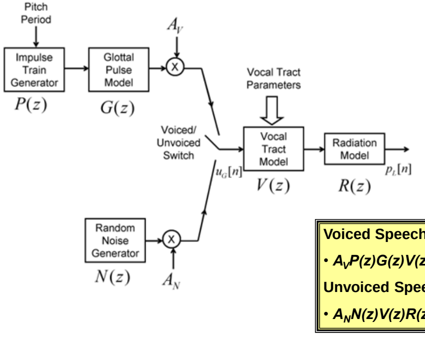
Digital Speech Processing— Lecture 9

Short-Time Fourier Analysis Methods- Introduction

General Discrete-Time Model of Speech Production



Voiced Speech:

- ***A_VP(z)G(z)V(z)R(z)***

Unvoiced Speech:

- ***A_NN(z)V(z)R(z)***

Short-Time Fourier Analysis

- represent signal by **sum of sinusoids** or complex exponentials as it leads to convenient solutions to problems (formant estimation, pitch period estimation, analysis-by-synthesis methods), and insight into the signal itself
- such **Fourier representations** provide
  - convenient means to determine response to a sum of sinusoids for linear systems
  - clear evidence of signal properties that are obscured in the original signal

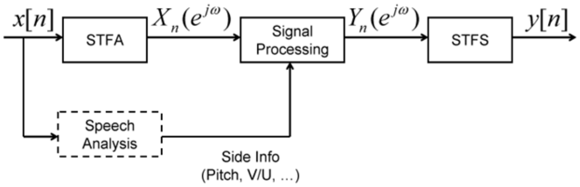
Why STFT for Speech Signals

- steady state sounds, like vowels, are produced by **periodic excitation of a linear system** => speech spectrum is the product of the excitation spectrum and the vocal tract frequency response
- speech is a **time-varying signal** => need more sophisticated analysis to reflect time varying properties
  - changes occur at syllabic rates (~10 times/sec)
  - over fixed time intervals of 10-30 msec, properties of most speech signals are relatively constant (when is this not the case)

Overview of Lecture

- define **time-varying Fourier transform** (*STFT*) analysis method
- define **synthesis method** from time-varying FT (filter-bank summation, overlap addition)
- show how time-varying FT can be viewed in terms of a **bank of filters model**
- **computation methods** based on using FFT
- **application** to vocoders, spectrum displays, format estimation, pitch period estimation

Frequency Domain Processing



- **Coding**:
  - transform, subband, homomorphic, channel vocoders
- **Restoration/Enhancement/Modification**:
  - noise and reverberation removal, helium restoration, time-scale modifications (speed-up and slow-down of speech)

Frequency and the *DTFT*

- sinusoids

$$x(n) = \cos(\omega_0 n) = (e^{j\omega_0 n} + e^{-j\omega_0 n})/2$$

where ω_0 is the *frequency* (in radians) of the sinusoid

- the Discrete-Time Fourier Transform (*DTFT*)

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

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

where ω is the *frequency variable* of $X(e^{j\omega})$

DTFT and DFT of Speech

- The DTFT and the DFT for the infinite duration signal could be calculated (the DTFT) and approximated (the DFT) by the following:

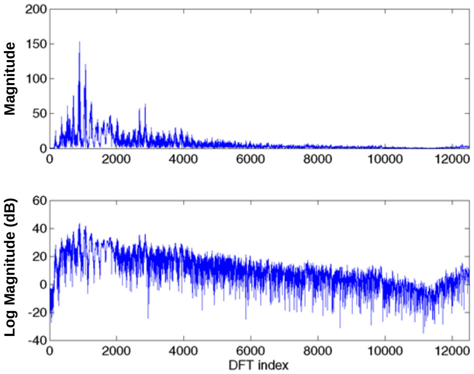
$$X(e^{j\omega}) = \sum_{m=-\infty}^{\infty} x(m)e^{-j\omega m} \quad (DTFT)$$

$$X(k) = \sum_{m=0}^{L-1} x(m)w(m)e^{-j(2\pi/L)km}, \quad k = 0,1,\ldots,L-1$$

$$= X(e^{j\omega})\Big|_{\omega=(2\pi k/L)} \quad (DFT)$$

- using a value of L=25000 we get the following plot

25000-Point DFT of Speech



Short-Time Fourier Transform (STFT)

Short-Time Fourier Transform

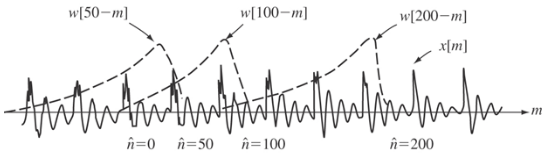
- speech is not a **stationary signal**, i.e., it has properties that **change with time**
- thus a **single representation** based on all the samples of a speech utterance, for the most part, has no meaning
- instead, we define a **time-dependent Fourier transform** (TDFT or STFT) of speech that changes periodically as the speech properties change over time

Definition of STFT

$$\boxed{X_{\hat{n}}(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} x(m)w(\hat{n}-m)e^{-j\hat{\omega}m}}$$

both $\hat{n}$ and $\hat{\omega}$ are variables

- $w(\hat{n}-m)$ is a real window which determines the portion of $x(\hat{n})$ that is used in the computation of $X_{\hat{n}}(e^{j\hat{\omega}})$

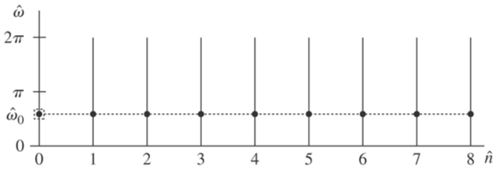


Short Time Fourier Transform

- STFT is a function of two variables, the time index, $\hat{n}$, which is discrete, and the frequency variable, $\hat{\omega}$, which is continuous

$$X_{\hat{n}}(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} x(m)w(\hat{n}-m)e^{-j\hat{\omega}m}$$

$$= DTFT\,(x(m)w(\hat{n}-m)) \Rightarrow \hat{n} \text{ fixed, } \hat{\omega} \text{ variable}$$



Short-Time Fourier Transform

- alternative form of STFT (based on change of variables) is

$$X_{\hat{n}}(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} w(m)\,x(\hat{n}-m)e^{-j\hat{\omega}(\hat{n}-m)}$$

$$= e^{-j\hat{\omega}\hat{n}} \sum_{m=-\infty}^{\infty} x(\hat{n}-m)w(m)e^{j\hat{\omega}m}$$

- if we define

$$\tilde{X}_{\hat{n}}(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} x(\hat{n}-m)w(m)e^{j\hat{\omega}m}$$

- then $X_{\hat{n}}(e^{j\hat{\omega}})$ can be expressed as (using $m' = -m$)

$$X_{\hat{n}}(e^{j\hat{\omega}}) = e^{-j\hat{\omega}\hat{n}}\tilde{X}_{n}(e^{j\hat{\omega}}) = e^{-j\hat{\omega}\hat{n}}\, DTFT\left[x(\hat{n}+m)w(-m)\right]$$

STFT-Different Time Origins

- the STFT can be viewed as having two different time origins
  1. time origin tied to signal $x(n)$

$$X_{\hat{n}}(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} x(m)w(\hat{n}-m)e^{-j\hat{\omega}m}$$

$$= DTFT\left[x(m)w(\hat{n}-m)\right], \quad \hat{n} \text{ fixed, } \hat{\omega} \text{ variable}$$

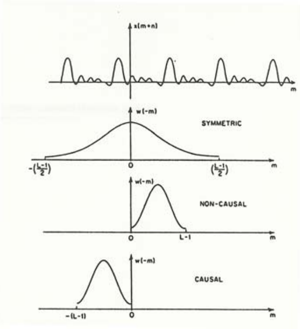
  2. time origin tied to window signal $w(-m)$

$$X_{\hat{n}}(e^{j\hat{\omega}}) = e^{-j\hat{\omega}\hat{n}} \sum_{m=-\infty}^{\infty} x(\hat{n}+m)w(-m)e^{-j\hat{\omega}m}$$

$$= e^{-j\hat{\omega}\hat{n}}\tilde{X}(e^{j\hat{\omega}})$$

$$= e^{-j\hat{\omega}\hat{n}}\ DTFT\left[w(-m)x(\hat{n}+m)\right], \quad \hat{n} \text{ fixed, } \hat{\omega} \text{ variable}$$

Time Origin for STFT



$m = -\hat{n} \Rightarrow x[0]$

Time origin tied to window $w[-m]\,x[\hat{n}+m]$

Interpretations of STFT

- there are 2 distinct interpretations of $X_{\hat{n}}(e^{j\hat{\omega}})$

1. assume $\hat{n}$ is fixed, then $X_{\hat{n}}(e^{j\hat{\omega}})$ is simply the normal Fourier transform of the sequence $w(\hat{n}-m)\,x(m),\ -\infty < m < \infty$ => for fixed $\hat{n}$, $X_{\hat{n}}(e^{j\hat{\omega}})$ has the same properties as a normal Fourier transform
2. consider $X_{\hat{n}}(e^{j\hat{\omega}})$ as a function of the time index $\hat{n}$ with $\hat{\omega}$ fixed. Then $X_{\hat{n}}(e^{j\hat{\omega}})$ is in the form of a convolution of the signal $x(\hat{n})e^{-j\hat{\omega}\hat{n}}$ with the window $w(\hat{n})$. This leads to an interpretation in the form of linear filtering of the frequency modulated signal $x(\hat{n})e^{-j\hat{\omega}\hat{n}}$ by $w(\hat{n})$.

- we will now consider each of these interpretations of the STFT in a lot more detail

Fourier Transform Interpretation

- consider $X_{\hat{n}}(e^{j\hat{\omega}})$ as the normal Fourier transform of the sequence $w(\hat{n}-m)\,x(m), -\infty < m < \infty$ for fixed $\hat{n}$.
- the window $w(\hat{n}-m)$ slides along the sequence $x(m)$ and defines a new STFT for every value of $\hat{n}$
- what are the conditions for the existence of the STFT
  - the sequence $w(\hat{n}-m)\,x(m)$ must be absolutely summable for all values of $\hat{n}$
    - since $|\,x(\hat{n})\,| \leq L$ (32767 for 16-bit sampling)
    - since $|\,w(\hat{n})\,| \leq 1$ (normalized window levels)
    - since window duration is usually finite
  - $w(\hat{n}-m)\,x(m)$ is absolutely summable for all $\hat{n}$

Frequencies for STFT

- the STFT is periodic in ω with period 2π, i.e.,

  $X_{\hat{n}}(e^{j\hat{\omega}}) = X_{\hat{n}}(e^{j(\hat{\omega}+2\pi k)}), \ \forall k$
- can use any of several frequency variables to express STFT, including

  --$\hat{\omega} = \hat{\Omega}T$ (where T is the sampling period for $x(m)$) to represent analog radian frequency, giving $X_{\hat{n}}(e^{j\hat{\Omega}T})$

  --$\hat{\omega} = 2\pi\hat{f}$ or $\hat{\omega} = 2\pi\hat{F}T$ to represent normalized frequency ($0 \le \hat{f} \le 1$) or analog frequency ($0 \le \hat{F} \le F_s = 1/T$), giving $X_{\hat{n}}(e^{j2\pi\hat{f}})$ or $X_{\hat{n}}(e^{j2\pi\hat{F}T})$

Signal Recovery from STFT

- since for a given value of $\hat{n}, X_{\hat{n}}(e^{j\hat{\omega}})$ has the same properties as a normal Fourier transform, we can recover the input sequence exactly
- since $X_{\hat{n}}(e^{j\hat{\omega}})$ is the normal Fourier transform of the windowed sequence $w(\hat{n}-m)x(m)$, then

$$w(\hat{n}-m)x(m) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X_{\hat{n}}(e^{j\hat{\omega}})e^{j\hat{\omega}m}d\hat{\omega}$$

- assuming the window satisfies the property that $w(0) \neq 0$ (a trivial requirement), then by evaluating the inverse Fourier transform when $m = \hat{n}$, we obtain

$$x(\hat{n}) = \frac{1}{2\pi w(0)}\int_{-\pi}^{\pi} X_{\hat{n}}(e^{j\hat{\omega}})e^{j\omega\hat{n}}d\hat{\omega}$$

Signal Recovery from STFT

$$x(\hat{n}) = \frac{1}{2\pi w(0)}\int_{-\pi}^{\pi} X_{\hat{n}}(e^{j\hat{\omega}})e^{j\omega\hat{n}}d\hat{\omega}$$

- with the requirement that $w(0) \neq 0$, the sequence $x(\hat{n})$ can be recovered exactly from $X_{\hat{n}}(e^{j\hat{\omega}})$, if $X_{\hat{n}}(e^{j\hat{\omega}})$ is known for all values of $\hat{\omega}$ over one complete period

  - sample-by-sample recovery process
  - $X_{\hat{n}}(e^{j\hat{\omega}})$ must be known for every value of $\hat{n}$ and for all $\hat{\omega}$

□can also recover sequence $w(\hat{n}-m)x(m)$ but can't guarantee that $x(m)$ can be recovered since $w(\hat{n}-m)$ can equal 0

Properties of STFT

$X_{\hat{n}}(e^{j\hat{\omega}}) = DTFT[w(\hat{n}-m)x(m)]$ $\quad \hat{n}$ fixed, $\hat{\omega}$ variable

- relation to short-time power density function

$$S_{\hat{n}}(e^{j\hat{\omega}}) = |X_{\hat{n}}(e^{j\hat{\omega}})|^2 = X_{\hat{n}}(e^{j\hat{\omega}}) \cdot X^*_{\hat{n}}(e^{j\hat{\omega}}) = DTFT\left[R_{\hat{n}}(k)\right] \quad \hat{n} \text{ fixed}$$

$$R_{\hat{n}}(k) = \sum_{m=-\infty}^{\infty} w(\hat{n}-m)x(m)w(\hat{n}-m-k)x(m+k) \Leftrightarrow S_{\hat{n}}(e^{j\hat{\omega}})$$

- Relation to regular $X(e^{j\hat{\omega}})$ (assuming it exists)

$$X(e^{j\hat{\omega}}) = DTFT\left[x(m)\right] = \sum_{m=-\infty}^{\infty} x(m)e^{-j\hat{\omega}m}$$

$$X_{\hat{n}}(e^{j\hat{\omega}}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W(e^{-j\theta})X(e^{j(\hat{\omega}-\theta)})e^{-j\theta\hat{n}}\,d\theta$$

$$\uparrow$$

$$\left[w(\hat{n}-m)□x(m) \leftrightarrow W(e^{-j\theta})e^{-j\theta\hat{n}} * X(e^{j\theta})\right]$$

Properties of STFT

- assume $X(e^{j\hat{\omega}})$ exists

$$X(e^{j\hat{\omega}}) = DTFT\left[x(m)\right] = \sum_{m=-\infty}^{\infty} x(m)e^{-j\hat{\omega}m}$$

$$X_{\hat{n}}(e^{j\hat{\omega}}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W(e^{-j\theta})X(e^{j(\hat{\omega}-\theta)})e^{-j\theta\hat{n}}d\theta$$

- limiting case

$$w(\hat{n}) = 1 \quad -\infty < \hat{n} < \infty \Leftrightarrow W(e^{j\hat{\omega}}) = 2\pi\delta(\hat{\omega})$$

$$X_{\hat{n}}(e^{j\hat{\omega}}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} 2\pi\delta(-\theta)X(e^{j(\hat{\omega}-\theta)})e^{-j\theta\hat{n}}d\theta = X(e^{j\hat{\omega}})$$

i.e., we get the same thing no matter where the window is shifted

Alternative Forms of STFT

Alternative forms of $X_{\hat{n}}(e^{j\hat{\omega}})$

1. real and imaginary parts

$$X_{\hat{n}}(e^{j\hat{\omega}}) = \text{Re}\left[X_{\hat{n}}(e^{j\hat{\omega}})\right] + j\,\text{Im}\left[X_{\hat{n}}(e^{j\hat{\omega}})\right]$$
$$= a_{\hat{n}}(\hat{\omega}) - j\,b_{\hat{n}}(\hat{\omega})$$
$$a_{\hat{n}}(\hat{\omega}) = \text{Re}\left[X_{\hat{n}}(e^{j\hat{\omega}})\right]$$
$$b_{\hat{n}}(\hat{\omega}) = -\text{Im}\left[X_{\hat{n}}(e^{j\hat{\omega}})\right]$$

- when $x(m)$ and $w(\hat{n}-m)$ are both real (usually the case) can show that $a_{\hat{n}}(\hat{\omega})$ is symmetric in $\hat{\omega}$, and $b_{\hat{n}}(\hat{\omega})$ is anti-symmetric in $\hat{\omega}$

2. magnitude and phase

$$X_{\hat{n}}(e^{j\hat{\omega}}) = |X_{\hat{n}}(e^{j\hat{\omega}})|e^{j\theta_{\hat{n}}(\hat{\omega})}$$

- can relate $|X_{\hat{n}}(e^{j\hat{\omega}})|$ and $\theta_{\hat{n}}(\hat{\omega})$ to $a_{\hat{n}}(\hat{\omega})$ and $b_{\hat{n}}(\hat{\omega})$

Role of Window in STFT

- The window $w(\hat{n}-m)$ does the following:
  1. chooses portion of $x(m)$ to be analyzed
  2. window shape determines the nature of $X_{\hat{n}}(e^{j\hat{\omega}})$
- Since $X_{\hat{n}}(e^{j\hat{\omega}})$ (for fixed $\hat{n}$) is the normal FT of $w(\hat{n}-m)x(m)$, then if we consider the normal FT's of both $x(n)$ and $w(n)$ individually, we get

$$X(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} x(m)e^{-j\hat{\omega}m}$$

$$W(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} w(m)e^{-j\hat{\omega}m}$$

Role of Window in STFT

- then for fixed $\hat{n}$, the normal Fourier transform of the product $w(\hat{n}-m)x(m)$ is the convolution of the transforms of $w(\hat{n}-m)$ and $x(m)$
- for fixed $\hat{n}$, the FT of $w(\hat{n}-m)$ is $W(e^{-j\hat{\omega}})e^{-j\hat{\omega}\hat{n}}$--thus

$$X_{\hat{n}}(e^{j\hat{\omega}}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W(e^{-j\theta})e^{-j\theta\hat{n}}\, X(e^{j(\hat{\omega}-\theta)})d\theta$$

- and replacing θ by $-\theta$ gives

$$X_{\hat{n}}(e^{j\hat{\omega}}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W(e^{j\theta})e^{j\theta\hat{n}}\, X(e^{j(\hat{\omega}+\theta)})d\theta$$

Interpretation of Role of Window

- $X_{\hat{n}}(e^{j\hat{\omega}})$ is the convolution of $X(e^{j\hat{\omega}})$ with the FT of the shifted window sequence $W(e^{-j\hat{\omega}})e^{-j\hat{\omega}\hat{n}}$
- $X(e^{j\hat{\omega}})$ really doesn't have meaning since $x(\hat{n})$ varies with time; consider $x(\hat{n})$ defined for window duration and extended for all time to have the <u>same</u> properties $\Rightarrow$ then $X(e^{j\hat{\omega}})$ does exist with properties that reflect the sound within the window (can also consider $x(\hat{n})=0$ outside the window and define $X(e^{j\hat{\omega}})$ appropriately--but this is another case)

<u>Bottom Line</u>: $X_{\hat{n}}(e^{j\hat{\omega}})$ is a smoothed version of the FT of the part of $x(\hat{n})$ that is within the window w.

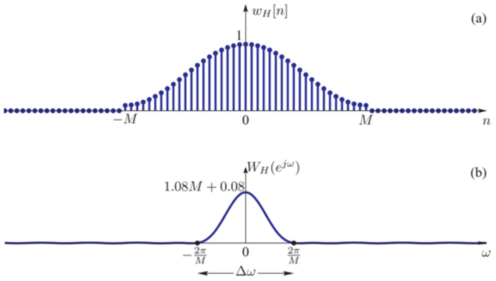
Windows in STFT

- for $X_{\hat{n}}(e^{j\hat{\omega}})$ to represent the short-time spectral properties of $x(\hat{n})$ inside the window => $W(e^{j\theta})$ should be much narrower in frequency than significant spectral regions of $X(e^{j\hat{\omega}})$--i.e., almost an impulse in frequency
- consider rectangular and Hamming windows, where width of the main spectral lobe is inversely proportional to window length, and side lobe levels are essentially independent of window length

Rectangular Window: flat window of length *L* samples; first zero in frequency response occurs at *F_S/L*, with sidelobe levels of -14 dB or lower

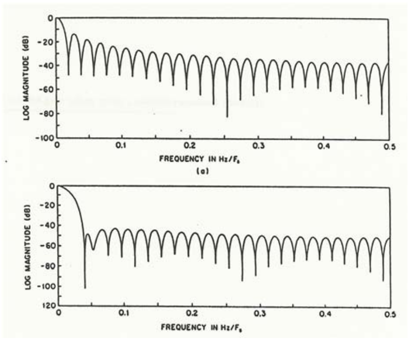
Hamming Window: raised cosine window of length *L* samples; first zero in frequency response occurs at *$2F_S$/L*, with sidelobe levels of -40 dB or lower

Windows

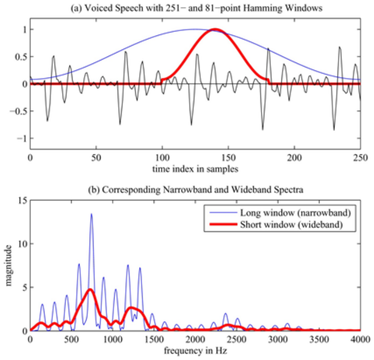


L=2M+1-point Hamming window and its corresponding DTFT

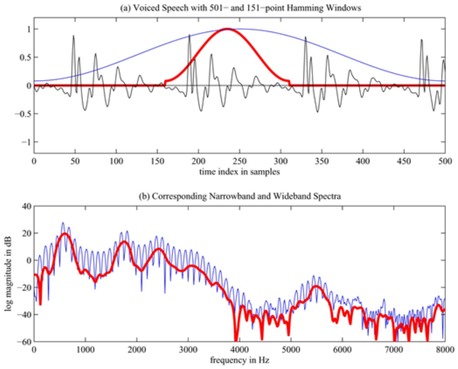
Frequency Responses of Windows



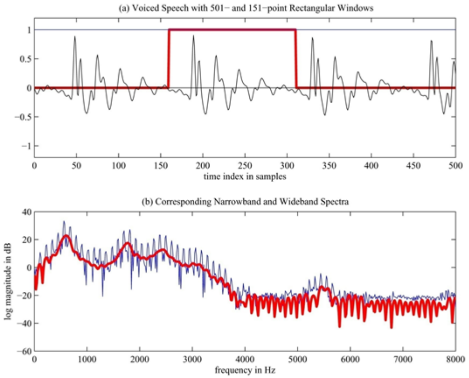
Effect of Window Length-HW



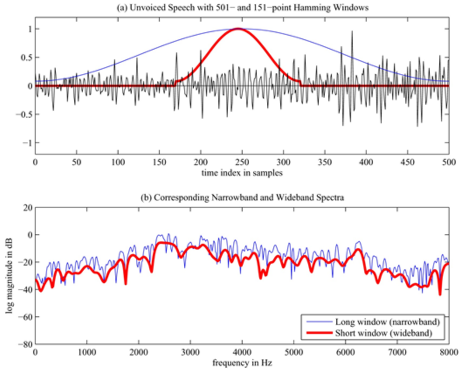
Effect of Window Length-HW



Effect of Window Length-RW



Effect of Window Length-HW



Relation to Short-Time Autocorrelation

□ $X_{\hat{n}}(e^{j\hat{\omega}})$ is the discrete-time Fourier transform of $w[\hat{n}-m]x[m]$ for each value of $\hat{n}$, then it is seen that

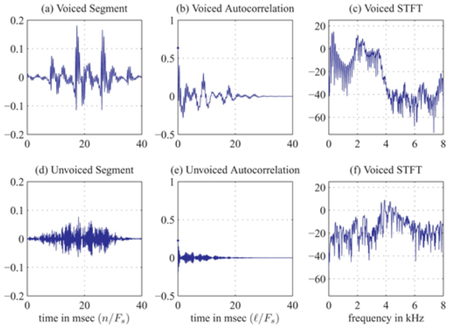
$$S_{\hat{n}}(e^{j\hat{\omega}}) = |X_{\hat{n}}(e^{j\hat{\omega}})|^2 = X_{\hat{n}}(e^{j\hat{\omega}})\, X^*_{\hat{n}}(e^{j\hat{\omega}})$$

is the Fourier transform of

$$R_{\hat{n}}(l) = \sum_{m=-\infty}^{\infty} w[\hat{n}-m]x[m]w[\hat{n}-l-m]x[m+l]$$

which is the short-time autocorrelation function of the previous chapter. Thus the above equations relate the short-time spectrum to the short-time autocorrelation,

Short-Time Autocorrelation and STFT



Summary of FT view of STFT

- interpret $X_{\hat{n}}(e^{j\omega})$ as the normal Fourier transform of the sequence $w(\hat{n}-m)x(m), -\infty < m < \infty$
- properties of this Fourier transform depend on the window
  - frequency resolution of $X_{\hat{n}}(e^{j\omega})$ varies inversely with the length of the window => want long windows for high resolution
  - want $x(n)$ to be relatively stationary (non-time-varying) during duration of window for most stable spectrum => want short windows

⇔ as usual in speech processing, there needs to be a compromise between good temporal resolution (short windows) and good frequency resolution (long windows)

Linear Filtering Interpretation of STFT

Linear Filtering Interpretation

1. modulation-lowpass filter form (n rather than $\hat{n}$)

$$X_n(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} x(m)e^{-j\hat{\omega}m}w(n-m)$$

$$= w(n) * \left(x(n)e^{-j\hat{\omega}n}\right), \quad n \text{ variable, } \hat{\omega} \text{ fixed}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} W(e^{j\theta})X(e^{j(\theta+\hat{\omega})})e^{j\theta n}d\theta$$

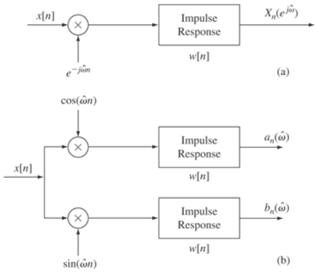
2. bandpass filter-demodulation

$$X_n(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} w(m)x(n-m)e^{-j\hat{\omega}(n-m)}$$

$$= e^{-j\hat{\omega}n}\sum_{m=-\infty}^{\infty}(w(m)e^{j\hat{\omega}m})x(n-m)$$

$$= e^{-j\hat{\omega}n}[(w(n)e^{j\hat{\omega}n}) * x(n)], \quad n \text{ variable, } \hat{\omega} \text{ fixed}$$

Linear Filtering Interpretation



1. modulation-lowpass filter form:

$$X_n(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} x(m)e^{-j\hat{\omega}m}w(n-m),$$

n variable, $\hat{\omega}$ fixed

$$= \left(x(n)e^{-j\hat{\omega}n}\right) * w(n)$$

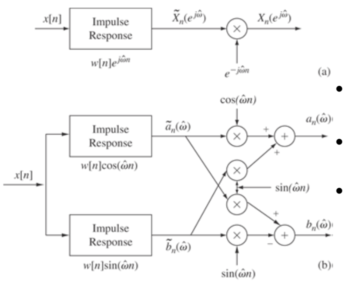
$$= \left(x(n)\cos(\hat{\omega}n)\right) * w(n) - j\left(x(n)\sin(\hat{\omega}n)\right) * w(n)$$

$$= a_n(\hat{\omega}) - jb_n(\hat{\omega})$$

Linear Filtering Interpretation

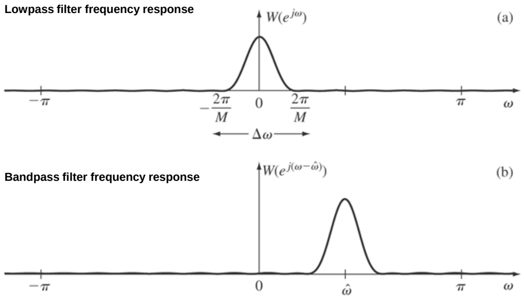
2. bandpass filter-demodulation form

$$X_n(e^{j\hat{\omega}}) = e^{-j\hat{\omega}n}\left[\left(w(n)e^{j\hat{\omega}n}\right) * x(n)\right], \quad n \text{ variable, } \hat{\omega} \text{ fixed}$$



- complex bandpass filter output modulated by signal $e^{-j\hat{\omega}n}$
- if $W(e^{j\theta})$ is lowpass, then filter is bandpass around $\theta = \hat{\omega}$
- all real computation for lower half structure

Linear Filtering Interpretation


Lowpass filter frequency response

Bandpass filter frequency response

Linear Filtering Interpretation

- assume normal FT of $x(n)$ exists

  $x(n) \leftrightarrow X(e^{j\theta})$ (recall that $\hat{\omega}$ is a particular frequency)

  $x(n)e^{-j\hat{\omega}n} \leftrightarrow X(e^{j(\theta+\hat{\omega})})$

=> spectrum of $x(n)$ at frequency $\hat{\omega}$ is shifted to zero frequency;

- since the STFT is a convolution, the FT of the STFT is the product of the individual FT's, i.e.,

  $X(e^{j(\theta+\hat{\omega})}) \cdot W(e^{j\theta})$

- if $W(e^{j\theta})$ resembles a narrow band lowpass filter, i.e., $W(e^{j\theta}) = 1$ for small θ and is 0 otherwise, then

  $X(e^{j(\theta+\hat{\omega})}) \cdot W(e^{j\theta}) \approx X(e^{j\hat{\omega}})$

Summary-STFT

Short-Time Fourier Transform (STFT)

$$X_{\hat{n}}(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} x[m]w[\hat{n}-m]e^{-j\hat{\omega}m},$$

$$-\infty < \hat{n} < \infty,\ 0 \le \hat{\omega} < 2\pi$$

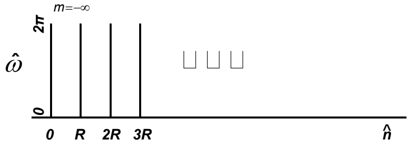
Fixed value of $\hat{n}$, varying $\hat{\omega}$ -- DFT Interpretation

Fixed value of $\hat{\omega}$, varying $\hat{n}$ -- Filter Bank Interpretation

Summary

Short-Time Fourier Transform (STFT)

$$X_{\hat{n}}(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} x[m]w[\hat{n}-m]e^{-j\hat{\omega}m},\ -\infty < \hat{n} < \infty,\ 0 \le \hat{\omega} < 2\pi$$



$$\text{DFT: } X_{\hat{n}}(e^{j\hat{\omega}}) = \sum_{m=\hat{n}-L+1}^{\hat{n}} \left(x[m]w[\hat{n}-m]\right)e^{-j\hat{\omega}m}$$

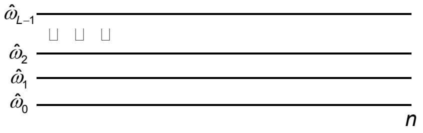
$$X_{\hat{n}}(e^{j\hat{\omega}}) = \text{DFT}\left(x[m]w[\hat{n}-m]\right)$$

$$0 \le \hat{\omega} < 2\pi,\quad \hat{n} = 0, R, 2R, \ldots$$

Summary – Modulation/Lowpass Filter

Short-Time Fourier Transform (STFT)

$$X_{\hat{n}}(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} x[m]w[\hat{n}-m]e^{-j\hat{\omega}m},\ -\infty < \hat{n} < \infty,\ 0 \le \hat{\omega} < 2\pi$$



$$\text{Filter Bank: } X_n(e^{j\hat{\omega}}) = \sum_{m=n-L+1}^{n} \left(x[m]e^{-j\hat{\omega}m}\right)w[n-m]$$

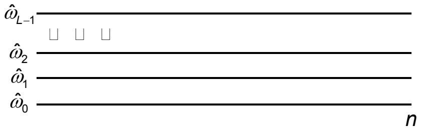
$$X_n(e^{j\hat{\omega}}) = \left(x[n]e^{-j\hat{\omega}n}\right)w[n-m]$$

$$= \left(x[n]e^{-j\hat{\omega}n}\right) * w[n]$$

Summary – Bandpass Filter/Demodulation

Short-Time Fourier Transform (STFT)

$$X_{\hat{n}}(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} x[m]w[\hat{n}-m]e^{-j\hat{\omega}m},\ -\infty < \hat{n} < \infty,\ 0 \le \hat{\omega} < 2\pi$$



$$\text{Filter Bank: } X_n(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} \left(x[n-m]e^{-j\hat{\omega}(n-m)}\right)w[m]$$

$$X_n(e^{j\hat{\omega}}) = e^{-j\hat{\omega}n}\left[(w[n]e^{j\hat{\omega}n}) * x[n]\right]$$

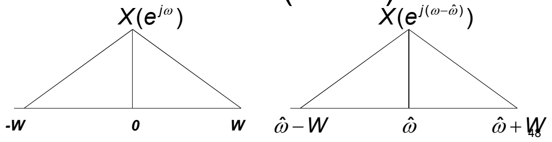
Summary – Modulation

Modulation

$$x[n] \square e^{j\hat{\omega}n} \leftrightarrow X(e^{j\omega}) * FT(e^{j\hat{\omega}n})$$

$$= X(e^{j\omega}) * \delta(\omega - \hat{\omega})$$

$$= X(e^{j(\omega-\hat{\omega})})$$



STFT Magnitude Only

- for many applications you only need the magnitude of the STFT(not the phase)
- in such cases, the bandpass filter implementation is less complex, since

$$| X_n(e^{j\hat{\omega}}) | = \left[a_n^2(\hat{\omega}) + b_n^2(\hat{\omega}) \right]^{1/2}$$
$$= | \tilde{X}_n(e^{j\hat{\omega}}) | = \left[\tilde{a}_n^2(\hat{\omega}) + \tilde{b}_n^2(\hat{\omega}) \right]^{1/2}$$

Sampling Rates of STFT

Sampling Rates of STFT

- need to sample STFT in both time and frequency to produce an unaliased representation from which *x(n)* can be exactly recovered
- sampling rates lower than the theoretical minimum rate can be used, in either time or frequency, and *x(n)* can still be exactly recovered from the aliased (under-sampled) short-time transform
  - this is useful for spectral estimation, pitch estimation, formant estimation, speech spectrograms, vocoders
  - for applications where the signal is modified, e.g., speech enhancement, cannot undersample STFT and still recover modified signal exactly

Sampling Rate in Time

- to determine the sampling rate in time, we take a linear filtering view
  1. $X_n(e^{j\hat{\omega}})$ is the output of a filter with impulse response $\tilde{w}(n)$
  2. $W(e^{j\hat{\omega}})$ is a lowpass response with effective bandwidth of B Hertz
- thus the effective bandwidth of $X_n(e^{j\hat{\omega}})$ is B Hertz => $X_n(e^{j\hat{\omega}})$ has to be sampled at a rate of $2B$ samples/second to avoid aliasing

Example: Hamming Window

$$w(n) = 0.54 - 0.46\cos(2\pi n/(L-1)) \quad 0 \le n \le L-1$$
$$= 0 \qquad \text{otherwise}$$

$\Rightarrow B \approx \frac{2F_s}{L}$(Hz); for $L = 400$, $F_s = 10{,}000$ Hz => $B = 50$ Hz => need rate of 100/sec (every 100 samples) for sampling rate in time

Sampling Rate in Frequency

- since $X_n(e^{j\hat{\omega}})$ is periodic in $\hat{\omega}$ with period 2π, it is only necessary to sample over an interval of length 2π
- need to determine an appropriate finite set of frequencies, $\hat{\omega}_k = 2\pi k/N,\ k = 0,1,\ldots,N-1$ at which $X_n(e^{j\hat{\omega}})$ must be specified to exactly recover $x(n)$
- use the Fourier transform interpretation of $X_n(e^{j\hat{\omega}})$
  1. if the window $w(n)$ is time-limited, then the inverse transform of $X_n(e^{j\hat{\omega}})$ is time-limited
  2. the sampling theorem requres that we sample $X_n(e^{j\hat{\omega}})$ in the frequency dimension at a rate of at least twice its ('symmetric') "time width"
  3. since the inverse Fourier transform of $X_n(e^{j\hat{\omega}})$ is the signal $x(m)w(n-m)$ and this signal is of duration L samples (the duration of $w(n)$), then according to the sampling theorem $X_n(e^{j\hat{\omega}})$ must be sampled (in frequency) at the set of frequencies
     $\hat{\omega}_k = \frac{2\pi k}{L},\ k = 0,1,\ldots,L-1$ (where $L/2$ is the effective width of the window)
     in order to exactly recover $x(n)$ from $X_n(e^{j\hat{\omega}_k})$

- thus for a Hamming window of duration L=400 samples, we require that the STFT be evaluated at at least 400 uniformly spaced frequencies around the unit circle

"Total" Sampling Rate of STFT

- the "total" sampling rate for the STFT is the product of the sampling rates in time and frequency, i.e.,
  SR = *SR*(time) x *SR*(frequency)
  = *2B x L* samples/sec
  B = frequency bandwidth of window (Hz)
  L = time width of window (samples)
- for most windows of interest, *B* is a multiple of F_S/L, i.e.,
  $B = C\ F_S/L$ (Hz), *C*=1 for Rectangular Window
  C=2 for Hamming Window
  $SR = 2C\ F_S$ samples/second
- can define an 'oversampling rate' of
  $SR/F_S = 2C$ = oversampling rate of STFT as compared to conventional sampling representation of *x(n)*
  for RW, *2C*=2; for HW *2C*=4 => range of oversampling is 2-4
  this *oversampling* gives a *very flexible representation* of the speech signal

Mathematical Basis for Sampling the STFT

- assume sample in time at $\hat{n} = n_r = rR, -\infty < r < \infty$

  and in frequency at $\hat{\omega} = \hat{\omega}_k = \left(\frac{2\pi}{N}\right)k,\ k = 0,1,\ldots,N-1$
- sample values

$$X_{rR}(e^{j\frac{2\pi}{N}k}) = \sum_{m=-\infty}^{\infty} w[rR-m]x[m]e^{-j\frac{2\pi}{N}km}$$

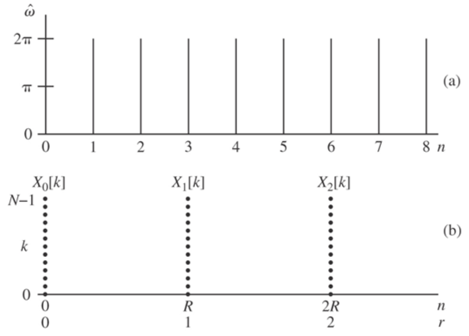
$$= e^{-j\frac{2\pi}{N}krR}\tilde{X}_{rR}(e^{j\frac{2\pi}{N}k})$$

$$\tilde{X}_{rR}(e^{j\frac{2\pi}{N}k}) = \sum_{m=-\infty}^{\infty} x[rR+m]w(-m)e^{-j\frac{2\pi}{N}km} \quad (\text{set } m = rR+m'; m = m')$$

- define DFT-type notation

$$X_r(k) = X_{rR}(e^{j\frac{2\pi}{N}k}) = e^{-j\frac{2\pi}{N}krR}\tilde{X}_r(k)$$

Sampling the STFT



Sampling the STFT

- DFT Notation

$$X_r[k] = X_{rR}(e^{j\frac{2\pi}{N}k}) = e^{-j\frac{2\pi}{N}krR}\tilde{X}_r[k]$$

- let $w[-m] \neq 0$ for $0 \leq m \leq L-1$ (finite duration window with no zero-valued samples)

$$\tilde{X}_r[k] = \sum_{m=0}^{L-1} x[rR+m]w[-m]\, e^{-j\frac{2\pi}{N}km}$$

$(r$ fixed, $0 \leq k \leq N-1)$

- if $L \leq N$ then (DFT defined with no aliasing $\Rightarrow$ can recover sequence exactly using inverse DFT)

$$x[rR+m]w[-m] = \frac{1}{N}\sum_{k=0}^{N-1}\tilde{X}_r[k]\, e^{j\frac{2\pi}{N}km}$$

$(r$ fixed, $0 \leq m \leq N-1)$

- if $R \leq L$ (IDFT defined with no aliasing), then all samples can be recovered from $X_r[k]$ $(R > L \Rightarrow$gaps in sequence$)$

What We Have Learned So Far

1. $X_{\hat{n}}(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} x(m)w(\hat{n}-m)e^{-j\hat{\omega}m}$

□function of $\hat{n} = n$ for sampled $\hat{\omega}$ (looks like a time sequence)

□function of $\hat{\omega} = \omega$ for sampled $\hat{n}$ (looks like a transform)

$X_{\hat{n}}(e^{j\hat{\omega}})$ (no sampling rate reduction) defined for $\hat{n} = 1,2,3,\ldots;\ 0 \leq \hat{\omega} \leq \pi$

2. $X_{\hat{n}}(e^{j\hat{\omega}}) = \text{DTFT}\left[x(m)w(\hat{n}-m)\right] \Rightarrow \hat{n}$ fixed, $\hat{\omega}$ variable

with time origin tied to $x(\hat{n})$

$X_{\hat{n}}(e^{j\hat{\omega}}) = e^{-j\hat{\omega}\hat{n}}\text{DTFT}\left[x(\hat{n}+m)w(-m)\right] \Rightarrow \hat{n}$ fixed, $\hat{\omega}$ variable

with time origin tied to $w(-m)$

3. Interpretations of $X_{\hat{n}}(e^{j\hat{\omega}})$

1. $\hat{n}$ fixed, $\hat{\omega} = \omega$ variable; $X_{\hat{n}}(e^{j\hat{\omega}}) = \text{DTFT}\left[x(m)w(\hat{n}-m)\right] \Rightarrow$DFT View

2. $\hat{n} = n$ variable, $\hat{\omega}$ fixed; $X_{\hat{n}}(e^{j\hat{\omega}}) = x(n)e^{-j\hat{\omega}n} * w(n) \Rightarrow$Linear Filtering view $\Rightarrow$ filter bank implementation

What We Have Learned So Far

4. Signal Recovery from STFT

$$x(m)w(\hat{n}-m) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X_{\hat{n}}(e^{j\hat{\omega}})e^{j\hat{\omega}m}d\hat{\omega}$$

$$x(\hat{n}) = \frac{1}{2\pi w(0)}\int_{-\pi}^{\pi} X_{\hat{n}}(e^{j\hat{\omega}})e^{j\hat{\omega}\hat{n}}d\hat{\omega}$$

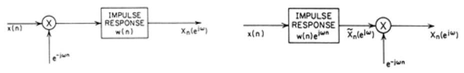
5. Linear Filtering Interpretation

1. modulation-lowpass filter $\Rightarrow$ $X_n(e^{j\hat{\omega}}) = w(n) * \left[x(n)e^{-j\hat{\omega}n}\right]$,

$\hat{n} = n$ variable, $\hat{\omega}$ fixed $\qquad X_n(e^{j\hat{\omega}}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W(e^{j\theta})X(e^{j(\theta+\hat{\omega})})e^{j\theta n}d\theta$

2. bandpass filer-demodulation $\Rightarrow$ $X_n(e^{j\hat{\omega}}) = e^{-j\hat{\omega}n}\left[\left(w(n)e^{j\hat{\omega}n}\right) * x(n)\right]$,

$\hat{n} = n$ variable, $\hat{\omega}$ fixed



What We Have Learned So Far

6. Sampling Rates in Time and Frequency

1. time: $W(e^{j\omega})$ has bandwidth of B Hertz $\Rightarrow$ $2B$ samples/sec rate

Hamming Window: $B = \frac{2F_S}{L}$(Hz)

2. frequency: $\tilde{w}(n)$ is time limited to L samples $\Rightarrow$ inverse of $X_n(e^{j\omega})$ is also time limited $\Rightarrow$ need to sample in frequency at twice the (effective) time width of the time-limited sequence $\Rightarrow$ L frequency samples

3. total Sampling Rate: $2B \cdot L$ samples/sec

- $B =$ frequency bandwidth of the window (Hz)
- $L =$ effective time width of the window (samples)

$B = C \cdot F_S / L$(Hz)$\Rightarrow$ Sampling Rate $= 2B \cdot L = 2CF_S$ samples/second

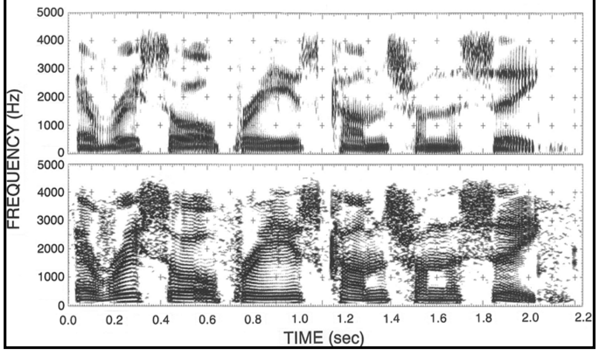
- for Rectangular Window, $C = 1$
- for Hamming Window, $C = 2$

Spectrographic Displays

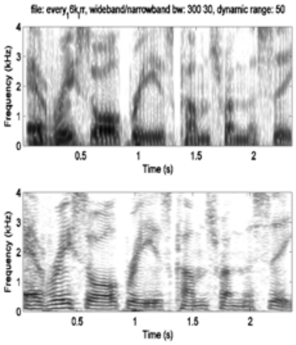
Spectrographic Displays

- ***Sound Spectrograph***-one of the earliest embodiments of the time-dependent spectrum analysis techniques
  - 2-second utterance repeatedly modulates a variable frequency oscillator, then bandpass filtered, and the average energy at a given time and frequency is measured and used as a crude measure of the STFT
  - thus energy is recorded by an ingenious electro-mechanical system on special electrostatic paper called teledeltos paper
  - result is a two-dimensional representation of the time-dependent spectrum-with vertical intensity being spectrum level at a given frequency, and horizontal intensity being spectral level at a given time-with spectrum magnitude being represented by the darkness of the marking
  - wide bandpass filters (300 Hz bandwidth) provide good temporal resolution and poor frequency resolution (resolve pitch pulses in time but not in frequency)—called wideband spectrogram
  - narrow bandpass filters (45 Hz bandwidth) provide good frequency resolution and poor time resolution (resolve pitch pulses in frequency, but not in time)—called narrowband spectrogram

Conventional Spectrogram (Every salt breeze comes from the sea)



Digital Speech Spectrograms

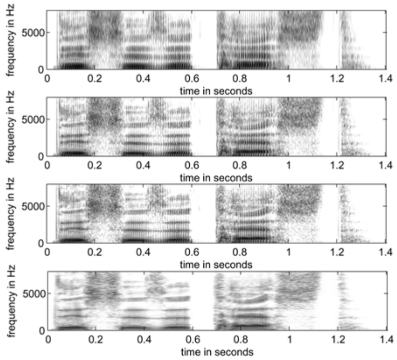


- **wideband spectrogram**
  - follows broad spectral peaks (formants) over time
  - resolves most individual pitch periods as vertical striations since the IR of the analyzing filter is comparable in duration to a pitch period
  - what happens for low pitch males—high pitch females
  - for unvoiced speech there are no vertical pitch striations
- **narrowband spectrogram**
  - individual harmonics are resolved in voiced regions
  - formant frequencies are still in evidence
  - usually can see fundamental frequency
  - unvoiced regions show no strong structure

Digital Speech Spectrograms

- Speech Parameters (“This is a test”):
  - sampling rate: 16 kHz
  - speech duration: 1.406 seconds
  - speaker: male
- Wideband Spectrogram Parameters:
  - analysis window: Hamming window
  - analysis window duration: 6 msec (96 samples)
  - analysis window shift: 0.625 msec (10 samples)
  - FFT size: 512
  - dynamic range of spectral log magnitudes: 40 dB
- Narrowband Spectrogram Parameters:
  - analysis window: Hamming window
  - analysis window duration: 60 msec (960 samples)
  - analysis window shift: 6 msec (96 samples)
  - FFT size: 1024
  - dynamic range of spectral log magnitudes: 40 dB

Digital Speech Spectrograms

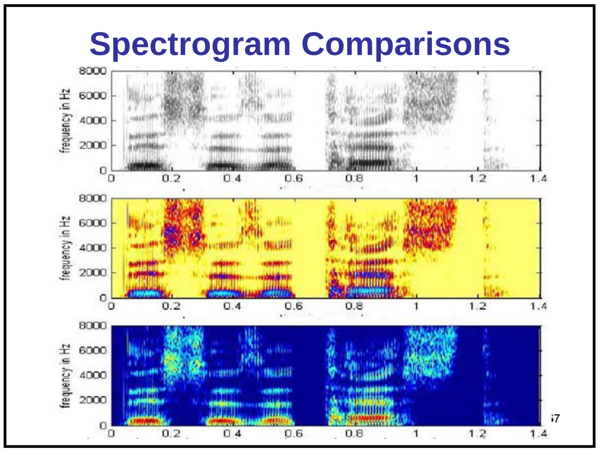


Top Panel:
3 msec (48 samples) window

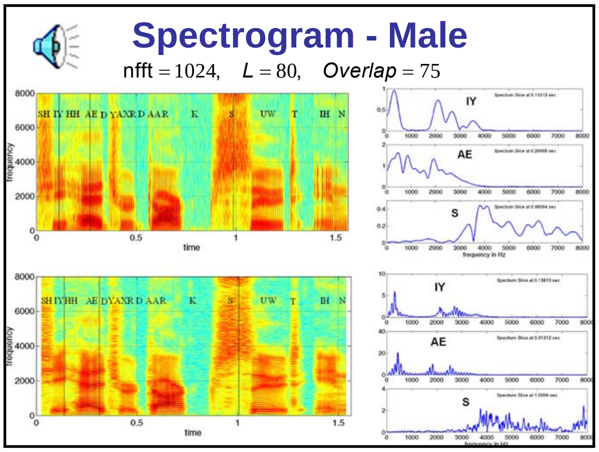
Second Panel:
6 msec (96 samples) window

Third Panel:
9 msec (144 sample) window

Fourth Panel:
30 msec (480 sample) window

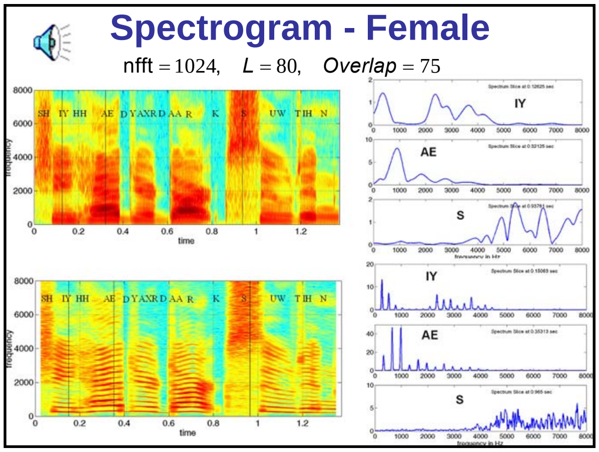


Spectrogram Comparisons



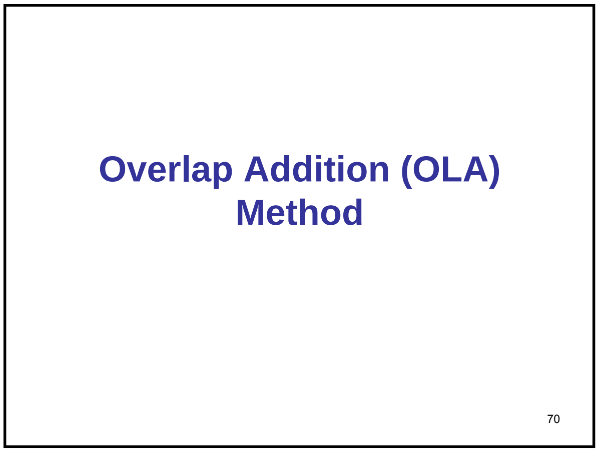
Spectrogram - Male

nfft $= 1024$, $L = 80$, $Overlap = 75$

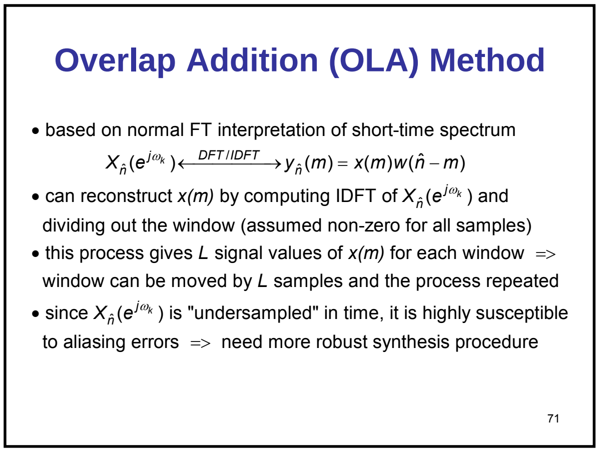


Spectrogram - Female

nfft $= 1024$, $L = 80$, $Overlap = 75$



Overlap Addition (OLA) Method

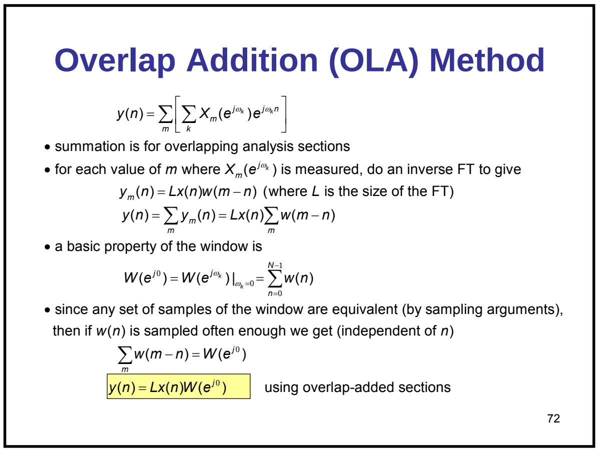


Overlap Addition (OLA) Method

- based on normal FT interpretation of short-time spectrum

$$X_{\hat{n}}(e^{j\omega_k}) \xleftrightarrow{DFT/IDFT} y_{\hat{n}}(m) = x(m)w(\hat{n}-m)$$

- can reconstruct *x(m)* by computing IDFT of $X_{\hat{n}}(e^{j\omega_k})$ and dividing out the window (assumed non-zero for all samples)
- this process gives *L* signal values of *x(m)* for each window => window can be moved by *L* samples and the process repeated
- since $X_{\hat{n}}(e^{j\omega_k})$ is "undersampled" in time, it is highly susceptible to aliasing errors => need more robust synthesis procedure



Overlap Addition (OLA) Method

$$y(n) = \sum_m \left[\sum_k X_m(e^{j\omega_k}) e^{j\omega_k n} \right]$$

- summation is for overlapping analysis sections
- for each value of *m* where $X_m(e^{j\omega_k})$ is measured, do an inverse FT to give

$$y_m(n) = Lx(n)w(m-n) \text{ (where } L \text{ is the size of the FT)}$$

$$y(n) = \sum_m y_m(n) = Lx(n) \sum_m w(m-n)$$

- a basic property of the window is

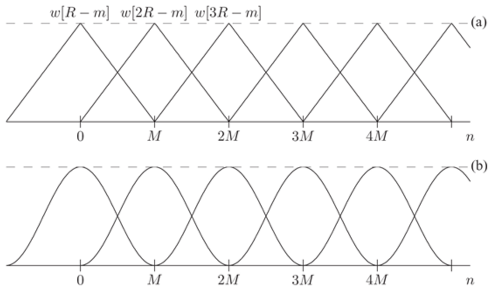
$$W(e^{j0}) = W(e^{j\omega_k})|_{\omega_k=0} = \sum_{n=0}^{N-1} w(n)$$

- since any set of samples of the window are equivalent (by sampling arguments), then if $w(n)$ is sampled often enough we get (independent of n)

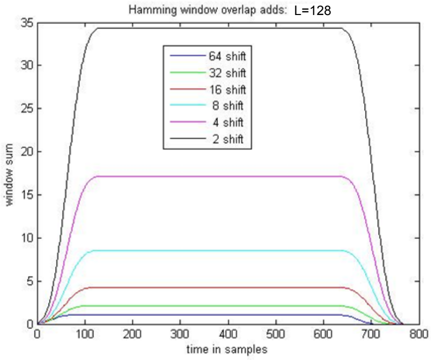
$$\sum_m w(m-n) = W(e^{j0})$$

$$\boxed{y(n) = Lx(n)W(e^{j0})} \quad \text{using overlap-added sections}$$

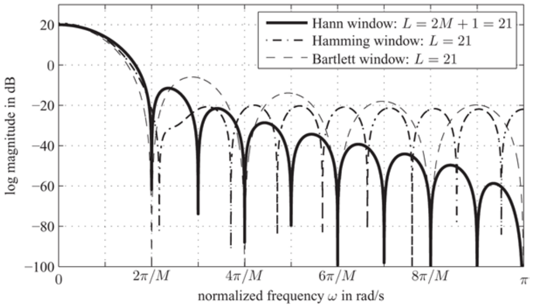
Overlap Addition of Bartlett and Hann Windows



Overlap Addition of Hamming Window

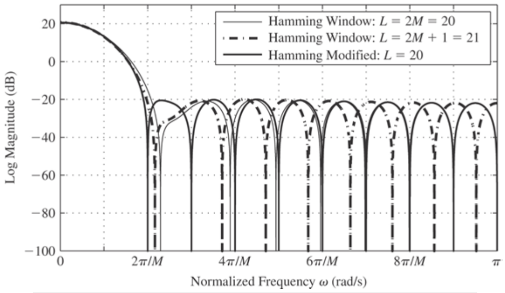


Window Spectra



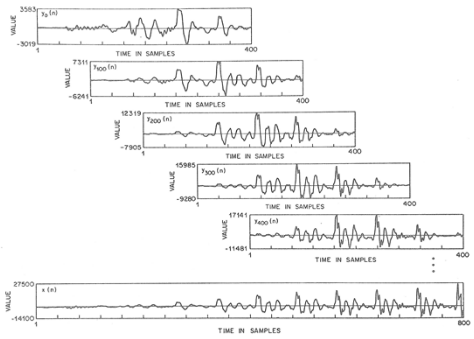
DTFTof Bartlett (triangular), Hann and Hamming windows

Hamming Window Spectra

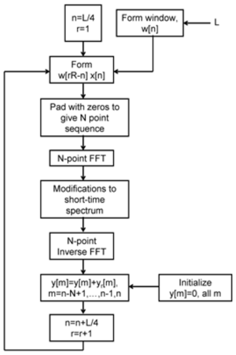


DTFTs of even-length, odd-length and modified odd-to-even length Hamming windows; zeros spaced at 2π/*R* give perfect reconstruction using OLA (even-length window)

Overlap Addition (OLA) Method



Overlap Addition (OLA) Method



- *w(n)* is an *L*-point Hamming window with *R=L/4*
- assume *x(n)=0* for *n<0*
- time overlap of 4:1 for HW
- first analysis section begins at *n=L/4*

Overlap Addition (OLA) Method

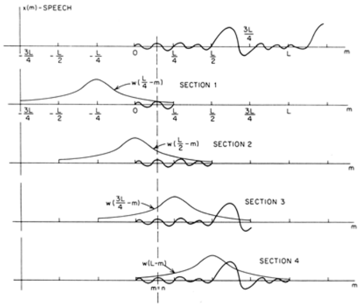


Fig. 6.17 Reconstruction procedure for $w(n)$ using an L-point Hamming window.

- 4-overlapping sections contribute to each interval
- *N*-point FFT's done using *L* speech samples, with *N-L* zeros padded at end to allow modifications without significant aliasing effects
- for a given value of *n*

y(n)=x(n)w(R-n)+x(n)w(2R-n)+x(n)w(3R-n)+x(n)w(4R-n)=x(n)[w(R-n)+w(2R-n)+w(3R-n)+w(4R-n)]=x(n) W(e^{j0})/R

Filter Bank Summation (FBS)

Filter Bank Summation

- the filter bank interpretation of the STFT shows that for any frequency ω_k, $X_n(e^{j\omega_k})$ is a lowpass representation of the signal in a band centered at ω_k ($n=\hat{n}$ for FBS)

$$X_n(e^{j\omega_k}) = e^{-j\omega_k n}\sum_{m=-\infty}^{\infty} x(n-m)w_k(m)\, e^{j\omega_k m}$$

where $w_k(m)$ is the lowpass window used at frequency ω_k (we have generalized the structure to allow a different lowpass window at each frequency ω_k).

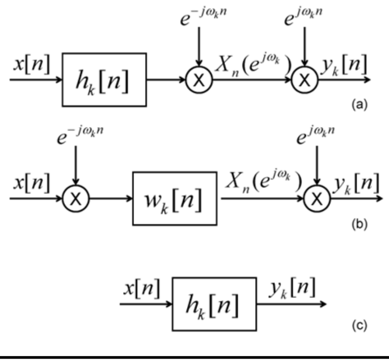
Filter Bank Summation

- define a bandpass filter and substitute it in the equation to give

$$h_k(n) = w_k(n)\, e^{j\omega_k n}$$

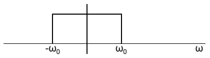
$$X_n(e^{j\omega_k}) = e^{-j\omega_k n}\sum_{m=-\infty}^{\infty} x(n-m)h_k(m)$$

Filter Bank Summation



Filter Bank Interpretation of STFT (case: w$_k$[n]=w[n])

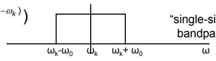
$$w[n] \leftarrow - - - - \rightarrow W(e^{j\omega})$$



$$h[n] = w[n]e^{j\omega_k n} \leftarrow - - - - \rightarrow H(e^{j\omega}) = W(e^{j\omega}) \otimes FT(e^{j\omega_k n})$$

$$FT(e^{j\omega_k n}) = \sum_{n=-\infty}^{\infty} e^{j\omega_k n}e^{-j\omega n} = \sum_{n=-\infty}^{\infty} e^{-j(\omega-\omega_k)n} = \delta(\omega-\omega_k)$$

$$H(e^{j\omega}) = W(e^{j\omega}) \otimes \delta(\omega-\omega_k) = W(e^{j(\omega-\omega_k)})$$



"single-sided, bandpass"

$$v_k[n] = X_n(e^{j\omega_k}) = e^{-j\omega_k n}\cdot\left[x[n] * h[n]\right]$$

$$V_k(e^{j\omega}) = \left[H(e^{j\omega})\cdot X(e^{j\omega})\right] \otimes FT(e^{-j\omega_k n})$$

$$= \left[H(e^{j\omega})\cdot X(e^{j\omega})\right] \otimes \delta(\omega+\omega_k)$$

$$= \left[X(e^{j\omega})\cdot W(e^{j(\omega-\omega_k)})\right] \otimes \delta(\omega+\omega_k)$$

$$= X(e^{j(\omega+\omega_k)})\cdot W(e^{j\omega})$$

"lowpass"

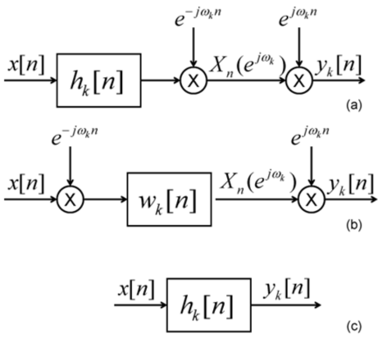
Filter Bank Summation

- thus $X_n(e^{j\omega_k})$ is obtained by bandpass filtering $x(n)$ followed by modulation with the complex exponential $e^{-j\omega_k n}$. We can express this in the form

$$y_k(n) = X_n(e^{j\omega_k})e^{j\omega_k n} = \sum_{m=-\infty}^{\infty} x(n-m)h_k(m)$$

- thus $y_k(n)$ is the output of a bandpass filter with impulse response $h_k(n)$

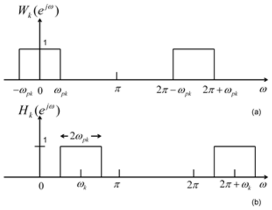
Filter Bank Summation



Filter Bank Summation

- a practical method for reconstructing $x(n)$ from the STFT is as follows
  1. assume we know $X_n(e^{j\omega_k})$ for a set of N frequencies $\{\omega_k\}$, $k = 0,1,\ldots,N-1$
  2. assume we have a set of N bandpass filters with impulse responses

     $h_k(n) = w_k(n)e^{j\omega_k n}$, $k = 0,1,\ldots,N-1$
  3. assume $w_k(n)$ is an ideal lowpass filter with cutoff frequency ω_{pk}



- the frequency response of the bandpass filter is

$$H_k(e^{j\omega}) = W_k(e^{j(\omega-\omega_k)})$$

Filter Bank Summation

- consider a set of N bandpass filters, uniformly spaced, so that the entire frequency band is covered

$$\omega_k = \frac{2\pi k}{N},\ k = 0,1,\ldots,N-1$$

- also assume window the same for all channels, i.e.,

$$w_k(n) = w(n),\ \ k = 0,1,\ldots,N-1$$

- if we add together all the bandpass outputs, the composite response is

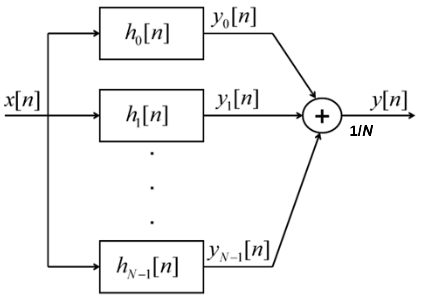
$$\tilde{H}(e^{j\omega}) = \sum_{k=0}^{N-1} H_k(e^{j\omega}) = \sum_{k=0}^{N-1} W(e^{j(\omega-\omega_k)})$$

- if $W(e^{j\omega_k})$ is properly sampled in frequency ($N \geq L$), where L is the window duration, then it can be shown that

$$\boxed{\frac{1}{N}\sum_{k=0}^{N-1} W(e^{j(\omega-\omega_k)}) = w(0)\quad \forall\omega}$$

FBS Formula

Filter Bank Summation



Filter Bank Summation

- derivation of FBS formula

$$w(n) \xleftrightarrow{FT\,/\,IFT} W(e^{j\omega})$$

- if $W(e^{j\omega})$ is sampled in frequency at N uniformly spaced points, the inverse discrete Fourier transform of the sampled version of $W(e^{j\omega_k})$ is (recall that sampling $\Rightarrow$ multiplication $\Leftrightarrow$ convolution $\Rightarrow$ aliasing)

$$\frac{1}{N}\sum_{k=0}^{N-1} W(e^{j\omega_k})e^{j\omega_k n} = \sum_{r=-\infty}^{\infty} w(n+rN)$$

- an aliased version of $w(n)$ is obtained.

Filter Bank Summation

- If $w(n)$ is of duration L samples, then
  $w(n) = 0,\ n < 0,\ \ n \ge L$
- and no aliasing occurs due to sampling in frequency of $W(e^{j\omega})$. In this case if we evaluate the aliased formula for $n = 0$, we get

$$\frac{1}{N}\sum_{k=0}^{N-1} W(e^{j\omega_k}) = w(0)$$

- the FBS formula is seen to be equivalent to the formula above, since (according to the sampling theorem) any set of N uniformly spaced samples of $W(e^{j\omega})$ is adequate

Filter Bank Summation

- the impulse response of the composite filter bank system is

$$\tilde{h}(n) = \sum_{k=0}^{N-1} h_k(n) = \sum_{k=0}^{N-1} w(n)e^{j\omega_k n} = N\,w(0)\delta(n)$$

- thus the composite output is

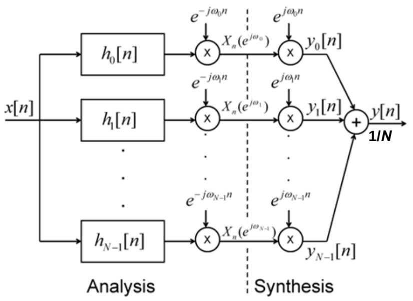
$$y(n) = x(n) * \tilde{h}(n) = N\,w(0)\,x(n)$$

- thus for FBS method, the reconstructed signal is

$$y(n) = \sum_{k=0}^{N-1} y_k(n) = \sum_{k=0}^{N-1} X_n(e^{j\omega_k})e^{j\omega_k n} = N\,w(0)\,x(n)$$

- if $X_n(e^{j\omega_k})$ is sampled properly in frequency, and is independent of the shape of $w(n)$

Filter Bank Summation



Filter Bank Summation

$$\begin{aligned}
y(n) &= \sum_{k=0}^{N-1} y_k(n) = \sum_{k=0}^{N-1} X_n(e^{j\frac{2\pi}{N}k})e^{j\frac{2\pi}{N}kn} \\
&= \sum_{k=0}^{N-1}\left[\sum_m x(m)w(n-m)e^{-j\frac{2\pi}{N}km}\right]e^{j\frac{2\pi}{N}kn} \\
&= \sum_m x(m)w(n-m)\sum_{k=0}^{N-1} e^{j\frac{2\pi}{N}k(n-m)} \\
&= \sum_m x(m)w(n-m)\sum_{r=-\infty}^{\infty} N\delta(n-m-rN)
\end{aligned}$$

$$\boxed{y(n) = N\sum_{r=-\infty}^{\infty} w(rN)x(n-rN)}$$

- $w(n) \ne 0$ for $0 \le n \le L-1 \Rightarrow$ if $N \ge L$ then need only $r = 0$ term

$$\boxed{y(n) = Nw(0)x(n)}$$

- if $N < L$ then in order for $y(n) = x(n)$ you need the condition $w(rN) = 0,\ r = \pm 1, \pm 2, \ldots$
- 'undersampled' representation can still work--at least in theory

Summary of FBS Method

- perfect reconstruction of *x(n)* from $X_n(e^{j\omega})$ is possible using FBS under the following conditions:
  1. *w(n)* is a finite duration filter/window
  2. $X_n(e^{j\omega})$ is sampled properly in both time and frequency
- perfect reconstruction of *x(n)* from $X_n(e^{j\omega_k})$ is also possible using FBS under the following condition:
  $W(e^{j\omega})$ is perfectly bandlimited
- To avoid time aliasing, $X_n(e^{j\omega_k})$ must be evaluated at at least L uniformly spaced frequencies, where L is the window duration
  -since window of length L samples has frequency bandwidth of from $2\pi/L$ (for RW) to $4\pi/L$ (for HW), the bandpass filters in FBS overlap in frequency since the analysis frequencies are $2\pi k/L,\ k = 0,1,\ldots,L-1$
- there is a way (at least theoretically) for $X_n(e^{j\omega_k})$ to be evaluated in non-overlapping bands and for which *x(n)* can still be exactly recovered

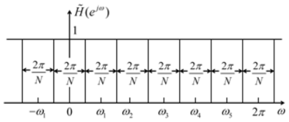
FBS Reconstruction in Non-Overlapping Bands

- assume window length for all bands is L samples
- assume the same window is used for N equally spaced frequency bands with analysis frequencies

$$\omega_k = \frac{2\pi k}{N},\ k = 0,1,\ldots,N-1$$

- where N can be less than L
- assume *w(n)* is an ideal lowpass filter with cutoff frequency

$$\omega_p = \frac{\pi}{N}$$



example with *N*=6 equally spaced ideal filters

FBS Reconstruction in Non-Overlapping Bands

- the composite impulse response for the FBS system is

$$\tilde{h}(n) = \sum_{k=0}^{N-1} w(n) e^{j\omega_k n} = w(n) \sum_{k=0}^{N-1} e^{j\omega_k n}$$

- defining a composite of the terms being summed as

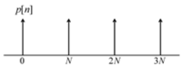
$$p(n) = \sum_{k=0}^{N-1} e^{j\omega_k n} = \sum_{k=0}^{N-1} e^{j2\pi kn/N}$$

- we get for $\tilde{h}(n)$

$$\tilde{h}(n) = w(n)\,p(n)$$

- it is easy to show that *p(n)* is a periodic train of impulses of the form

$$p(n) = N \sum_{r=-\infty}^{\infty} \delta(n - rN)$$

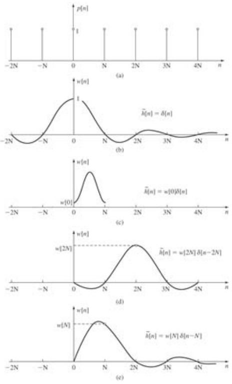


- giving for $\tilde{h}(n)$ the expression

$$\tilde{h}(n) = N \sum_{r=-\infty}^{\infty} w(rN)\,\delta(n - rN)$$

- thus the composite impulse response is the window sequence sampled at intervals of *N* samples

FBS Reconstruction in Non-Overlapping Bands



impulse response of ideal lowpass filter
with cutoff frequency *π/N*

- for ideal LPF we have

$$w(n) = \frac{\sin(\pi n / N)}{\pi n},\ w(rN) = \frac{\sin(\pi r)}{\pi rN} = \frac{1}{N}\delta(r)$$

giving $\tilde{h}(n) = \delta(n)$

- other cases where perfect reconstruction is obtained
  1. $w(n)$ is of finite length $L \le N$ and causal (no images)
  2. $w(n)$ has length $> N$ and has the property

$$\begin{aligned} w(n) &= 1/N, && \text{for } n = r_0 N \\ &= 0 && \text{for } n = rN\ (r \ne r_0,\ r = 0, \pm 1, \pm 2, \ldots) \end{aligned}$$

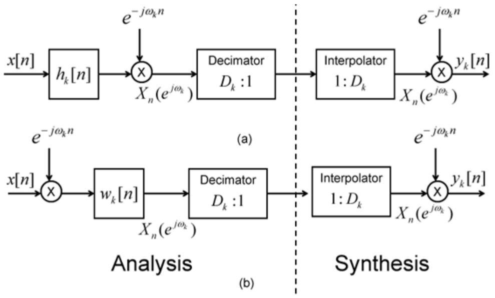
giving $\tilde{h}(n) = p(n)w(n) = \delta(n - r_0 N)$

$$\tilde{H}(e^{j\omega}) = e^{-j\omega r_0 N} \Rightarrow y(n) = x(n - r_0 N)$$

Summary of FBS Reconstruction

- for perfect reconstruction using FBS methods
  1. *w(n)* does not need to be either time-limited or frequency-limited to exactly reconstruct *x(n)* from $X_n(e^{j\omega_k})$
  2. *w(n)* just needs equally spaced zeros, spaced *N* samples apart for theoretically perfect reconstruction
- exact reconstruction of the input is possible with a number of frequency channels less than that required by the sampling theorem
- key issue is how to design digital filters that match these criteria

Practical Implementation of FBS



FBS and OLA Comparisons

FBS and OLA Comparisons

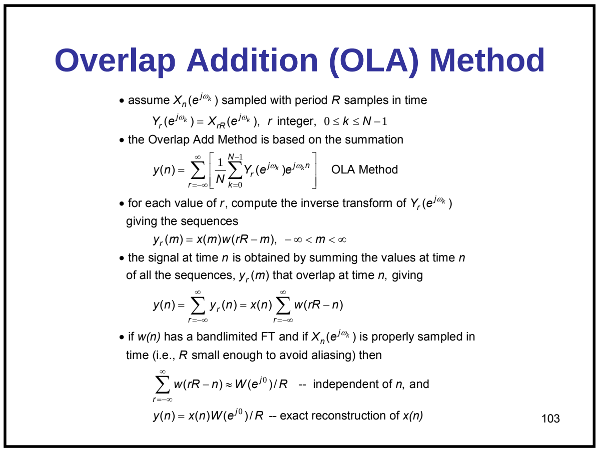
- filter bank summation method $\xleftrightarrow{\text{duals}}$ overlap addition method
  - -- one depends on sampling relation in frequency
  - -- one depends on sampling relation in time
- FBS requires sampling in frequency be such that the window transform $W(e^{j\omega})$ obeys the relation

$$\frac{1}{N}\sum_{k=0}^{N-1} W(e^{j(\omega-\omega_k)}) = w(0) \quad \text{any } \omega$$

- OLA requires that sampling in time be such that the window obeys the relation

$$\sum_{r=-\infty}^{\infty} w(rR - n) = W(e^{j0})/R \quad \text{any } n$$

- the key to Short-Time Fourier Analysis is the ability to modify the short-time spectrum (via quantization, noise enhancement, signal enhancement, speed-up/slow-down,etc) and recover an "unaliased" modified signal



Overlap Addition (OLA) Method

- assume $X_n(e^{j\omega_k})$ sampled with period R samples in time

  $Y_r(e^{j\omega_k}) = X_{rR}(e^{j\omega_k})$, r integer, $0 \le k \le N-1$
- the Overlap Add Method is based on the summation

  $$y(n) = \sum_{r=-\infty}^{\infty} \left[\frac{1}{N} \sum_{k=0}^{N-1} Y_r(e^{j\omega_k}) e^{j\omega_k n} \right] \quad \text{OLA Method}$$
- for each value of r, compute the inverse transform of $Y_r(e^{j\omega_k})$ giving the sequences

  $y_r(m) = x(m)w(rR-m), \ -\infty < m < \infty$
- the signal at time n is obtained by summing the values at time n of all the sequences, $y_r(m)$ that overlap at time n, giving

  $$y(n) = \sum_{r=-\infty}^{\infty} y_r(n) = x(n) \sum_{r=-\infty}^{\infty} w(rR-n)$$
- if *w(n)* has a bandlimited FT and if $X_n(e^{j\omega_k})$ is properly sampled in time (i.e., R small enough to avoid aliasing) then

  $\sum_{r=-\infty}^{\infty} w(rR-n) \approx W(e^{j0})/R$ -- independent of n, and

  $y(n) = x(n)W(e^{j0})/R$ -- exact reconstruction of *x(n)*