AUTOMATIC SPEECH RECOGNITION

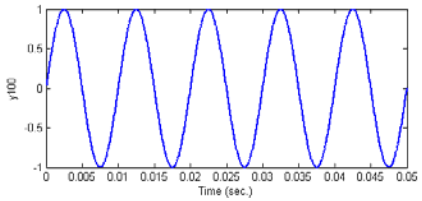
Lecture 2 & 3

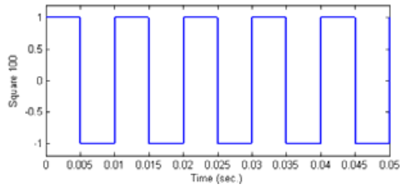
Speech Signal Fundamentals

Department of Computer Engineering, Chulalongkorn University
Spoken Language Systems Research Group

Signal

- Periodic Signal





- Non-periodic Signal
- Random Signal

2110432 Auto. Speech. Recog. | First Semester 2007 | Lecture 2 & 3
ATIWONG SUCHATO

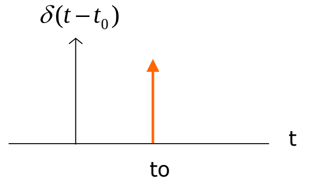
Department of Computer Engineering, Chulalongkorn University
Spoken Language Systems Research Group

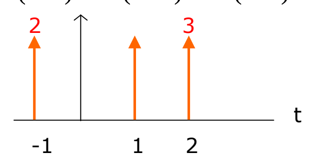
Impulse

$$\delta(t-t_0)=\begin{cases}\infty & ; \ t=t_0 \\ 0 & ; \ t\neq t_0\end{cases}$$

and

$$\int_{-\infty}^{\infty}\delta(t-t_0)\,dt=1$$





2110432 Auto. Speech. Recog. | First Semester 2007 | Lecture 2 & 3
ATIWONG SUCHATO

Department of Computer Engineering, Chulalongkorn University
Spoken Language Systems Research Group

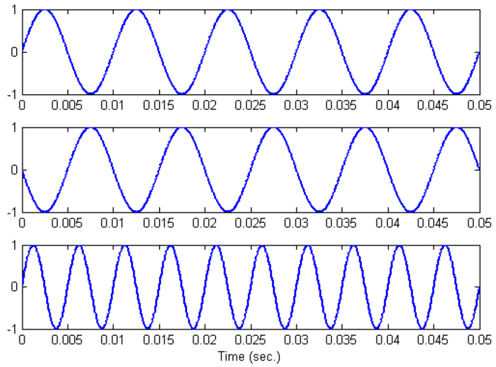
Sinusoidal Signal

$$y = A\sin(2\pi f t+\phi)$$

$f = 100Hz.$, $\phi = 0$

$f = 100Hz.$, $\phi = \pi$

$f = 200Hz.$, $\phi = 0$



$A = 1$

2110432 Auto. Speech. Recog. | First Semester 2007 | Lecture 2 & 3
ATIWONG SUCHATO

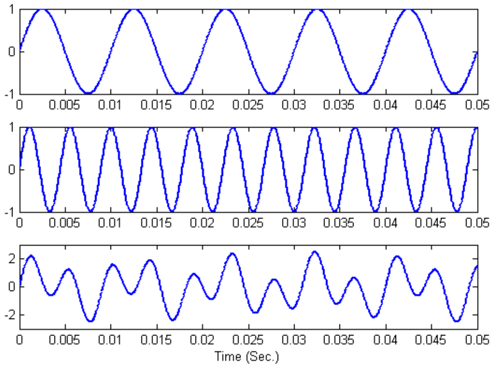
Linear Combination of Signals

$$s = a_1 y_1 + a_2 y_2 + \cdots + a_N y_N$$

$y_1 = \sin(2\pi(100)t)$

$y_2 = \sin(2\pi(225)t)$

$s = y_1 + 1.5y_2$



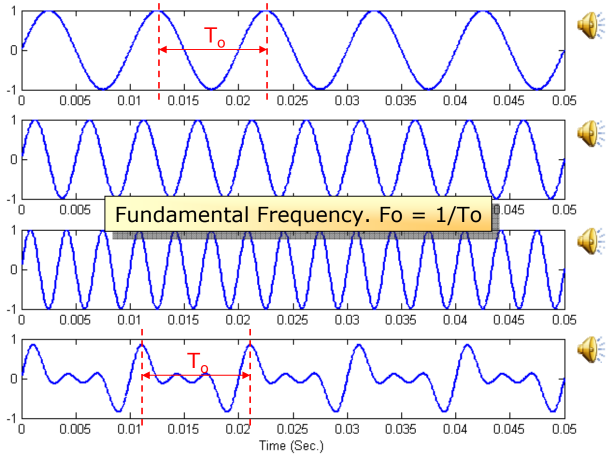
Fundamental Frequency and Harmonics

$y_1 = \sin(2\pi(100)t)$

$y_2 = \sin(2\pi(200)t)$

$y_3 = \sin(2\pi(300)t)$

$s = 0.3y_1 + 0.4y_2 + 0.3y_3$



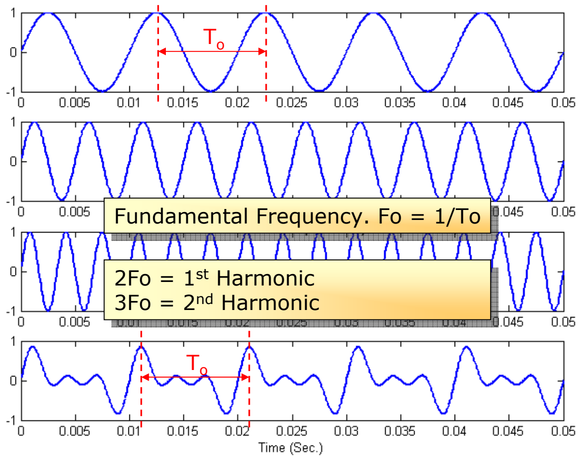
Fundamental Frequency and Harmonics

$y_1 = \sin(2\pi(100)t)$

$y_2 = \sin(2\pi(200)t)$

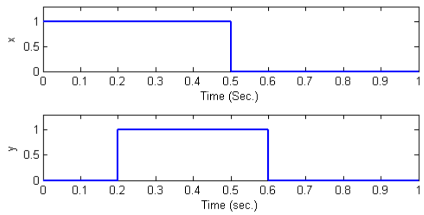
$y_3 = \sin(2\pi(300)t)$

$s = 0.3y_1 + 0.4y_2 + 0.3y_3$



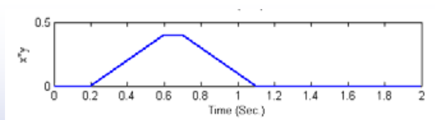
Convolution

$$x(t) * y(t) = \int_{-\infty}^{\infty} x(\tau) y(t-\tau) d\tau$$



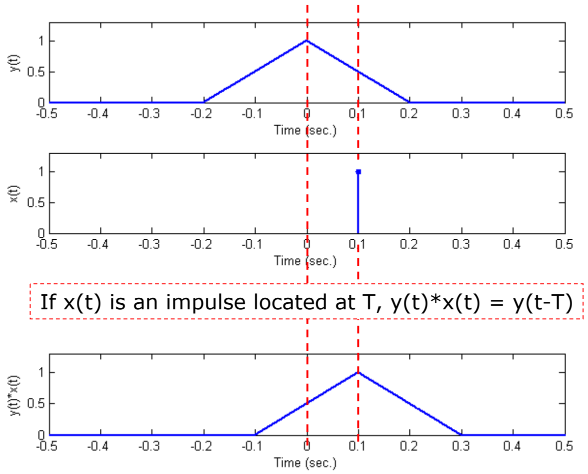
$$x(t) = \begin{cases} 0 & ; \ -\infty < t < 0, \ \ 0.5 \le t < \infty \\ 1 & ; \ 0 \le t < 0.5 \end{cases}$$

$$y(t) = \begin{cases} 0 & ; \ -\infty < t < 0.2, \ \ 0.6 \le t < \infty \\ 1 & ; \ 0.2 \le t < 0.6 \end{cases}$$



$$x(t) * y(t) = \begin{cases} 0 & ; -\infty < t < 0.2 \\ t - 0.2 & ; 0.2 \le t < 0.6 \\ 0.4 & ; 0.6 \le t < 0.7 \\ -t - 1.1 & ; 0.7 \le t < 1.1 \\ 0 & ; 1.1 \le t < \infty \end{cases}$$

Convolution



Fourier Transform

$$x(t) \Leftrightarrow X(f)$$

Inverse Fourier Transform

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df$$

Fourier Transform

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt$$

Magnitude and Phase

Fourier transform can be a complex number.

$$x(t) \Leftrightarrow X(f) = a + jb$$

$$|X(f)| = \sqrt{a^2 + b^2}$$

$$\angle X(f) = \arctan(\frac{b}{a})$$

Fourier Transform of an Impulse

$$x(t) = \delta(t - t_0)$$

$$X(f) = \int_{-\infty}^{\infty} \delta(t - t_0) e^{-j2\pi ft} dt = e^{-j2\pi ft_o}$$

$$\therefore \delta(t - t_0) \Leftrightarrow e^{-j2\pi t_0 f}$$

Magnitude and Phase of Complex Expo.

Useful Relationship

$$e^{jx} = \cos(x) + j\sin(x)$$

$$\cos(x) = \frac{e^{jx} + e^{-jx}}{2}$$

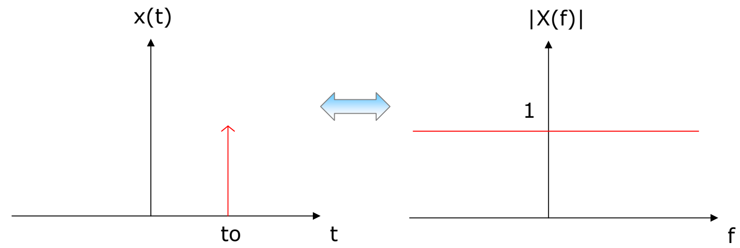
$$\sin(x) = \frac{e^{jx} - e^{-jx}}{j2}$$

$$|e^{jx}| = |\cos(x) + j\sin(x)| = \sqrt{\cos(x)^2 + \sin(x)^2} = 1$$

$$\angle e^{jx} = \angle(\cos(x) + j\sin(x)) = \arctan\left(\frac{\sin(x)}{\cos(x)}\right) = x$$

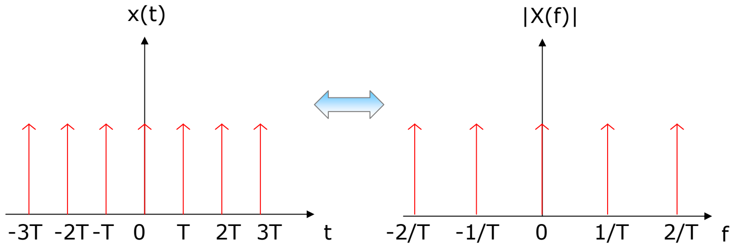
Fourier Transform of an Impulse

$$\delta(t - t_0) \Leftrightarrow e^{-j2\pi t_0 f}$$



Fourier Transform of Impulse Train

$$\sum_{n=-\infty}^{\infty} \delta(t - nT) \Leftrightarrow \frac{1}{T} \sum_{k=-\infty}^{\infty} \delta\left(f - \frac{k}{T}\right)$$



Fourier Transform of Pure Sine

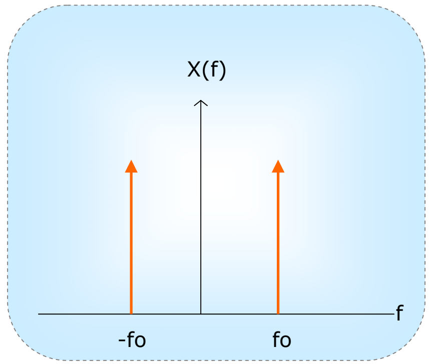
$$x(t) = \cos(2\pi f_0 t)$$

$$X(f) = \int_{-\infty}^{\infty} \cos(2\pi f_0 t) e^{-j2\pi f t} dt$$

$$= \int_{-\infty}^{\infty} \frac{e^{j2\pi f_0 t} + e^{-j2\pi f_0 t}}{2} e^{-j2\pi f t} dt$$

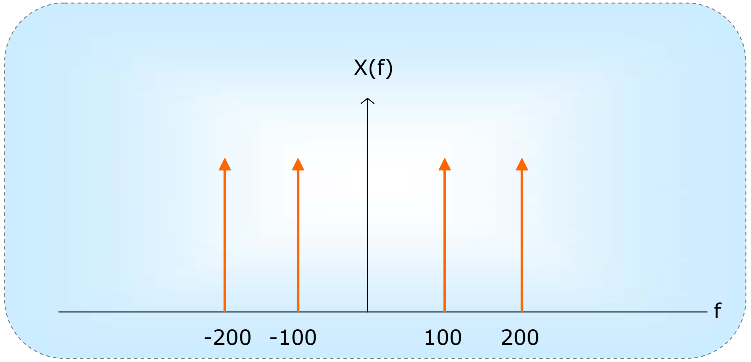
$$= \int_{-\infty}^{\infty} \frac{e^{-j2\pi(f - f_0)t} + e^{-j2\pi(f + f_0)t}}{2} dt$$

$$= \frac{1}{2}\left(\delta(f - f_0) + \delta(f + f_0)\right)$$



Frequency Components

$$x(t) = \cos(2\pi(100)t) + \cos(2\pi(200)t)$$



Department of Computer Engineering, Chulalongkorn University
Spoken Language Systems Research Group

Fourier Series

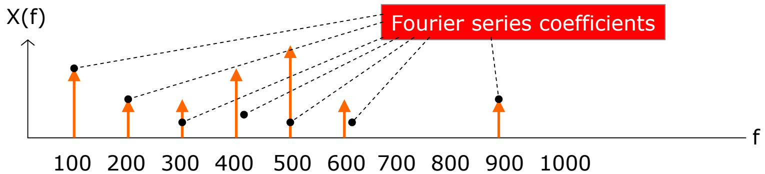
- Linear combination of sinusoidal signals with frequency:

  Fo

  nFo; n = 2,3,4,…

  results in periodic signals.

- Their Fourier transforms are discrete.



2110432 Auto. Speech. Recog. | First Semester 2007 | Lecture 2 & 3
ATIWONG SUCHATO

Department of Computer Engineering, Chulalongkorn University
Spoken Language Systems Research Group

Some Important Properties

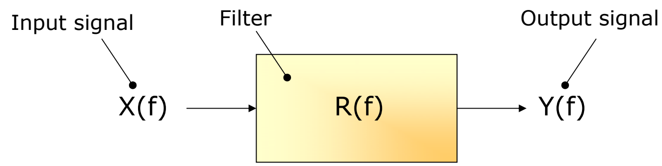
$$x(t) * y(t) \Leftrightarrow X(f)Y(f)$$

$$x(t)y(t) \Leftrightarrow X(f) * Y(f)$$

2110432 Auto. Speech. Recog. | First Semester 2007 | Lecture 2 & 3
ATIWONG SUCHATO

Department of Computer Engineering, Chulalongkorn University
Spoken Language Systems Research Group

Linear System



Impulse Response → response of the system when the input is an impulse.

Transfer Function (R(f))

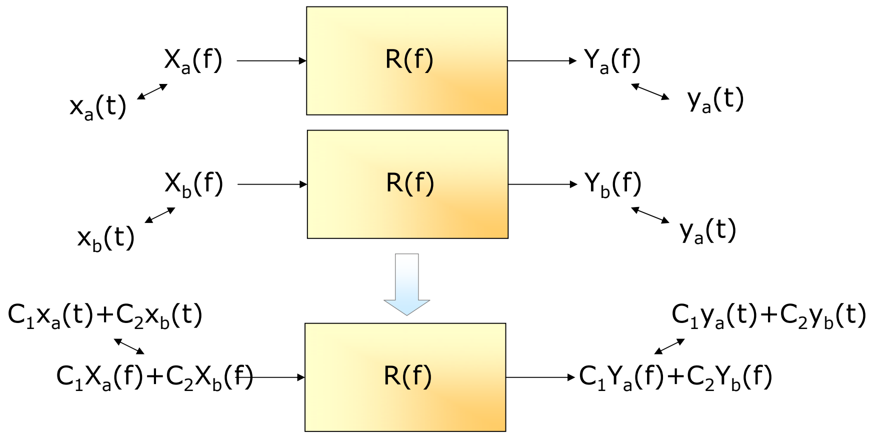
→ R(f) = Y(f)/X(f)

→ R(f) is the Fourier transform of the impulse response of the filter.

$$Y(f) = X(f)R(f) \Leftrightarrow y(t) = x(t) * r(t)$$

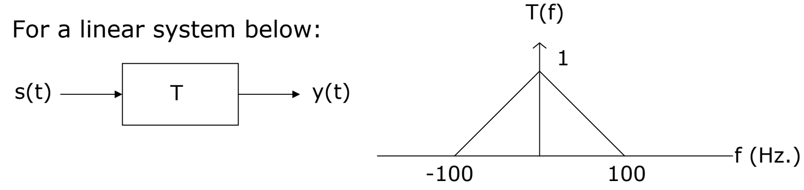
2110432 Auto. Speech. Recog. | First Semester 2007 | Lecture 2 & 3
ATIWONG SUCHATO

Linear System



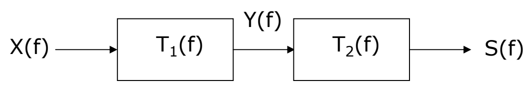
Example

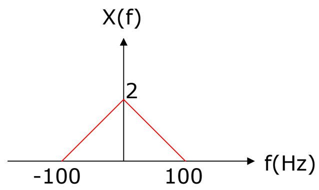
- Find the Fourier transform of:
  $s(t) = 3\cos(2\pi(50)t) - \sin(2\pi(100)t)$
- For a linear system below:

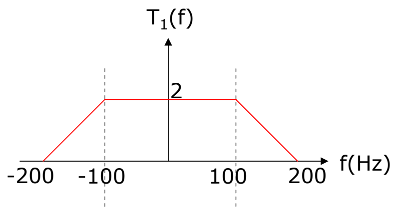


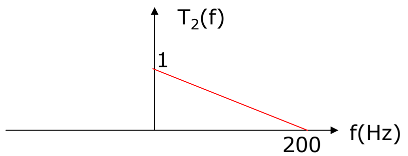
Find y(t).

Example









Sketch S(f)

Window Effect

$x(t) = \cos(2\pi f_0 t)$ → x(t) has infinite length.

But in real life, we can never have a signal of infinite length.

A portion of sine wave with finite length can be considered as a windowed version of the infinite length sine wave.

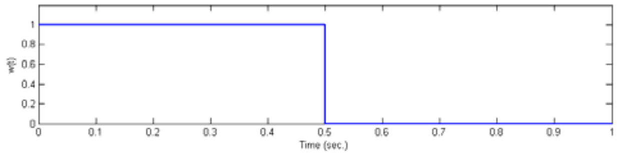
E.g.: A sine wave from t1 to t2, $x_w(t)$ can be written as:

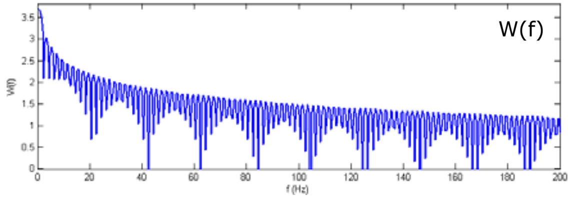
$$x_w(t) = \cos(2\pi f_0 t) w(t)$$

Where: $$w(t) = \begin{cases} 1 & ; t_1 \le t < t_2 \\ 0 & ; otherwise \end{cases}$$

Square Window

$$w(t)=\begin{cases}1 & ;0\le t<0.5\,\text{sec.}\\ 0 & ;otherwise\end{cases}$$

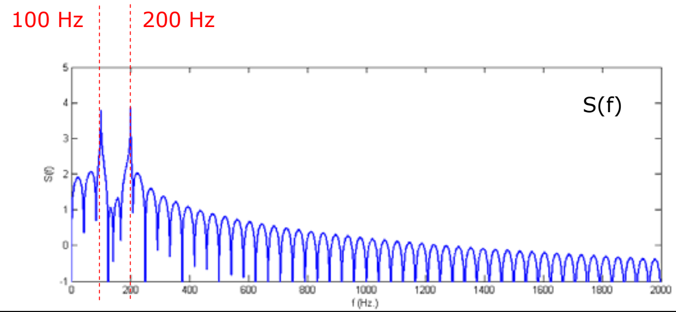




Window Effect

$$s(t)=(\sin(2\pi(100)t)+\sin(2\pi(200)t))w(t)$$

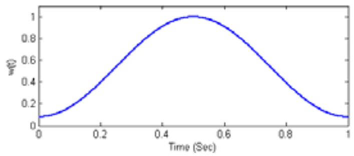
w(t) is a 1-second-long square window

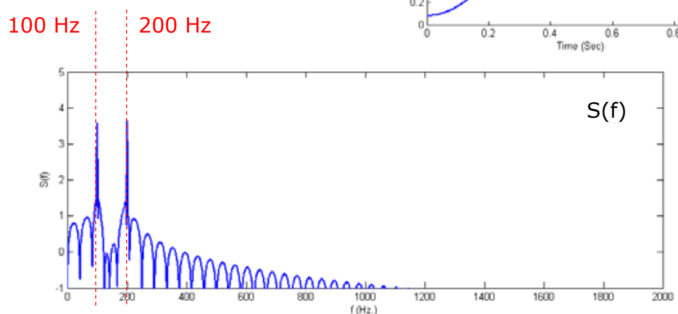


Window Effect

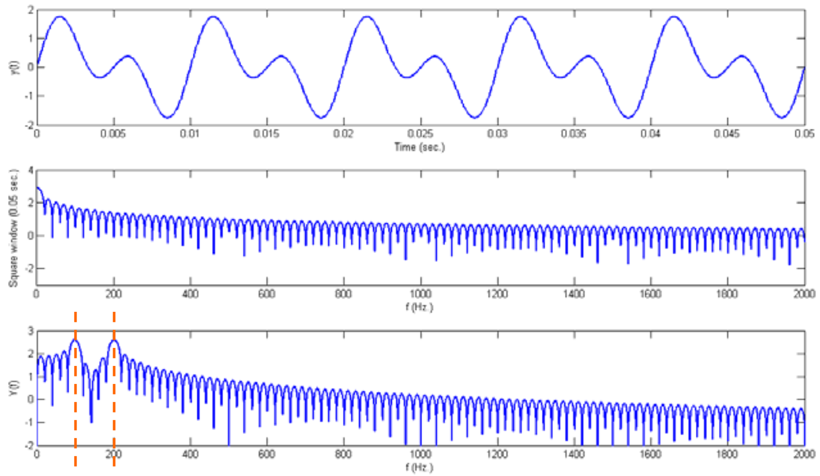
$$s(t)=(\sin(2\pi(100)t)+\sin(2\pi(200)t))w(t)$$

w(t) is a 1-second-long hamming window

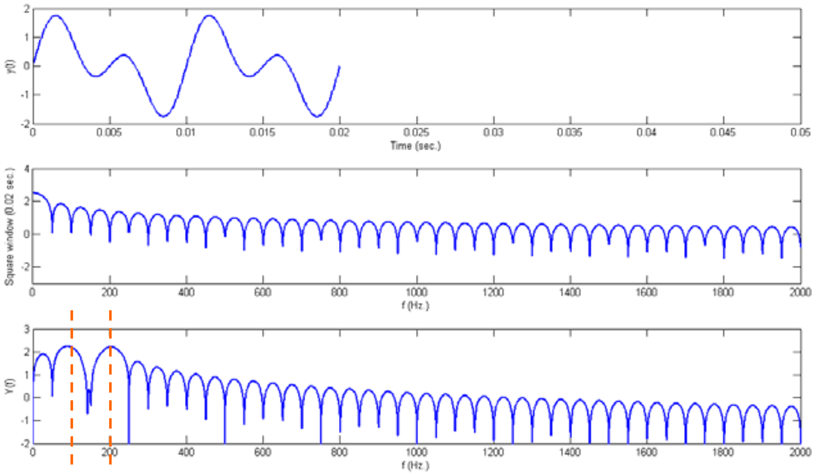




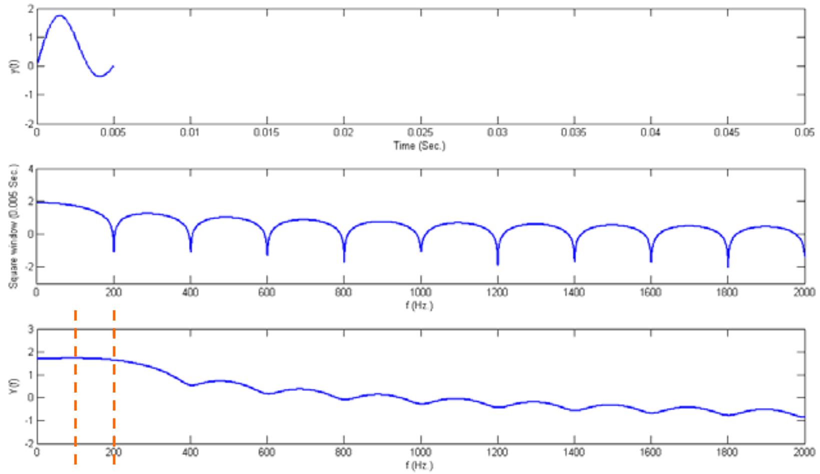
Window Length



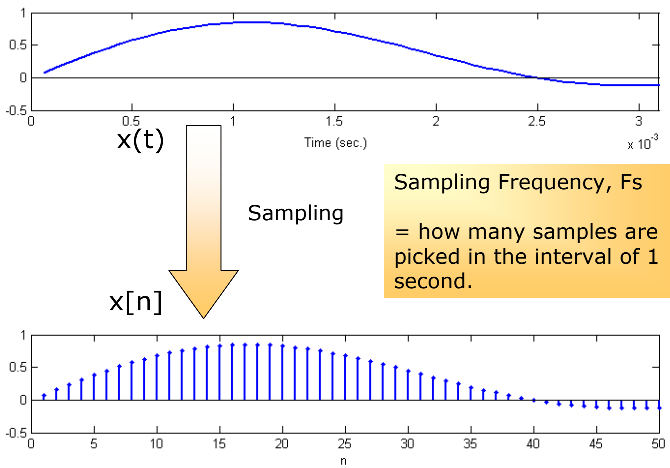
Window Length



Window Length



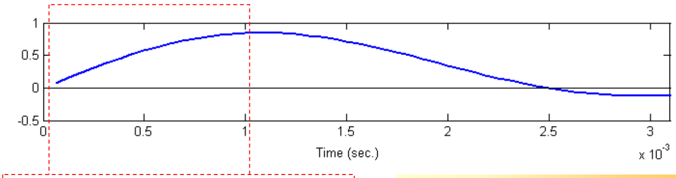
Discrete-time Signal



Sampling Frequency, Fs

= how many samples are picked in the interval of 1 second.

Sampling

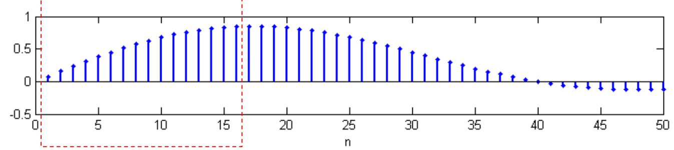


In 1 x 10-3 sec.
16 samples are picked

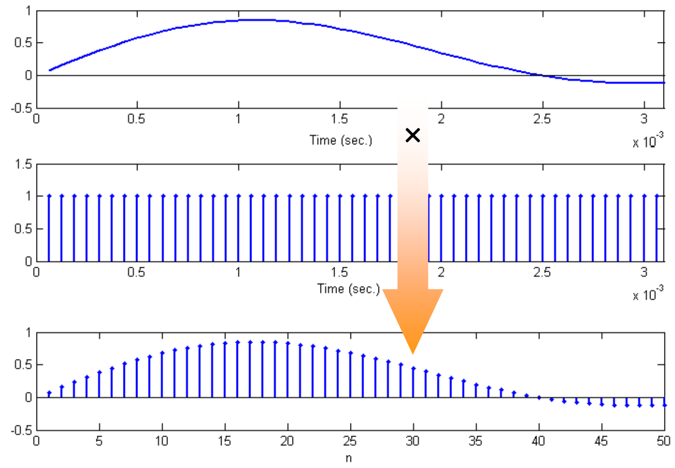
∴Fs = 16,000 samples/sec.
= 16 kHz.

Sampling Frequency, Fs

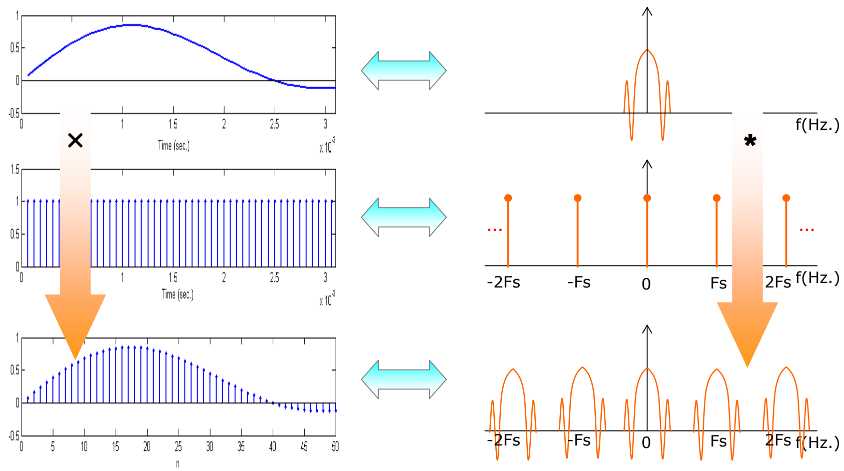
= how many samples are picked in the interval of 1 second.



Sampling



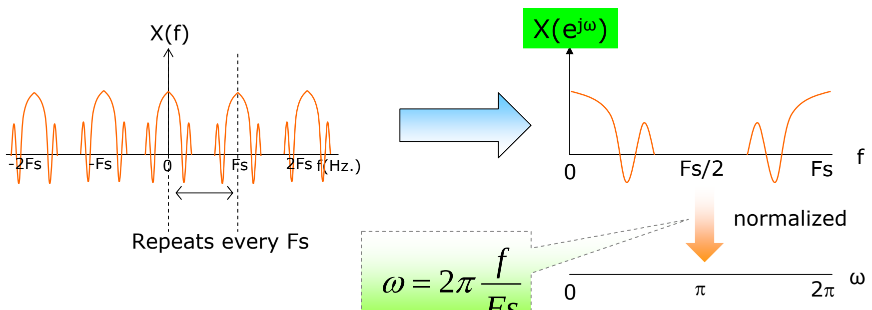
Sampling



Normalized Frequency

Fourier Transform

Discrete-time Fourier Transform



$$\omega = 2\pi \frac{f}{Fs}$$

f : frequency (Hz.)

ω : Normalized frequency (Radian per sample)

Discrete-time Fourier Transform

$$x[n] \Leftrightarrow X(e^{j\omega})$$

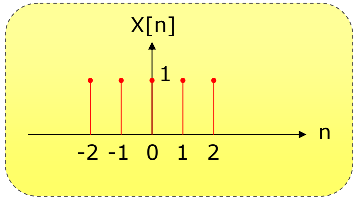
Discrete-time Fourier Transform

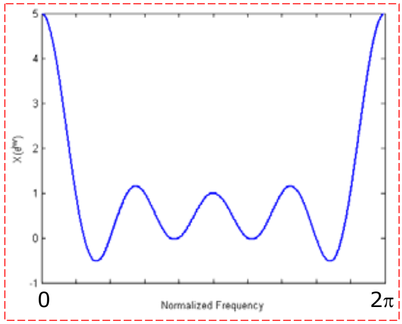
$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n] e^{-j\omega n}$$

Discrete-time Fourier Transform

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega n}$$

Example





$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n}$$

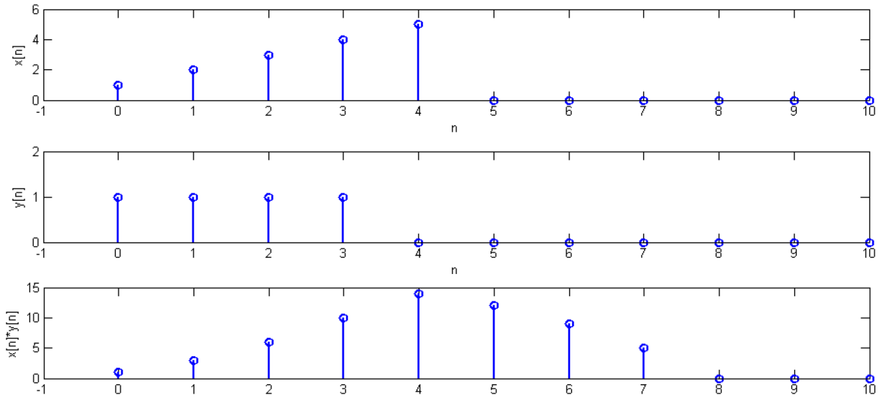
$$= x[-2]e^{-j\omega(-2)} + x[-1]e^{-j\omega(-1)} + x[0]e^{-j\omega(0)} + x[1]e^{-j\omega(1)} + x[2]e^{-j\omega(2)}$$

$$= e^{j\omega(2)} + e^{j\omega} + 1 + e^{-j\omega} + e^{-j\omega(2)}$$

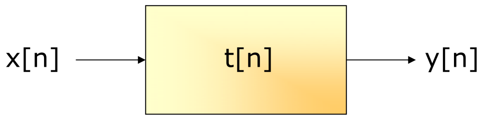
$$= 1 + 2\cos(\omega) + 2\cos(2\omega)$$

Discrete-time Convolution

$$x[n] * y[n] = \sum_{m=-\infty}^{\infty} x[m]y[n-m]$$



Discrete-time Filter



$$Y(e^{j\omega}) = X(e^{j\omega})T(e^{j\omega}) \Leftrightarrow y[n] = x[n] * t[n]$$

Family of Fourier Transform

- "Discreteness" in one domain implies "Periodicity" in the other domain.
- "Continuity" in one domain implies "Aperiodicity" in the other domain.

Transform	Time	Freq.
Cont. Fourier Trans.	Cont. & aperiodic	Cont. & aperiodic
Fourier Series	Cont. & periodic	Disc. & aperiodic
DTFT	Disc. & aperiodic	Cont. & periodic
DFT	Disc. & periodic	Disc & aperiodic

Discrete Fourier Transform

$$x[n] \Leftrightarrow X[k]$$

DFT

$$X[k]=\sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi kn}{N}} \qquad k=0,1,\dots,N-1$$

IDFT

$$X[n]=\frac{1}{N}\sum_{k=0}^{N-1} x[k]e^{j\frac{2\pi kn}{N}} \qquad n=0,1,\dots,N-1$$

Discrete Fourier Transform

$$X(f)=\int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt$$

$$\omega = 2\pi\frac{f}{Fs}$$

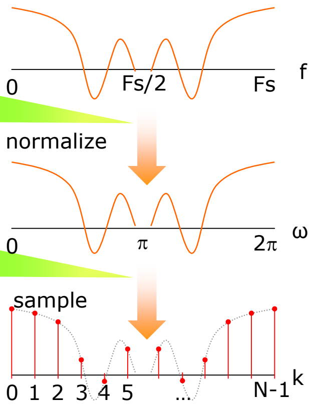
normalize

$$X(e^{j\omega})=\sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n}$$

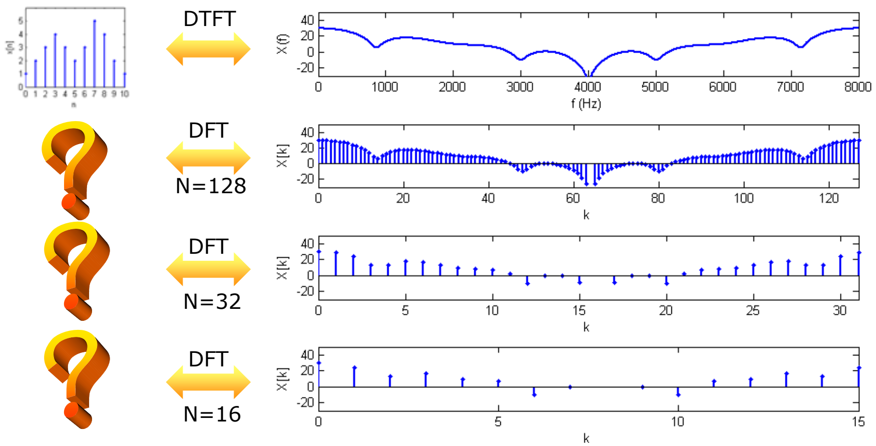
$$k=\frac{\omega}{2\pi}N$$

sample

$$X[k]=\sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi kn}{N}}$$



N DFT



Matlab Demo

- Basic operations
  - Variables
  - Colon operator
  - Matrix manipulation
- Arrays and Cells
- Plotting
- Loading/Saving MS wav
- Help & lookfor
- Control structures