

SHORT-TIME FOURIER TRANSFORM(STFT)

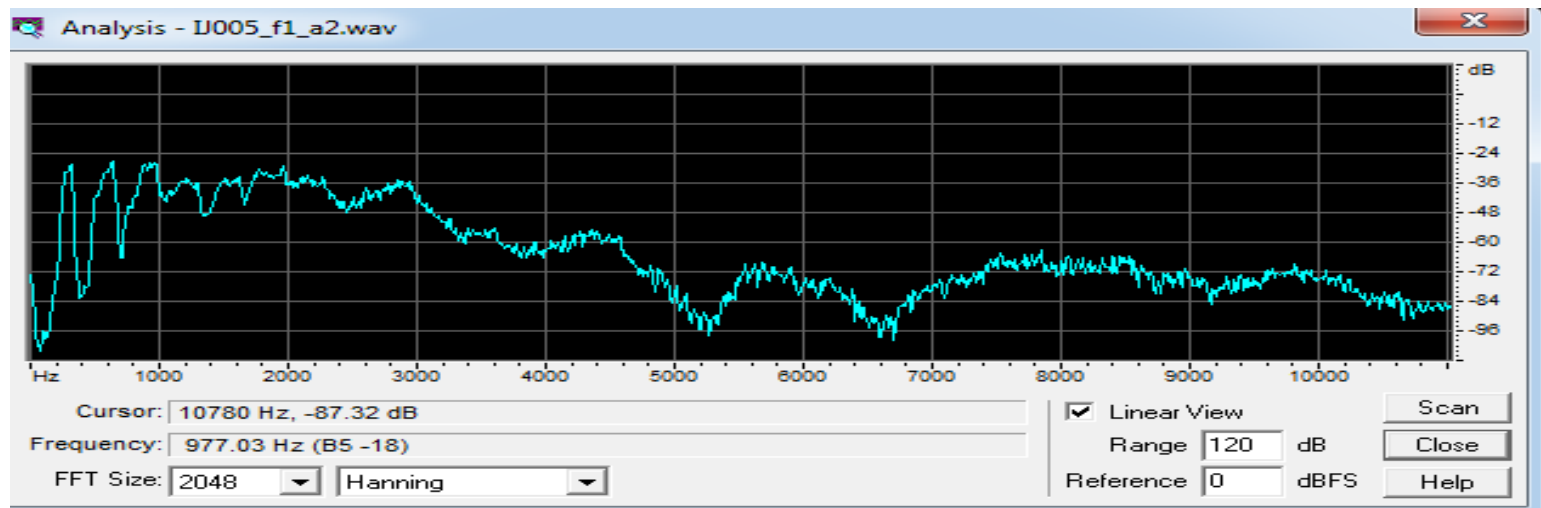
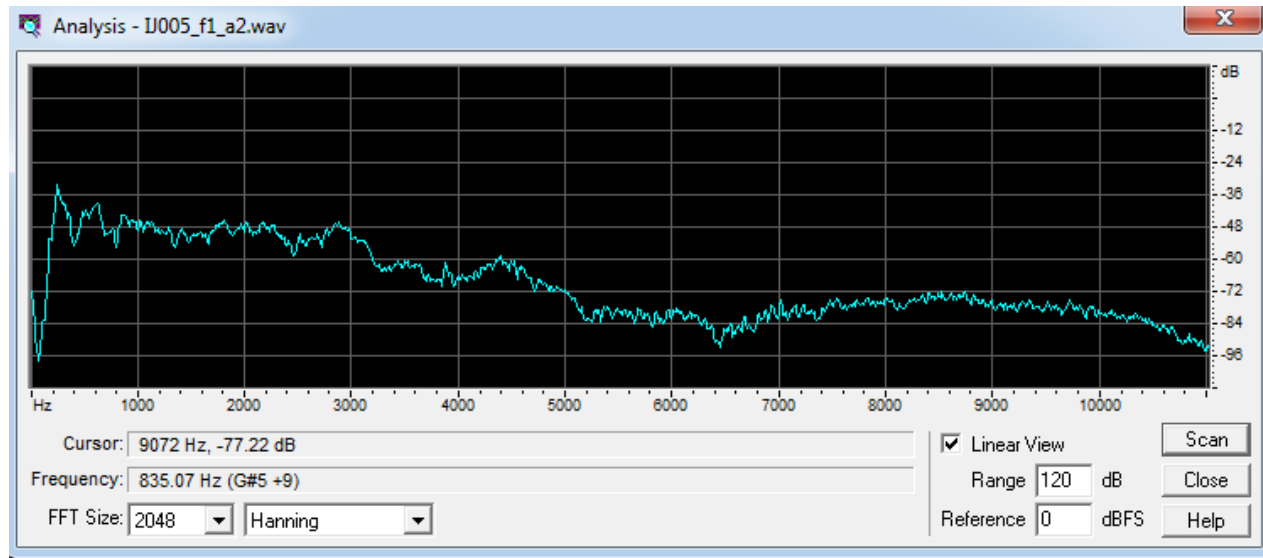
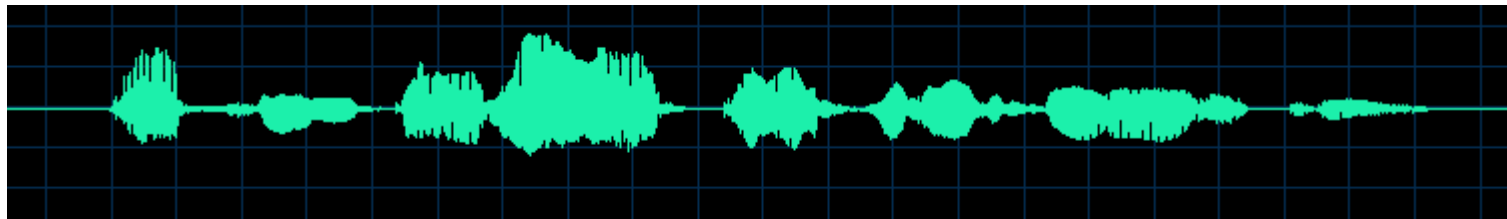
STFT → (a) Analysis (b) synthesis

(a) Analysis:- FT view and Filtering view

(b) Synthesis:-Filter bank summation (FBS)
Method and OLA Method

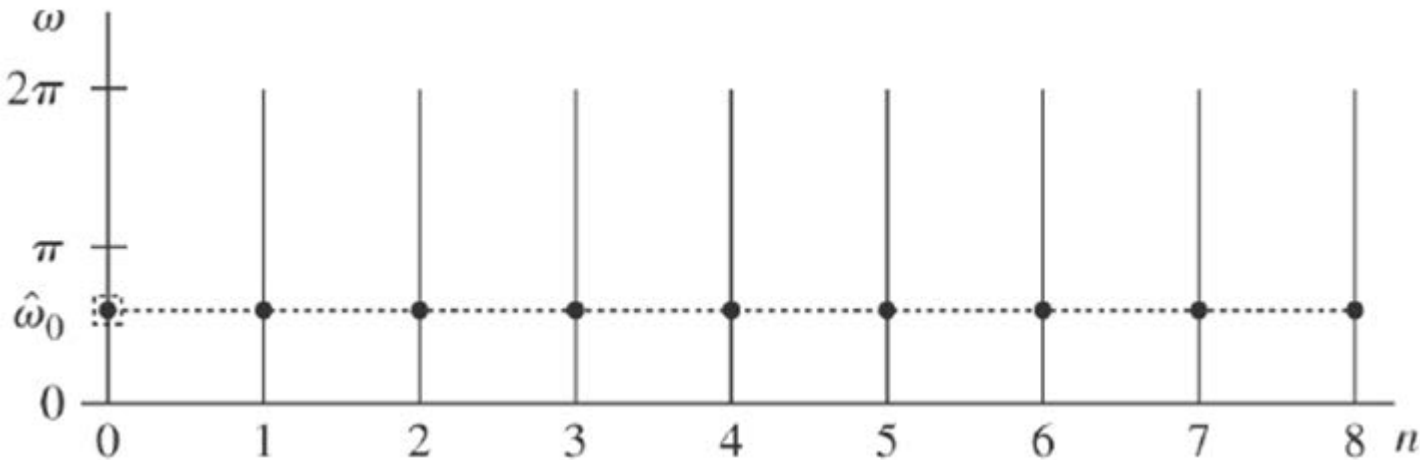
Short-Time Fourier Transform

- ❑ 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
- ❑ Define a **time-dependent Fourier transform (TDFT or STFT)** of speech that changes periodically as the speech properties change over time



STFT is a function of two variables, the time index, n which is discrete and the frequency variable ω which is continuous

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



Different time origins of STFT

STFT can be viewed as having two different time origins

1. Time origin tied to signal

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

2. Time origin tied to window

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

STFT Analysis

DFT view

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

□ $w[n]$ is non zero only in the interval $[0, N-1]$ where N is the window length

□ Time reversing the analysis window $w[m]$ and multiplying it with $x[m]$

$$\begin{array}{ccc} X(n, k) & = & X(n, \omega) \Big|_{\omega = \frac{2\pi}{N}k} \\ \text{DFT} & & \text{STFT} \end{array}$$

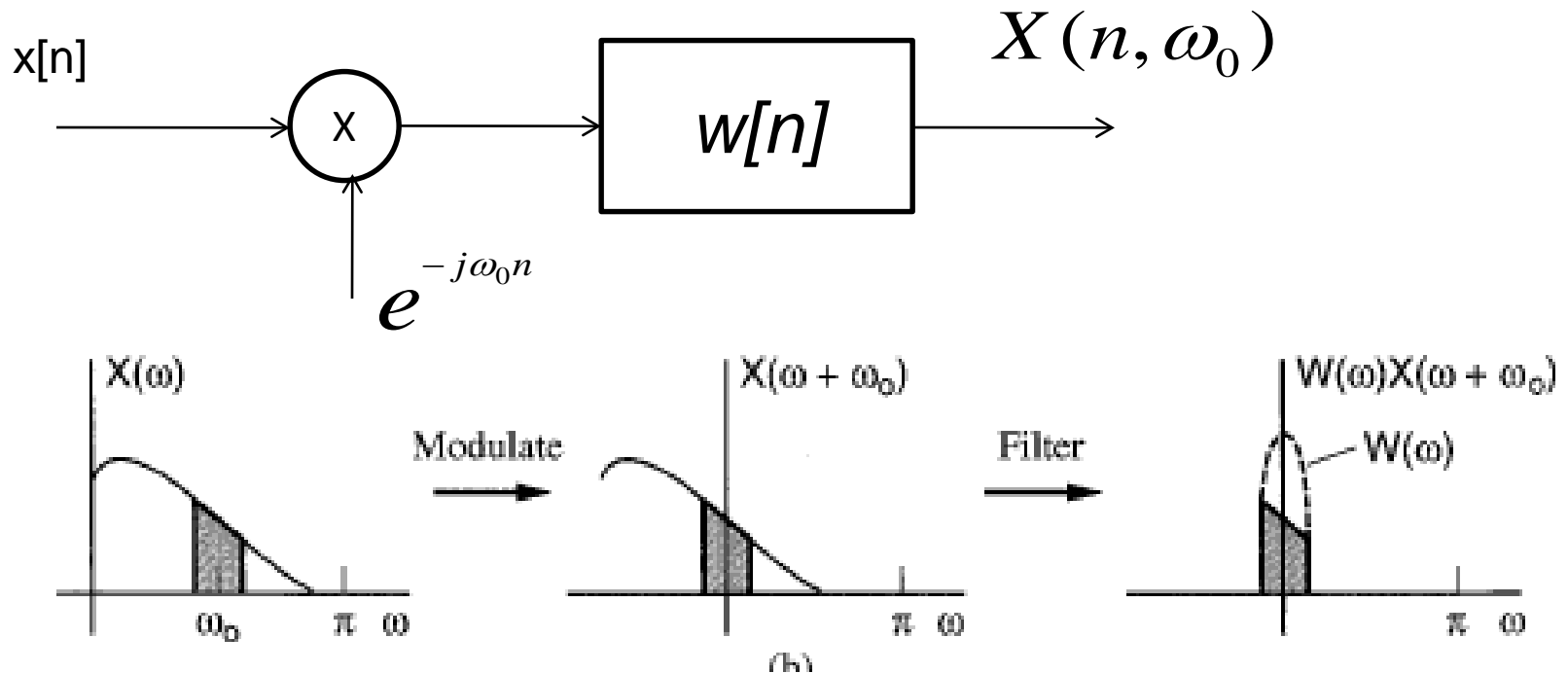
$$X(n, k) = \sum_{m=-\infty}^{\infty} x[m]w[n-m]e^{-j\frac{2\pi}{N}km}$$

Filtering view

$$X(n, \omega_0) = \sum_{m=-\infty}^{\infty} (x[m]e^{-j\omega_0 m})w[n-m]$$

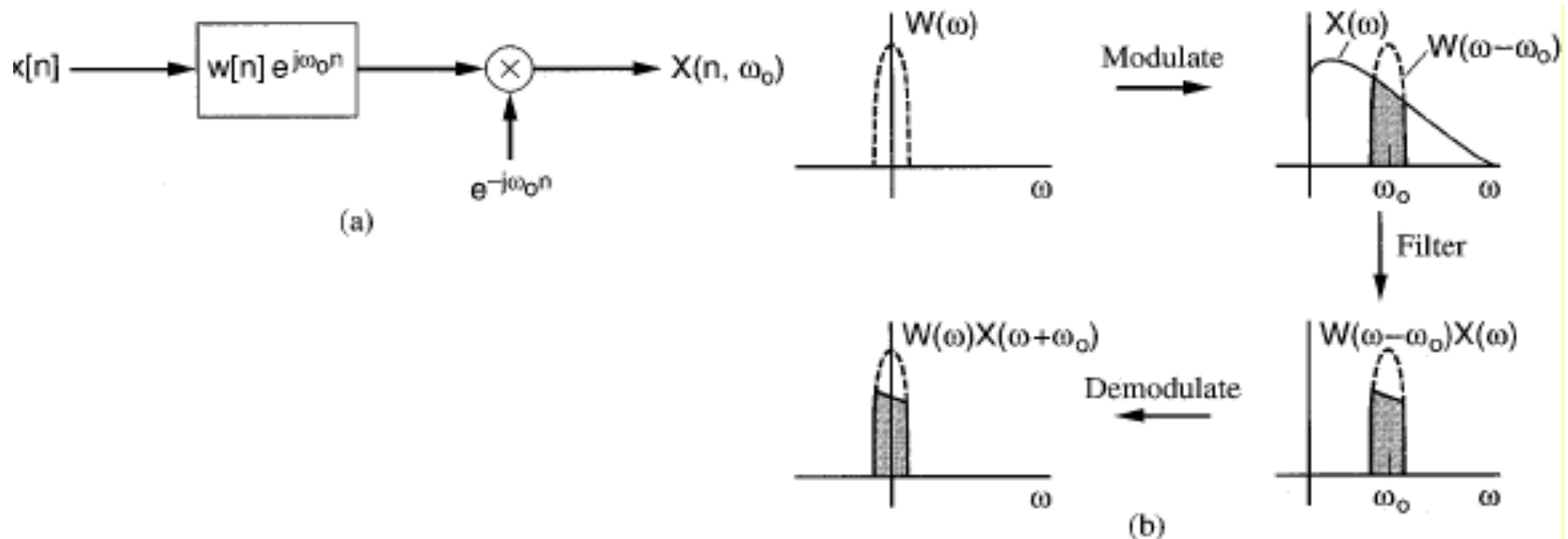
$$X(n, \omega_0) = (x[n]e^{-j\omega_0 n}) * w[n]$$

The signal $x[n]$ is first modulated with $e^{-j\omega_0 n}$, and then passed through a filter with impulsive response $w[n]$.



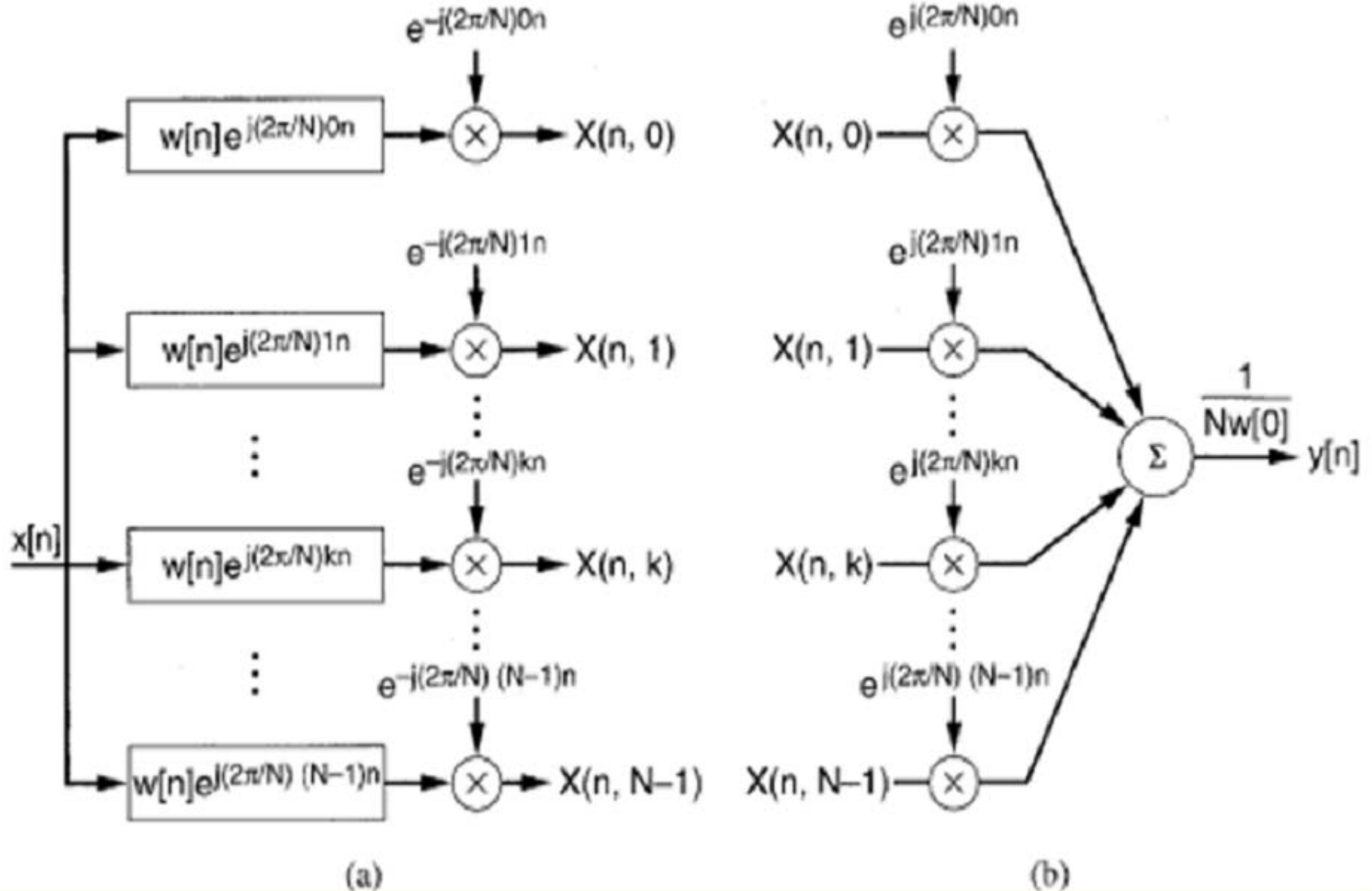
$$X(n, \omega_0) = e^{-j\omega_0 n} (x[n] * w[n] e^{j\omega_0 n})$$

That is, the sequence $x[n]$ is first passed through the filter $w[n]$ with a linear phase factor. The output is then modulated by $e^{-j\omega_0 n}$.



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

Analysis with the discrete STFT



General properties of the filtered sequence

1. If $x[n]$ has length N and $w[n]$ has length M then $X(n, \omega)$ has length $N+M-1$
2. The bandwidth of the sequence $X(n, \omega_0)$ is less than or equal to the bandwidth of $w[n]$
3. The sequence $X(n, \omega_0)$ has the spectrum centered at origin

Time-Frequency Resolution Tradeoffs

$$X(n, \omega) \xrightarrow{DFT} f_n[m] = x[m]w[n-m]$$

$$X(\omega) \xrightarrow{DFT} \text{ of } x[m]$$

$$W(-\omega)e^{j\omega n} \xrightarrow{DFT} \text{ of } w[n-m]$$

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

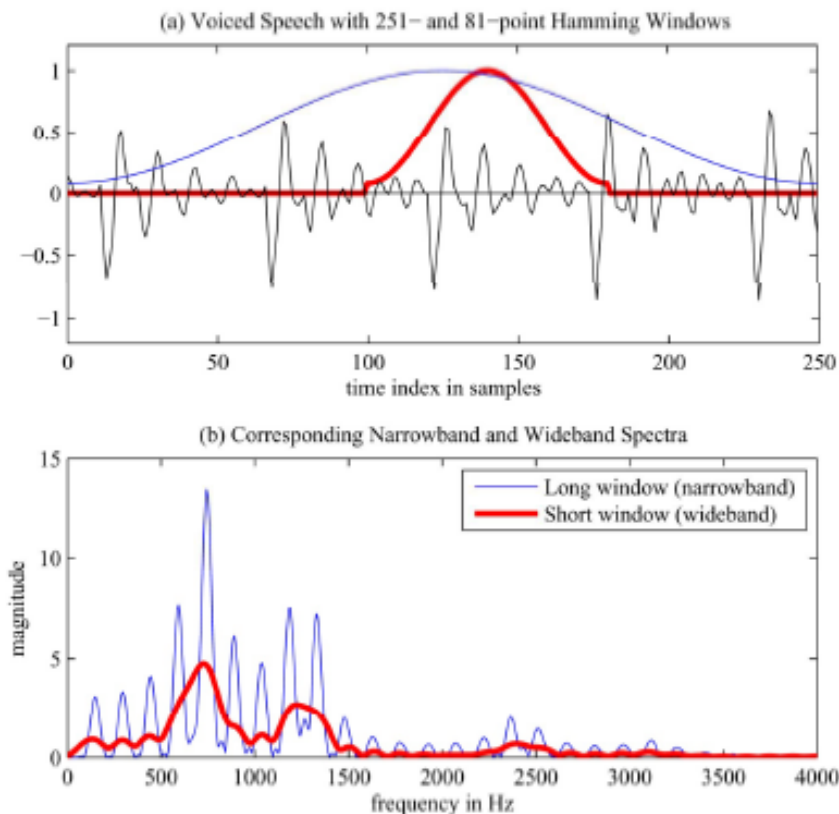
$$X(n, \omega) = X(\omega) \quad \text{then} \quad W(\omega) \quad \text{should be impulse}$$

A fundamental problem of STFT and other time-frequency analysis techniques is the selection of the windows to achieve a good tradeoff between time and frequency resolution.

Role of Window in STFT

The window $w(n-m)$ does the following:

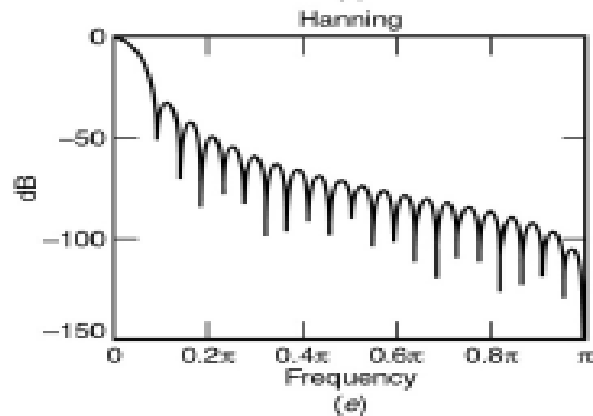
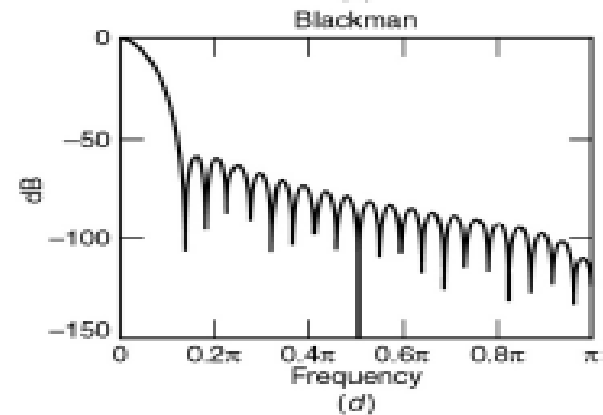
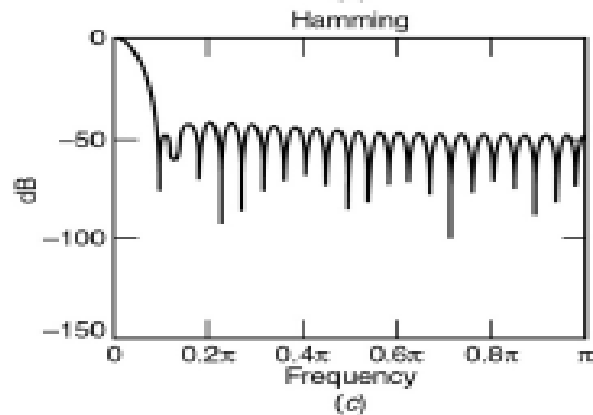
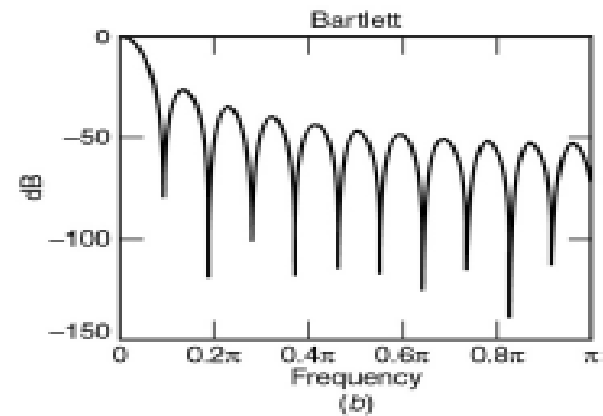
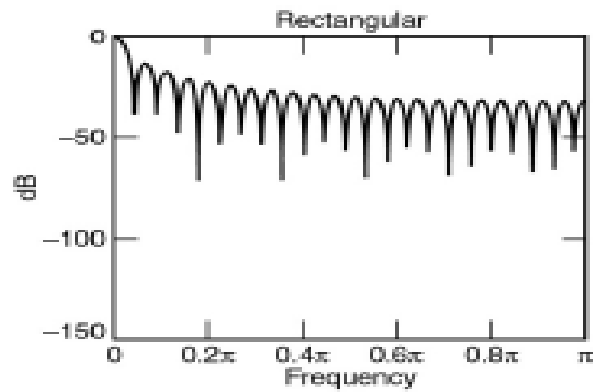
- ❑ Chooses portion of the signal to be analyzed
- ❑ Window shape determines the nature of the $X(n, \omega)$



Window Function for FIR Filter Design

Name of Window	Window function
Bartlett(triangular)	$1 - \frac{2 \left n - \frac{M-1}{2} \right }{M-1}$
Blackman	$0.42 - 0.5 \cos \frac{2\pi n}{M-1} + 0.08 \cos \frac{4\pi n}{M-1}$
Hamming	$0.54 - 0.46 \cos \frac{2\pi n}{M-1}$
Hanning	$\frac{1}{2} \left(1 - \cos \frac{2\pi n}{M-1} \right)$
Kaiser	$\frac{I_0 \left[\alpha \sqrt{(M-1)^2 - \left(n - \frac{M-1}{2} \right)^2} \right]}{I_0 \left[\alpha \left(\frac{M-1}{2} \right) \right]}$

Common Windows (Frequency)



STFT Synthesis

For a given value of n , $X(n, \omega)$ has the same properties as a normal Fourier transform, we can recover the input sequence exactly.

For each n , we take the inverse Fourier transform of $X(n, \omega)$ from the STFT.

Then obtain $f[m] = x[m]w[n-m]$.

Evaluating $f[m]$ at $m = n$, obtain $x[n]w[0]$. Assuming $w[0] \neq 0$

Then $x[n] = f[n] / w[0]$

$$f_n[m] = x[m]w[n-m]$$



DFT



IDFT at $m=n$



$$x[m]w[0]$$

If the STFT is unique representation of $x[n]$ then it invertible.

$$x[n] = \frac{1}{2\pi w[0]} \int_{-\pi}^{\pi} X(n, \omega) e^{i\omega n} d\omega$$

Synthesis equation for discrete-time STFT

With the requirement that $w[0] \neq 0$, the sequence $x[n]$ can be recovered exactly from $X(n, \omega)$, if $X(n, \omega)$ is known for all values of ω over one complete period

- ❑ **Sample-by-sample recovery process**

- ❑ **$X(n, \omega)$, must be known for every value of n and for all ω**

- ❑ To reduce the computational complexity, the STFT is not computed at every time sample, but rather at a certain time decimation rate. In some cases, the discrete STFT may not be invertible, i.e. there are certain constraints on the frequency-sampling and time-decimation rates.

- ❑ By selecting appropriate constraints on the frequency sampling and time decimation rates the discrete STFT is invertible

Short-Time Synthesis

- **Example 1.**
 - Consider the case when $w[n]$ is band limited with bandwidth of B .

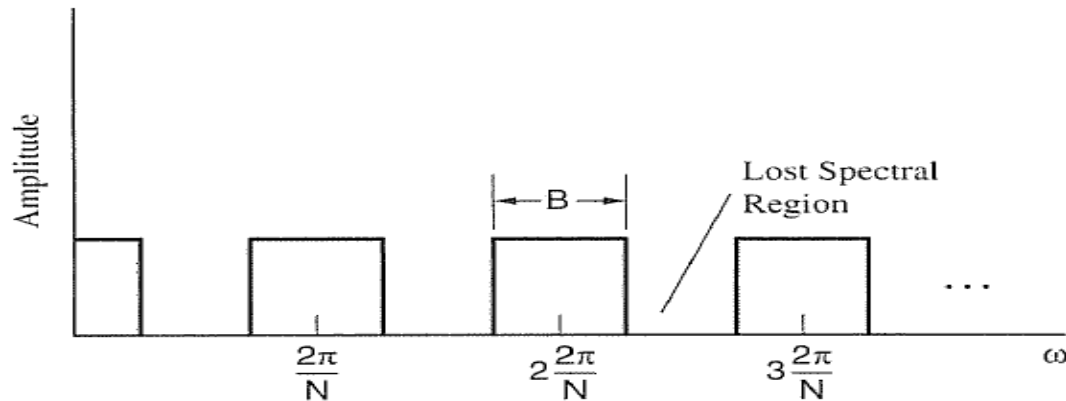


Figure 7.10 Undersampled STFT when the frequency sampling interval $\frac{2\pi}{N}$ is greater than the analysis-filter bandwidth B .

SOURCE: S.H. Nawab and T.F. Quatieri, "Short-Time Fourier Transform" [13].
©1987, Pearson Education, Inc. Used by permission.

If there are frequency components of $x[n]$ which do not pass through any of the filter regions of the discrete STFT then it is not a unique representation of $x[n]$, and $x[n]$ is not invertible.

Example 2.

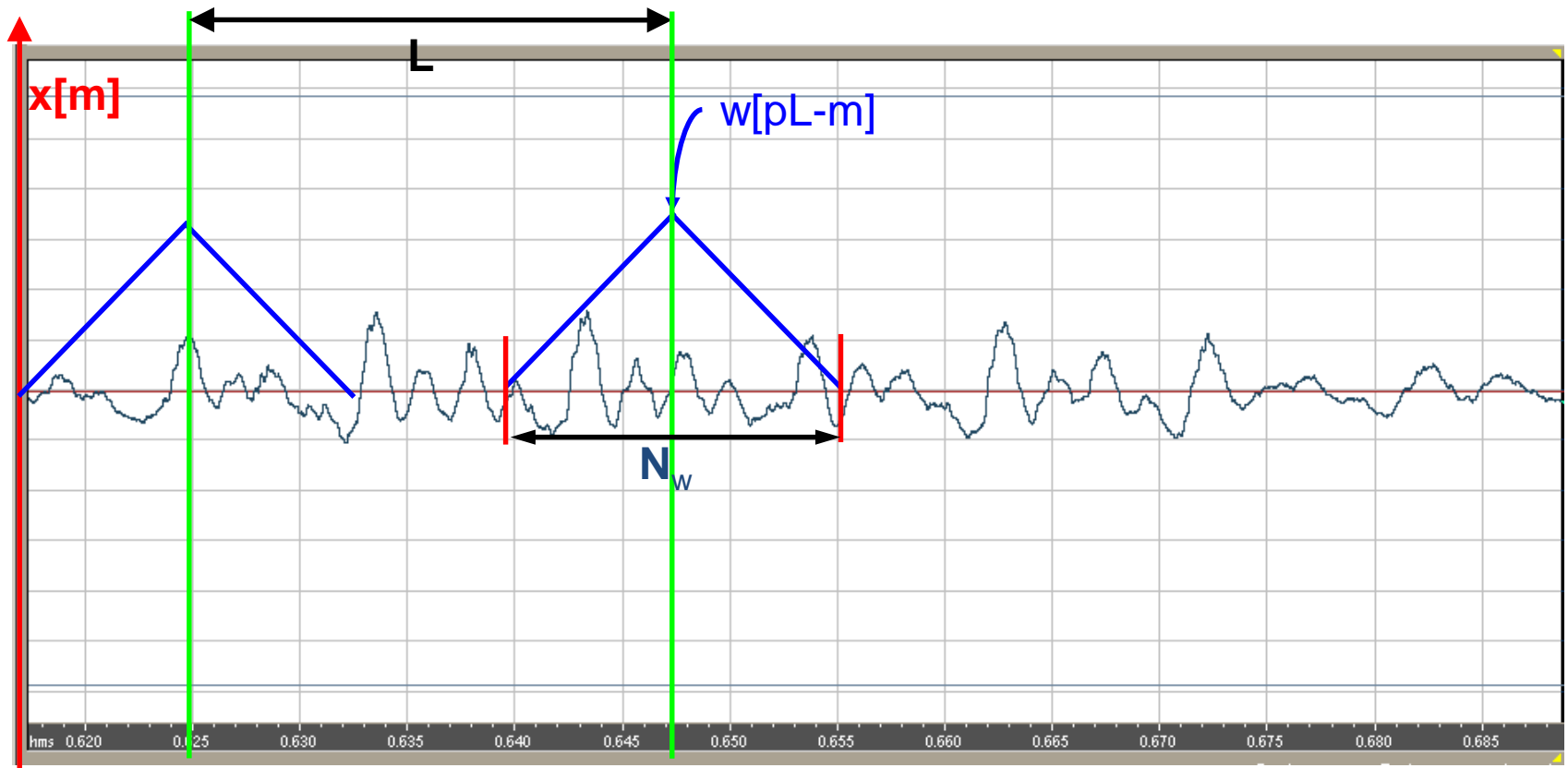
Consider $X(n,k)$ decimated in time by factor L , i.e., STFT is applied every L samples.

$w[n]$ is non-zero over its length N_w .

If $L > N_w$ then there are gaps in time where $x[n]$ is not considered. Thus in such cases again $x[n]$ is not invertible.

$x[n]$ is invertible if temporal decimation factor L is equal to or less than the size of the analysis window N_w and the frequency sampling interval $2\pi/N \leq 2\pi/N_w$

$$L > N_w$$



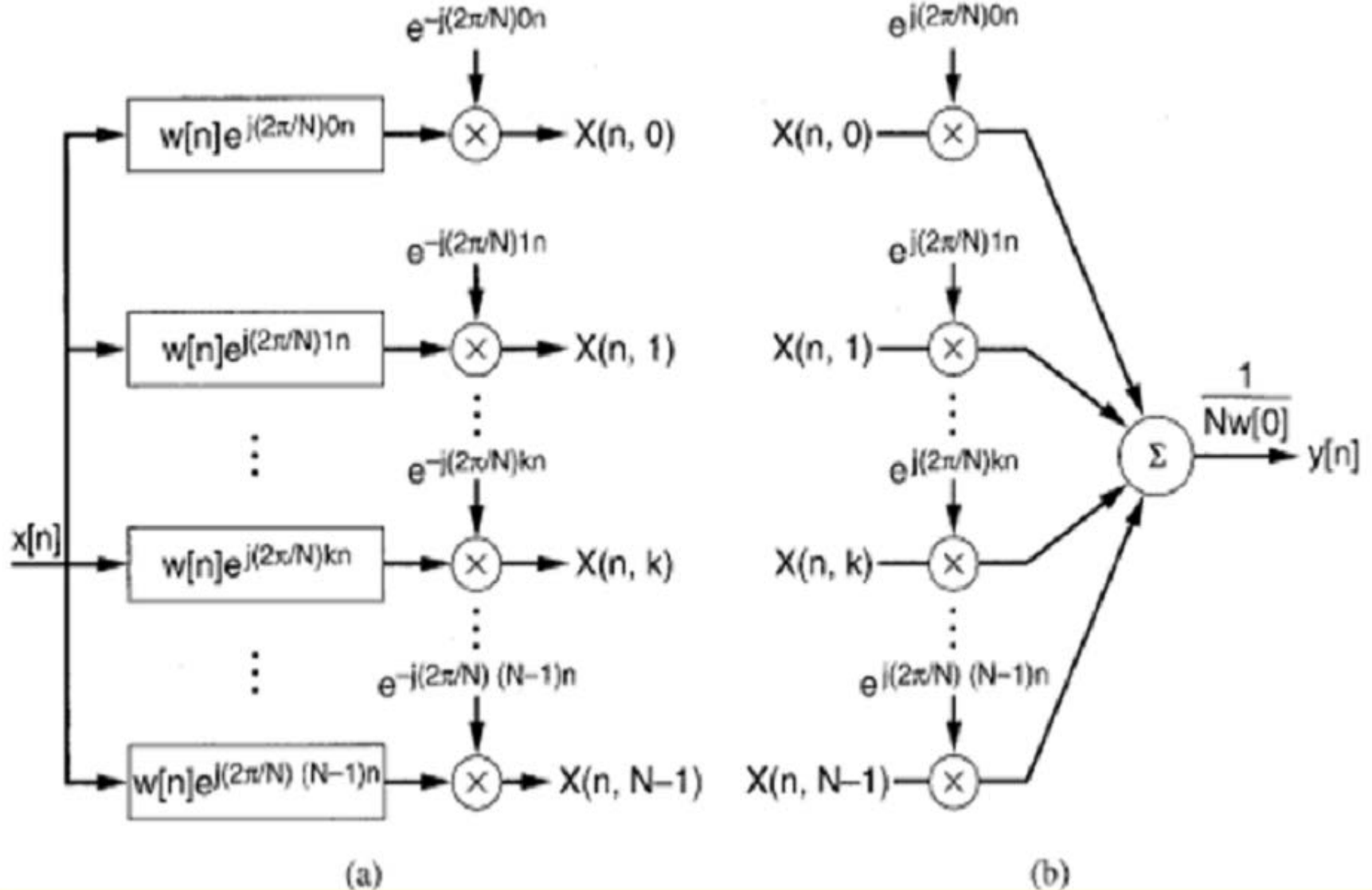
Two common methods for STFT synthesis

- ❑ Filter Bank Summation (FBS) method
- ❑ Overlap-Add (OLA) method.

Filter Bank Summation (FBS) Method

- Traditional short-time synthesis method that is commonly referred to as the Filter Bank Summation (FBS).
- FBS is best described in terms of the filtering interpretation of the discrete STFT.
 - The discrete STFT is considered to be the set of outputs of a bank of filters.
 - The output of each filter is modulated with a complex exponential
 - Modulated filter outputs are summed at each instant of time to obtain the corresponding time sample of the original sequence

Analysis with the discrete STFT



Filter Bank Summation (FBS) method

$$y[n] = \frac{1}{Nw[0]} \sum_{k=0}^{N-1} X(n, k) e^{\frac{j2\pi nk}{N}}$$

$$y[n] = \frac{1}{Nw[0]} \sum_{k=0}^{N-1} \underbrace{\left\{ \sum_{m=-\infty}^{\infty} x[m] w[n-m] e^{-j\frac{2\pi}{N} km} \right\}}_{X(n, k)} e^{j\frac{2\pi}{N} kn}$$

Interchanging summation operation this equation reduces to:

$$y[n] = \frac{1}{Nw[0]} x[n] * \sum_{k=0}^{N-1} w[n] e^{\frac{j2\pi nk}{N}}$$

Finite sum over the complex exponential reduce to an impulse train with period N

$$y[n] = \frac{1}{Nw[0]} x[n] * w[n] \sum_{r=-\infty}^{\infty} \delta[n - rN]$$

$y[n]$ is the output of the convolution of $x[n]$ with a product of the analysis window with a periodic impulse sequence

$$w[n] \sum_{r=-\infty}^{\infty} \delta[n-rN]$$

reduces to $\delta[n]$ if:

$$w[n] \sum_{r=-\infty}^{\infty} \delta[n-rN] = w[0]\delta[n]$$

$$y[n] = x[n]$$

❑ Window length $N_w \leq N$, or

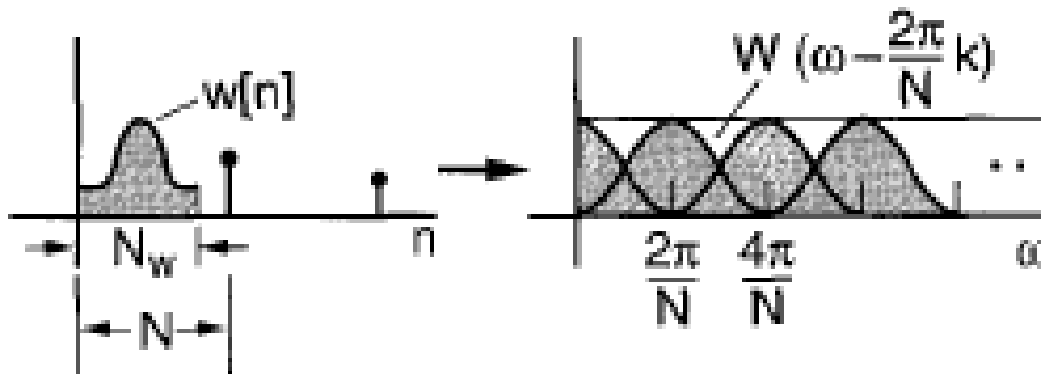
❑ For $N_w > N$, must have $w[rN] = 0$ for $r \neq 0$, that is

$$w[rN] = 0 \text{ for } r = -1, 1, -2, 2, \dots$$

This constraint is known as the **FBS constraint**.

$$\sum_{k=0}^{N-1} W\left(\omega - \frac{2\pi}{N}k\right) = Nw[0]$$

This expression states that the frequency responses of the analysis filters should sum to a constant across the entire bandwidth.



Shifted version of the Fourier transform of the analysis window were required to add up to a constant

Generalized FBS Method

- Note:
$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\sum_{r=-\infty}^{\infty} f[n, n-r] X(r, \omega) \right] e^{j\omega n} d\omega$$
- “Smoothing” function $f[n, m]$ is referred to as the ***time-varying synthesis*** filter.
- It can be shown that any $f[n, m]$ that fulfills the condition below makes the synthesis equation above valid

$$\sum_{m=-\infty}^{\infty} f[n-m] w[m] = 1$$

- Basic FBS method can be obtained by setting the synthesis filter to be a non-smoothing filter:

$$f[n, m] = \delta[m]$$

Consider the discrete STFT with decimation factor L .
Generalized FSB of the synthesized signal is given by:

$$y[n] = \frac{L}{N} \sum_{r=-\infty}^{\infty} \sum_{k=0}^{N-1} f[n, n-rL] X(rL, k) e^{j \frac{2\pi}{N} nk}$$

Furthermore, consider time invariant smoothing filter:

$$f[n, m] = f[m]$$

That is:

$$f[n, n-rL] = f[n-rL]$$

Generalized FBS Method

Thus

$$y[n] = \frac{L}{N} \sum_{r=-\infty}^{\infty} \sum_{k=0}^{N-1} f[n-rL] X(rL, k) e^{j \frac{2\pi}{N} nk}$$

This equation holds when the following constrain is satisfied by the analysis and synthesis filters as well as the temporal decimation and frequency sampling factors:

$$L \sum_{r=-\infty}^{\infty} f[n-rL] w[rL-n+pN] = \delta[p], \quad \forall n$$

- For $f[m] = \delta[m]$ and $L=1$ this method reduces to the basic FBS method.

Generalized FBS Method

If $L > 1$ $f[n]$ is an interpolating filter \Rightarrow Interpolation FBS
Methods:

1. Helical Interpolation (Partnoff)
2. Weighted Overlap-add Method (Crochiere)

Overlap-Add (OLA) method

- ❑ Take inverse DFT for each fixed time in the discrete STFT. Instead of dividing out the analysis window from each of the resulting short time sections perform an overlap add operation between the short sections
- ❑ Overlap and add operation effectively eliminates the analysis window

$$x[n] = \frac{1}{2\pi W[0]} \int_{-\pi}^{\pi} X(n, \omega) e^{j\omega n} d\omega$$

If $x[n]$ is averaged over many short-time segments and normalized by $W(0)$ then

$$x[n] = \frac{1}{2\pi W[0]} \int_{-\pi}^{\pi} \sum_{p=-\infty}^{\infty} X(p, \omega) e^{j\omega p} d\omega$$

where

$$W(0) = \sum_{n=-\infty}^{\infty} w[n]$$

Discrete version of OLA is given by:

$$y[n] = \frac{1}{W(0)} \sum_{p=-\infty}^{\infty} \underbrace{\left\{ \frac{1}{N} \sum_{k=0}^{N-1} X(p, k) e^{j\frac{2\pi}{N}kn} \right\}}_{\text{IDFT: } f_p[n] = x[n]w[p-n]}$$

$$y[n] = \frac{1}{W(0)} \sum_{p=-\infty}^{\infty} x[n] w[p - n]$$

$$y[n] = x[n] \left(\frac{1}{W(0)} \right) \sum_{p=-\infty}^{\infty} w[p - n]$$

$$y[n] = x[n] \quad \text{if}$$

$$\sum_{p=-\infty}^{\infty} w[p - n] = W(0)$$

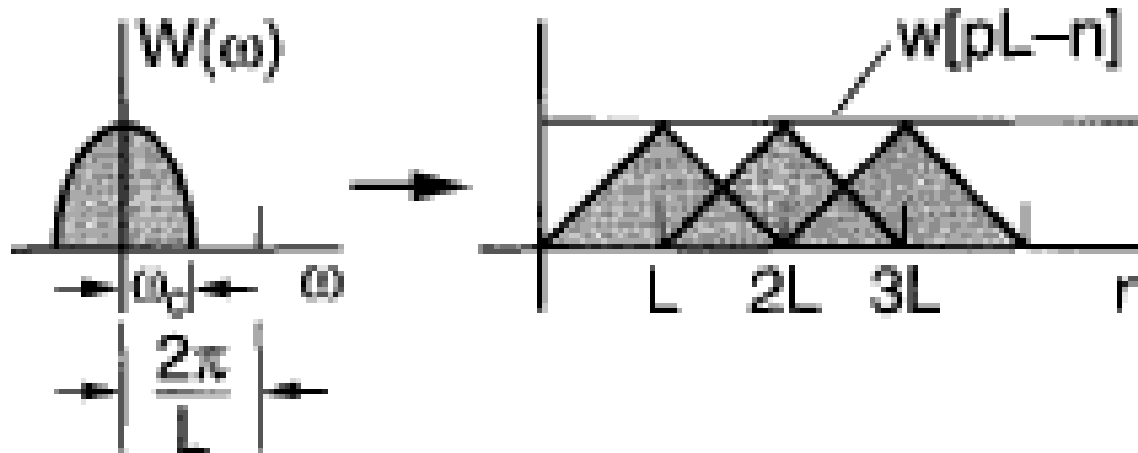
Sum of values of a sequence must be the first value of its Fourier Transform

$$\sum_{p=-\infty}^{\infty} w[p - n] = W(0)$$

For decimation in time by factor of $L \rightarrow \sum_{p=-\infty}^{\infty} w[pL - n] = \frac{W(0)}{L}$

$$y[n] = \frac{L}{W(0)} \sum_{p=-\infty}^{\infty} \left[\frac{1}{N} \sum_{k=0}^{N-1} X(pL, k) e^{j2\pi k / N} \right]$$

The above equation depicts general constrain imposed by OLA method. It requires that the sum of all the analysis windows (obtained by sliding $w[n]$ with L -point increments) to add up to a constant



For OLA methods it can be shown that its constrained is satisfied by all-finite-bandwidth analysis windows whose maximum frequency is less than $2\pi/L$ (where L is temporal decimation factor).

This finite bandwidth can be relaxed provided

$$W(\omega - 2\pi k / L) = 0 \quad \text{at} \quad \omega = 2\pi k / L$$

Analogous to FBS constrain for $N_w > N$ where the window $w[n]$ is required to take on value zero at $n = \pm N, \pm 2N, \pm 3N, \dots$

FBS Method

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

Adding Frequency component for each n

$$\sum_{k=0}^{N-1} W\left(\omega - \frac{2\pi}{N}k\right) = Nw[0]$$

Constraint

$$N_w < N \longrightarrow y[n] = x[n]$$

OLA Method

$$y[n] = \frac{1}{W[0]} \sum_{p=-\infty}^{\infty} x[n] w[pL - n]$$

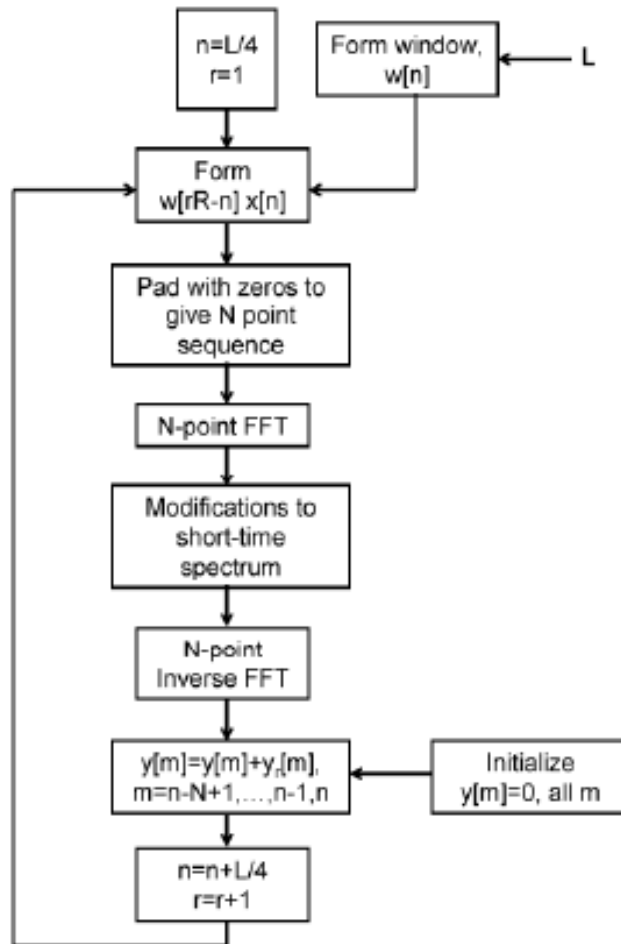
Adding time component for each n

$$\sum_{p=-\infty}^{\infty} w[pL - n] = \frac{W[0]}{L}$$

Constraint

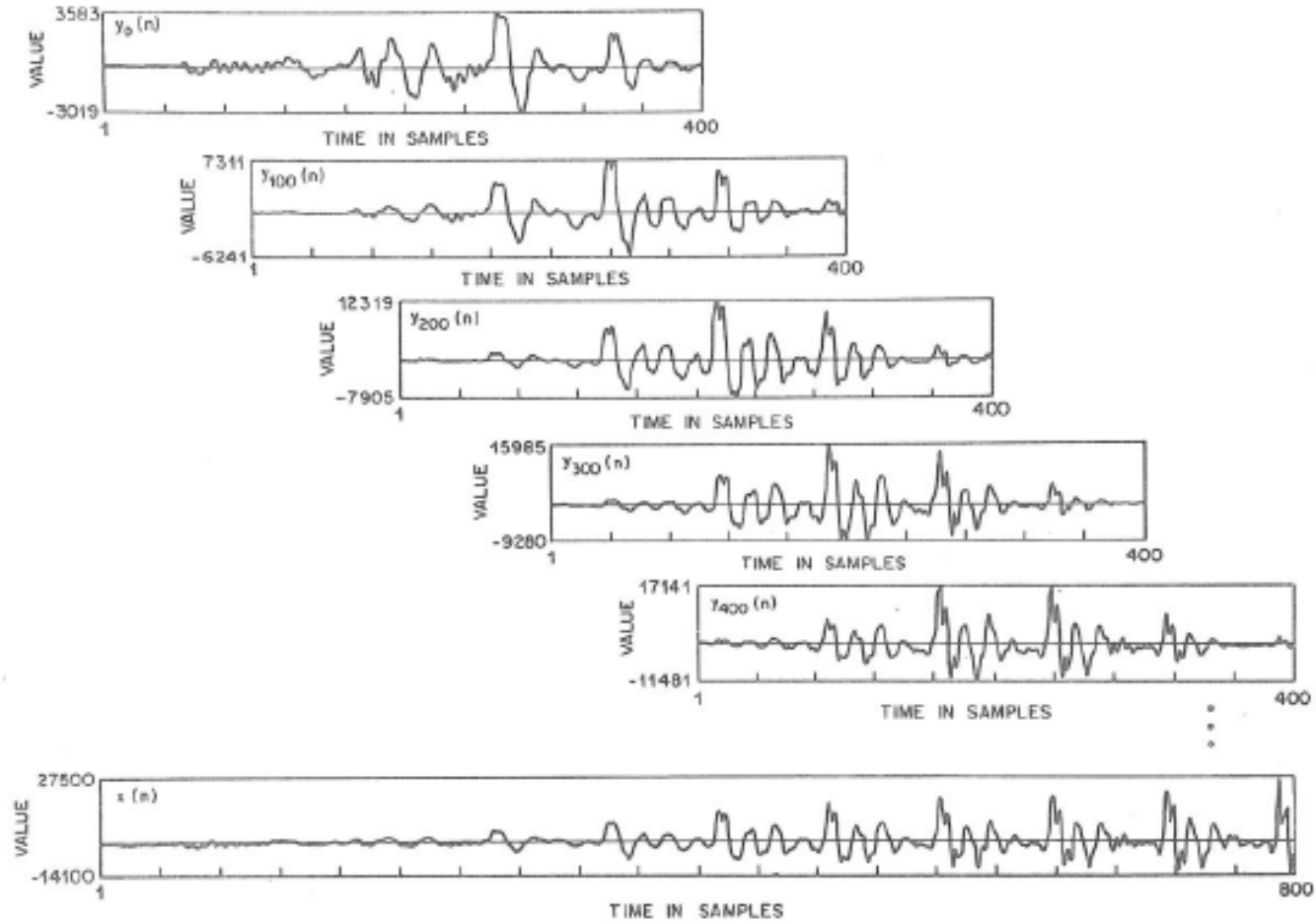
$$\omega_c < \frac{2\pi}{N} \longrightarrow y[n] = x[n]$$

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 $n=L/4$

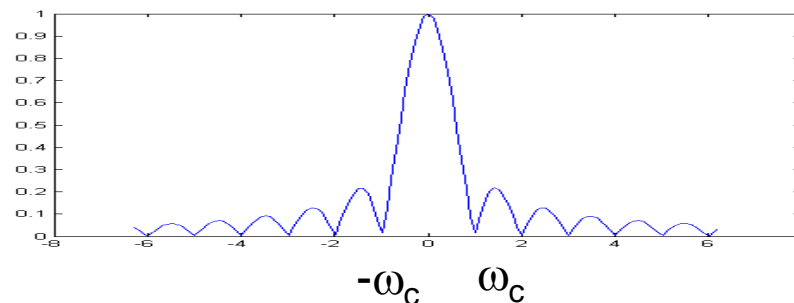
Overlap Addition (OLA) Method



Time-Frequency Sampling

- Summary of sampling issues for those two methods that gives motivation for our earlier statement that sufficient but not necessary conditions for invertability of the discrete STFT are:
 1. The analysis window is non-zero over its finite length N_w .
 2. The temporal decimation factor $L \leq N_w$
 3. The frequency sampling interval $2\pi/N \leq 2\pi/N_w$

- Consider windowed/short-time signal:
 - $f_n[m] = w[m]x[n-m]$, and
 - $X(n, \omega)$ – Fourier transform of $f_n[m]$
 - Analysis window duration of N_w
- From Fourier transform point of view:
 - Reconstruction of $f_n[m]$ from $X(n, k)$ requires a frequency sampling of at least $2\pi/N_w$ or finer.
- From Time-domain point of view:
 - Time decimation interval L is required to meet Nyquist criterion based on the bandwidth of the window $w[n]$.
 - This implies sampling of $X(n, k)$ at a time interval $L \leq 2\pi/\omega_c$ to avoid frequency-domain aliasing of the time sequence $X(n, \omega)$
 - ω_c is the bandwidth of $W(\omega)$ $[-\omega_c, \omega_c]$



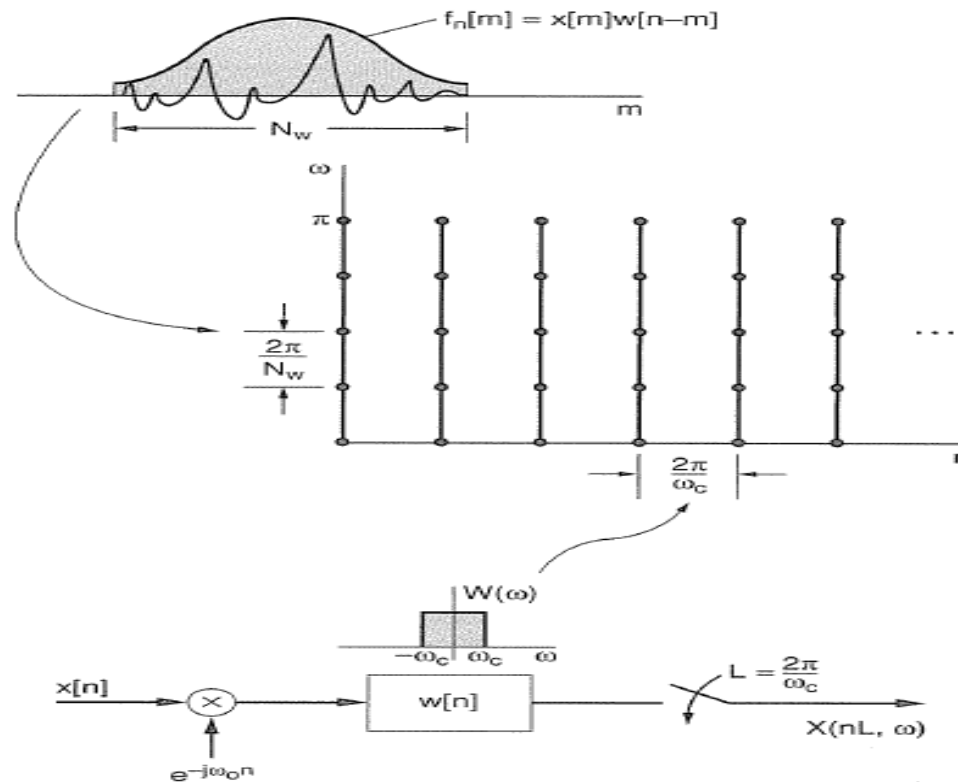


Figure 7.14 Time-frequency sampling constraints from the perspective of classical time- and frequency-domain aliasing. The time sampling must satisfy the Nyquist criterion to avoid aliasing in frequency (but the OLA constraint allows relaxing the finite filter bandwidth constraint), while the frequency sampling must be fine enough to avoid aliasing in time (but the FBS constraint allows relaxing the finite window duration condition).

Time-Frequency Sampling

- Sufficient (but not necessary) conditions for signal reconstruction are:
 1. Window is non-zero over its lengths N_w
 2. Temporal decimation factor $L \leq N_w (2\pi/\omega_c)$
 3. Frequency sampling interval $2\pi/N \leq 2\pi/N_w$
- To avoid aliasing:
 - I. ***In the time domain*** - by ensuring condition 3.
 - II. ***In the frequency domain*** - by ensuring condition 2.

Time-Frequency Sampling

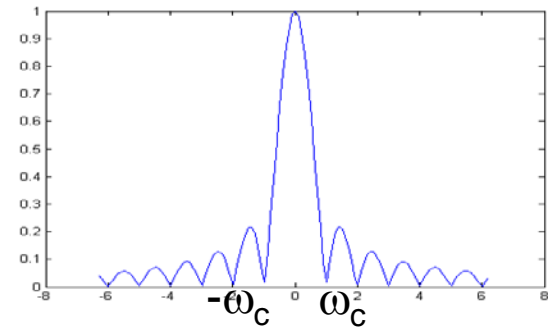
- Implication on the use of practical windows:

I. Rectangular window, N_w

\Rightarrow Assuming bandwidth equal
to the extent of the main lobe

$$B = [-2\pi/N_w, 2\pi/N_w] = 4\pi/N_w$$

$$\Rightarrow L_w \leq \frac{2\pi}{B} \leq \frac{N_w}{2} \quad ; 50\% \text{ Overlap in windows}$$



II. Hamming Window, N_w

\Rightarrow Bandwidth $B = 8\pi/N_w$

$$\Rightarrow L_w \leq \frac{2\pi}{B} \leq \frac{N_w}{4} \quad 75\% \text{ Overlap in windows}$$

Summary

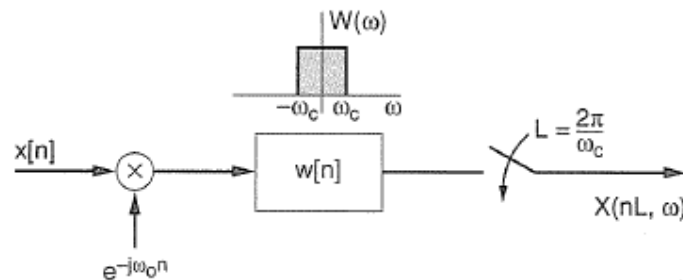
- **OLA Method (DFT of order N)**

1. No time aliasing if window length N_w so that:
 $2\pi/N \leq 2\pi/N_w$
2. No frequency-domain aliasing occurs if decimation factor L is small enough so that filter bandwidth
 $\omega_c = (2\pi/L)$
3. If zeros are allowed in $W(\omega)$ then condition 2 can be relaxed. In this case we can under-sample in frequency and still recover the sequence.

Summary

- **FBS Method**

1. No frequency-domain aliasing occurs if the decimation factor L meets the Nyquist criterion, i.e., $L \leq N_w (2\pi/\omega_c)$ where ω_c is the $w[n]$ bandwidth.
2. Not time-domain aliasing occurs if $2\pi/N \leq 2\pi/N_w$
 $\Rightarrow N_w \leq N$.
3. If zeros in $w[n]$ are allowed then condition 2 can be relaxed. In this case we can under-sample in time and still recover the sequence.




Short-Time Fourier Transform Magnitude (STFTM)

- STFTM discards (possibly) phase information, which has numerous uses in application areas:
 - Time-scale modification
 - Speech Enhancement
- In all these applications phase information estimation of speech is difficult (e.g., presence of noise in the signal)
- Furthermore, a number of techniques have been developed to obtain phase estimate from a STFT magnitude.

- Squared-Magnitude and Autocorrelation Relationship:


$$r[n, m] = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(n, \omega)|^2 e^{j\omega n} d\omega$$

Short-time autocorrelation



$$|X(n, \omega)|^2 = \sum_{m=-\infty}^{\infty} r[n, m] e^{-j\omega n}$$

Short-time magnitude



– m-autocorrelation “lag”

Short-Time Fourier Transform Magnitude (STFTM)

- Furthermore, the autocorrelation $r[n,m]$ is given by the convolution of the short-time signal:

$$r[n,m] = f_n[m] * f_n[-m]$$

where

$$f_n[m] = x[m]w[n-m]$$

Signal Representation

- Under what conditions STFTM can be used to represent a sequence uniquely?
- Note that:

$$|\mathcal{F}\{x[n]\}| = |\mathcal{F}\{-x[n]\}|$$

⇒ Ambiguity, thus STFTM is not unique representation for all cases.

- However, by imposing certain mild restrictions on:
 - the analysis window and
 - the signal,unique signal representation is indeed possible with the discrete-time STFTM.

Reconstruction from Time-Frequency Samples

- To carry out STFTM analysis on a digital computer, *discrete* STFTM must be applied.
- Uniqueness theory of STFTM can be easily extended to discrete STFTM.
 - Uniqueness of STFTM based on the short-time autocorrelation functions.
 - Autocorrelation functions can be obtained even if the STFTM is sampled in frequency (discrete STFTM) with adequate frequency sampling.
- To consider effects of temporal decimation with factor L , we note that adjacent short-time sections now have an overlap of $N_w - L$ instead of $N_w - 1$.

Application of STFT Analysis and Synthesis

- ❑ Signal Estimation from Modified STFT or STFTM
- ❑ Time Scale Modification and Enhancement of Speech
- ❑ Noise Reduction

Time-Scale Modification and Enhancement of Speech

Goal

- ❑ To either speed up or slow down a speech signal while maintaining the approximate pitch

Applications

- Change voice mail playback
- Court stenographers-play proceedings quicker
- Sound effects
- Etc...

Time-Scale Modification

- **Methods:**
 - **Cut & Paste (Fairbanks method):**
 - Discard or duplicate frames, in order to speed up or slow down the articulation respectively.
 - Problem:
 - Pitch period mismatch at adjacent frames causes distortion.
 - **Pitch-synchronous OLA (Scott & Gerber)**
 - Select frame size & location synchronous to pitch periods. Problem of pitch period mismatch is avoided.
 - Problem:
 - Pitch synchronization is not always easy.
 - **STFTM Synthesis**
 - To avoid pitch synchronization problems use only the magnitude of STFT (i.e., STFTM)
 1. Compute $|X(nL, \omega)|$ at an appropriate frame interval – decimation rate L (e.g., $L=128$ at $F_s=10000$ Hz, and N is several T_0 long)
 2. Modify decimation rate with new rate M (e.g., $M=L/2$) for a speed-up of factor of $\frac{1}{2}$:
 $|Y(nM, \omega)| = |X(nL, \omega)|$
 3. Apply the Least-Squared Error iterative estimation algorithm until $|Y(nM, \omega)|$ converged.
 - Problem:
 - Occasional reverberant characteristic of synthesized signal are perceived due to lack of STFT phase control.

Time-Scale Modification

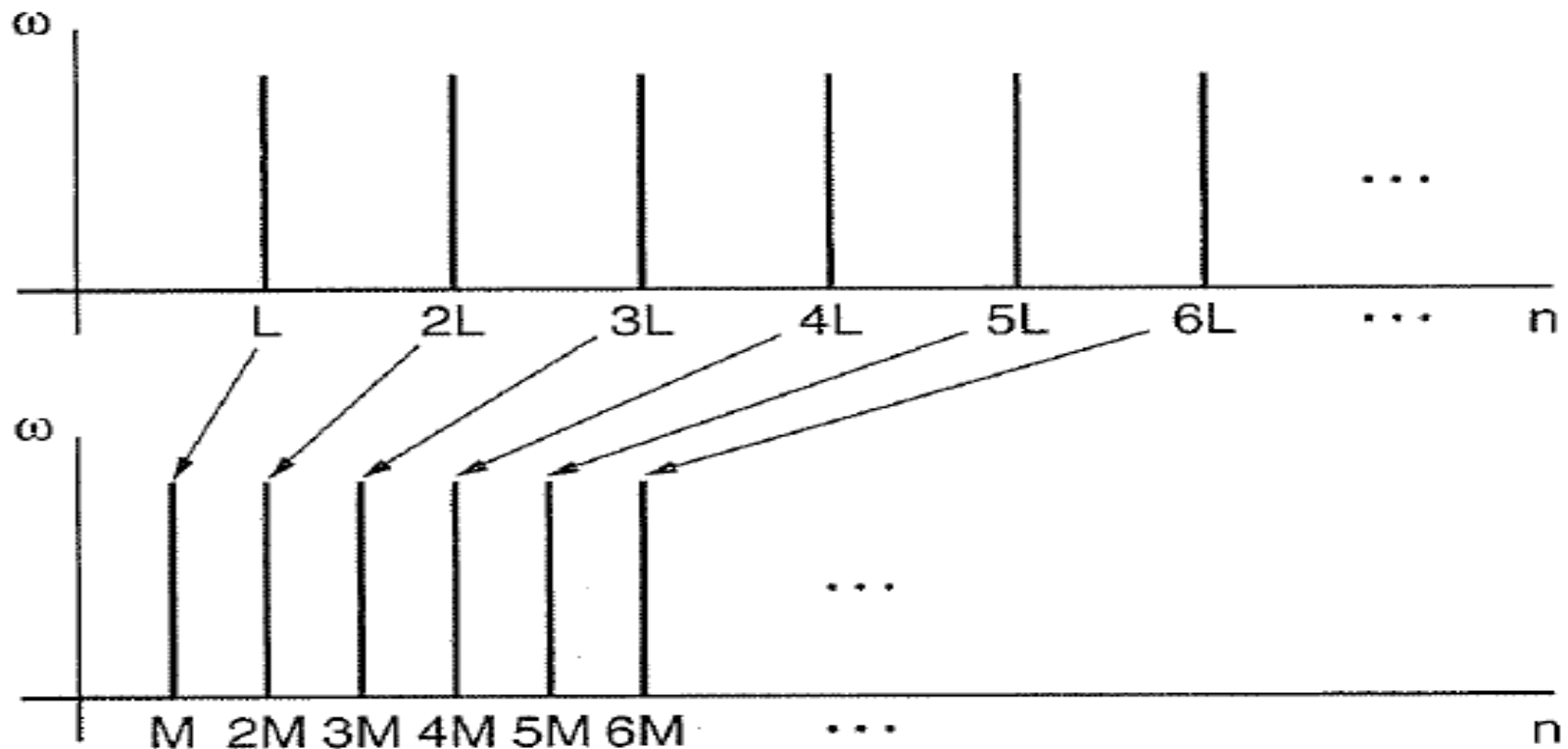
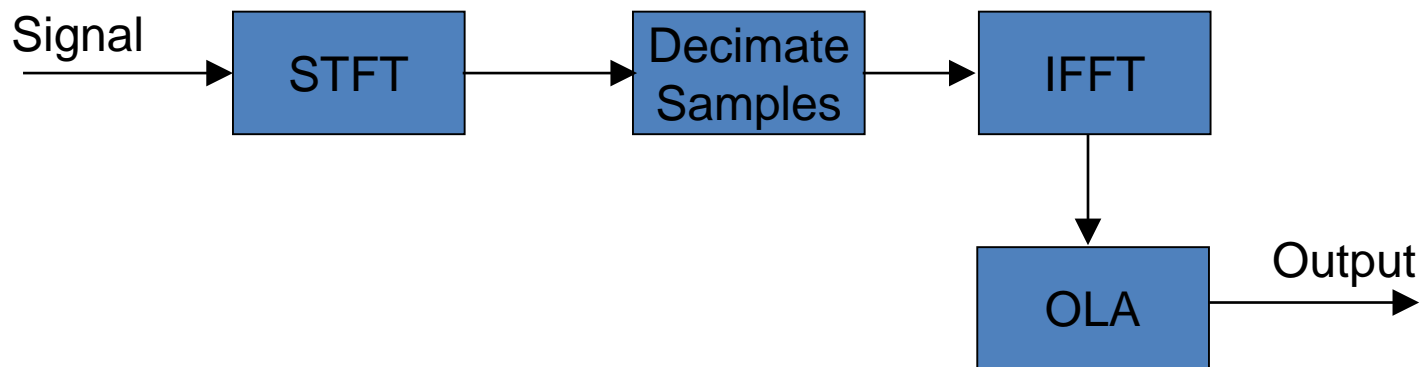


Figure 7.28 Alternative modified STFT for time-scale modification where no frames are discarded. In the example, $M = \frac{L}{2}$.

Short Time Fourier Transform



Noise Reduction

- A number of techniques developed to remove/reduce additive noise:
- Noise corrupted signal is given by:

$$y[n]=x[n]+b[n]$$

– STFT Synthesis:

- Subtract Noise spectrum $\hat{S}_b(\omega)$

$$\hat{X}(nL, \omega) = \left[|Y(nL, \omega)|^2 - \alpha \hat{S}_b(\omega) \right]^{\frac{1}{2}} e^{-j\angle Y(nL, \omega)}$$

$$\text{if } |Y(nL, \omega)|^2 - \alpha \hat{S}_b(\omega) < 0 \Rightarrow |Y(nL, \omega)|^2 - \alpha \hat{S}_b(\omega) = 0$$

- Original phase spectrum $\angle Y(nL, \omega)$ is retained because phase of the noise can not be reliably estimated in general.
- Factor α is a control of the degree of noise reduction.

Noise Reduction

- STFTM Synthesis:
 - Ignore phase and use Sequential Extrapolation or Least-Squared Error estimation method to construct clean signal.

Signal Representation

- Suppose $x[n]$ is the sum of two signals: $x_1[n]$ and $x_2[n]$ occupying different regions of the n -axis.
- Furthermore, suppose that the gap of zeros between $x_1[n]$ and $x_2[n]$ is large enough so that there is no analysis window position for which the corresponding short-time section includes non-zero samples of both $x_1[n]$ and $x_2[n]$.
- Because of the ambiguity condition STFTM of:
 - $x_1[n] + x_2[n]$
 - $x_1[n] - x_2[n]$, and
 - $-x_1[n] + x_2[n]$
 is the same.

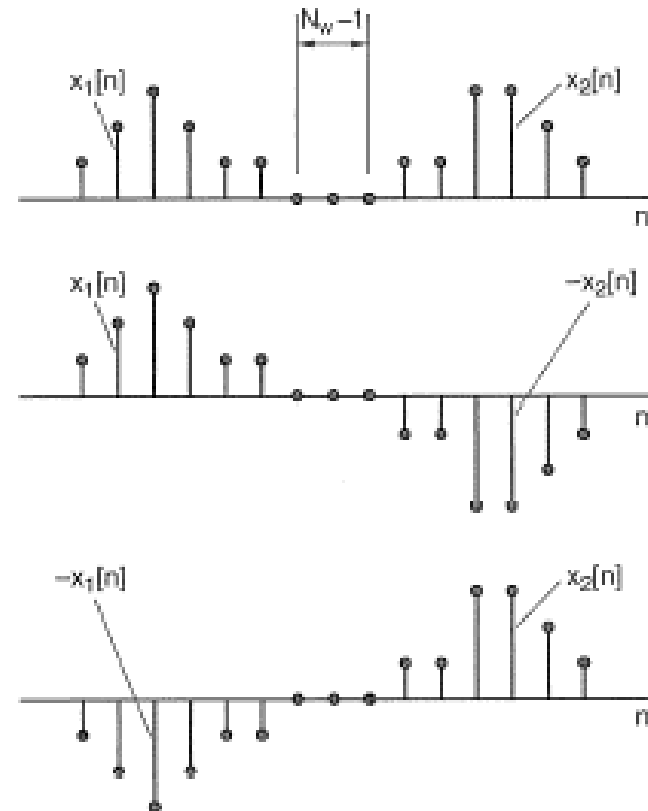


Figure 7.15 Three sequences with the same STFTM.

SOURCE: S.H. Nawab and T.F. Quatieri, "Short-Time Fourier Transform" [13]. ©1987, Pearson Education, Inc. Used by permission.

Signal Representation

- Any uniqueness conditions must include a restriction on the length of zero gaps between non-zero portions of the signal $x[n]$.
- Sufficient uniqueness conditions are the following:
 1. The analysis window $w[n]$ is known sequence of finite length N_w with no zeros over its durations.
 2. The sequence $x[n]$ is one-sided with at most N_w-2 consecutive zero samples, and the sign of its first non-zero value is known.

Signal Representation

- If the successive STFTM correspond to overlapping signal segments then:
 - If short-time spectral magnitude of signal segment at time n is known then
 - Spectral magnitude of the adjacent section at time $n+1$ must be consistent in the region of overlap with the known short-time section.
- ⇒ If the analysis window were non-zero and of length N_w , then after dividing out the analysis window, the first N_w-1 samples of the segment at time $n+1$, must equal the last N_w-1 of the segment at time n (as illustrated in the next slide)
- ⇒ If the last sample of a segment can be extrapolated from its first N_w-1 values, one could repeat this process to obtain the entire signal $x[n]$.

Signal Representation

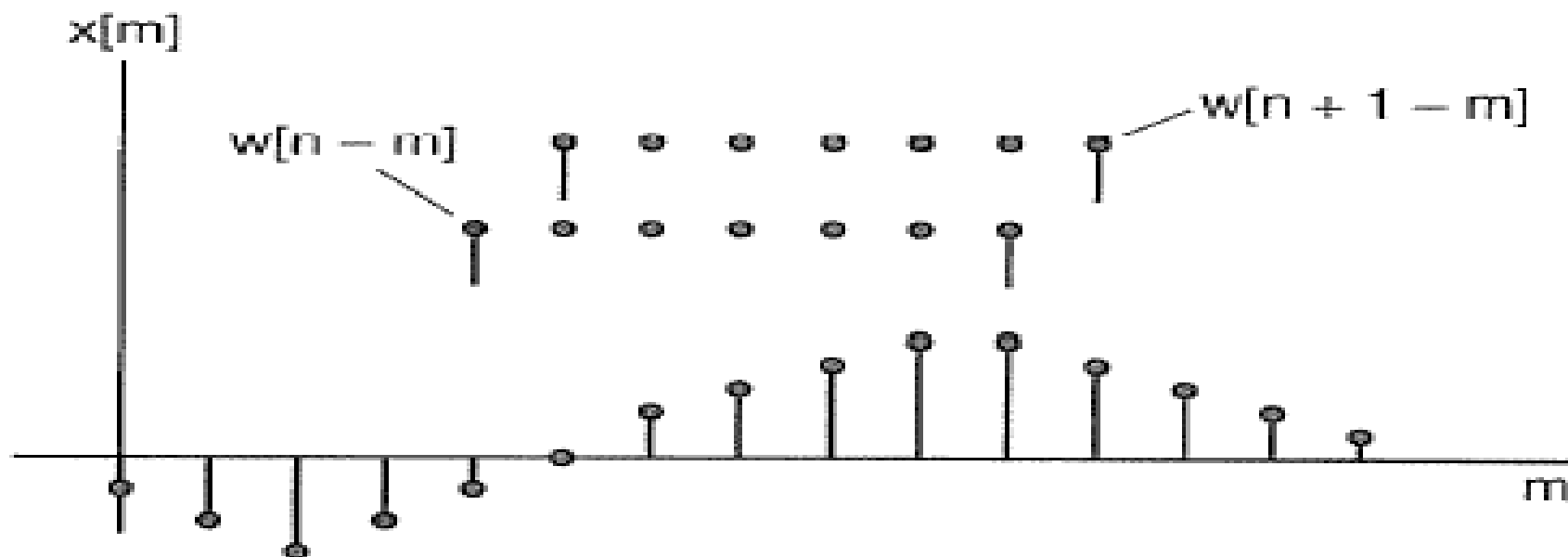


Figure 7.16 Illustration of the consistency required among adjacent short-time sections. Note the samples that are common to the adjacent sections. A rectangular analysis window is assumed.

SOURCE: S.H. Nawab and T.F. Quatieri, "Short-Time Fourier Transform" [13]. ©1987, Pearson Education, Inc. Used by permission.

Signal Representation

- To develop the procedure for extrapolating the next sample of a sequence using its STFTM, assume that the first N_w-1 samples under the analysis window positioned at time n are known.
 - The sequence $x[n]$ has been obtained up to some time $n-1$ from its STFTM.
- Goal is to compute sample $x[n]$ from these initial samples and the STFT magnitude, $|X(n,\omega)|$, or equivalently $r[n,m]$.

Signal Representation

- Note that $r[n, N_w-1]$, the maximum lag of autocorrelation, is given by the product of the first and last value of the segment:

$$r[n, N_w - 1] = \underbrace{(w[0]x[n - 0])}_{\text{first of next}} \underbrace{(w[N_w - 1]x[n - (N_w - 1)])}_{\text{last of present}}$$

$$\Rightarrow x[n] = \frac{r[n, N_w - 1]}{w[0]w[N_w - 1]x[n - (N_w - 1)]}$$

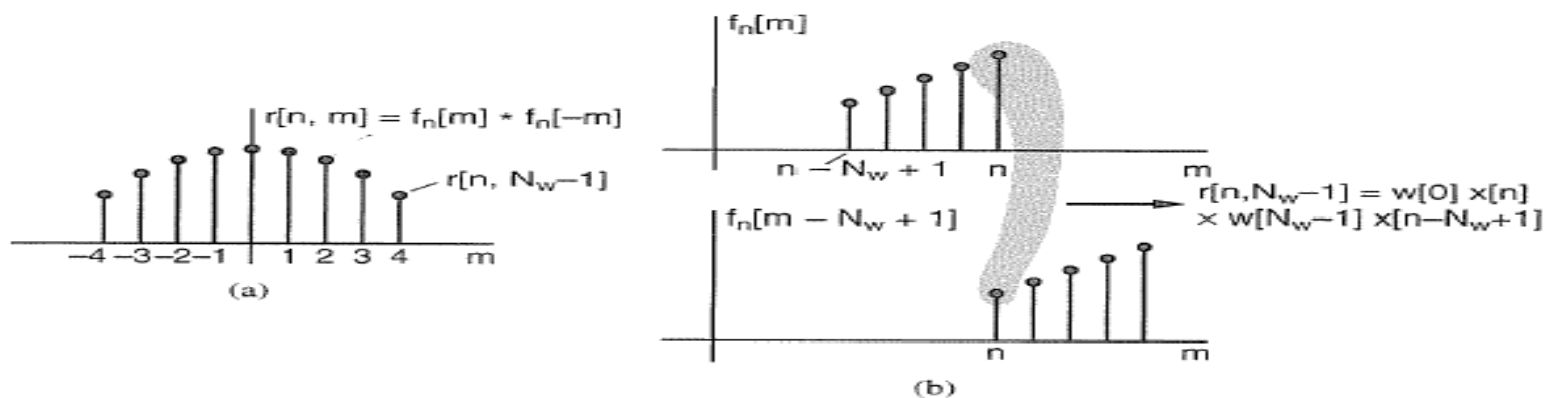


Figure 7.17 Computation of the last non-zero autocorrelation sample of a five-point sequence: (a) autocorrelation function; (b) product of first and last sequence values.

Signal Representation

- Note that:

$$|X(n, \omega)|^2 = \sum_{m=-\infty}^{\infty} r[n, m] e^{-j\omega n}$$

- If the first value of the short-time section, $x[n-(N_w-1)]$ happens to be equal to zero, must find the first non-zero value within the section and again use the product relation as depicted in the last expression.
- Note that such a sample can be found because it was assumed that there are at most N_w-2 consecutive zero samples between any two non-zero samples of $x[n]$.

Signal Representation

- Sequential extrapolation algorithm
 1. Initialize with $x[0]$
 2. Update time n
 3. Compute $r[n, N_w - 1]$ from the inverse DFT of $|X(n, k)|^2$.

4. Compute:

$$x[n] = \frac{r[n, N_w - 1]}{w[0]w[N_w - 1]x[n - (N_w - 1)]}$$

5. Return to step (2) and repeat

Reconstruction from Time-Frequency Samples

- Sufficient uniqueness conditions for the partial overlap case:
 1. The analysis window $w[n]$ is a known sequence of finite length N_w , with no zeros over its duration.
 2. The sequence $x[n]$ is one-sided with, at most $N_w - 2L$ consecutive zero samples.
L consecutive samples of $x[n]$ (from the first non-zero sample) are known.
This is a sufficient but not a necessary condition.

Signal Estimation from the Modified STFT or STFTM

- Synthesis of a signal from a time-frequency function of a modified STFT or STFTM required in many applications.
- Modification may arise due to:
 1. Quantization errors (e.g., from speech coding)
 2. Time-varying filtering
 3. Speech Enhancement
 4. Signal Rate modifications
- Limitations:
 - Modifications in frequency should result in time modification that are restricted within an analysis window (Figure 7.18 next slide)
 - Overlapping sections must undergo similar modifications (Figure 7.19)

Signal Estimation from the Modified STFT or STFTM

- Example 7.5. Removal of interfering tone.
 - Consider modifying a valid $X(n, \omega)$ of short time $f_n[m] = x[m]w[n-m]$ segment by inserting a zero gap where there is known to lie an unwanted interfering sine wave component.
 - Removal of the interfering signal with $H(n, \omega)$.
 - Resulting frequency representation is: $Y(n, \omega) = X(n, \omega)H(n, \omega)$
 - Inverse transforming it to obtain modified short-time sequence $g_n[m]$ is non-zero beyond the extent of the original short-time segment $f_n[m] = x[m]w[n-m]$.

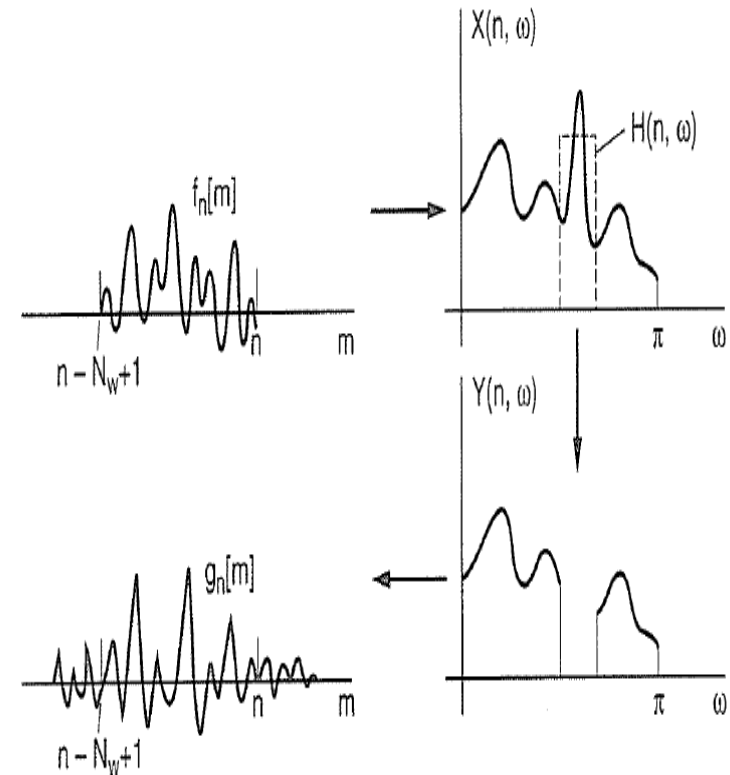


Figure 7.18 Schematic of violation of STFT duration constraint after modification. The original STFT $X(n, \omega)$ is modified by a filter $H(n, \omega)$ that removes an interfering sinewave component.

Signal Estimation from the Modified STFT or STFTM

- Example 7.6
 - At time n :
 - Suppose a time-decimated STFT, $X(nL, \omega)$ is multiplied by a linear phase factor $e^{j\omega n_0}$ to obtain $Y(nL, \omega) = X(nL, \omega)e^{j\omega n_0}$
 - At time $(n+1)$
 - $X((n+1)L, \omega)$ is multiplied by a negative of this linear phase factor $e^{-j\omega n_0}$ to obtain $Y((n+1)L, \omega) = X((n+1)L, \omega)e^{-j\omega n_0}$
 - Overlapping sections of inverse Fourier Transforms denoted by $g_{nL}[m]$ and $g_{(n+1)L}[m]$ are not consistent.

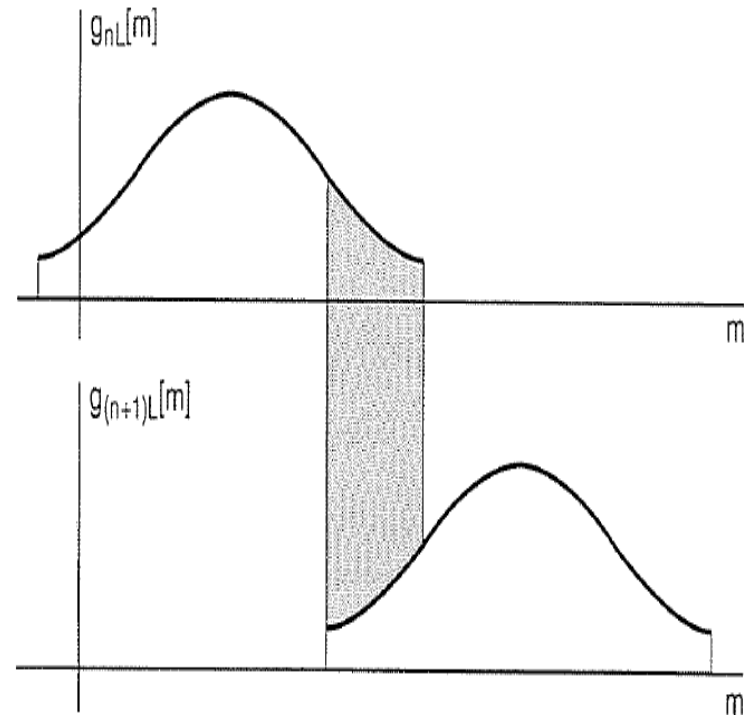


Figure 7.19 Consistency must be satisfied in adjacent short-time segments after modification for a valid STFT. This figure illustrates the violation of the consistency constraint with linear phase modification. After dividing out the window, the resulting sequences are not equal in their region of overlap.

Heuristic Application of STFT Synthesis Methods

- Although modifications of the STFT or STFTM may violate some principles, results may be “reasonable”.
- Resulting effect of modifying STFT (FBS and OLA) with another time-frequency function can be shown to be a time-varying convolution between $x[n]$ and a function $\hat{h}[n,m]$: $x[n] * \hat{h}[n,m]$.
- Let $X(n,\omega)$ be modified by a function $H(n,\omega)$:
$$Y(n,\omega) = X(n,\omega)H(n,\omega)$$
- This corresponds to a new short-time segment:
$$g_n[m] = f_n[n] * h[n,m]$$
- $h[n,m]$ – time varying system impulse response (Chapter 2).

Heuristic Application of STFT Synthesis Methods

- Consider FBS method (discretization in frequency to obtain):

$$Y(n,k) = Y(n,\omega) \Big|_{\omega = \frac{2\pi}{N}k} = X(n,k)H(n,k)$$

- N-point IDFT of $H(n,k)$:

$$\tilde{h}[n,m] = \sum_{l=-\infty}^{\infty} h[n,m-lN], \quad \text{periodic over } N$$

- Then resulting sequence can be written as:

$$y[n] = \sum_{m=-\infty}^{\infty} x[n-m] \hat{h}[n,m]$$

where

$$\hat{h}[n,m] = w[n] \sum_{l=-\infty}^{\infty} h[n,m-lN]$$

Heuristic Application of STFT Synthesis Methods

- Using OLA method, it can be shown (see Exercise 7.11) that:

$$\hat{h}[n,m] = w[n] * \sum_{l=-\infty}^{\infty} h[n,m-lN]$$

- Contrasting FBS with OLA
 - FBS: multiplication \Rightarrow instantaneous change
 - OLA: convolution \Rightarrow smoothing

Heuristic Application of STFT Synthesis Methods

- Example 7.7
 - Suppose we want to deliberately introduce reverberation into a signal $x[n]$ by convolution with the filter:

$$h[n] = \delta[n] + \alpha \delta[n-n_o]$$

- Fourier transform of which is:

$$H(\omega) = 1 + \alpha e^{-j\omega n_o}$$

- STFT of resulting signal is given by:

$$Y(n, \omega) = X(n, \omega) H(\omega)$$

where

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

Example 7.7 (cont.)

- Using OLA method (7.21):

$$y[n] = \frac{1}{W(0)} \sum_{p=-\infty}^{\infty} \left[\frac{1}{n} \sum_{k=0}^{N-1} Y(p, k) e^{j \frac{2\pi}{N} kn} \right]$$

- It is then possible to express $y[n]$ in terms of original sequence:

$$y[n] = \frac{1}{W(0)} \sum_{p=-\infty}^{\infty} x[m] \underbrace{\left[\frac{1}{N} \sum_{k=0}^{N-1} H(k) e^{j \frac{2\pi}{N} k(n-m)} \right]}_{\text{IDFT} \rightarrow \sum_{r=-\infty}^{\infty} h[n-m+rN]} \underbrace{\left[\sum_{p=-\infty}^{\infty} w[p-m] \right]}_{W(0)}$$

$$= \sum_{p=-\infty}^{\infty} x[m] \hat{h}[n-m]$$

Example 7.7 (cont.)

- Where

$$\hat{h}[n] = \sum_{r=-\infty}^{\infty} h[n+rN] = \sum_{r=-\infty}^{\infty} (\delta[n+rN] + \alpha \delta[n-n_o+rN])$$

is periodic extension of $h[n]$, over N , of which we only consider interval $[0, N-1]$.

- This implies that original reverberated signal is obtained only when $n_o < N$, otherwise temporal alias will occur (as illustrated in 7.20).

Example 7.7 (cont.)

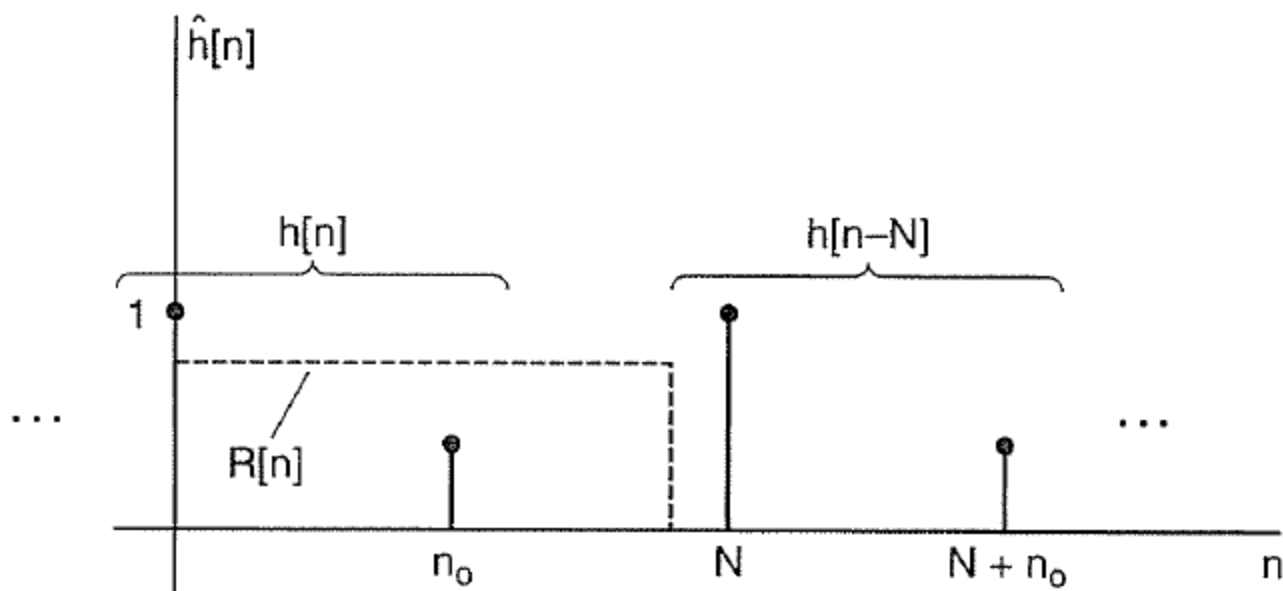


Figure 7.20 Illustration of the echo-generating function $\hat{h}[n]$ that results from the OLA method in Example 7.7. $R[n]$ is the rectangular function, non-zero in the interval $[0, N - 1]$, invoked by the inverse DFT.