Basic Signal Processing (I)

Computational Social Intelligence - Lecture 16

Prof. Alessandro Vinciarelli
School of Computing Science &
Institute of Neuroscience and Psychology

http://www.dcs.gla.ac.uk/vincia
Alessandro.Vinciarelli@glasgow.ac.uk

University of Glasgow

EPSRC
Engineering and Physical Sciences Research Council

FNSNF

Texts (see Moodle)

This lecture is based on the following text (available on Moodle):

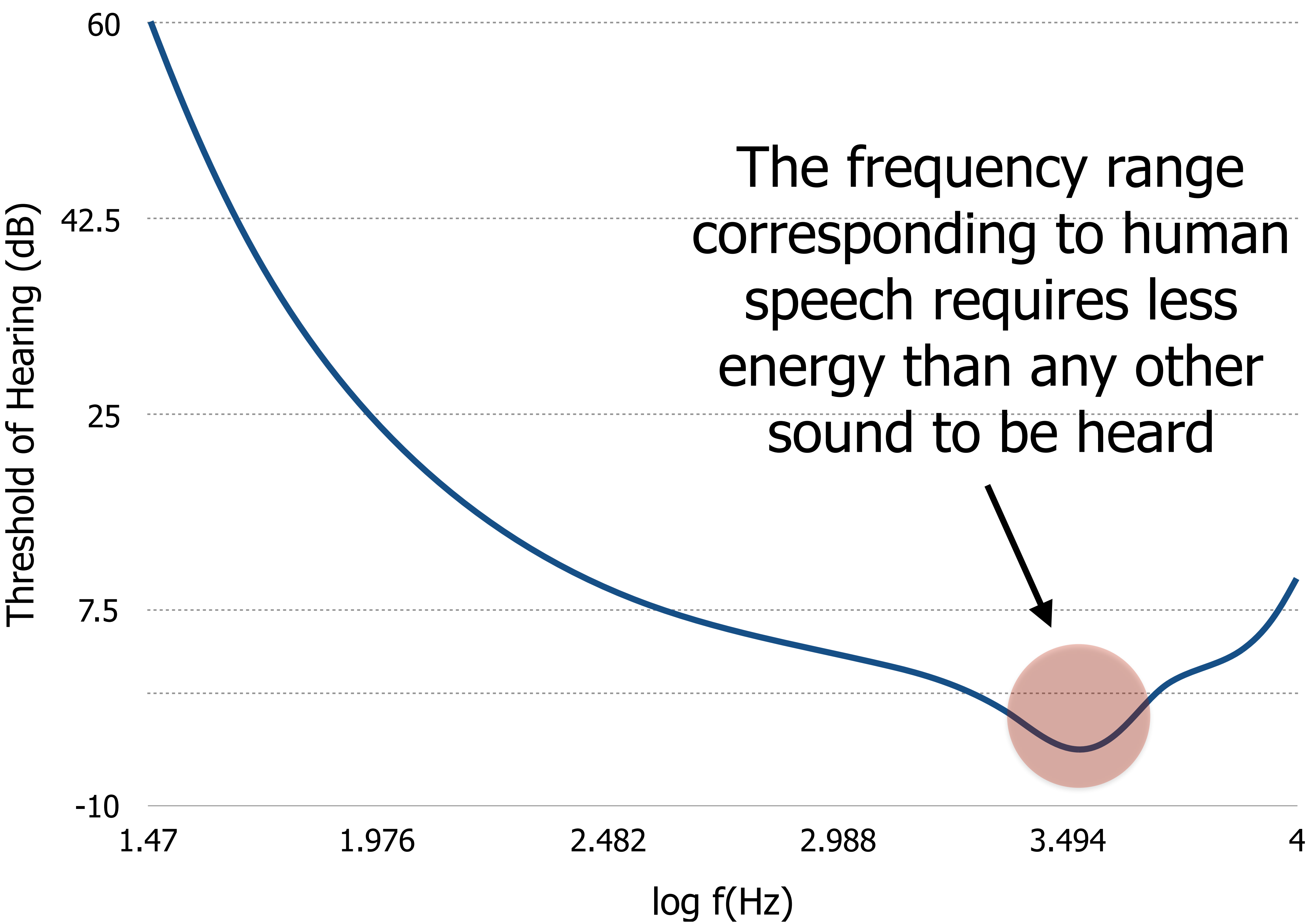
- F.Camastra and A.Vinciarelli, "Machine Learning for Audio, Image and Video Processing", Springer Verlag, Chapter 2, pp. 38-46, 2008.

Outline

- Introduction
- Time Domain Processing
- Short-Term Analysis
- Conclusions

Outline

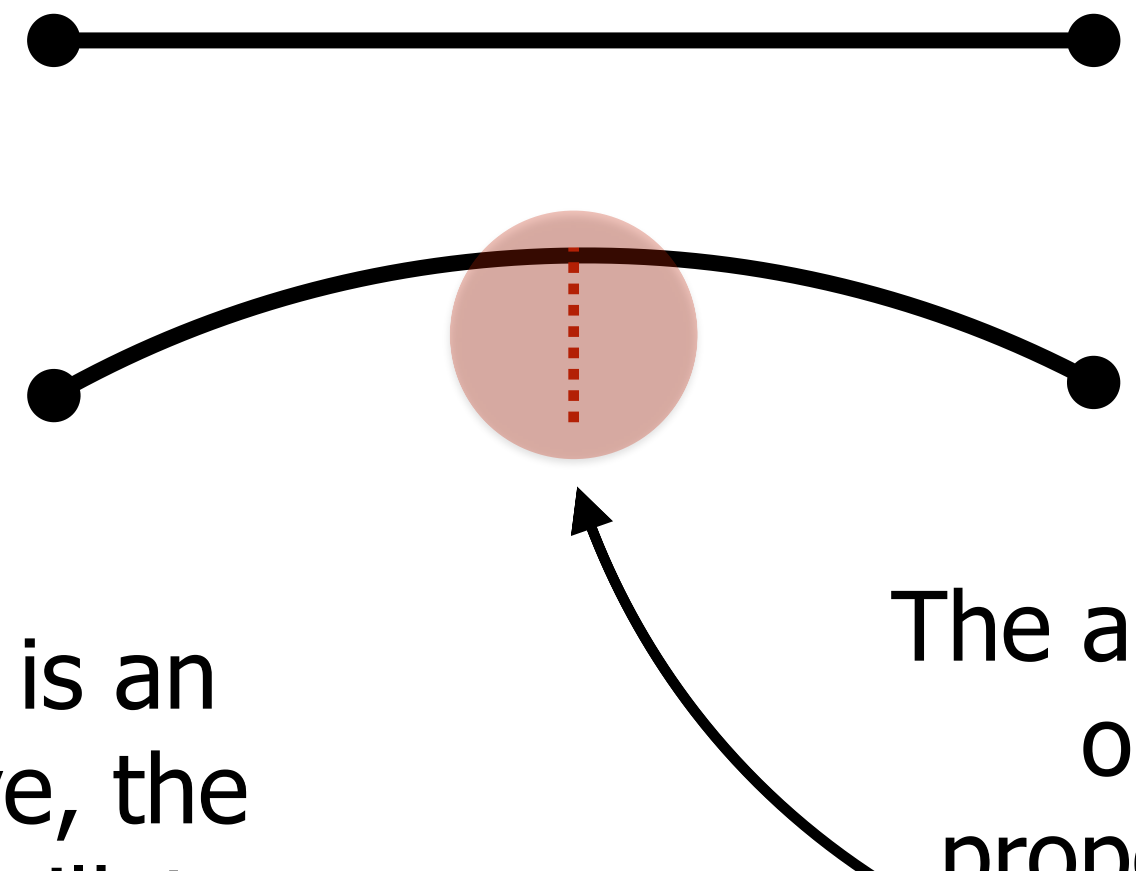
- Introduction
- Time Domain Processing
- Short-Term Analysis
- Conclusions



60
42.5
25
7.5
-10
Threshold of Hearing (dB)
The frequency range corresponding to human speech requires less energy than any other sound to be heard
1.47
1.976
2.482
2.988
3.494
4
log f(Hz)

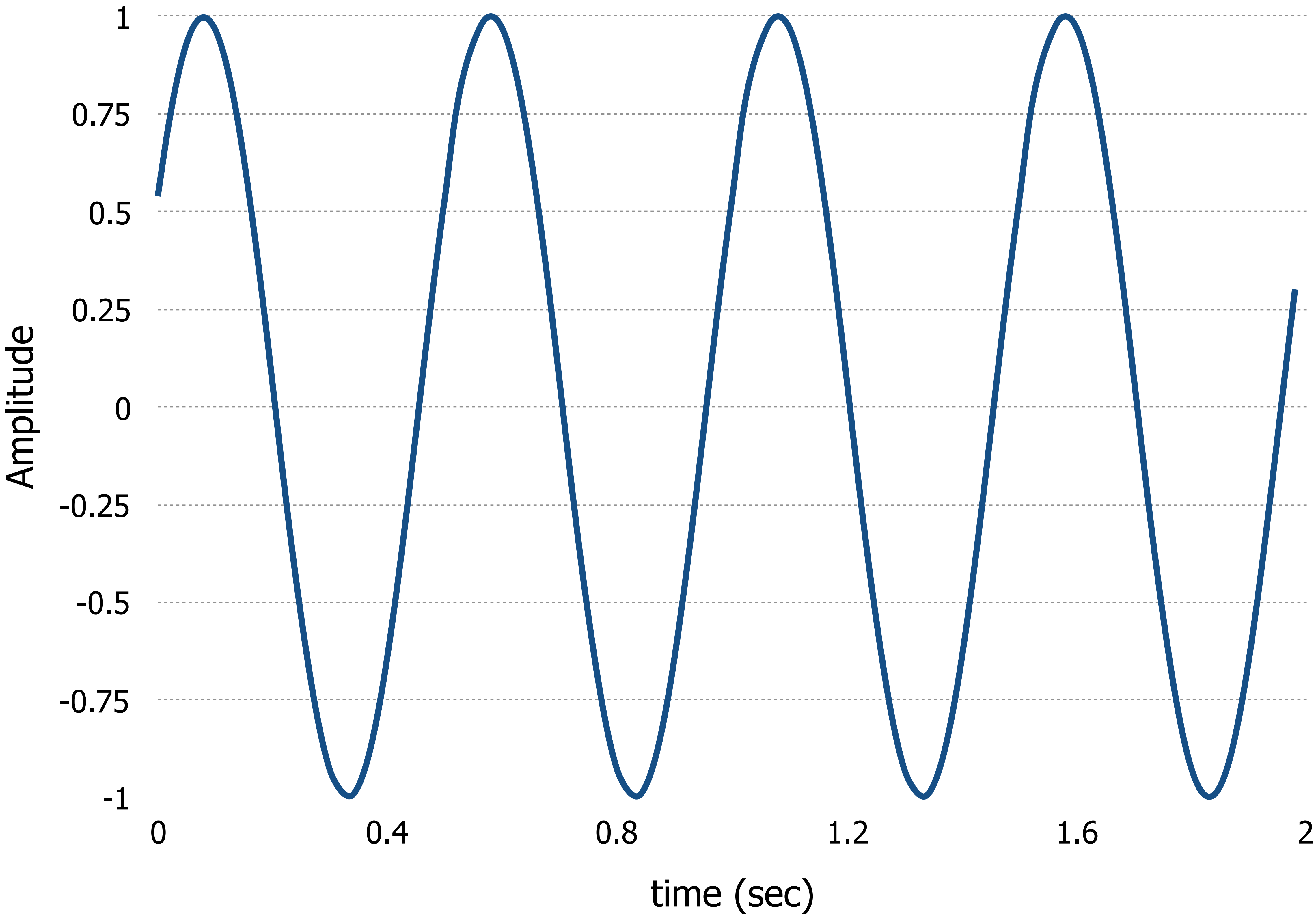
A microphone measures at regular time steps the the oscillations of an elastic membrane

When there is no acoustic wave, the membrane is in equilibrium

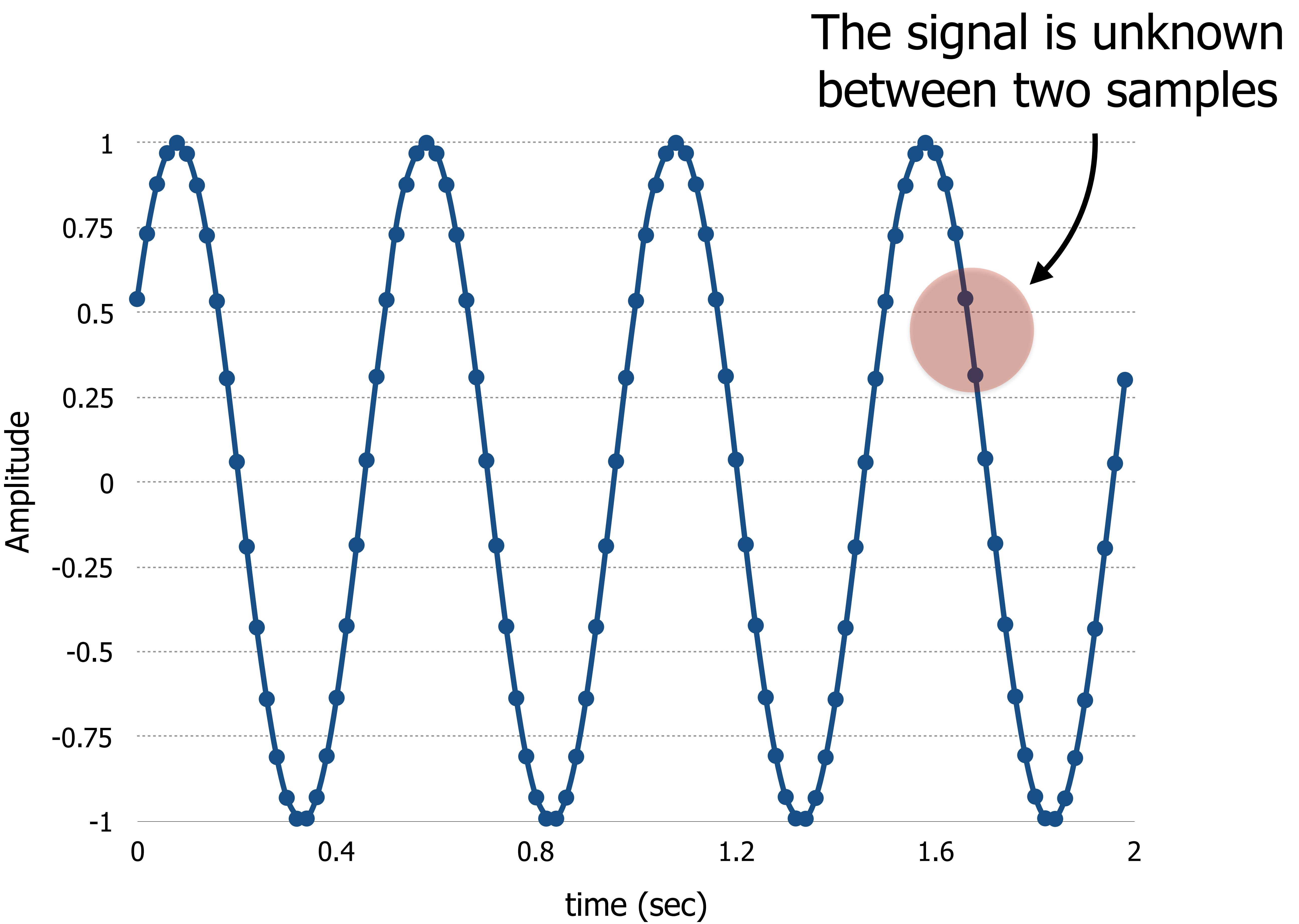


When there is an acoustic wave, the membrane oscillates

The amplitude of the oscillation is proportional to the amplitude of the acoustic wave



1
0.75
0.5
0.25
0
-0.25
-0.5
-0.75
-1
Amplitude
0
0.4
0.8
1.2
1.6
2
time (sec)


The signal is unknown
between two samples
1
0.75
0.5
0.25
0
-0.25
-0.5
-0.75
-1
Amplitude
0
0.4
0.8
1.2
1.6
2
time (sec)

The sample “k”

The signal at time “kT”

$$s[k] = s(kT)$$

$$k = -\infty, \ldots, \infty$$

The index of the sample

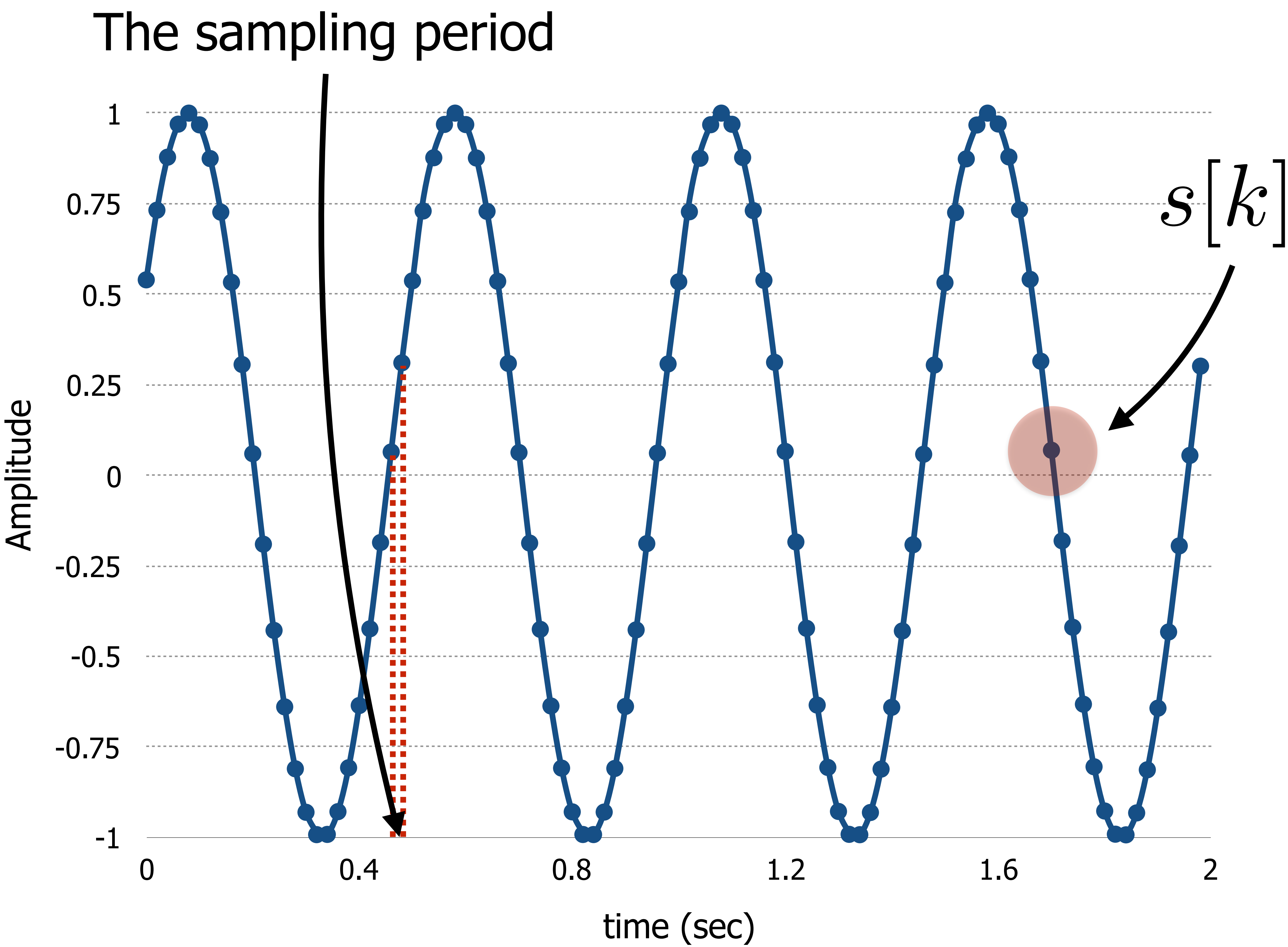
The sampling period
(time interval between
consecutive samples)

There are no infinite signals in reality

All samples with k<0 are conventionally null

$$s[k] = 0 : k < 0$$
$$s[k] = 0 : k > K$$

All samples with k<K (where KT is the time the last measurement is done) are conventionally null


The sampling period
$s[k]$
1
0.75
0.5
0.25
0
-0.25
-0.5
-0.75
-1
Amplitude
0
0.4
0.8
1.2
1.6
2
time (sec)

The sampling frequency is the inverse of the sampling period

It must be at least twice as much as the highest frequency expected to be observed in the signal

$$F = \frac{1}{T}$$

It is the number of samples per unit of time

Recap

- A signal is a measurable physical quantity that changes over time and it is continuous;
- A digital signal is a sequence of physical measurements collected at regular time steps (the sampling period) and it is discrete;
- Nothing can be said about what happens in the signal between two consecutive samples.

Outline

- Introduction
- Time Domain Processing
- Short-Term Analysis
- Conclusions

The convolution between the signal “s” and the window “w”

The sum extends over all samples of the signal

$$y[k] = \sum_{n=-\infty}^{\infty} s[n]w[k-n]$$

The window

The signal

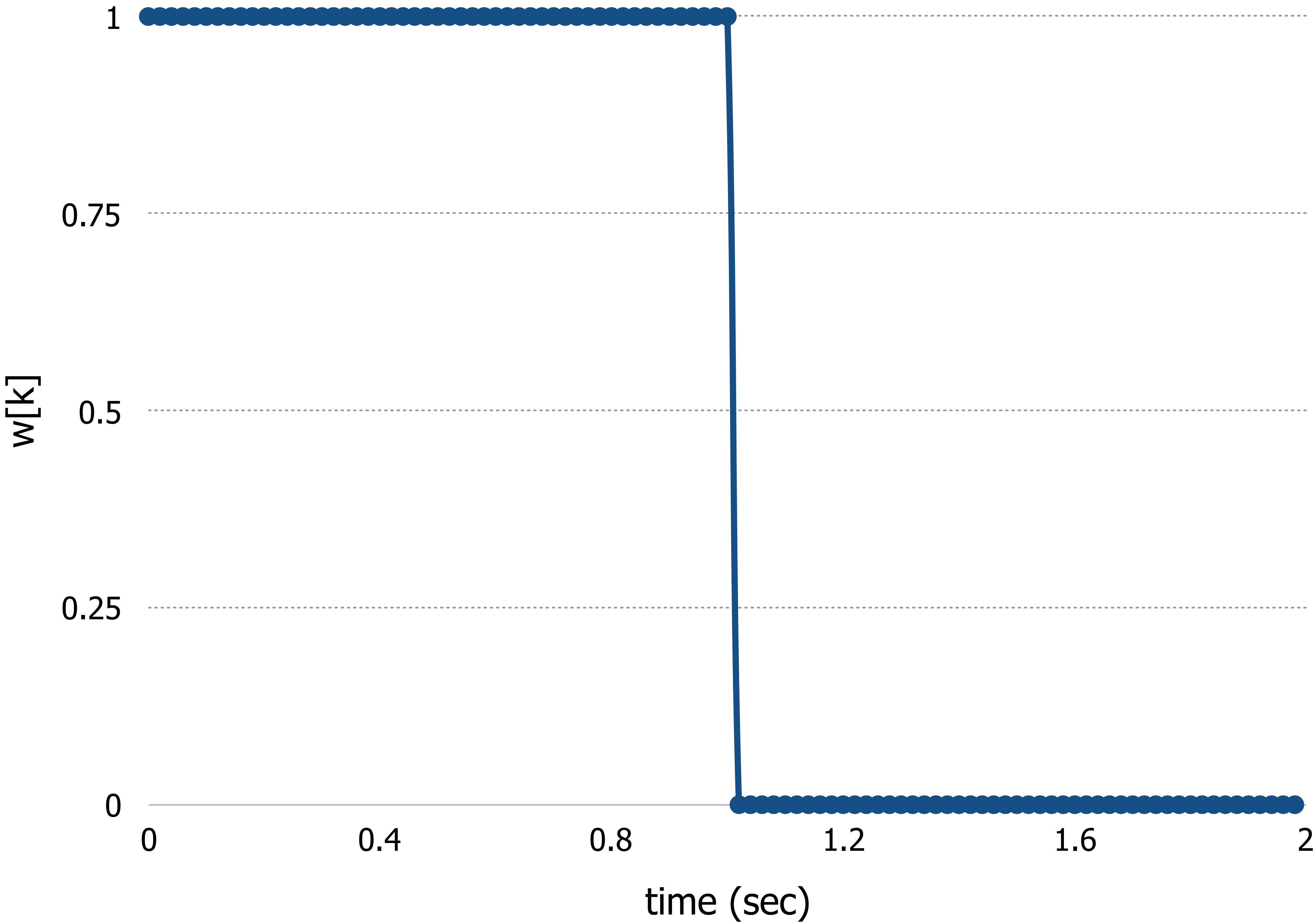
The window length

The rectangular window

$$w[n] = \begin{cases} 1 & : \quad 0 \le n \le N-1 \\ 0 & : \quad n < 0 \\ 0 & : \quad n > N-1 \end{cases}$$

When the index is negative, the window signal is null

When the index is larger than the length, the window signal is null



1
0.75
0.5
0.25
0
w[k]
0
0.4
0.8
1.2
1.6
2
time (sec)

The convolution between the signal “s” and the window “w”

The sum extends over all samples of the signal

$$y[k] = \sum_{n=-\infty}^{\infty} s[n]w[k-n]$$

The window

The signal

Any product in which "n" is less than or equal to "k" is null

$$k - n \geq 0 \Rightarrow n \leq k$$

$$k - n \leq N - 1 \Rightarrow n \geq k - N + 1$$

Only products in which "n" is greater than or equal to "k-N+1" is null

The convolution between the signal “s” and the window “w”

The sum extends over a finite number of samples

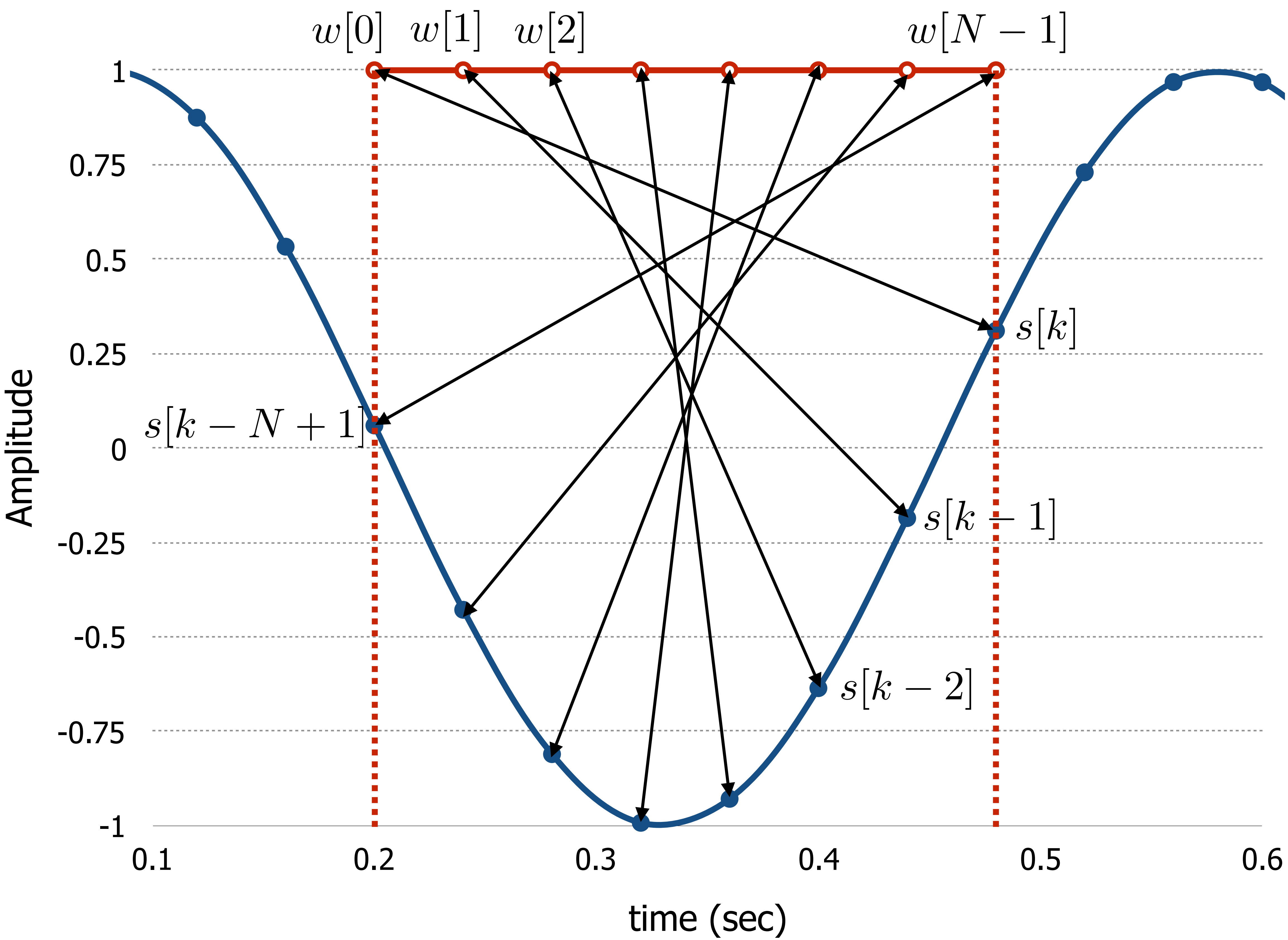
$$y[k] = \sum_{n=k-N+1}^{k} s[n]w[k-n]$$

$$y[k] = \sum_{n=k-N+1}^{k} s[n]w[k-n] =$$

$$= s[k-N+1]w[N-1]+$$

$$+s[k-N+2]w[N-2]+\ldots$$

$$+\ldots+s[k-1]w[1]+s[k]w[0]$$



$w[0]$
$w[1]$
$w[2]$
$w[N-1]$
$s[k]$
$s[k-N+1]$
$s[k-1]$
$s[k-2]$
1
0.75
0.5
0.25
0
-0.25
-0.5
-0.75
-1
0.1
0.2
0.3
0.4
0.5
0.6
Amplitude
time (sec)

The convolution between the signal “s” and the rectangular window “w”

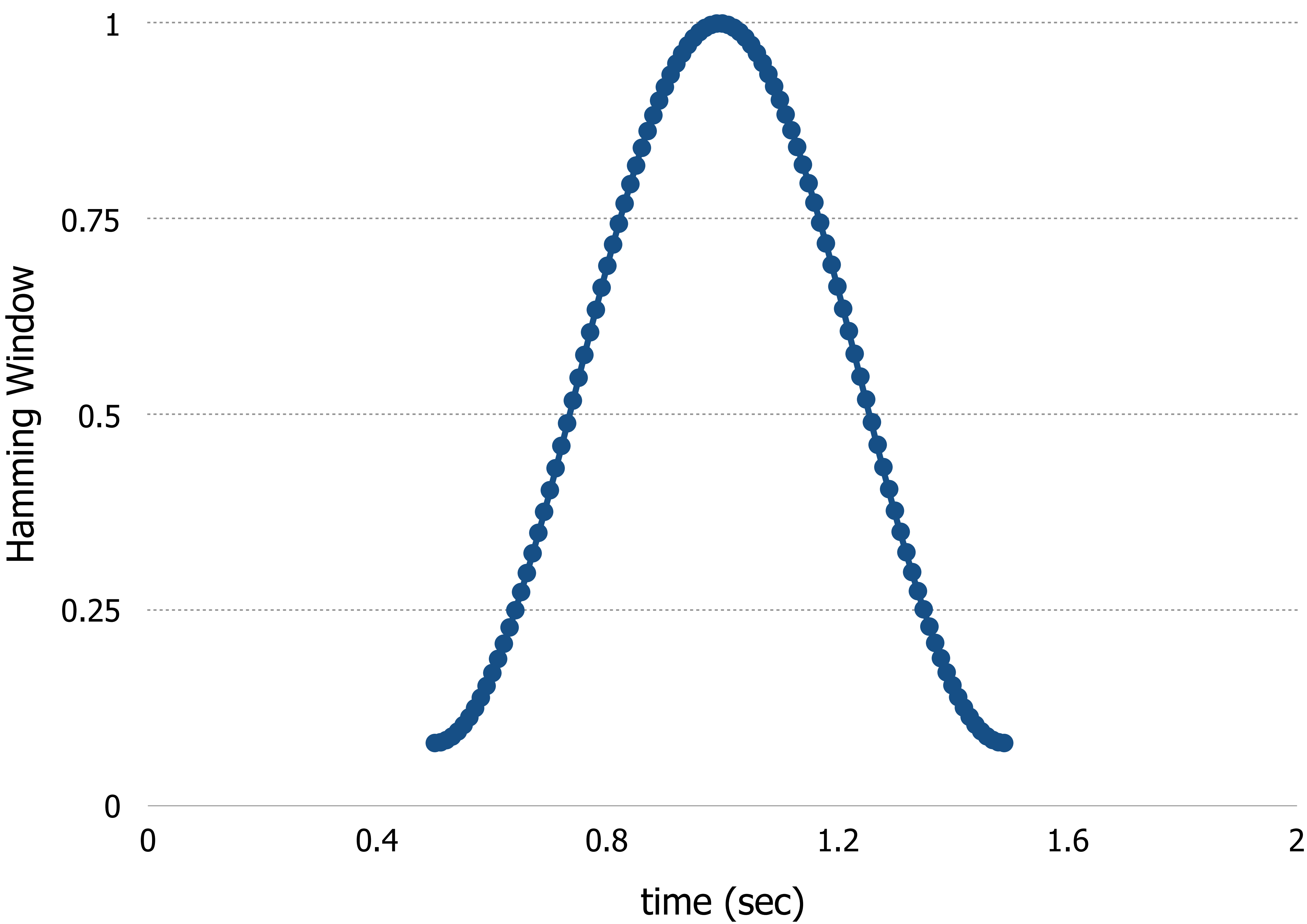
Sample “k” of the convolution is the sum of the samples between “k-N+1” and “k”

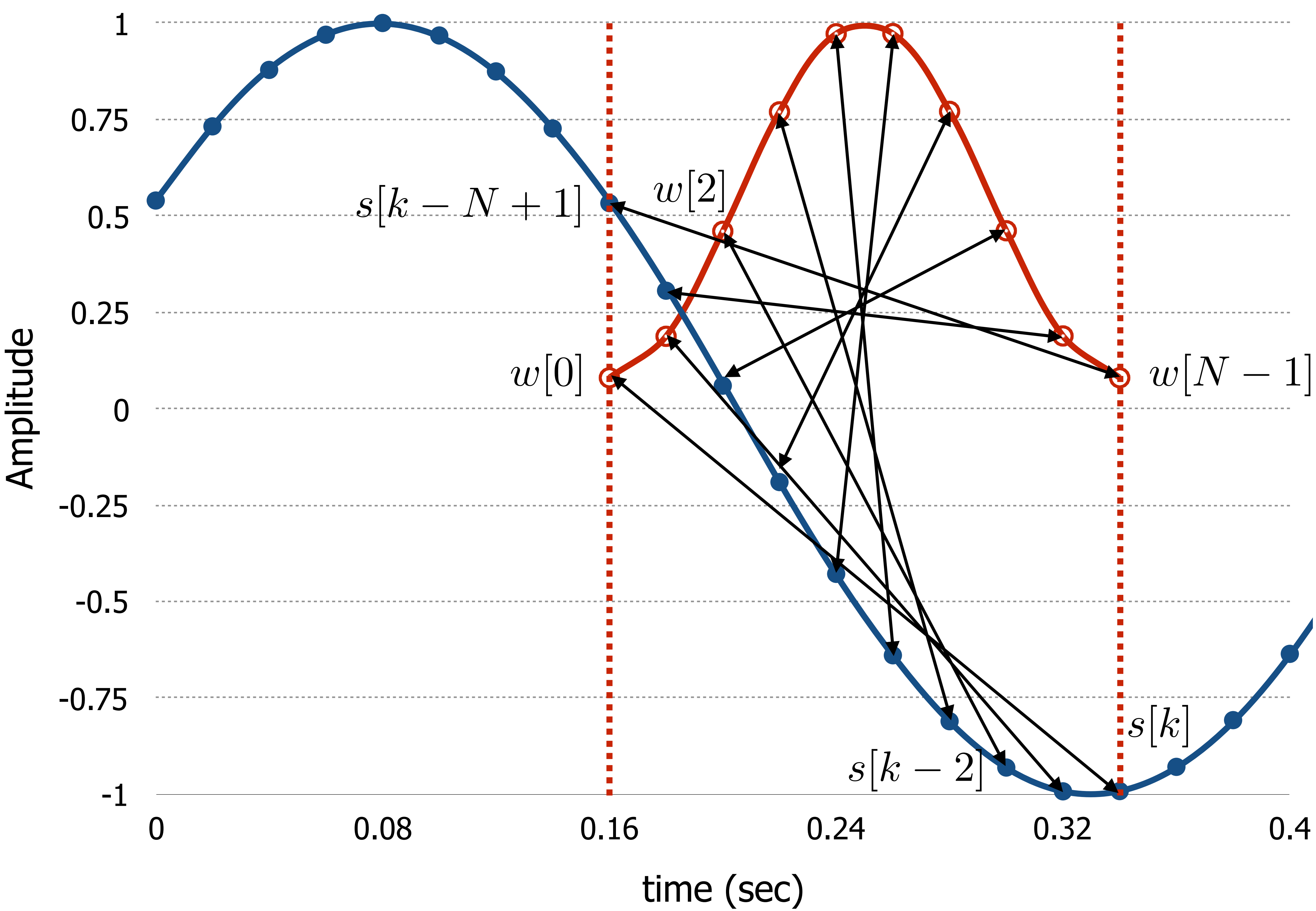
$$y[k] = \sum_{n=k-N+1}^{k} s[n]$$

Only samples for which "n" is between 0 and "N-1" are not null

The Hamming window

$$w[n] = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) & : \quad 0 \le n \le N-1 \\ 0 & : \quad n < 0 \\ 0 & : \quad n > N-1 \end{cases}$$


Hamming Window
time (sec)
1
0.75
0.5
0.25
0
0
0.4
0.8
1.2
1.6
2



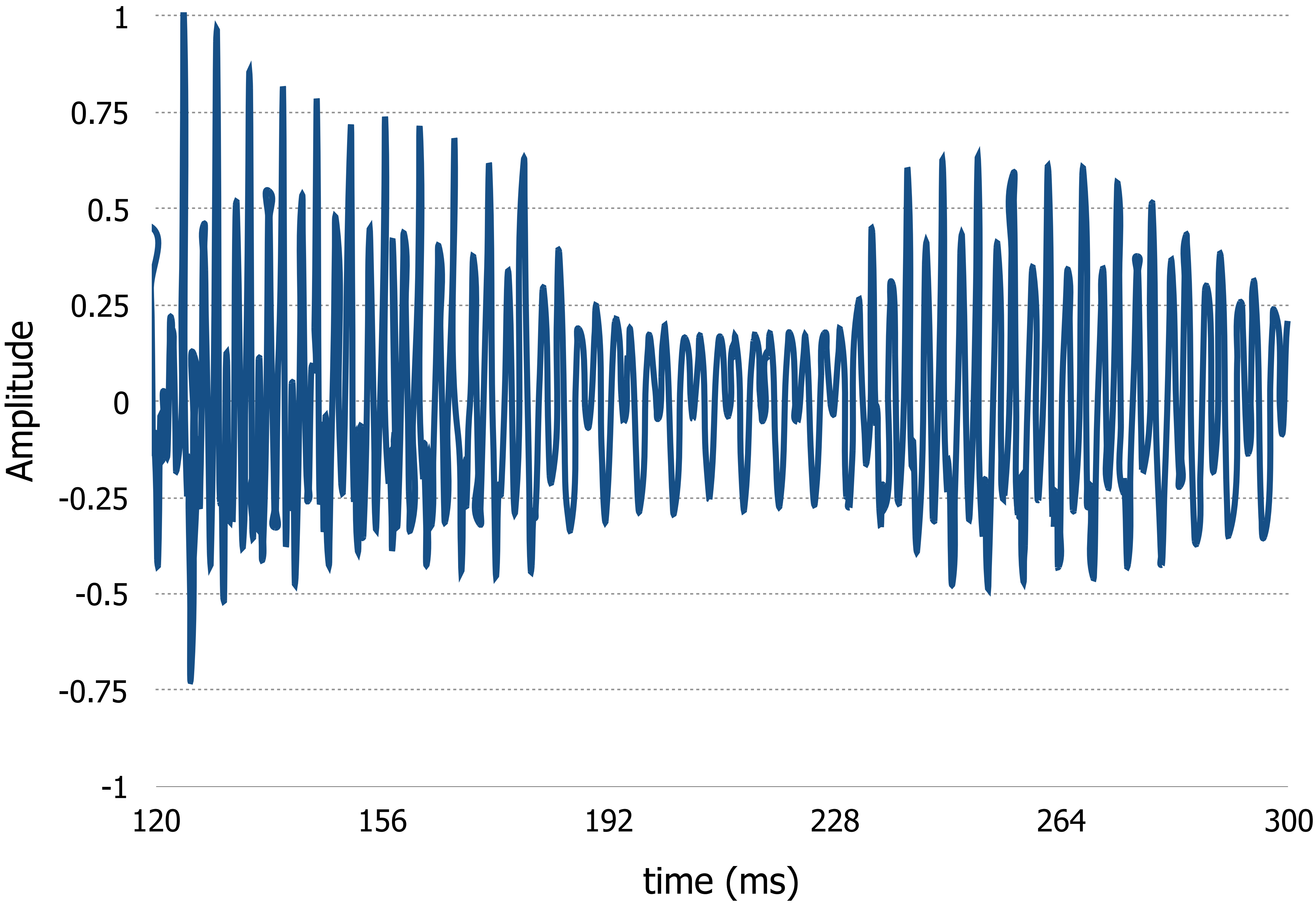
1
0.75
0.5
0.25
0
-0.25
-0.5
-0.75
-1
Amplitude
0
0.08
0.16
0.24
0.32
0.4
time (sec)
$s[k-N+1]$
$w[2]$
$w[0]$
$w[N-1]$
$s[k]$
$s[k-2]$

Short-Term Properties

- Any property should be measured over an interval that is short enough to ensure that the signal characteristics are stable;
- However, the interval should still be long enough to capture meaningful atomic units;
- In the case of speech, the typical interval length is 20-30 ms, the time one speaker keeps the articulators in a stable configuration.

Outline

- Introduction
- Time Domain Processing
- Short-Term Analysis
- Conclusions


Amplitude
1
0.75
0.5
0.25
0
-0.25
-0.5
-0.75
-1
120
156
192
228
264
300
time (ms)

Any short-term property that can be extracted from the signal

A function of sample "n"

$$x[k] = \sum_{n=-\infty}^{\infty} f(s[n])w[k-n]$$

A window signal

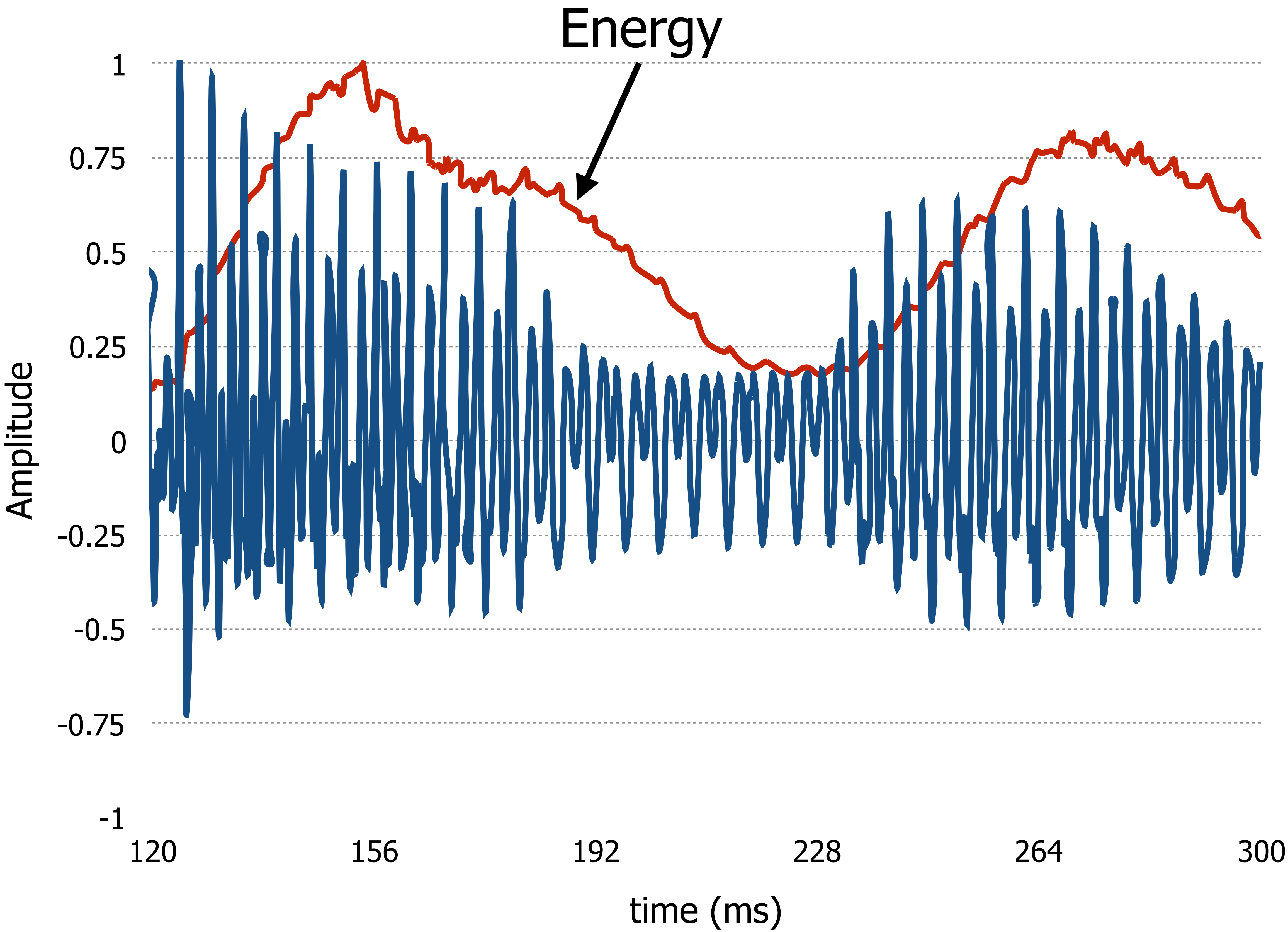
Convolution between f(s[n]) and a window

The function is the square of the sample divided by the length of the window

The Energy

$$E[k] = \sum_{n=-\infty}^{\infty} \frac{(s[n])^2}{N} w[k-n]$$

The energy is the average of the samples' squares in the window ending at sample "k"



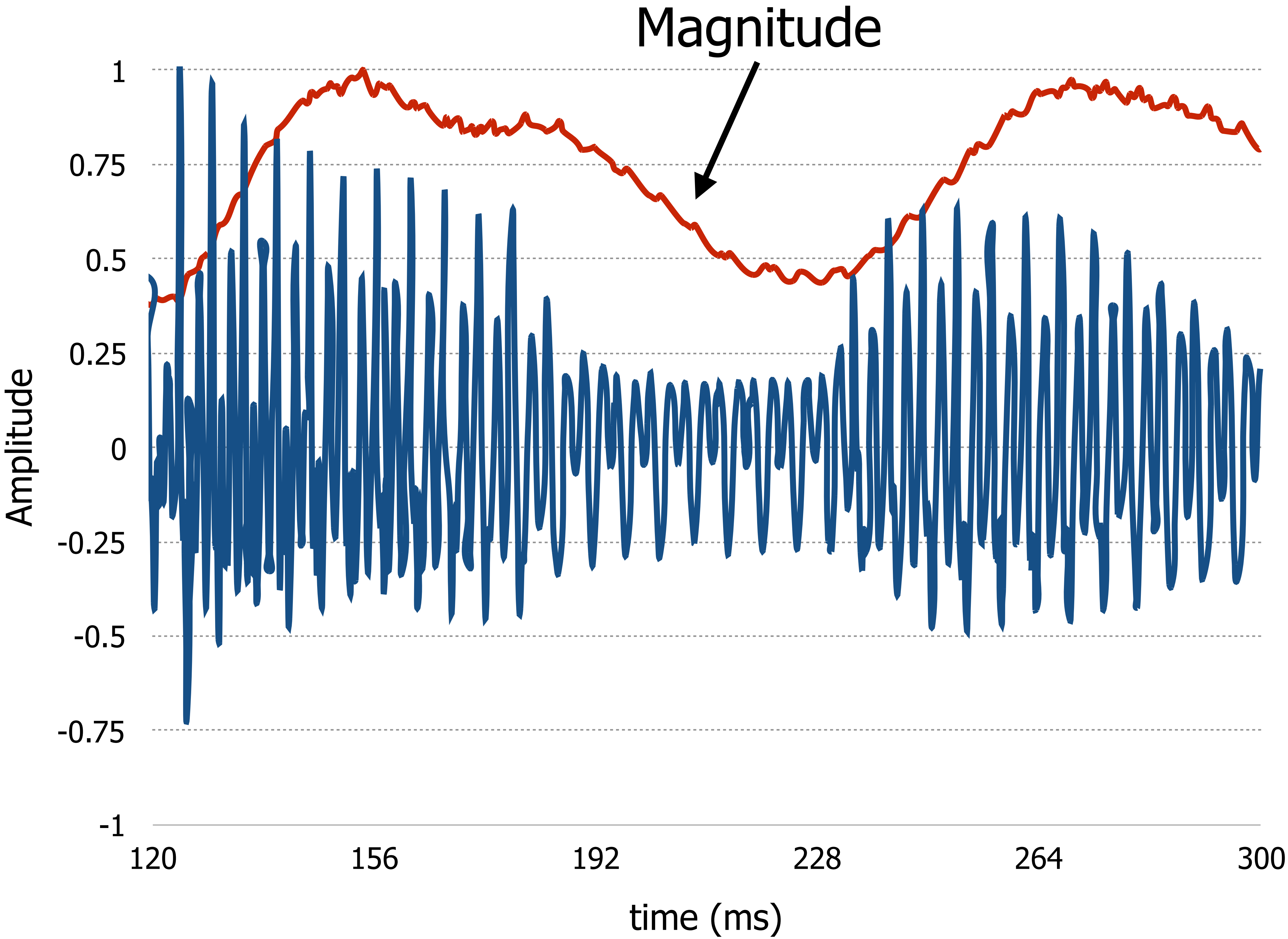
Energy
1
0.75
0.5
0.25
0
-0.25
-0.5
-0.75
-1
Amplitude
120
156
192
228
264
300
time (ms)

The function is the
absolute value of the
sample divided by the
length of the window

The Magnitude

$$M[k] = \sum_{n=-\infty}^{\infty} \frac{|s[n]|}{N} w[k-n]$$

The magnitude is the
average of the samples'
absolute value in the
window ending in
sample "k"



Magnitude
Amplitude
1
0.75
0.5
0.25
0
-0.25
-0.5
-0.75
-1
120
156
192
228
264
300
time (ms)

Outline

- Introduction
- Time Domain Processing
- Short-Term Analysis
- Conclusions

Conclusions

- Speech signals can be analysed in the time domain through convolution operations;
- In most cases, the processing takes place in the frequency domain (after performing Fourier transform);
- The main reason for analysing speech is that it is the main form of communication between people.