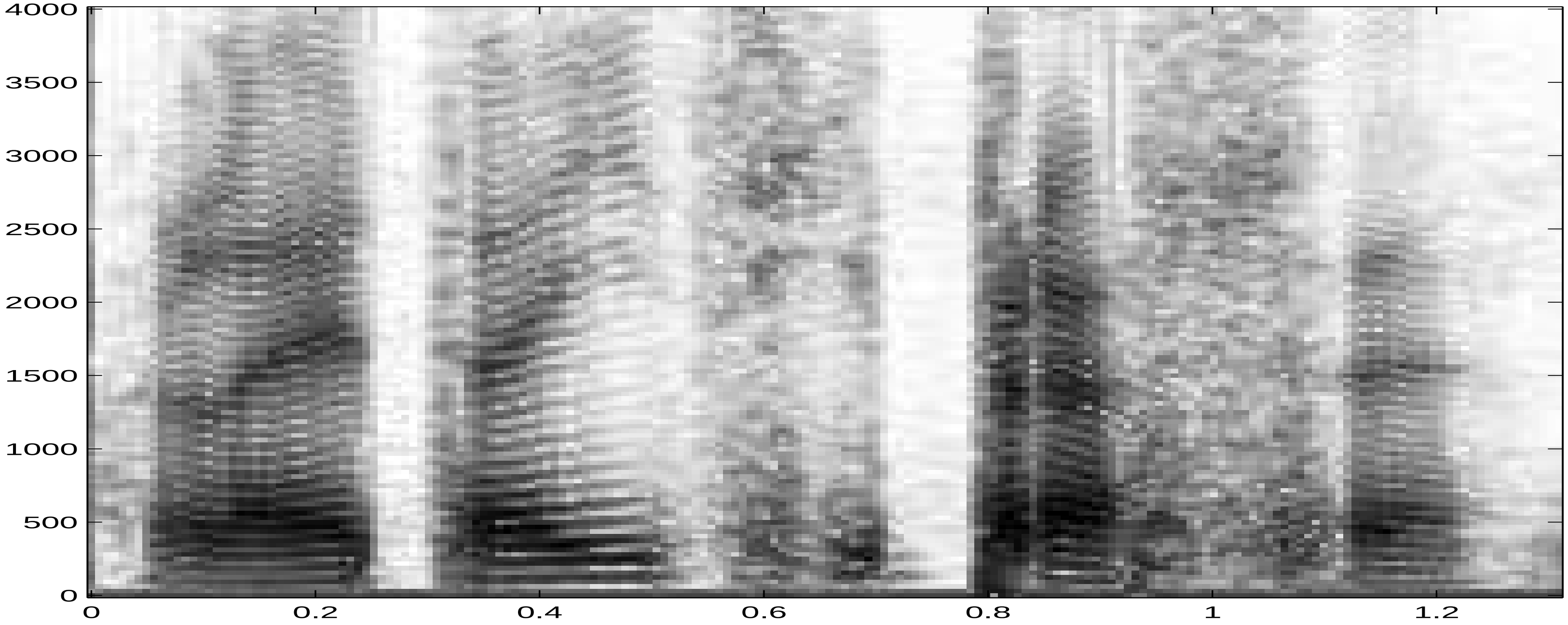
Spectrogram



Spectrogram

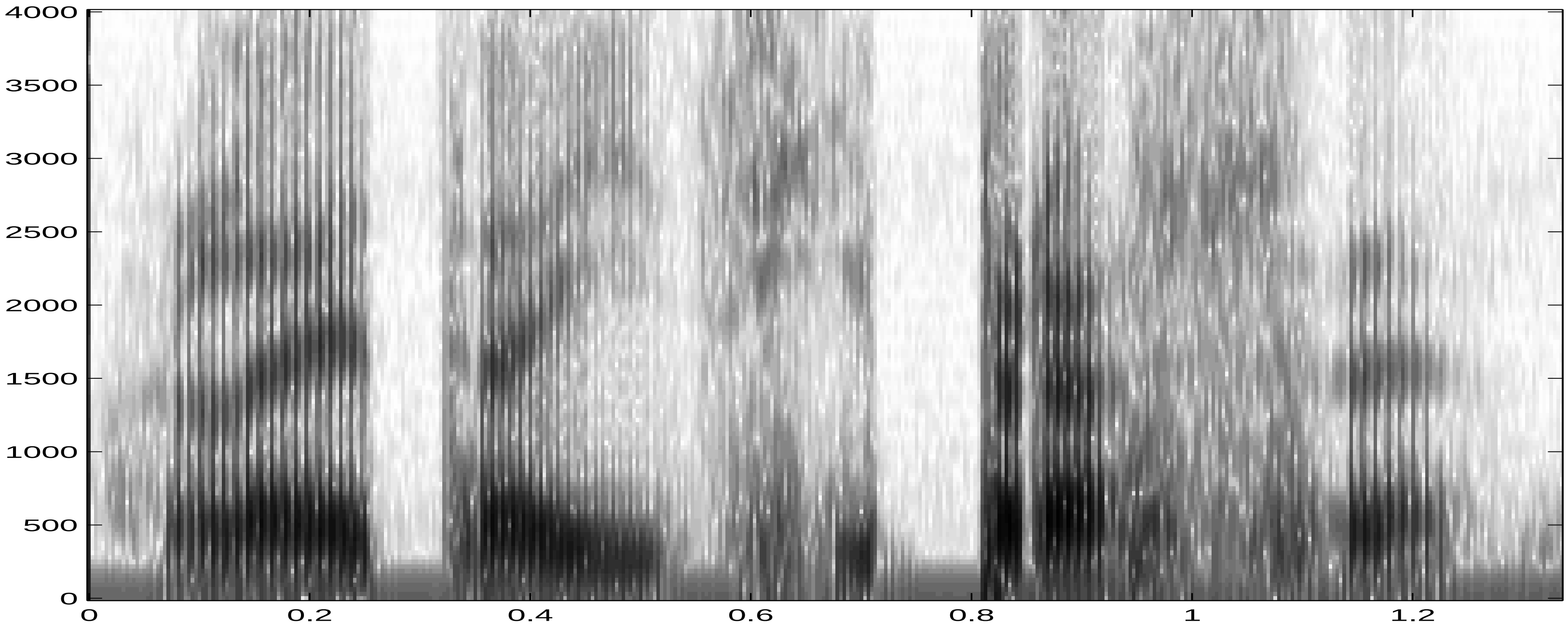
long-term/narrow-band:

```
specgram(s,256,8000,hamming(256),200);
```

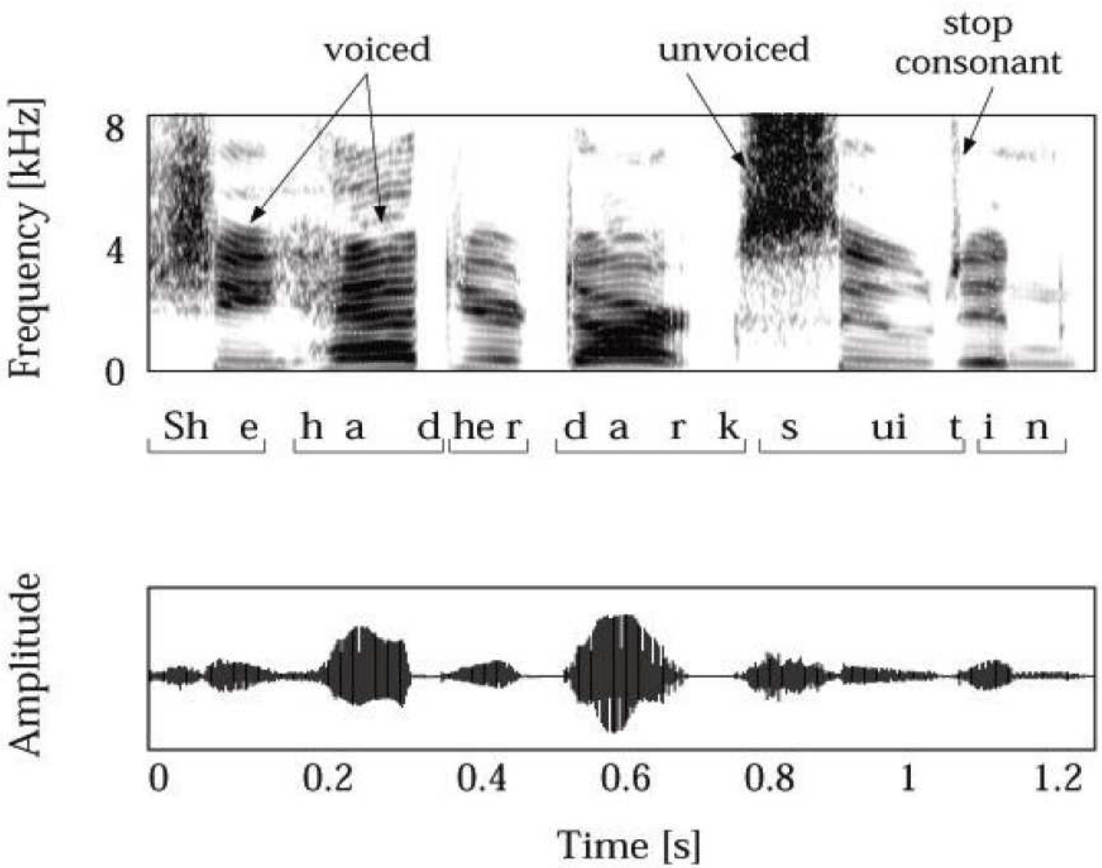


Spectrogram

short-term/wide-band: `specgram(s,256,8000,hamming(50));`

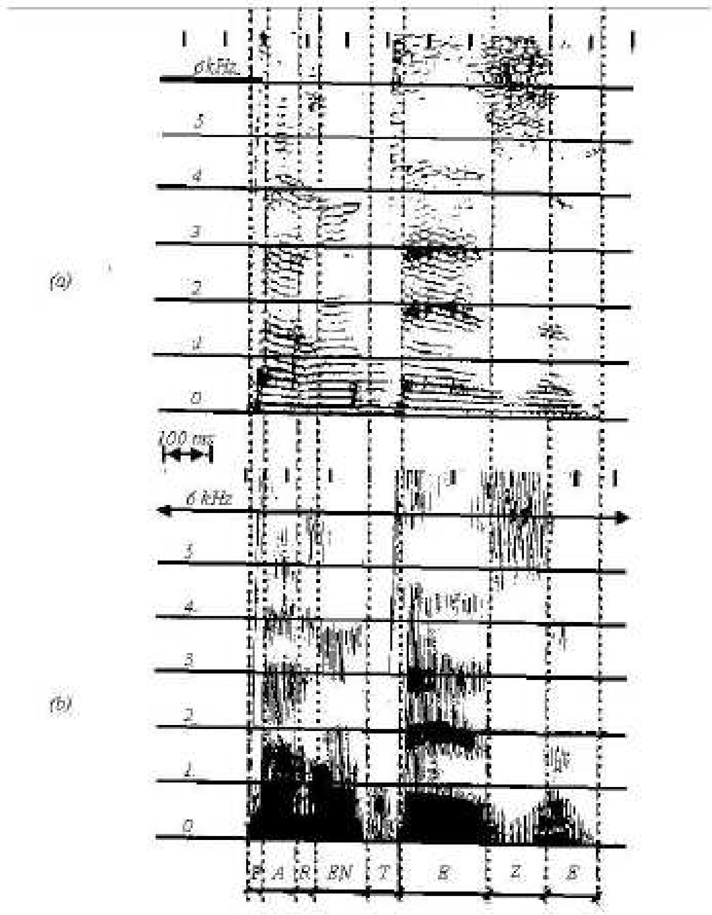


Spectrogram



Spectrogram

Spectrogram of the word: “parenthese”: (a) narrow-band, (b) wide-band.



Multi-rate Signal Processing (MSP)

- Involves the change of sampling rate while the signal is in the digital domain.
- It can reduce the algorithmic and HW complexity or increase the resolution ... by introducing additional signal samples.
- MSP – two basic soperations - up-sampling and down-sampling:
  - down-sampling - increasing the sampling period - decreasing samppling frequency and data-rate of the digital signal. A sampling rate reduction by integer L:

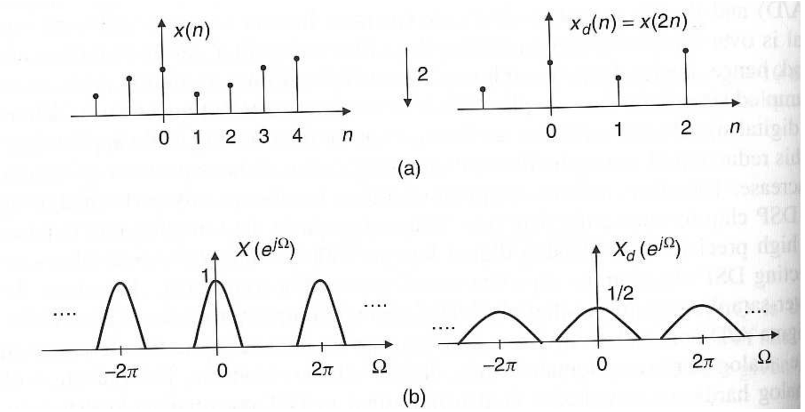
$$x_d[n] = x[nL]$$

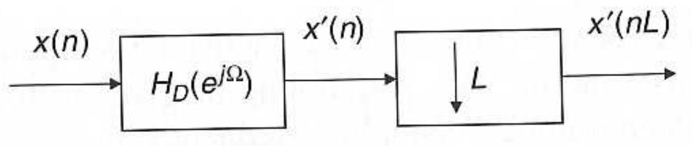
  - up-sampling - decreasing the sampling period - increasing sampling frequency. A sampling rate increase by integer L:

$$y_e[n] = \begin{cases} x[nL] & \text{for } \ n \ is \ integer - multiple \ of \ L \\ 0 & \text{elsewhere} \end{cases}$$

Multi-rate Signal Processing (MSP)

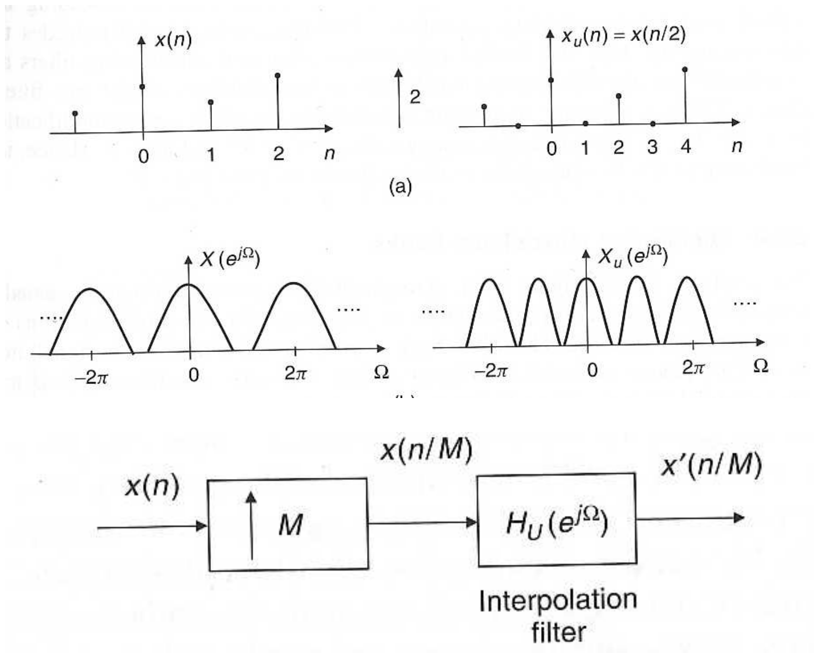
$$H_D(e^{j\Omega}) = \begin{cases} 1, & 0 \leq |\Omega| \leq \pi/L \\ 0, & \pi/L \leq |\Omega| \leq \pi \end{cases}$$





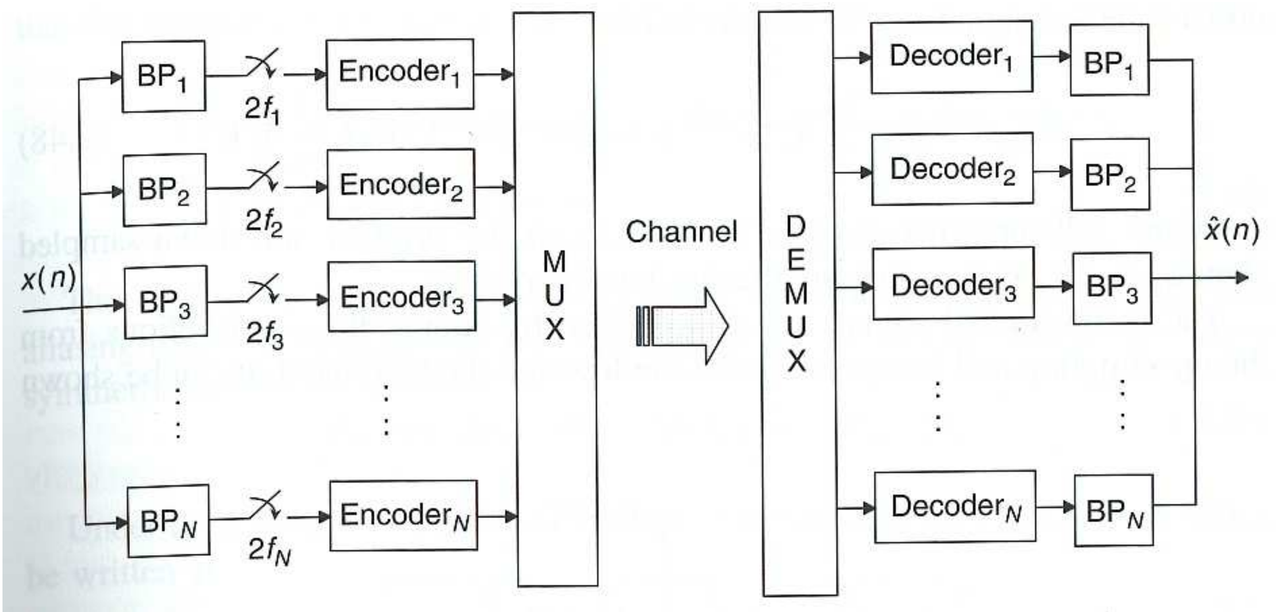
Multi-rate Signal Processing (MSP)

$$H_U(e^{j\Omega}) = \begin{cases} M, & 0 \le |\Omega| \le \pi/M \\ 0, & \pi/M \le |\Omega| \le \pi \end{cases}$$



Quadrature Mirror Filter Banks

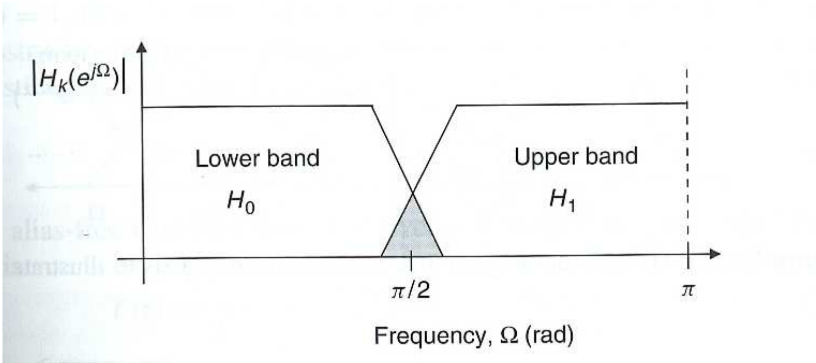
- The analysis of the signal in perceptual audio coding system: filter banks, frequency-domain transformations, combination of both
- Filter bank: to decompose the signal into several filter banks
  $\Rightarrow$ subbnand coding

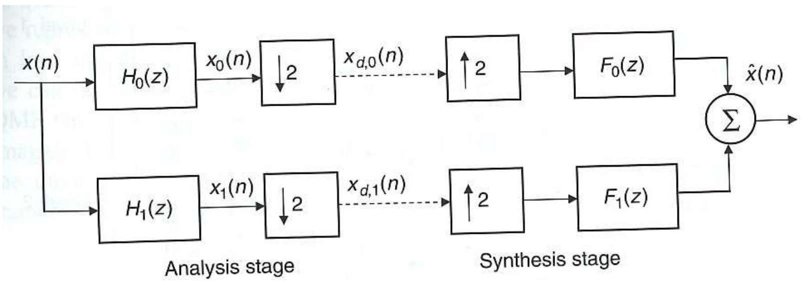


Quadrature Mirror Filter Banks

Important aspect: aliasing between the different sub-bands – due to imperfect frequency responses of the digital filters:

- Solved in 1977 by proposing perfect reconstruciton filter bank known as QMF
- QMF: anti-aliasing filters, down/up-sampling stages, interpolation filters





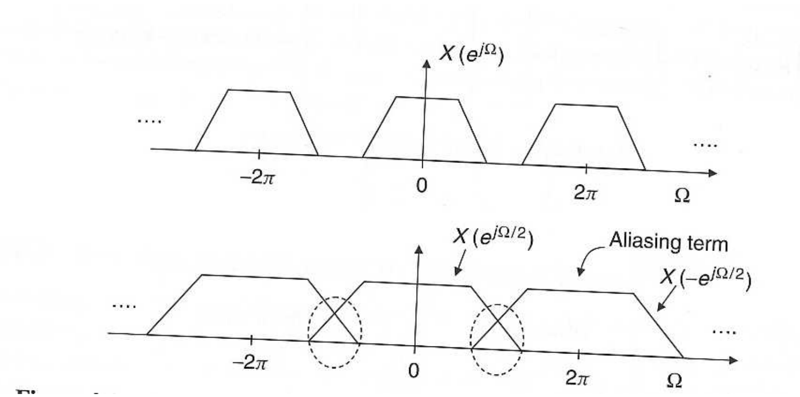
Quadrature Mirror Filter Banks

How to:

- Signal is filtered and down-sampled:

$$X_{d,k}(e^{j\Omega}) = X_k(e^{j\Omega/2}) + X_k(e^{j(\Omega-2\pi)/2})$$

- The reconstructed signal $\hat{x}[n]$ – derived by adding the contributions from the up-sampling and interpolations of the low and high band.



Quadrature Mirror Filter Banks

It can be shown that the reconstructed signal in Z-domain:

$$\hat{X}(z) = 1/2\Big(H_0(z)F_0(z) + H_1(z)F_1(z)\Big)X(z)+$$

$$+1/2\Big(H_0(-z)F_0(z)H_1(-z)F_1(z)\Big)X(-z)$$

The aliasing term can be cancelled by designing filters to have the following mirror symetries:

$$F_0(z) = H_1(-z) \qquad F_z(z) = -H_0(-z)$$

Under these conditions, the overall transfer function of QMF can be:

$$T(z) = 1/2\Big(H_0(z)F_0(z) + H_1(z)F_1(z)\Big)$$

If $T(z) = 1$ the filter bank allows for perfect reconstruction. Perfect delay-less reconstruction is not realizable.

Quadrature Mirror Filter Banks

For example: the choice of first-order FIR filter:

$$H_0(z) = 1 + z^{-1} \qquad H_1(z) = 1 - z^{-1}$$

results in alias-free reconstruction. The overall $T(z)$ function will be:

$$T(z) = 1/2\Big((1 + z^{-1})^2 - (1 - z^{-1})^2\Big) = 2z^{-1}$$

The signal will be reconstructed within the delay of one sample and with overall gain of 2.

Quadrature Mirror Filter Banks

QMF can be cascaded to form tree structures. The theory of QMF has been associated with wavelet transform theory.