Speech (Signal) Processing (Part IV)

CSCI B659: DEEP LEARNING FOR SPEECH PROCESSING

SEPTEMBER 10, 2019

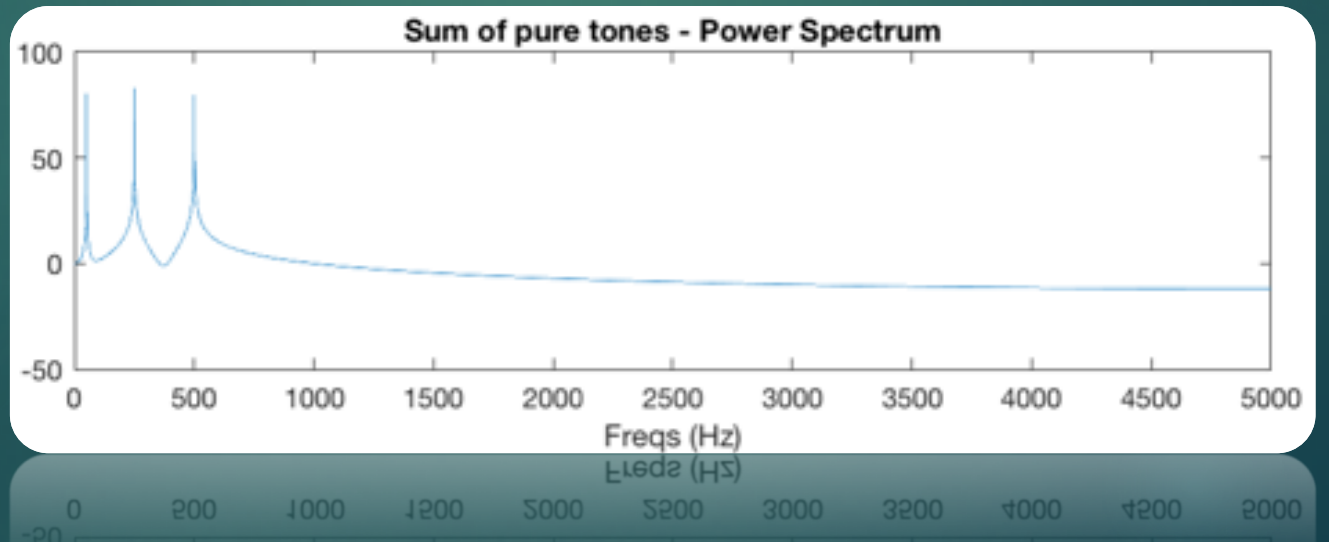
LECTURE #5

Learning Objectives

- **You the student will be able to:**
  - Understand the concept of filters and filterbanks
  - Understand windowing functions (T-F resolution)
  - Describe impulse response and convolution
  - Understand conversion from T-F to time domain (Filtering view)

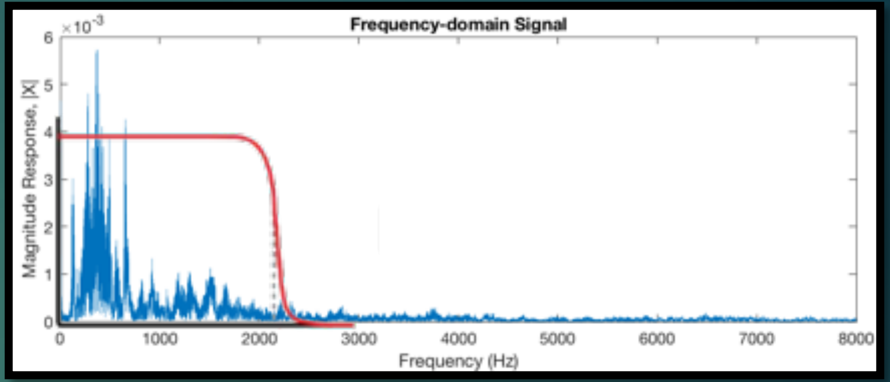
Filters

- Filters are used to process (and modify) signals in the (time-) frequency domain
- Often times it is desired to isolate certain frequencies
  - E.g. removal of narrowband noise
  - Can isolate a single frequency or a band of frequencies (i.e. subband)

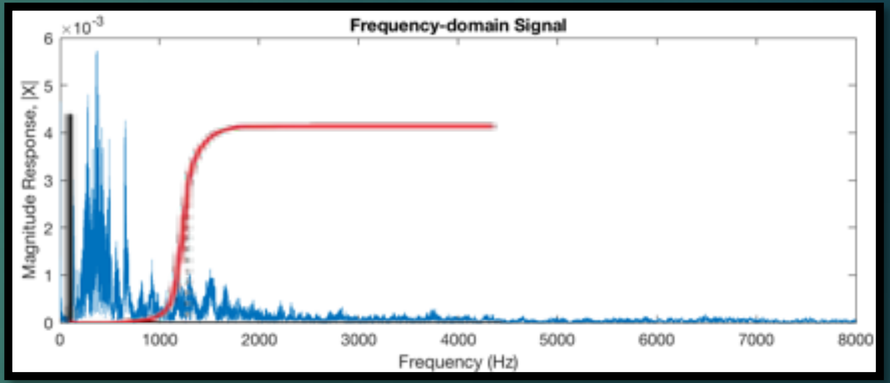


Common Filters

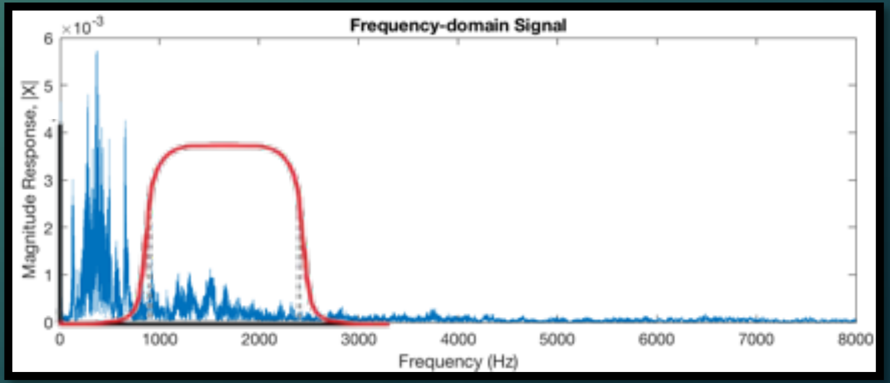
- Low Pass Filter
  - Retains low frequency content



- High Pass Filter
  - Retains high frequency content



- Band Pass Filter
  - Retains content that falls within a frequency band (or range)



DEMO

Filtering: Frequency Domain

- *H[k]* filters the frequency response of a signal *X[k]*

Mathematical representation:

$$Y[k] = H[k]X[k]$$

$X[k]$: Input Signal
$H[k]$: Filter frequency response
$Y[k]$: Filtered signal (or filter output)

Block Diagram:

Filtering: Time domain

- Filtering in the time domain is performed using ***convolution***

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

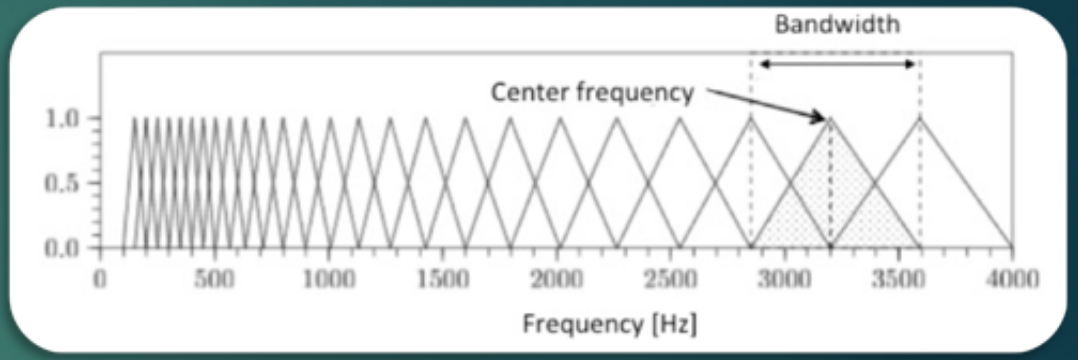
$x[n]$: input time-domain signal
$h[n]$: impulse response
$y[n]$: output time-domain signal

- Each sample of *h* must be multiplied against each value of x, for all values of *n*
  - This leads to a large number of multiplications and additions
  - Hence, filtering is usually performed in the frequency domain

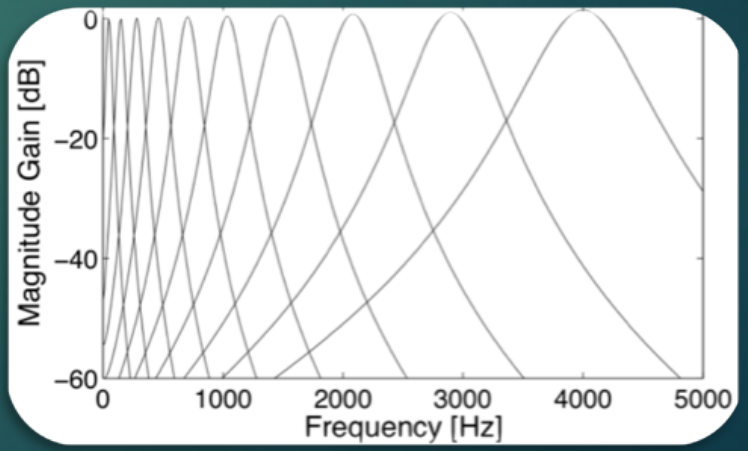
Filterbanks

- A set of filters
- Common filter banks
  - Mel scale
  - Gammatone
    - Model frequency processing of the human ear
  - More on these later
- At a given time, results in *M* frequency outputs
  - *M* is the number of filters in the filter bank
  - Each filter is separately applied to the input signal

Mel-scale Filterbank



Gammatone Filterbank



Filtering (Convolutional) View of STFT

- STFT calculation looks a lot like convolution between x*[n]* and w*[n]*, with an extra exponential term

STFT

$$X[n,k] = \sum_{m=-\infty}^{\infty} x[m]w[n-m]e^{\frac{-j2\pi km}{N}}$$

$$= \sum_{m=-\infty}^{\infty} \left(x[m]e^{-j\frac{2\pi}{N}km}\right) w[n-m]$$

$$= w[n] * \left(x[n]e^{-j\frac{2\pi}{N}kn}\right)$$

Convolution

$$y[n] = h[n] * x[n]$$

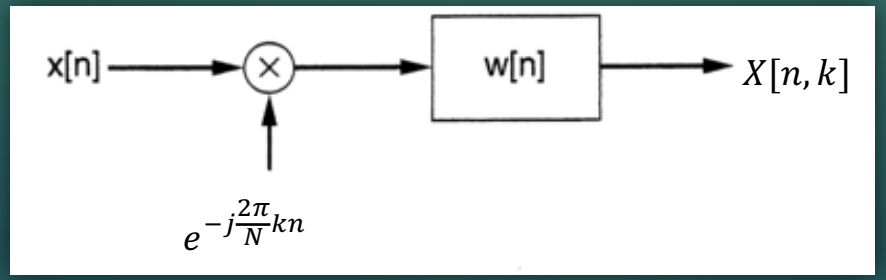
$$= \sum_{k=-\infty}^{\infty} x[k]h[n-k]$$

Filtering View of STFT (cont.)

- STFT calculation looks a lot like convolution between *x[n]* and *w[n]*, with an extra exponential term

$$\underline{\text{STFT}}$$
$$X[n,k] = w[n] * \left(x[n]e^{-j\frac{2\pi}{N}kn}\right)$$

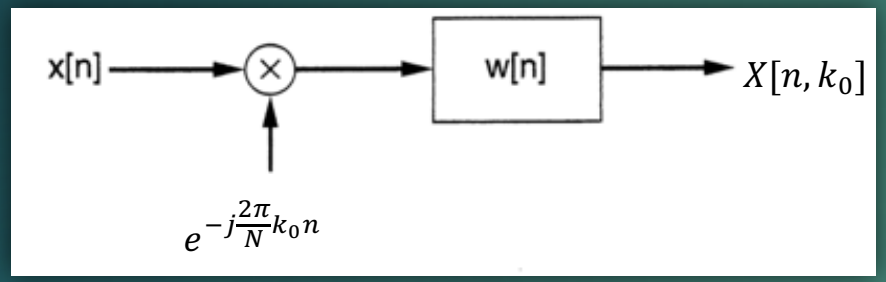
- Block diagram model of STFT calculation



w[n] is a filter

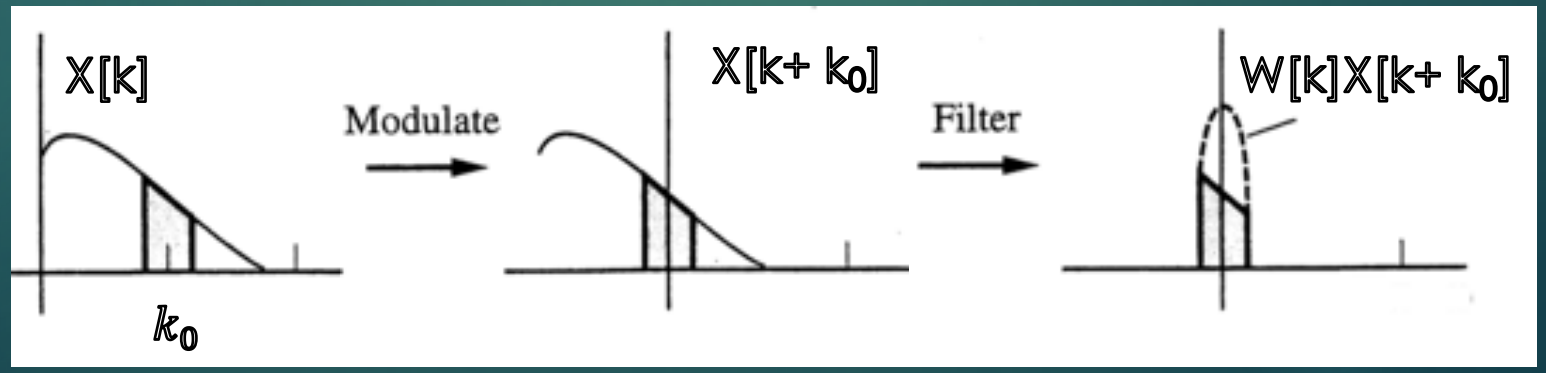
Filtering View of STFT (cont.)

- Calculations in Time domain: $w[n] * \left(x[n]e^{-j\frac{2\pi}{N}kn}\right)$



Property	Time Domain	Frequency Domain
Modulation	$e^{-j\frac{2\pi k_0}{N}n}x[n]$	$X[k+k_0]$
Convolution	$x_1[n] * x_2[n]$	$X_1[k]X_2[k]$

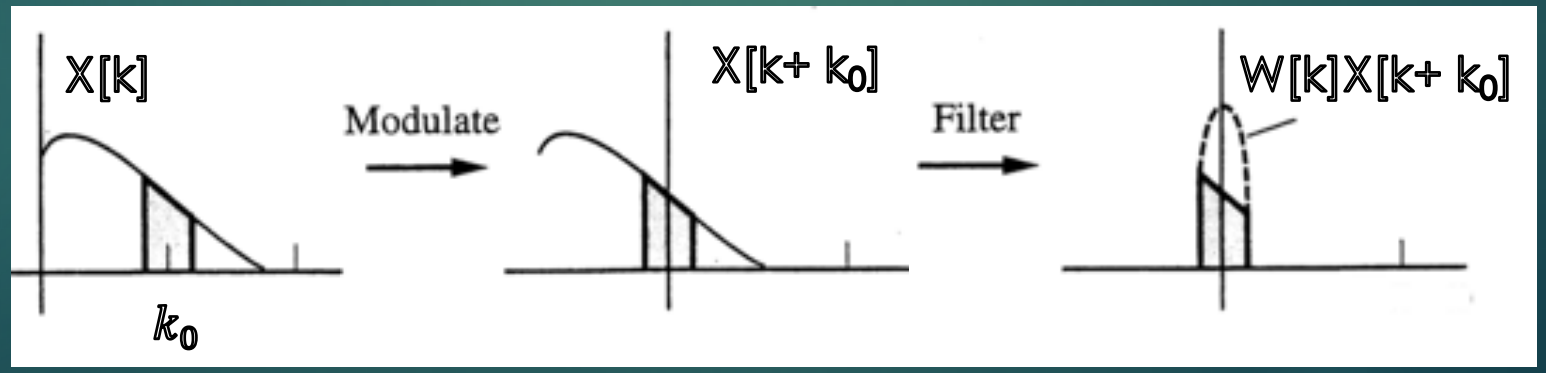
- Calculations in Frequency domain: $W[k]X[k+k_0]$



Notes

- We notice the following:
  - The spectral shape is based on the frequency response of the windowing function
  - Everything is centered around ***baseband*** (0 Hz)

- The filter (window) response is based on the selected filter
  - This also impacts T-F resolution

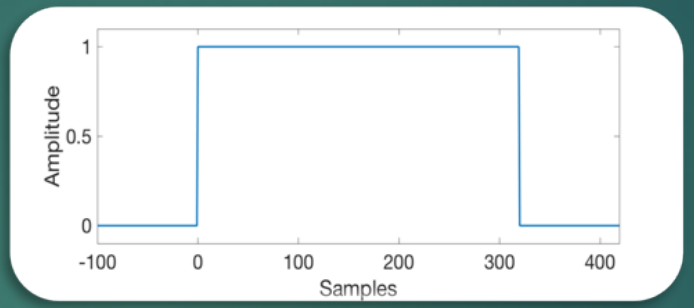


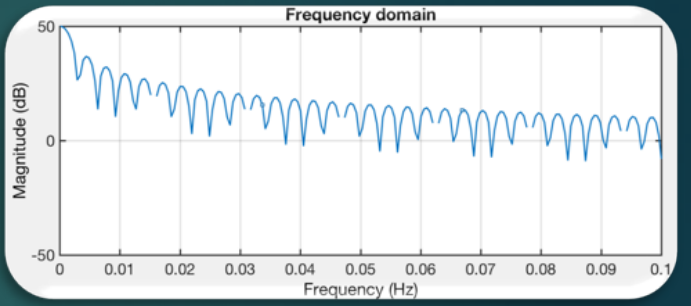
Analysis Windowing Functions: *w[n]*

- There are several windowing functions that can be used
- Each window has it's pros and cons

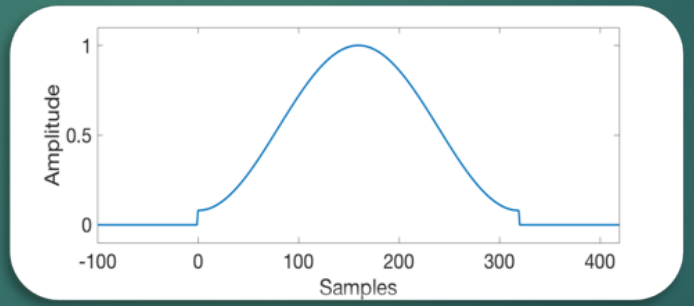
Time | **Frequency**

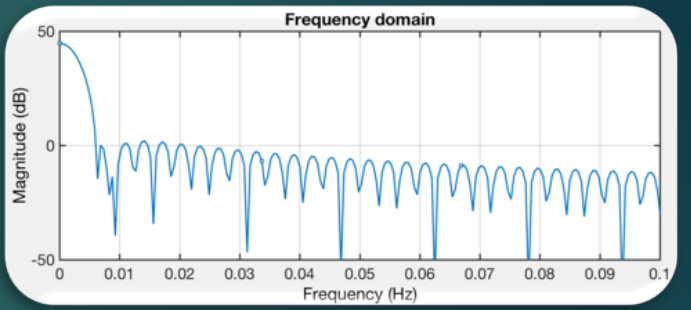
Rectangular



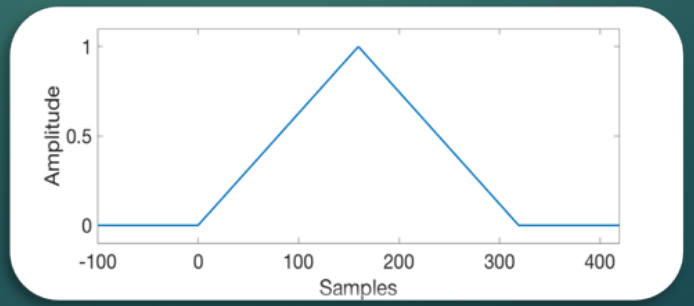


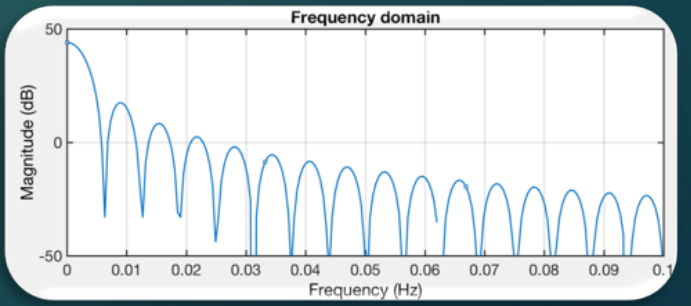
Hamming



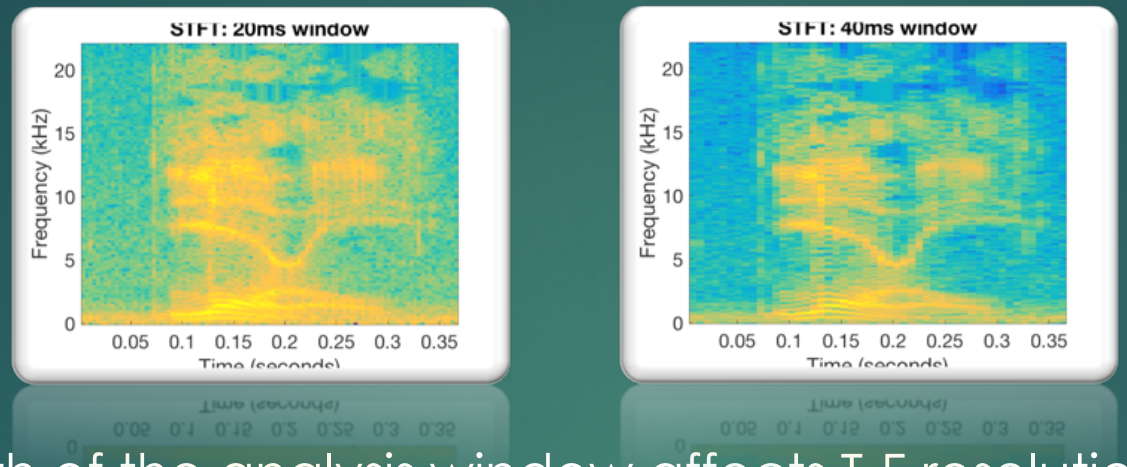


Bartlett





Revisit: T-F Resolution Tradeoff



- The length of the analysis window affects T-F resolution
  - ***Long window length***: Fine frequency detail, poor temporal structure
  - ***Short window length***: Poor frequency detail, fine temporal structure

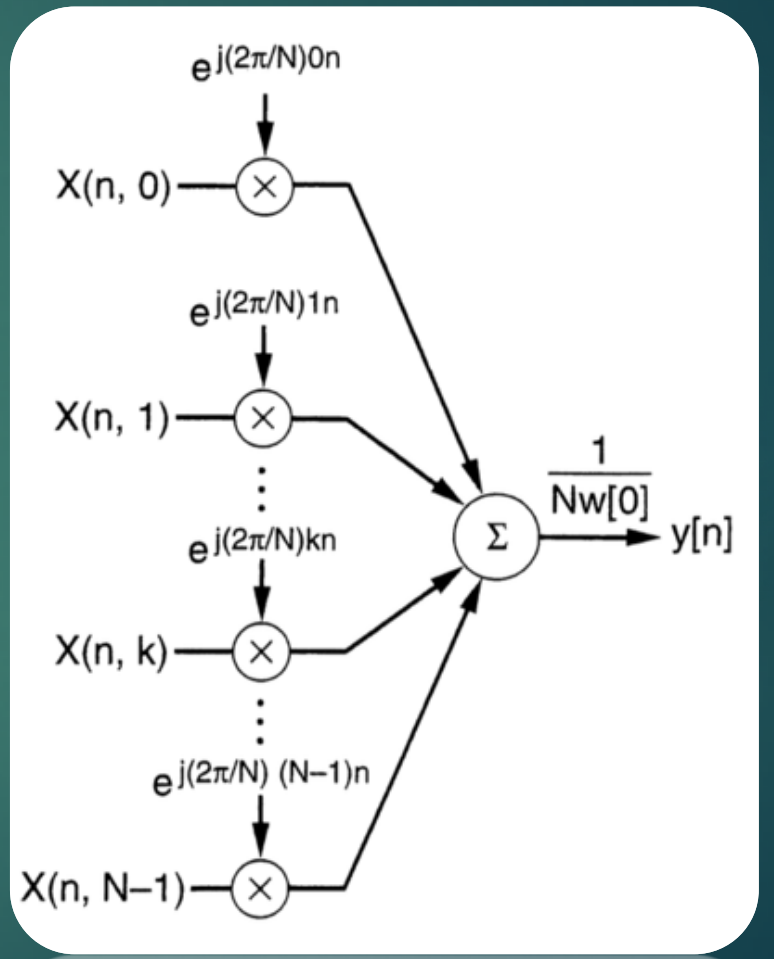
- This is based directly on the type of windowing function and its corresponding length (see board)
  - Long window in time => short bandwidth in frequency
  - Short window in time => long bandwidth in frequency

Synthesis: T-F to Time domain

Filterbank Summation (FBS)

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

- Filtering view approach
- Adding frequency components for each *n*
- Steps:
  - Demodulate *X[n,k]*
  - Sum over all *k*
  - Scale by a factor
  - Repeat for all *n*



STFT and Analysis Window

- Analysis window impacts frequency response (and time signal)
  - X[n,k] depends on main and sidelobes of the windowing frequency response
  - This means that more information may be present than desired
- See whiteboard example

Next Class

- Topic:
  - Supervised Learning