Understanding sound and speech as periodic signals

Outline today

Recap periodic basis functions
Analysis of frequency content (power spectrum)
Simple convolution filters
Phenomenology - sound and speech signals
Acquire audio signals,
Understand the spectrogram
analysis of pitch

DTU Compute
Department of Applied Mathematics and Computer Science

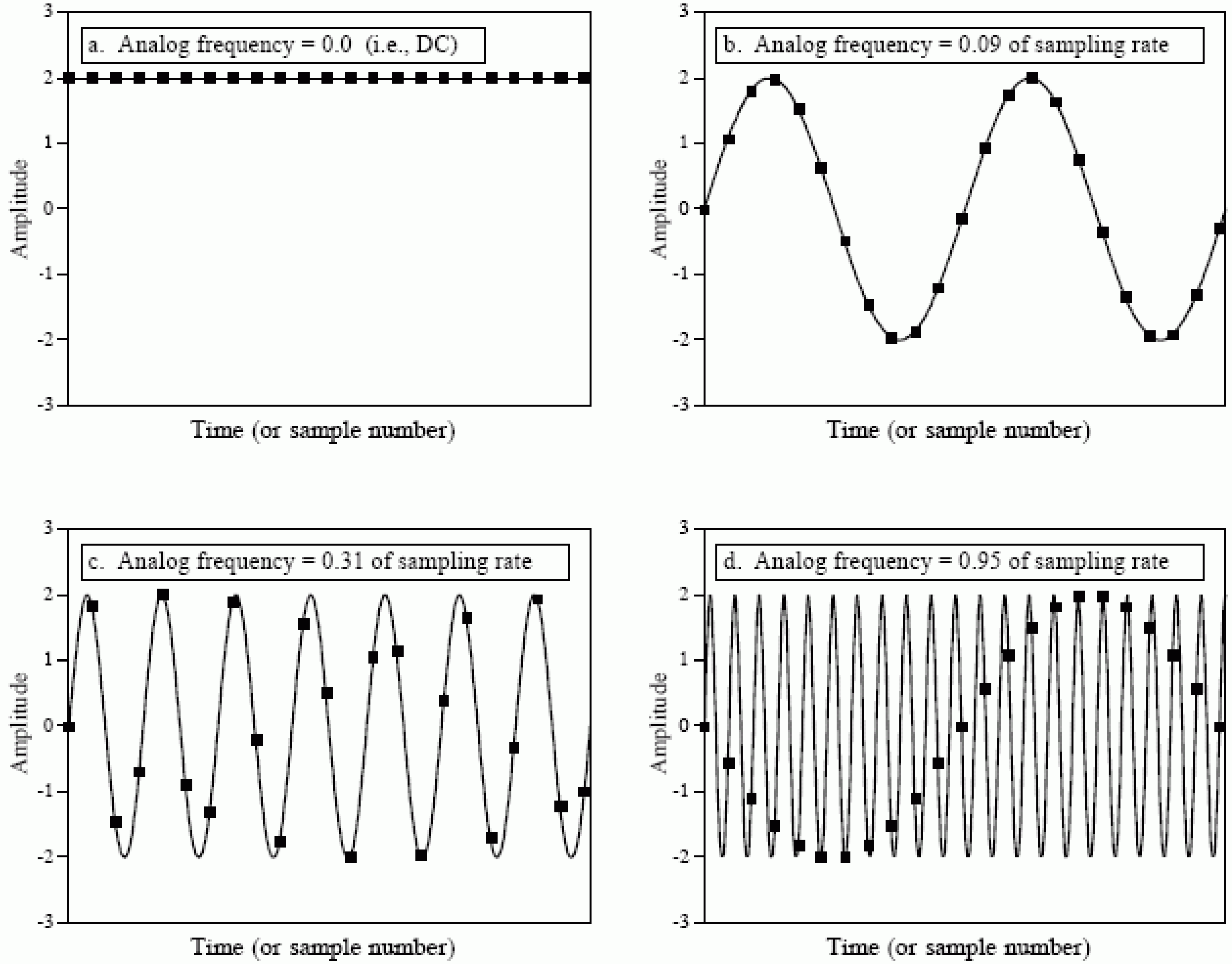


FIGURE 3-3
Illustration of proper and improper sampling. A continuous signal is sampled *properly* if the samples contain all the information needed to recreate the original waveform. Figures (a), (b), and (c) illustrate *proper sampling* of three sinusoidal waves. This is certainly not obvious, since the samples in (c) do not even appear to capture the shape of the waveform. Nevertheless, each of these continuous signals forms a unique one-to-one pair with its pattern of samples. This guarantees that reconstruction can take place. In (d), the frequency of the analog sine wave is greater than the Nyquist frequency (one-half of the sampling rate). This results in *aliasing*, where the frequency of the sampled data is different from the frequency of the continuous signal. Since aliasing has corrupted the information, the original signal cannot be reconstructed from the samples.

http://www.dspguide.com/ch3/2.htm

Sampling – The Nyquist rate

A simple analog signal is the harmonic oscillation:

$$x_a(t) = A\sin(2\pi F\,t + \theta)$$

The frequency F is measured in cycles per second (hertz).
The discrete signal obtained with a sampling rate $F_s = \frac{1}{T}$ is:

$$x_d(n) = A\sin\left(2\pi\frac{F}{F_s}n + \theta\right)$$

Note, that if the frequency increases $F' = F + kF_s$ for some integer k, the discrete signal becomes:

$$x'_d(n) = A\sin\left(2\pi\frac{F + kF_s}{F_s}n + \theta\right) = A\sin\left(2\pi\frac{F}{F_s}n + \theta\right)$$

Such higher frequencies, F', are called aliases of the frequence F as $x'_d(n) = x_d(n)$.
The Shannon-Nyquist sampling theorem states that the continuous signal can only be properly sampled when the sampling frequency is at least twice the frequency of the signal, ie. $F_s > 2F$ (The Nyquist rate).

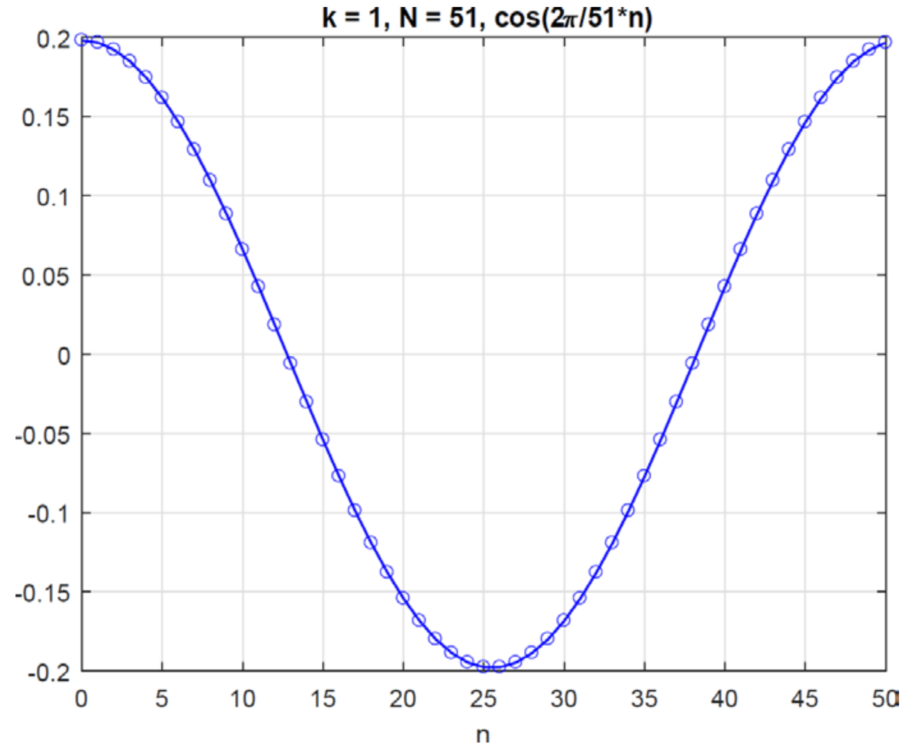
Periodic functions as basis vectors

Let us consider signals sampled on the interval $n \in [0, N-1]$.
The fundamental frequency functions

$$x_1(n) = A \sin\left(\frac{2\pi}{N} n\right)$$

$$y_1(n) = B \cos\left(\frac{2\pi}{N} n\right)$$

are periodic with $x(n+N) = x(n), \; y(n+N) = y(n)$

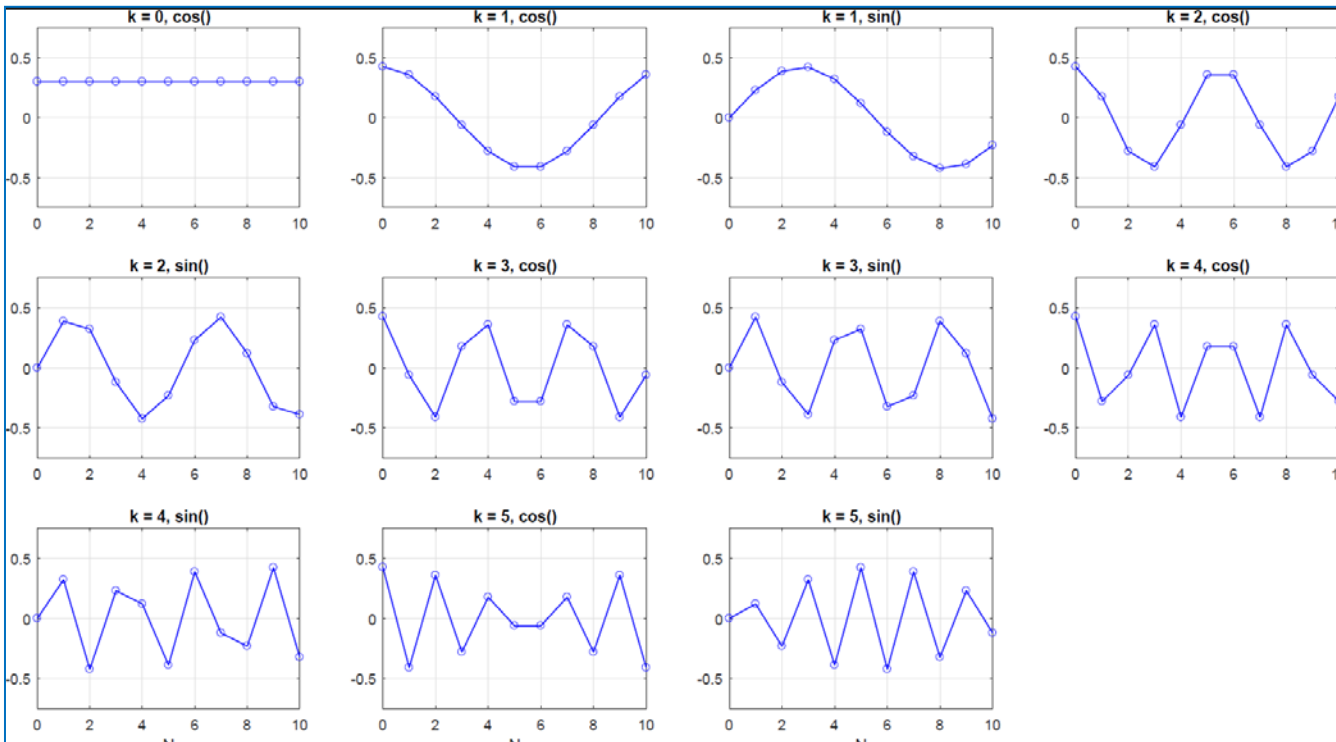


The 'higher' harmonics are given by

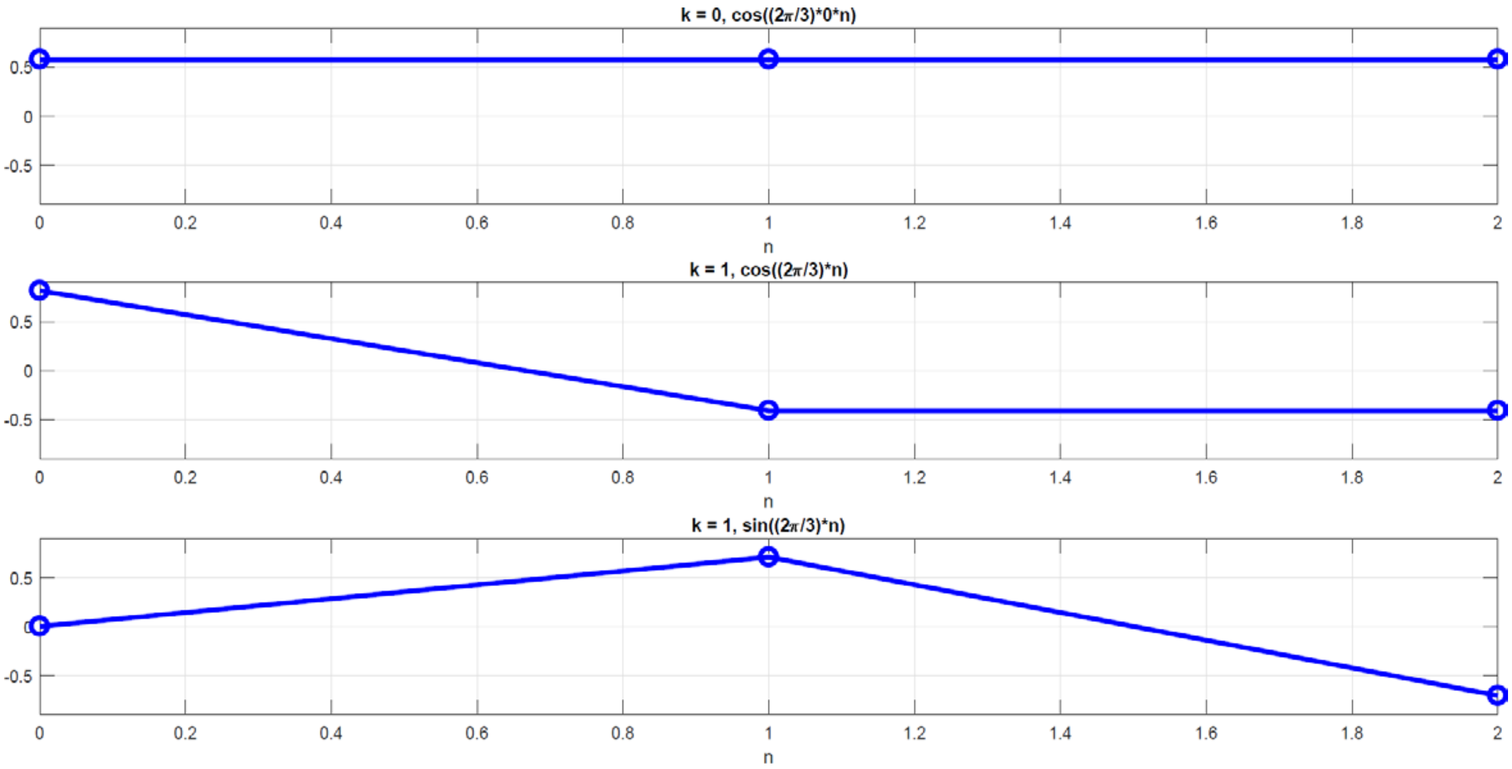
$$x_k(n) = A \sin\left(\frac{2\pi k}{N} n\right)$$

$$y_k(n) = B \cos\left(\frac{2\pi k}{N} n\right)$$

are periodic with $x(n + \frac{N}{k}) = x(n)$, $y(n + \frac{N}{k}) = y(n)$, i.e. their frequencies are k times the higher that the fundamental frequency.



N = 3



~ (1, 1, 1)

~ (1,-0.5,-0.5)

~ (0 , -1 , 1)

Orthogonal?

Periodic functions as basis vectors

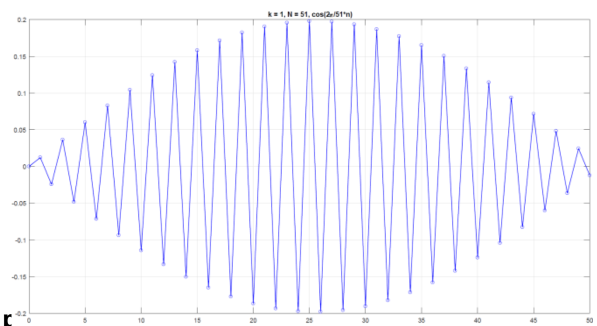
What is the highest frequency harmonic, i.e., the largest k?
For N even

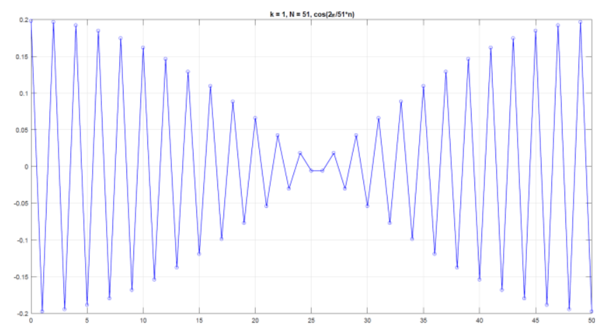
$$y_{N/2}(n) = B\cos\left(\frac{2\pi\frac{N}{2}}{N}n\right) = B\cos\left(\frac{2\pi\frac{N}{2}}{N}n\right) = (-1)^n B$$

For N odd

$$x_{(N-1)/2}(n) = A\sin\left(\frac{2\pi\frac{N-1}{2}}{N}n\right) = A\sin\left(\frac{2\pi\frac{N-1}{2}}{N}n\right) = (-1)^n A\sin\left(\frac{\pi n}{N}\right)$$

$$y_{(N-1)/2}(n) = B\cos\left(\frac{2\pi\frac{N-1}{2}}{N}n\right) = B\cos\left(\frac{2\pi\frac{N-1}{2}}{N}n\right) = (-1)^n B\cos\left(\frac{\pi n}{N}\right)$$





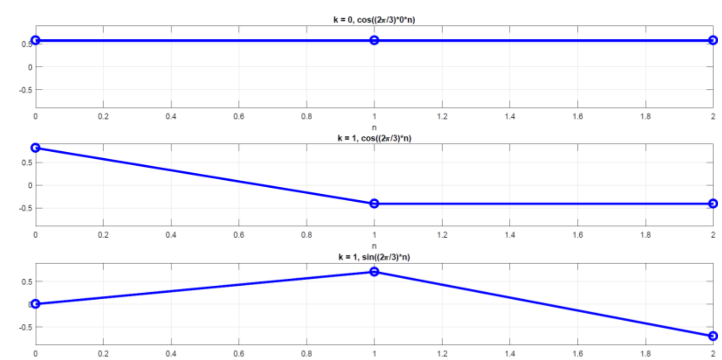
Periodic functions as basis vectors

The range of k's is $k = [0, N/2]$ for even N and $k = [0, (N-1)/2]$ for odd. With $k = 0$ being the constant function

$$x_{(N-1)/2}(n) = B\cos\left(\frac{2\pi 0}{N} n\right) = B$$

So in total we have precisely N discrete sampled harmonic functions or vectors if they are visualized as points in $\mathbb{R}^N$.

$$(\mathbf{x}_k)_n = x_k(n)$$



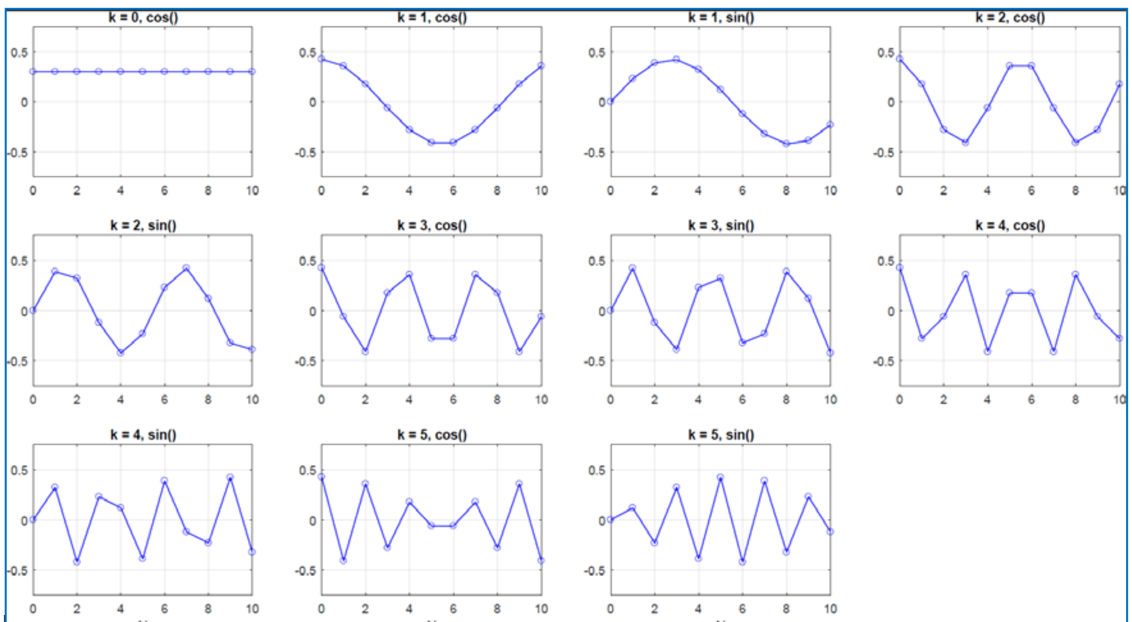
Periodic functions as basis vectors

In the exercise we will show that harmonics are mutually orthogonal, so when properly normalized and numbered from 0 to $N-1$

$$\mathbf{u}_{2k} = \mathbf{x}_k / \|\mathbf{x}_k\|$$

$$\mathbf{u}_{2k+1} = \mathbf{y}_k / \|\mathbf{y}_k\|$$

we have a complete orthonormal basis set!
Hence, we can form a *basis matrix* $\mathbf{U} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \ldots \quad \mathbf{u}_N)$ where $\|\mathbf{u}\| = 1$



Periodic functions as basis vectors

Remind from last week: If $\mathbf{v}$ is a unit vector $||\mathbf{v}|| = 1$ then we can find the projection of $\mathbf{x}$,

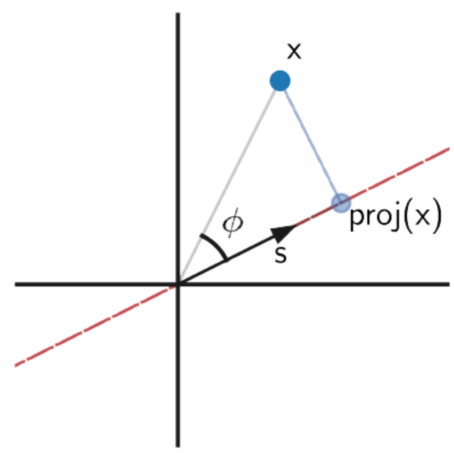
$$\text{proj}_v(\mathbf{x}) = \mathbf{v}||\mathbf{x}|| \cos(\theta) = \mathbf{v}(\mathbf{v}^\top \mathbf{x}).$$

- think of $\mathbf{v}^\top \mathbf{x}$ as the *"coordinate"* of $\mathbf{x}$ along $\mathbf{v}$.

So if we have a *basis matrix* $\mathbf{U} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_D)$ where $||\mathbf{u}_d|| = 1$

$$\mathbf{U}^\top \mathbf{x} = \begin{bmatrix} \mathbf{u}_1^\top \mathbf{x} \\ \mathbf{u}_2^\top \mathbf{x} \\ \vdots \\ \mathbf{u}_D^\top \mathbf{x} \end{bmatrix}$$

$$\text{proj}_v(x) = \underbrace{\frac{v}{||v||}}_{\text{unit factor}} \cdot \underbrace{||x||}_{\text{original length}} \cdot \underbrace{\cos(\phi)}_{\text{fraction parallel to factor}}$$



So $\mathbf{U}^\top \mathbf{x}$ is the vector of *coordinates* in the new periodic basis $\mathbf{U}$!

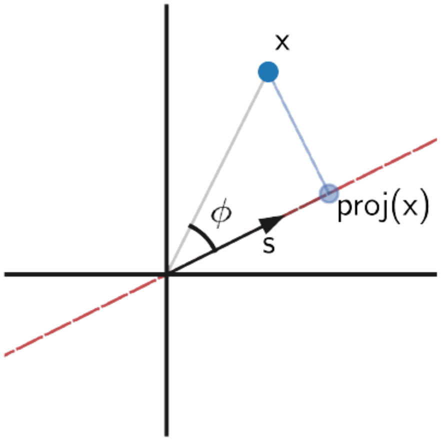
Periodic functions as basis vectors

- Given the periodic function basis we can decompose the "energy" in the sample vector $(\mathbf{x})_j = x_j$, with individual contributions given by the projections

$$\mathbf{z}_m = \left(\mathbf{U}^\top \mathbf{x}\right)_m = \mathbf{u}_m^\top \mathbf{x} = \sum_{j=0}^{N-1} \cos\left(\frac{\pi m j}{N}\right) x(j)$$

$$\sum_k \mathbf{z}_k^2 = \sum_k \left(\mathbf{U}^\top \mathbf{x}\right)_k^2 = \left(\mathbf{U}^\top \mathbf{x}\right)^\top \mathbf{U}^\top \mathbf{x} = \mathbf{x}^\top \mathbf{U}\mathbf{U}^\top \mathbf{x} = \|\mathbf{x}\|^2$$

- Where we remembered $\mathbf{U}\mathbf{U}^\top = \mathbf{U}^\top \mathbf{U} = \mathbf{I}$



Analysis of random signals and filters

- Simplest random signal is the identical, independent distributed (iid.) signal $x(j) \sim \mathcal{N}(0, \sigma^2)$

$$\operatorname{cov}(x(j), x(j')) = \mathbb{E}\{(x(j) - \mu(j))(x(j') - \mu(j'))\} = \sigma^2 \quad \boldsymbol{\Sigma} = \mathbb{E}\{\mathbf{x}\mathbf{x}^\top\} = \sigma^2\mathbf{I}$$

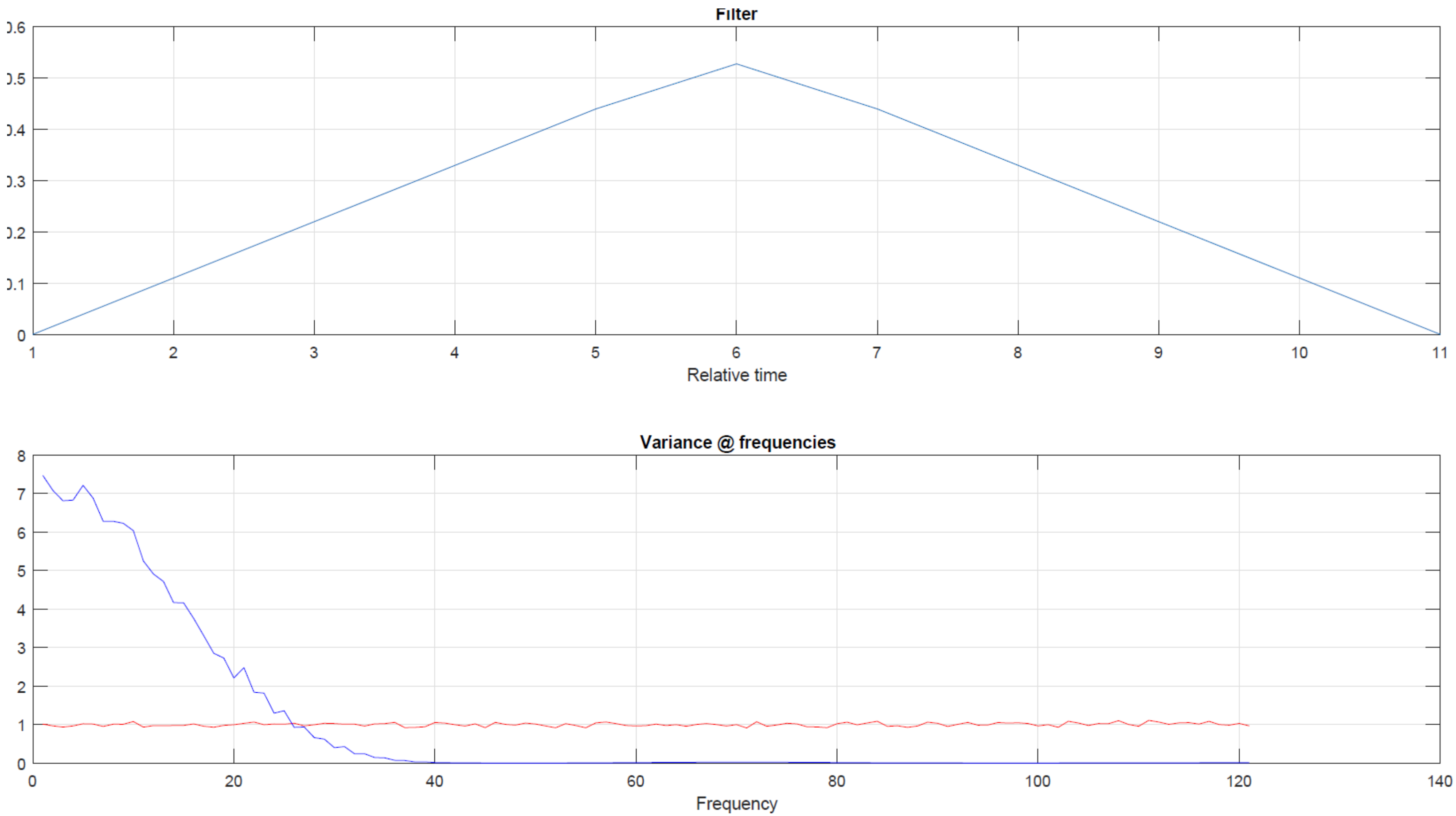
- We can project the iid. signal on the basis $\mathbf{z} = \mathbf{U}^\top \mathbf{x}$

$$\operatorname{cov}(z(k), z(k')) = \mathbb{E}\{z(k)z(k')\}$$

$$\mathbb{E}\{\mathbf{U}^\top \mathbf{x} \left(\mathbf{U}^\top \mathbf{x}\right)^\top\} = \mathbf{U}^\top \mathbb{E}\{\mathbf{x}\mathbf{x}^\top\}\mathbf{U} = \sigma^2 \mathbf{U}^\top \mathbf{I} \mathbf{U} = \sigma^2 \mathbf{I}$$

- Conclusion: the iid. random signal has evenly distributed energy over the set of harmonics

Local averaging convolution (filter)



Analysis of random signals and filters

- We can construct a simple random periodic signal using the basis

$$x(n) = a\cos\left(\frac{2\pi mn}{N}\right) + b\sin\left(\frac{2\pi mn}{N}\right) = c\cos\left(\frac{2\pi mn}{N} + \phi\right)$$

Where the amplitude is $c = \sqrt{a^2 + b^2}$ and the phase is given by $\tan\phi = \frac{a}{b}$

- Let $a_m, \_m \sim \mathcal{N}(0, \sigma^2)$ to create a signal of frequency $F = \frac{m}{N}$, and with random amplitude and phase

The signal can be written in matrix notation

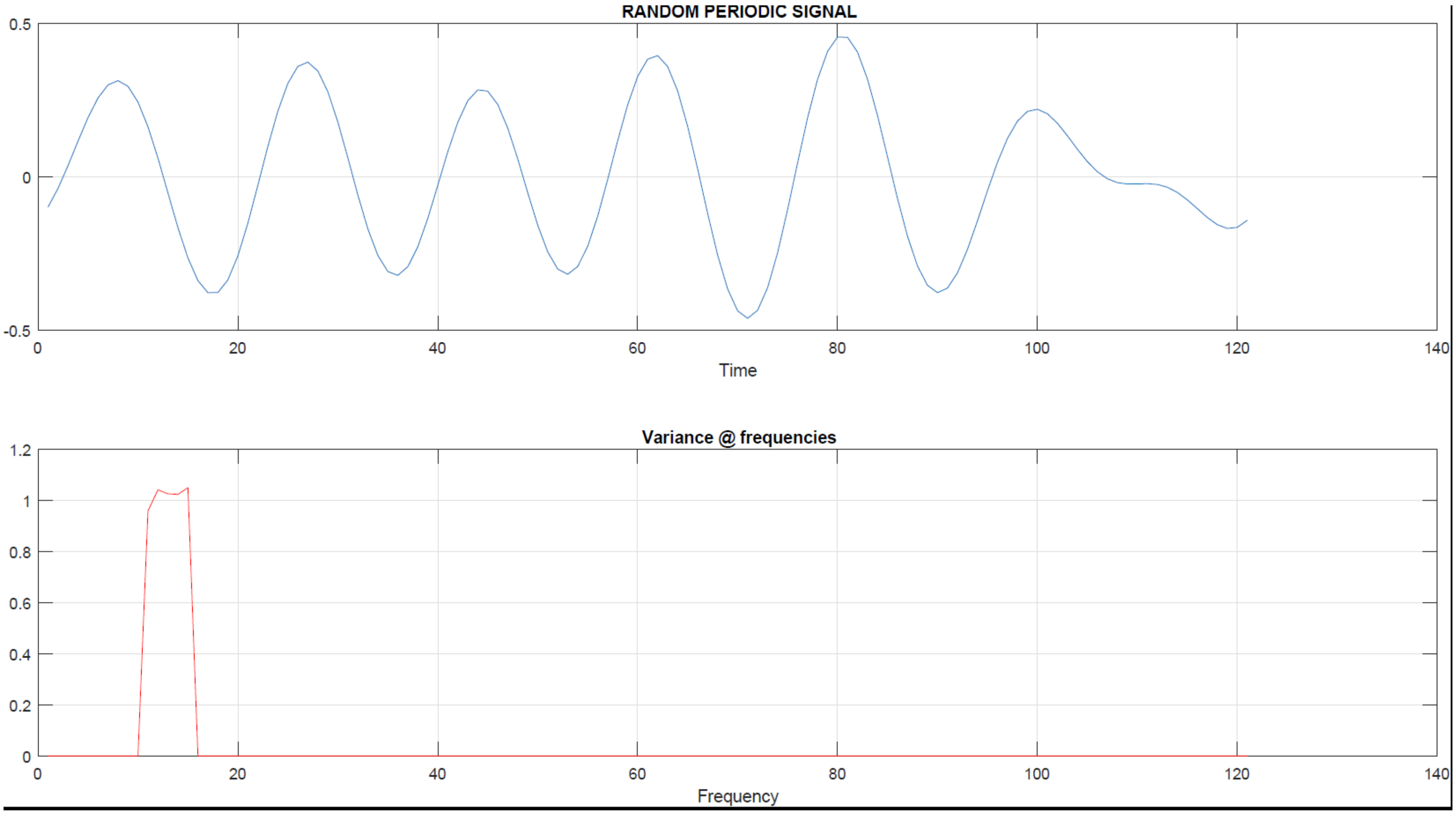
$$\mathbf{x} = a\mathbf{u}_m + b\mathbf{u}_{m+1}$$

Hence if we project on the basis

$$\mathbf{z} = \mathbf{U}^\top(a\mathbf{u}_m + b\mathbf{u}_{m+1}) = (0, 0, 0, 0, a, b, 0, 0 \ldots 0)^\top$$

The energy of the signal is confined - as expected - to the two basis functions at the given frequency.

Random periodic signal



Filters - circular convolutions

Local weighted average or difference

$$w(m) = h_{-1}x\left(\mathrm{mod}(m-1,N)\right) + h_0 x\left(\mathrm{mod}(m,N)\right) + h_1 x\left(\mathrm{mod}(m+1,N)\right)$$

e.x.

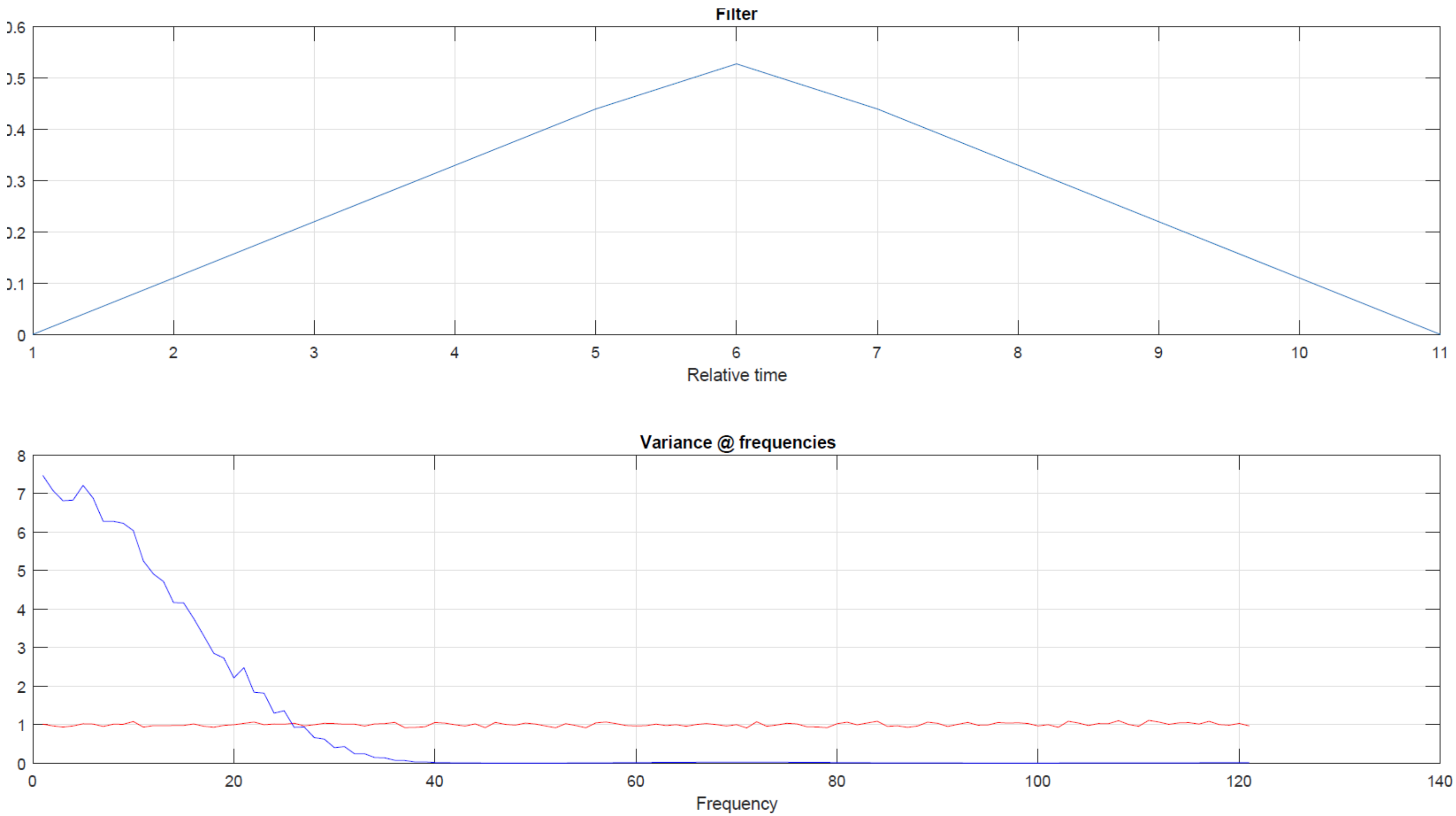
$$w(m) = 0.25x(m-1) + 0.5x(m) + 0.25x(m+1)$$

$$w(m) = \sum_{t=-L}^{L} h_t x\left(\mathrm{mod}(m+t,N)\right) \equiv \sum_{j=1}^{N} \tilde{h}(\mathrm{mod}(m-j,N))x(j) \equiv \sum_{j=1}^{N} H_{m,j}x(j)$$

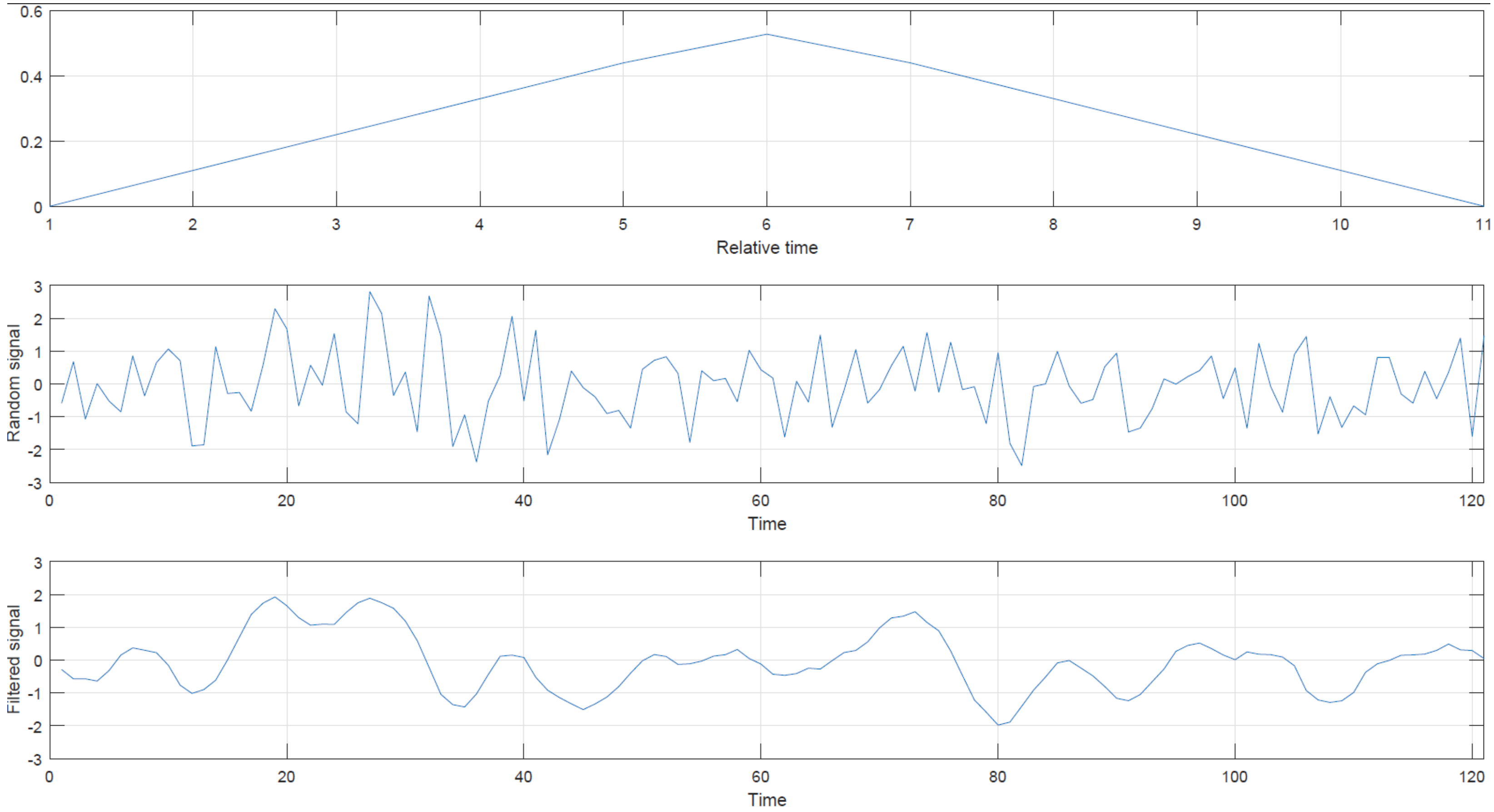
Where the matrix H is banded

$$\mathbf{w} = \mathbf{H}\mathbf{x}$$

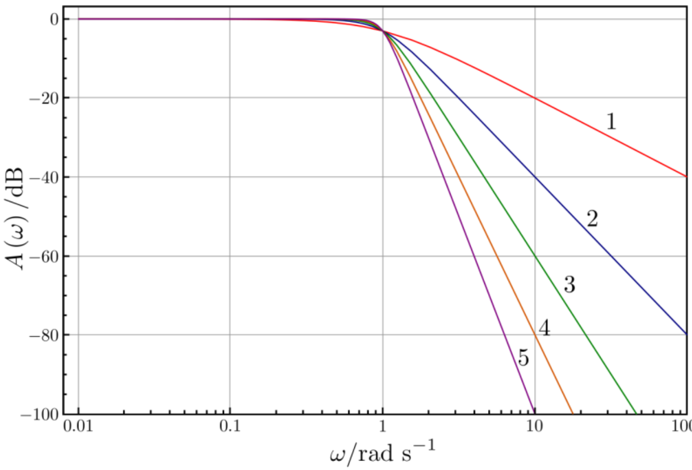
Local averaging convolution (filter)



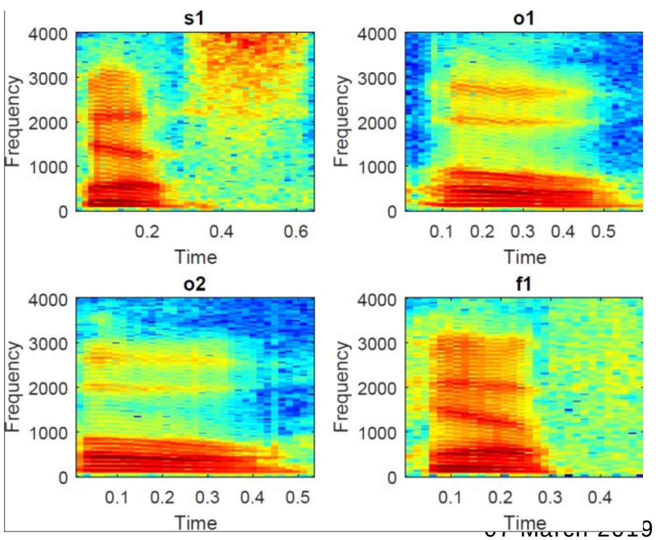
Local averaging convolution (filter)



Production tools



The “butterworth” filters are based on fast implementation of the projections based on the Fast Fourier Transform (Nlog(N) operations instead of N^2)



The spectrogram function applies band power calculation for overlapping windows in a long non-stationary signal

https://www.dspguide.com/ch33/6.htm

Session 2

Frequency vibrating string

$$f_n = \frac{n}{2L}\sqrt{\frac{T}{\mu}}$$

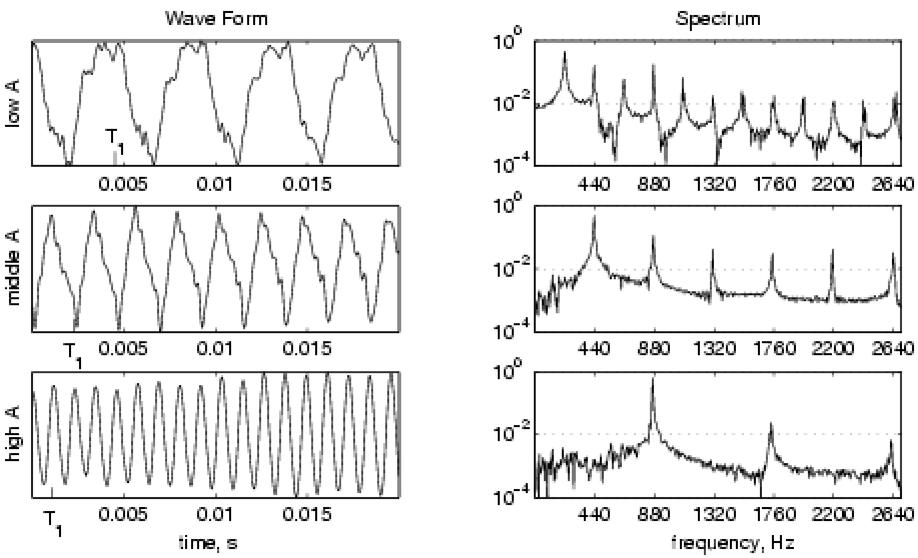
T tension,

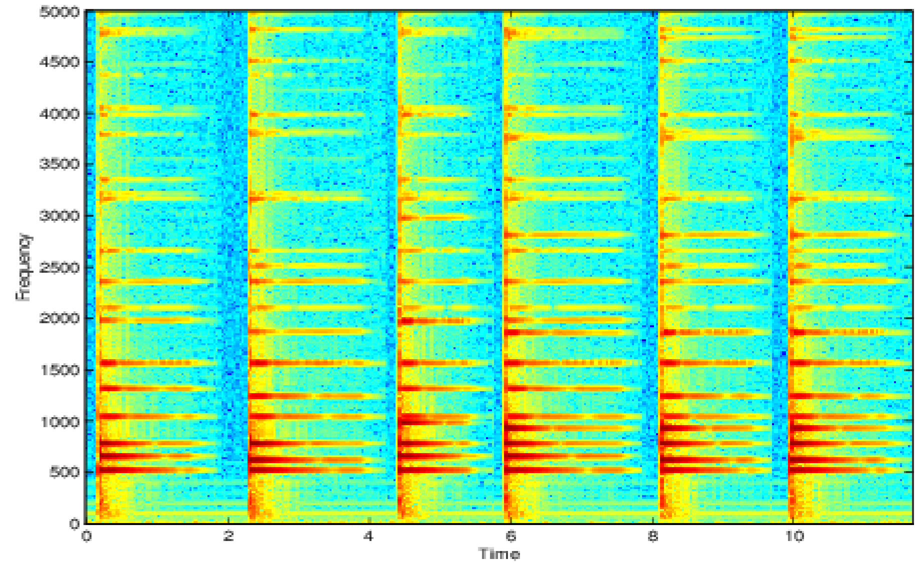
mu mass/length

L length

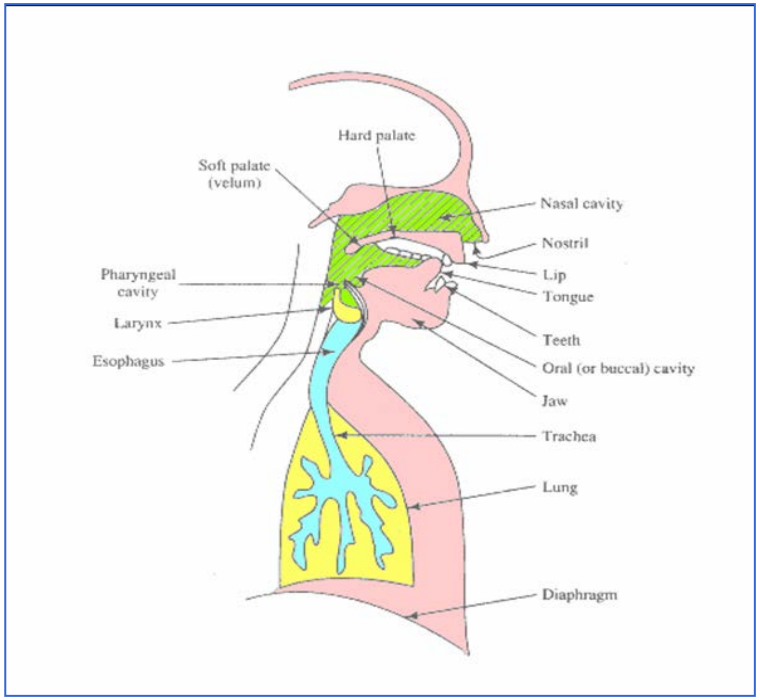
“Tones”

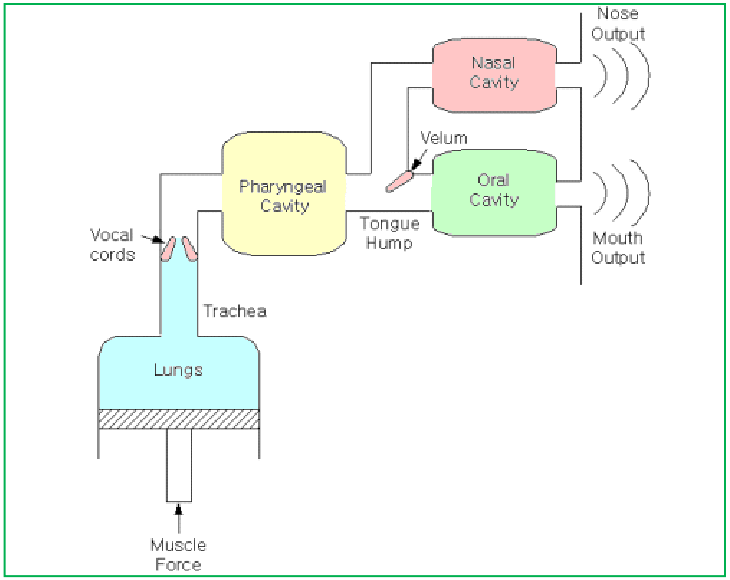
- Single instruments
- Vibrating strings.. piano



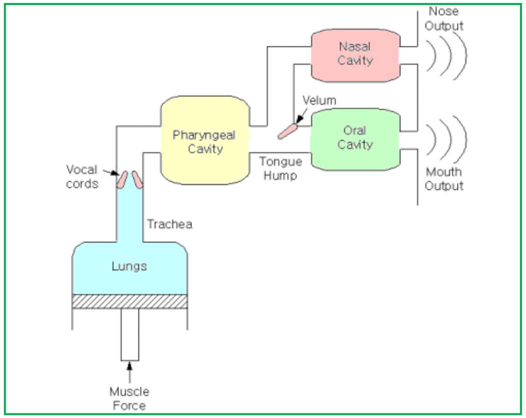


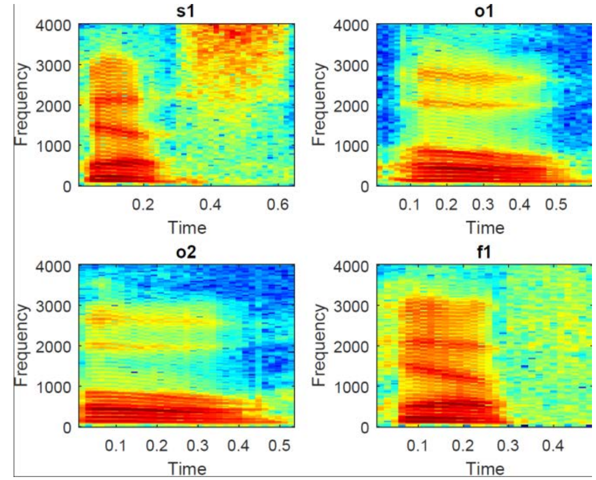
Speech production

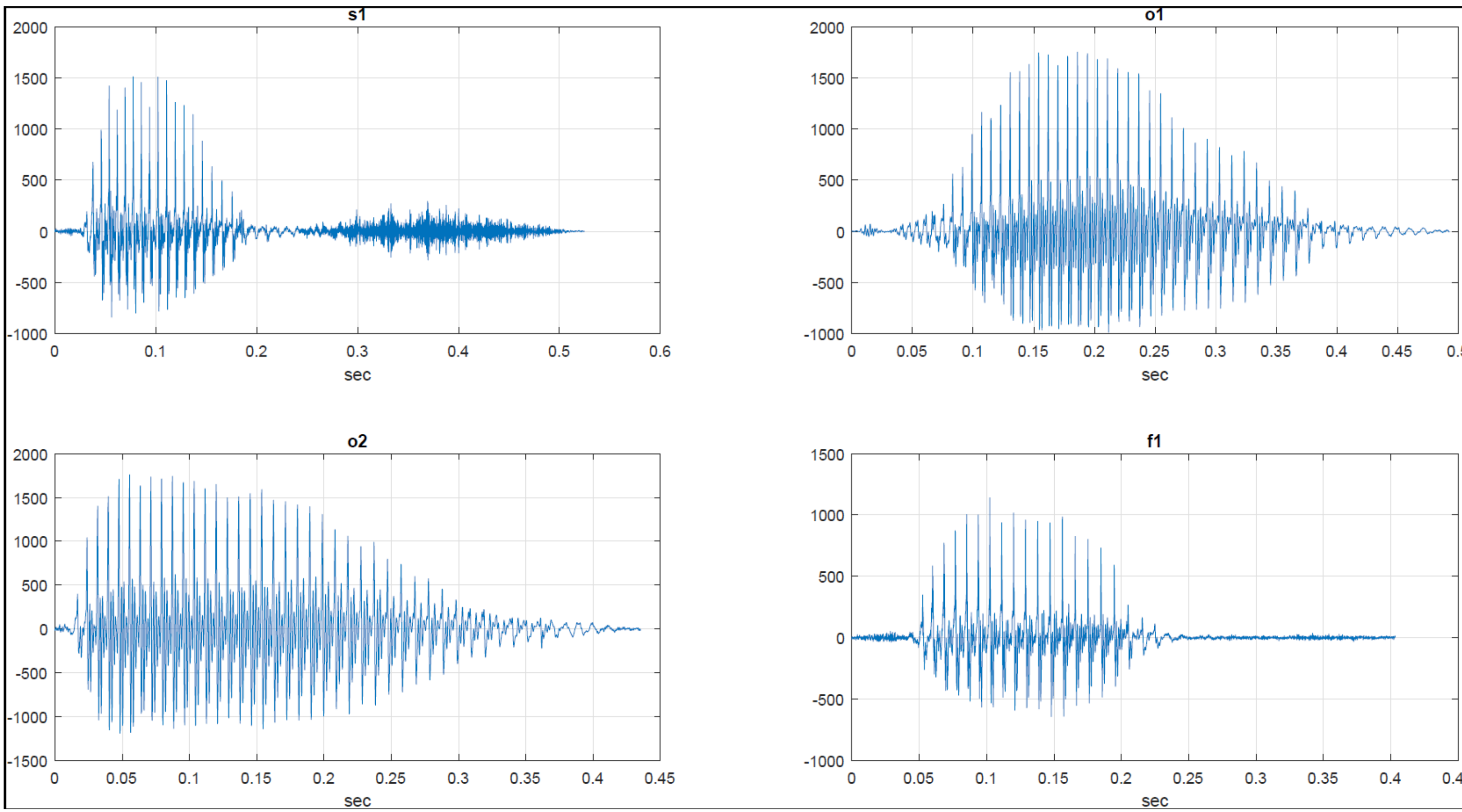




https://www.youtube.com/watch?v=uTOhDqhCKQs


Nose Output
Nasal Cavity
Velum
Pharyngeal Cavity
Oral Cavity
Tongue Hump
Mouth Output
Vocal cords
Trachea
Lungs
Muscle Force


s1
o1
o2
f1
Frequency
Time


s1
o1
o2
f1
sec

Power distribution over time and frequency