CS 6351 DATA COMPRESSION

THIS LECTURE:
TRANSFORMS PART III

Instructor: Abdou Youssef

OBJECTIVES OF THIS LECTURE

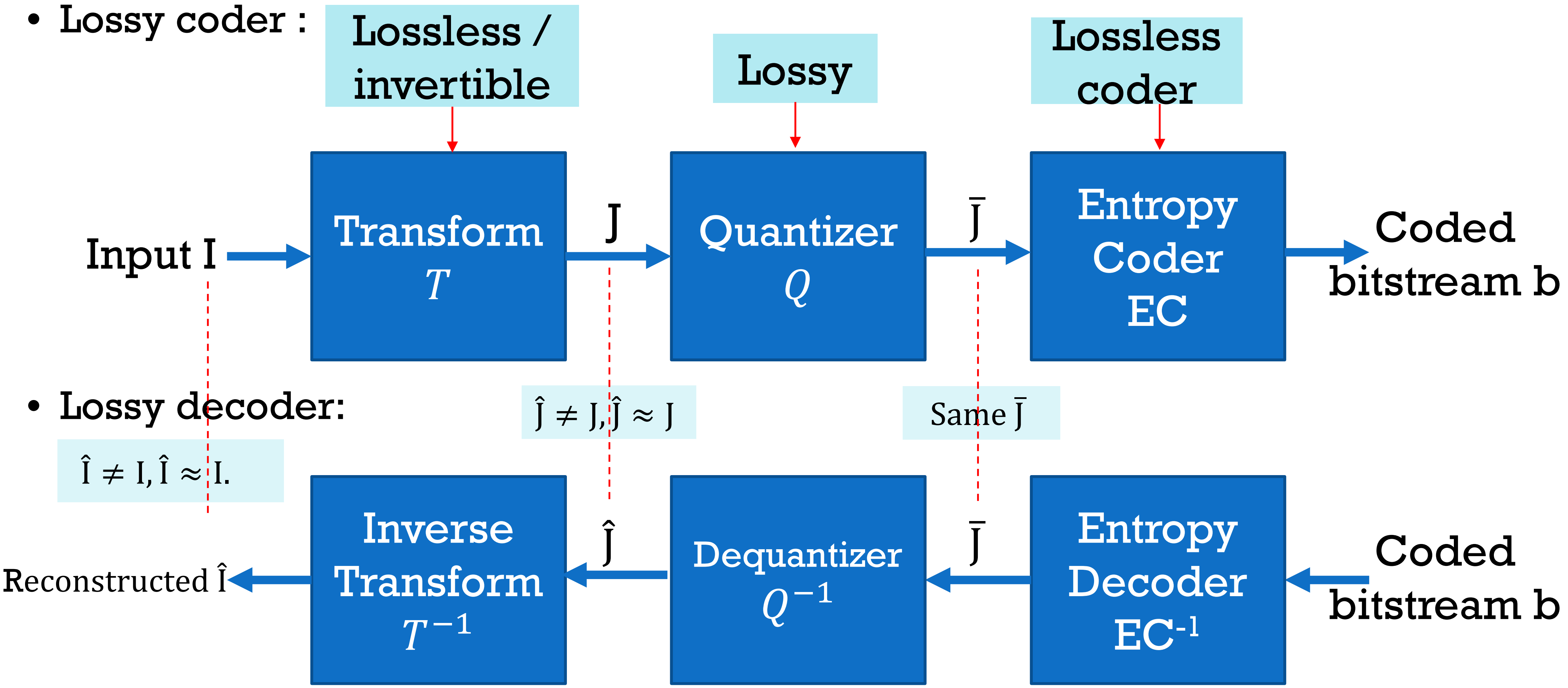
By the end of this lecture, you will be able to:

- Approach the analysis and understanding of transforms from the frequency perspective
- Define and explain frequencies in natural signals, and compute and analyze the frequencies using Fourier series and Fourier transform
- Relate frequencies to **audio-visual** human senses, especially the varying sensitivity of humans to different frequencies
- Connect Fourier series with data compression and vector spaces
- Derive the Discrete Fourier Transform (DFT) of digital signals from the "Analog" Fourier series/transform
- Identify the strengths and limitations of the Fourier transform with respect to data compression
- Derive the Discrete Cosine Transform (DCT) from DFT, compare and contrast the two, and argue the reasons why DCT is the best among the transforms studied so far

OUTLINE

- Periodic functions and frequencies
- Fourier series and Fourier transform of analog signals
- Sensitivity of human senses to different frequencies
- Relation between Fourier series/transform/frequencies and lossy compression and vector spaces
- DCT in comparison to DFT
- DCT in comparison with all the transforms we have studied so far

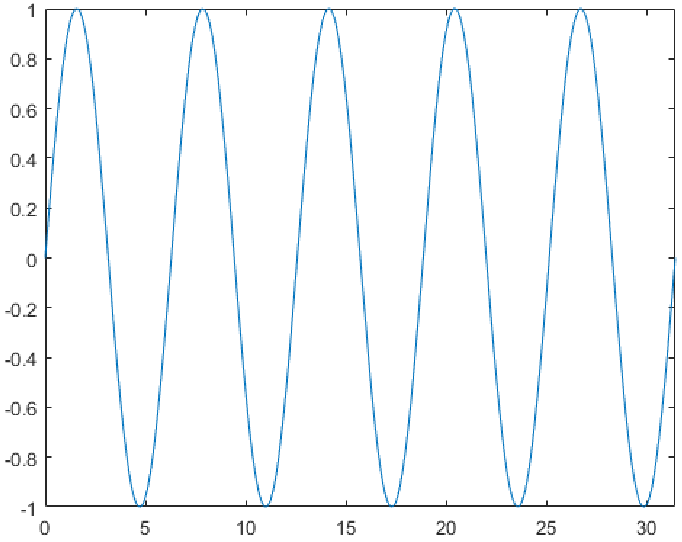
RECALL: GENERAL SCHEME OF LOSSY COMPRESSION

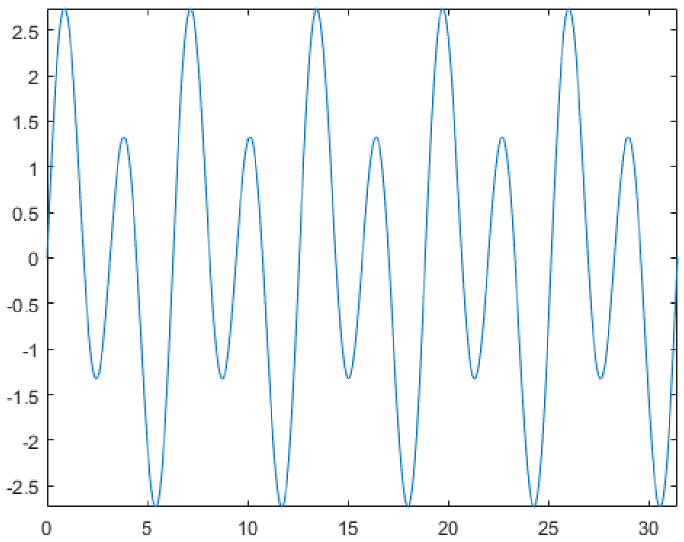


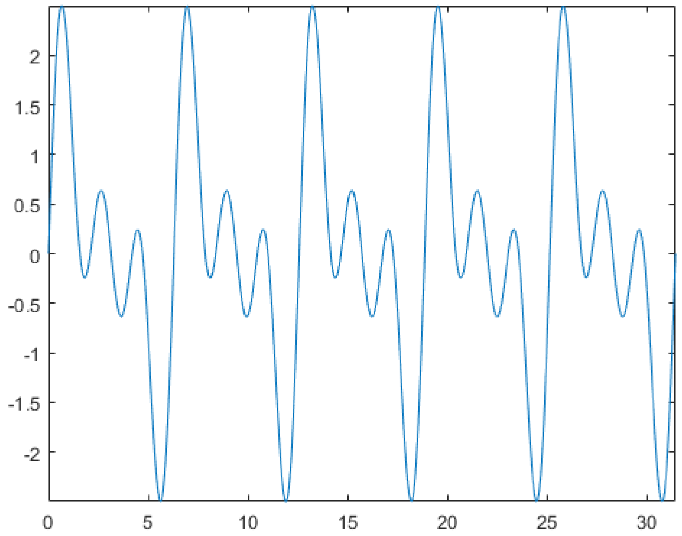
The entire data loss is limited to the quantizer

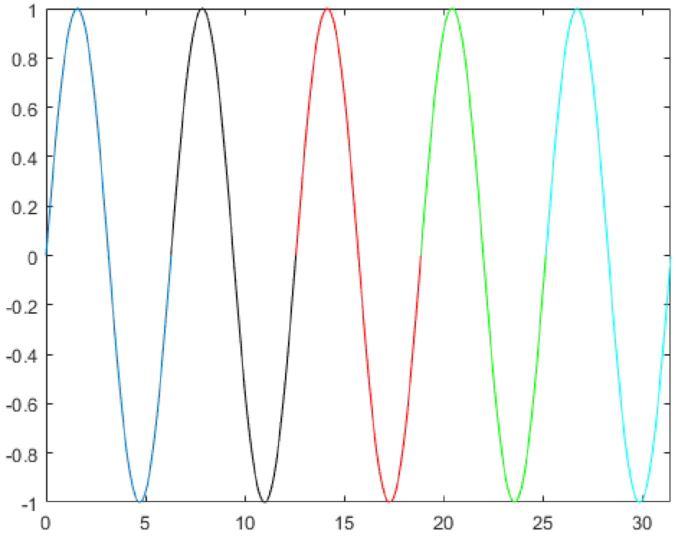
PERIODIC FUNCTIONS
-- INFORMAL DEFINITION --

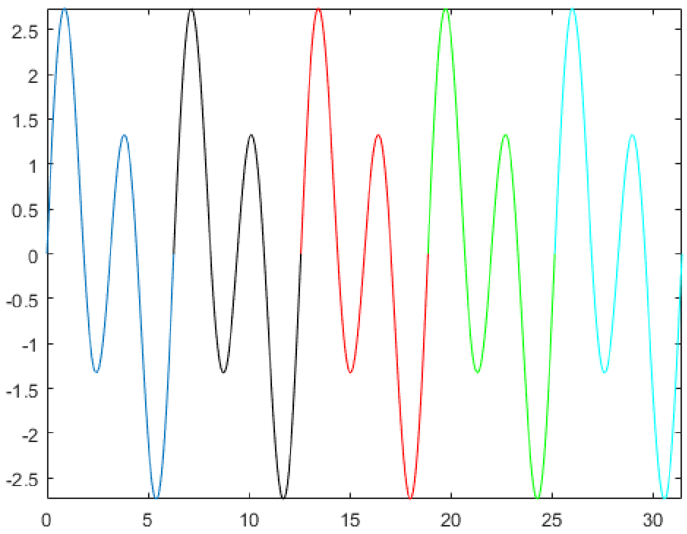
- A function (or signal) is periodic if it is made of perfectly repeating patterns

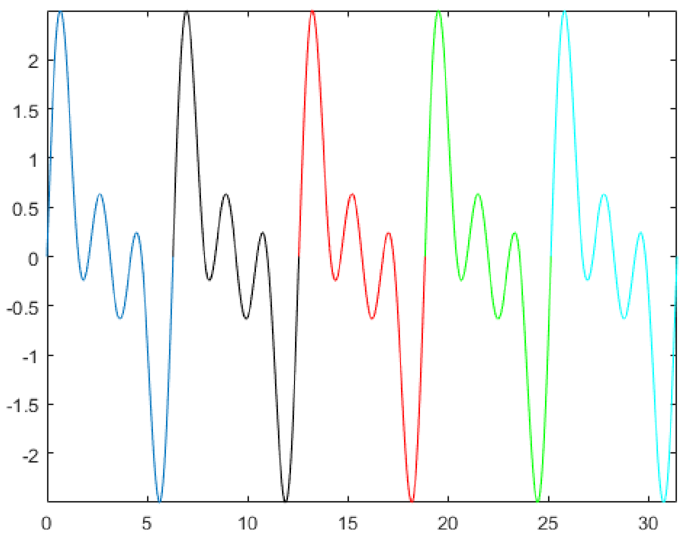










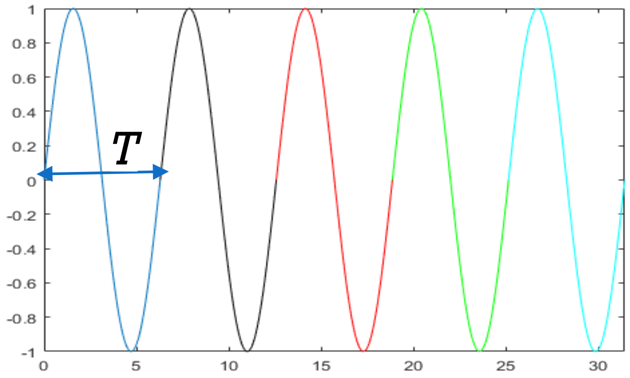


PERIODIC FUNCTIONS
-- FORMAL DEFINITION AND EXAMPLES --

- **Periodic functions**: Formally, a function $g(t)$ is called ***periodic*** if there exists a positive number T such that $g(t+T) = g(t) \quad \forall\, t$
- **Period**: The ***period*** of a periodic function $g(t)$ is the **smallest** positive number T such that $g(t+T) = g(t) \quad \forall\, t$.

> - The period T is the **width** of the shortest repeating pattern
> - The repeating pattern is the curve of $g(t)$ for $0 \le t \le T$

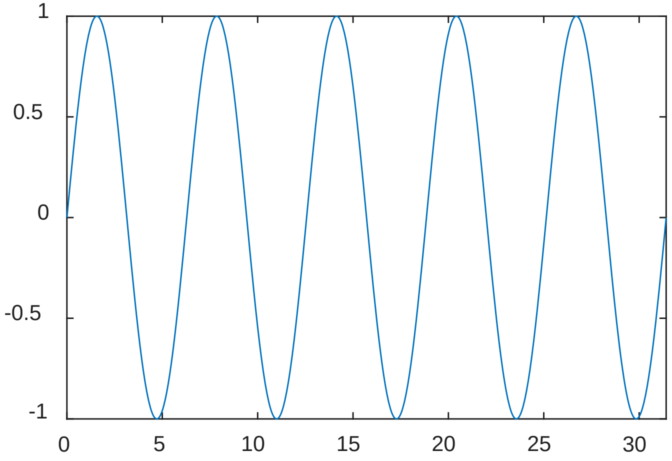
- Examples:
  - $g(t) = \sin t$ is periodic of period 2π because $\sin(t+2\pi) = \sin t \quad \forall\, t$
  - Similarly, $g(t) = \cos t$ is periodic of period 2π
  - $g(t) = \sin t + 4\cos t$, period = 2π
  - $g(t) = \sin(2t)$ is periodic of periodic π because
    - $g(t+\pi) = \sin\big(2(t+\pi)\big) = \sin(2t+2\pi) = \sin(2t) = g(t)$



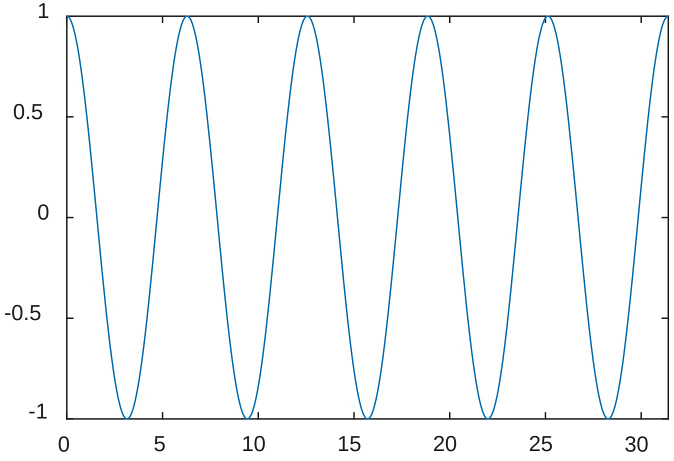
PERIODIC FUNCTIONS

-- EXAMPLES IN PLOTS--

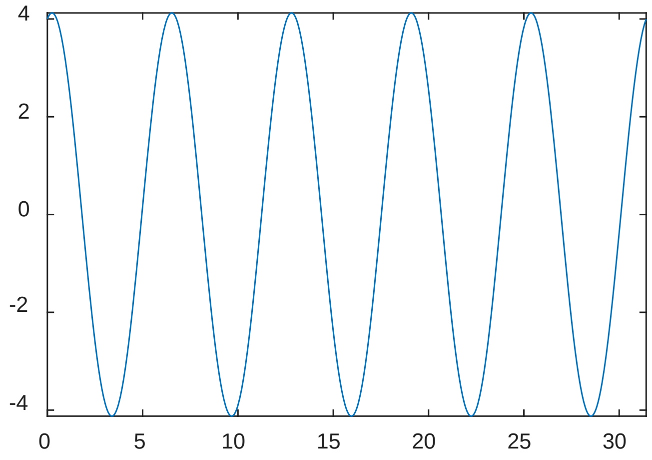
sin(t)



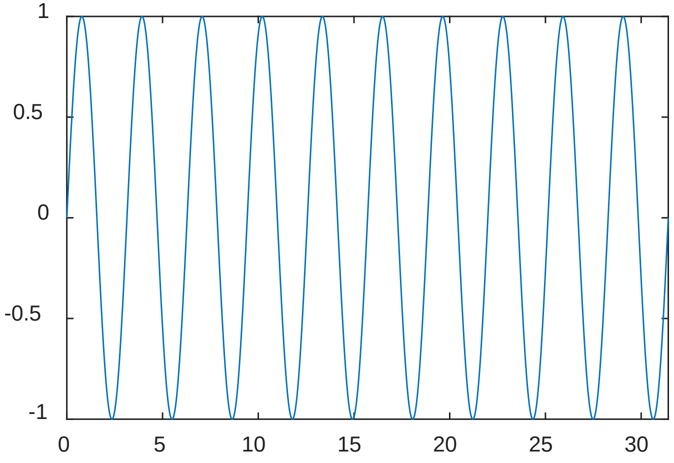
cos(t)



sin(t)+4cos(t)

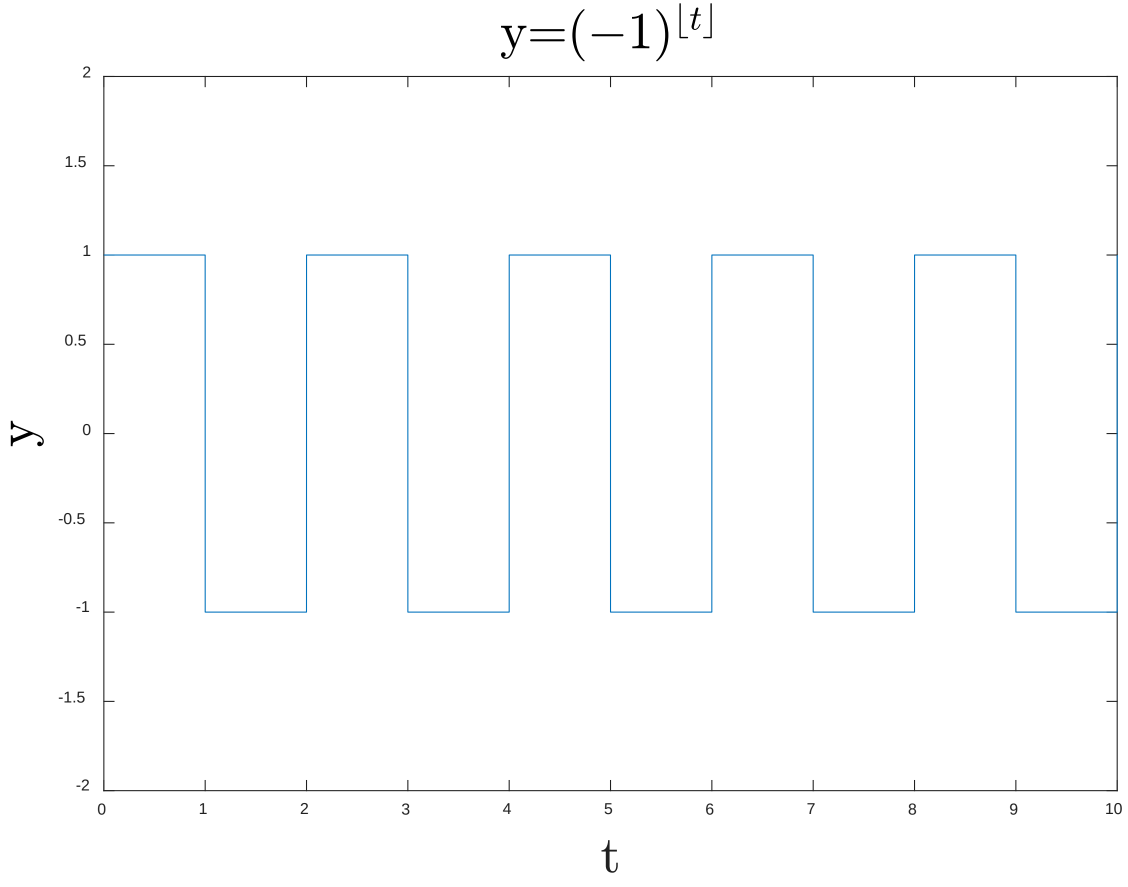


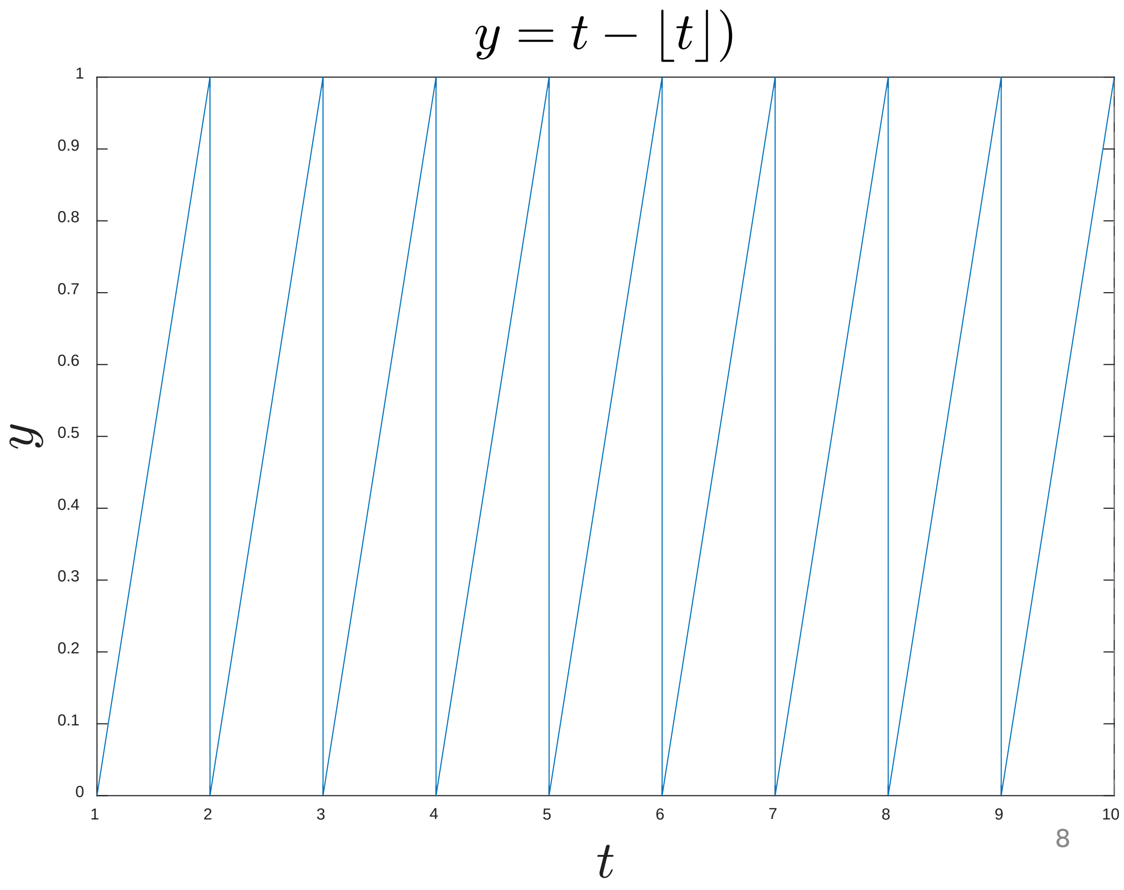
sin(2t)



PERIODIC FUNCTIONS
-- MORE EXAMPLES (NON-TRIGONOMETRIC) --

- $g(t) = (-1)^{\lfloor t \rfloor}$ is a periodic function of period 2
- $g(t) = t - \lfloor t \rfloor$ is a periodic function of period 1





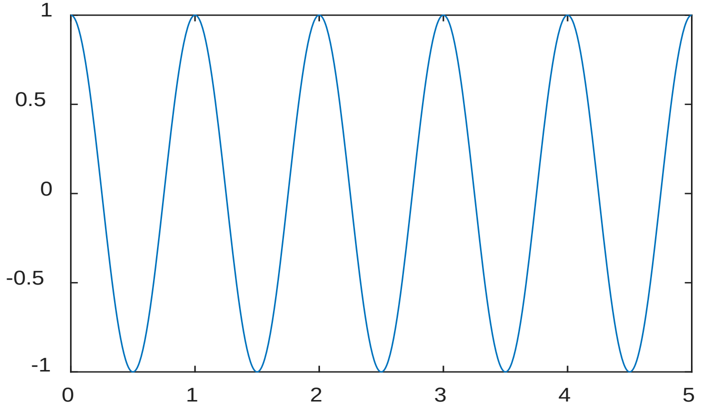
FREQUENCY

- Let $g(t)$ be a periodic function of period T
- **Frequency of a periodic function**: the ***frequency*** f of $g(t)$ is the number of repeated patterns per unit length of t.

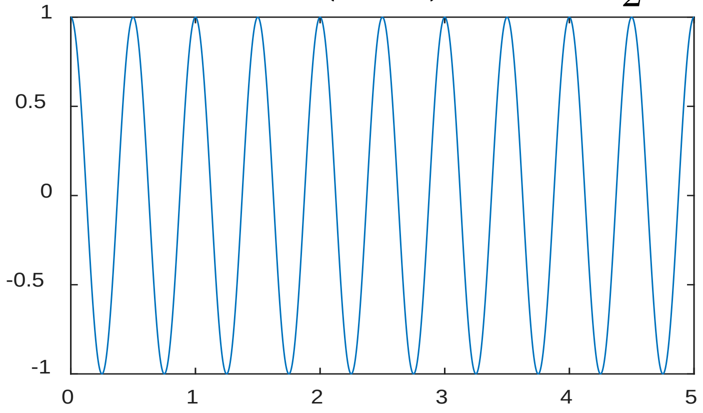
FREQUENCY

-- GRAPHICAL EXAMPLES --

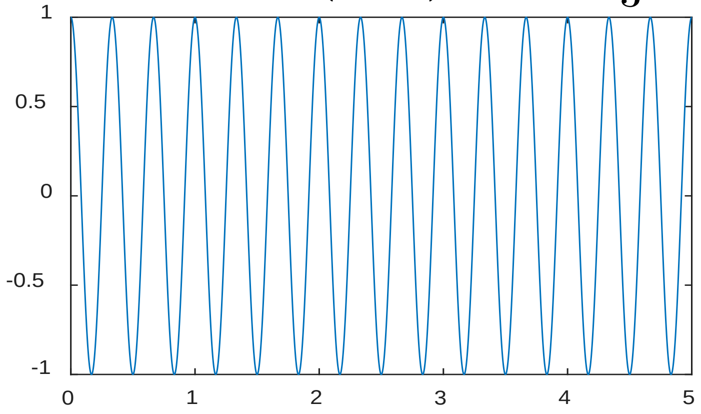
$y = cos(2\pi t),\, T = 1$



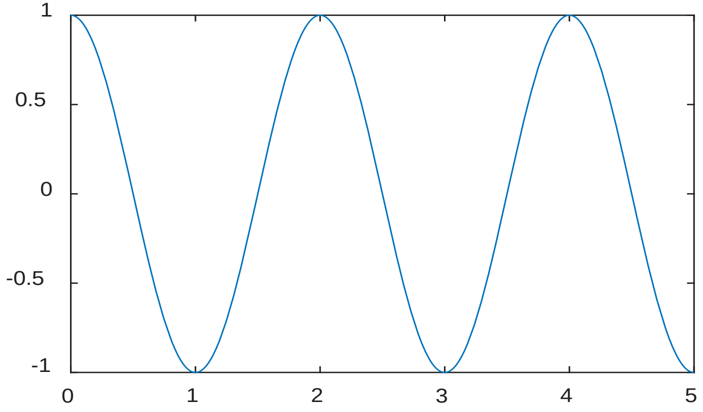
$y = cos(4\pi t),\, T = \frac{1}{2}$



$y = cos(6\pi t),\, T = \frac{1}{3}$



$y = cos(\pi t),\, T = 2$



FREQUENCY
-- FORMALLY, IN TERMS OF THE PERIOD--

- Let $f(t)$ be a periodic function of period T
- **Frequency of a periodic function**: the ***frequency*** of $g(t)$ is the number of repeated patterns per unit length of t.
- **Frequency in term of the period**: The ***frequency*** f of a periodic function/signal $g(t)$ of period T is

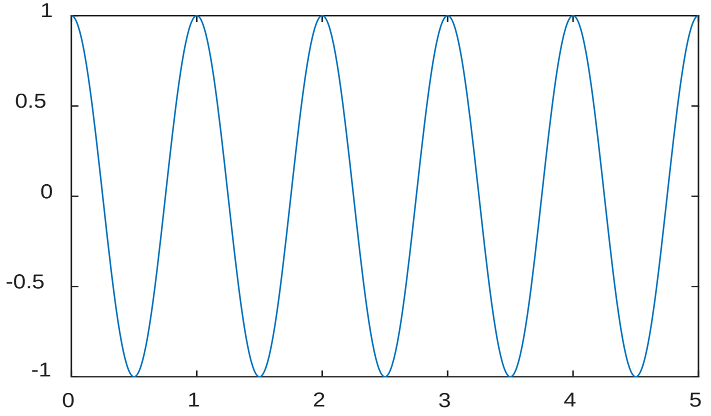
$$\boldsymbol{f = \frac{1}{T}}$$

- **Units of frequencies**: Frequencies are measures in units called Hertz (H) or kilohertz (1000 hertz)

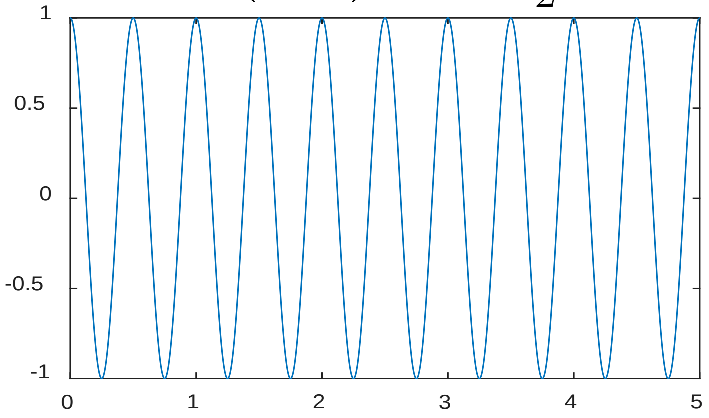
FREQUENCY

-- GRAPHICAL EXAMPLES, SHOWING FREQUENCY f --

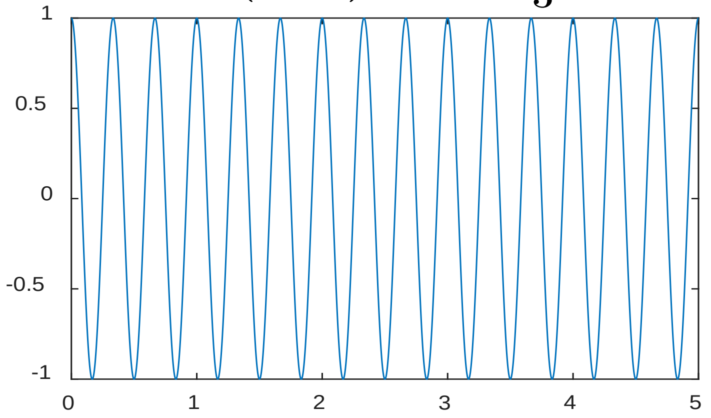
$y = cos(2\pi t),\ T = 1, f = 1$



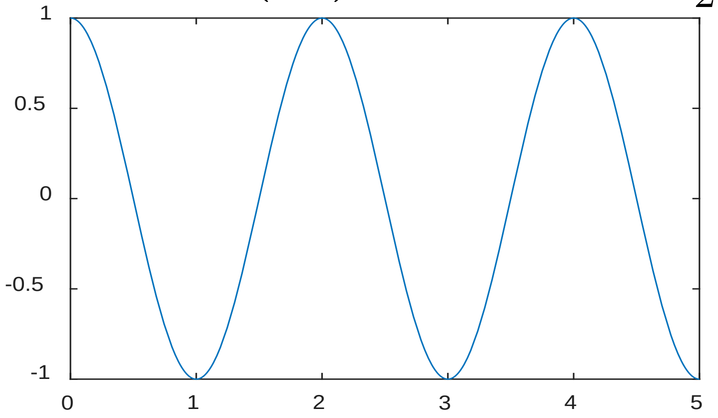
$y = cos(4\pi t), T = \frac{1}{2}, f = 2$



$y = cos(6\pi t), T = \frac{1}{3}, f = 3$



$y = cos(\pi t), T = 2, f = \frac{1}{2}$



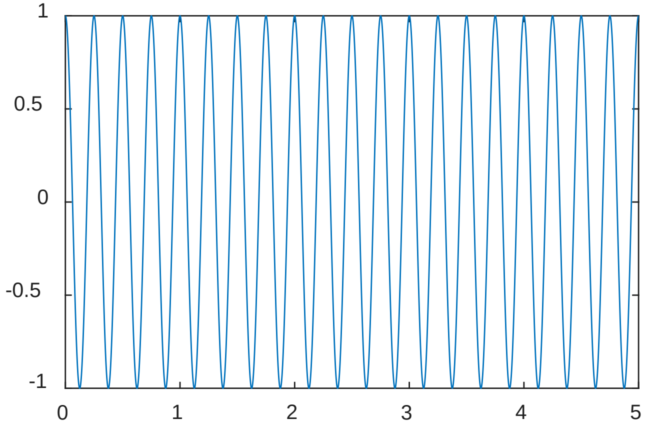
OBSERVATIONS ABOUT FREQUENCIES

- The higher the frequency, the "busier" the signal
- Above a certain frequency, we can no longer see the waves
  - That is, the graph becomes a single blob of ink
- On the opposite end of the spectrum, the smaller the frequency, the less busy the signal
- Below a certain frequency, the signal looks like a straight line, that is,
  - We can't see any waves, any changes
  - The signal looks like a constant

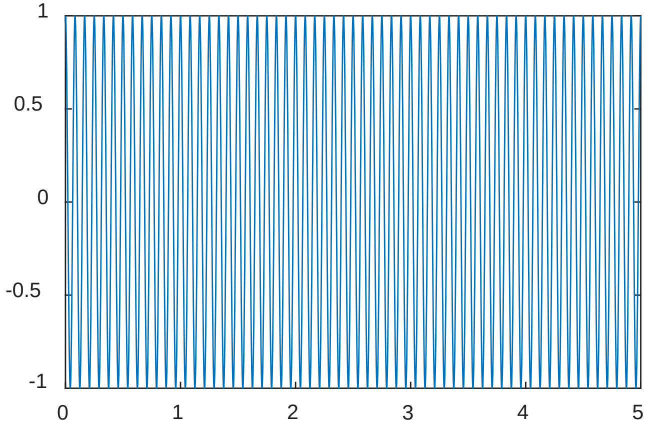
OBSERVATIONS ABOUT FREQUENCIES

-- ILLUSTRATION OF HIGHER FREQUENCIES --

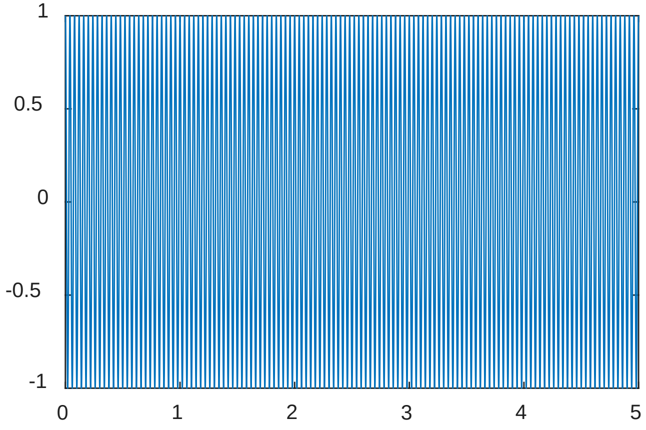
$y = cos(8\pi t), f = 4$



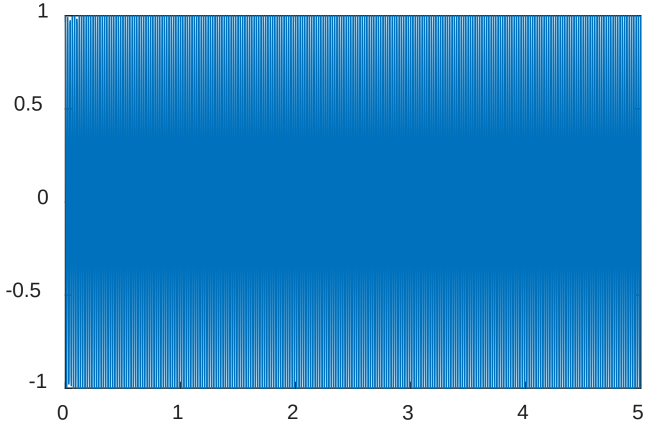
$y = cos(24\pi t), f = 12$



$y = cos(50\pi t), f = 25$

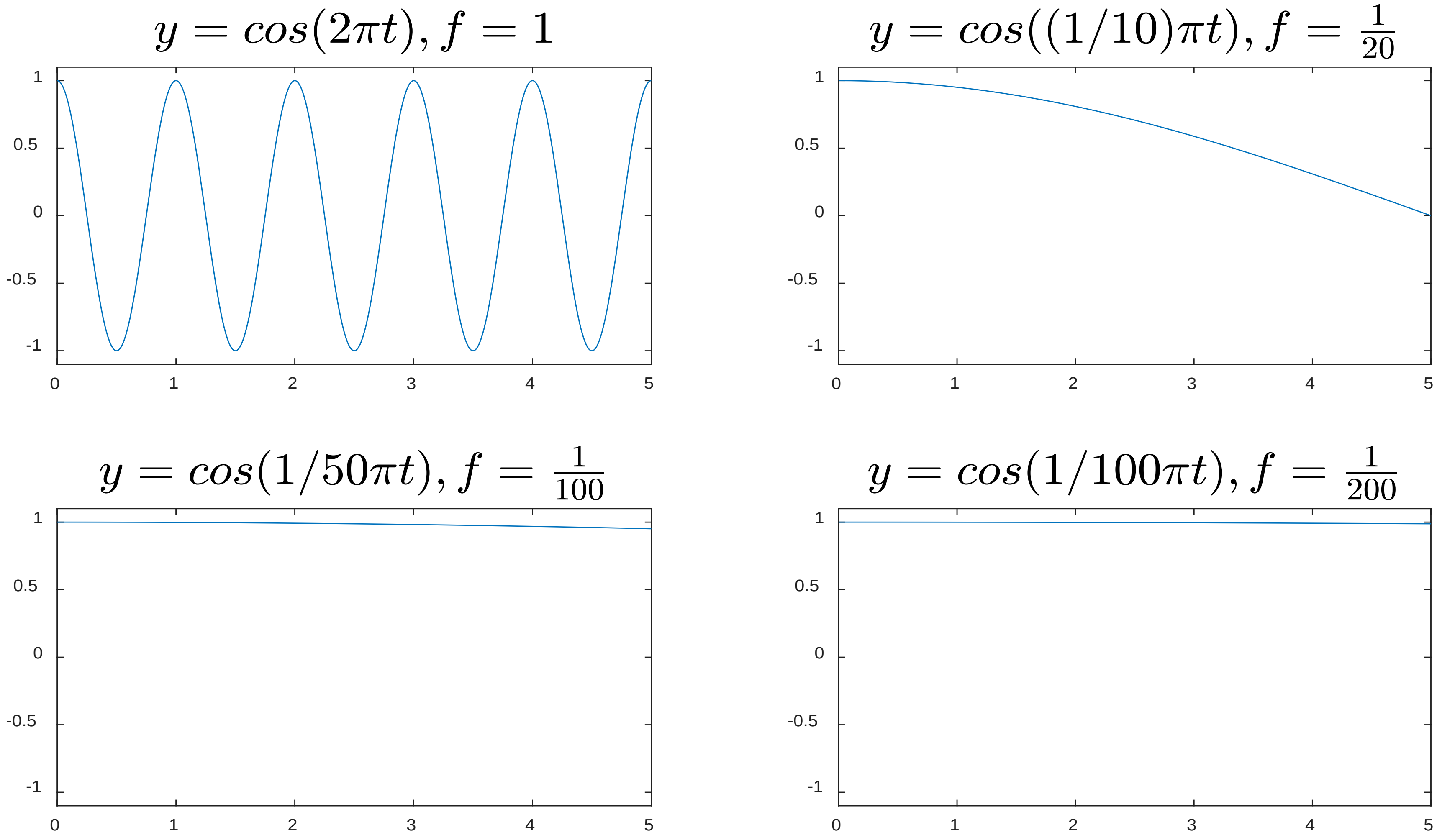


$y = cos(100\pi t), f = 50$



OBSERVATIONS ABOUT FREQUENCIES

-- ILLUSTRATION OF LOWER FREQUENCIES --



Break

FREQUENCIES OF $sin(\alpha t)$ AND $\cos(\alpha t)$

- Since $\sin(\alpha t) = \sin(\alpha t + 2\pi) = \sin(\alpha\left(t + \frac{2\pi}{\alpha}\right))$, we conclude:
  - The period of $\sin(\alpha t)$ is $T = \frac{2\pi}{\alpha}$
  - The frequency of $\sin(\alpha t)$ is $f = \frac{1}{T} = \frac{\alpha}{2\pi}$
- Same for $\cos(\alpha t)$: its period is $T = \frac{2\pi}{\alpha}$, and its frequency is $f = \frac{\alpha}{2\pi}$
- In particular, the frequency of $\sin(k\pi t)$ and $\cos(k\pi t)$ is $f = \frac{k\pi}{2\pi} = \frac{k}{2}$
- So, the greater k, the greater the frequency of $\sin(k\pi t)$ and $\cos(k\pi t)$
- Similarly, the frequency of $\sin(kt)$ and $\cos(kt)$ is $f = \frac{k}{2\pi}$, and so, the greater k, the greater the frequency of $\sin(kt)$ and $\cos(kt)$

WHAT ABOUT REAL LIFE SIGNALS?

- So far, we have been dealing with "ideal" signals: $\sin(\alpha t)$ and $\cos(\alpha t)$
- What about real-world signals, such as sound/speech signals, images, ECG signals, seismic signals, radiations from outer space, etc.
- Certainly, they are not sine or cosine waves (except occasionally)
- So, what is the relevance of all that work we have done with sin and cos?

WHAT ABOUT REAL LIFE SIGNALS?
-- FOURIER'S GREAT DISCOVERY IN 1807 --

- Fourier discovered what we now know as Fourier series
- **Theorem [Fourier 1807)**: For every function/signal $g(t)$ (or $x(t)$) defined for $t \in [0, T]$ and satisfying very mild conditions (typically satisfied by most kinds of real-world signals), we can express $x(t)$ as:

$$x(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos\frac{2\pi}{T} k\, t + b_k \sin\frac{2\pi}{T} k\, t)$$

$$a_k = \frac{1}{T}\int_0^T x(t) \cos\frac{2\pi}{T} k\, t\, dt \qquad b_k = \frac{1}{T}\int_0^T x(t) \sin\frac{2\pi}{T} k\, t\, dt$$

- The sum on the right hand side in the formula above is called a Fourier Series corresponding to $x(t)$

WHAT ABOUT REAL LIFE SIGNALS?

-- FOURIER SERIES IN TERMS OF COMPLEX NUMBER--

- Recall from our review of complex numbers that every complex number $z = a + ib$ can be expressed in polar coordinates as
  - $z = re^{i\theta} = r\cos\theta + ir\sin\theta$
  - $a = r\,cos\,\theta\,, b = r\,sin\,\theta$
- So, $e^{i\theta} = \cos\theta + i\sin\theta$, and thus $\boldsymbol{\cos\theta = \frac{e^{i\theta}+e^{-i\theta}}{2}}$ and $\boldsymbol{\sin\theta = \frac{e^{i\theta}-e^{-i\theta}}{2i}}$
- **Therefore, the series** $x(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k \cos\frac{2\pi}{T}k\,t + b_k \sin\frac{2\pi}{T}k\,t)$ can be expressed in terms of $e^{\frac{2\pi}{T}ikt}$ as follows:

$$x(t) = \sum_{k=-\infty}^{\infty} y_k e^{\frac{2\pi}{T}ikt}, \text{ where } y_k = \frac{1}{T}\int_0^T x(t)e^{-\frac{2\pi}{T}ikt}dt$$

FOURIER TRANSFORM
-- FOR CONTINUOUS SIGNALS, DISCRETE FREQUENCIES --

- **Therefore, the series** $x(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k \cos\frac{2\pi}{T}k\,t + b_k \sin\frac{2\pi}{T}k\,t)$ can be expressed in terms of $e^{\frac{2\pi}{T}ikt}$ as follows:

$$x(t) = \sum_{k=-\infty}^{\infty} y_k e^{\frac{2\pi}{T}ikt}, \text{ where } y_k = \frac{1}{T}\int_0^T x(t)e^{-\frac{2\pi}{T}ikt}dt$$

- The **Fourier Transform** is a mapping: $\boldsymbol{x(t) \rightarrow (y_k)_k}$
- Where y_k is the coefficient of $e^{\frac{2\pi}{T}ikt}$ in $x(t) = \sum_{k=-\infty}^{\infty} y_k e^{\frac{2\pi}{T}ikt}$
- Observe that $e^{\frac{2\pi}{T}ikt}$ (and $\cos\frac{2\pi}{T}k\,t$ and $b_k \sin\frac{2\pi}{T}k\,t$) are periodic of period $\frac{T}{k}$, and of frequency $f_k = \frac{k}{T}$ (referred to as the k^{th} frequency)
- For that reason, y_k is referred to as the $\boldsymbol{k^{th}}$ **frequency content** of $x(t)$, or the $\boldsymbol{k^{th}}$ **frequency component** of $x(t)$

PROVEN FACTS ABOUT FOURIER TRANSFORM/SERIES

- For any integrable function/signal $x(t)$, the Fourier series $\sum_{k=-\infty}^{\infty} y_k e^{\frac{2\pi}{T}ikt}$ exists
- The Fourier series $\sum_{k=-\infty}^{\infty} y_k e^{\frac{2\pi}{T}ikt}$ is not guaranteed to converge, and even if it converges, it is not guaranteed that $x(t) = \sum_{k=-\infty}^{\infty} y_k e^{\frac{2\pi}{T}ikt}$
- However, under mild conditions satisfied by (nearly) all real-life signals, the Fourier series $\sum_{k=-\infty}^{\infty} y_k e^{\frac{2\pi}{T}ikt}$ does converge, and converges to $x(t)$, that is,

$$x(t) = \sum_{k=-\infty}^{\infty} y_k e^{\frac{2\pi}{T}ikt}$$

- **Theorem**: Under those mild conditions where $x(t) = \sum_{k=-\infty}^{\infty} y_k e^{\frac{2\pi}{T}ikt}$, we have

$$y_k \to 0 \text{ when } |k| \to \infty$$

IMPLICATIONS OF FOURIER TRANSFORM/SERIES

- $\boldsymbol{x(t) = \sum_{k=-\infty}^{\infty} y_k e^{\frac{2\pi}{T}ikt}}$ **and** $\boldsymbol{y_k \rightarrow 0}$ **when** $\boldsymbol{|k| \rightarrow \infty}$
- Therefore:
  - Every (analog) signal $x(t)$ can be represented by a discrete sequence $(y_k)_k$
  - The transform is completely invertible: $(y_k)_k \rightarrow x(t) = \sum_{k=-\infty}^{\infty} y_k e^{\frac{2\pi}{T}ikt}$
- Furthermore, since $y_k \rightarrow 0$ when $|k| \rightarrow \infty$,
  - we can drop the y_k's for $|k|$ large enough (like $|k|$ >50, or let's say $|k| > N$ for some N), and
  - the remaining frequency contents $(y_k)_{-N \leq k \leq N}$ are enough to reconstruct a good approximation $\hat{x}(t) = \sum_{k=-N}^{N} y_k e^{\frac{2\pi}{T}ikt} \approx x(t)$
- Therefore, the Fourier Transform lends itself to lossy compression

CONNECTION WITH VECTOR SPACES (1/2)
-- VECTOR SPACE OF ANALOG SIGNALS --

- Recall from vectors spaces that the set V of analog signals, endowed with
  - the addition (+) of signals/functions (where $(g+h)(t) = g(t) + h(t)$), and
  - the multiplication (.) of a signal by any number (where $(ag)(t) = a.g(t)$, changing volume),

  forms a vector space, where the zero O is the zero signal/function (i.e., complete silence in the case of an audio signal)

- Recall also that some vector spaces have bases, such as the vector space $\mathbb{R}^n$
  - The canonical basis $\{e_1, e_2, \dots, e_n\}$, where $e_i = [0, \dots 0,1,0, \dots, 0]^T$
  - $\{f_1, f_2, \dots, f_n\}$ which are the columns of A^{-1} for any invertible linear transform of matrix A

- The question was/is whether the vector space V of analog signals has a basis

CONNECTION WITH VECTOR SPACES (2/2)
-- BASIS FROM THE FOURIER SERIES --

- Thanks to Fourier analysis, we know that for (nearly) all real-life signals $x(t)$, we have

$$x(t) = \sum_{k=-\infty}^{\infty} y_k e^{\frac{2\pi}{T}ikt}$$

- Therefore, the following special signals $\ldots, e^{-\frac{2\pi}{T}i3t}, e^{-\frac{2\pi}{T}i2t}, e^{-\frac{2\pi}{T}it}, 1, e^{\frac{2\pi}{T}it}, e^{\frac{2\pi}{T}i2t}, e^{\frac{2\pi}{T}i3t}, \ldots$ form a **basis** for the vector space of real-world signals. Denote that basis by $\left\{\boldsymbol{e^{\frac{2\pi}{T}ikt}}\right\}_{-\infty\leq \boldsymbol{k}\leq\infty}$

- Similarly, the following special signals $1, \cos\frac{2\pi}{T}t, \cos\frac{2\pi}{T}2t, \cos\frac{2\pi}{T}3t, \ldots, \sin\frac{2\pi}{T}t, \sin\frac{2\pi}{T}2t, \sin\frac{2\pi}{T}3t, \ldots$ form another basis for the vector space of real-world signals, because

$$x(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty}\left(a_k \cos\frac{2\pi}{T}k\,t + b_k \sin\frac{2\pi}{T}k\,t\right)$$

What is the dimension of the vector space V of analog signals?

- Furthermore, when we express a signal $x(t)$ as a linear combination of the vectors of the basis $\left\{\boldsymbol{e^{\frac{2\pi}{T}ikt}}\right\}_{-\infty\leq \boldsymbol{k}\leq\infty}$, i.e., $x(t) = \sum_{k=-\infty}^{\infty} y_k e^{\frac{2\pi}{T}ikt}$, the coefficients $(y_k)_k$ are nearly 0 for large $|k|$

CONNECTION WITH VECTOR SPACES AND WITH THE HUMAN AUDIO-VISUAL SYSTEMS (1/2)

- We expressed in previous lectures the **desire for** a vector space **basis** that aligns (somehow) with the human eyes and ears
- So the question now is:
  - **Does the basis $\left\{e^{\frac{2\pi}{T}ikt}\right\}_{-\infty\leq k\leq\infty}$, or the basis $\left\{cos\frac{2\pi}{T}kt, \sin\frac{2\pi}{T}kt\right\}_{k=0,1,2,\ldots}$, align with the human eyes/ears?**
- The answer is YES, YES

CONNECTION WITH VECTOR SPACES AND WITH THE HUMAN AUDIO-VISUAL SYSTEMS (2/2)

- YES, YES, the bases $\{e^{\frac{2\pi}{T}ikt} \mid -\infty \le k \le \infty\}$ and $\left\{\cos\frac{2\pi}{T}kt, \sin\frac{2\pi}{T}kt \;\middle|\; k = 0,1,2,\ldots\right\}$, both align with the human eyes/ears
- Experiments and studies have shown that
  - the human eyes & ears are more sensitive to low-frequency contents than to high frequency contents, i.e., the $(y_k)_k$ for smaller $|k|$ are more important to our eyes & ears than the $(y_k)_k$ for larger $|k|$
  - our sensitivity peaks at some very low frequency k_{peak}
  - our sensitivity decreases as the frequency increases beyond k_{peak}, i.e., the importance of y_k decreases as k increases, for all $k > k_{peak}$
- Implications to lossy compression (later)

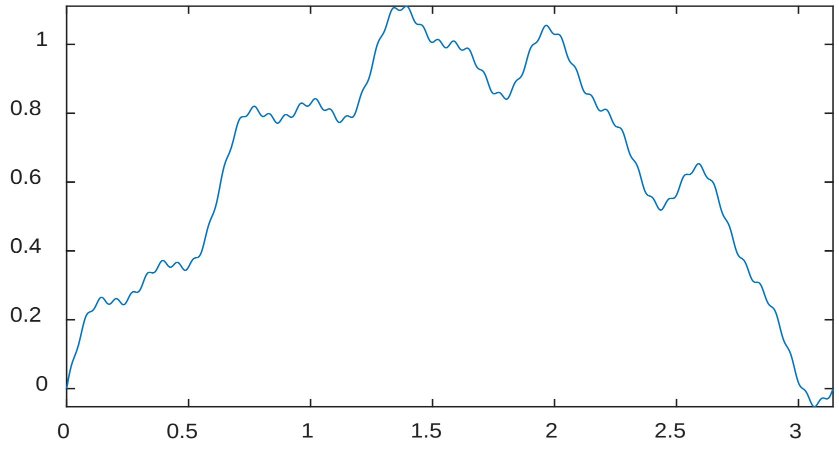
VISUAL ILLUSTRATIONS OF OUR SENSITIVITY TO DIFFERENT FREQUENCIES (1/4)

- Take an original signal $\sin(t) + 0.1\sin(10t) + 0.05\sin(20t) + 0.01\sin(100t)$
- This signal is the combination of 4 different frequencies
- Drop terms corresponding to different frequencies, and see the corresponding effect

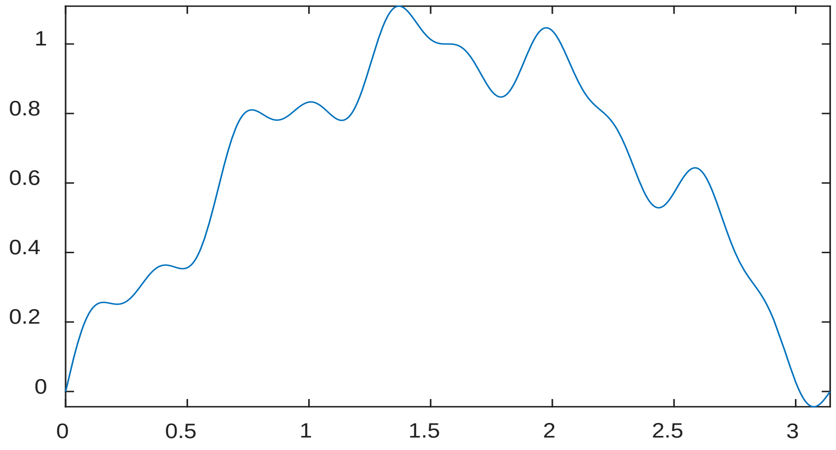
Frequencies Dropped	Signal
None (keep original signal)	$\sin(t) + 0.1\sin(10t) + 0.05\sin(20t) + 0.01\sin(100t)$
Highest frequency (100)	$\sin(t) + 0.1\sin(10t) + 0.05\sin(20t)$
2nd highest frequency (20)	$\sin(t) + 0.1\sin(10t) + \quad 0.01\sin(100t)$
Two highest frequencies (20, 100)	$\sin(t) + 0.1\sin(10t)$
2nd lowest frequency (10)	$\sin(t) + \quad 0.05\sin(20t) + 0.01\sin(100t)$
2nd & 3rd lowest frequencies (10, 20)	$\sin(t) + \quad 0.01\sin(100t)$
Three highest frequencies (10,20,100)	$\sin(t)$
Lowest frequency (1)	$0 + \quad 0.1\sin(10t) + 0.05\sin(20t) + 0.01\sin(100t)$

VISUAL ILLUSTRATIONS OF OUR SENSITIVITY TO DIFFERENT FREQUENCIES (2/4)

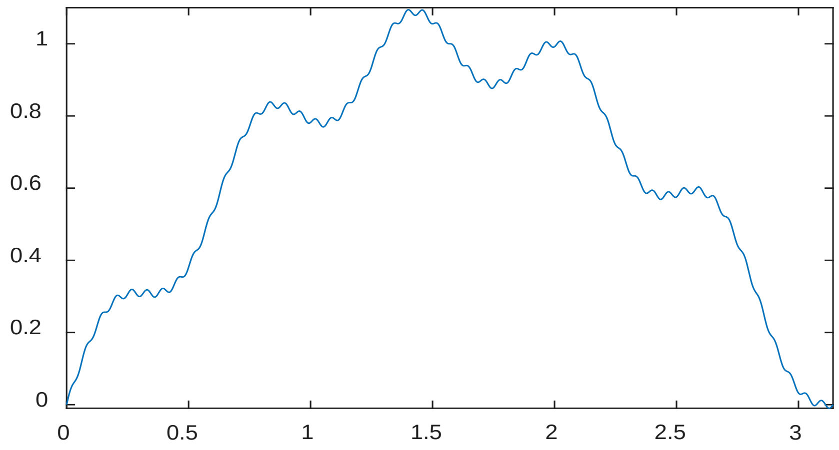
$$\sin(t) + 0.1\sin(10t) + 0.05\sin(20t) + 0.01\sin(100t)$$



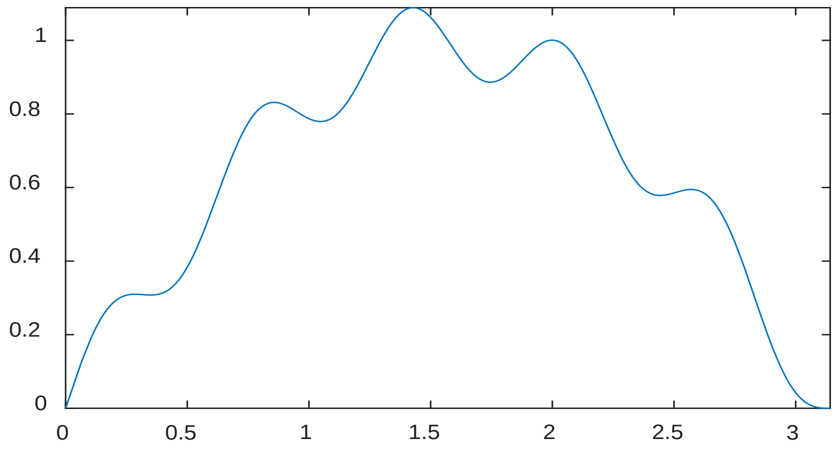
$$\sin(t) + 0.1\sin(10t) + 0.05\sin(20t) + 0$$



$$\sin(t) + 0.1\sin(10t) + 0 + 0.01\sin(100t)$$

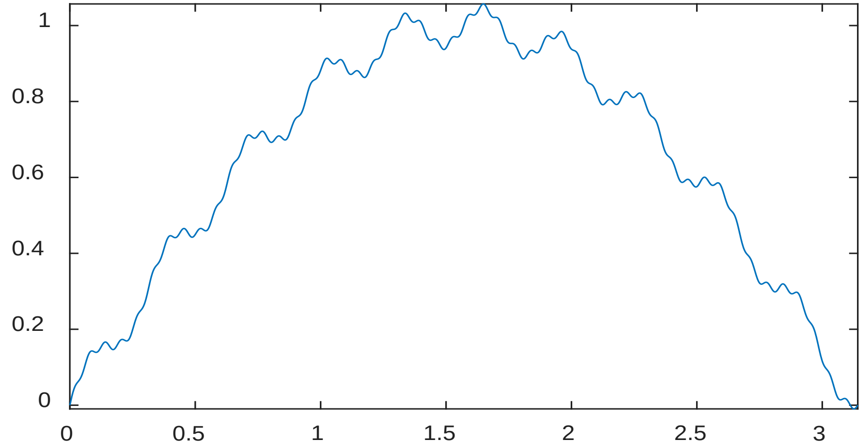


$$\sin(t) + 0.1\sin(10t) + 0 + 0$$

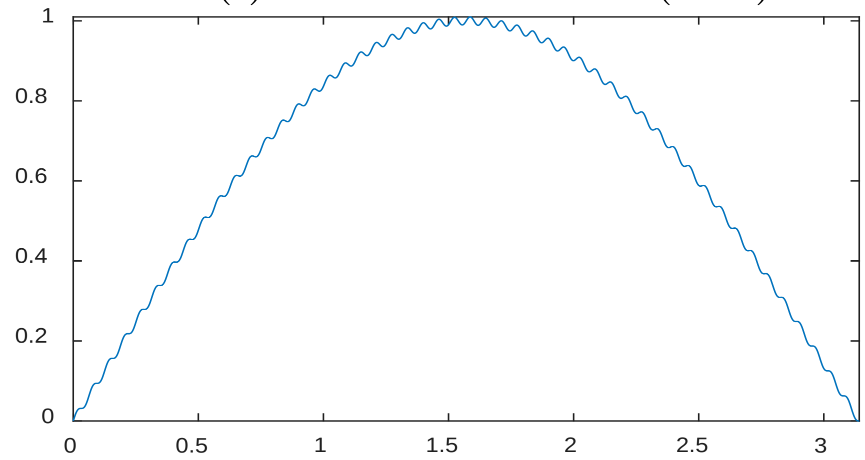


VISUAL ILLUSTRATIONS OF OUR SENSITIVITY TO DIFFERENT FREQUENCIES (3/4)

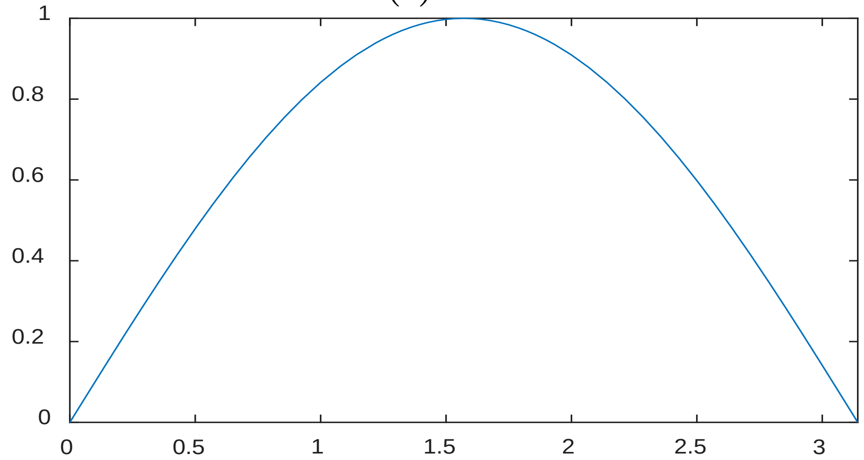
$\sin(t) + \quad 0.05\sin(20t) + 0.01\sin(100t)$



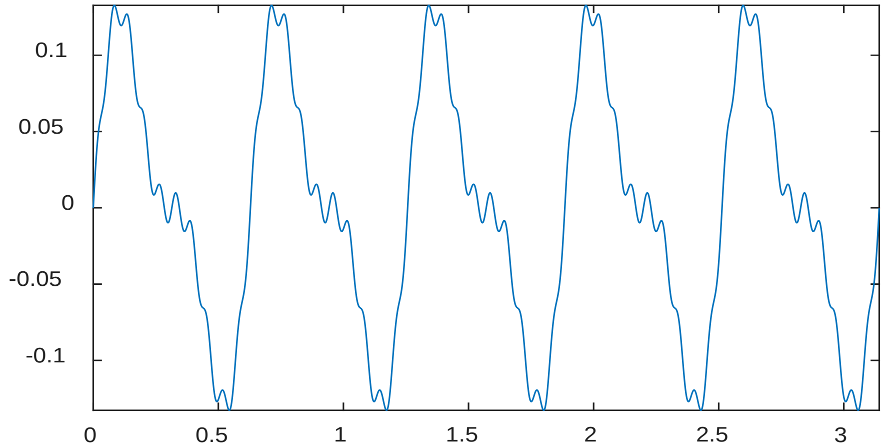
$\sin(t) + \quad 0.01\sin(100t)$



$\sin(t)$



$0 + \quad 0.1\sin(10t) + 0.05\sin(20t) + 0.01\sin(100t)$



VISUAL ILLUSTRATIONS OF OUR SENSITIVITY TO DIFFERENT FREQUENCIES (4/4)

- Notice how the dropping of higher frequencies did not affect much the overall shape of the plot
- Even if multiple high frequency components are deleted simultaneously, the overall shape of the plot was largely retained
- The visual loss due to dropping high frequency components was primarily a loss of fine details (usually **<u>imperceptible</u>** to human senses)
- The biggest (worst) effect was observed when the lowest frequency content was dropped
  - The shape of the plot was changed dramatically

IMPLICATIONS FOR LOSSY COMPRESSION

- So, to recall and recap, for human eyes and ears:
  - the low-frequency contents are more important than high frequency contents, i.e., the $(y_k)_k$ for smaller $|k|$ are more important than those for larger $|k|$
  - The importance of the individual y_k's peaks at some very low frequency k_{peak}
  - The importance of y_k decreases as k increases, for all $k > k_{peak}$
- Therefore, for great lossy compression performance:
  - Not only drop the y_k's for large $|k|$ (i.e., $|k| >$ N) and retain $\sum_{k=-N}^{N} y_k e^{\frac{2\pi}{T}ikt}$
  - But also quantize increasingly aggressively the y_k's of increasing k's $\leq N$, to greatly save on bits with hardly any perceptible effect on the reconstructed signal
  - And quantize lightly the y_k's of small k's, to largely preserve the important audio/visual features of the signal

$\hat{y}_k$ is the quantized-then-dequantized y_k

  - The reconstructed (analog) signal will be $\hat{x}(t) = \sum_{k=-N}^{N} \hat{y}_k e^{\frac{2\pi}{T}ikt}$

Break

TREATMENT OF DISCRETE SIGNALS (1/4)

- So far, all the analysis and implications assumed analog signals (because, as you recall, calculus is easier for continuous functions)

- But in practice, signals have become digital

- Therefore, we need to transition somehow to digital signals, while retaining the benefits of all that work we did on analog signals

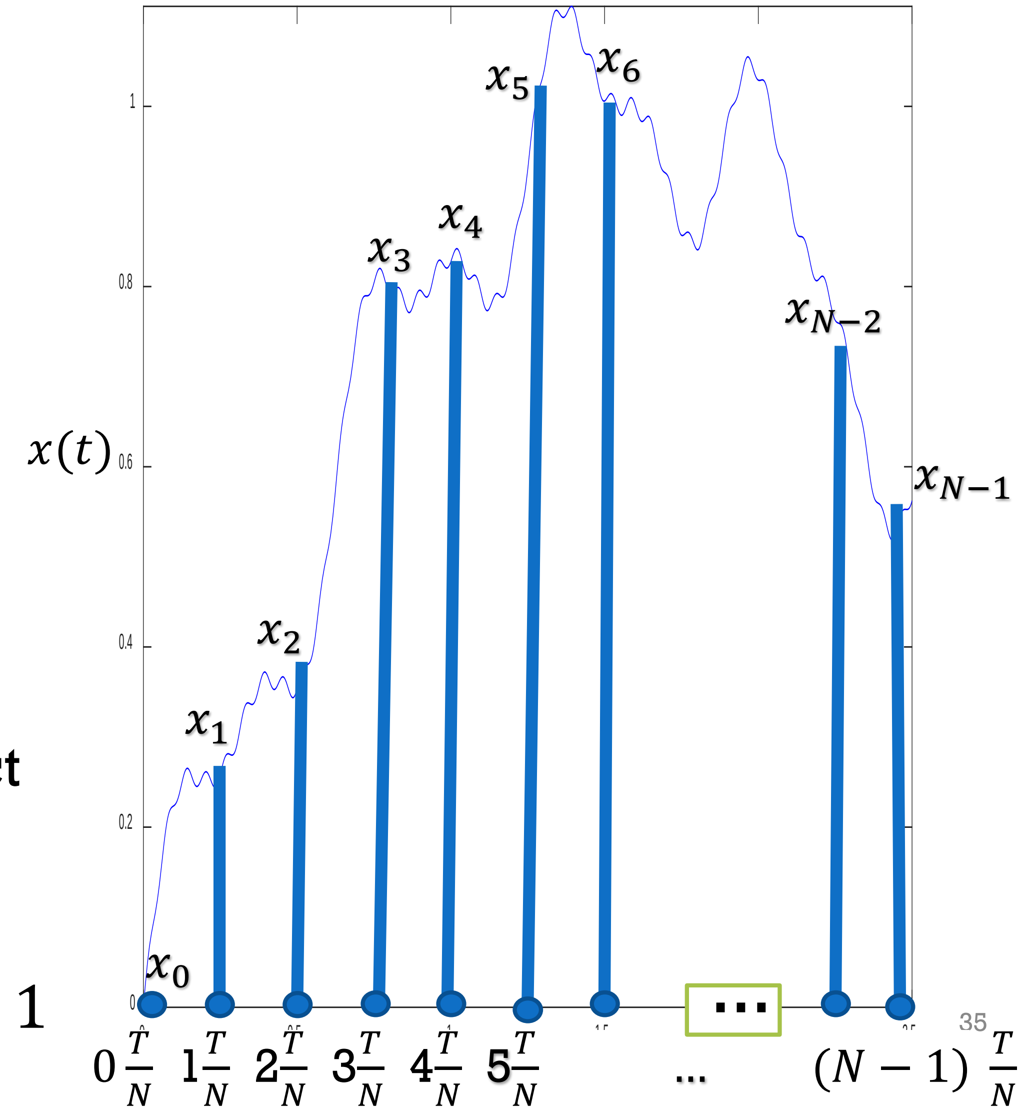
- For that, we will digitize analog signals appropriately, as will be done next

TREATMENT OF DISCRETE SIGNALS: DIGITIZATION (2/4)

- Take analog $x(t)$, recorded over the period $[0, T]$):
- Sample N values from analog signal $x(t)$:

  $x_l = x(l\frac{T}{N})$, for $l = 0,1,2, \dots, N-1$

  - $x_l = x\left(l\frac{T}{N}\right) = \sum_k y_k e^{\frac{2\pi}{T}ikt}$ for $t = l\frac{T}{N}$
  - Therefore, $x_l = \sum_k y_k e^{\frac{2\pi}{T}ikl\frac{T}{N}} = \sum_k y_k e^{\frac{2\pi}{N}ikl}$, for $l = 0,1,2, \dots, N-1$
- Since the sequence (x_l) is discrete and finite, there is no need to keep an ∞ of y_k's
- Rather $y_0, y_1, \dots, y_{N-1}$ are sufficient (to reconstruct (x_l) for $l = 0,1,2, \dots, N-1$)
- So, $x_l = x\left(l\frac{T}{N}\right) = \sum_{k=0}^{N-1} y_k e^{\frac{2\pi}{N}ikl}, l = 0,1, \dots, N-1$



TREATMENT OF DISCRETE SIGNALS (3/4)
-- DISCRETE FOURIER TRANSFORM --

- From last slide: $x_l = x\left(l\frac{T}{N}\right) = \sum_{k=0}^{N-1} y_k e^{\frac{2\pi}{N}ikl}$, for $l = 0,1,2,\ldots,N-1$
- This can be put in matrix form:

$$\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix} = \begin{bmatrix} b_{00} & b_{01} & \cdots & b_{0,N-1} \\ b_{01} & b_{02} & \cdots & b_{0,N-1} \\ \vdots & \ddots & & \vdots \\ b_{N-1,0} & b_{N-1,1} & \cdots & b_{N-1.N-1} \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix}$$

$$x = By$$

$$b_{kl} = e^{\frac{2\pi}{N}ikl}$$

- Note the change in notation: from $x = x(t)$ to x being a vector in $\mathbb{R}^N$

-- DISCRETE FOURIER TRANSFORM --

- From $x = By$, we conclude that $y = Ax$ for $A = B^{-1}$

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix} = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0,N-1} \\ a_{01} & a_{02} & \cdots & a_{0,N-1} \\ \vdots & \ddots & & \vdots \\ a_{N-1,0} & a_{N-1,1} & \cdots & a_{N-1.N-1} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix}$$

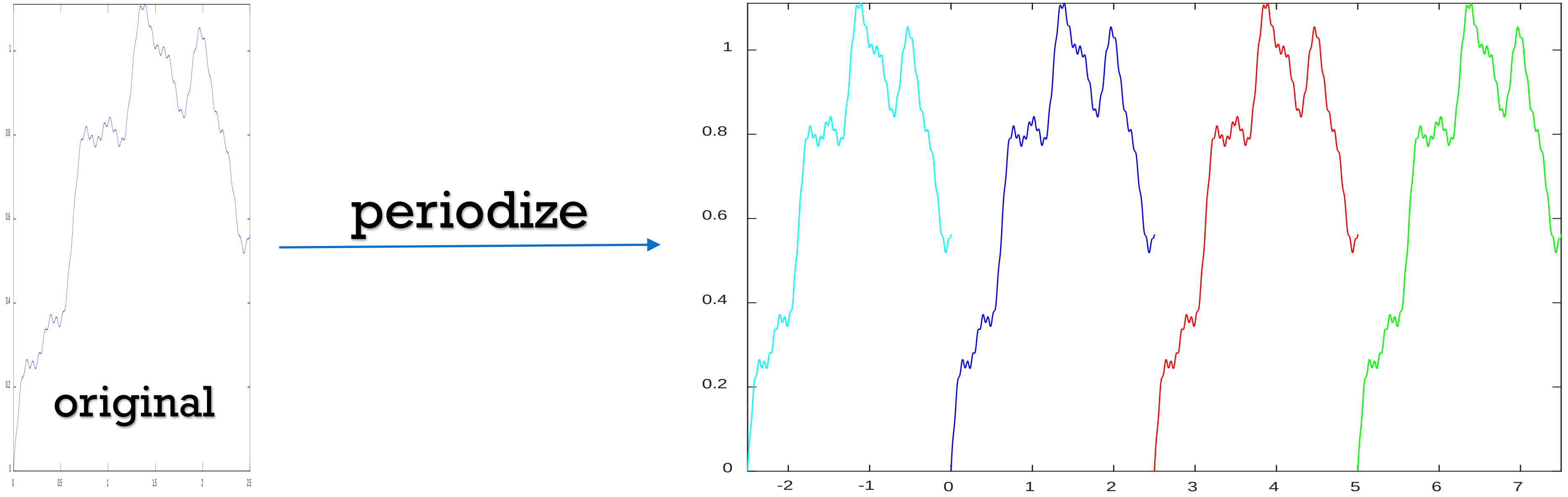
$$x = Ay$$

$$\boxed{a_{kl} = \frac{1}{N} e^{-\frac{2\pi}{N}ikl}}$$

- That is exactly the Discrete Fourier Transform we defined two lectures ago, with a slight adjustment: $a_{kl} = \sqrt{\frac{1}{N}} e^{-\frac{2\pi}{N}ikl}$ and $b_{kl} = \sqrt{\frac{1}{N}} e^{\frac{2\pi}{N}ikl}$

WHY DISCRETE COSINE TRANSFORM (DCT) INSTEAD OF DISCRETE FOURIER TRANSFORM (DFT) (1/5)

- When we took an analog $x(t)$, recorded over the period $[0, T]$), and we expressed $x(t)$ as the Fourier series $x(t) = \sum_{k=-\infty}^{\infty} y_k e^{\frac{2\pi}{T}ikt}$, we ended up "***periodizing***" $x(t)$ because $\sum_{k=-\infty}^{\infty} y_k e^{\frac{2\pi}{T}ikt}$ is periodic of period T



- Now, observe the major discontinuities (breaks) in the periodized $x(t)$ (for $-\infty < t < \infty$): we see (breaks) from each period to the next

WHY DCT INSTEAD OF DFT (2/5)
-- DISCONTINUITIES LEAD TO RINGING ARTIFACTS --

- Those breaks between periods are big, sharp (sudden) changes
- Big sharp changes mean high frequency, i.e., need high-frequency components to "capture"/represent them
- Therefore, big sharp changes manifest themselves in the Fourier series as large high-frequency contents y_k
- This means when we eliminate or quantize heavily all the high frequency contents, those contents that represent those breaks are gone
- As a result, the reconstructed signal won't be that good
  - You get what is known as boundary artifacts, also known as Gibbs phenomenon
  - In practical/visual terms, those distortions appear like ringing/echoing

WHY DCT INSTEAD OF DFT: RINGING VISUALIZED (3/5)

original

reconstructed

WHY DCT INSTEAD OF DFT (4/5)

-- HOW TO AVOID RINGING ARTIFACTS --

- Note that if the original signal $x(t)$ recorded over $[0, T]$ happens to satisfy $x(0) = x(T)$ or $x(0) \approx X(T)$, then the discontinuities between periods won't exist or will be minor
- So, this gives us an idea:
    1. Modify the signal in some systematic fashion so its <u>end value</u> is (approximately) equal to its <u>start value</u>
    2. Then apply the Fourier transform
- But this leaves another drawback of DFT …

WHY DCT INSTEAD OF DFT (5/5)
-- ANOTHER DRAWBACK OF DFT --

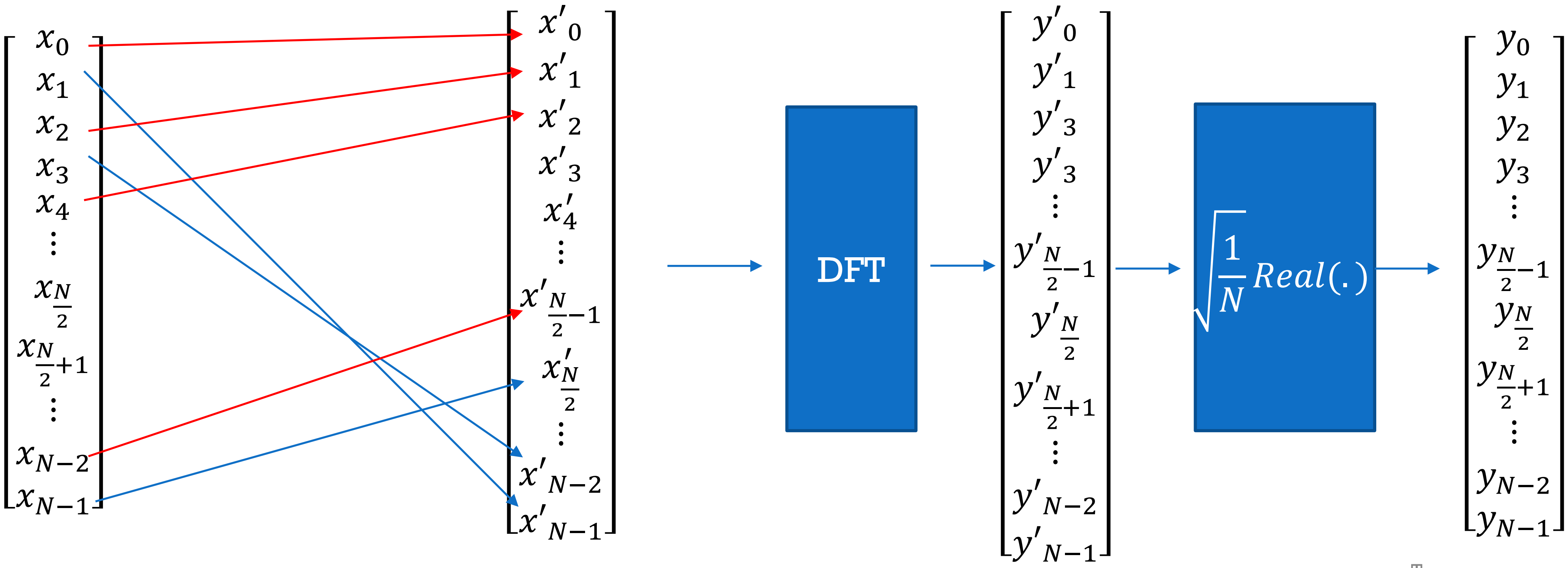
- But this leaves another major drawback of DFT …
  - It introduces complex numbers
  - Computationally, since a complex number is actually two real numbers, it doubles memory and computation time
  - Furthermore, when you quantize and apply the inverse transform, the reconstructed signal usually has complex numbers, which can’t be displayed/played (ouch!)
- Therefore, we need an alternative transform that
  - Deals only with real number: no complex numbers
  - Mitigates the discontinuity/boundary problems caused by periodization
  - And yet preserves the frequency insight/benefits of DFT

DISCRTE COSINE TRANSFORM (DCT)

- We will see that DCT preserves much of the advantages of DFT (frequency perspective) without the drawbacks of DFT (boundary problems and complex numbers)
- By why DFT in the first place?
  - Because of its great mathematical/analytical power and insight
- How does DCT accomplish those desirable goals?
  - We'll see that next

RELATION OF DCT TO DFT (1/2)

- Take the digital signal $x = [x_0, x_1, \ldots, x_{N-1}]^T$ and assume N is even
- Shuffle x to become $x' = [x'_0, x'_1, \ldots, x'_{N-1}]^T$ where $x'_k = x_{2k}$ and $x'_{N-k-1} = x_{2k+1}$ for $k = 0,1,2, \ldots, \frac{N}{2} - 1$. Then apply DFT to get y', then take $y = \sqrt{\frac{1}{N}} Real(y')$, you get DCT.



RELATION OF DCT TO DFT (2/2)

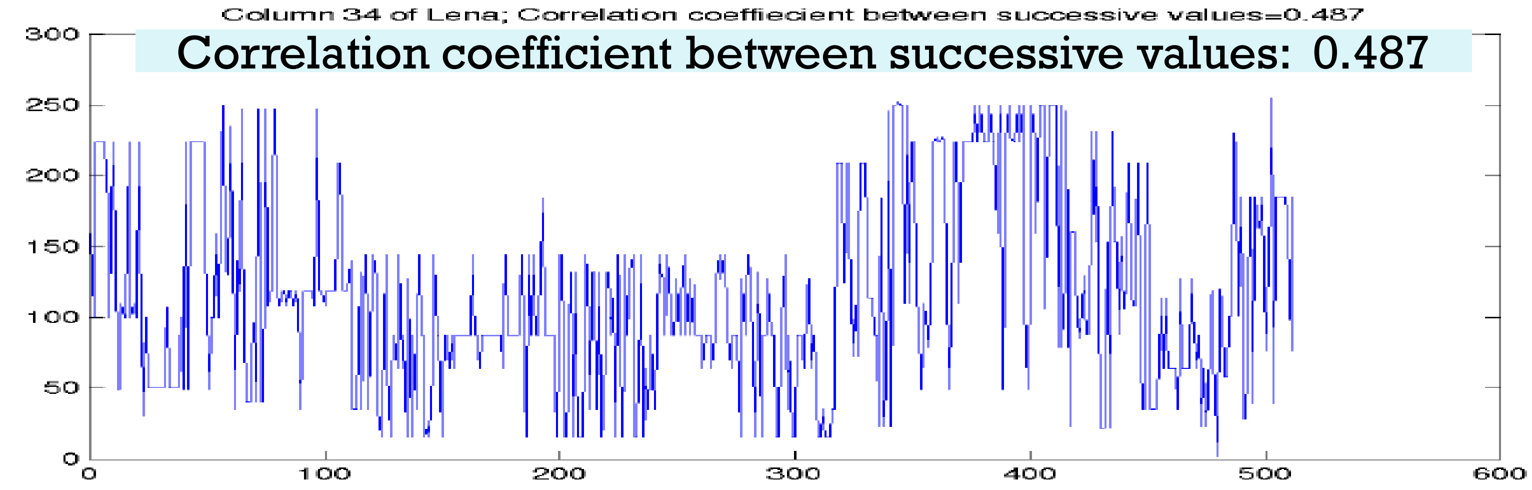
- The shuffling of $x = [x_0, x_1, \dots, x_{N-1}]^T$ to get $x' = [x'_0, x'_1, \dots, x'_{N-1}]^T$ causes the last sample of x' to be nearly equal to the first sample of x':
  - $x'_{N-1} = x_1 \approx x_0 = x'_0 \Rightarrow \boldsymbol{x'_{N-1} \approx x'_0}$
- Therefore, the discontinuities (breaks) between periods are minor, which is what we want
- DCT deals with real numbers, which is what we want
- DCT uses DFT (mathematically rather computationally), which inherits the frequency-perspective benefits, which is what we want

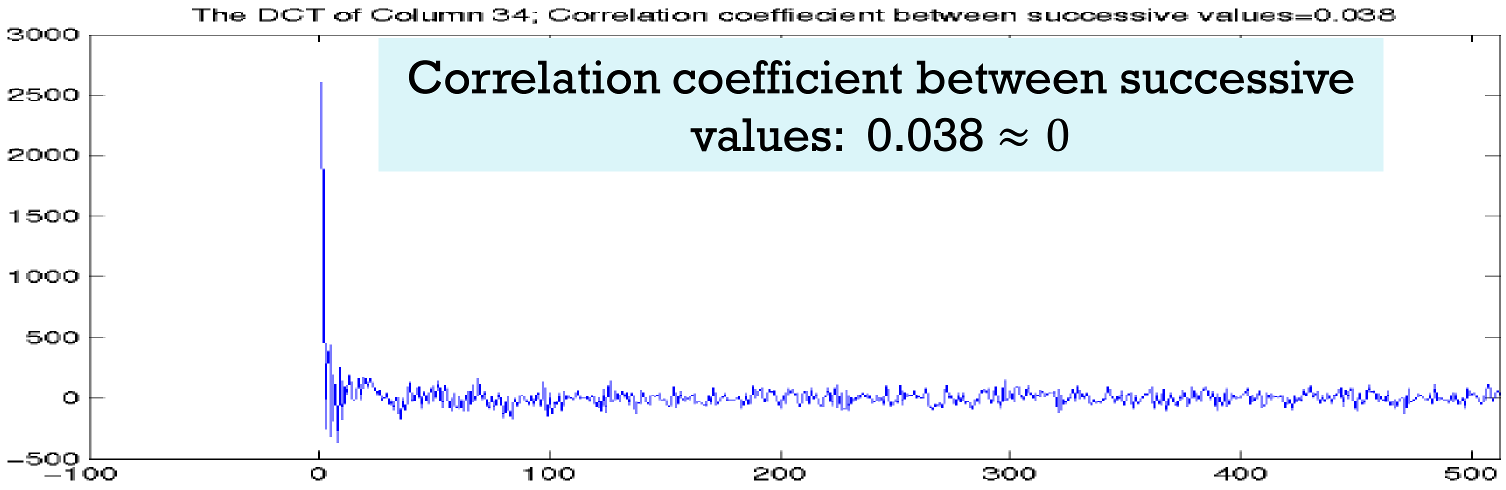
STATISTICAL PERSPECTIVE (1/4)
-- DECORRELATION OF DATA --

- Decorrelation of data leads to energy compaction, that is, concentrating the visual contents into a few coefficients
- Decorrelation of data minimizes the distortions caused by scalar quantization
- Does the DFT/DCT decorrelate the data, or greatly reduce the correlation?
  - Yes
  - See the illustration next

STATISTICAL PERSPECTIVE (2/4)

-- ILLUSTRATION OF DATA-DECORRELATION --





STATISTICAL PERSPECTIVE (3/4)

-- TOTAL DECORRELATION/ENERGY-COMPACTION --

- We just saw that DCT decorrelates the data greatly, but not completely
- Is there a transform that decorrelates completely?
- Yes: The Karhunen-Loeve (KL) transform
  - Decorrelates the signal data completely, and thus compacts the energy into the minimum number of coefficients
  - Compacts the energy the most, that is
    - For every k, the MSE_k between the original signal and the one reconstructed from the k most important coefficients of a transform is minimized by KL

KL transform is the same as Principal Component Analysis (PCA) in Machine Learning

- But, KL has one major drawback
  - The KL matrix A depends on the input signal
  - So, it takes time to derive A, and takes memory to store A for the decoder

STATISTICAL PERSPECTIVE (4/4)
-- DCT VS. KL VS. DFT --

- It turned out, the matrix of DCT is a good approximation of the KL matrix
- Therefore, DCT is a good approximation of KL
- Therefore
  - DCT is nearly optimal for decorrelation and energy compaction
  - Its matrix is data-independent: no need to derive a separate matrix for each new signal, and no need to compute/store that matrix
  - It deals with real numbers rather than complex numbers
  - It inherits from DFT the frequency-perspective and the ensuing insight, and speed of computation
- In fact, there is a divide-and-conquer algorithm (by Cooley and Tukey) for computing DFT (and thus DCT) in $O(n \log n)$ instead of $O(n^2)$ time

DCT VS. THE OTHER TRANSFORMS

- The Walsh-Hadamard transform is the integer approximation of DCT
- It was good when hardware was expensive and basic, since in Walsh-Hadamard
  - Only integer arithmetic is needed
  - And indeed, only additions and subtractions are needed, and no multiplication or division, which are more expensive
- But now, hardware is cheap and fast, even for floating-point arithmetic
- And DCT is mathematically better at energy compaction, decorrelation, and frequency exploitation, than Walsh -Hadamard
- The same can be said about DCT vs. the Haar transform

WRAP-UP

- We saw many transforms, and studied them from different perspectives: computation, vector space, frequency, and statistical perspectives
- DFT gives great mathematical & neurological insight
- For lossy data compression, DCT inherits the benefits of DFT without the practical drawbacks of DFT
- DCT is the best of all worlds among the transforms we have studied
- No wonder that DCT was adopted for lossy compression standards
- Next lecture we study the JPEG and MPEG standards
- (We will see later on that despite its advantages, DCT is not perfect and has its drawbacks, and we'll study better alternatives)

NEXT LECTURE

- Next week: the midterm

- The week after that: JPEG and MPEG standards