LING 446

Fundamentals for Speech Signal Processing and Analysis

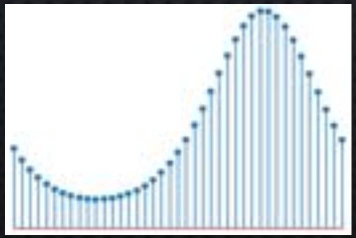
Yan Tang

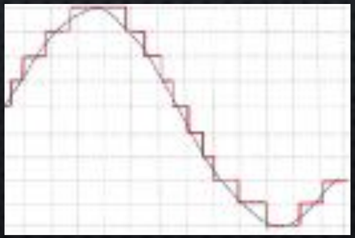
Department of Linguistics, UIUC

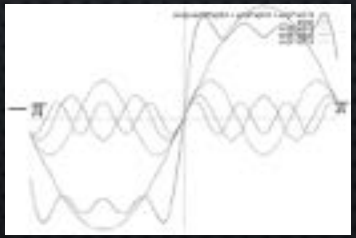
Week 2: Basic signal representations and properties

Last week…

- Applications of DSP
- Motivations of learning DSP
- Definition of signal
- Analog vs digital signal
    - Advantages of performing DSP
- General workflow of DSP systems
    - Three fundamental ideas of DSP

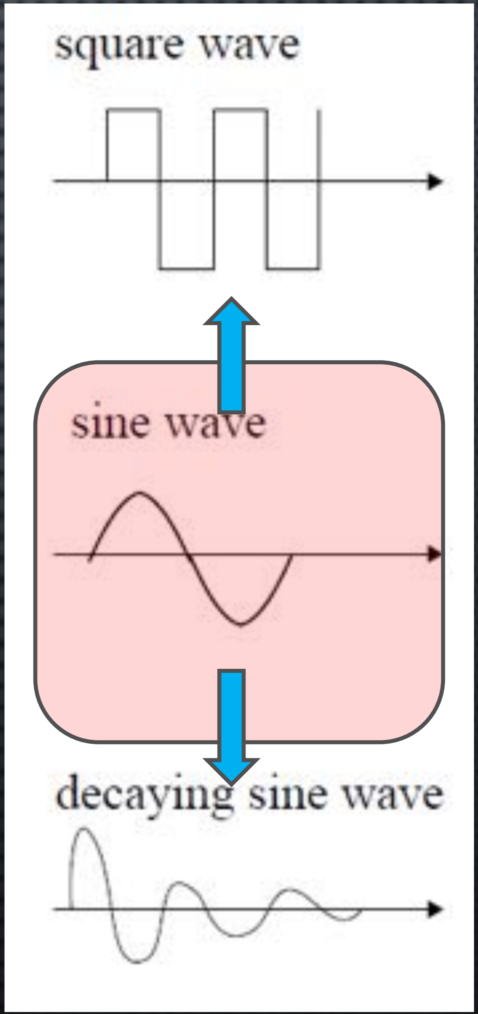




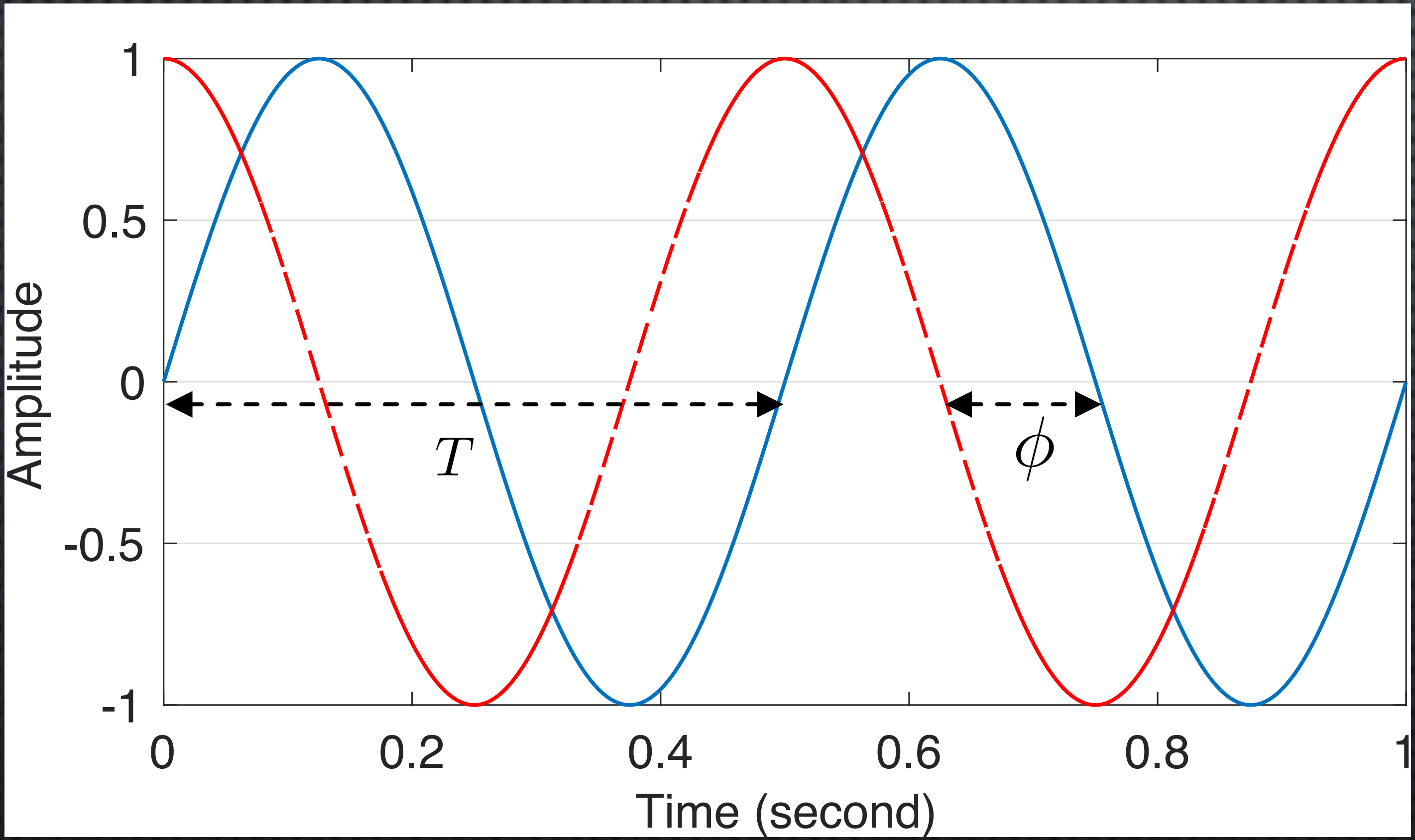


Waveform of signals

- Some examples
  - Sinusoid: useful for describing many types of continuous signal
  - Square wave: useful for describing binary signals in communication
  - Decaying sine wave: useful for pitch-periods of speech
- Key idea of DSP
  - Describe complex signal in terms of the sum of basis functions which are simpler to interpret



Properties of continuous signals



Peak amplitude: ***A***

frequency: ***f, ω***

phase: ***ϕ***

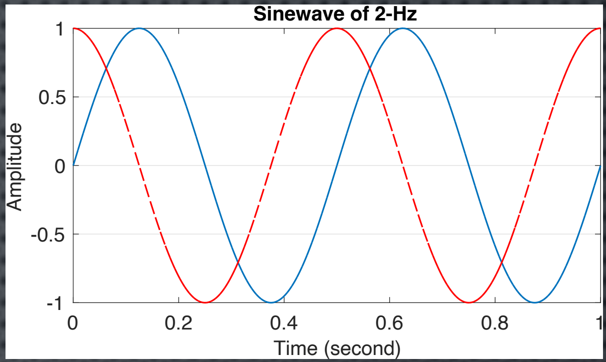
period: ***T***

Periodical signal

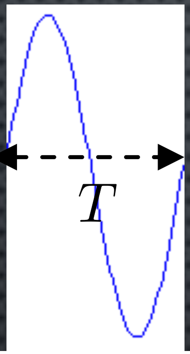
$$x(t) = A \cdot \sin(2\pi \cdot f \cdot t + \emptyset)$$

$$x(t) = A \cdot \sin(\omega \cdot t + \emptyset)$$

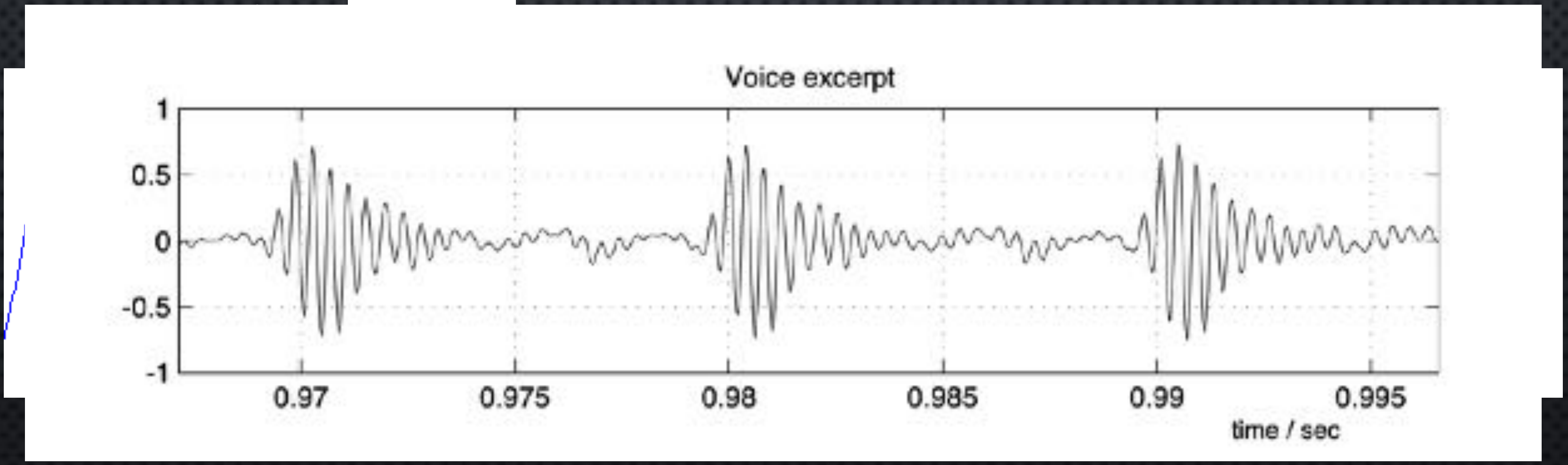
What is periodicity?



1 period



- periodic indicates repeativeness
- Period ***T***: the time a signal takes to repeat itself once



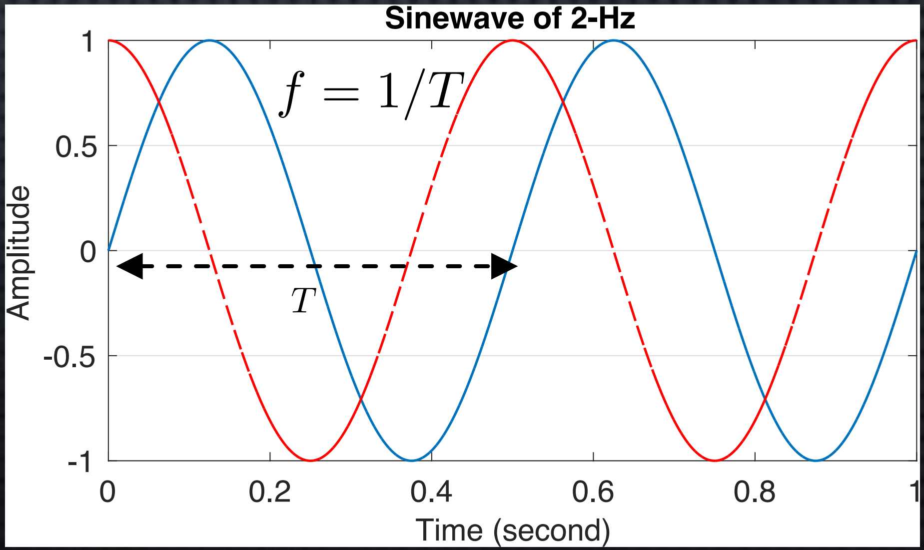
In fact, speech waveforms are not periodic (later we will see how to handle this)

What is frequency?

- Frequency indicates how frequently a signal repeats itself in a unit time

- An inverse function of ***T***

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



$x(t) = A \cdot \sin(2\pi \cdot f \cdot t + \emptyset)$

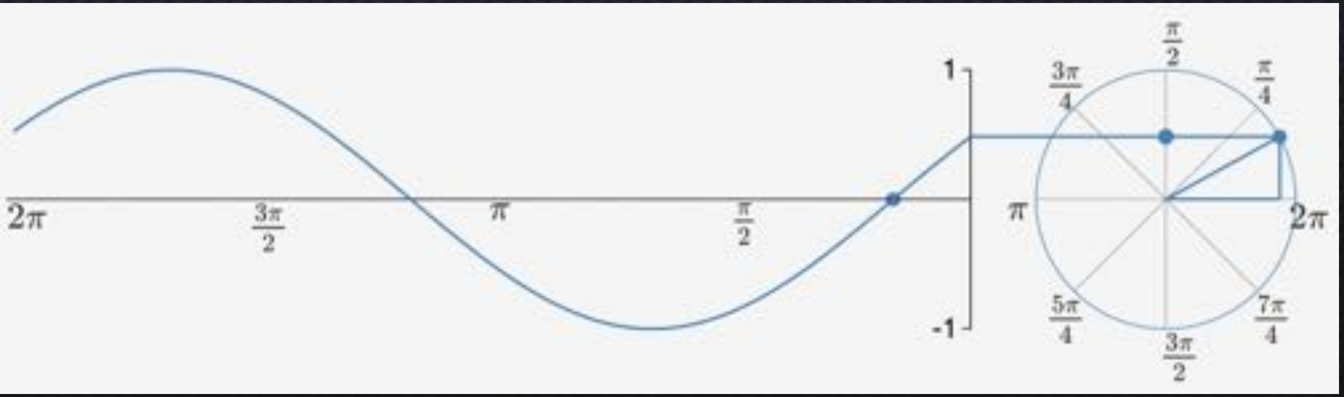
$x(t) = A \cdot \sin(\omega \cdot t + \emptyset)$

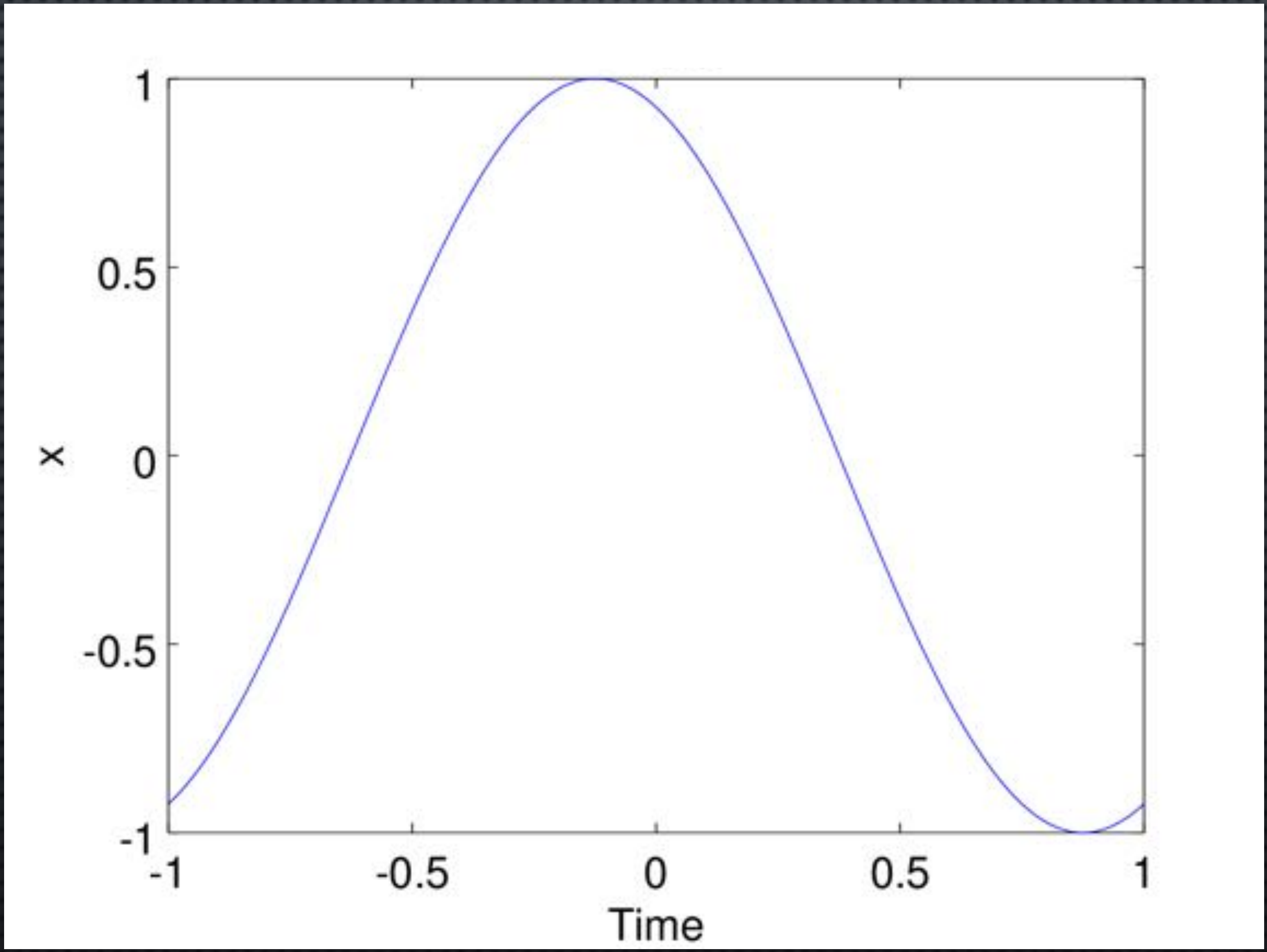
Relationship between *f* and **ω**

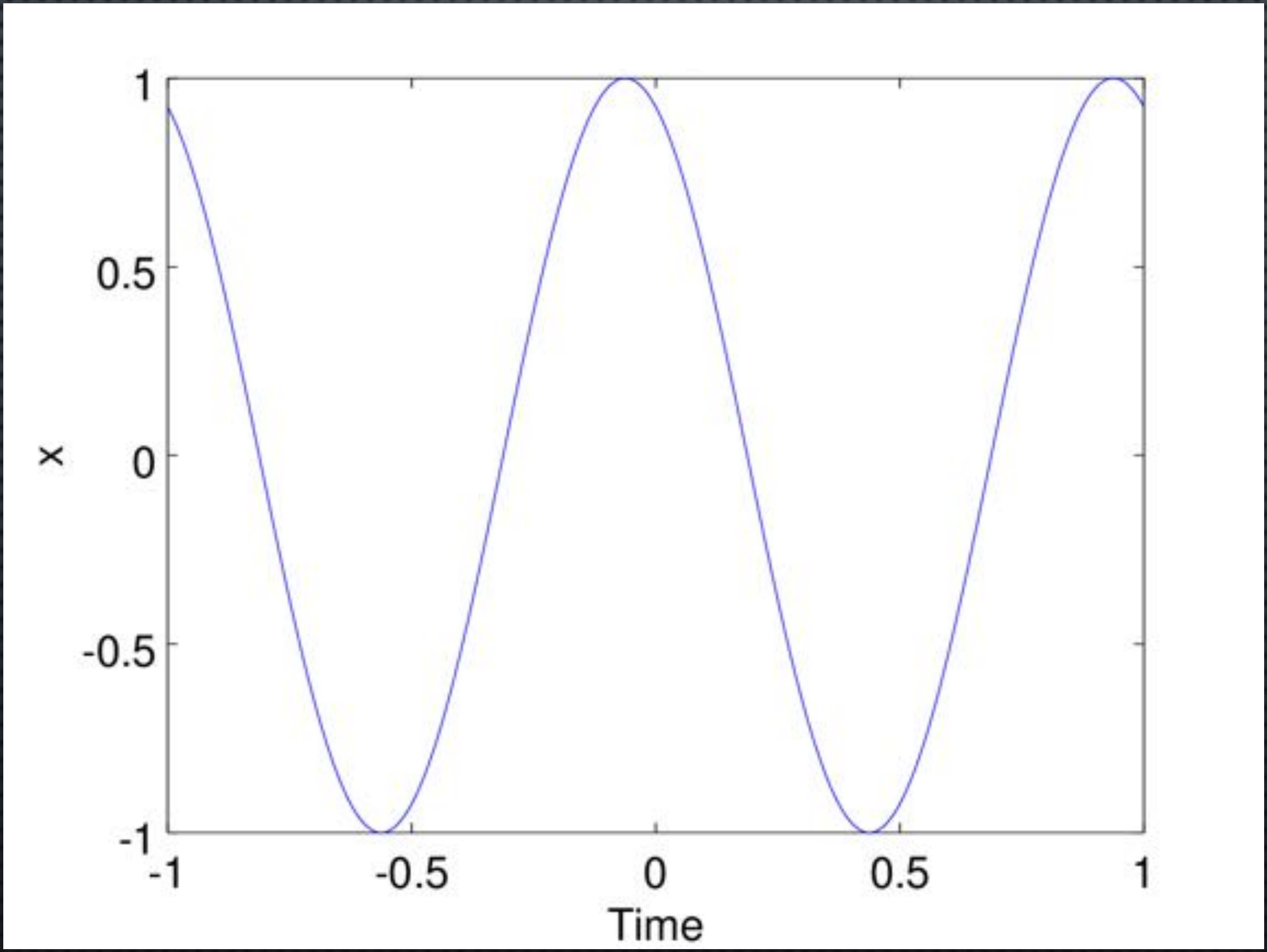
- ***f*** : regular/linear frequency (cycle/s, or Hz)
  - The number of times the signal repeats itself per second
- **ω:** angular/radial/circular frequency (radians/s)
  - Angular displacement per/second

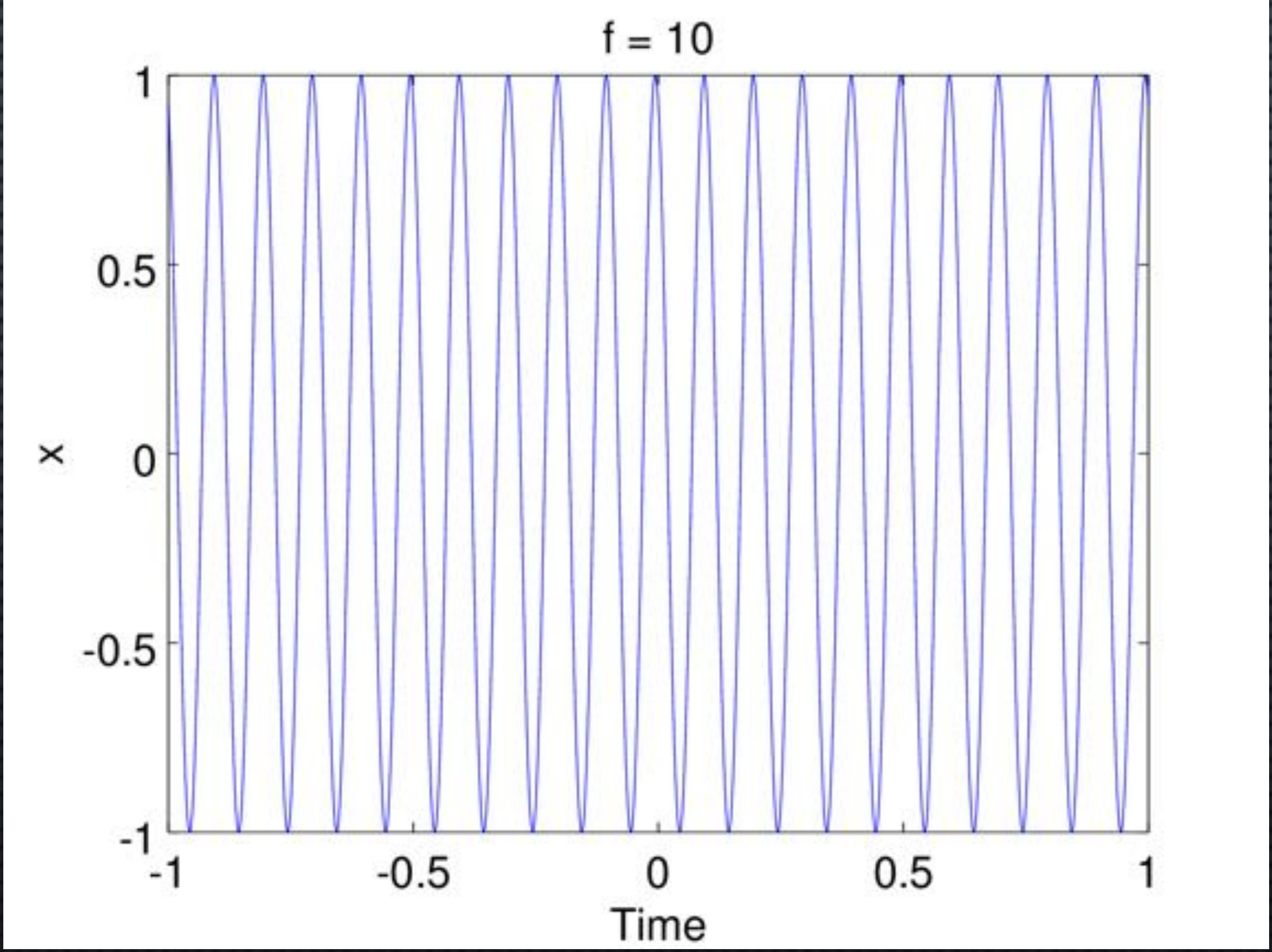
$$\omega = 2\pi \cdot f$$

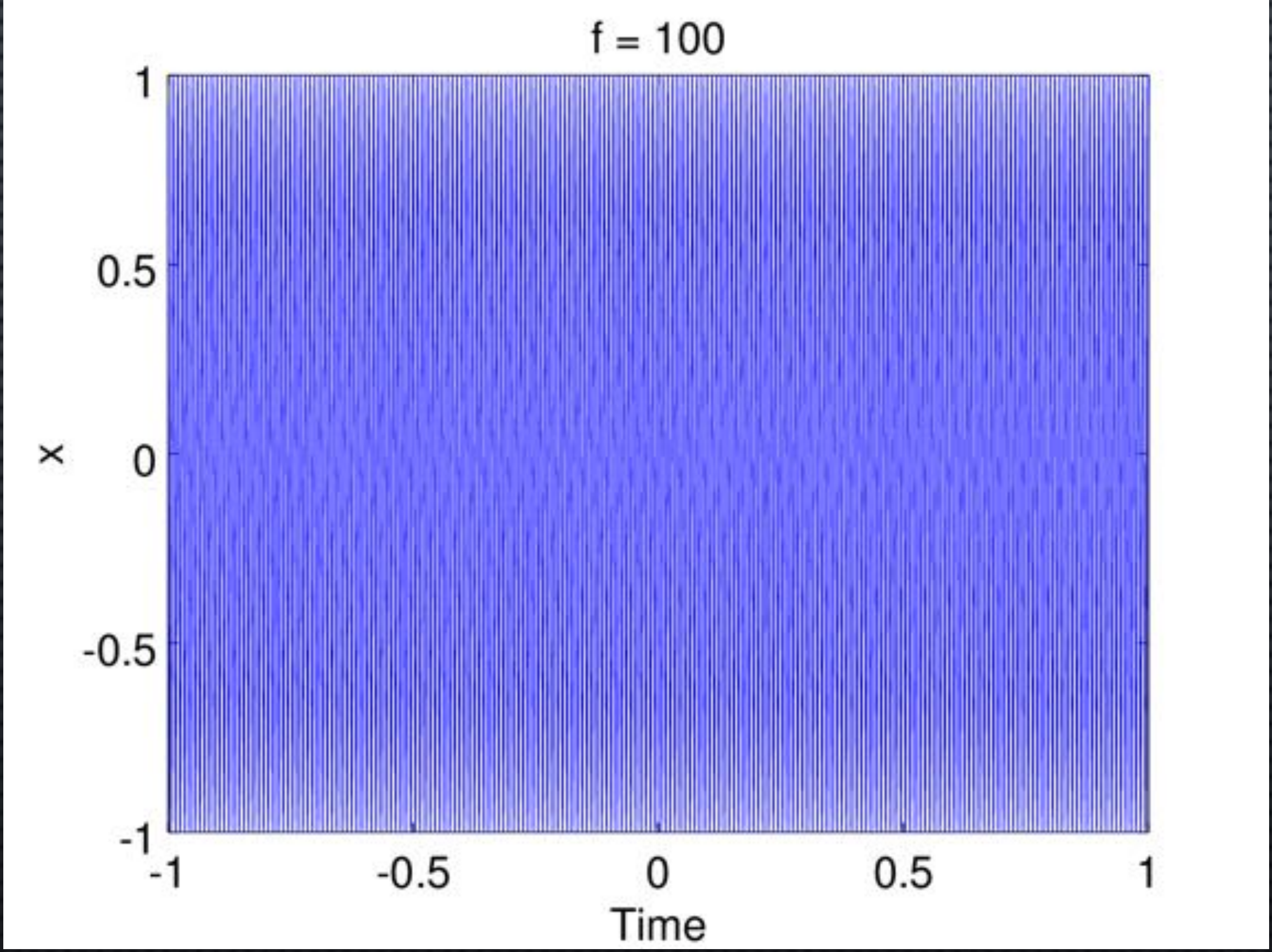
1 Hz ≈ 6.28 rad/s; 1 radian ≈ 57.3°




1
0.5
0
-0.5
-1
x
-1
-0.5
0
0.5
1
Time


1
0.5
0
-0.5
-1
x
-1
-0.5
0
0.5
1
Time

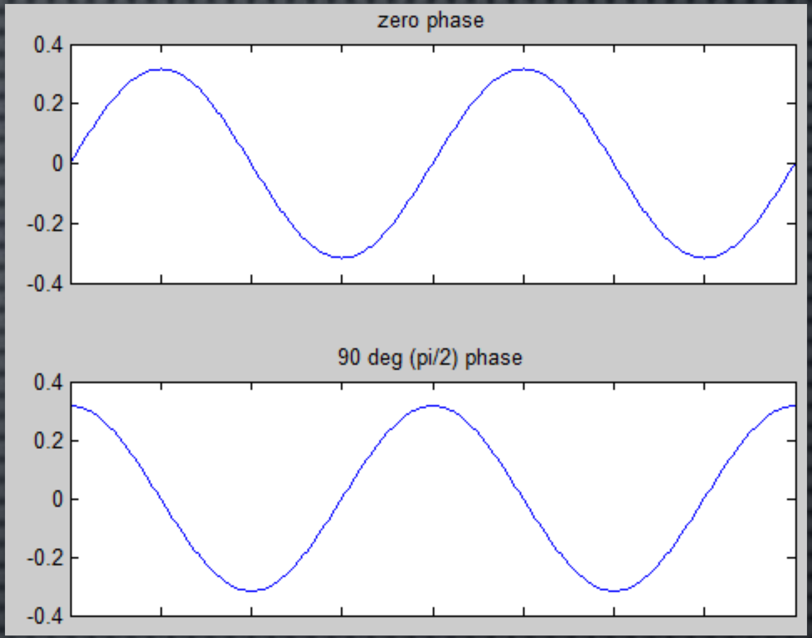

f = 10
1
0.5
0
-0.5
-1
x
-1
-0.5
0
0.5
1
Time

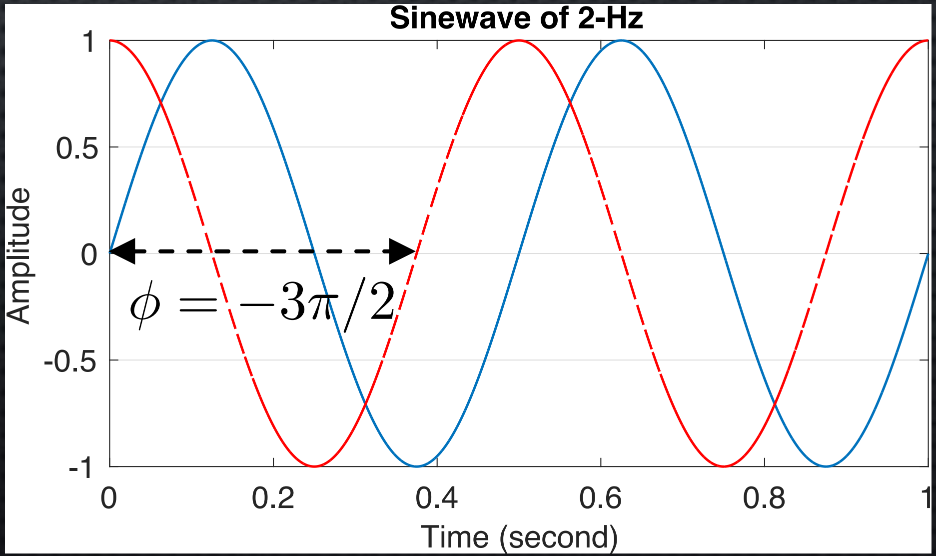

f = 100
1
0.5
x
0
-0.5
-1
-1
-0.5
0
0.5
1
Time

What is phase?

- For a periodic waveform, phase defines the point within the period at which the waveform starts (relative to 0)

- Measured in degrees or radians (360 degrees = 2*pi radians)





Phase in degrees and radians

- ϕ: Phase in radians
- d: phase in angle

$$x(t) = A \cdot \sin(2\pi \cdot f \cdot t + \emptyset)$$

$$x(t) = A \cdot \sin(\omega \cdot t + \emptyset)$$

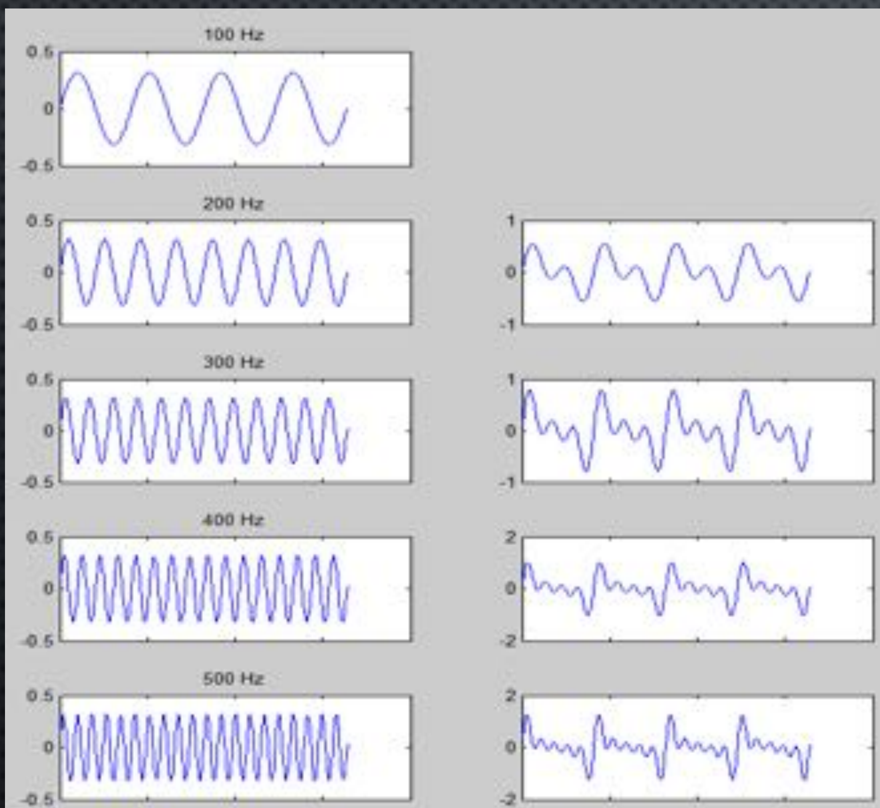
$$d = \frac{360 \cdot \phi}{2\pi}$$

$$\phi = \frac{2\pi \cdot d}{360}$$

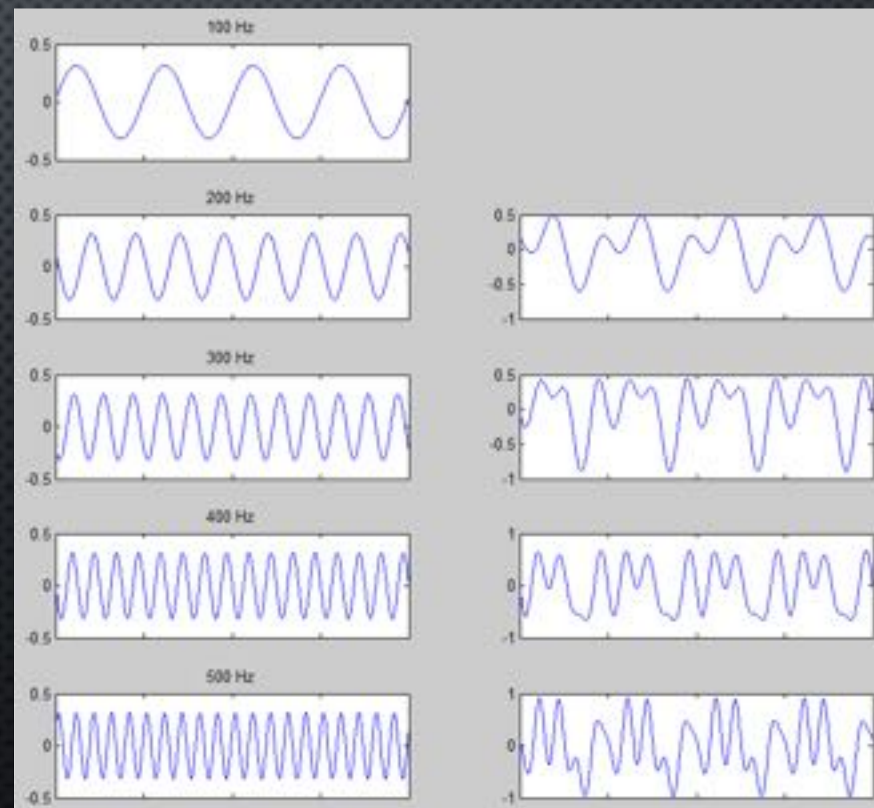
Phase in complex signals

Phase is required to be able to decompose waveforms exactly

phase = 0



phase = random



Properties of continuous signals

$$x(t) = A \cdot \sin(2\pi \cdot f \cdot t + \emptyset)$$

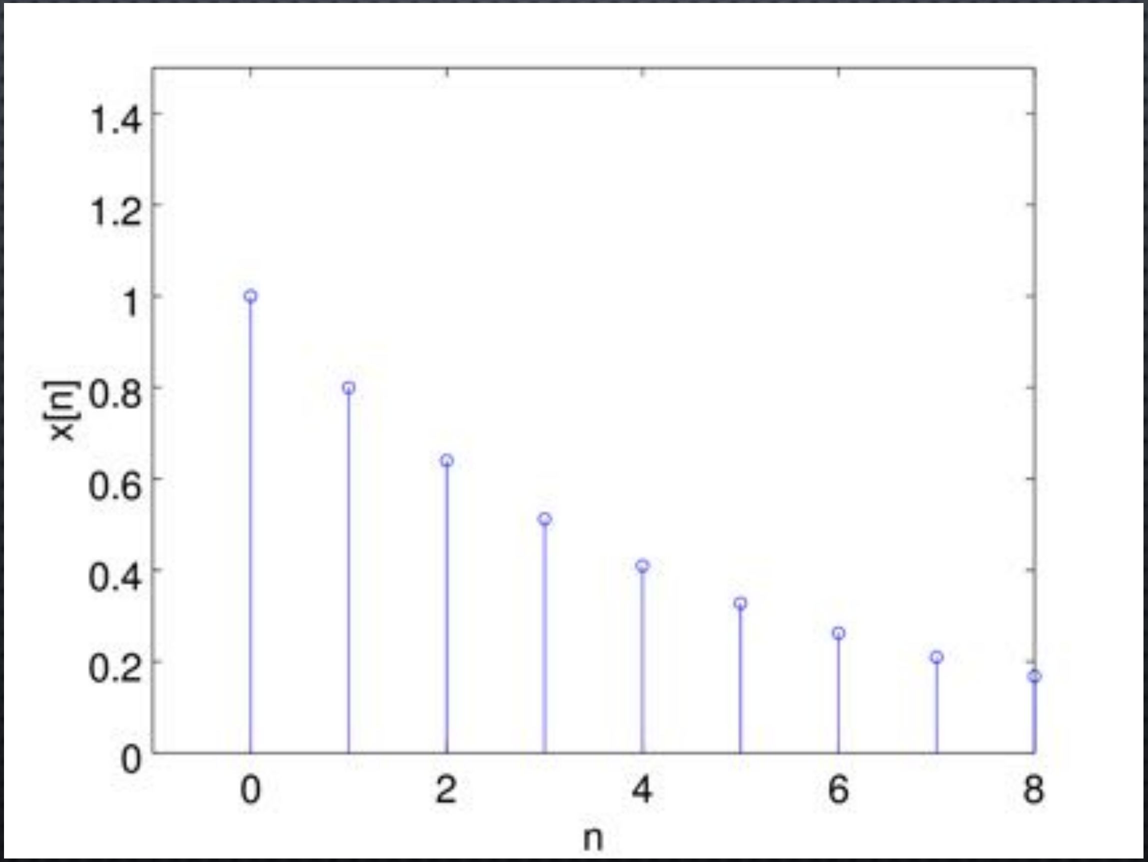
- For every fixed value of *f*, x(t) is periodic with period T = 1/*f*
- Continuous signals with distinct frequencies are themselves distinct
- Increasing *f* results is an increase in the rate of oscillation. As *t* is continuous, we can increase *f* without limit

Continuous vs Discrete time signals

- What does a digital signal look like?
- For a continuous-time signal x(t), you can find a value for x at any value of t
- Discrete time signals are defined only at specific values of time

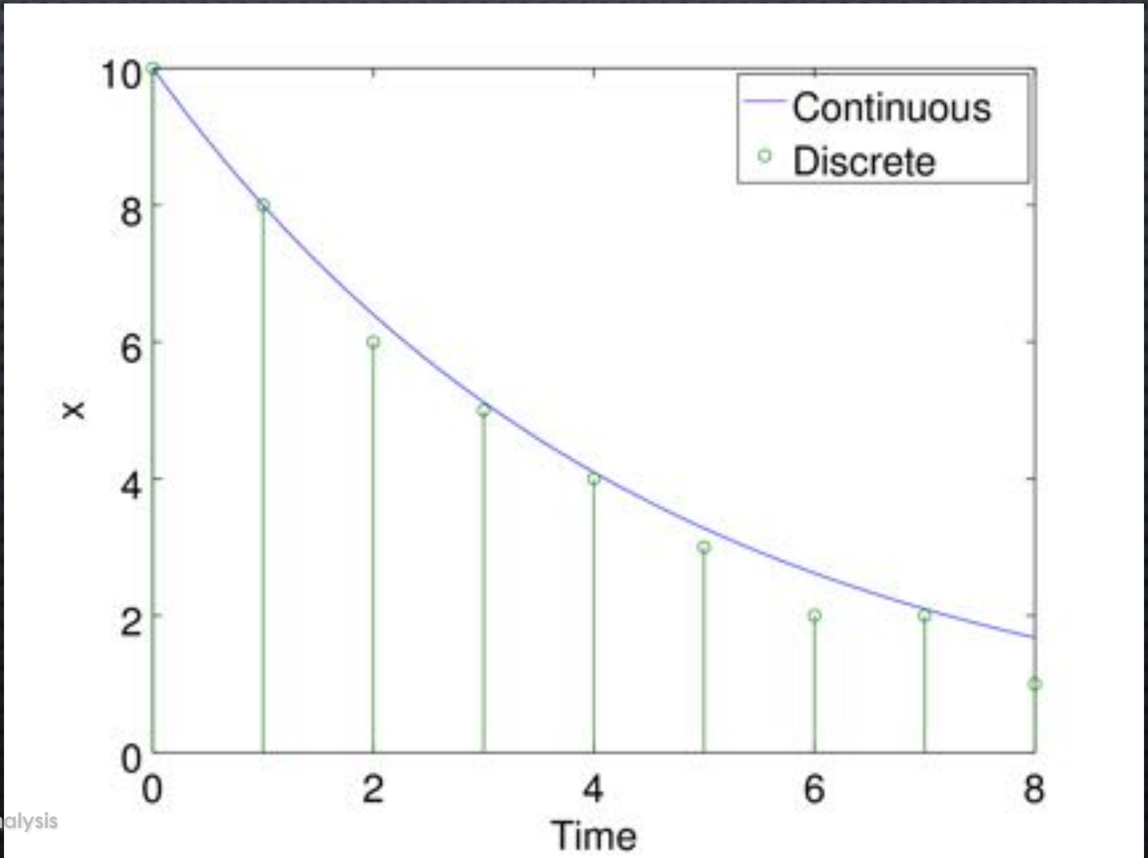
Continuous vs Discrete time signals

- Digital signals are often drawn using stem plots:



Continuous vs Discrete time signals

- Due to the finite resolution in time, the resolution in level is also finite



$$x(t) = A \cdot \sin(2\pi \cdot f \cdot t + \emptyset)$$

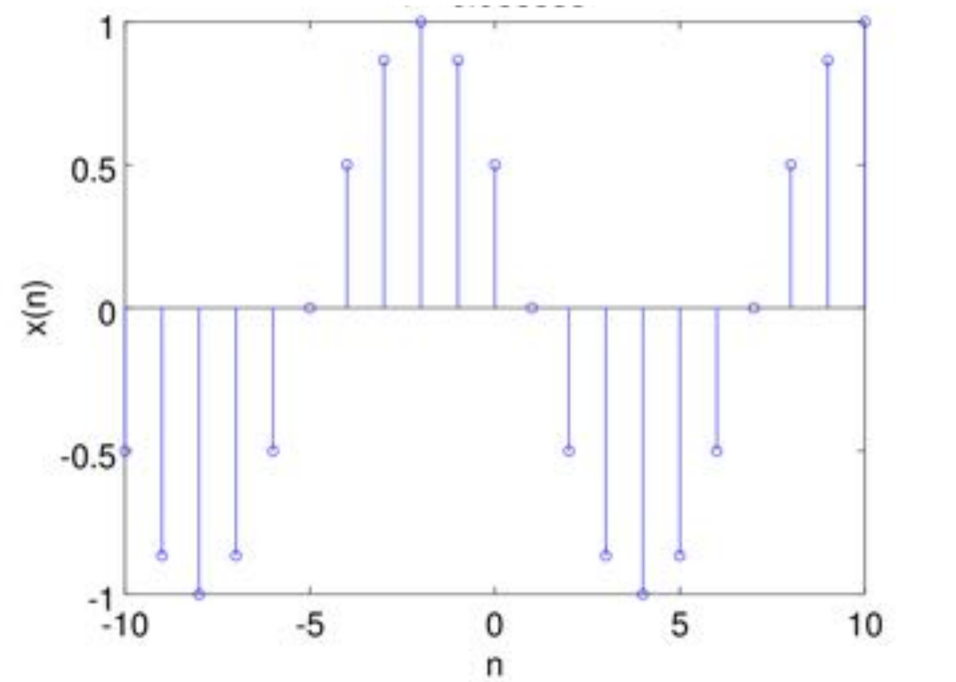
$$x(t) = A \cdot \sin(\omega \cdot t + \emptyset)$$

Properties of discrete time signals

- A discrete time sinusoid can be expressed by:

$$x(n) = A \cdot \cos(\omega' \cdot n + \emptyset)$$

$f' = 1/12 \approx 0.08333$



NOTE:

$\omega' = 2\pi \cdot f'$, where $f' = f/fs$

f': cycles/sample

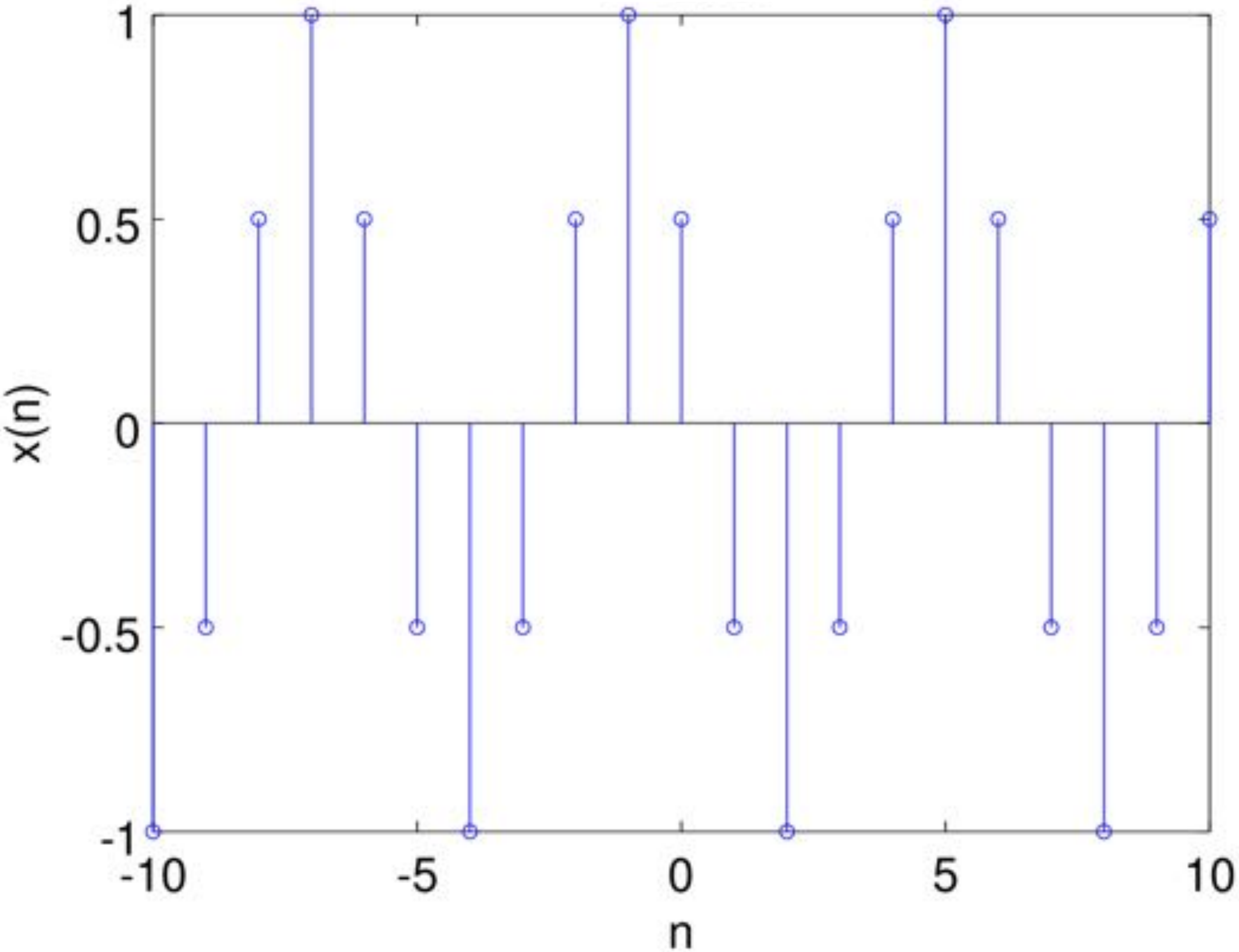
fs: sampling frequency

Properties of discrete time signals

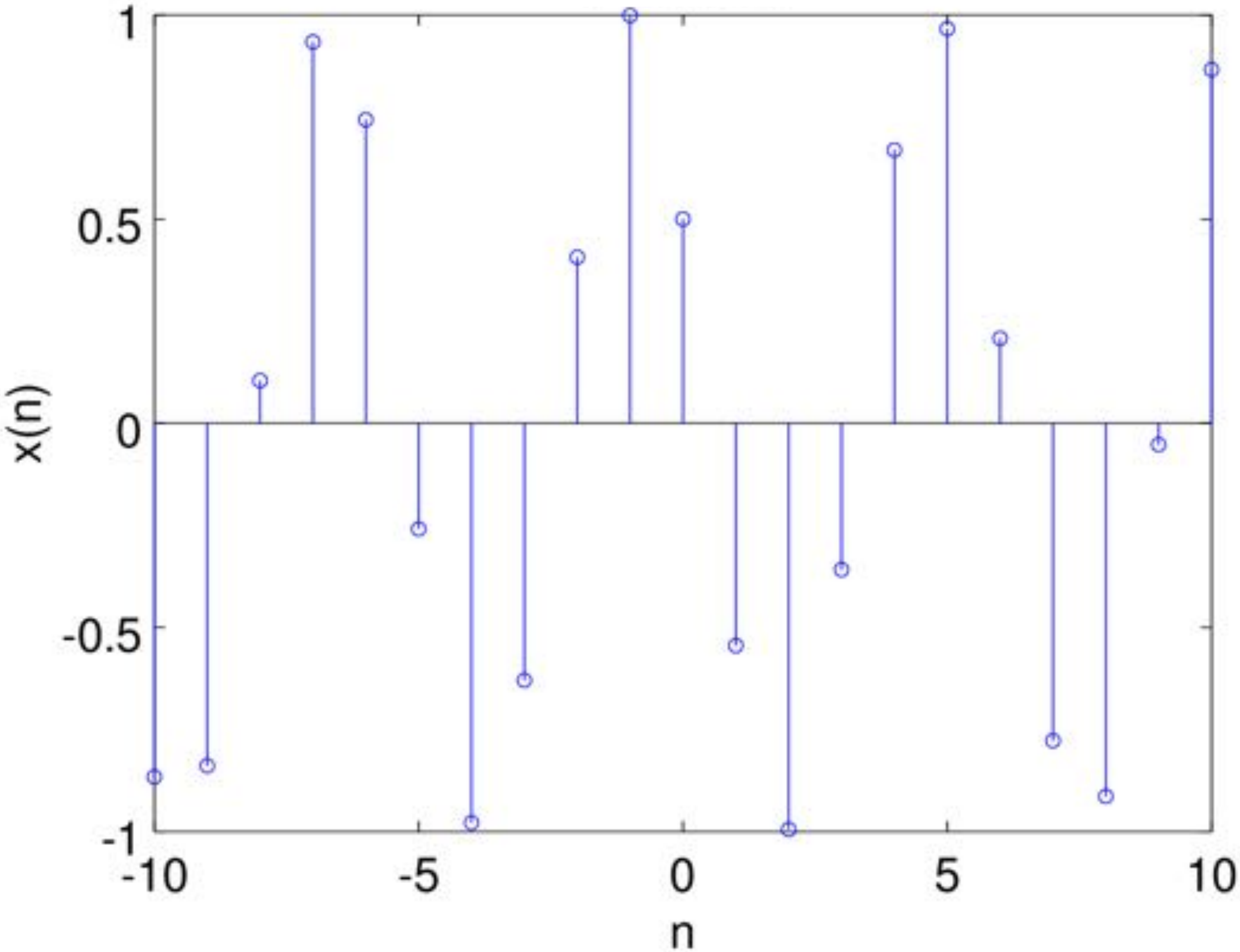
$$x(n) = A \cdot \cos(\omega' \cdot n + \emptyset) = A \cdot \cos(2\pi \cdot f' \cdot n + \emptyset)$$

- A discrete-time signal is periodic if f' is a rational number
  - Fundamental period N
- Discrete time signals whose frequencies are separated by an integer multiple of 2π are identical
- The highest rate of oscillation is when ω= π

$f' = 1/6$



$f' = 2.1/12$

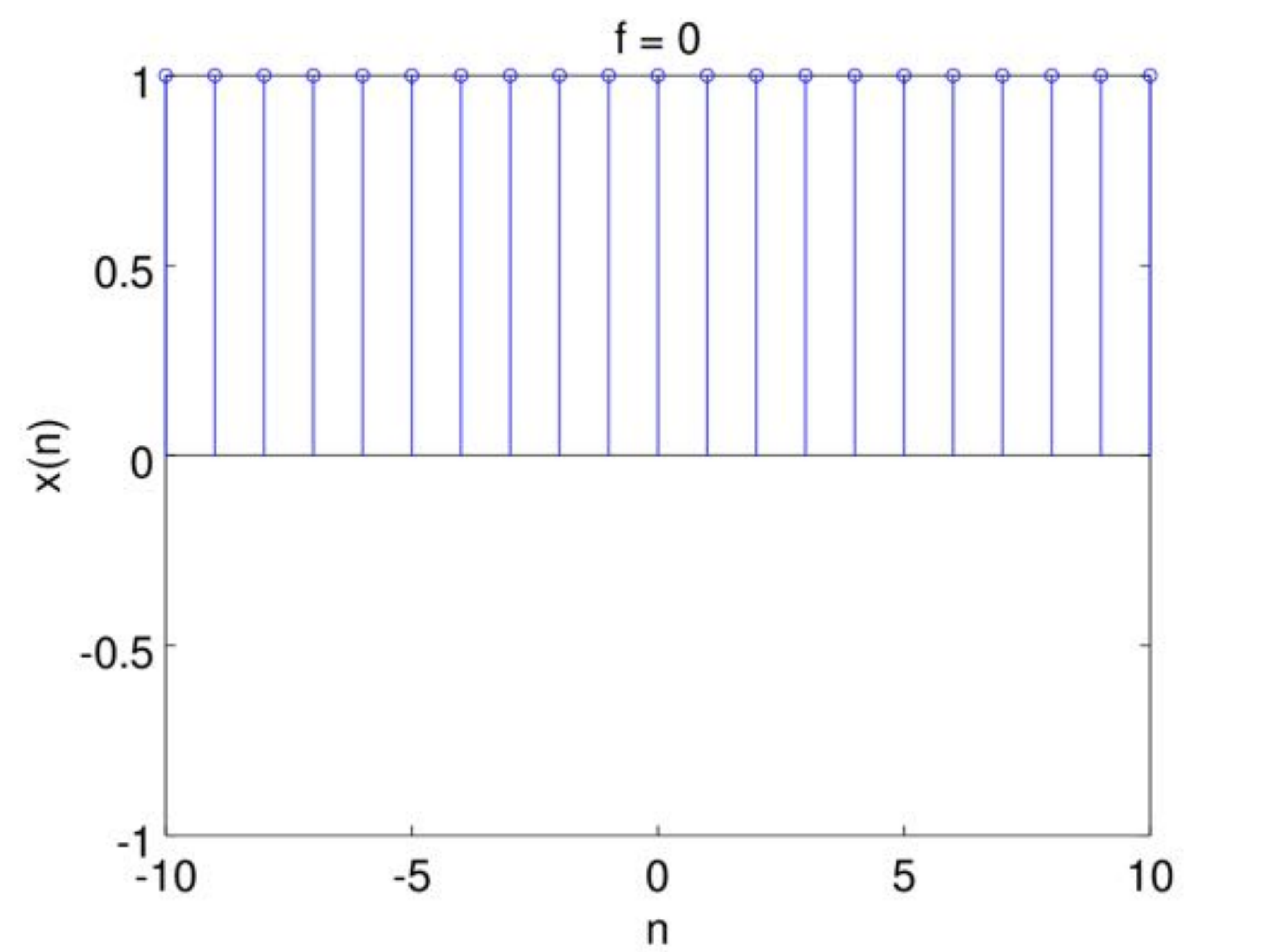


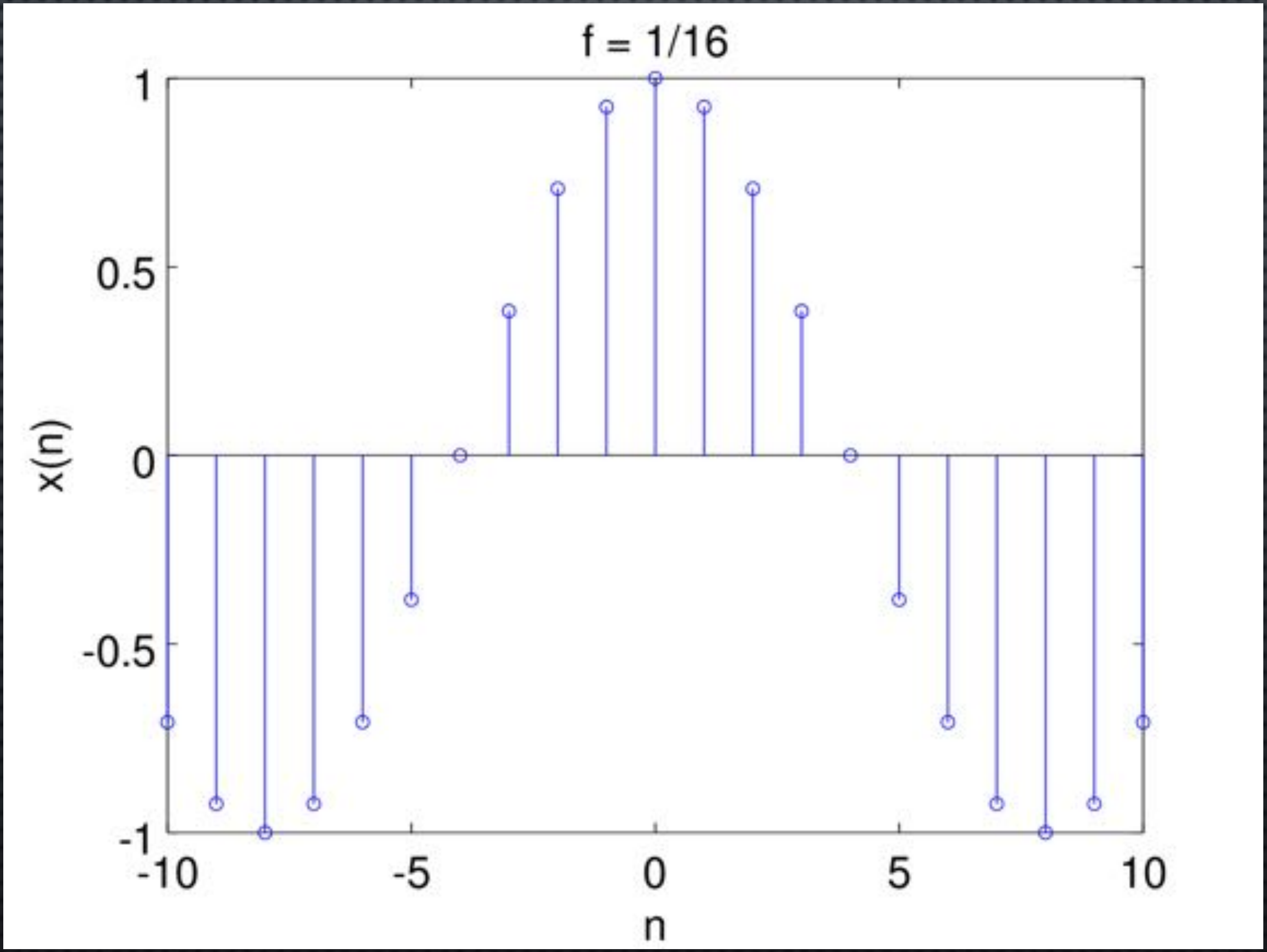
Properties of discrete time signals

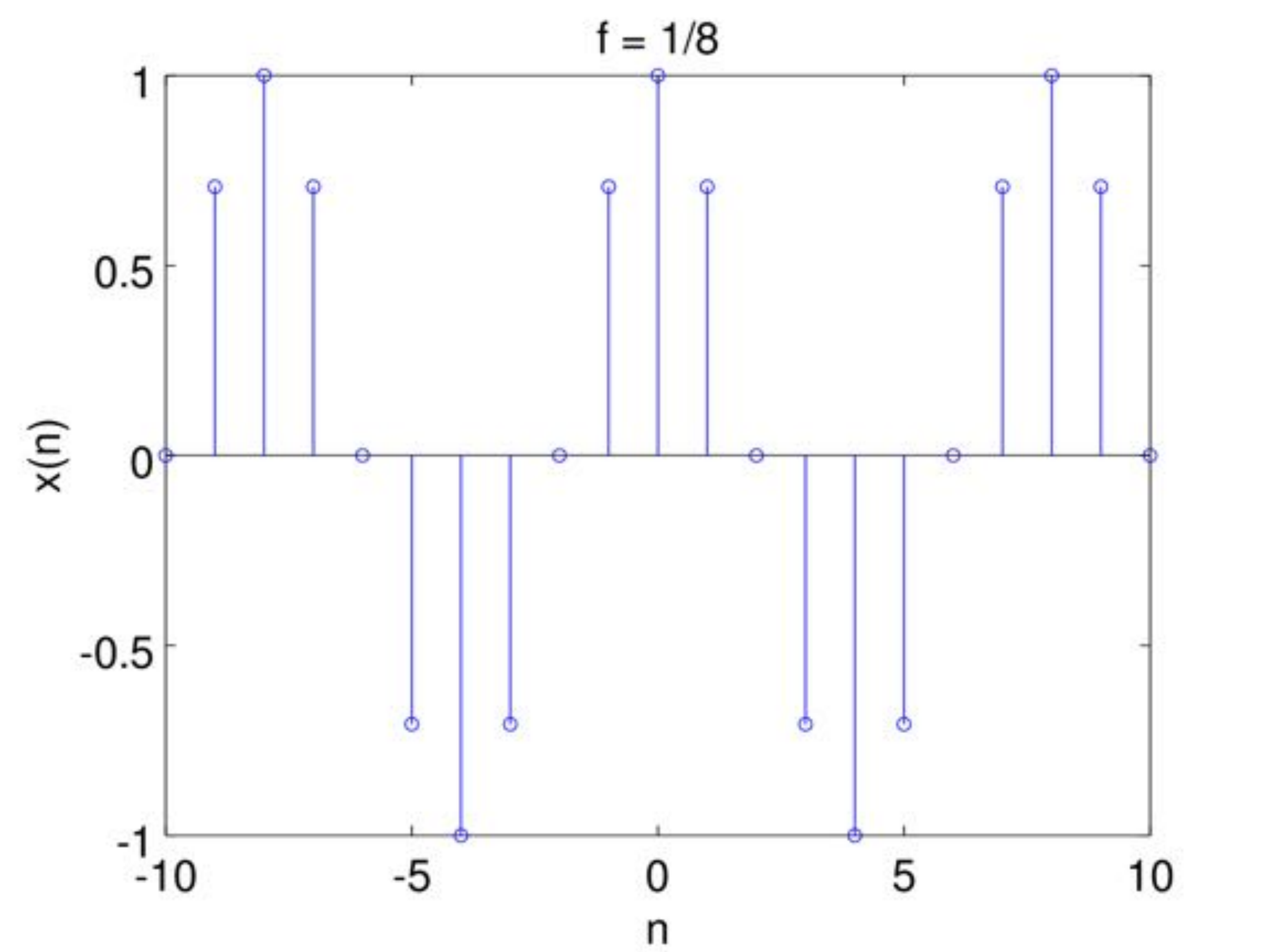
- A discrete-time signal is periodic if *f* is a rational number
- Discrete time signals whose frequencies are separated by an integer multiple of 2π are identical
- The highest rate of oscillation is when $\omega = \pi$.

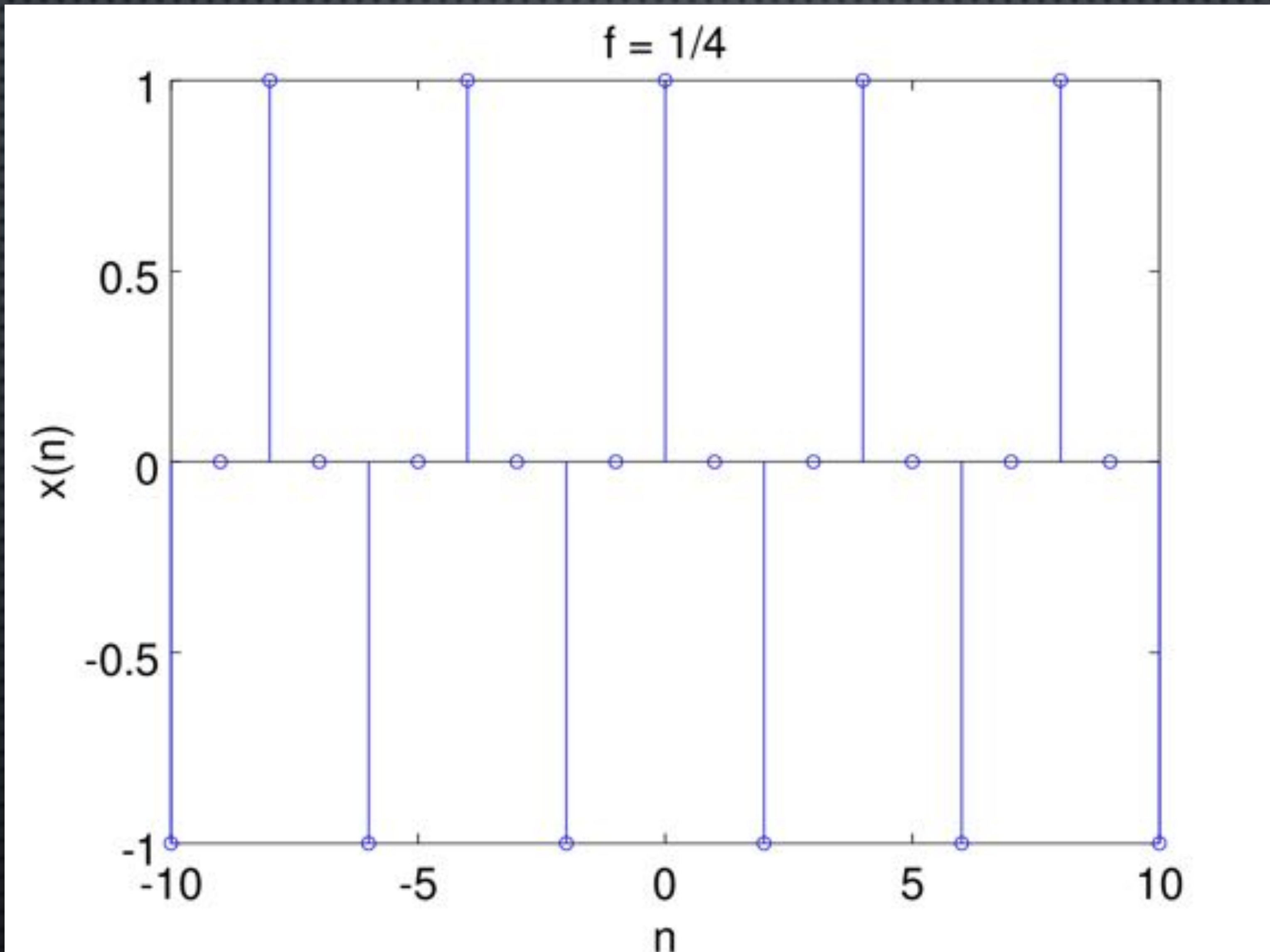
Properties of discrete time signals

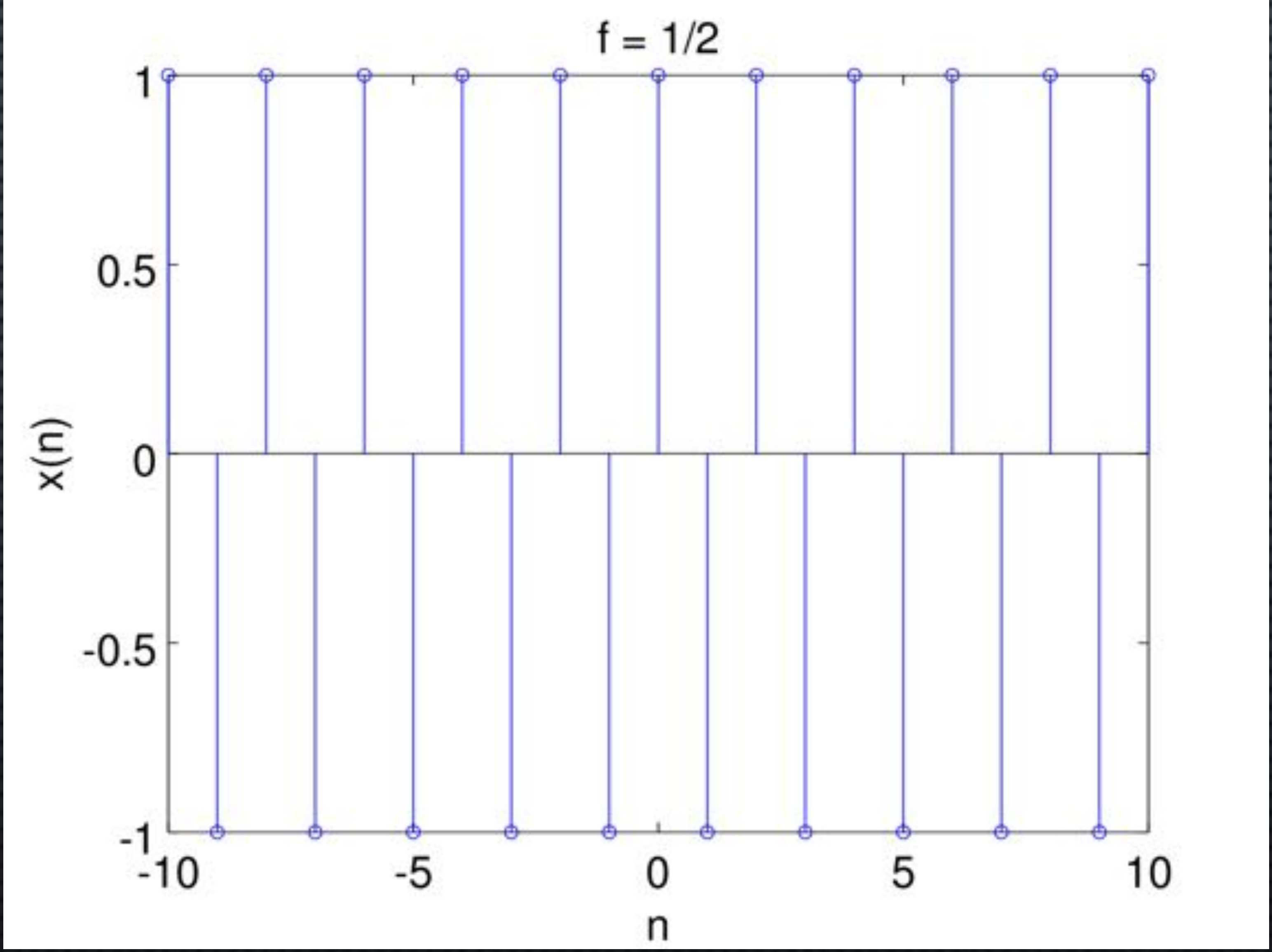
- A discrete-time signal is periodic if *f* is a rational number.
- Discrete time signals whose frequencies are separated by an integer multiple of 2π are identical.
- The highest rate of oscillation is when $\omega = \pi$ or $f' = \frac{1}{2}$


f = 0
1
0.5
0
-0.5
-1
x(n)
-10
-5
0
5
10
n


f = 1/16
x(n)
n
1
0.5
0
-0.5
-1
-10
-5
0
5
10


f = 1/8
1
0.5
0
-0.5
-1
x(n)
-10
-5
0
5
10
n


f = 1/4
1
0.5
0
-0.5
-1
x(n)
-10
-5
0
5
10
n



f = 1/2
1
0.5
0
-0.5
-1
x(n)
-10
-5
0
5
10
n

Properties of discrete time signals

- What happens between $\pi \leq \omega \leq 2\pi$?

- This is called aliasing…
  - If the angular frequency of a discrete time signal increased from π to 2π, its rate of oscillation will decrease!

Summary of key points

- For a continuous signal x(t), we can find the value of x at any point in time

- Discrete time signals are defined only at specific values of time

Summary of key points

Continuous signal	Discrete signal
For every fixed value of *f*, x(t) is periodic	A discrete-time signal is periodic if *f'* is a rational number
Continuous signals with distinct frequencies are themselves distinct	Discrete-time signals whose frequencies are separated by an integer multiple of 2π are identical
Increasing *f* results in an increase in the rate of oscillation. As *t* is continuous, we can increase *f* without limit	The highest rate of oscillation is when $\omega = \pi$

LING 446

Fundamentals for Speech Signal Processing and Analysis

Yan Tang

Department of Linguistics, UIUC

Week 3: Sampling and quantisation

Last week…

$$x(t) = A \cdot \sin(2\pi \cdot f \cdot t - \phi)$$

- Properties of continuous periodic signals
  - Amplitude, frequency, period and phase
  - Relationship between regular and angular frequency: $\omega = 2\pi \cdot f$
  - Conversion between phase in degree and radians: $d = \frac{360 \cdot \phi}{2\pi}$ and $\phi = \frac{2\pi \cdot d}{360}$

Last week…

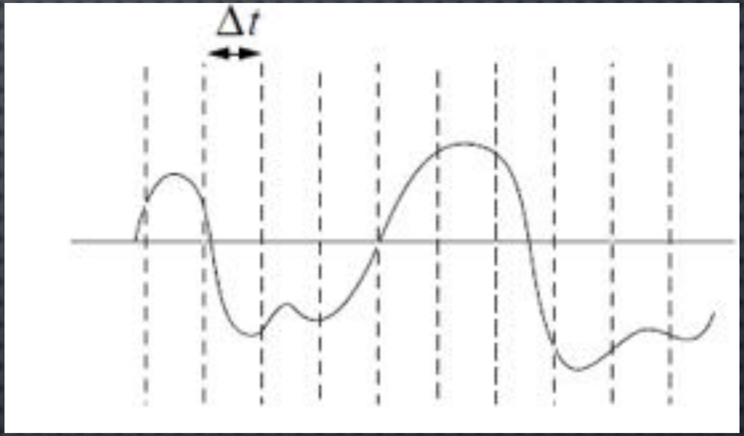
$$x(n) = A \cdot \cos(\omega' \cdot n - \phi)$$

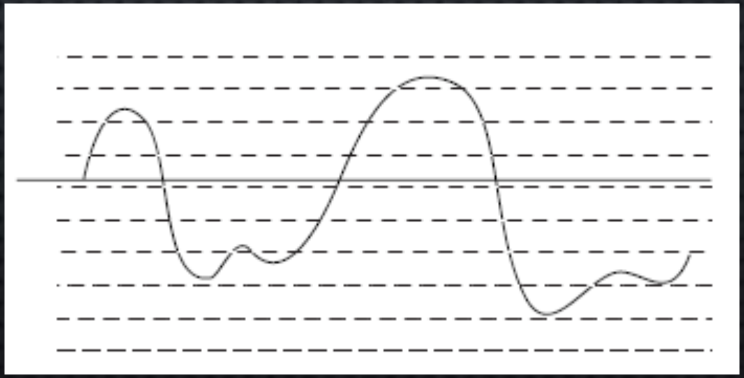
$$\omega' = 2\pi \cdot f'$$

- Properties of discrete periodic signals
  - Frequency f' : cycle/sample. Must be a rational number
  - Signals with frequencies of $(\omega_N + k \cdot 2\pi)$ are identical
  - The highest effective angular frequency is $\omega' = \pi$, i.e. $f' = 1/2$

Digitalisation

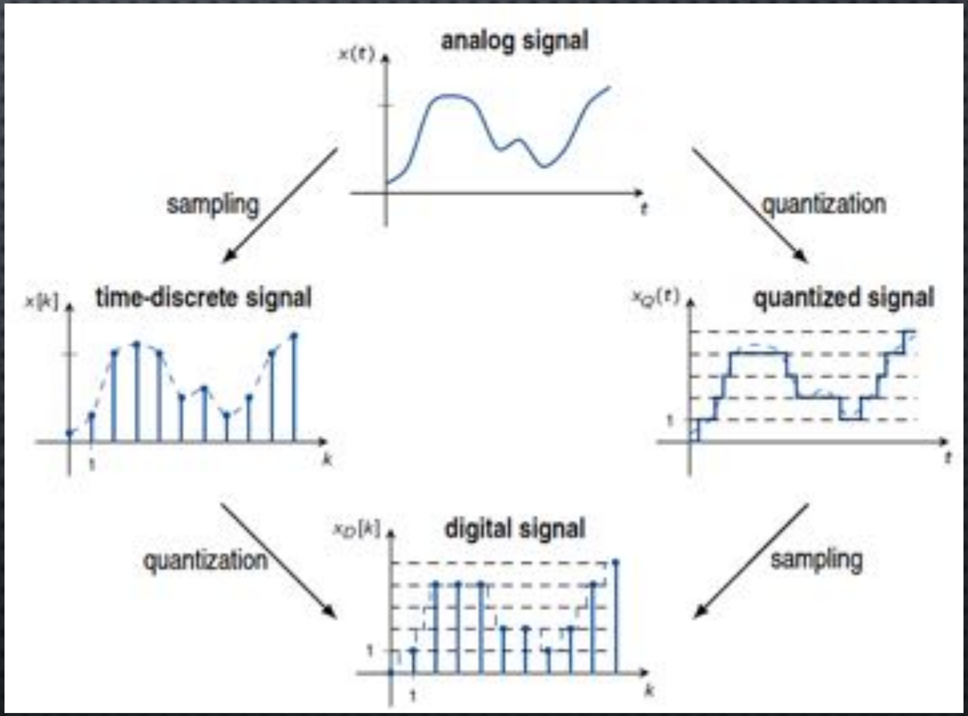
- DSP requires that the signal is *sampled* **in time and** *quantised* **in amplitude**
- Both have the potential to lose information or add artefacts to the signal
- Often inevitable; the consequences should be aware of





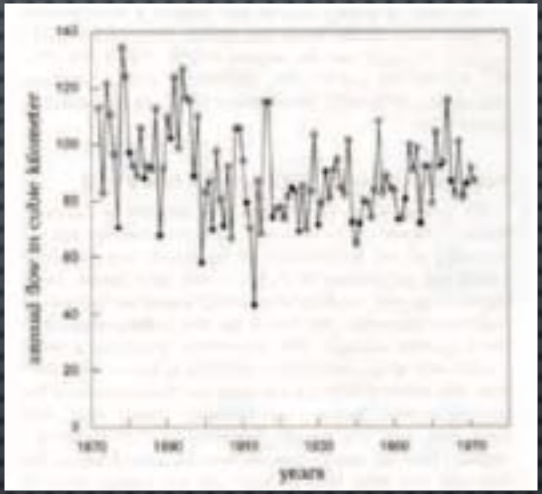
Sampling

- Sampling is the process of breaking up a signal in time
- Digital systems deal with discrete data rather than continuous data
- This process has a few consequences for our signal



DEMO 1

The earliest application of DSP sampling



The Palermo Stone

- Ancient Egyptian back to 25th century BC
- Recorded the data of annual flood level
- Resembles the current approach of data sampling

Relationship between continuous time and sample number

- The number of samples taken in each second is called the *sampling frequency* or *sampling rate* (usually denoted as fs)

$$t = n\Delta_s = \frac{n}{f_s}$$

$\boldsymbol{t}$: continuous time in seconds

$\boldsymbol{f_s}$: sampling frequency in *Hz*

$\boldsymbol{n}$: number of samples

Δ_s: the sampling period, i.e. $\frac{1}{f_s}$

Consequence of all this?

Only frequencies below half of the sampling rate will be accurately represented after sampling!

e.g. To sample a 100 Hz sine wave, an $f_s > 200$ Hz is required

But why? What will happen if f_s does not meet the above rule?

DEMO 2

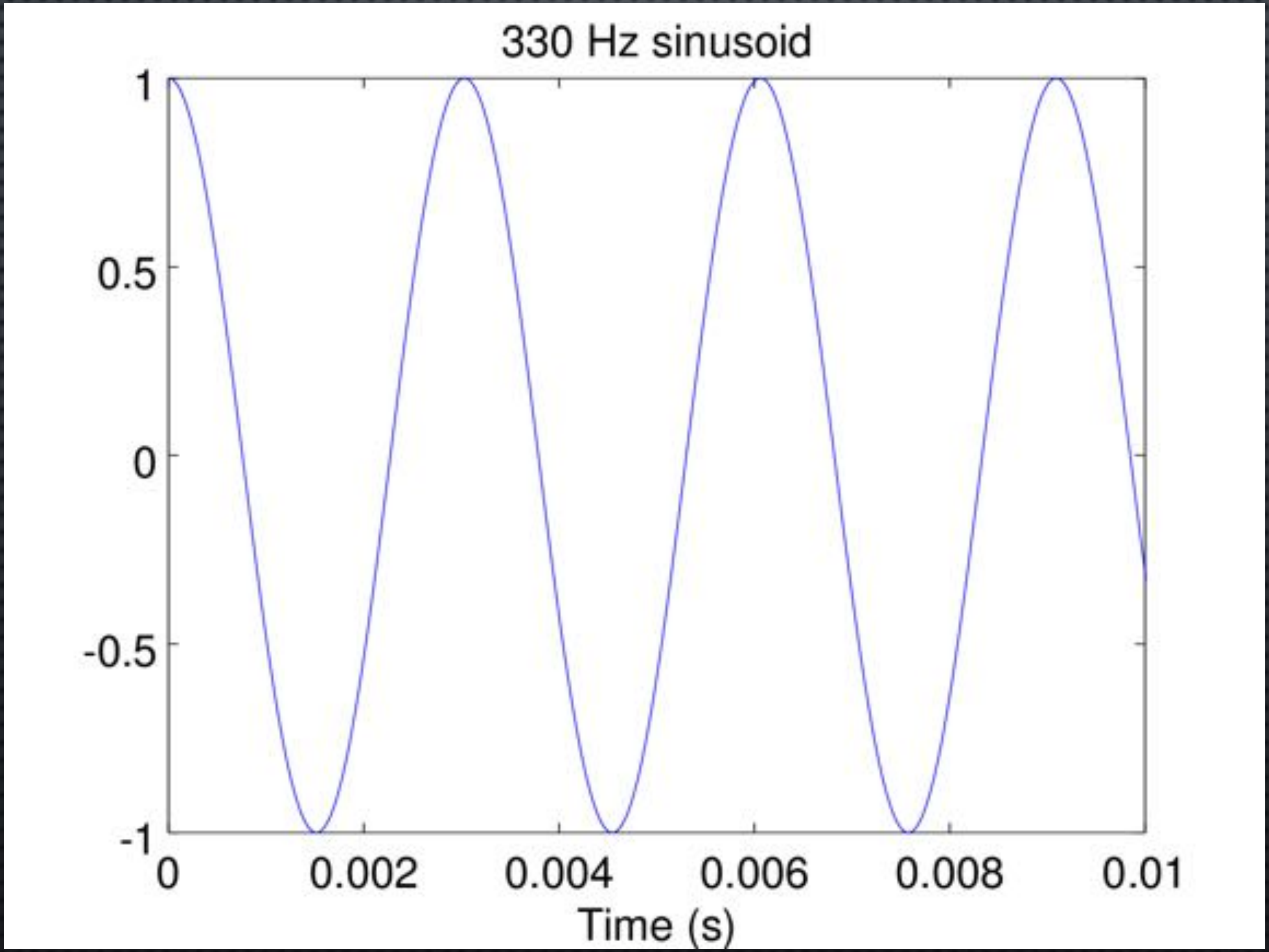
Nyquist frequency

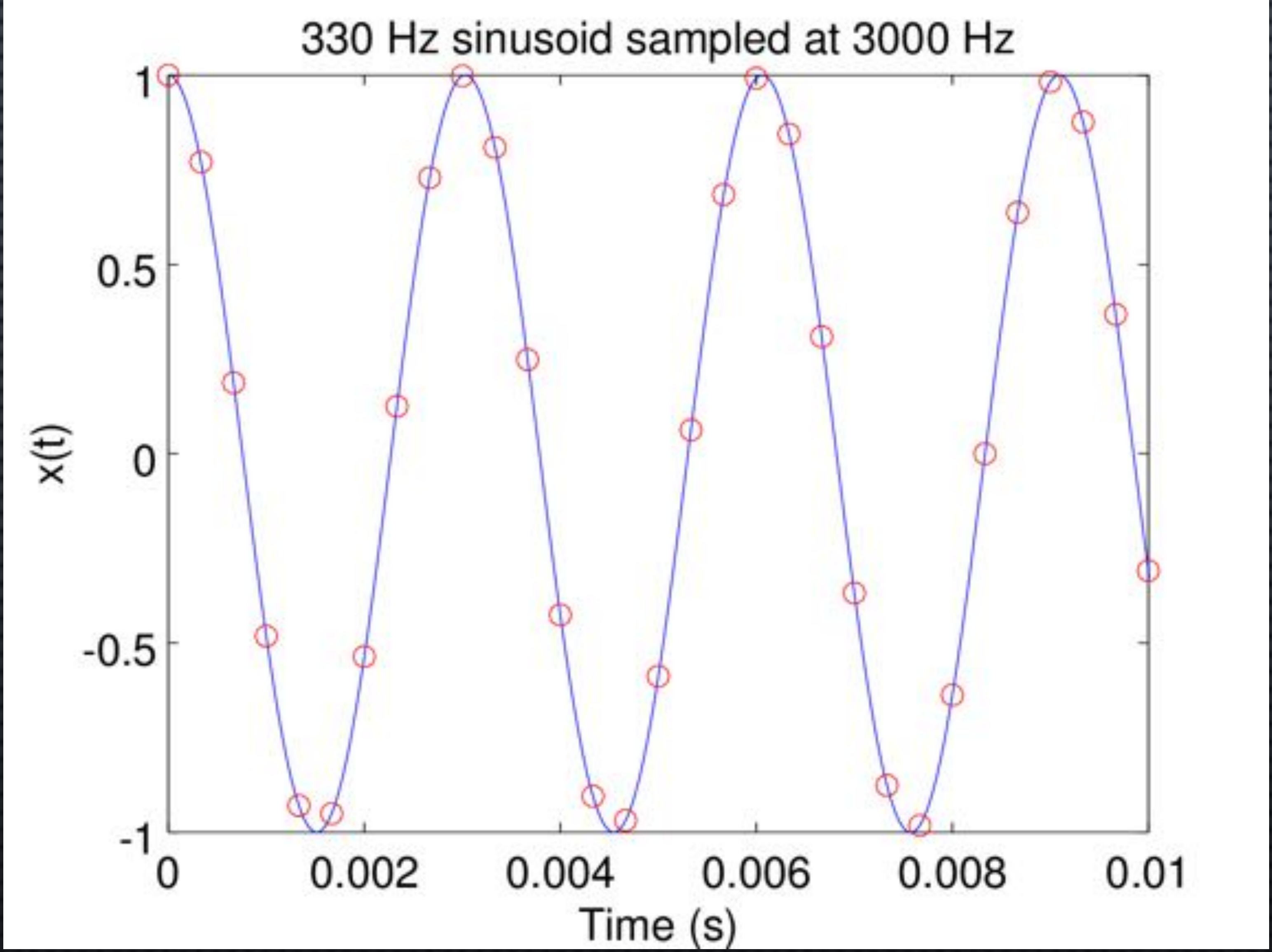
- This frequency ($f_s/2$) is called the *Nyquist frequency* or *Nyquist limit*

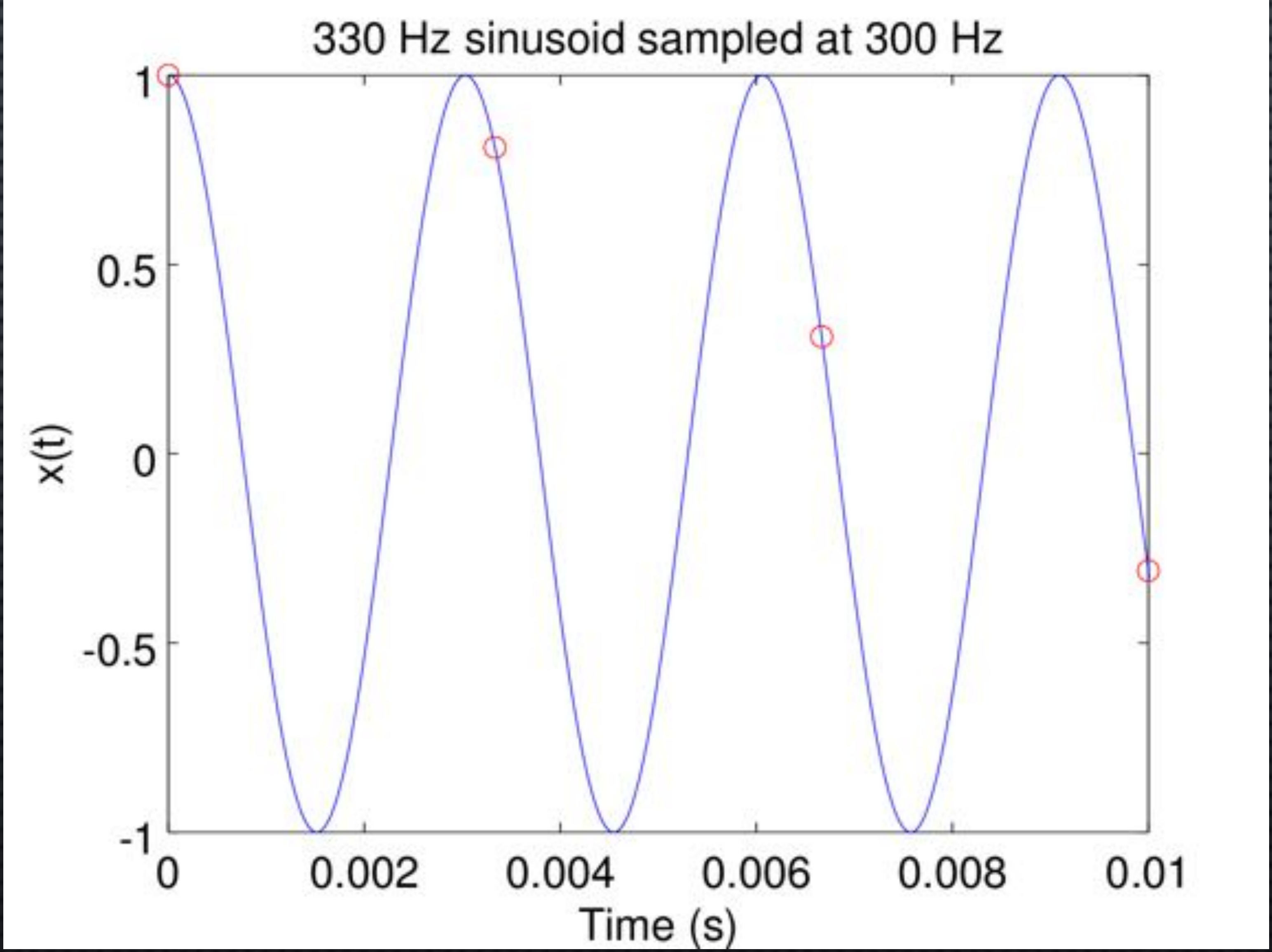
← After this guy: Harry Nyquist

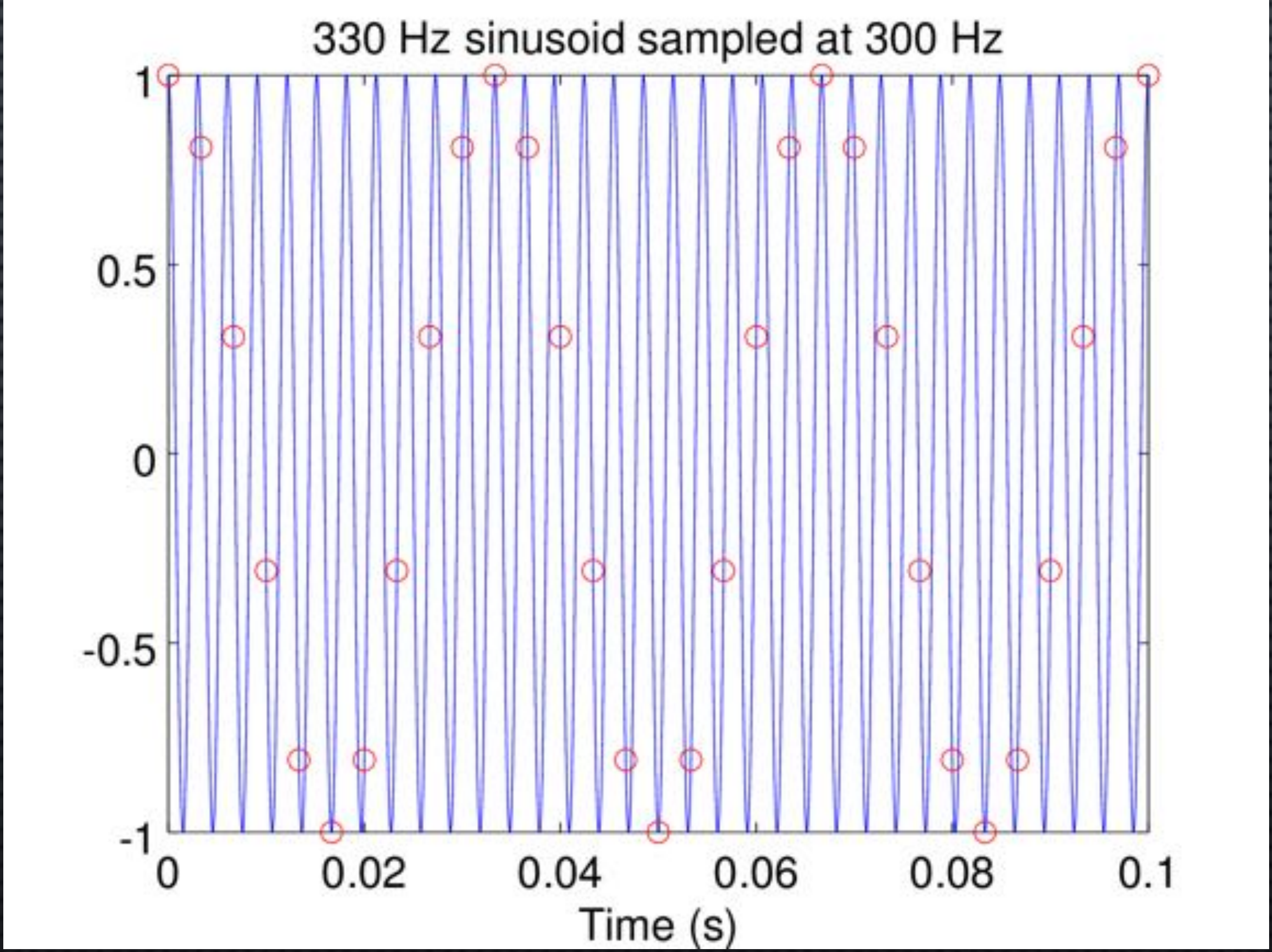
Nyquist-Shannon (Sampling) theorem

A continuous (analogue) signal can be recovered from a sampled signal if the signal is sampled at a rate $f_s > 2f$


330 Hz sinusoid
1
0.5
0
-0.5
-1
0
0.002
0.004
0.006
0.008
0.01
Time (s)


330 Hz sinusoid sampled at 3000 Hz
x(t)
1
0.5
0
-0.5
-1
0
0.002
0.004
0.006
0.008
0.01
Time (s)


330 Hz sinusoid sampled at 300 Hz
x(t)
1
0.5
0
-0.5
-1
0
0.002
0.004
0.006
0.008
0.01
Time (s)


330 Hz sinusoid sampled at 300 Hz
1
0.5
0
-0.5
-1
0
0.02
0.04
0.06
0.08
0.1
Time (s)

Demo 3

Without limiting the frequency of the input signal, the sampled signal – which should not exist – is a mirrored image of the signal at the Nyquist Frequency/Limit

i.e. Given a sampling rate of f_s Hz, a sampled sine wave at a frequency of f (where $\frac{f_s}{2} < f \leq f_s$) is indistinguishable from a sine wave at a frequency of $(f - f_s)$ Hz

What will happen when $f > f_s$?

Sinewave aliasing

The Sampled Sinewave Theorem:

Given a sampling rate of f_s Hz, and an integer k, a sine wave at a frequency of f $(0 < f \leq \frac{f_s}{2})$ is indistinguishable from a sine wave at a frequency of $(f + k \cdot f_s)$ after being sampled by f_s

$f_s = 20$ Hz, what are the aliases for a sine wave of 4 Hz (only considering positive values of k)?

e.g. given an $f_s = 20$ Hz, a sine wave with a frequency of 4 Hz is indistinguishable from sine waves at 24 Hz, 44 Hz, 64 Hz….

DEMO 4

The wagon wheel effect

Phenomenon: on film or television, wheels sometimes seem to spin backwards, although the vehicle is going forward.

- Under-sampling
- Aliasing

DEMO 5

Question

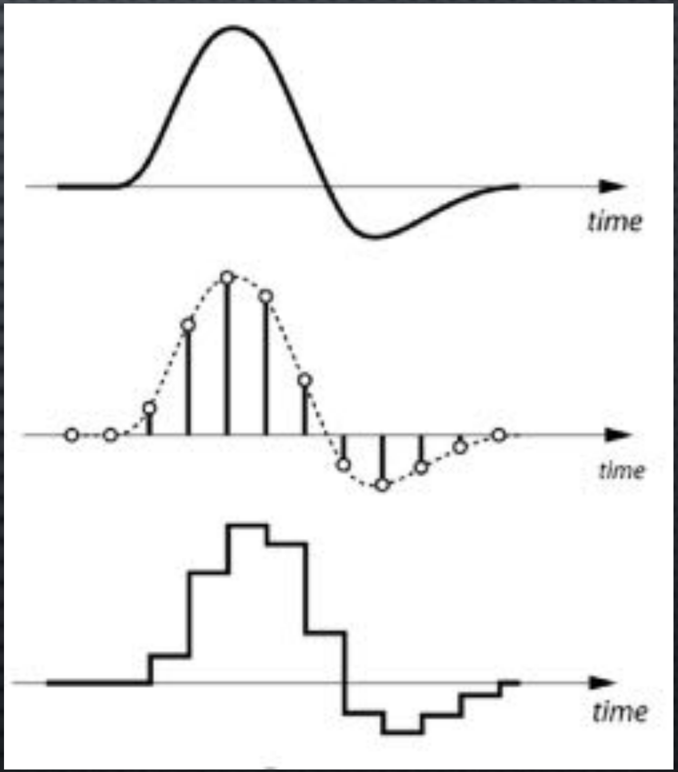
What’s the minimum sampling frequency we need for ideal audio?

- The range of human hearing: 20 - 20000 Hz
- We need to sample at the double that rate, i.e. at least 40 kHz
- CD audio is sampled at 44.1kHz. Professional audio is often sampled at 48 kHz.

Quantisation

- As well as sampling, we also need to decimate in level.

- The process of converting a discrete-time continuous amplitude signal into a digital signal by expressing each sample as a finite number of digits is called ***quantisation***.



Why quantise samples?

- Consider the signal:

$$x(n) = \begin{cases} 0.9^n, & n \geq 0 \\ 0, & n < 0 \end{cases}$$

- How many significant digits do we need to represent this signal?

Why quantise samples?

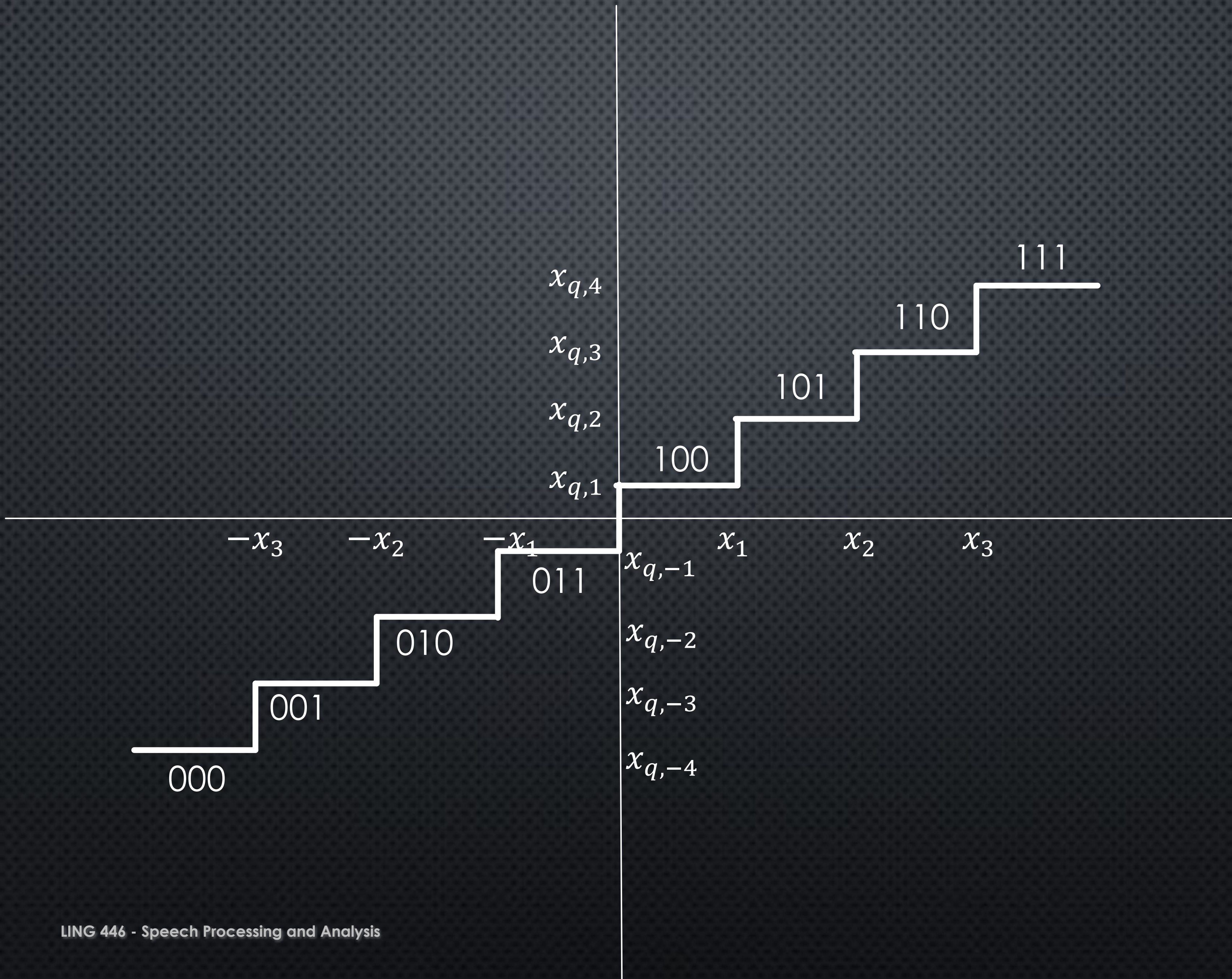
$$x(n) = \begin{cases} 0.9^n, & n \geq 0 \\ 0, & n < 0 \end{cases}$$

n	x(n)
0	1
1	0.9
2	0.81
3	0.729
4	0.6561
5	0.59049
6	0.531441
7	0.4782969
8	0.43046721
9	0.387420489

- *n* significant digits are need for each $x(n)$

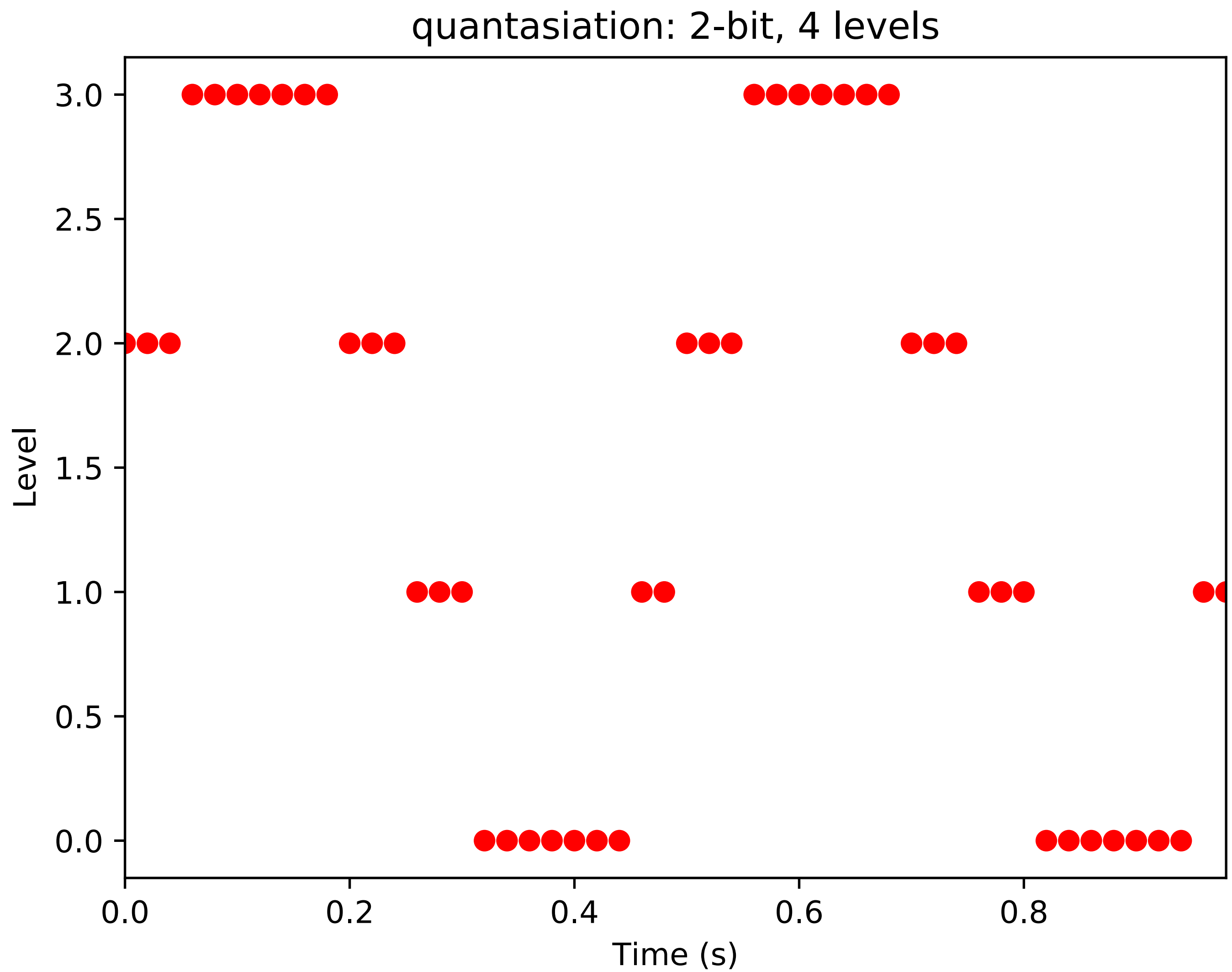
Why quantise samples?

n	x(n)	$x_q(n)$ (Rounding)	$e_q(n)$ (Rounding)
0	1	1	0
1	0.9	0.9	0
2	0.81	0.8	-0.01
3	0.729	0.7	-0.029
4	0.6561	0.7	0.0439
5	0.59049	0.6	0.00951
6	0.531441	0.5	-0.031441
7	0.4782969	0.5	0.0217031
8	0.43046721	0.4	-0.03046721
9	0.387420489	0.4	0.012579511


111
110
101
100
011
010
001
000
$x_{q,4}$
$x_{q,3}$
$x_{q,2}$
$x_{q,1}$
$x_{q,-1}$
$x_{q,-2}$
$x_{q,-3}$
$x_{q,-4}$
$-x_3$
$-x_2$
$-x_1$
x_1
x_2
x_3

Quantisation levels

- The values allowed in the digital signal
- The distance between two quantisation levels is called the *quantisation step*
- In a digital system, the number of bits (quantisation resolution) determines the number of levels – ($2^{n\text{-bit}}$) levels
  - e.g. 8 bits -> 256 levels

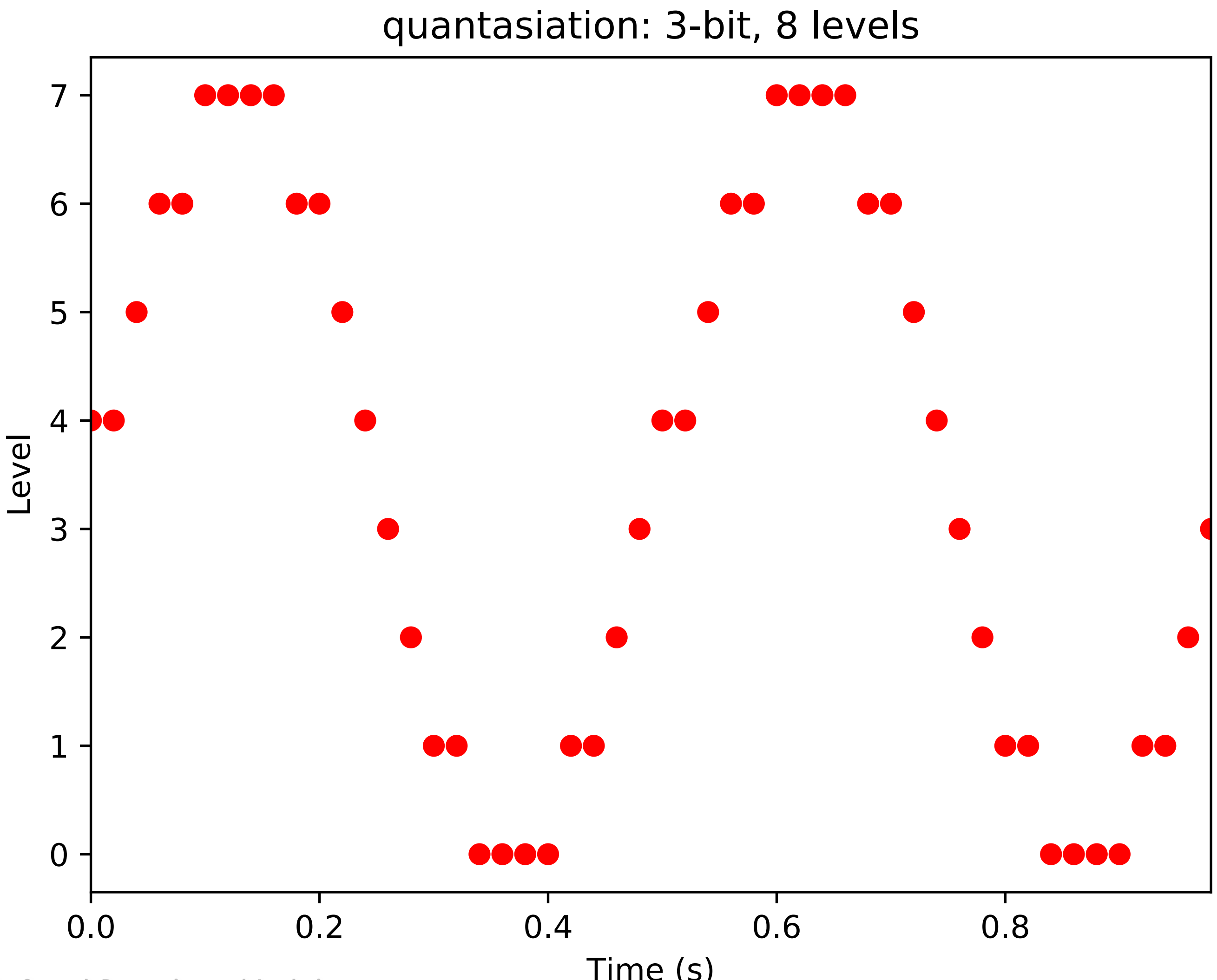

quantasiation: 2-bit, 4 levels
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Level
0.0
0.2
0.4
0.6
0.8
Time (s)

1

0

1

0


quantasiation: 3-bit, 8 levels
Level
Time (s)
0
1
2
3
4
5
6
7
0.0
0.2
0.4
0.6
0.8

Quantisation error

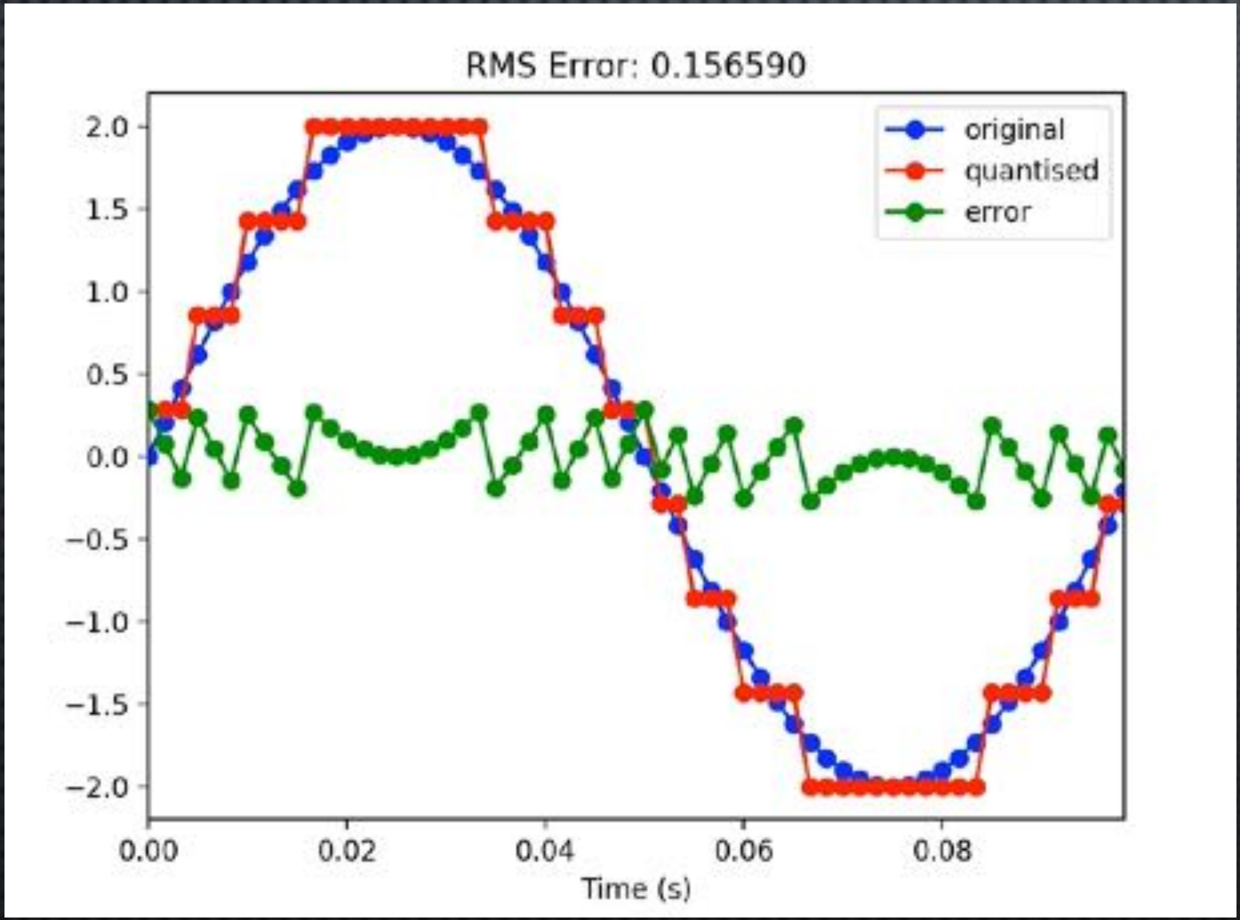
- The quantisation error is a sequence, e_q, defined as the difference between the quantised values and the original values:

$$e_q(n) = x_q(n) - x(n)$$

Quantisation error is unavoidable

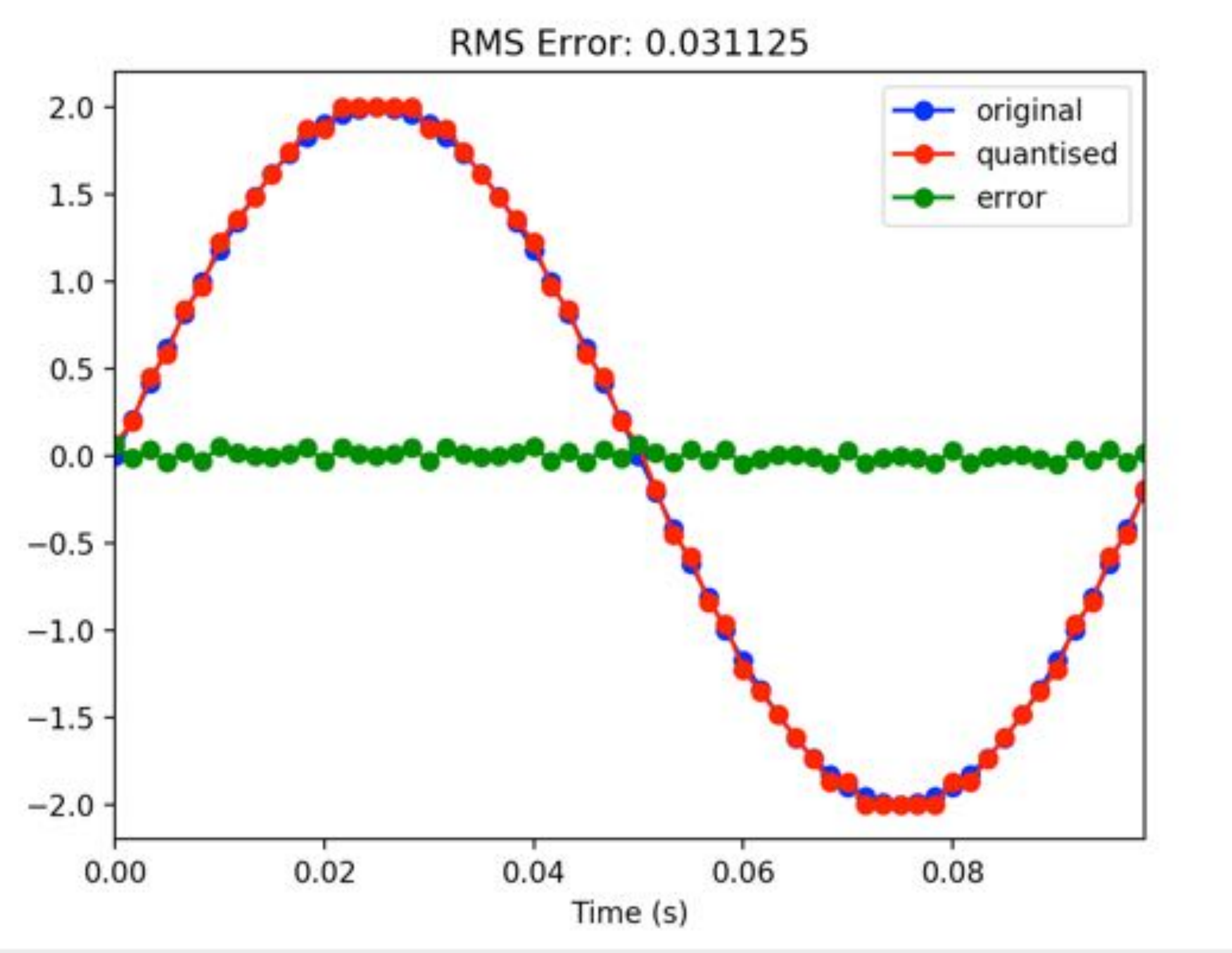
- There will always be quantisation noise in the signal
- It is always possible to talk about the signal-to-noise ratio (SNR) of a "clean" digital signal
- Make use of enough bits for high resolution

ERRORS: 3-BIT 8-LEVEL QUANTISATION

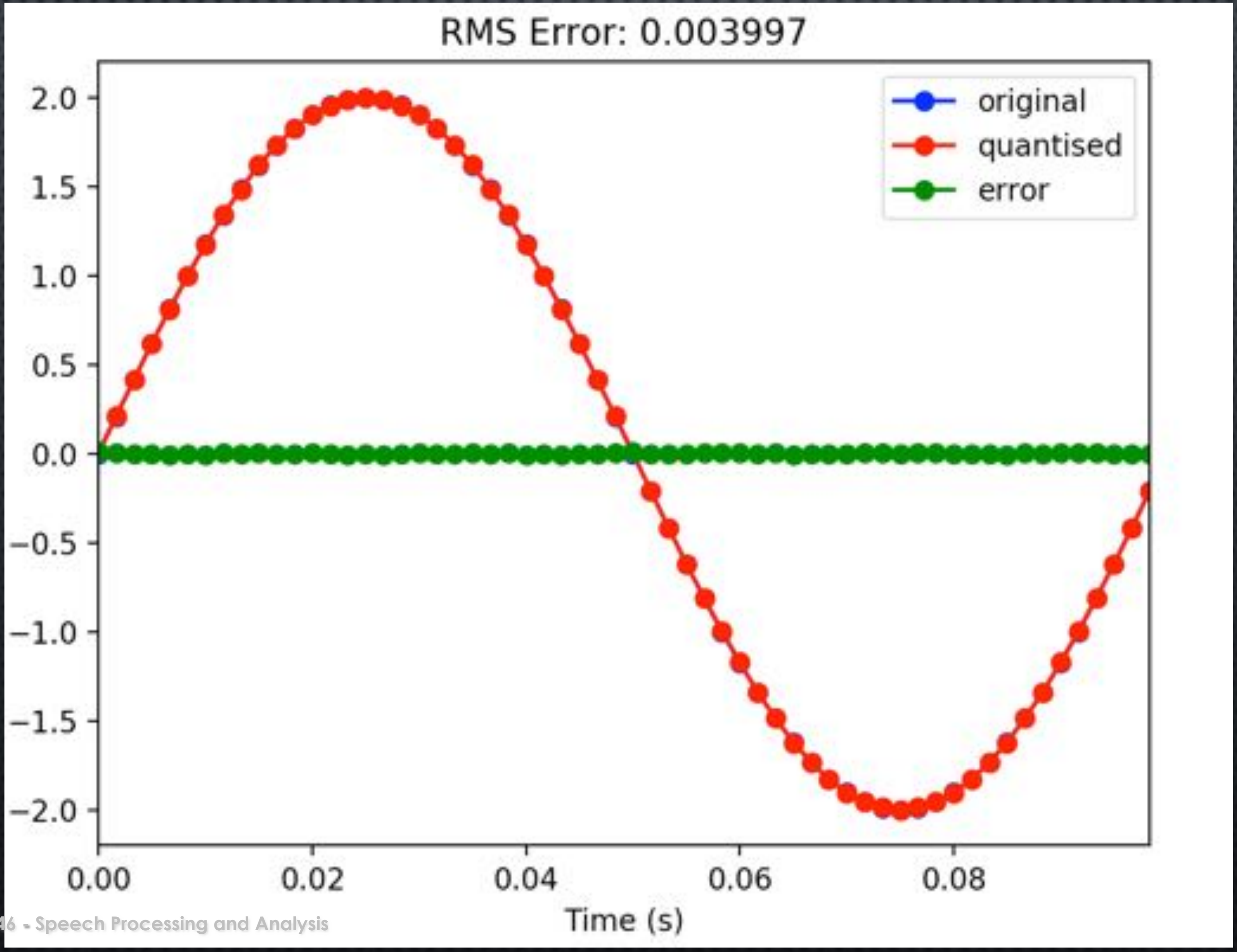


$$RMS_e = \sqrt{\frac{\sum_{n=1}^{N} e_q^2(n)}{N}}$$

ERRORS: 5-BIT 32-LEVEL QUANTISATION



ERRORS: 8-BIT 256-LEVEL QUANTISATION

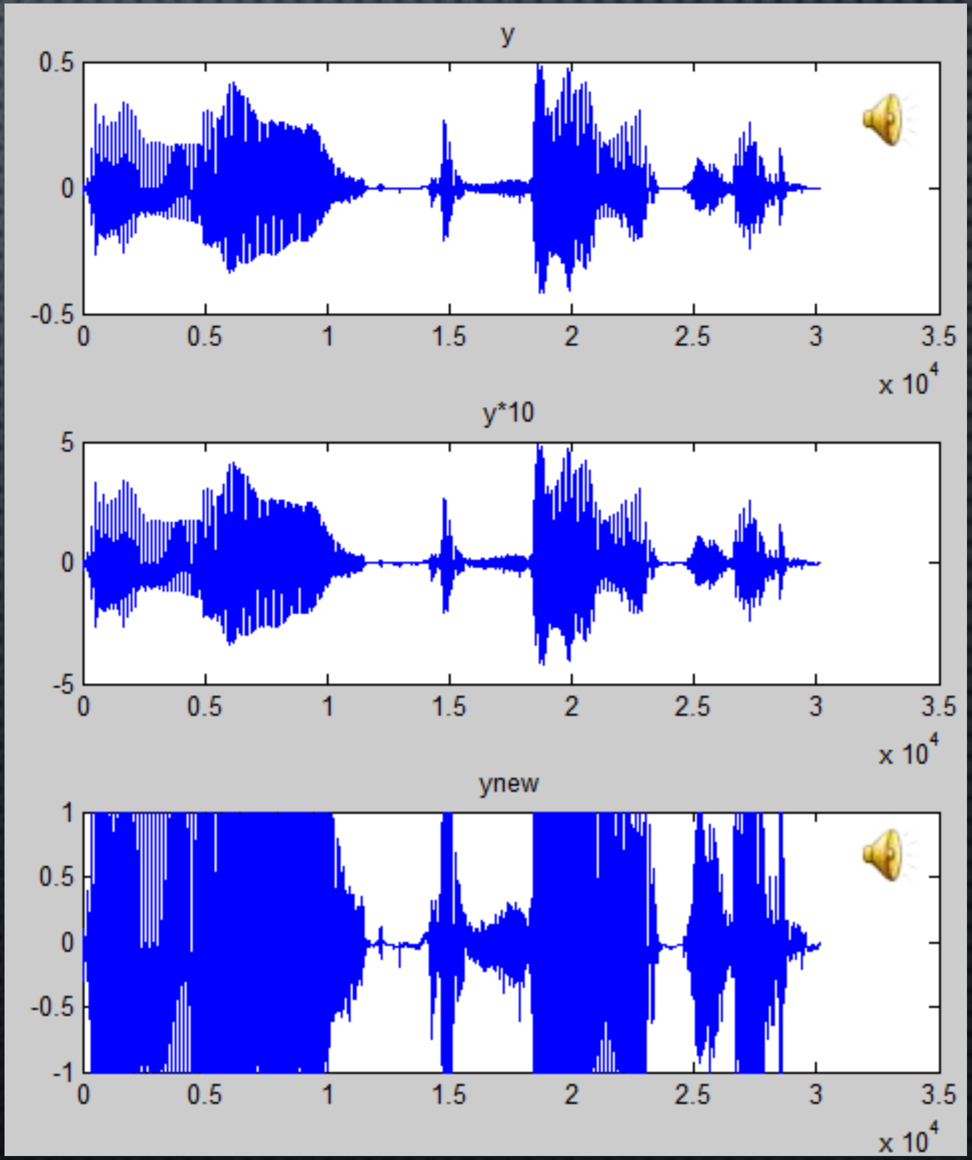


How many steps do we need?

- Make sure that there are a sufficient number of steps so that the difference is inaudible to our ears
  - 16-bit resolution is typcally used: 65536 levels!!!
- In reality, even the best analogue system cannot truly represent an irrational voltage level, as the noise floor will always limit the accuracy of the representation.
- What makes the errors produced by quantisation more challenging to handle is that, it is in fact distortion, e.g. clipping

Clipping in waveform

- Clipping leads to the loss of dynamic range of amplitude
- Reduction of modulation depth in speech

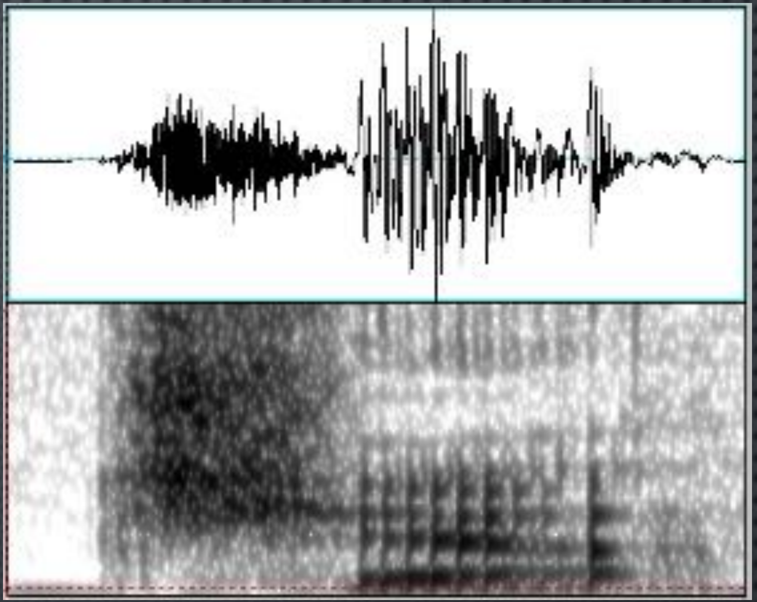


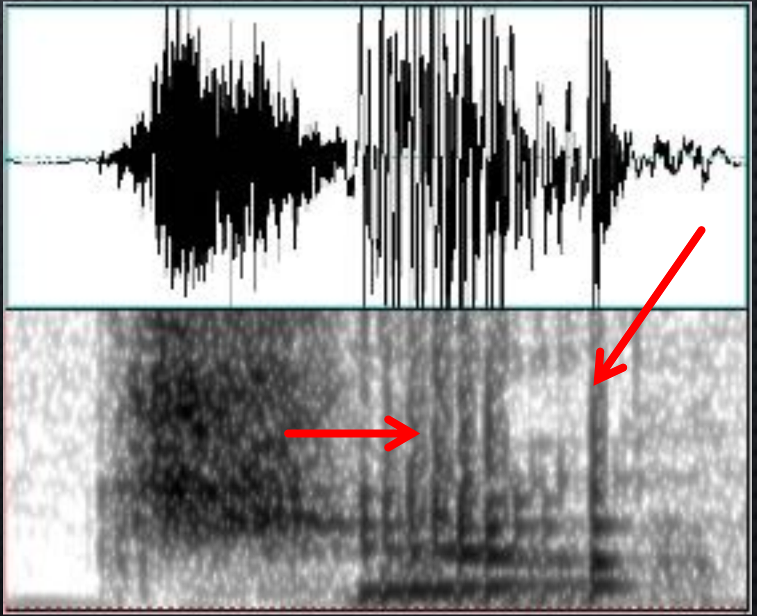
Clipping in spectrum

Clipping produces visible artefacts in the spectrogram, particularly at high frequencies

- Spurious appearance of bursts
- Extra 'vertical' noise can reduce the salience of 'horizontal' structure, disguising evidence of formants

1-kHz puretone:
unclipped vs. clipped





Signal acquisition: from analogue to digital

- Set gain level
  - High enough to make the best use of a limited number of bits
  - Low enough to avoid clipping

- Set sampling frequency high enough to capture information that the human ear can perceive
  - Set it high: you can always downsample – carefully! – later
  - Too high and excessive storage requirements

LING 446

Fundamentals for Speech Signal Processing and Analysis

Yan Tang

Department of Linguistics, UIUC

Week 5: Sound intensity, pressure and loudness

Last week…

- Audio formats and codecs
  - Uncompressed and compressed
  - Lossy and lossless
  - How and why MP3 works
  - Audio file size
- WAV format
  - History and usage
  - Data structure
- Python API for WAV file reading and writing

Sound level

- To quantify how strong (in terms of energy per unit) a sound signal is
- Sound intensity (SI): I
  - *Sound energy* flux in a specific direction and sense through an area perpendicular to that direction, divided by the area.
  - Measured in the direction of the sound wave propagation
- Sound pressure (SP), p
  - *Sound force* applied perpendicular to the surface of an object, divided by the area

Sound intensity, I

- *Power* per unit area (e.g. m^2) carried by a wave

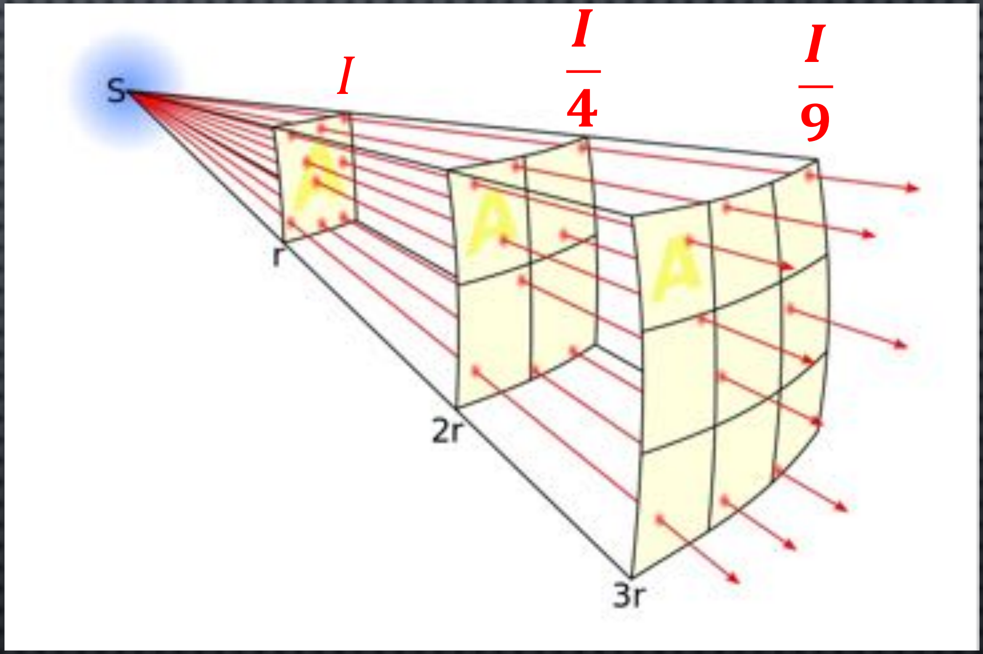
$$I = \frac{P}{A}$$

  - $\boldsymbol{P}$: power (in watts, W: $\frac{kg \cdot m^2}{s^3}$), the rate at which energy is transferred by the wave
  - $\boldsymbol{A}$: area in m^2

- Unit: W/m^2
  - Human ears can detect sound intensities as low as 0.000000000001 W/m^2 and up to 50 W/m^2 or more

Inverse-square law

- An “ideal” acoustic space: free field
  - free of reflection
- Sound intensity is inversely proportional to distance squared



$$I = \frac{P}{A} = \frac{P}{4\pi r^2} \quad \therefore I \propto \frac{1}{r^2}$$

Image source: https://en.wikipedia.org/wiki/Inverse-square_law

Sound pressure, p

- *Force* per unite area (e.g. m^2) carried by a wave

$$p = \frac{F}{A}$$

  - $\boldsymbol{F}$: force (in Newtons, N: $\frac{kg \cdot m}{s^2}$), interaction between the wave and ambience that cause the wave to accelerate/decelerate.
  - $\boldsymbol{A}$: area in m^2

- Unit: N/m^2 or *Pascal, Pa*: $\frac{kg}{m \cdot s^2}$
  - Human ears can deal with sound pressures of 0.00002 pascals up to 200 pascals

Sound intensity level (SIL) and sound pressure level (SPL)

- The range of sound level the auditory system can sense is extraordinarily wide!
  - 10^{-12} to 10^{2} for sound intensity
  - 10^{-5} to 10^{2} for sound pressure
- Idea of *decibels*: express *SIL* and *SPL* on a logarithmic scale, as a ratio of comparing to a reference

$$decibels = 10 \cdot log_{10} \frac{I}{I_{ref}} = 10 \cdot log_{10}(\frac{p}{p_{ref}})^2 = 20 \cdot log_{10}(\frac{p}{p_{ref}})$$

I_{ref}: $10^{-12}\, W/m^2$

p_{ref}: $2 \times 10^{-5}\, N/m^2$

$$decibels = 10 \cdot log_{10}\frac{I}{I_{ref}} = 10 \cdot log_{10}(\frac{p}{p_{ref}})^2 = 20 \cdot log_{10}(\frac{p}{p_{ref}})$$

I_{ref}: $10^{-12}\ W/m^2$

p_{ref}: $2 \times 10^{-5}\ N/m^2$

Decibels

- If the intensity of a sound, I, is $10^{-10}\ W/m^2$, what is the SIL level?
- If the pressure of a sound, p, is $2 \times 10^{-4}\ N/m^2$, what is the SPL level?
- What would happen if the I (or p) were the same as the I_{ref} (or p_{ref})?
  - What is the implication?

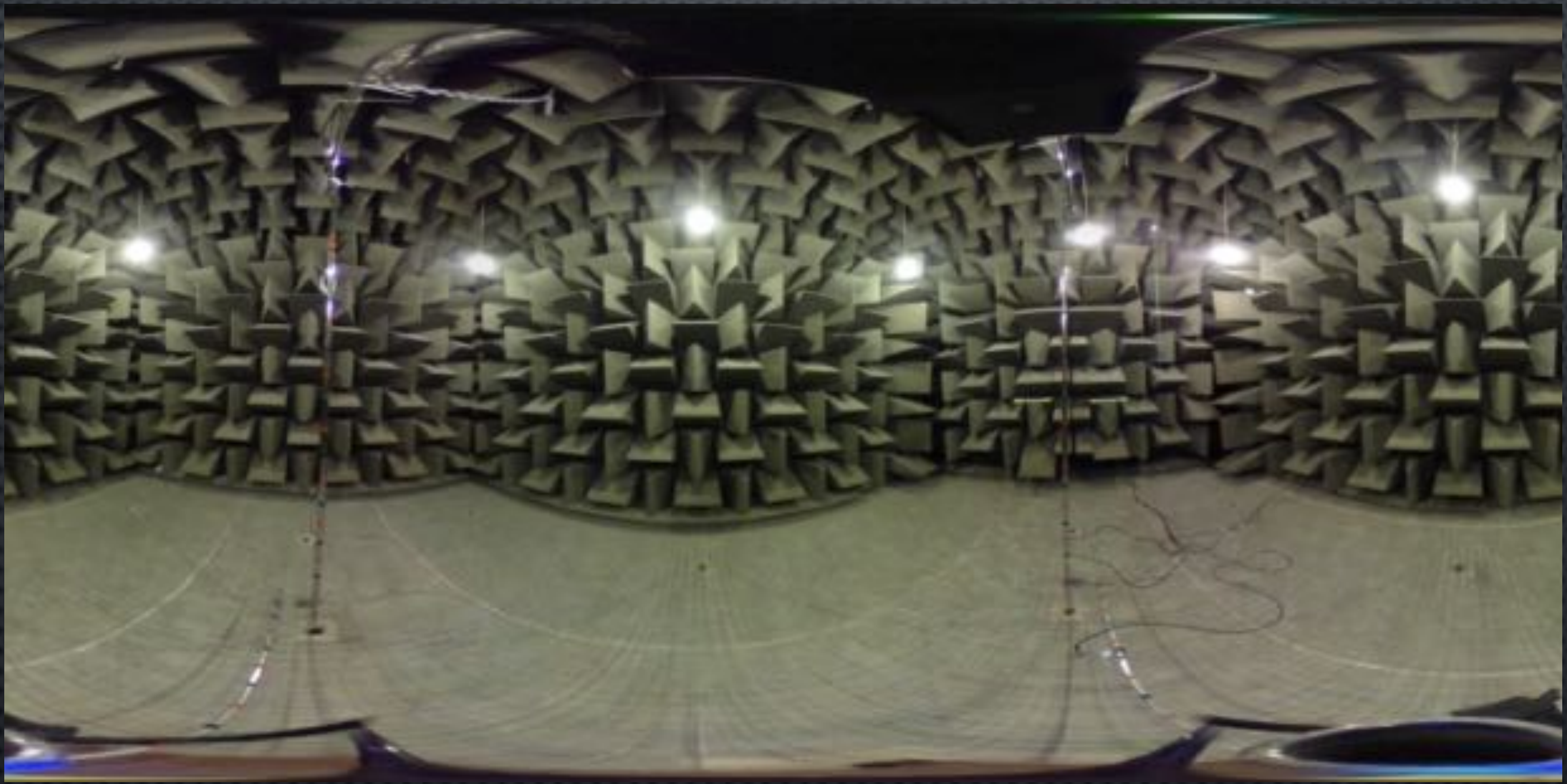
SPL in common scenarios

Sound sources (noise) Examples with distance	Sound pressure Level L_p dB SPL
Jet aircraft, 50 m away	140
Threshold of pain	130
Threshold of discomfort	120
Chainsaw, 1 m distance	110
Disco, 1 m from speaker	100
Diesel truck, 10 m away	90
Kerbside of busy road, 5 m	80
Vacuum cleaner, distance 1 m	70
Conversational speech, 1 m	60
Average home	50
Quiet library	40
Quiet bedroom at night	30
Background in TV studio	20
Rustling leaves in the distance	10
Hearing threshold	0

Sound pressure p N/m^2 = Pa Sound field quantity	Sound intensity I W/m^2 Sound energy quantity
200	100
63.2	10
20	1
6.3	0.1
2	0.01
0.63	0.001
0.2	0.000 1
0.063	0.000 01
0.02	0.000 001
0.006 3	0.000 000 1
0.002	0.000 000 01
0.000 63	0.000 000 001
0.000 2	0.000 000 000 1
0.000 063	0.000 000 000 01
0.000 02	0.000 000 000 001

Image srouce:http://www.sengpielaudio.com/TableOfSoundPressureLevels.htm

SPL in common scenarios



- Anechoic chamber at the Acoustics Research Centre, University of Salford, UK
- Background noise level: -12.4 dBA

The quietest place on the earth!

- Anechoic chamber at Microsoft Research.
- Background noise level: -20.6 dBA

The relationship between I and p

$$I = \frac{p^2}{\rho \cdot c} \quad \therefore I \propto p^2$$

- $\boldsymbol{\rho}$: density of air, kg/m^3
- $\boldsymbol{c}$: sound speed in air, m/s

- Acoustic impedance:
  - $(\rho \cdot c)$, $pa \cdot s/m$
  - Both ρ are c variables to temperature
  - When T= 20 ℃, ≈ 413.3 $pa \cdot s/m$

ρ and c as functions of temperature

Effect of temperature on properties of air

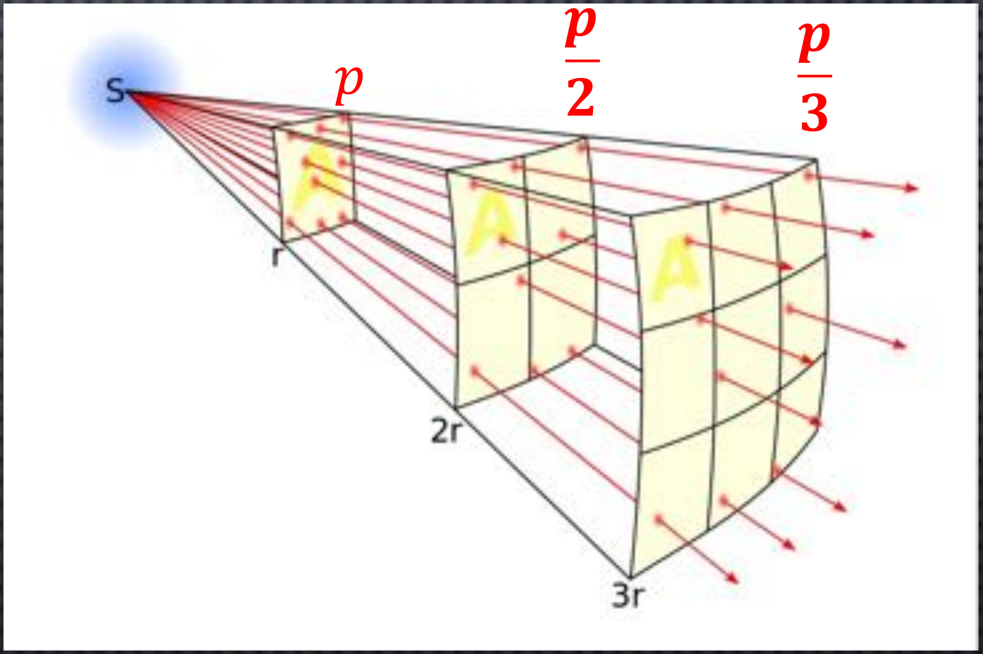
Temperature T (°C)	Speed of sound c (m/s)	Density of air ρ (kg/m^3)	Characteristic specific acoustic impedance z_0 (Pa·s/m)
35	351.88	1.1455	403.2
30	349.02	1.1644	406.5
25	346.13	1.1839	409.4
20	343.21	1.2041	413.3
15	340.27	1.2250	416.9
10	337.31	1.2466	420.5
5	334.32	1.2690	424.3
0	331.30	1.2922	428.0
−5	328.25	1.3163	432.1
−10	325.18	1.3413	436.1
−15	322.07	1.3673	440.3
−20	318.94	1.3943	444.6
−25	315.77	1.4224	449.1

https://en.wikipedia.org/wiki/Speed_of_sound

Validation of the relationship between I and p

Inverse-distance law

- Pressure is proportional to the square root of the intensity
- Sound pressure is inversely proportional to the distance



$$p = \frac{\sqrt{P\rho c\pi^{-1}}}{2r} \quad \therefore p \propto \frac{1}{r}$$

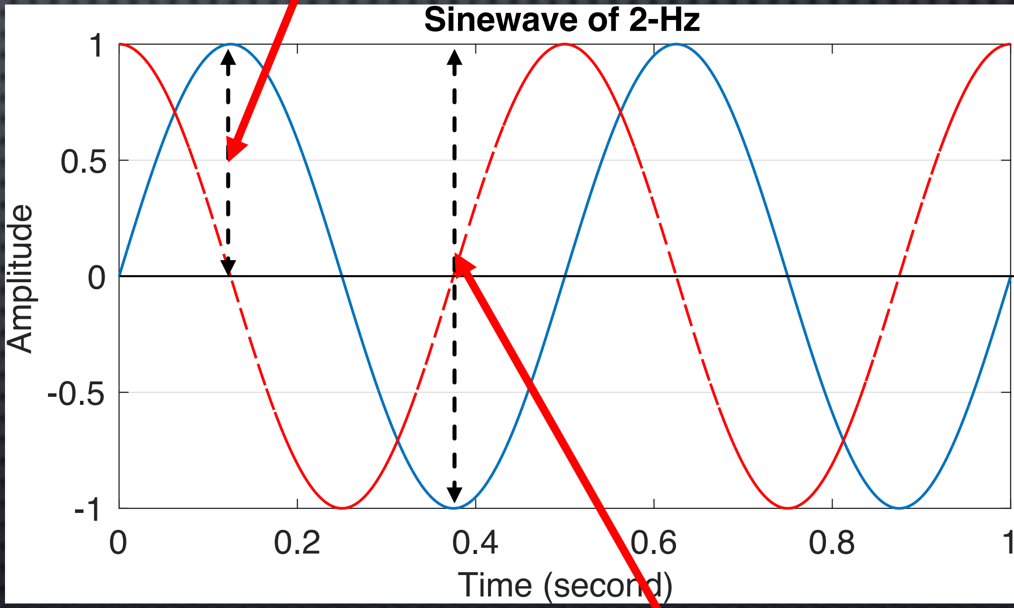
$$I = \frac{P}{A} = \frac{P}{4\pi r^2} \qquad I = \frac{p^2}{\rho \cdot c}$$

Image source: https://en.wikipedia.org/wiki/Inverse-square_law

Theoretical value of SPL

- What does the amplitude of a signal represent on a computer?
  - Assumption: the instant *pressure* in pascal at that time or N/m^2
  - Pressure variation or Amplitude pressure
    - $A_P = \frac{A_{p2p}}{2}$

peak amplitude



Peak-to-peak amplitude

Theoretical value of SPL

- Root-mean-square (RMS) amplitude

$$A_{rms} = \sqrt{\frac{\sum_{i=1}^{N} s_i^2}{N}} \approx 0.707A_p \approx 0.354A_{p2p}$$

N: number of samples

s: amplitude

1-ms 10-ms 50-ms

- Sound perception is a process of energy accumulation or aggregation
- RMS pressure, p_{rms} *(i.e. A_{rms}),* is measured over a certain duration, instead of instant pressure

$$SPL = 20 \cdot log_{10}(\frac{p_{rms}}{p_{ref}})$$

$$SPL = 20 \cdot log_{10}(\frac{p_{rms}}{p_{ref}})$$

Considering the following scenario:

- Donny is going to listen to music in WAV format
- He computes the average SPL of the sound in the WAV file, which turns out to be 73 dB SPL
- He is then listening to the WAV over a pair of headphones

Q: *what's the average SPL of the music Donny actually hears?*

Q: *what'll the SPL of the music Donny hears be when it is played back via a loudspeaker 2 metres away from Donny?*

Theoretical SPL vs practical SPL

- Assumes that amplitude represents sound pressure measured at the listener’s position
  - Several factors may affect the sound level when reaching your ear, e.g. hardware, distance.
- In real situations, SPLs are measured using an SPL meter.

Digital signal level vs SPL

- Signal level can also be quantified using the notion of dB
  - We are interested in the relative level of a signal compared to other signals
  - The intensity of signal s can be computed as an absolute value in dB, e.g.

$$I_{dB} = 10 \cdot log_{10} A_{rms}^2$$

- Therefore, 60 dB is not the same as 60 dB SPL!!!

Perception of sound level

- "Sound with low intensities are perceived as '*soft*' and high intensities as '*loud*'"

- Which of the following sound is the loudest?

100 Hz

250 Hz

1k Hz

10k Hz

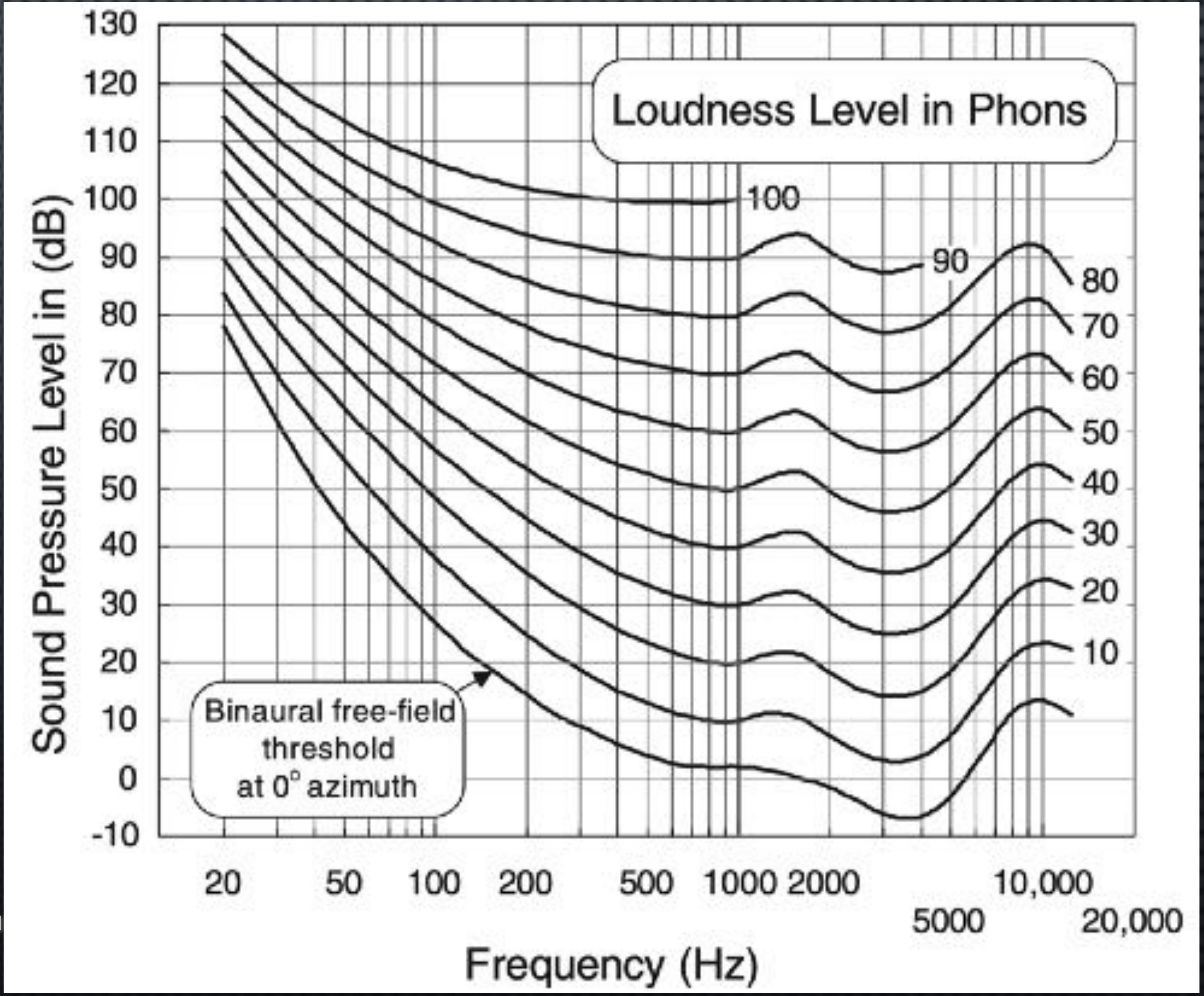
  - Sound signals at same sound level (e.g. rms=0.1) may be perceived differently loud!

Loudness

- The perception of sound level is called loudness

- Units:
  - ***Phon*** vs dB SPL
    - Referenced to dB SPL at *1k* Hz.
  - ***Sone*** L_s vs Phon L_p
    - $L_p = 40 + 10log_2(L_s)$ or $L_s = 2^{\frac{L_p - 40}{10}}$

Normal equal-loudness-level contours for pure tones



BS ISO 226:2003
ANSI S3.6-2004
ISO 389-7-2005

Loudness (≠, !=, ~=, ne) sound level!!!

- Objective measurement VS Subjective perception
  - Sound level is a physical parameter of sound
  - Loudness is percept associated with the physical aspect
  - Loudness may vary across the population
  - Sound level holds constant
- Relationship
  - Increasing level is associated with increasing loudness
  - *Not simple one-to-one correspondence*
- Other factors
  - Frequency range (as demonstrated)

LING 446

Fundamentals for Speech Signal Processing and Analysis

Yan Tang

Department of Linguistics, UIUC

Week 6: LTI systems and their properties

Last week…

- Sound intensity level (SIL) and sound pressure level (SPL)
  - Definition of decibel (dB)
  - $SIL(dB) = 10 \cdot log_{10} \frac{I}{I_{ref}}$
  - $SPL(dB) = 10 \cdot log_{10}(\frac{p}{p_{ref}})^2 = 20 \cdot log_{10}(\frac{p}{p_{ref}})$

I_{ref}: $10^{-12}\ W/m^2$
p_{ref}: $2 \times 10^{-5}\ N/m^2$

- Sound level: intensity I and pressure p
  - $I = \frac{p^2}{\rho \cdot c}$
- Theoretical vs practical SPL
- Loudness
  - Subjective perception vs objective measurement

Intro to linear time-invariant system

Definition of system

- What is a system?
  - Something which performs some operations on, or transformation of, an ***input*** signal to produce an ***output*** signal
  - Signal processing (SP): a component or pipeline which processes signals to create other signals
- Can you think of any examples in audio SP?
  - Amplifier, filters, artificial reverberation generator, etc

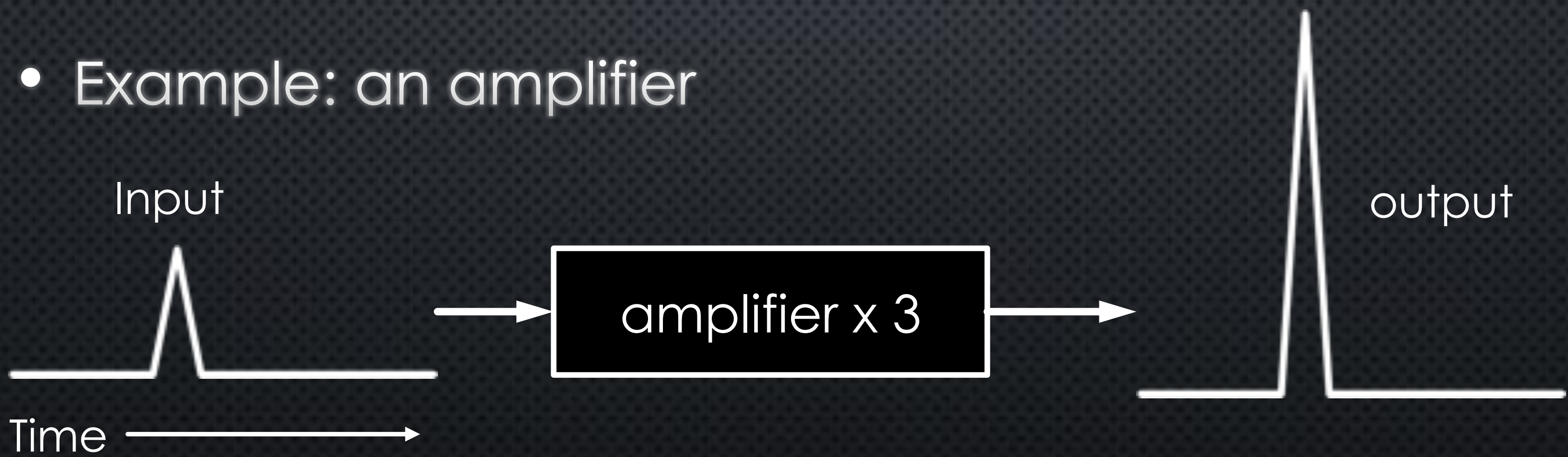
Some examples of systems

- Telephone network
- VoIP, e.g. Skype
- An analogue or digital filter
- A guitar effects pedal
- Radar
- An image processing algorithm
- The economy
- A car steering system
- etc……..

System diagrams

- A system operating on a signal can be presented as,

- Example: an amplifier

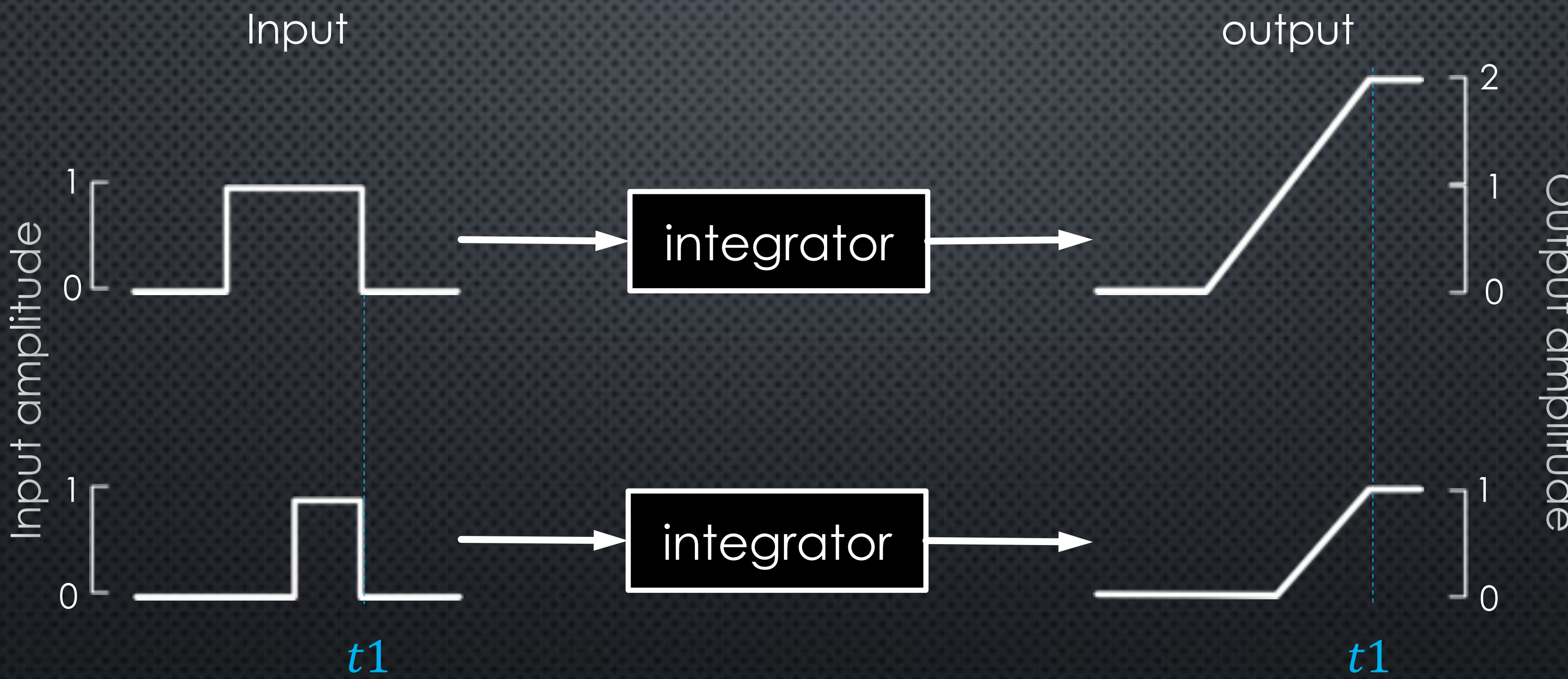


$$output(t) = 3 \times input(t)$$

“memory” of a system

- If a system has no “memory”:
  - It only needs to know its input at one time, t, to produce the corresponding output for t
  - Does not store any other information
  - E.g. chain and parallel systems
- A system with “memory”:
  - Its output at t depends on the input signal or/and output at one or more time moments in the past
  - $t-1, t-2, \ldots, t-n, n$ is positive integer
  - E.g. Feedback loop.

A system with "memory"



Another example in discrete time: $output(t) = \ input(t) + output(t-1)$

Most of systems in DSP have "memory"!

Causality of a system

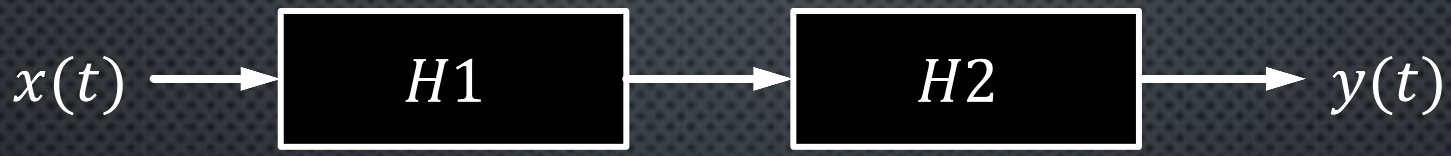
- What if the output of a system depends on the *future* of the input? i.e. $t + 1, t + 2$ and $t + n, n$ is positive integer
  - This system is not causal

- A system is *causal* if the output at time n only depends on the input or output up to time n

- Is every real world system casual?

Connecting systems

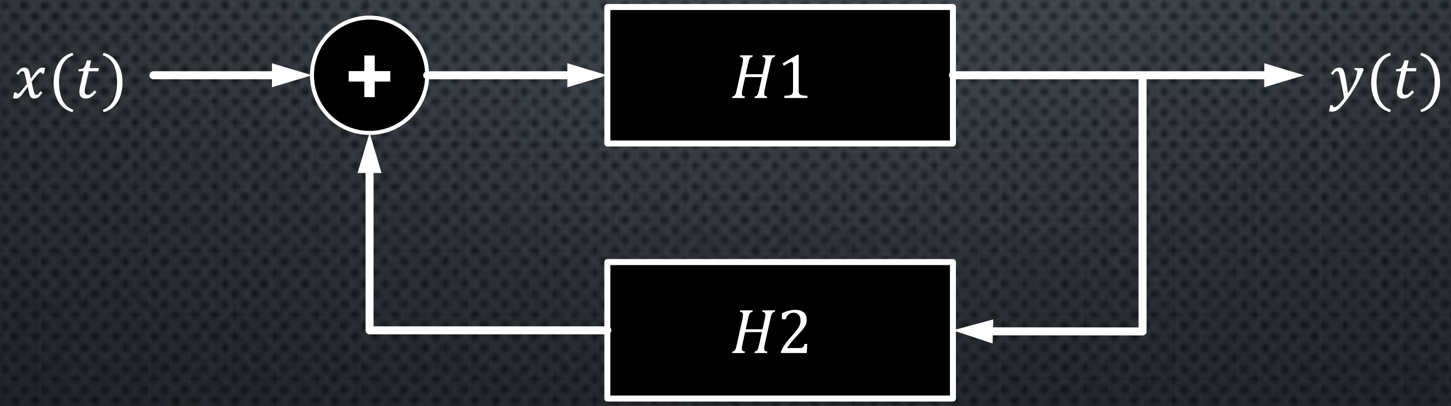
- Chain systems:



- Parallel system:

Connecting systems

- Feedback loop:



Linear time-invariant (LTI) system

- In DSP, we are concerned with systems that transform one digital signal into another, in particular with those that perform ***linear*** processing that does not change with time (***time-invariant***)
  - A good approximation of real word system

- Linearity means that we can understand the output of a system in **response to a complex input** (like speech) in terms of the **sum of simpler components**
  - It makes analysis easier

Properties of LTI system

- The properties of LTI system:
  - **Linearity** = *homogeneity* + *additivity*
  - **Time invariance**

Linearity: homogeneity

Homogeneity implies a proportionality between the level of the input and the level of the output

$$x(t) \longrightarrow \boxed{H} \longrightarrow y(t)$$

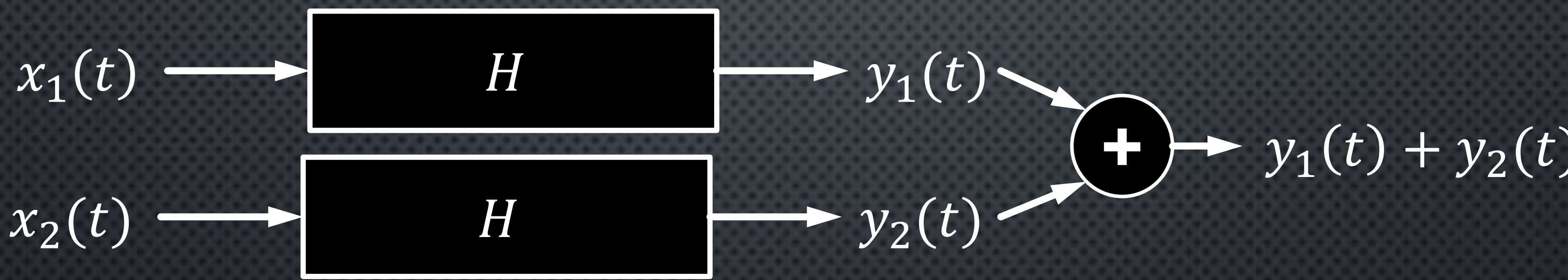
$$k \cdot x(t) \longrightarrow \boxed{H} \longrightarrow k \cdot y(t)$$

Is the following system homogenous?

- $y(t) = x(t) + 3x(t-1)$

Linearity: additivity

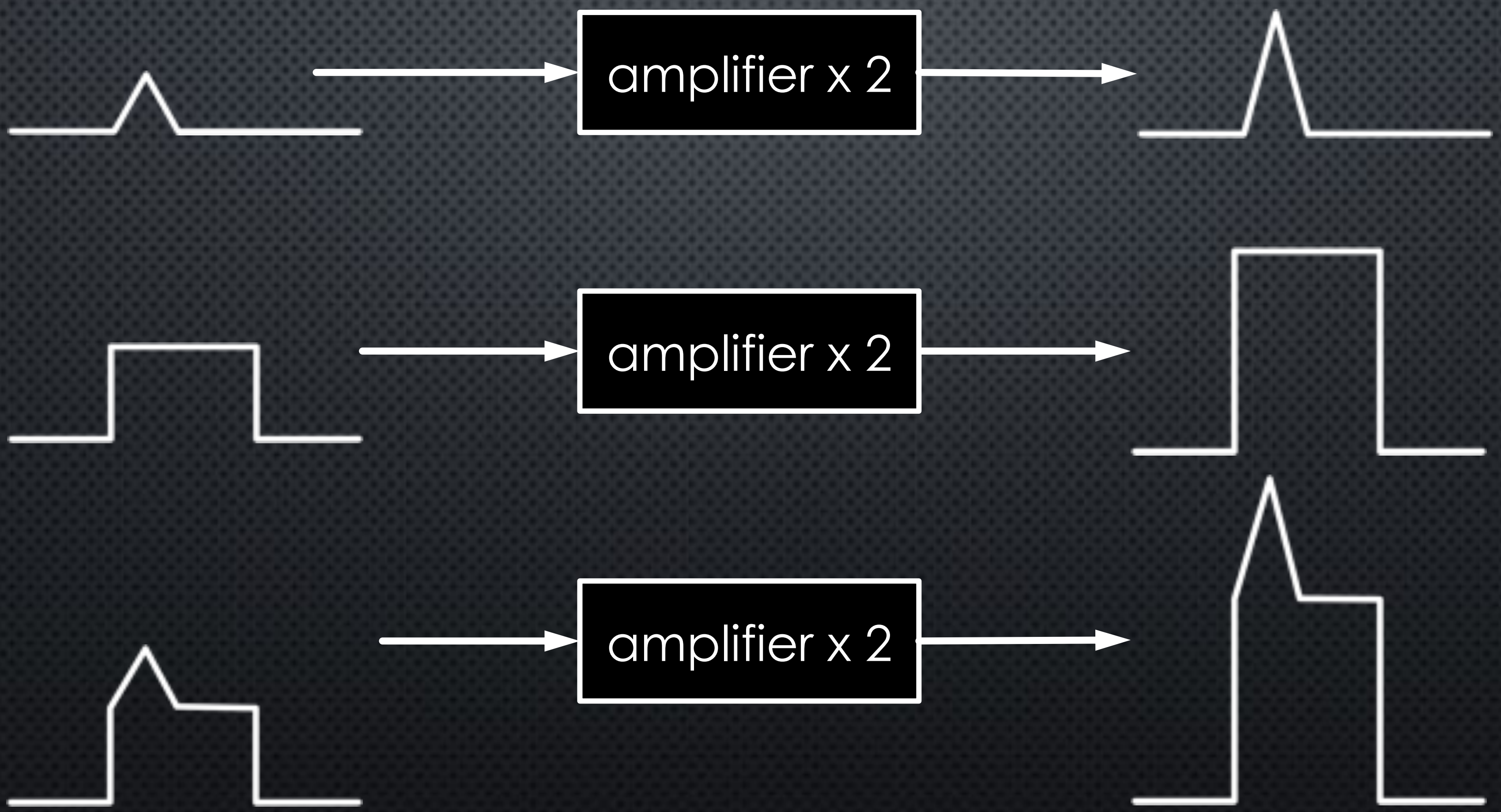
Given the output to each individual input, if the system is additive, the output to the sum of all the inputs is the sum of the separate outputs



Is the following system additive?

- $y(t) = x(t) + 3x(t-1)$

Linearity: additivity



Question

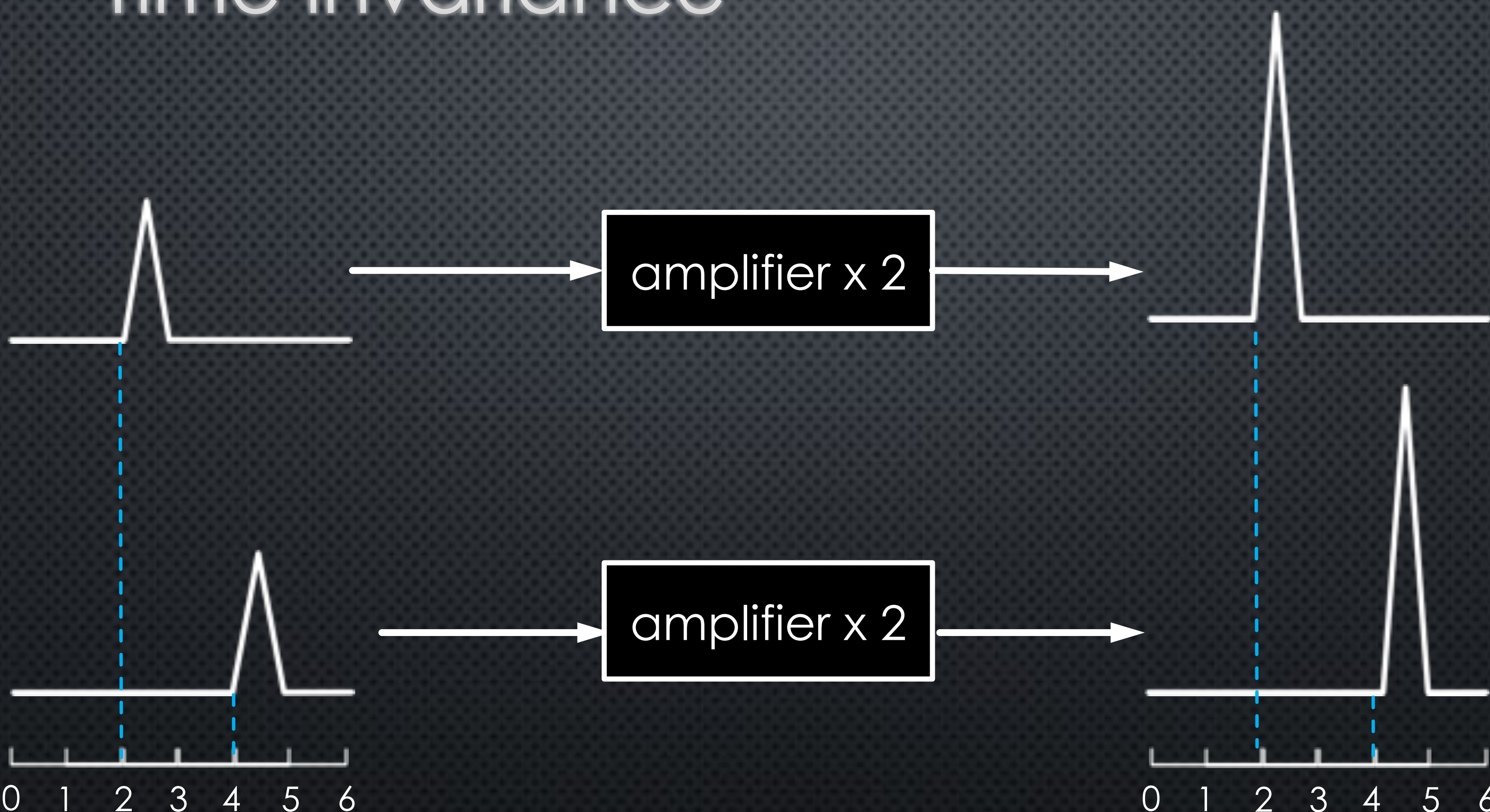
- Is this system linear or non-linear?

$$y(t) = 3x(t) + 5$$

Time invariance

The system behaves the same way, regardless of when the input is applied

Time invariance



Question

- Are these systems time-invariant?

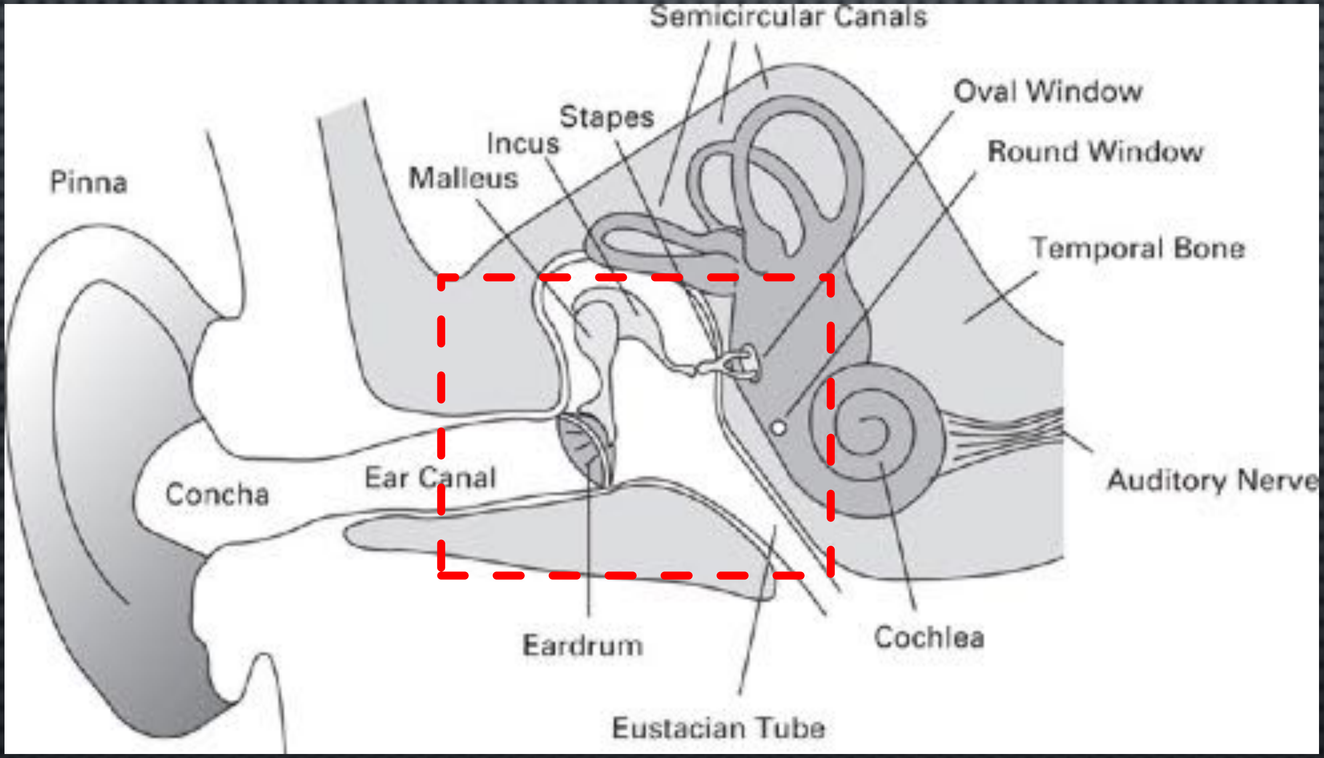
$$y(t) = 3x(t) + 5$$
$$y(t) = \ x(t) + 3x(t-1)$$

Question

- is the following an LTI system, where $x(n)$ is a discrete signal?

$$y(n) = \cos(2\pi n)\, x(n)$$

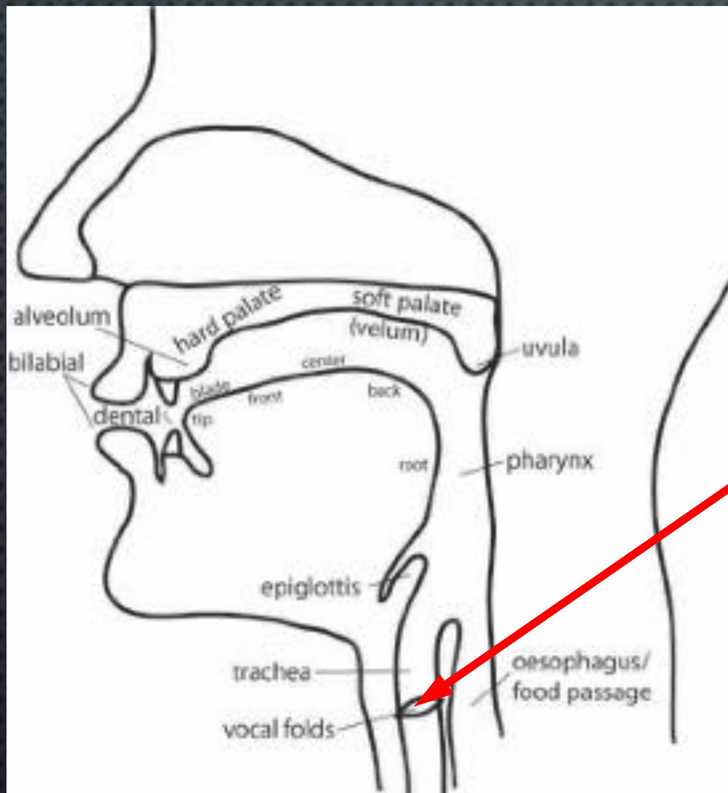
Linear system in real world



system

Sound press at eardrum —input→ **Eardrum and ossicles** —output→ Movement of stapes

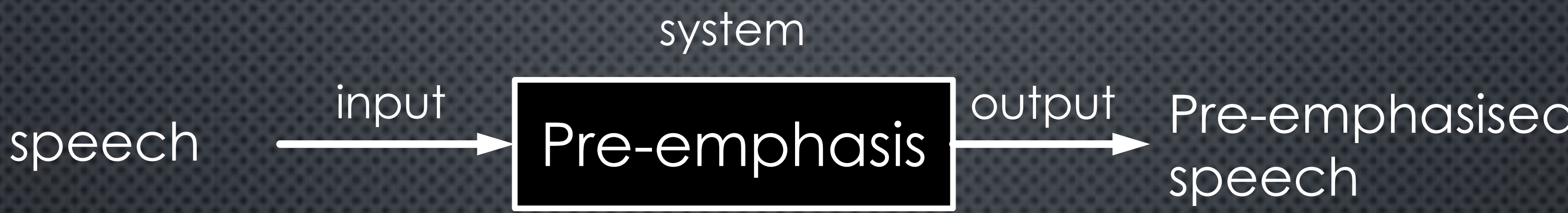
Time-variant system in real world



system

vocal cords vibration → input → **vocal tract and larynx activities** → output → sound radiating from the lips

LTI in DSP: pre-emphasis



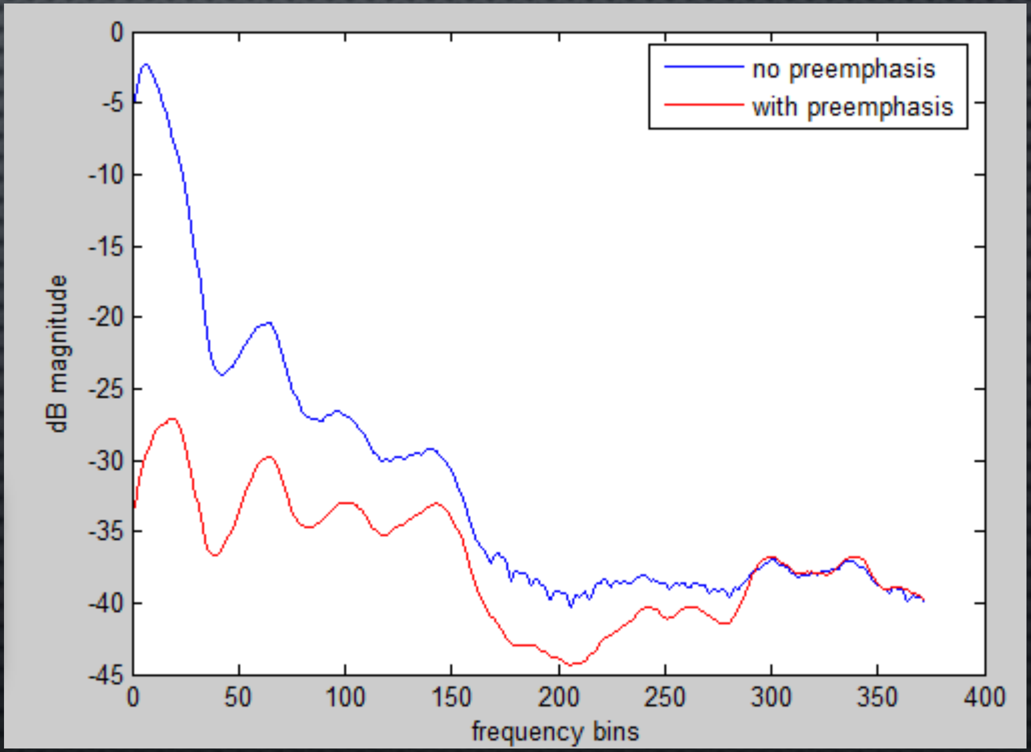
$$y(t) = x(t) - \alpha * x(t-1)$$

i.e. output y at time t is formed by taking the current input $x(t)$ and subtracting 'a' times the previous input, $x(t-1)$

```
x=[1 2 5 4 7];
preemphasis(x,0.98)
    1.0000    1.0200    3.0400   -0.9000    3.0800
```

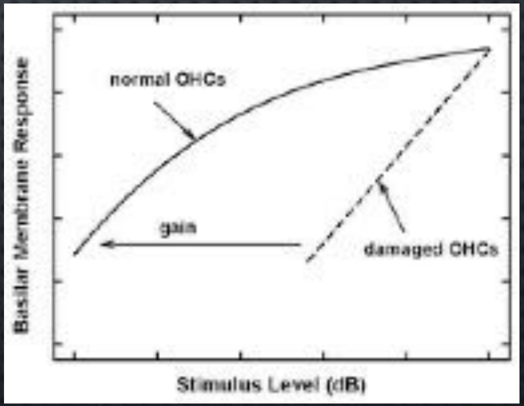
What does pre-emphasis do?

- The effect of changing spectral tilt to boost the high frequencies in the spectrum
- Pre-emphasis is an example of a **filter**
- Pre-processing of ASR



Relationship between linear and non-linear systems

- LTI system is an ideal model of a subset of systems; it is an approximation to real situations
- Techniques developed for linear systems can help understanding non-linear systems, e.g. filter
- A linear system may only be *linear* over a limited range of input, e.g. auditory system
  - Some parts of a non-linear system can be analysed and modelled as a linear system
- A time-variant system may be treated as a TI system in a short period of time, e.g. vocal tract



The non-linearity of outer hair cells in the cochlear (Bacon, 2002)