NBE-4070 : Basics of Biomedical Data Analysis

Stéphane Deny
Prof. in Neuroscience and Biomedial Engineering and Computer Science
Aalto University

Lecture 2: Fourier Transform, Wavelet Transforms, Spectrograms, High-pass, Low-pass filters

Quiz 1

Question 1

What does the standard deviation of a sample measure?

- ☐ a. A range of values where the true mean of the population probably resides.
- ☐ b. The spread of the data around the mean.

Question 2

What does a confidence interval indicate?

- ☐ a. A range of values where the true mean of the population probably resides.
- ☐ b. The spread of the data around the mean.

Question 3

What is a t-test?

- ☐ a. It is a statistical test to compare the standard deviations of two populations from their samples.
- ☐ b. It is a statistical test to compare the means of two populations from their samples.

Question 4

Select the correct statement:

- ☐ a. A statistically significant effect can correspond to a very small effect size
- ☐ b. When an effect is statistically significant, it always correspond to an important effect size

Statistical significance ≠ Effect size

J Grad Med Educ. 2012 Sep; 4(3): 279–282.
doi: 10.4300/JGME-D-12-00156.1

Using Effect Size—or Why the *P* Value Is Not Enough

Gail M. Sullivan, MD, MPH and Richard Feinn, PhD

$$\sigma(X) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

$$SE = \frac{\sigma(X)}{\sqrt{N}}$$

Why Isn't the *P* Value Enough?

Go to:

Statistical significance is the probability that the observed difference between two groups is due to chance. If the *P* value is larger than the alpha level chosen (eg, .05), any observed difference is assumed to be explained by sampling variability. With a sufficiently large sample, a statistical test will almost always demonstrate a significant difference, unless there is no effect whatsoever, that is, when the effect size is exactly zero; yet very small differences, even if significant, are often meaningless. Thus, reporting only the significant *P* value for an analysis is not adequate for readers to fully understand the results.

For example, if a sample size is 10 000, a significant *P* value is likely to be found even when the difference in outcomes between groups is negligible and may not justify an expensive or time-consuming intervention over another. The level of significance by itself does not predict effect size. Unlike significance tests, effect size is independent of sample size. Statistical significance, on the other hand, depends upon both sample size and effect size. For this reason, *P* values are considered to be confounded because of their dependence on sample size. Sometimes a statistically significant result means only that a huge sample size was used.[3]

A commonly cited example of this problem is the Physicians Health Study of aspirin to prevent myocardial infarction (MI).[4] In more than 22 000 subjects over an average of 5 years, aspirin was associated with a reduction in MI (although not in overall cardiovascular mortality) that was highly statistically significant: $P < .00001$. The study was terminated early due to the conclusive evidence, and aspirin was recommended for general prevention. However, the effect size was very small: a risk difference of 0.77% with $r^2 = .001$—an extremely small effect size. As a result of that study, many people were advised to take aspirin who would not experience benefit yet were also at risk for adverse effects. Further studies found even smaller effects, and the recommendation to use aspirin has since been modified.

Outline of the course

1. Mean, Standard Deviation, Standard Error, Confidence Intervals, T-test
2. Fourier Transform, Wavelet Transforms, Spectrograms, High-pass, Low-pass filters
3. Principal Component Analysis / Singular Value Decomposition
4. Linear Regression / Logistic Regression
5. Clustering Methods
6. Non-linear Methods: Independant Component Analysis, t-Stochastic Neighbour Embedding, Random Forests, Deep Networks
7. Invited lecture from the biomedical industry

Spend 5 minutes explaining to your neighbor

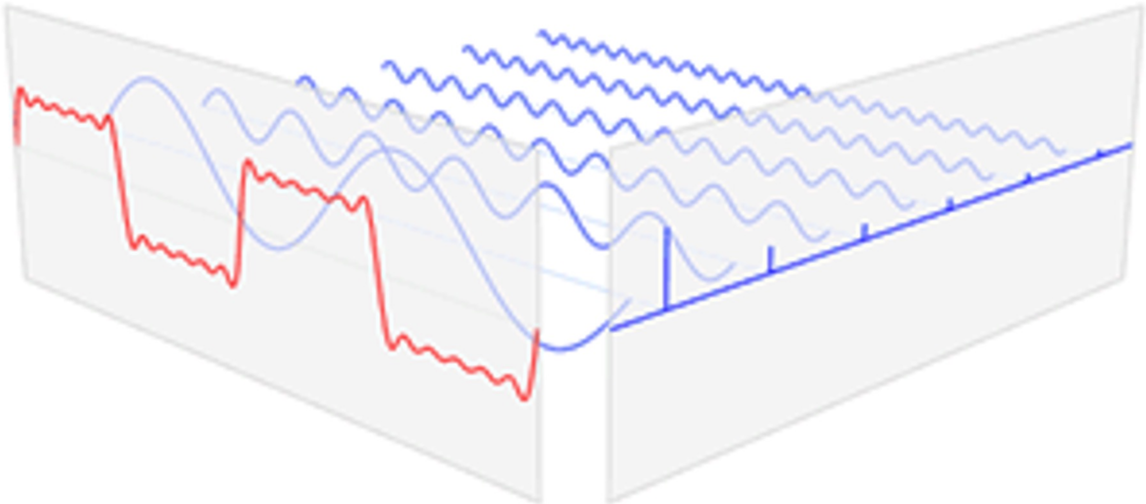
- What is the Fourier transform?

(switch roles)

- What can it be useful for?

Visualization of the Fourier Transform

The Fourier transform expresses a signal into a sum of sine and cosine functions.



source: https://en.wikipedia.org/wiki/Fourier_transform

Fourier Transform : definition

It is a change of basis, which re-expresses an input sequence X into a new basis of sine and cosine functions (i.e. *frequency domain*):

$$\widehat{x_k} = \sum_{n=0}^{N-1} x_n \left[\cos(\frac{2\pi k}{N} n) - i. \sin(\frac{2\pi k}{N} n) \right]$$

where $\widehat{x_k}$ is the k^{th} Fourier component

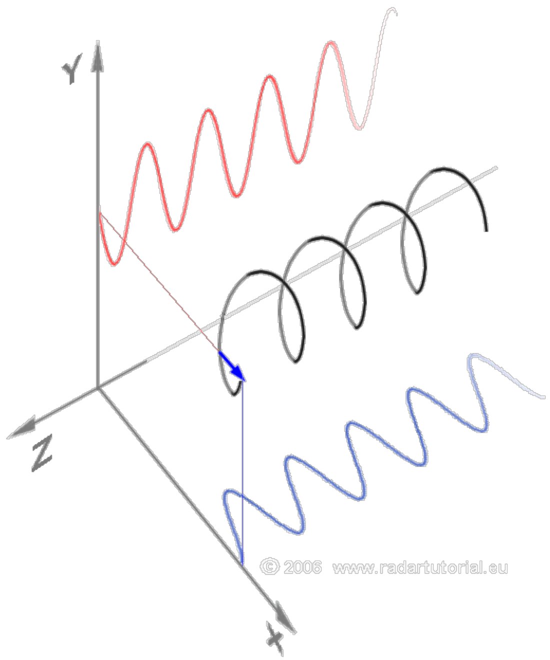
k is a frequency index

x_n is the n^{th} element of the sequence

N is the total size of the input sequence

Visualization of the Fourier Transform

Expressing the signal as complex sinusoids has the benefit of compactly representing all necessary information about each frequency, namely the amplitude and phase of the signal for this frequency

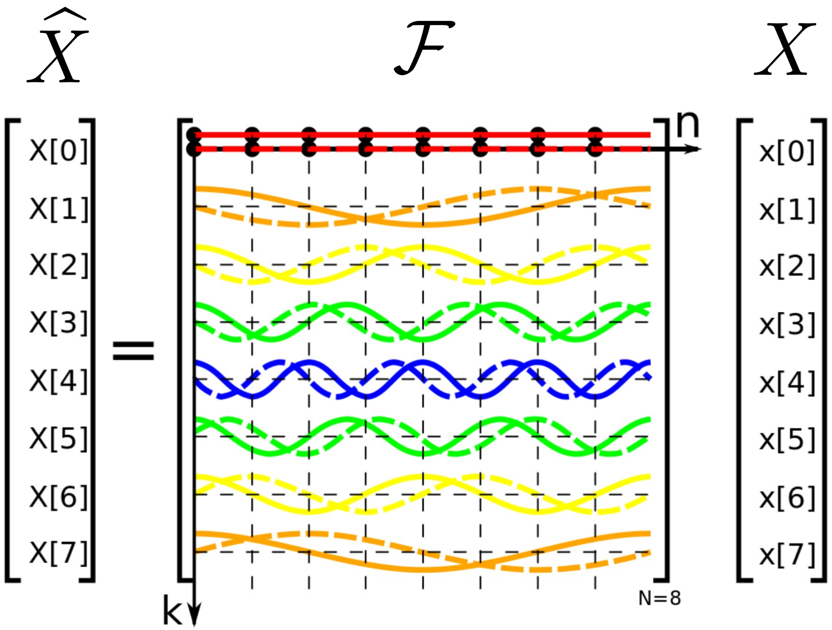


$$f_{n,k} = \cos(\frac{2\pi k}{N}n) - i.\sin(\frac{2\pi k}{N}n)$$

source: https://en.wikipedia.org/wiki/Fourier_transform

Discrete Fourier Transform (DFT) : definition

It is a change of basis, which re-expresses an input sequence X into a new basis of sine and cosine functions (i.e. *frequency domain*):



where $\widehat{X}$ is the Fourier representation

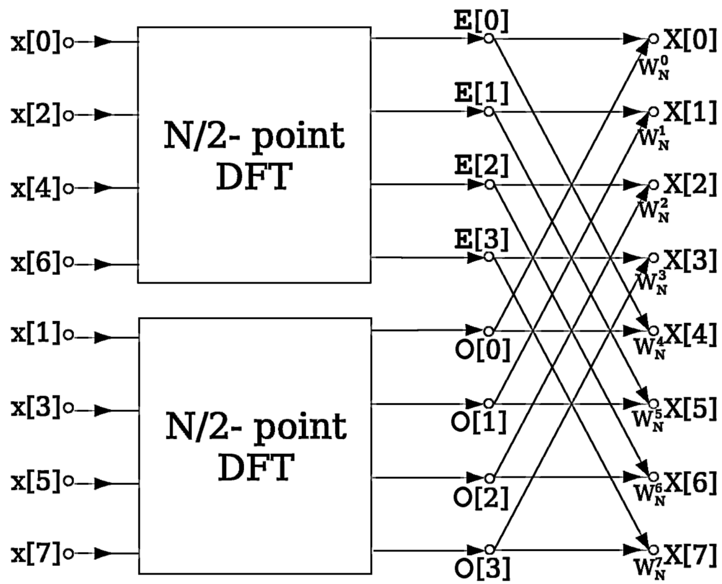
$\mathcal{F}$ is the Fourier transform

X is the input sequence

The real part (cosine wave) is denoted by a solid line, and the imaginary part (sine wave) by a dashed line.

Fast Fourier transform: definition

Fast Fourier transform (FFT) is an algorithm that computes the discrete Fourier transform (DFT) of a sequence very efficiently (complexity N.log(N) instead of N^2, where N is the size of the input sequence), by taking advantage of the very specific structure of the DFT transform:



source: https://en.wikipedia.org/wiki/Fast_Fourier_transform

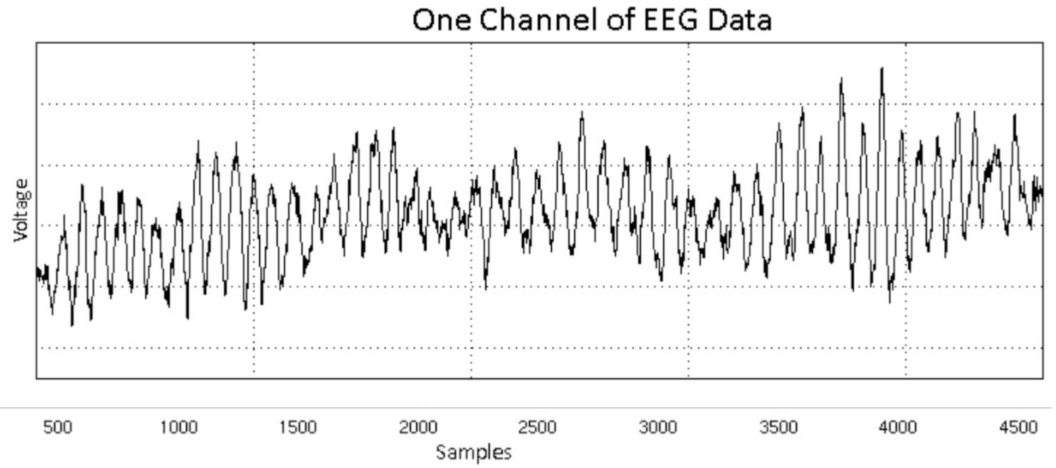
Fourier Spectrum: definition

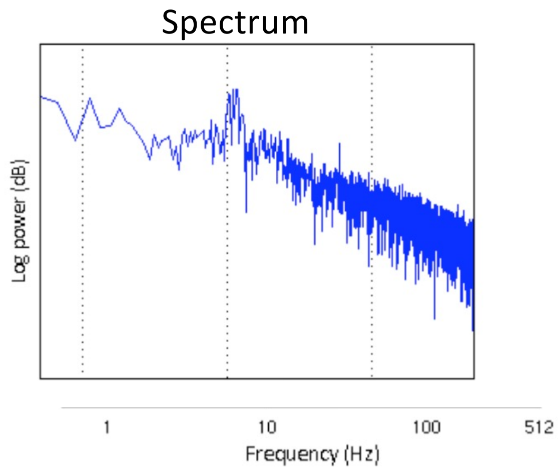
The Fourier spectrum of a signal or time series describes how the power of this signal is distributed over frequency. It is defined as:

$$\mathcal{S}(k) = |\widehat{x_k}|^2$$

where $\mathcal{S}(k)$ is the power spectrum at frequency k

$\widehat{x_k}$ is the Fourier component at frequency k





Low-pass, high-pass and band-pass filters

- A low-pass filter is a filter that passes signals with a frequency lower than a selected cutoff frequency and attenuates signals with frequencies higher than the cutoff frequency.

- A high-pass filter is a filter that passes signals with a frequency higher than a certain cutoff frequency and attenuates signals with frequencies lower than the cutoff frequency.

- A band-pass filter is the combination of a high-pass and a low-pass filter

Low-pass, high-pass and band-pass filters

Figure 3: original image

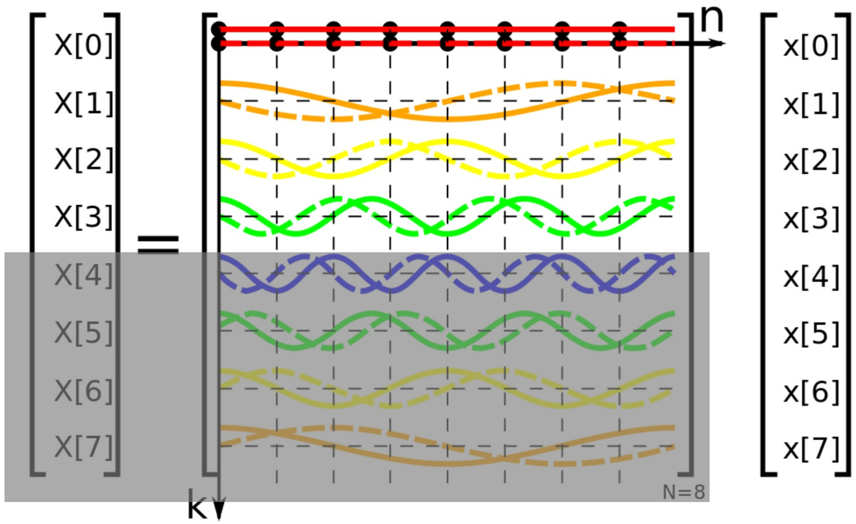
Figure 4: result of Gaussian low pass filter

Figure 5: Gaussian high pass filter

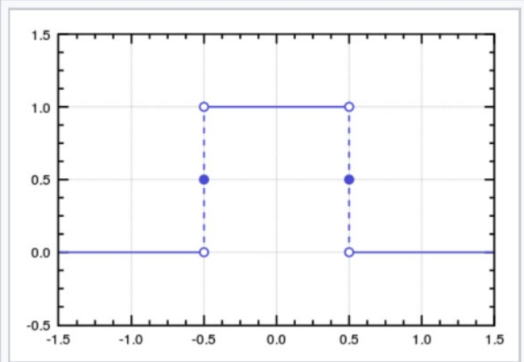
Low-pass filtering

Low-pass filtering consists in getting rid of the high-frequency components:

$$\widehat{X}_{trunc} \qquad \mathcal{F}_{trunc} \qquad X$$



low-pass filtering window



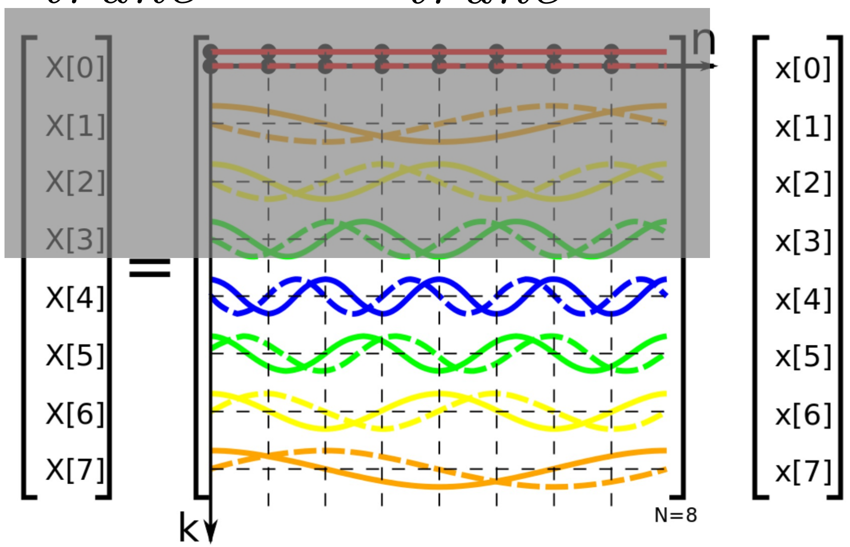
The rectangular function, the frequency response of the sinc filter.

$$X_{low-pass} = \mathcal{F}^{t*}_{trunc}\widehat{X}_{trunc}$$

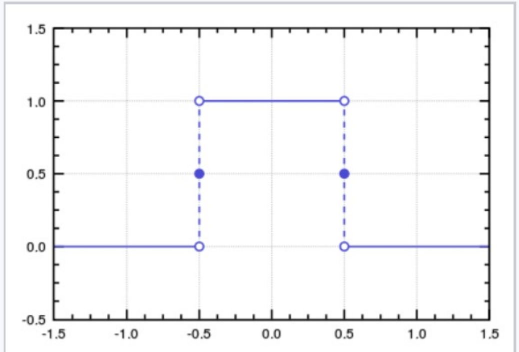
High-pass filtering

<u>High-pass filtering</u> consists in <u>getting rid of the low frequency components</u>:

$$\widehat{X}_{trunc} \qquad \mathcal{F}_{trunc} \qquad X$$



low-pass filtering window



The rectangular function, the frequency response of the sinc filter.

$$X_{high-pass} = \mathcal{F}^{t*}_{trunc}\widehat{X}_{trunc}$$

Ringing artifacts and how to avoid them

Ringing artifacts are oscillations in the signal that appear as a result of sharp low-pass filtering:

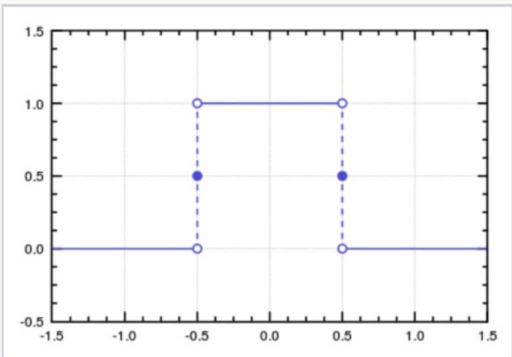
Same image without ringing artifacts.

Image showing ringing artifacts. 3 levels on each side of transition: overshoot, first ring, and (faint) second ring.

Ringing artifacts and how to avoid them

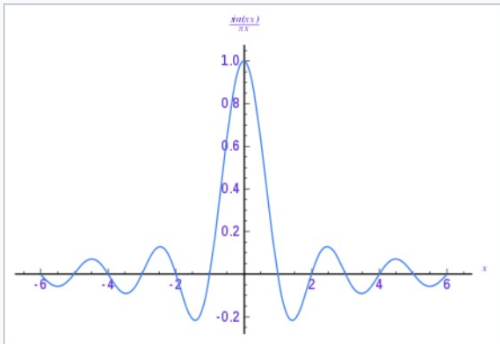
A sharp low-pass filtering window in frequency domain artificially creates oscillations in the temporal domain because the Fourier transform of a rectangular window is the 'sinc' function:

low-pass filtering window



The rectangular function, the frequency response of the sinc filter.

Fourier transform of this low-pass filtering window



The normalized sinc function, the impulse response of the sinc filter.

Ringing artifacts and how to avoid them

To avoid ringing artifacts, one can smoothly attenuate high-frequency components, a process called ‘windowing’. Many possible windows exist:

There are many possible low-pass filter types or configurations. The most commonly used are Butterworth, Chebyshev, and Bessel filters. The frequency responses of these filters differ, and they offer some key differentiators depending on the application (Figure 3).

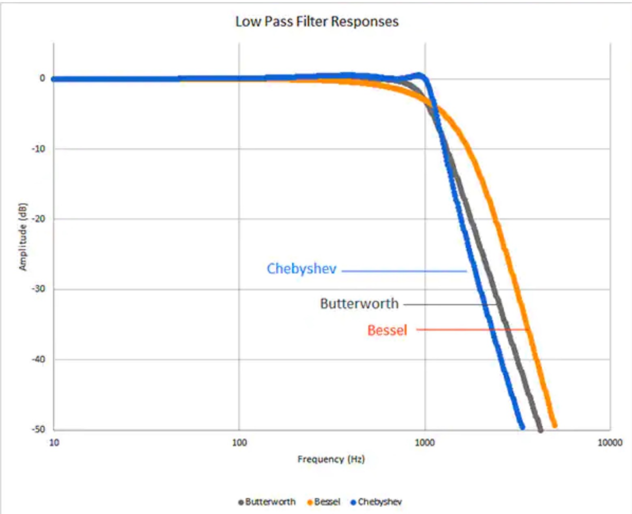


Figure 3: A comparison of Butterworth (gray), Chebyshev (blue), and Bessel (orange) filter frequency responses. The filter types differ in passband flatness, phase delay, and slope of the transition region. (Image source: Digi-Key Electronics)

The response of these filters to a pulse in the time domain is useful in understanding the appropriate filter type selection (Figure 4).

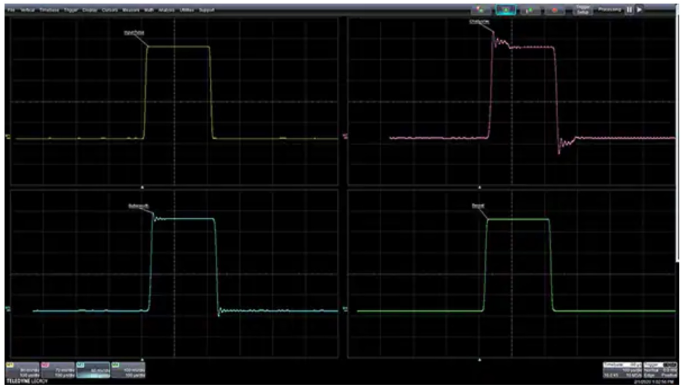


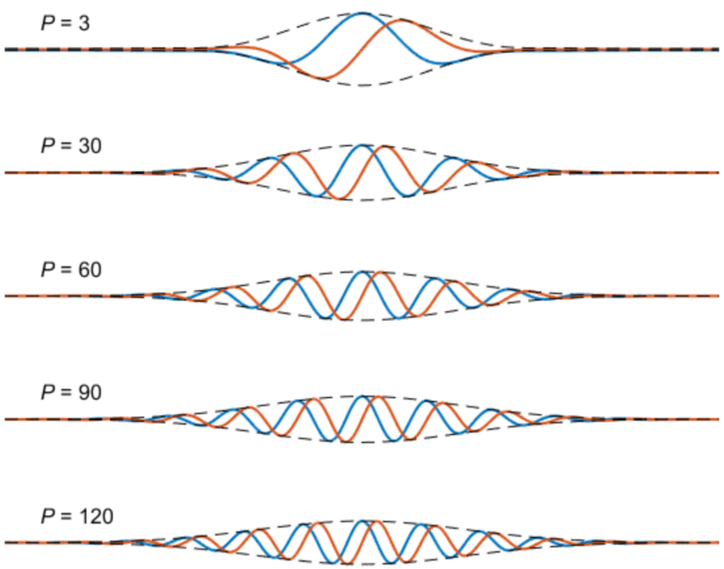
Figure 4: Filter response to an input pulse (upper left) shows the differences in time domain pulse response of the Chebyshev (upper right), Butterworth (lower left), and Bessel (lower right) filter types. (Image source: Digi-Key Electronics)

The Bessel filter’s linear phase response with frequency passes the pulse with minimal distortion, but it doesn’t have the amplitude flatness of the Butterworth filter or the sharp cutoff of the Chebyshev filters. The type of filter selected depends on the application:

- The Butterworth filter should be chosen if amplitude accuracy is the paramount concern
- The Chebyshev filter would be the filter of choice if the desired sampling rate is close to the signal bandwidth
- The Bessel filter is the best choice if pulse fidelity is the primary concern

Wavelet transform: definitions

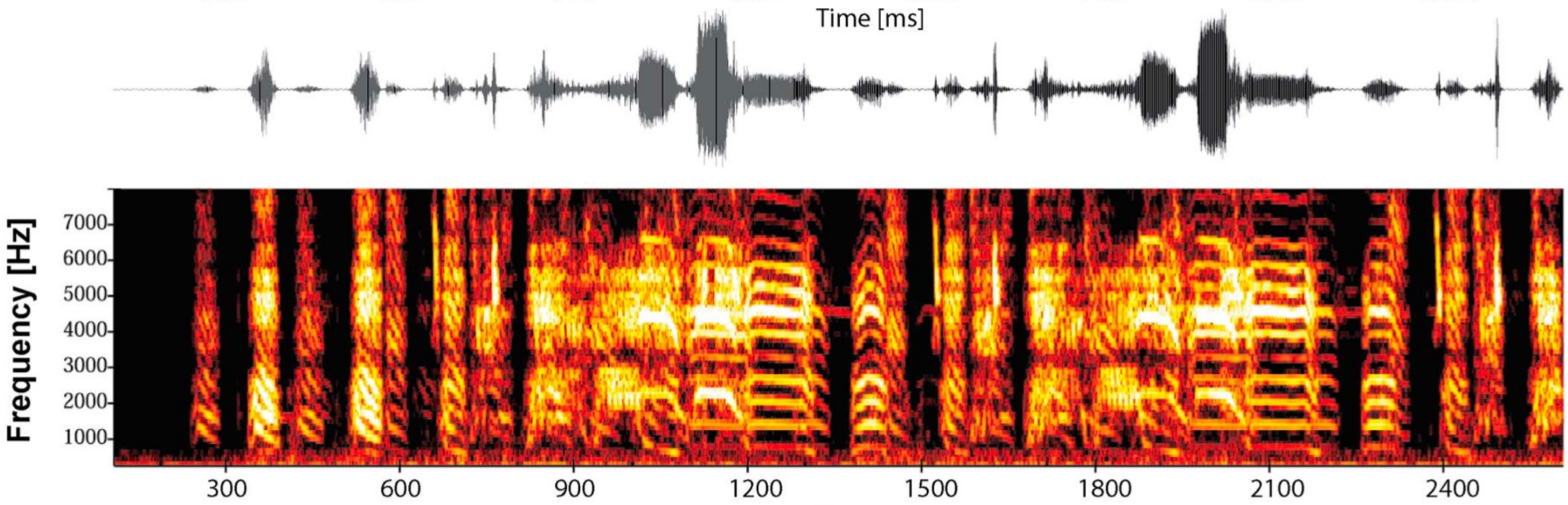
- A wavelet is a wave-like oscillation with an amplitude that begins at zero, increases or decreases, and then returns to zero one or more times.



Wavelet transform: definitions

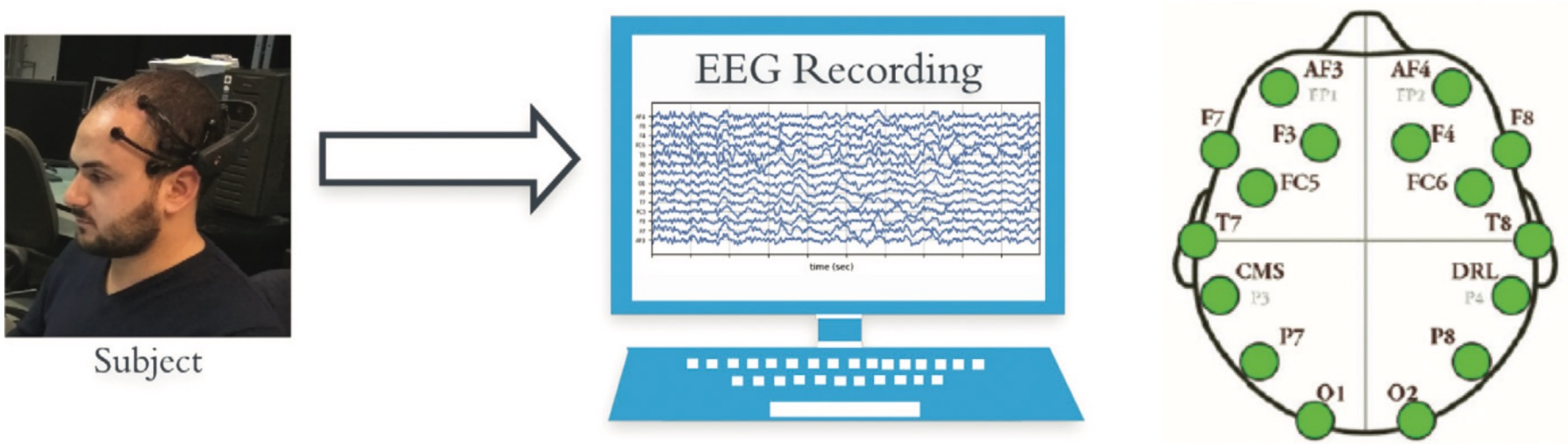
- A wavelet is a wave-like oscillation with an amplitude that begins at zero, increases or decreases, and then returns to zero one or more times.
- A wavelet transform is a projection of the signal onto a family of wavelets of increasing frequency.
- A wavelet transform is often used to compute spectrograms (also called scalograms), which are visual representation of the spectrum of frequencies of a signal as it varies with time.

Example: Spectrogram for syllable detection in the songbird



Fukushima & Margoliash 2014

Case study 1: EEG recordings



(a) EEG recording

(b) 10-20 System

Figure 1.
EEG recording setup: (a) a wireless device Emotiv Epoch mounted on a subject, transmitting EEG signal to a computer. (b) Electrode positions as 10–20 system, source: https://www.emotiv.com/.

Spectrum of an EEG recording

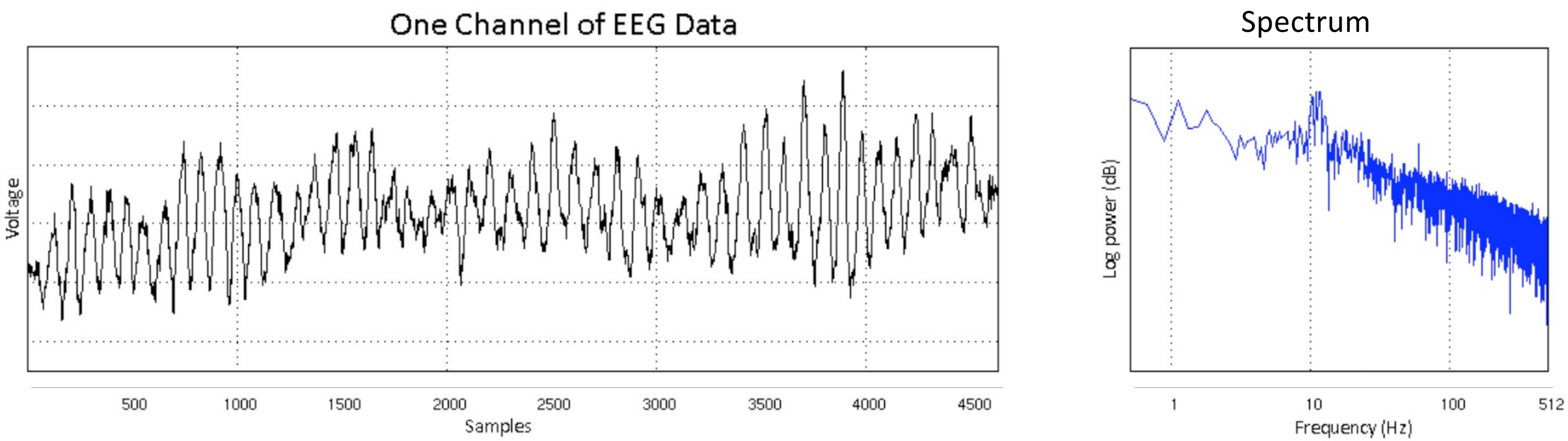


Figure 3. Example of the signal and corresponding spectra for 4.5 seconds of EEG data for one electrode. The top left panel shows the time-course of the raw EEG. The corresponding power spectrum is shown on the right panel, in logarithmic scale.

Artifacts in EEG recordings

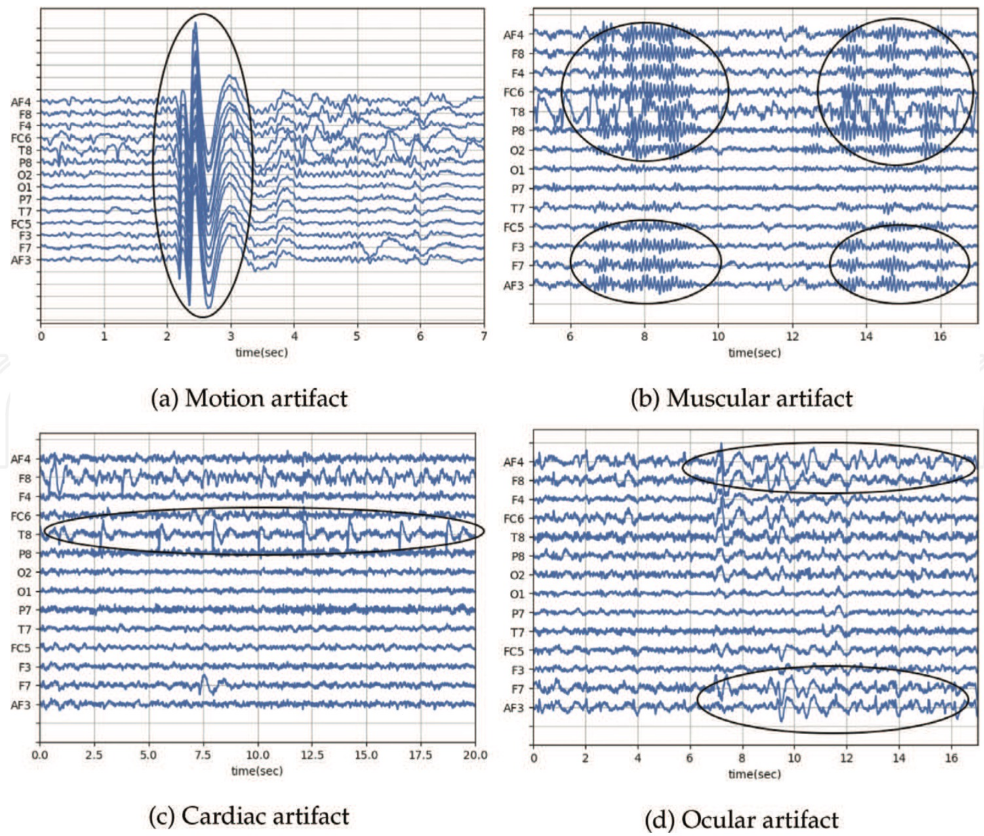


Figure 4.
Common type of artefacts in EEG. Corresponding artefacts are circled in the figure.

Artifact: an electroencephalographic wave that arises from sources other than the brain

EEG spectrograms

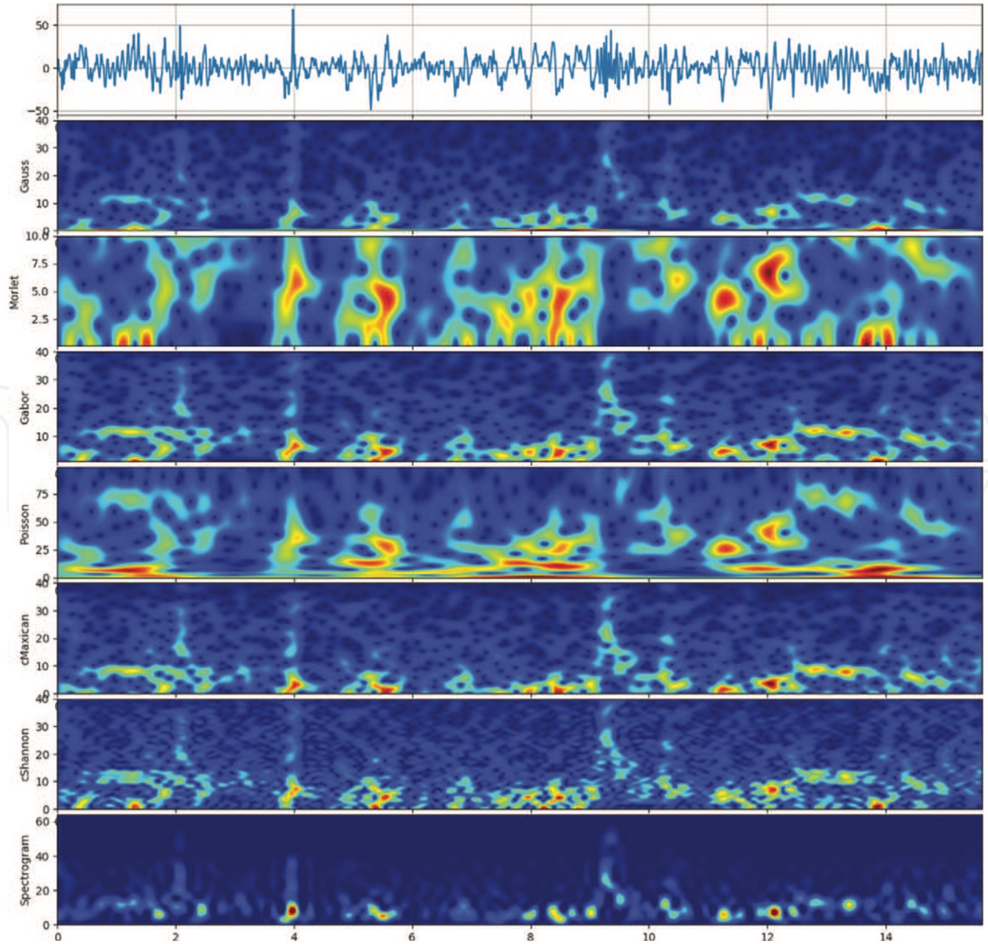
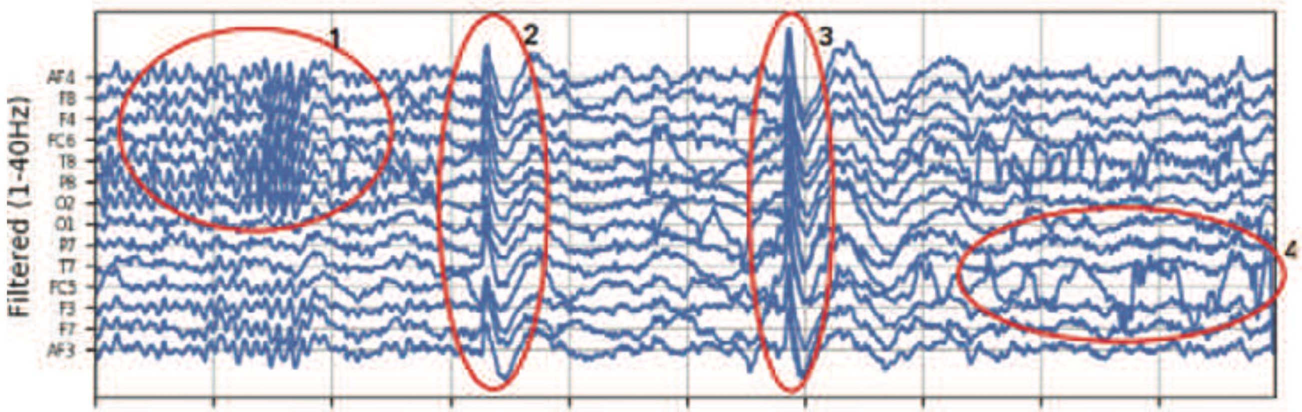
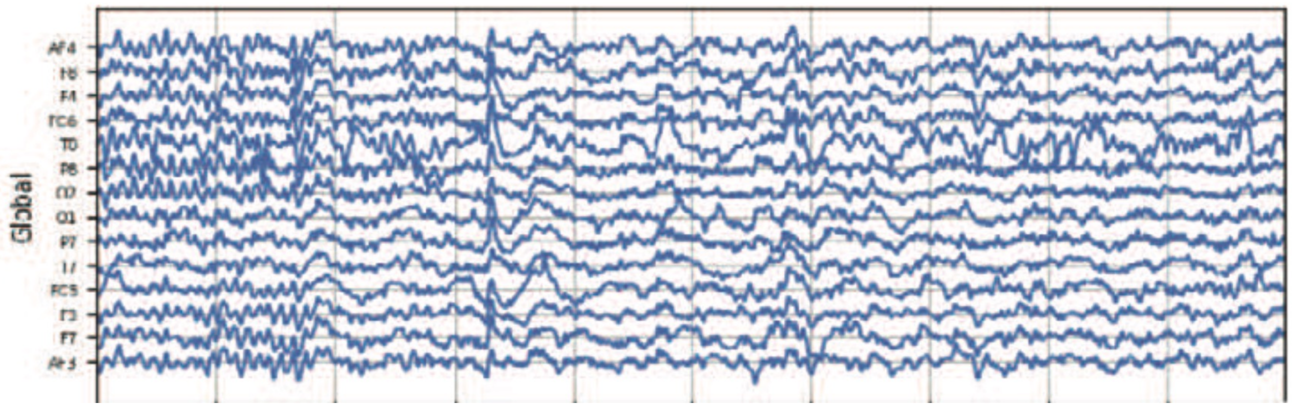


Figure 7.
Scalogram and spectrogram of a segment of signal channel EEG signal with six wavelet functions and STFT. Figure obtained using spkit python library - https://spkit.github.iohttps://spkit.github.io

Denoising EEG using the spectrogram



Denoised by removing wavelet coefficients above a certain magnitude



Band-pass analysis of EEG recordings

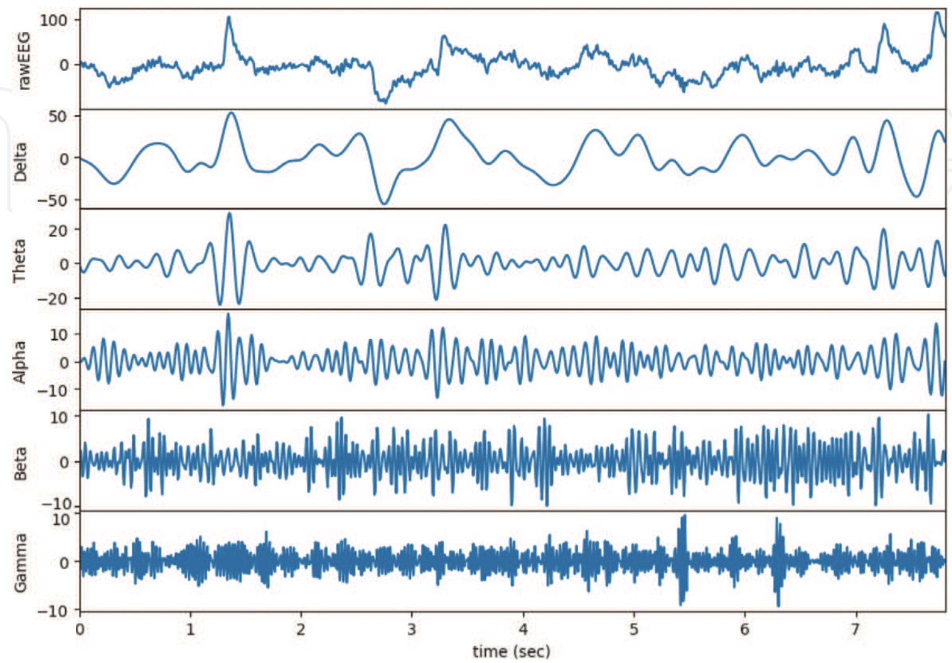
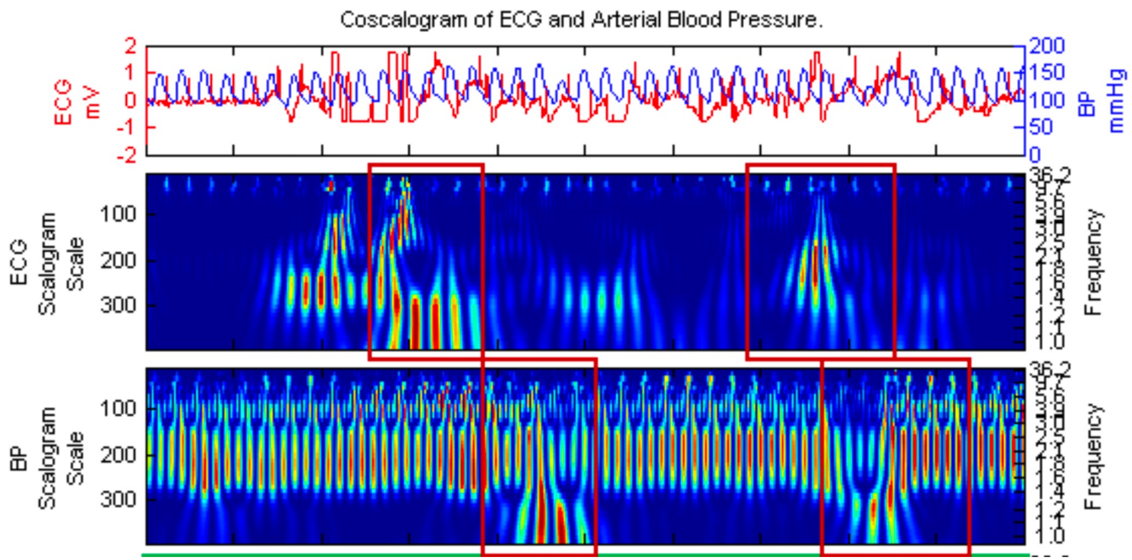


Figure 2.
The signal channel raw EEG signal and corresponding frequency bands: Delta (0.1 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 14 Hz), Beta (14 – 30 Hz), gamma (30 – 63 Hz).

Case study 2: Spectrogram for data analysis



Relation between electrocardiographs (ECG) and blood pressure (BP) measurements over the same time period.

In the plot presented, one can see ECG and BP signals, along with their scalograms. It is visible that two (framed) events, visible in the scalogram of ECG, have corresponding reaction in BP scalogram, but with some visible time delay.

Next lecture

- Principal Component Analysis

References

- https://en.wikipedia.org/wiki/Spectrogram
- https://en.wikipedia.org/wiki/Wavelet_transform
- https://en.wikipedia.org/wiki/Stochastic_process
- https://en.wikipedia.org/wiki/Fourier_transform
- https://en.wikipedia.org/wiki/Stationary_process
- https://en.wikipedia.org/wiki/Sine_and_cosine_transforms
- https://en.wikipedia.org/wiki/Fast_Fourier_transform
- https://www.mathworks.com/help/signal/ug/scalogram-computation-in-signal-analyzer.html
- https://mulloverthing.com/is-anti-aliasing-a-low-pass-filter/
- https://thewolfsound.com/what-is-aliasing-what-causes-it-how-to-avoid-it/
- https://en.wikipedia.org/wiki/Ringing_artifacts

Supplementary material (can be ignored)

Nyquist–Shannon sampling theorem

- The sampling theorem essentially says that a signal has to be sampled at least with twice the maximum frequency of the original signal, in order to avoid loosing information by confusing some frequencies with others.

DFT MATRIX

DFT matrix

From Wikipedia, the free encyclopedia

In applied mathematics, a **DFT matrix** is an expression of a discrete Fourier transform (DFT) as a transformation matrix, which can be applied to a signal through matrix multiplication.

Contents [hide]

Definition [edit]

An *N*-point DFT is expressed as the multiplication $X = Wx$, where x is the original input signal, W is the *N*-by-*N* square DFT matrix, and X is the DFT of the signal.

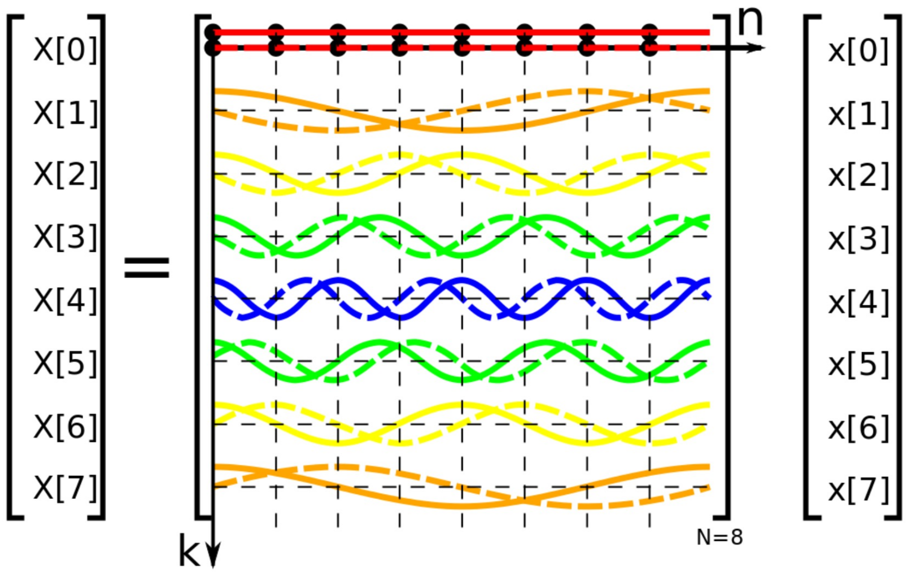
The transformation matrix W can be defined as $W = \left(\frac{\omega^{jk}}{\sqrt{N}}\right)_{j,k=0,\ldots,N-1}$, or equivalently:

$$W = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \omega^3 & \cdots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{2(N-1)} \\ 1 & \omega^3 & \omega^6 & \omega^9 & \cdots & \omega^{3(N-1)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \omega^{3(N-1)} & \cdots & \omega^{(N-1)(N-1)} \end{bmatrix},$$

where $\omega = e^{-2\pi i/N}$ is a primitive *N*th root of unity in which $i^2 = -1$. We can avoid writing large exponents for ω using the fact that for any exponent x we have the identity $\omega^x = \omega^{x \bmod N}$. This is the Vandermonde matrix for the roots of unity, up to the normalization factor. Note that the normalization factor in front of the sum ($1/\sqrt{N}$) and the sign of the exponent in ω are merely conventions, and differ in some treatments. All of the following discussion applies regardless of the convention, with at most minor adjustments. The only important thing is that the forward and inverse transforms have opposite-sign exponents, and that the product of their normalization factors be 1/*N*. However, the $1/\sqrt{N}$ choice here makes the resulting DFT matrix unitary, which is convenient in many circumstances.

DFT MATRIX

The following image depicts the DFT as a matrix multiplication, with elements of the matrix depicted by samples of complex exponentials:



The real part (cosine wave) is denoted by a solid line, and the imaginary part (sine wave) by a dashed line.

The top row is all ones (scaled by $1/\sqrt{8}$ for unitarity), so it "measures" the DC component in the input signal. The next row is eight samples of negative one cycle of a complex exponential, i.e., a signal with a fractional frequency of −1/8, so it "measures" how much "strength" there is at fractional frequency +1/8 in the signal. Recall that a matched filter compares the signal with a time reversed version of whatever we're looking for, so when we're looking for fracfreq. 1/8 we compare with fracfreq. −1/8 so that is why this row is a negative frequency. The next row is negative two cycles of a complex exponential, sampled in eight places, so it has a fractional frequency of −1/4, and thus "measures" the extent to which the signal has a fractional frequency of +1/4.

A word about aliasing and windowing

4

Performance Comparison of Data Windows

In this chapter, we compute several parameters of a window that are useful in choosing a suitable window for particular applications, such as power spectral estimation via discrete Fourier transform (DFT) and the design of FIR digital filters. A comprehensive comparison of the windows that were introduced in Chapter 3 is made based on the computed parameters. All the parameters listed are computed using the properties of windows in the continuous-time-domain.

The leakage that occurs in spectral estimation (via DFT) due to the prominent side lobes of the spectral window obviously degrades the accuracy of the results. Windows are weighting functions applied to the finite observation data to reduce the spectral leakage. There are four basic factors that need to be considered while choosing a window: (a) resolution or bandwidth, (b) stability, (c) leakage, and (d) smoothness. We shall now examine each of them in detail.

(a) *Resolution* refers to the ability of a spectrum estimate to represent fine structures in the frequency properties of the data, such as narrow peaks in the spectrum. Owing to the averaging involved in computing a spectrum estimate, a narrow peak in the periodogram is spread out into a broader peak. The width is roughly an image of the spectral window used in the estimate. Note that the width of the suitably defined spectral window is the bandwidth of the estimate. If the spectrum of a time series consists of two narrow peaks that are closer together than the bandwidth of the estimate used, we find that the two narrow peaks overlap, resulting in a single peak (which is broader). Thus, the estimate fails to resolve two narrow peaks that occur in close proximity to each other in the true spectrum.

(b) *Stability* of a spectrum estimate refers to the extent to which the estimates computed from different segments of a series concur, or the extent to which irrelevant fine structures in the periodogram are eliminated. Actually, resolution and stability are conflicting requirements since a high stability requires averaging over many periodograms, whereas this results in a reduced resolution.

(c) As discussed in one of the earlier chapters, *leakage* occurs because of the side lobes in the spectral window. This could be reduced by applying appropriate window functions. The smoothness of a spectrum is a less tangible property that would add a further conflict in requirements.

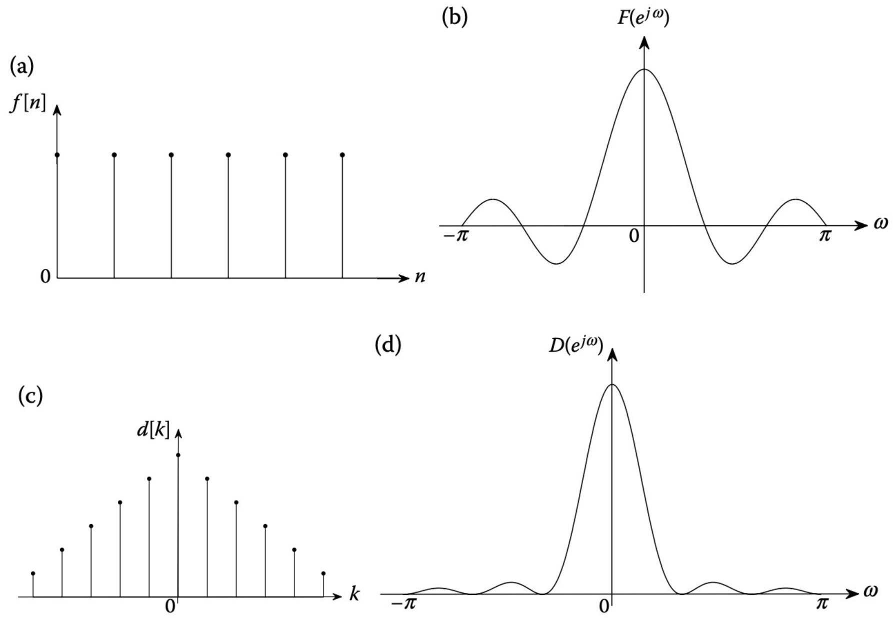


FIGURE 5.2
Interplay between various window functions. (a) Data window. (b) Frequency window. (c) Lag window. (d) Spectral window.

Spectrogram vs Scalogram

- A **spectrogram** is a visual representation of the spectrum of frequencies of a signal as it varies with time.

- A spectrogram can be generated by a bank of band-pass filters, by Fourier transform or by a wavelet transform (in which case it is also known as a **scaleogram** or **scalogram**).[2]

Spectrogram vs Scalogram

Scalogram Computation in Signal Analyzer

R2022b

The scalogram is the absolute value of the continuous wavelet transform (CWT) of a signal, plotted as a function of time and frequency. The scalogram can be more useful than the spectrogram for analyzing real-world signals with features occurring at different scales — for example, signals with slowly varying events punctuated by abrupt transients. Use the scalogram when you want better time localization for short-duration, high-frequency events, and better frequency localization for low-frequency, longer-duration events.

> **Note**
>
> You need a Wavelet Toolbox™ license to use the scalogram view.

The spectrogram is obtained by windowing the input signal with a *window* of constant length (duration) that is shifted in time and frequency. (See Spectrogram Computation in Signal Analyzer for more information.) The window used in the spectrogram is even, real-valued, and does not oscillate. Because the spectrogram uses a constant window, the time-frequency resolution of the spectrogram is fixed.

By contrast, the CWT is obtained by windowing the signal with a *wavelet* that is scaled and shifted in time. The wavelet oscillates and can be complex-valued. The scaling and shifting operations are applied to a prototype wavelet. The scaling used in the CWT both shrinks and stretches the prototype wavelet. Shrinking the prototype wavelet yields short duration, high-frequency wavelets that are good at detecting transient events. Stretching the prototype wavelet yields long duration, low-frequency wavelets which are good at isolating long-duration, low frequency events.

To compute the scalogram, **Signal Analyzer** performs these steps:

1. If the signal has more than 1 million samples, divide the signal into overlapping segments.
2. Compute the CWT of each segment to get its scalogram.
3. Display the scalogram segment by segment.

As implemented, the CWT uses L^1 normalization. Therefore, the amplitudes of the oscillatory components in a signal agree with the amplitudes of the corresponding wavelet coefficients.

Spectrogram vs Scalogram

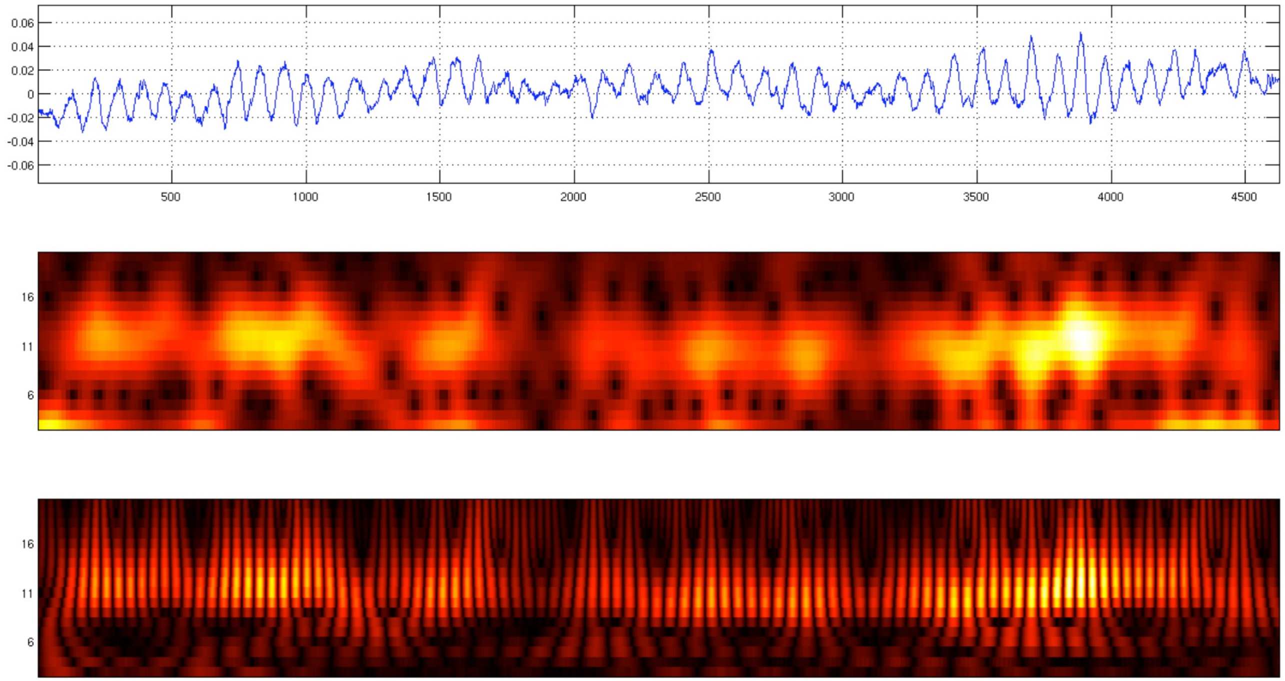


Figure. 4 Comparison of the spectrogram, wavelet scalogram and Hilbert spectrum of a same time series. Top Row: Original signal. Second row: Short time Fourier Transform spectrogram. Third row: Wavelet Scalogram, using a Morlet mother wavelet.