SHORT TIME FOURIER TRANSFORM (STFT)

Sec 11.3 in Mitra

The Fourier transforms (FT, DTFT, DFT, etc) do not clearly indicate how the frequency content of a signal changes over time.

That information is hidden in the phase — it is not revealed by the plot of the magnitude of the spectrum.

> To see how the frequency content of a signal changes over time, we can cut the signal into blocks and compute the spectrum of each block.

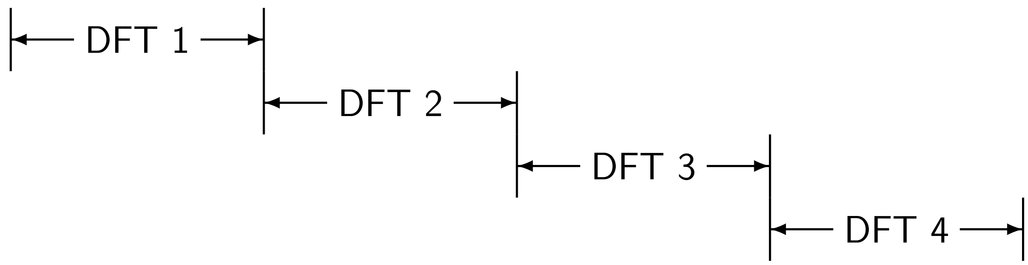
To improve the result,

1. blocks are overlapping,
2. each block is multiplied by a window that is tapered at its endpoints.

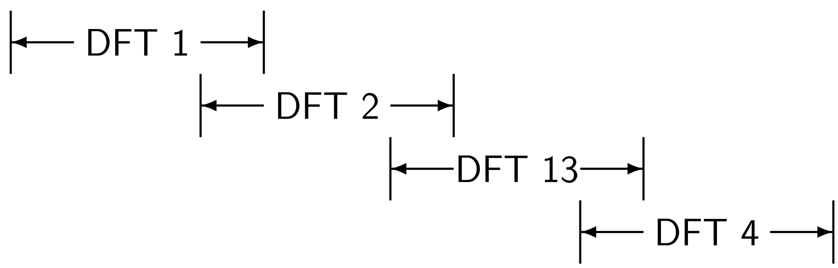
Several parameters must be chosen:

1. Block length, R.
2. The type of window.
3. Amount of overlap between blocks.
4. Amount of zero padding, if any.

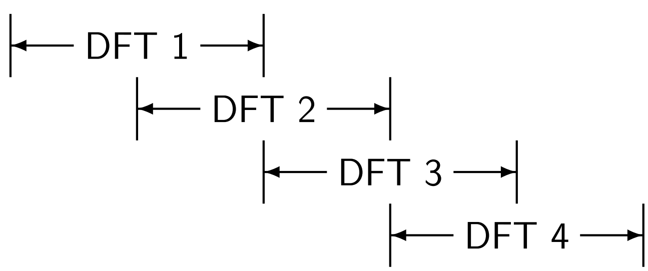
NO OVERLAP



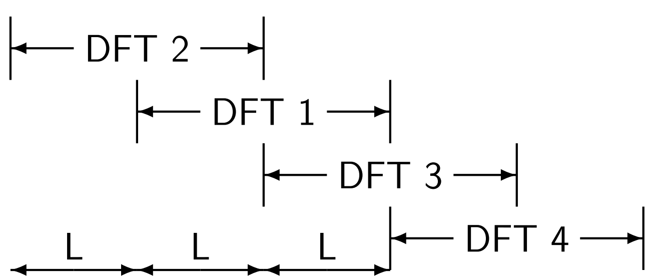
R/4 OVERLAP



R/2 OVERLAP



The parameter L



L is the number of samples between adjacent blocks.

SHORT TIME FOURIER TRANSFORM (STFT)

The short-time Fourier transform is defined as

$$\begin{aligned} X(\omega, m) &= \text{STFT}\{x(n)\} \\ &:= \text{DTFT}\{x(n+m)\,w(n)\} \\ &= \sum_{n=-\infty}^{\infty} x(n+m)\,w(n)\,e^{-j\omega n} \\ &= \sum_{n=0}^{R-1} x(n+m)\,w(n)\,e^{-j\omega n} \end{aligned}$$

where $w(n)$ is the window function of length R.

1. The STFT of a signal $x(n)$ is a function of two variables: time and frequency.

2. The block length is determined by the support of the window function $w(n)$.

3. A graphical display of the magnitude of the STFT, $|X(\omega, m)|$, is called the *spectrogram* of the signal. It is often used in speech processing.

4. The STFT of a signal is invertible.

5. One can choose the block length. A long block length will provide higher frequency resolution (because the main-lobe of the window function will be narrow). A short block length will provide higher time resolution because less averaging across samples is performed for each STFT value.

6. A *narrow-band* spectrogram is one computed using a relatively long block length R, (long window function).

7. A *wide-band* spectrogram is one computed using a relatively short block length R, (short window function).

To numerically evaluate the STFT, we sample the frequency axis ω in N equally spaced samples from $\omega = 0$ to $\omega = 2\pi$.

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

We then have the discrete STFT,

$$\begin{aligned}
X^d(k,m) &:= X(\frac{2\pi}{N}k, m) \\
&= \sum_{n=0}^{R-1} x(n+m)\, w(n)\, e^{-j\frac{2\pi}{N}kn} \\
&= \sum_{n=0}^{R-1} x(n+m)\, w(n)\, W_N^{-kn} \\
&= \mathsf{DFT}_N\{\{x(n+m)\, w(n)\}_{n=0}^{R-1}, \underbrace{0, \dots, 0}_{N-R}\}
\end{aligned}$$

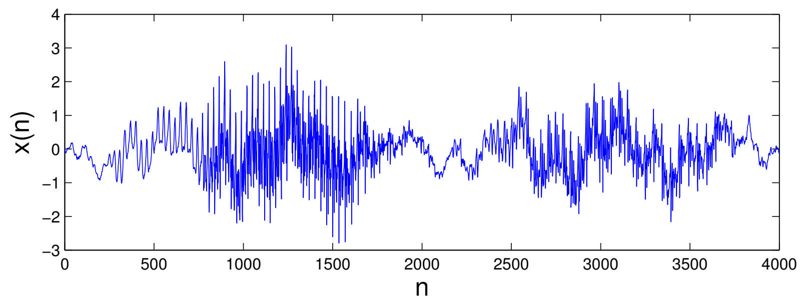
In this definition, the overlap between adjacent blocks is $R-1$. The signal is shifted along the window one sample at a time. That generates more points than is usually needed, so we also sample the STFT along the time direction. That means we usually evaluate

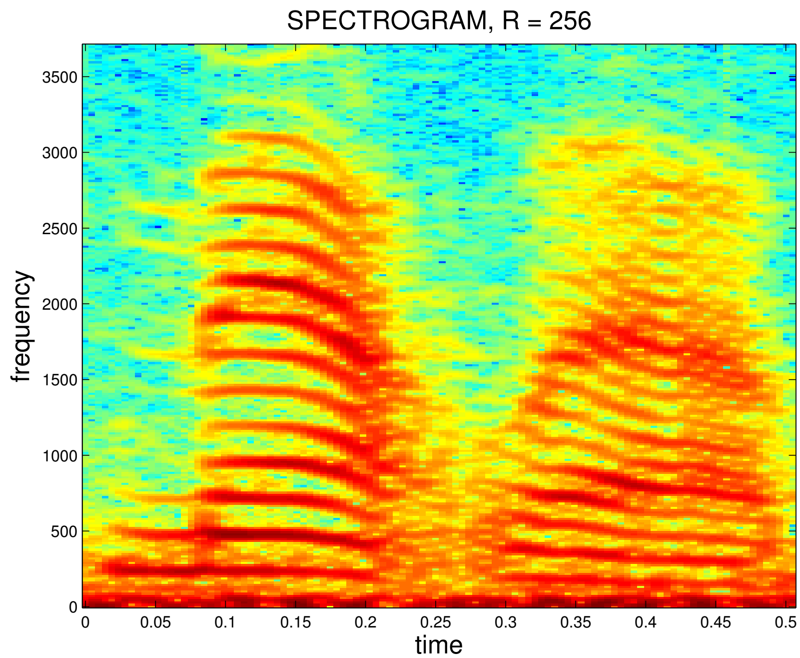
$$X^d(k, Lm)$$

where L is the time-skip. The relation between the time-skip, the number of overlapping samples, and the block length is

$$\text{overlap} = R - L.$$

SPECTROGRAM EXAMPLE





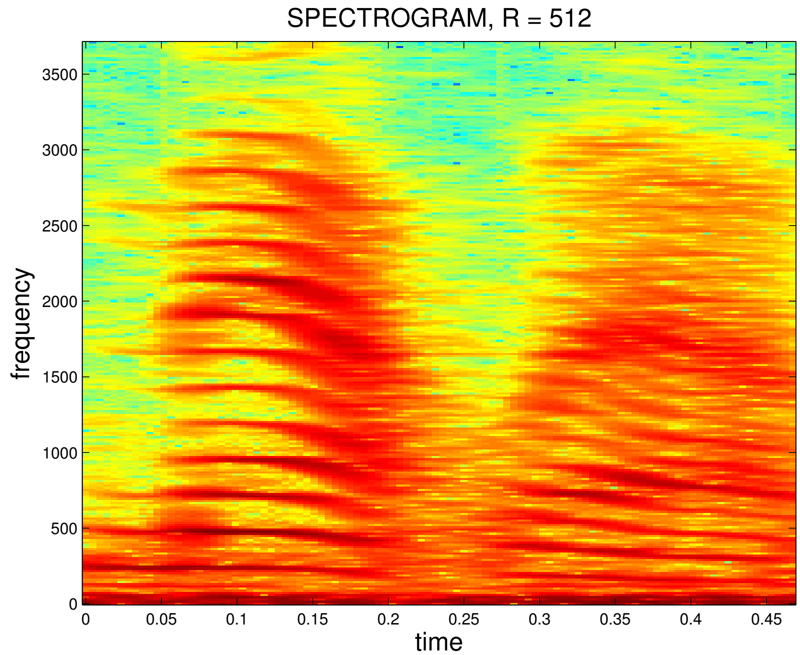
SPECTROGRAM EXAMPLE

The Matlab program for producing the figures on the previous page.

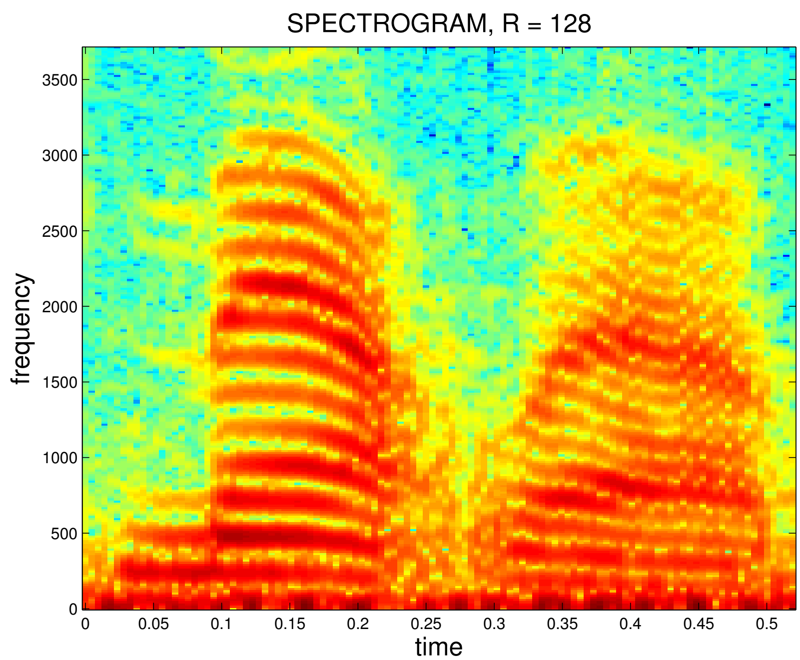
```
% ------ SPECTROGRAM EXAMPLE ------

% LOAD DATA
load mtlb;
x = mtlb;

figure(1), clf
plot(0:4000,x)
xlabel('n')
ylabel('x(n)')

% SET PARAMETERS
R = 256;                % R: block length
window = hamming(R);    % window function of length R
N = 512;                % N: frequency discretization
L = 35;                 % L: time lapse between blocks
fs = 7418;              % fs: sampling frequency
overlap = R - L;

% COMPUTE SPECTROGRAM
[B,f,t] = specgram(x,N,fs,window,overlap);

% MAKE PLOT
figure(2), clf
imagesc(t,f,log10(abs(B)));
colormap('jet')
axis xy
xlabel('time')
ylabel('frequency')
title('SPECTROGRAM, R = 256')
```

EFFECT OF WINDOW LENGTH R

Narrow-band spectrogram: better frequency resolution

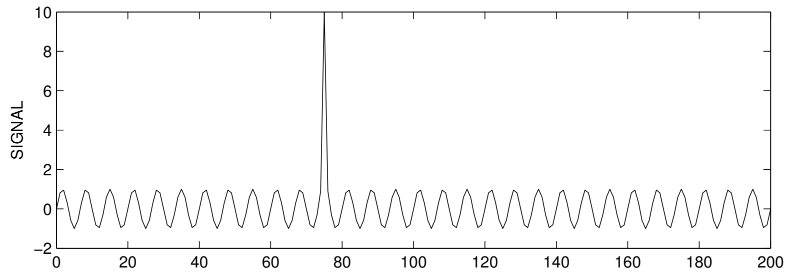


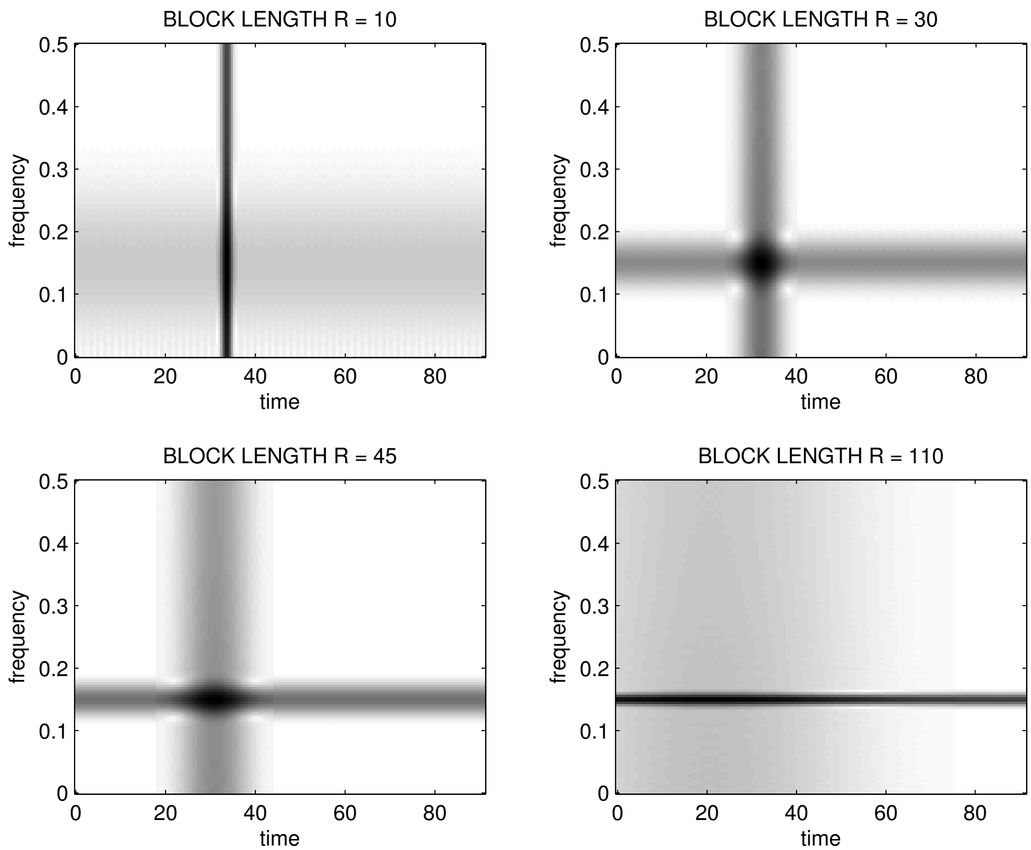
Wide-band spectrogram:: better time resolution



EFFECT OF WINDOW LENGTH R

Here is another example to illustrate the frequency/time resolution trade-off.



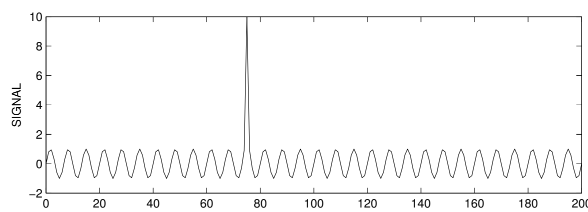


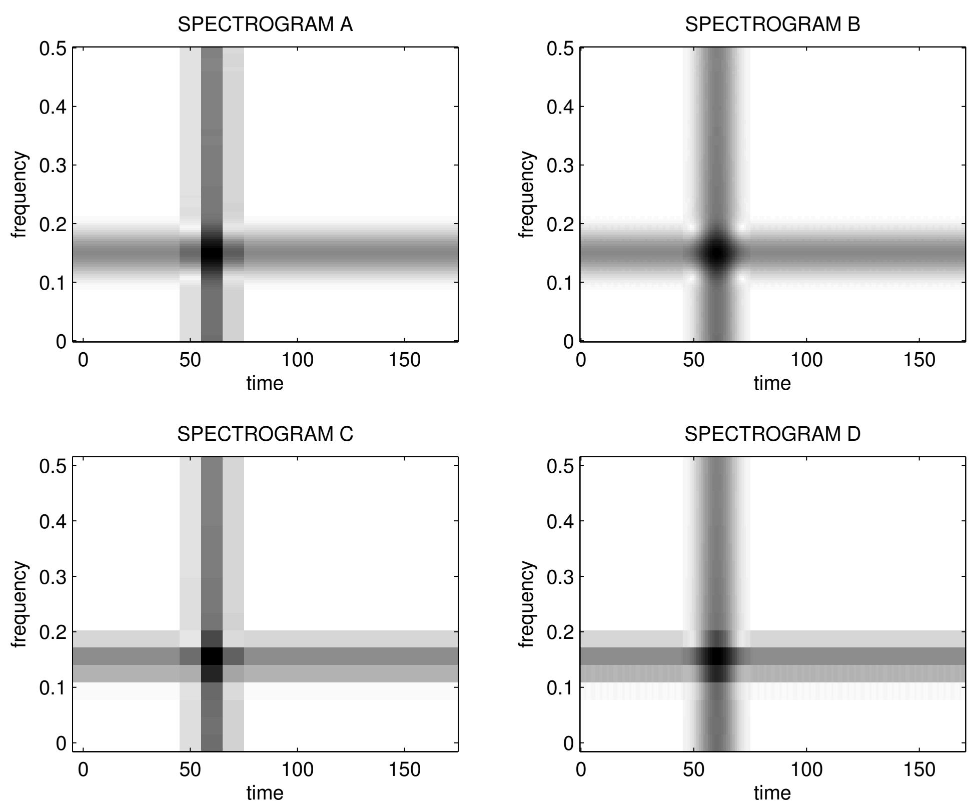
A spectrogram is computed with different parameters:

$$L \in \{1, 10\}, \qquad N \in \{32, 256\}$$

L = time lapse between blocks.
N = FFT length (Each block is zero-padded to length N.)
In each case, the block length is 30 samples.

For each of the four spectrograms, can you tell what L and N are?





L and N do not effect the time resolution or the frequency resolution. They only influence the 'pixelation'.