مقدمه‌ای بر

پردازش صوت

محیط‌های چندرسانه‌ای

۱۴۰۱-۱۱-۰۸۰۵

(بخش دهم)

دانشگاه شهید بهشتی

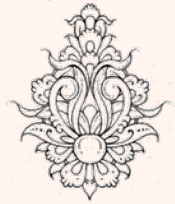
بهار ۱۳۹۴

احمد محمودی ازناوه

http://faculties.sbu.ac.ir/~a_mahmoudi/

فهرست مطالب

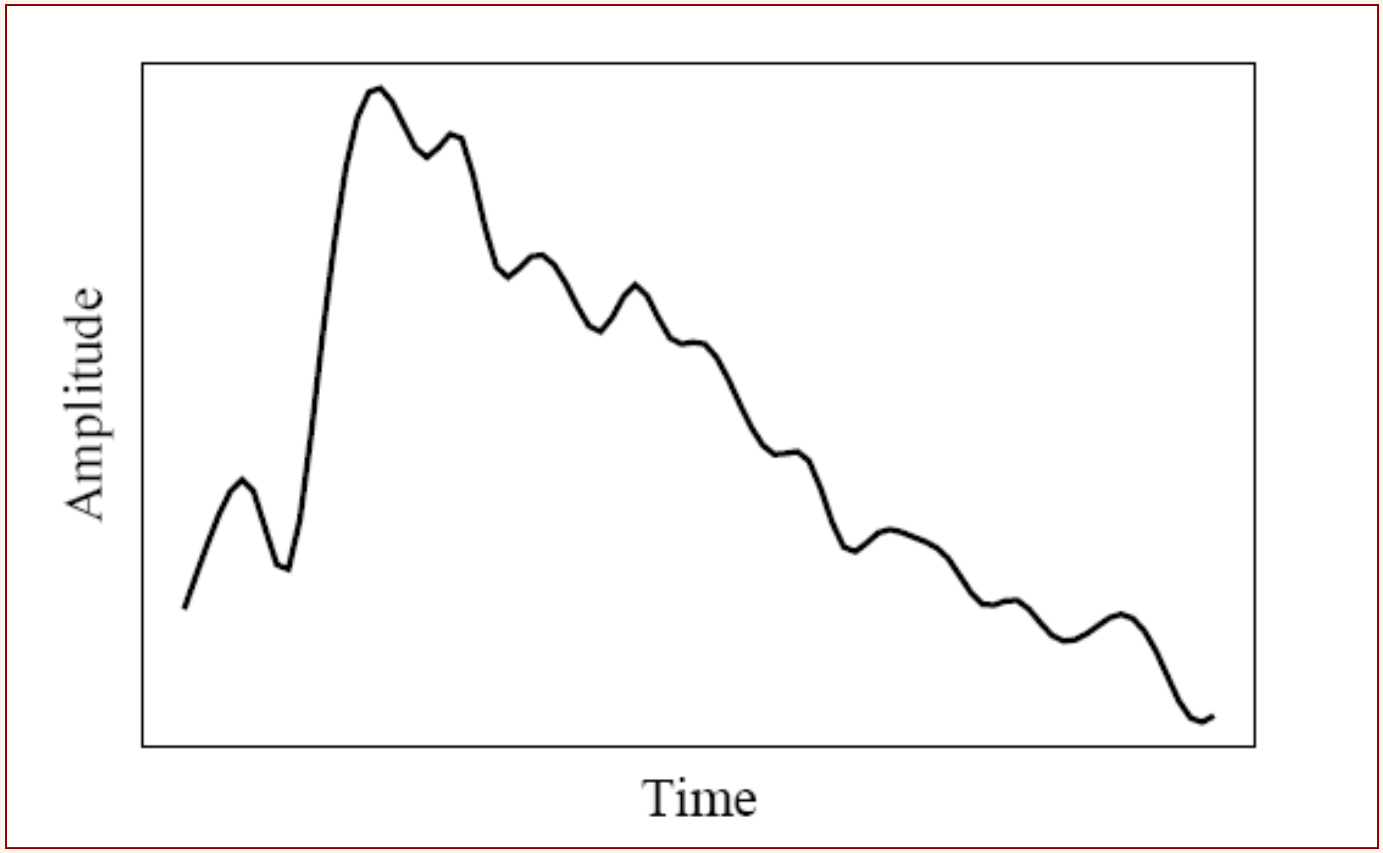
- نمونه‌برداری
- خطای چندی کردن
- تحلیل سیگنال صدا در فضای تبدیل
  - طیف‌نگاره (spectrogram)
  - تبدیل فوریه زمان کوتاه
- روش‌های چندی کردن
- معیار کیفیت و شدت صوت
- فشرده‌سازی در دامنه‌ی زمان



تعریف

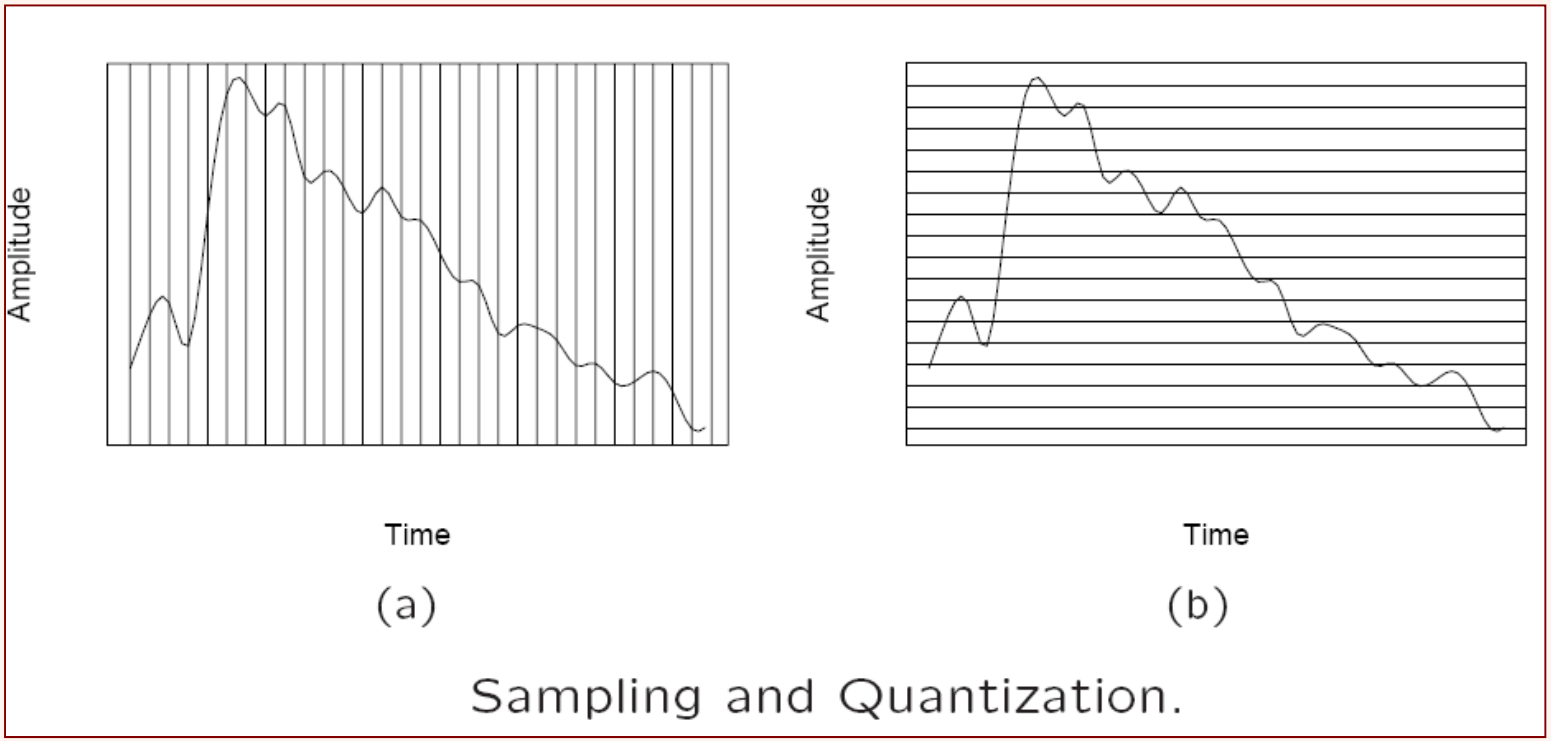
- سیگنال صوت سیگنالی یک بعدی است؛ این سیگنال دنباله‌ای از تغییرات فشار است.
- تفاوت‌هایی مابین صوت و تصویر یا حتی ویدئو وجود دارد.
  - به راحتی می‌توان فریمی از جریان داده‌ی ویدئویی را حذف کرد، با چنین عملی در مورد صوت داده‌هایی مهم از دست می‌رود.
  - حساسیت سیستم شنوایی انسان به گونه‌ای است که می‌باید این خصوصیات در پردازش سیگنال صوت لحاظ گردد.

سیگنال زمان‌پیوسته



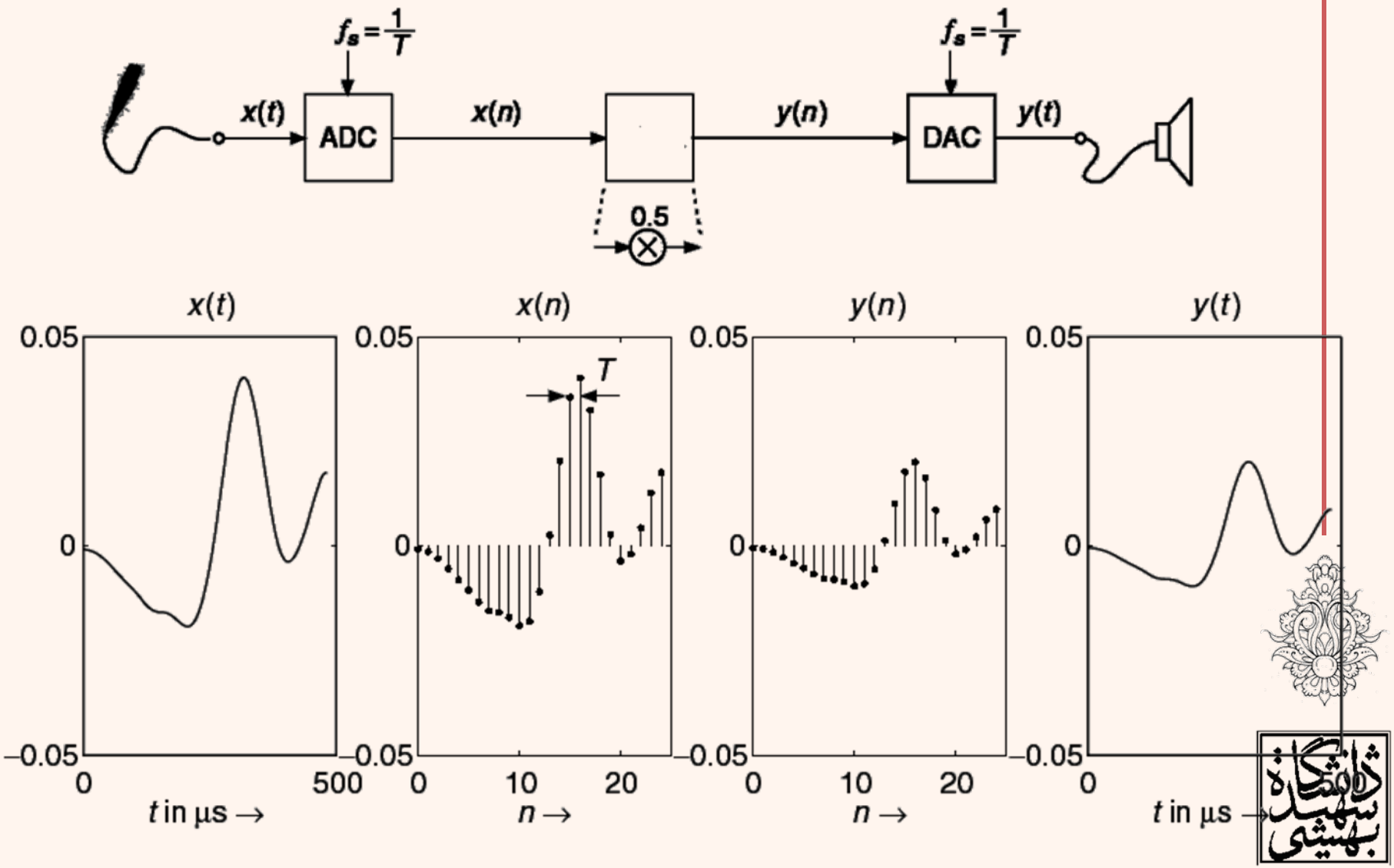
سیگنال گسسته

- شیوه‌ی اولیه برای تولید یک سیگنال گسسته از پیوسته، نمونه‌برداری و چندی‌کردن نمونه‌هاست.



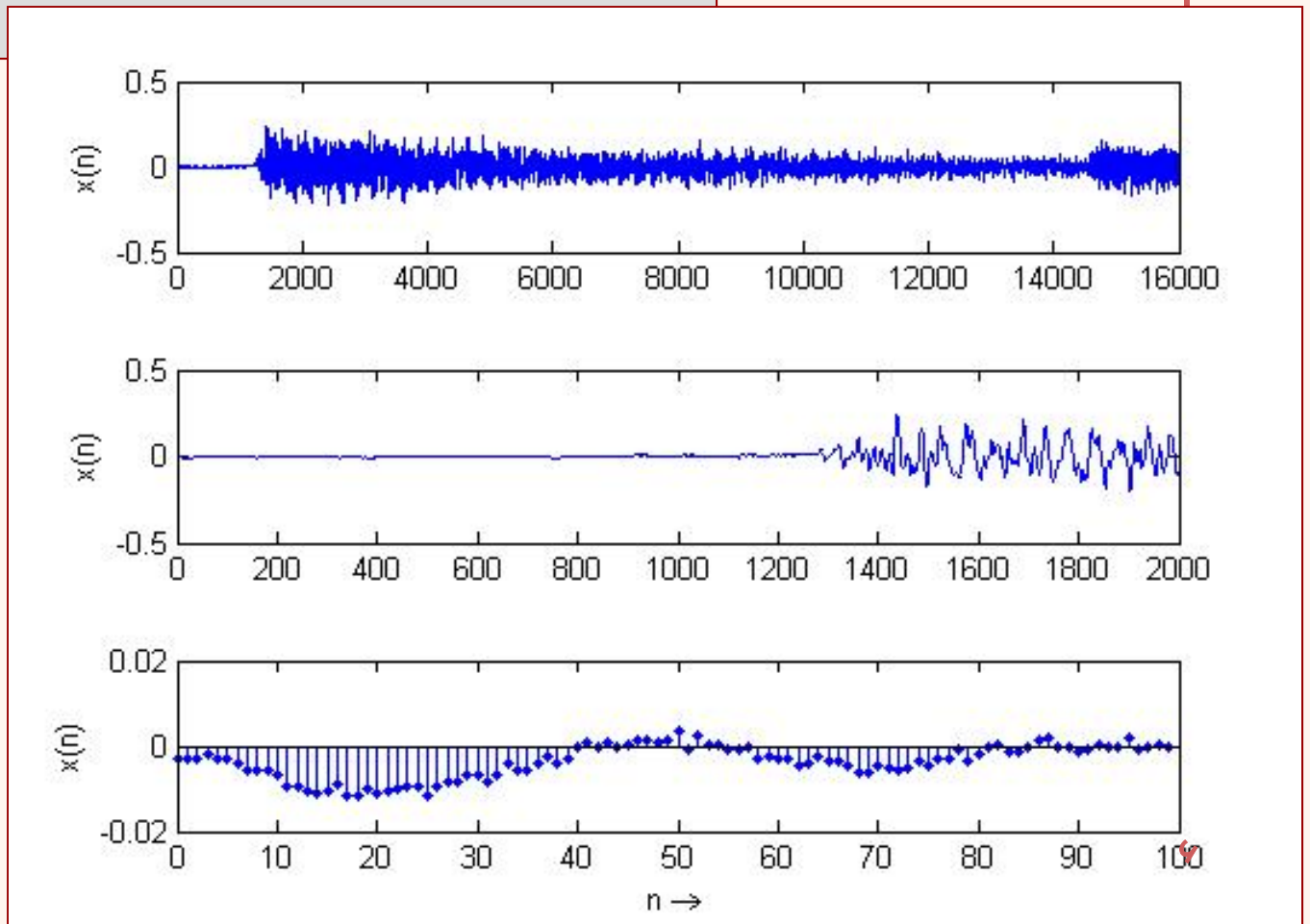
Sampling and Quantization.

مثال


$f_s = \frac{1}{T}$
x(t)
ADC
x(n)
y(n)
$f_s = \frac{1}{T}$
DAC
y(t)
0.5
x(t)
x(n)
y(n)
y(t)
0.05
0
−0.05
0
500
t in μs →
T
10
20
n →
t in μs →

مثال

```
[x,FS,NBITS]=wavread('Art.wav');
figure(1)
subplot(3,1,1);
plot(0:15999,x(1:16000));ylabel('x(n)');
subplot(3,1,2);
plot(0:1999,x(1:2000));ylabel('x(n)');
subplot(3,1,3);
stem(0:99,x(1:100),'.');ylabel('x(n)');
xlabel('n \rightarrow');
wavwrite(x,FS,NBITS,'out.wav');
sound(x,FS);
```



FS

exp1.wav

2*FS

exp1_2.wav

0.5*FS

exp1_3.wav

محیط‌های چندرسانه‌ای

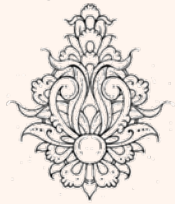
مثال

```
% Read input sound file into vector x(n) and
sampling frequency FS
[x,FS,NBITS]=wavread('Art.wav');
% Sample-by sample algorithm y(n)=a*x(n)
a=5;
y(1:length(x))=a.*x(1:length(x));
% Write y(n) into output sound file with number of
% bits Nbits and sampling frequency FS
wavwrite(y,FS,NBITS,'A_Art.wav');
```

Original(Art.wav)

Modified(A_Art.wav)

- قبل از انجام هرگونه تبدیل سیگنال پیوسته می‌باید به سیگنال گسسته تبدیل گردد.
- فرکانس نمونه‌برداری باید به گونه‌ای باشد که سیگنال اصلی به درستی ساخته شود.



Nyquist theorem

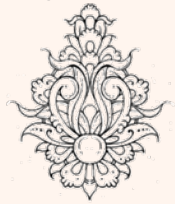
- بر اساس این نظریه فرکانس نمونه‌برداری از یک سیگنال به گونه‌ای که **دوباره‌سازی** سیگنال به درستی صورت گیرد، می‌باید به صورت زیر باشد:

$$f_s > 2f_{\max}$$

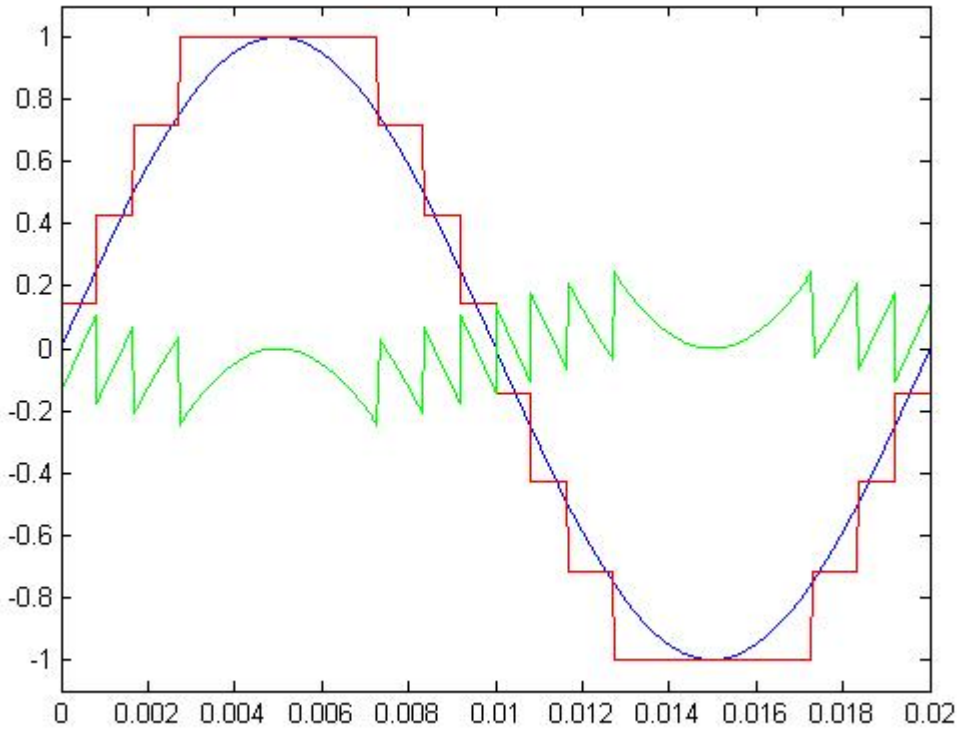
f_{max} is the highest frequency component

- چنان‌چه فرکانس نمونه‌ها مابین f_1 و f_2 محدود باشد رابطه‌ی زیر می‌باید برقرار باشد:

$$fs > 2\,(f_2 - f_1)$$



خطای چندی کردن



خطای چندی‌سازی : اختلاف بین سیگنال آنالوگ و نمایش دیجیتال آن است.

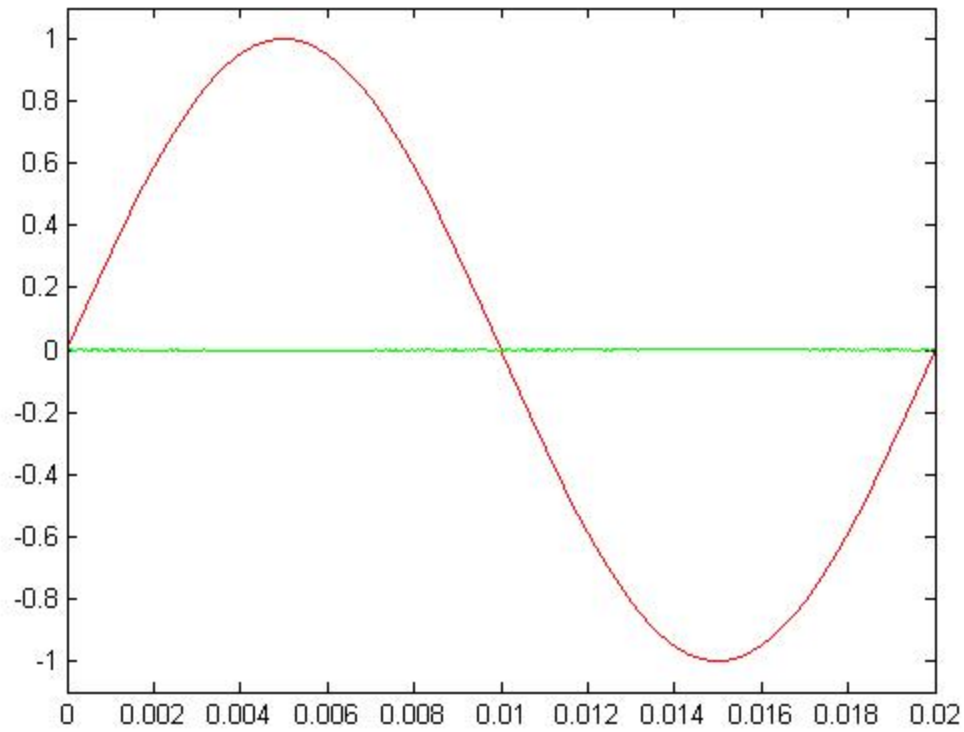
Quantization Noise

$$\text{N Bits} \Rightarrow 2^N \text{ levels}$$

http://web.mit.edu/hst.723/www/

http://web.mit.edu/hst.723/www/

خطای چندی کردن(ادامه...)



Bits	Levels	
3	8	3bits.wav
4	16	4bits.wav
5	32	5bits.wav
8	256	8bits.wav
16	65536	16bits.wav

http://web.mit.edu/hst.723/www/

با افزایش تعداد سطوح چندی کردن(تعداد بیت‌ها) نمایش دیجیتال سیگنال آنالوگ با خطای کم‌تری همراه خواهد بود.

خطای چندی کردن(ادامه...)

Time domain

frequency domain

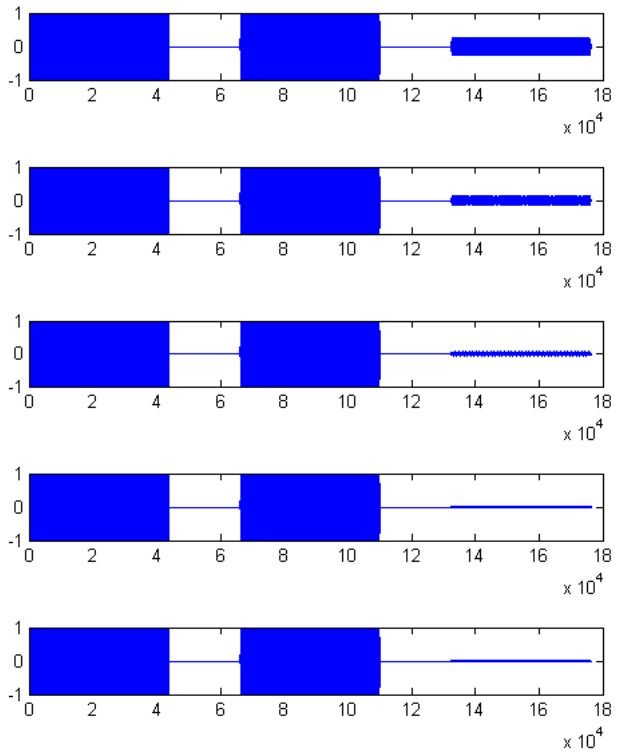
3bits.wav

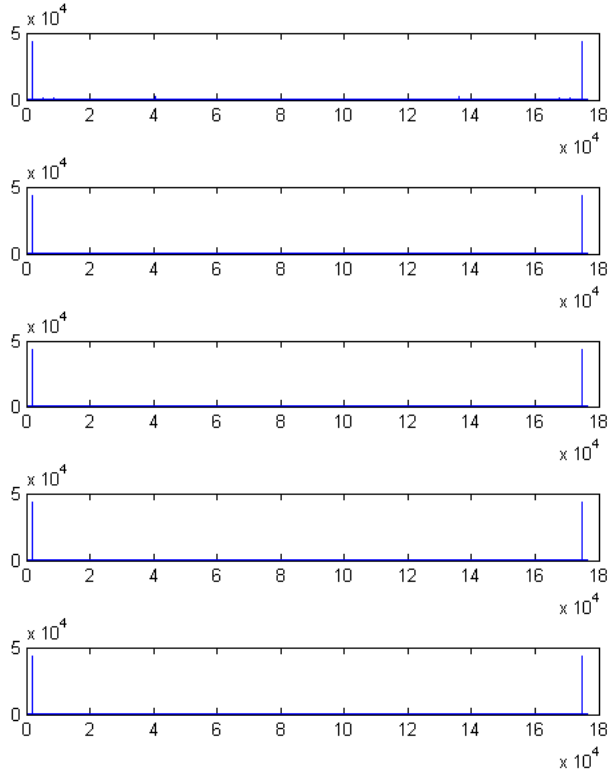
4bits.wav

5bits.wav

8bits.wav

16bits.wav

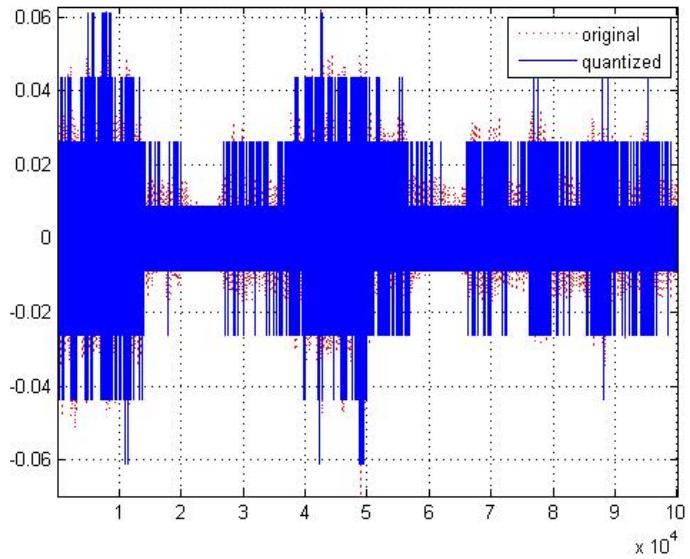


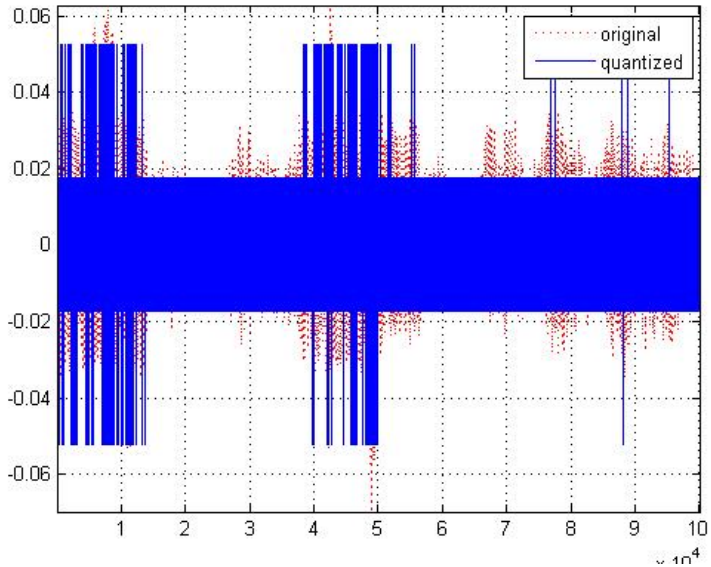


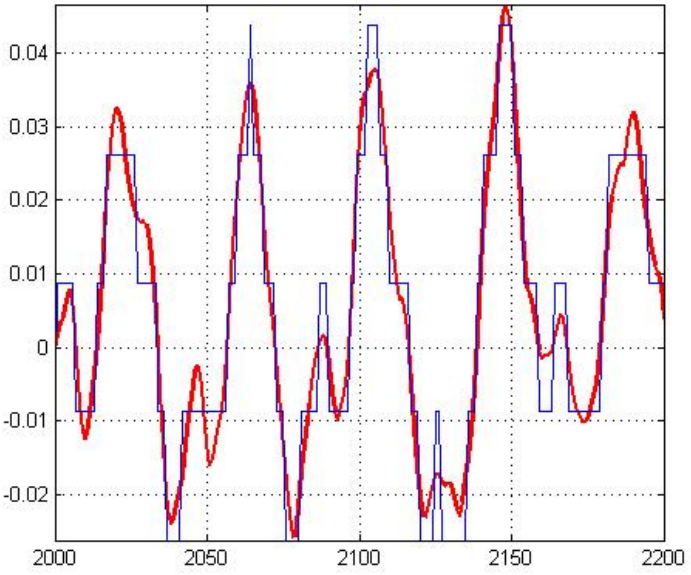
خطای چندی کردن(ادامه...)

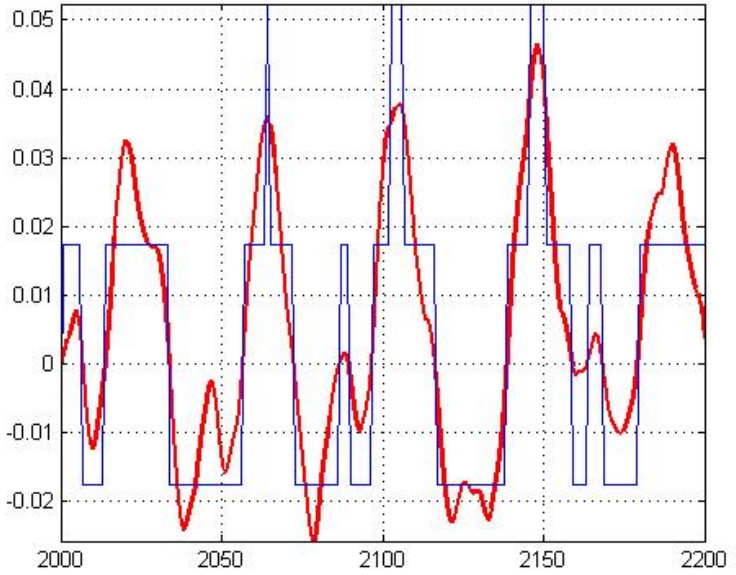
- هر چقدر سطوح چندی‌شدن بیشتر باشد میزان تفاوت با سیگنال اصلی کم‌تر خواهد بود.
- در این صورت از میزان پله‌ای شدن سیگنال چندی شده هم کم می‌شود.
- هر چه سیگنال پله‌ای‌تر باشد، دارای فرکانس‌های بالاتری خواهد بود.
- مثال

quantized to 256 linear levels		quant256.wav
quantized to 16 linear levels		quant16.wav
quantized to 4 linear levels		quant4.wav









mozartQ8.wav

Q8

Q4

mozartQ4.wav

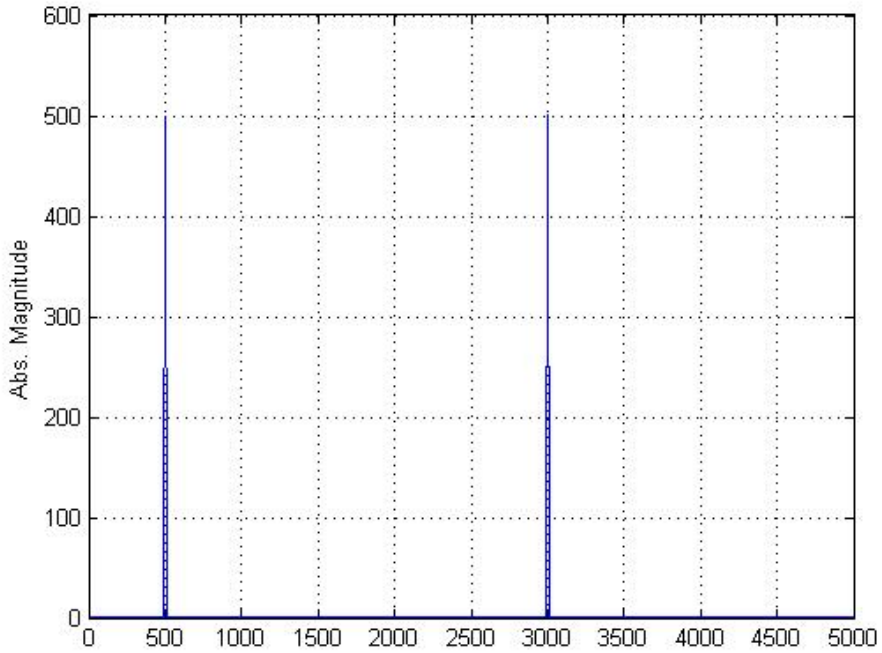
اصلی

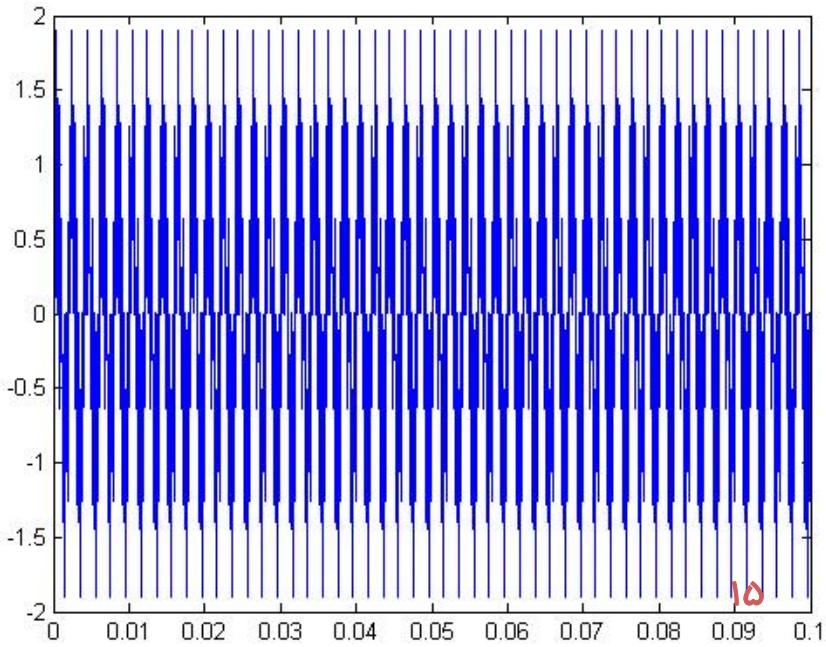
mozart.wav

```
% generate a complex signal
fs = 10000;
t = 0:1/fs:0.1;
x = sin(2*pi*500*t) + sin(2*pi*3000*
wavwrite(x,fs,16,'outsin1.wav');
plot(t,x)
figure;

% perform 1000-point transform
y = fft(x,1000);
amplitudes
y = y(1:500);
m = abs(y);

% plot spectrum 0..fs/2
f = (0:499)*fs/1000;
plot(f,m);                    % plot magn
ylabel('Abs. Magnitude'), grid on;
```

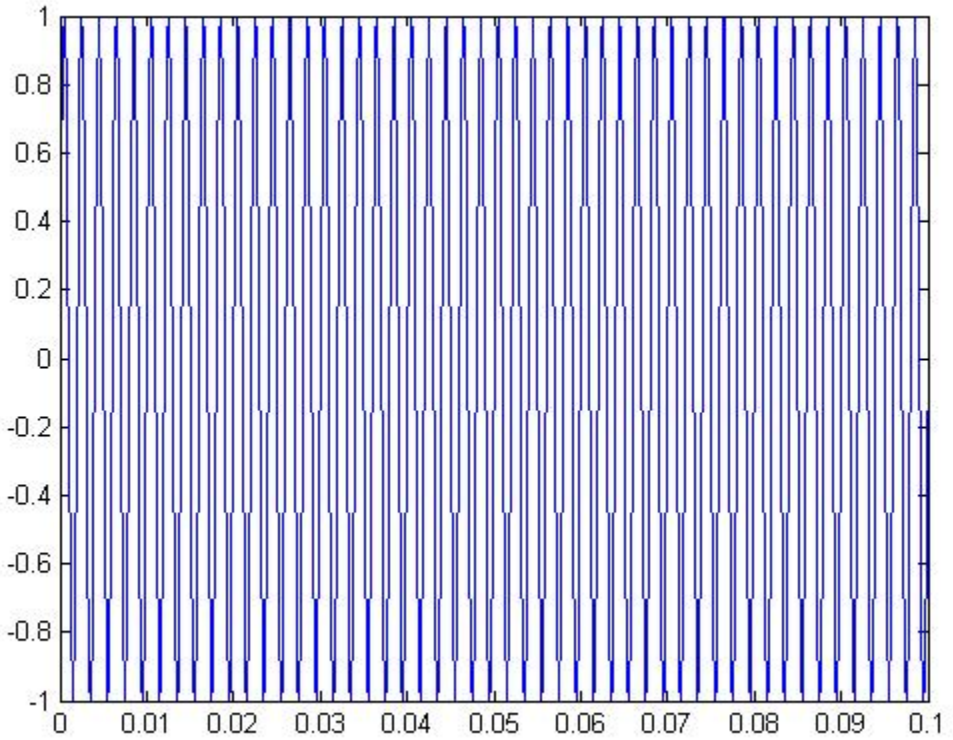


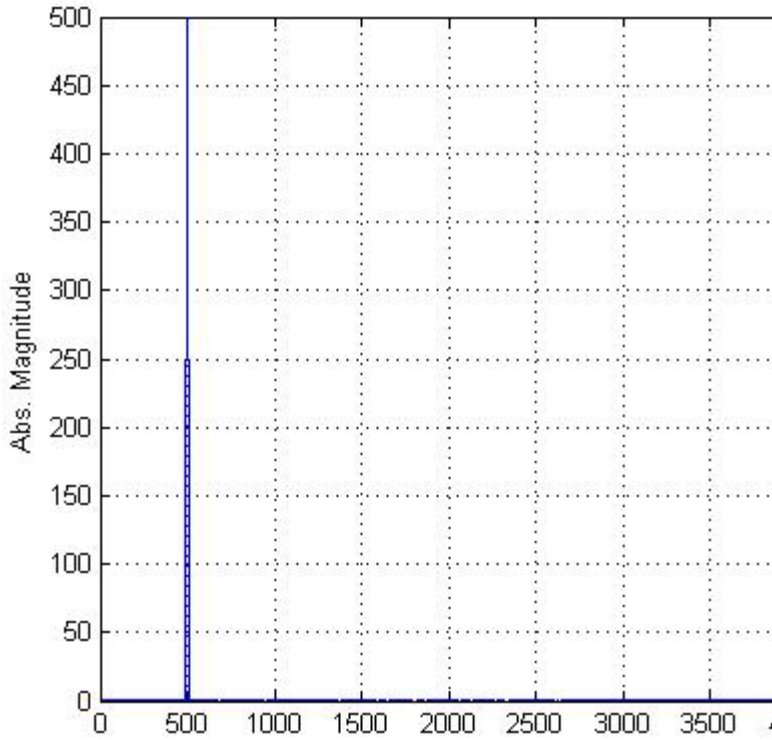


outsin1.wav

مثال

```
x = sin(2*pi*500*t);
```



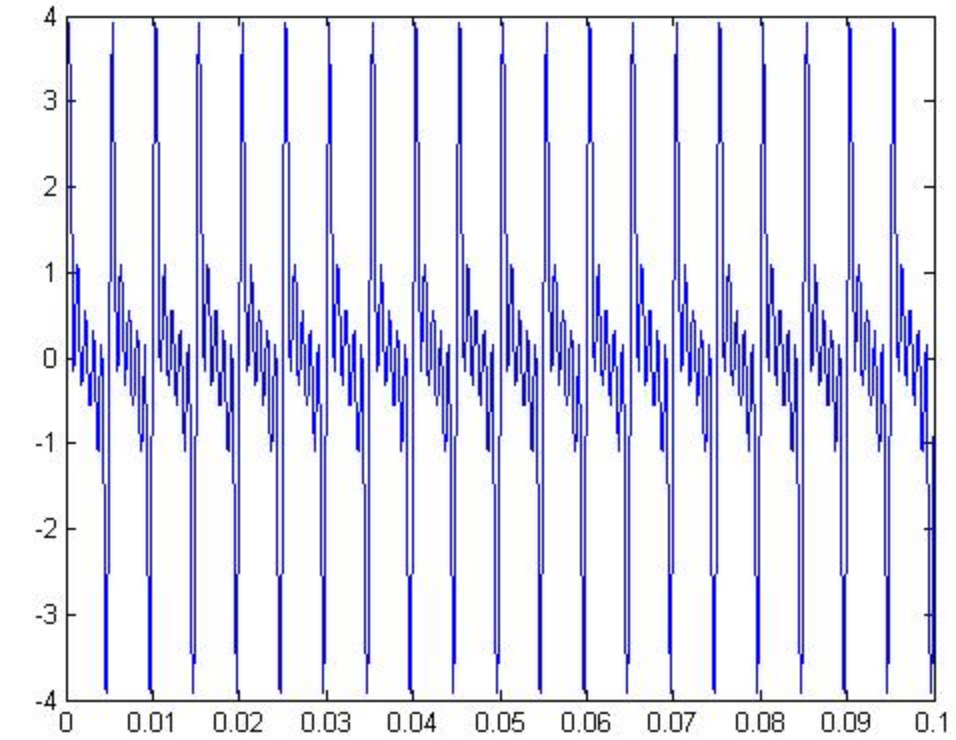


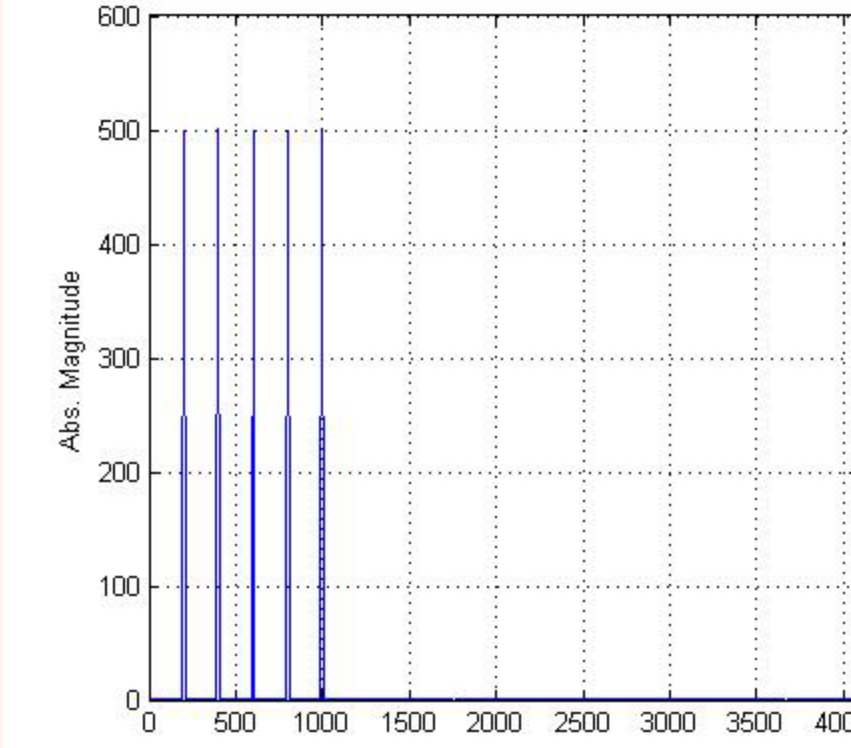
بهنیتی

outsin2.wav

مثال

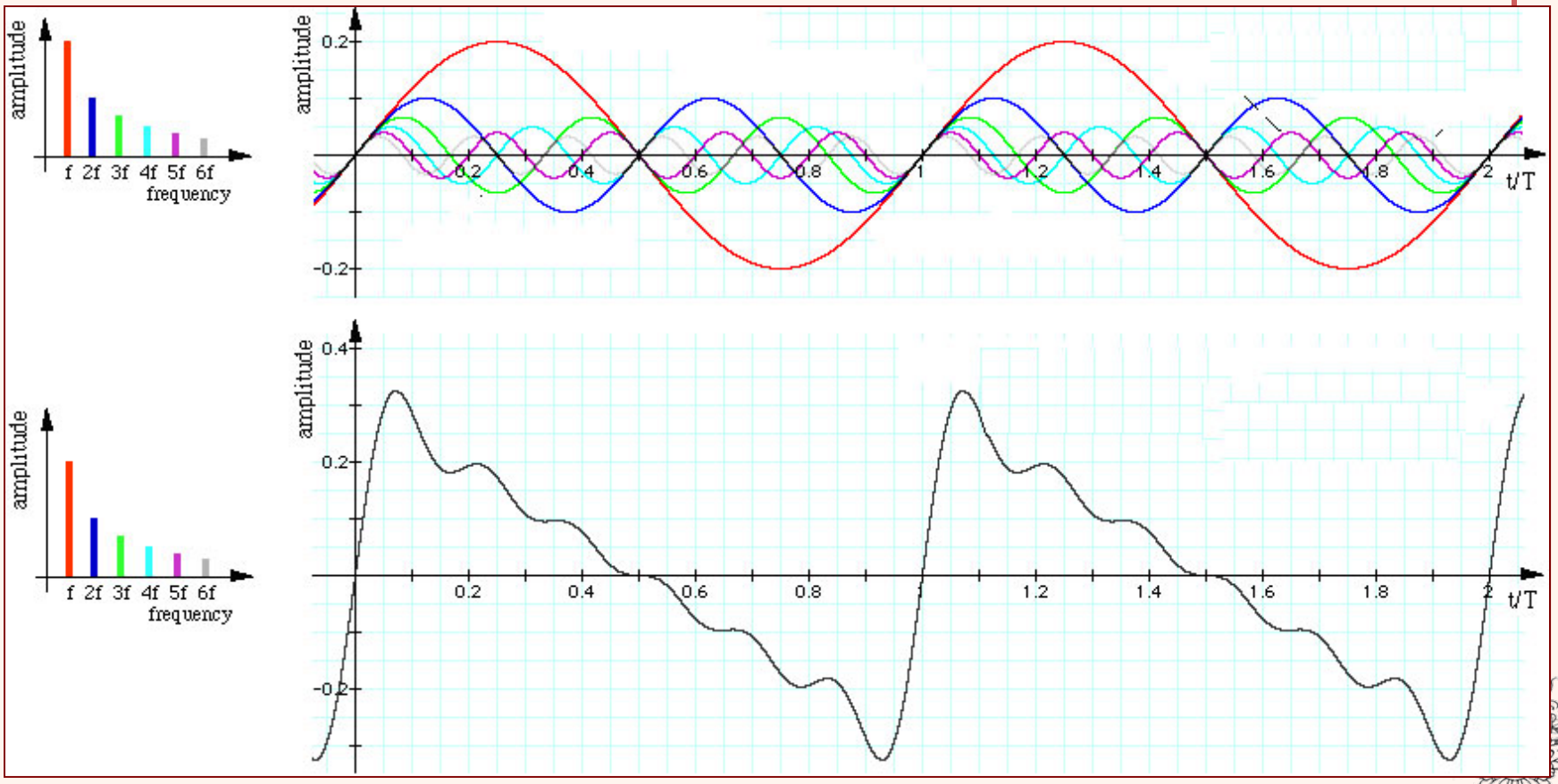
x = sin(2*pi*200*t) +sin(2*pi*400*t)+ sin(2*pi*600*t)+sin(2*pi*800*t)+sin(2*pi*1000*t);



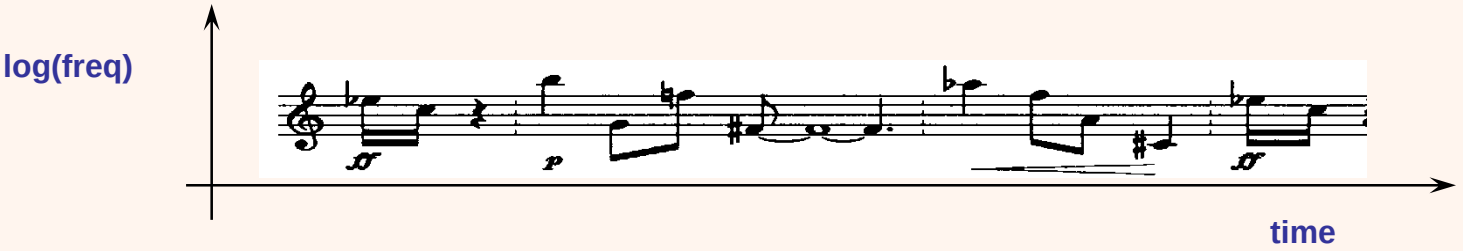
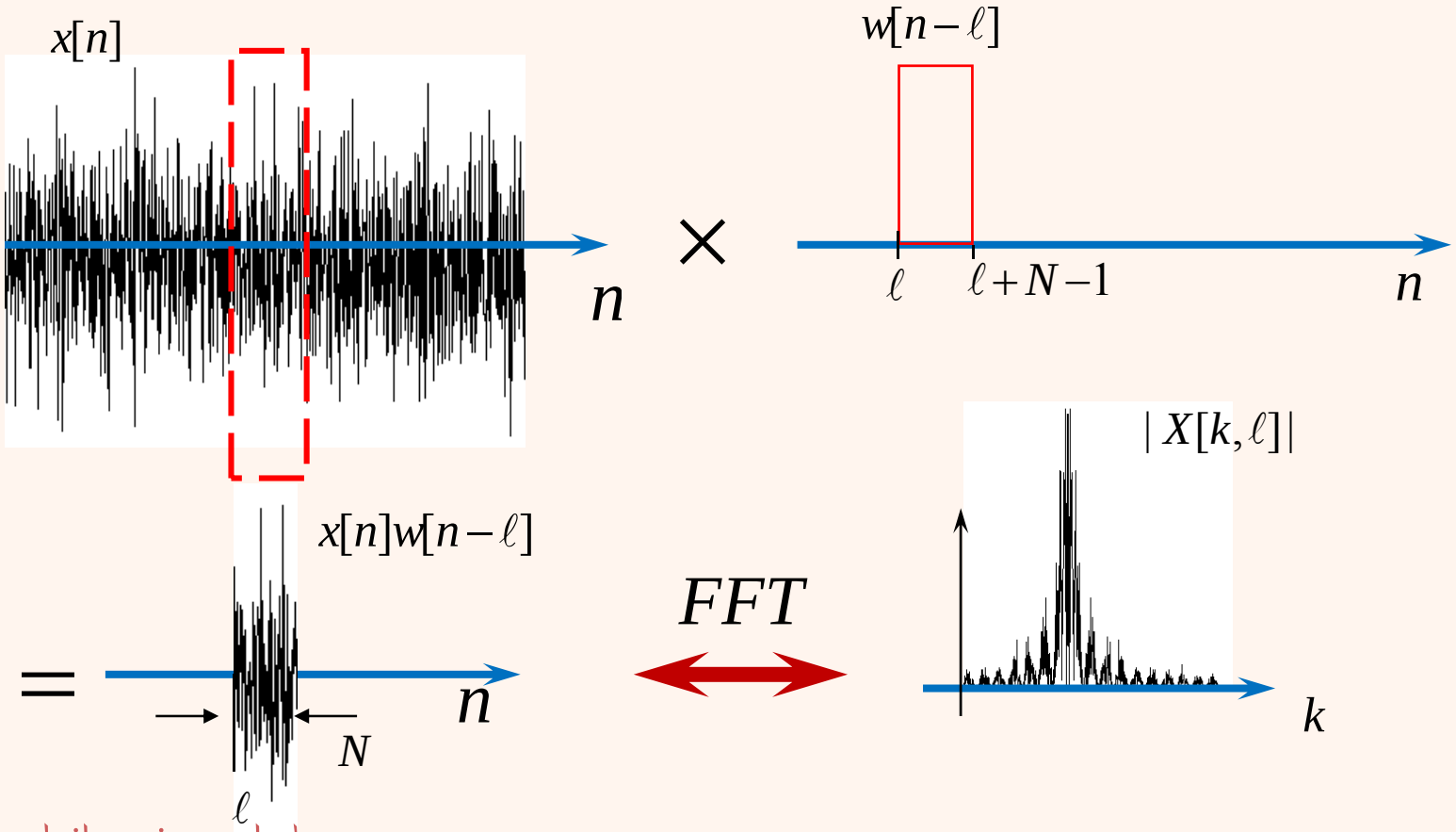


outsin3.wav

مثال

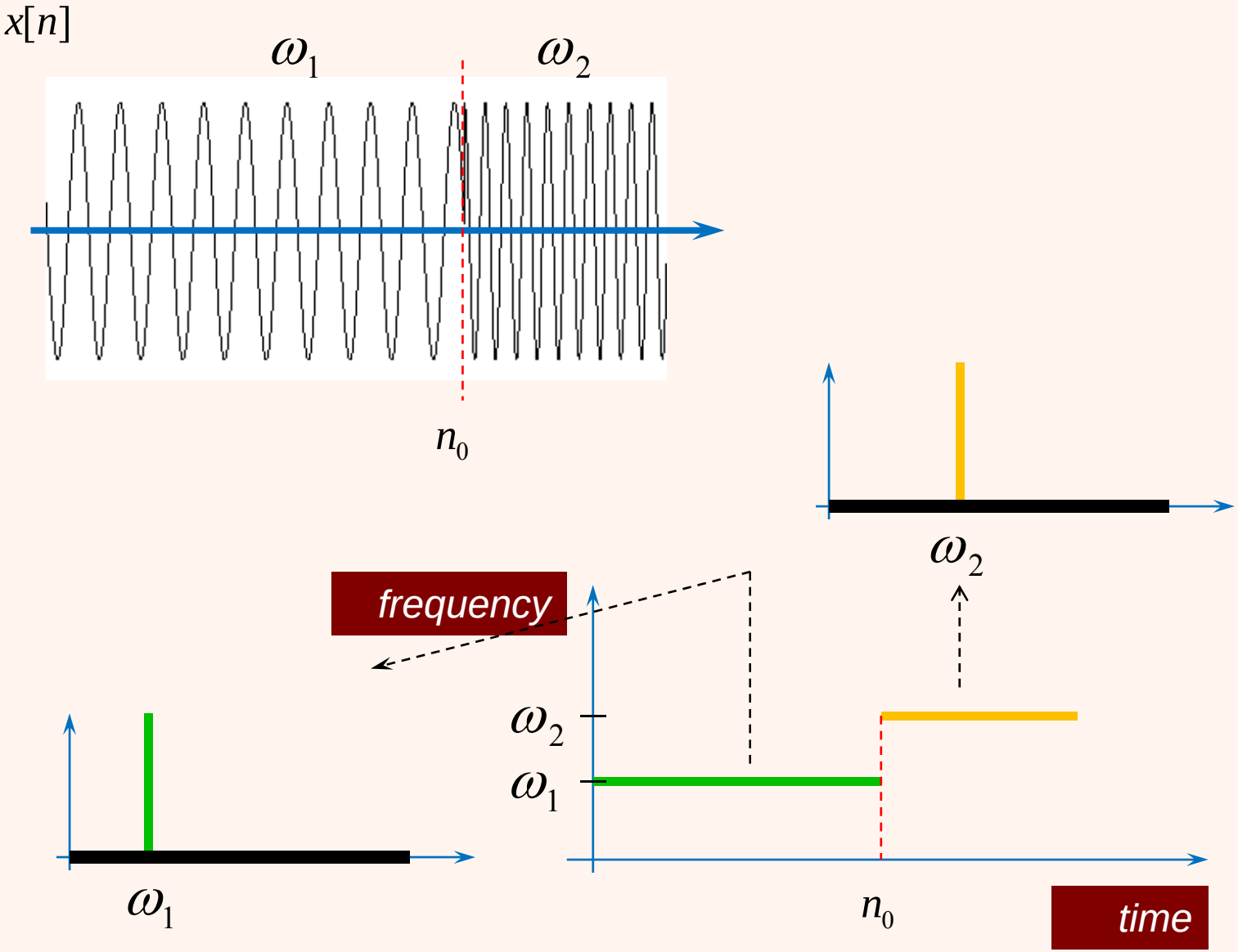

amplitude
f 2f 3f 4f 5f 6f
frequency
amplitude
0.2
-0.2
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
t/T
amplitude
f 2f 3f 4f 5f 6f
frequency
amplitude
0.4
0.2
-0.2
0.2
0.4
0.6
0.8
1.2
1.4
1.6
1.8
2
t/T

Short-time Fourier transform

تبدیل فوریه زمان کوتاه



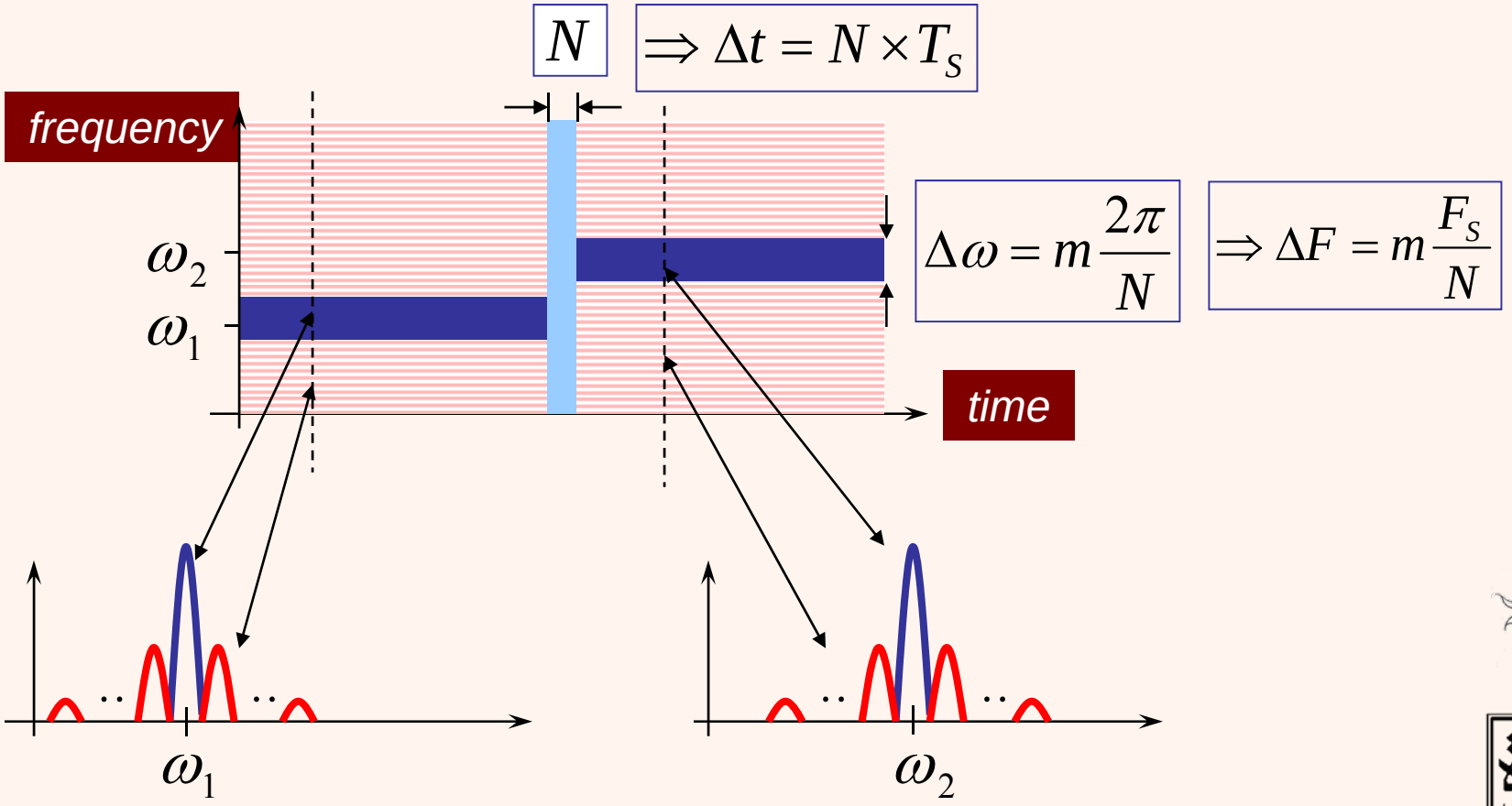
Spectrogram

حالت ایده‌آل



Spectrogram

حالت واقعی



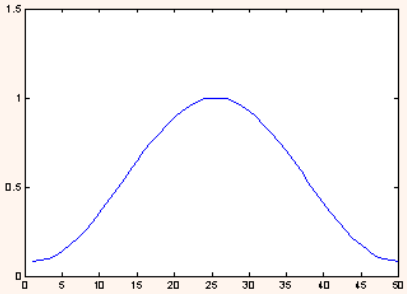
Windows

انواع پنجره

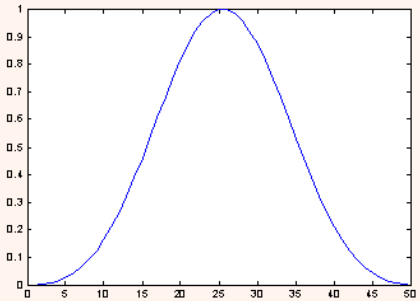
Rectangular **Hamming** **Blackman**

$w[n]$

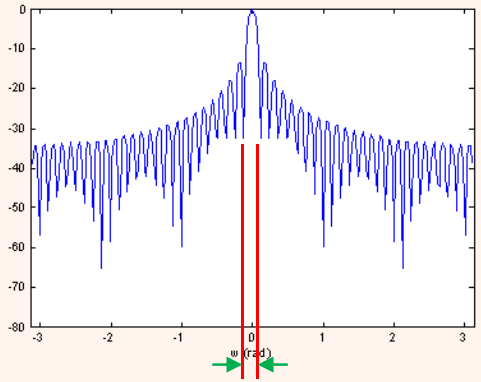
$w[n]$



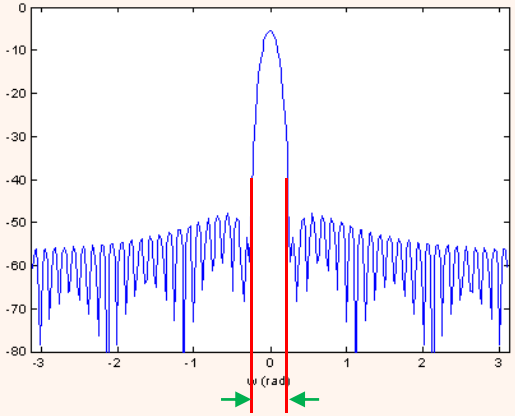
$w[n]$



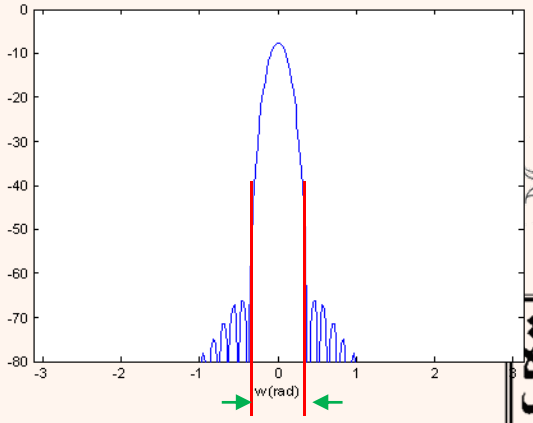
$|W_N(\omega)|$



$|W_N(\omega)|$



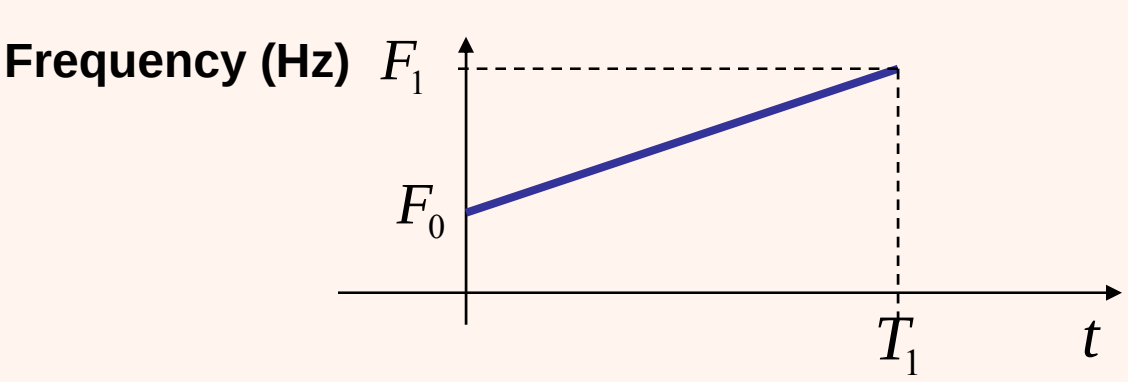
$|W_N(\omega)|$



- تابعی که فرکانس آن همراه با زمان تغییر کند را Chirp می‌گویند.

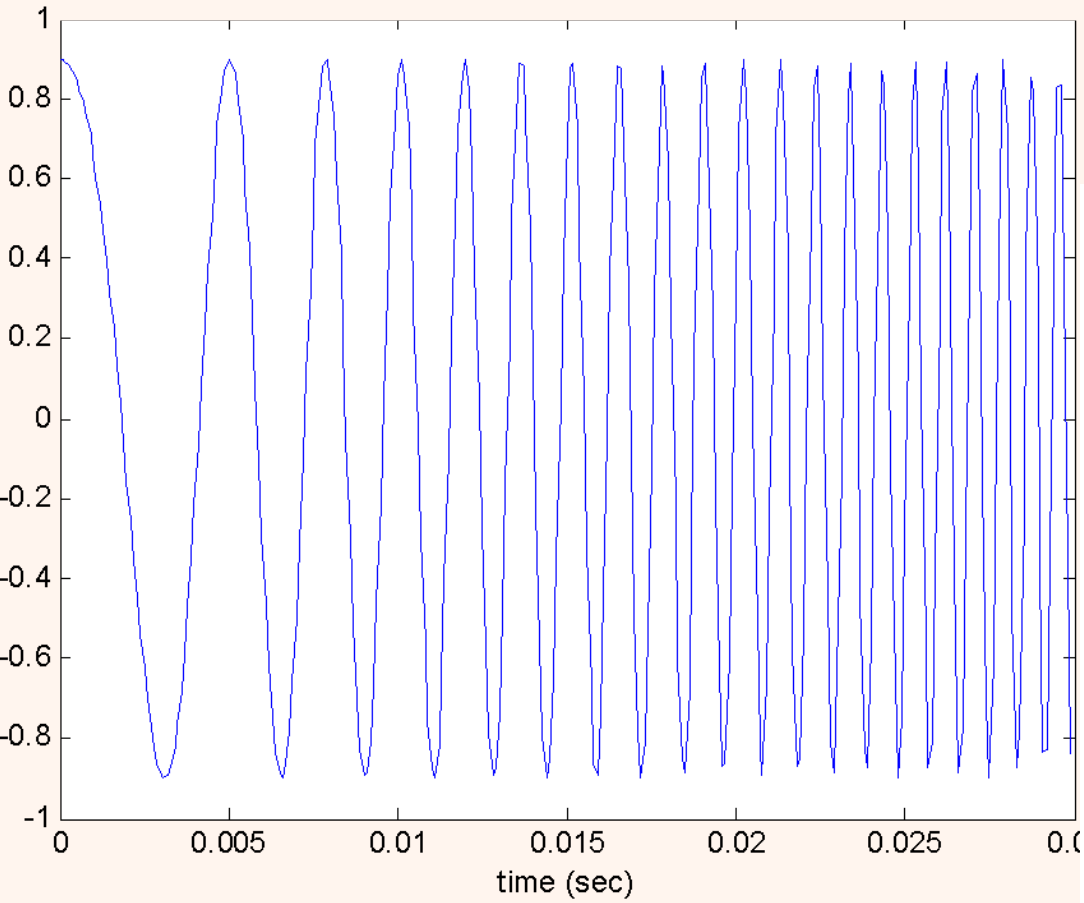
$$x(t) = A\cos\left(2\pi F(t)t + \alpha\right)$$

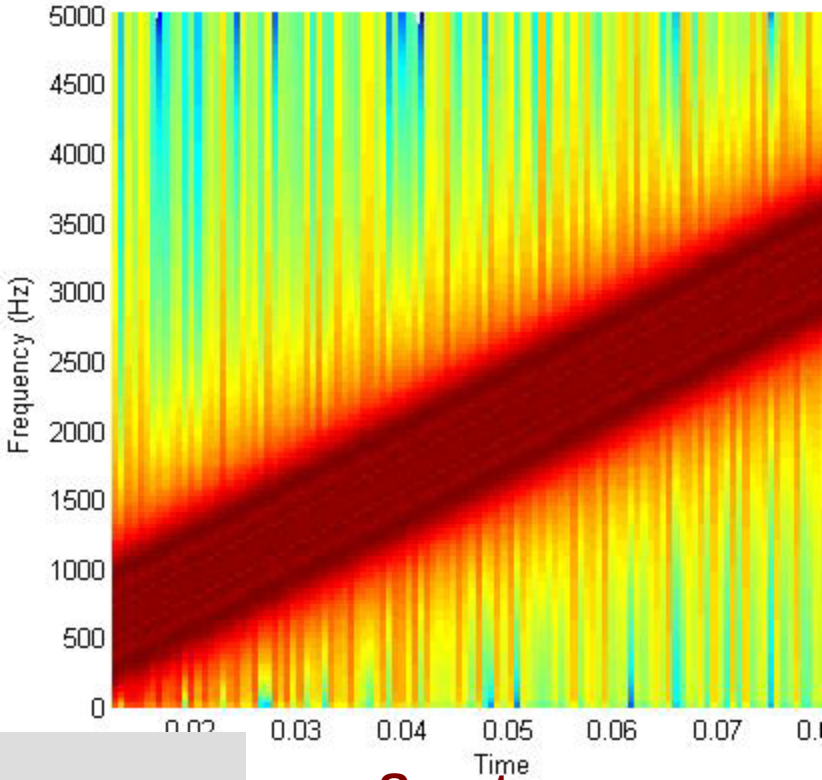
- اگر فرکانس خطی بر حسب زمان تغییر کند.



مثال

تغییرات فرکانس بر حسب زمان





Spectrogram

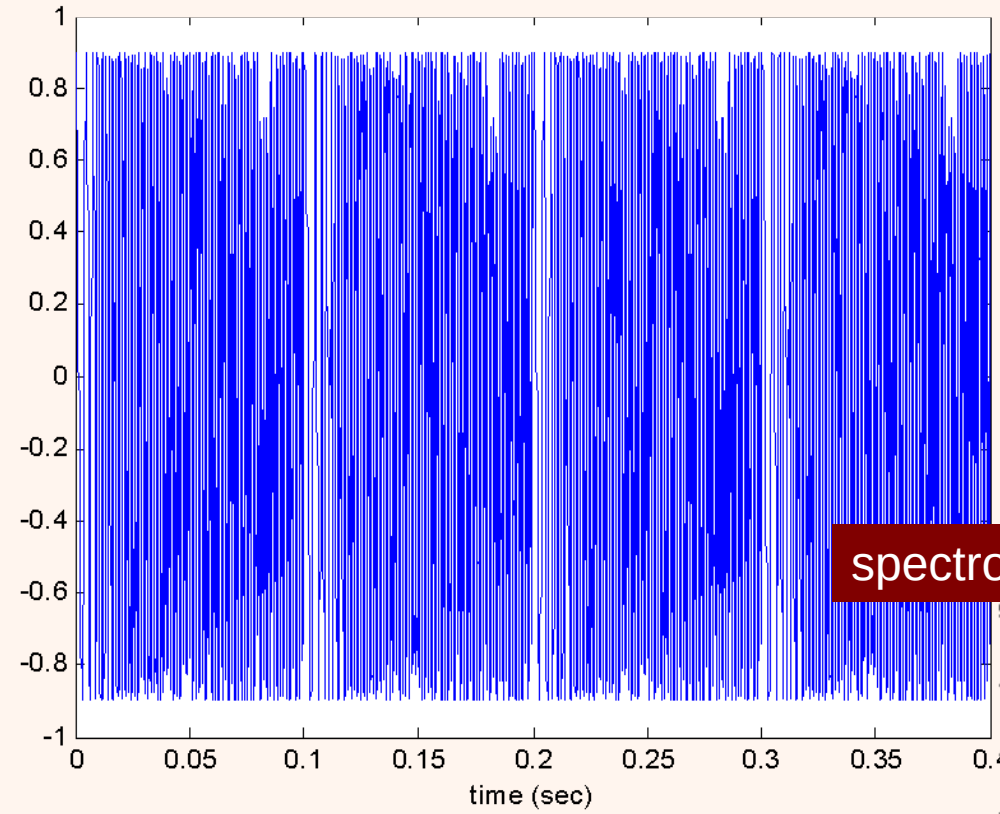
```
Fs=10000; Ts=1/Fs; t=(0:999)*Ts;
y=chirp(t, 100, t(1000), 4000);
plot(t(1:300), y(1:300));
sound(y,Fs);
wavwrite(y,Fs,16,'chirp.wav');
spectrogram(y, rectwin(256), 250, 256,Fs,'yaxis');
```

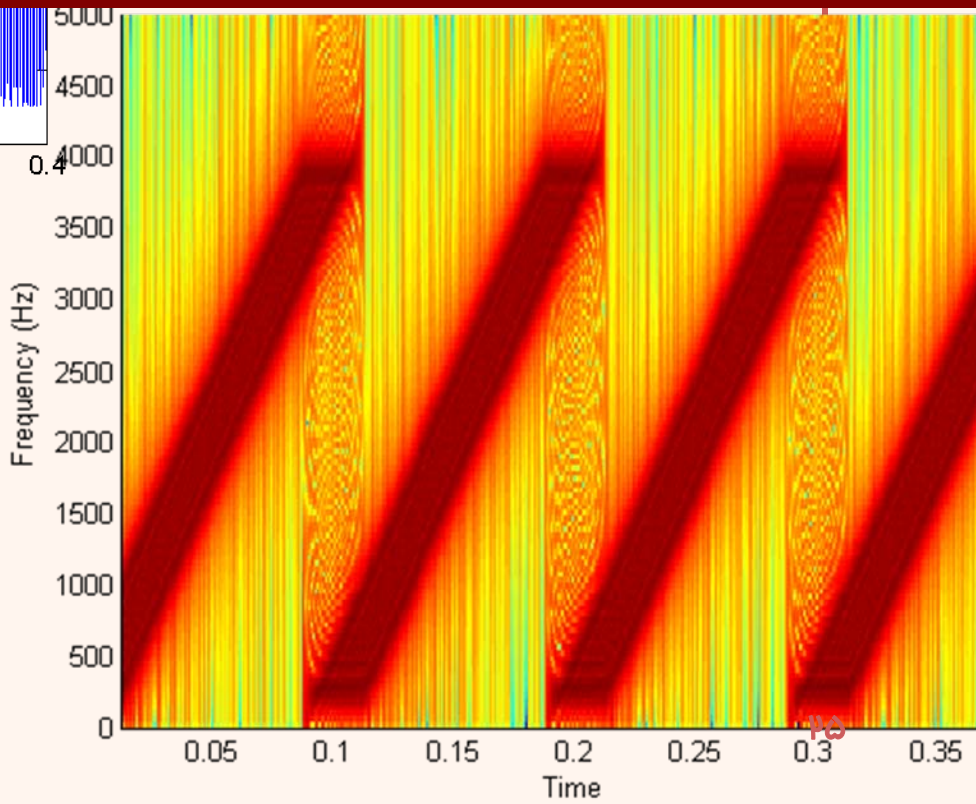
chirp.wav

مثال

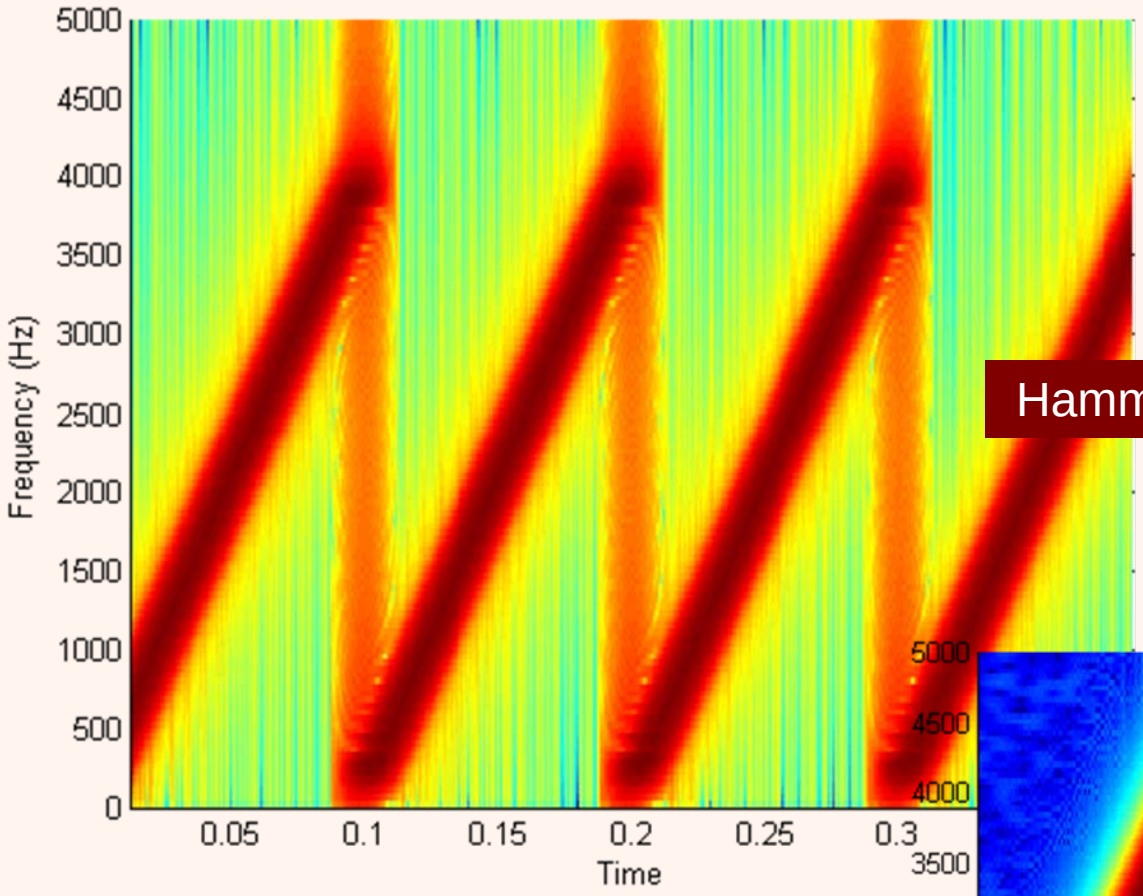
چهار تکرار تابع قبلی

```
spectrogram(y, rectwin(256), 250, 256,Fs,'yaxis');
```

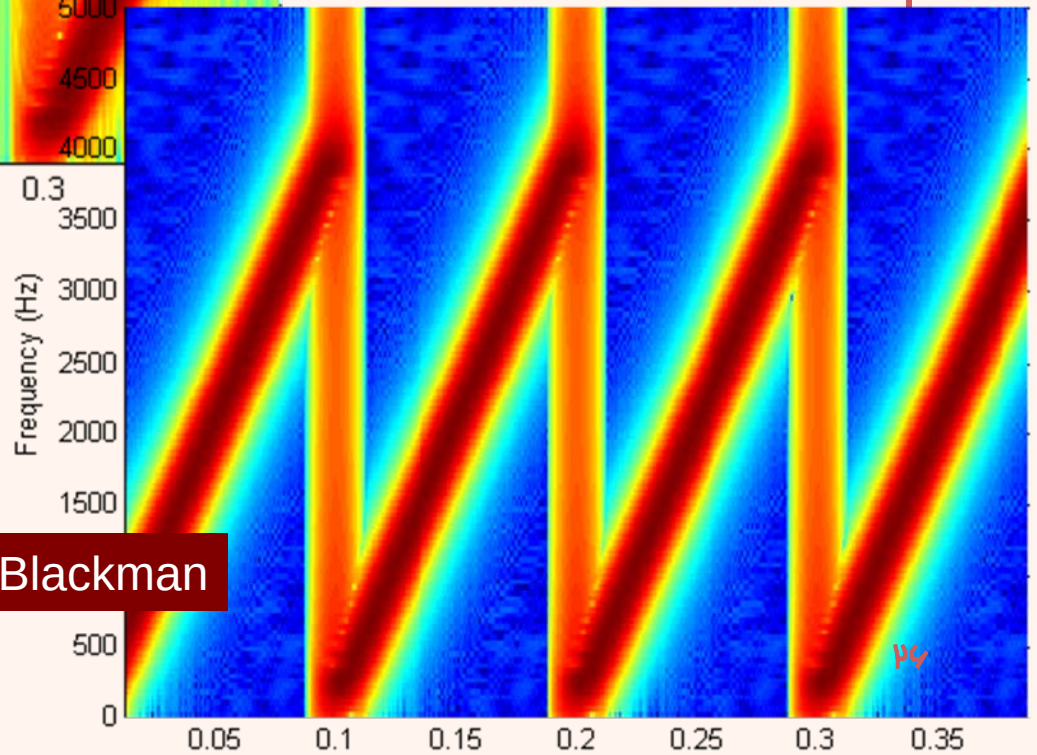




مثال



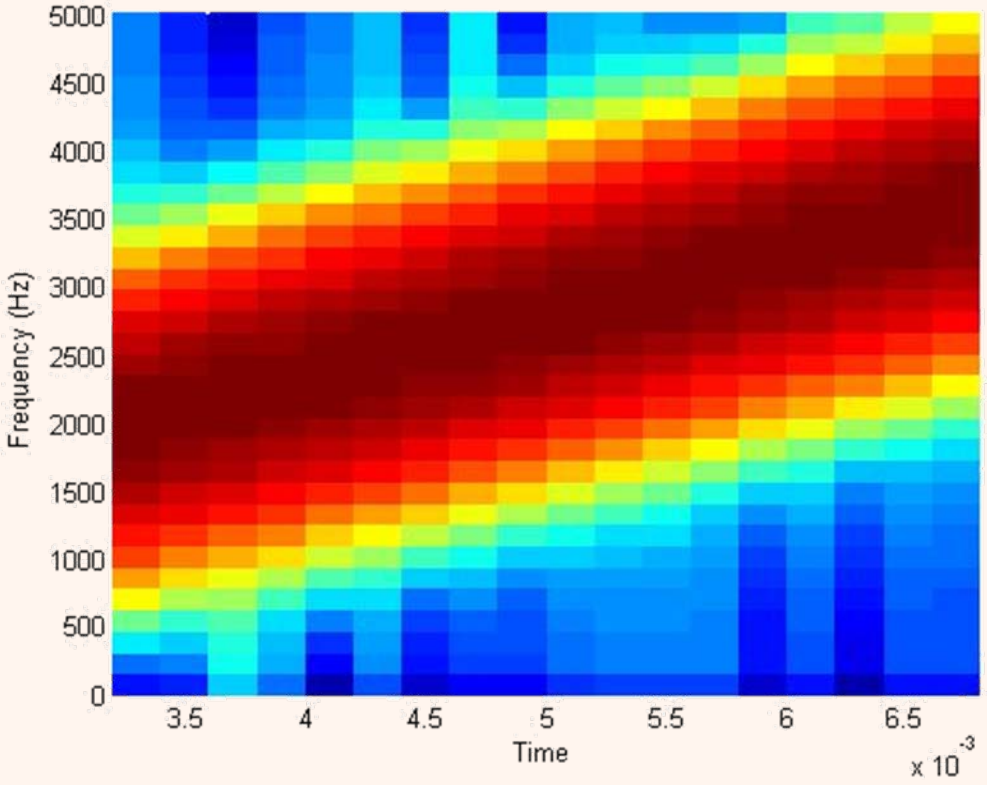
Hamming



Blackman

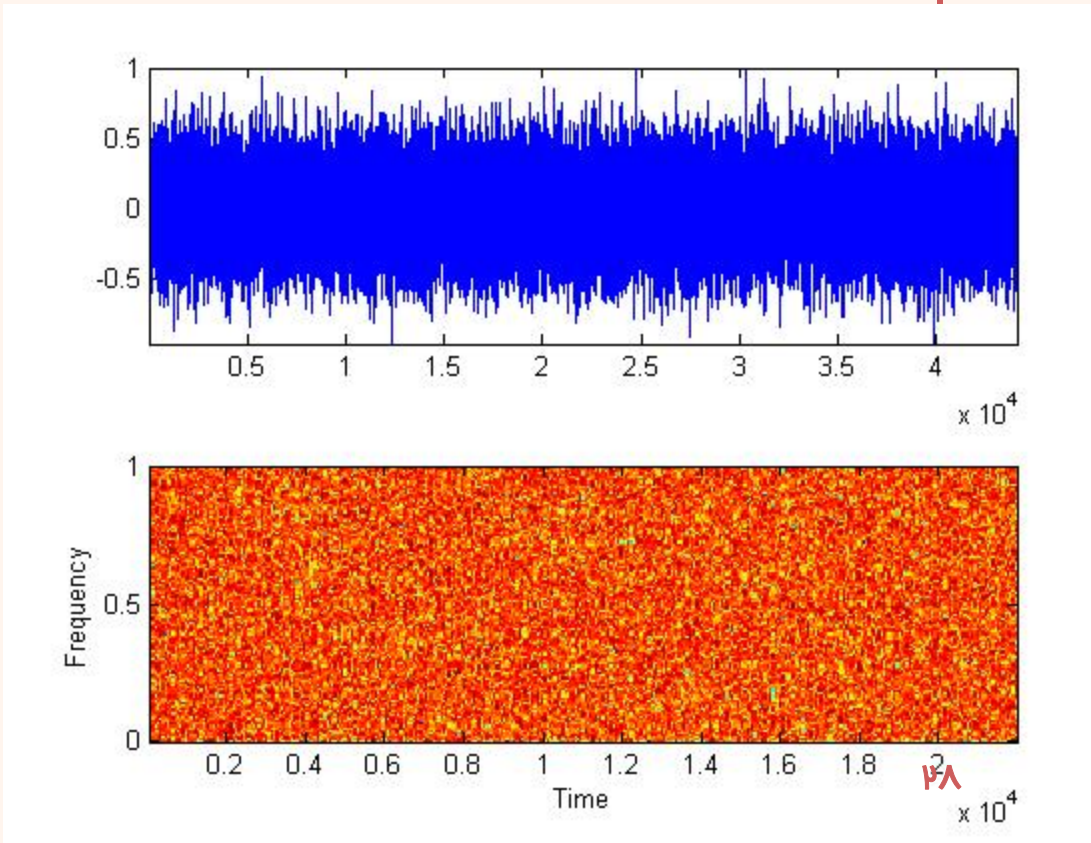
مثال

```
spectrogram(y, hamming(64), 62, 64,Fs,'yaxis');
```



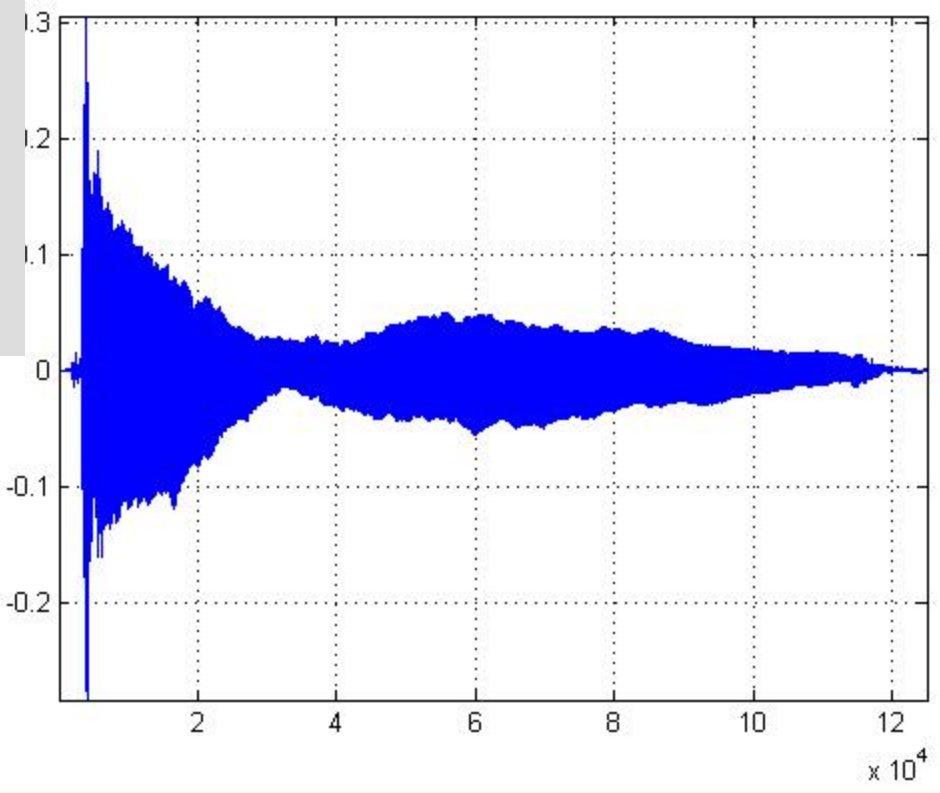
مثال

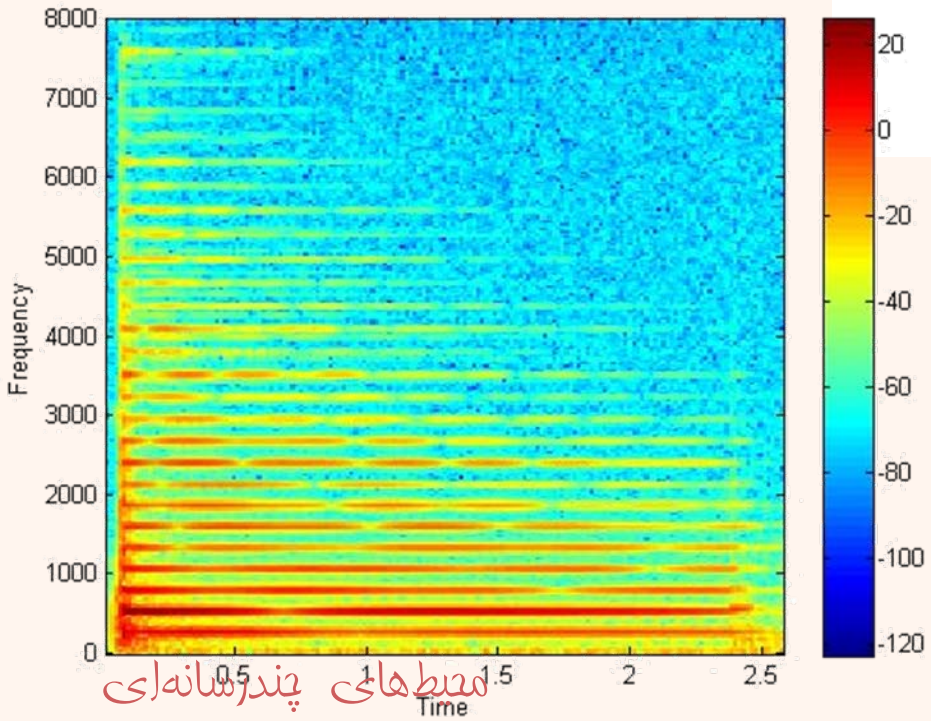
```
Fs = 44100;
n = randn(3*Fs, 1);
n = n / max(abs(n));
subplot(211)
plot(n);
subplot(212);
specgram(n);
sound(n,Fs);
wavwrite(n,Fs,16,'noise.wav');
```



noise.wav

```
[s, fs] = wavread('pia60.wav');
plot(s),grid on;
figure;
s2 = resample(s, 1, 3);
specgram(s2, 512, fs / 3);
sound(s,fs);
```

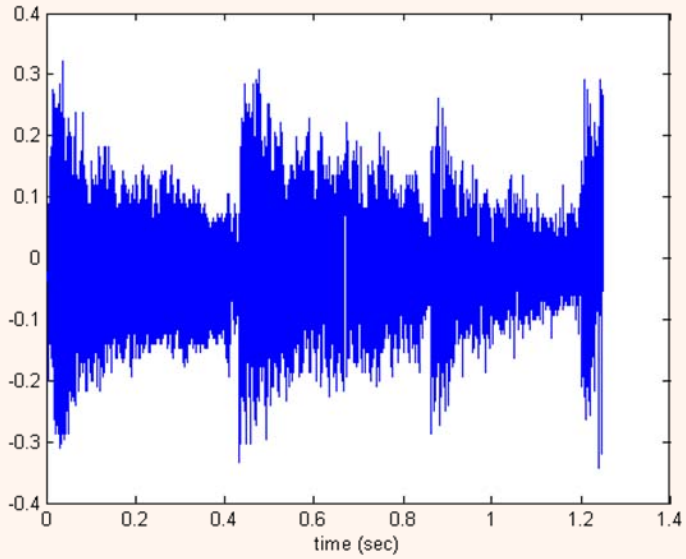


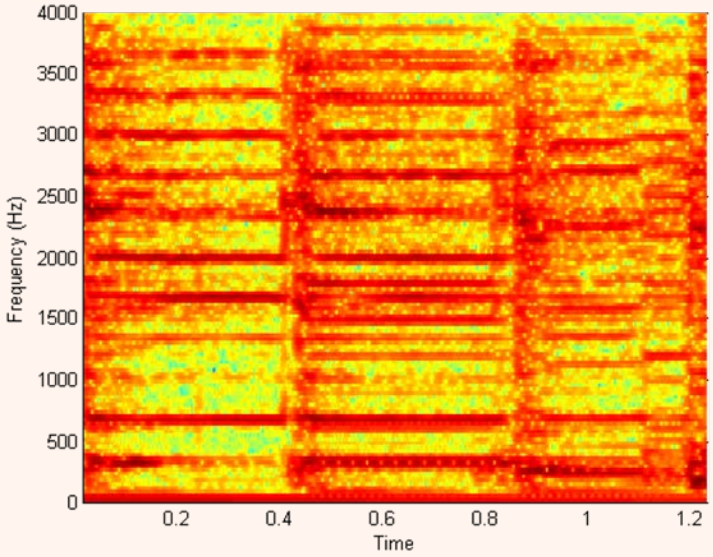


pia60.wav

مثال

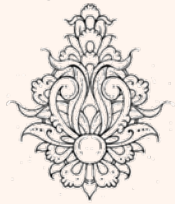
HENDRIX


0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
0
0.2
0.4
0.6
0.8
1
1.2
1.4
time (sec)


4000
3500
3000
2500
2000
1500
1000
500
0
Frequency (Hz)
0.2
0.4
0.6
0.8
1
1.2
Time

نگاهی دوباره به نمونه‌برداری

- چنان‌چه گفته شد، در نمونه‌برداری از سیگنال‌های پیوسته، برای این که سیگنال گسسته به دست آمده قابل تبدیل به سیگنال اولیه باشد، باید نرخ نمونه‌برداری حداقل **دو** برابر بالاترین مؤلفه‌ی فرکانسی موجود در سیگنال اولیه باشد.
  - برای مثال در مورد صوت انسان که حداکثر فرکانس آن ۴ کیلوهرتز است این فرکانس برابر ۸ کیلوهرتز باید باشد.

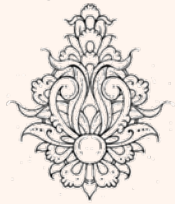


اندازه‌ی فایل صوتی

- اندازه‌ی یک فایل صوتی از رابطه‌ی زیر قابل محاسبه است:

$$S = R \times (b/8) \times C \times D$$

S file size	bytes
R sampling rate	samples per second
b resolution	bits
C channels	1 - mono, 2 - stereo
D recording duration	seconds



مثال

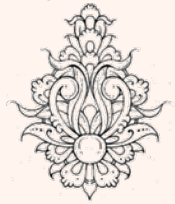
- اگر ده ثانیه از یک موزیک استریو با فرکانس 44.1 کیلوهرتز را با 16 بیت ذخیره‌سازی کنیم، اندازه‌ی فایل را حساب کنید:

$$S = 44100 \times (16/8) \times 2 \times 10$$

$$= 1{,}764{,}000 \text{bytes}$$

$$= 1722.7 \text{Kbytes}$$

$$= 1.68 \text{Mbytes}$$



Signal to Noise Ratio(SNR)

نسبت سیگنال به نویز

- معیاری است که نسبت توان سیگنال به نویز(پس‌زمینه) را نشان می‌دهد، هر چه این نسبت بیشتر باشد، کیفیت سیگنال بهتر است.

$$SNR = 10\log_{10}\frac{V_{signal}^2}{V_{noise}^2} = 20\log_{10}\frac{V_{signal}}{V_{noise}}$$

- واحد اندازه‌گیری بر حسب dB است.
- برای سیگنال‌های دیجیتال از واحد دیگری نیز استفاده می شود:

Signal to Quantization Noise Ratio(SQNR)

- در این حالت دقت بستگی به تعداد بیتی که سیگنال به واسطه‌ی آن نمایش داده می‌شود، دارد.

Signal to Quantization Noise Ratio (SQNR, SNqR)

- جدا از نویزی که معمولا همراه با سیگنال وجود دارد، نویز دیگری که حاصل از فرآیند چندی کردن اتفاق می‌افتد نیز می‌باید بررسی گردد.
- کیفیت سیگنال چندی‌شده با پارامتری با نام SQNR سنجیده می‌شود.

quantization noise

- در بدترین حالت خطا متناسب است با نیمی از بازه‌ی چندی شدن

$$SQNR = 20\log_{10}\frac{V_{signal}}{V_{quan-noise}}$$

خطای چندی کردن

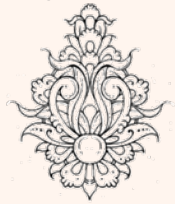
the *Peak signal-to-noise ratio, PSQNR: peak* signal and peak noise

- در یک مبدل ایده‌آل آنالوگ به دیجیتال که میزان نویز چندی شدن به صورت یکنواخت توزیع شده باشد، SQNR به طریق زیر به دست می‌آید:

$$PSQNR = 20\log_{10}\frac{2^{N-1}}{1/2}$$

$$= 20 \times N \times \log_{10} 2 \ = 6.02N(dB)$$

- **PSQNR is approximately 6 dB per bit.**
  - **N~SQNR**
  - **16-bit => 96 dB**



اندازه‌گیری شدت صوت

- صدا در واقع در اثر اختلاف فشار (SPL) منتقل می‌شود.
  - گوش انسان توانایی شنیدن صدا در بازه‌ی 20μPa تا 20Pa را دارد، به دلیل بزرگی این بازه برای اندازه‌گیری شدت صوت از مقیاس لگاریتمی استفاده می‌شود.

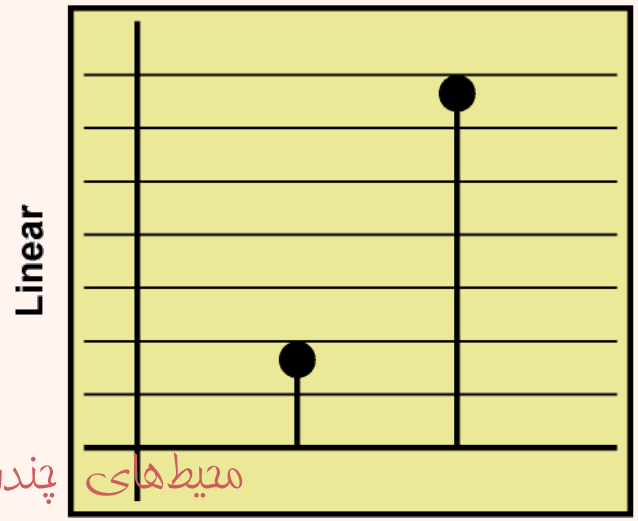
Threshold of hearing	0
Rustle of leaves	10
Very quiet room	20
Average room	40
Conversation	60
Busy street	70
Loud radio	80
Train through station	90
Riveter	100
Threshold of discomfort	120
Threshold of pain	140
Damage to ear drum	160

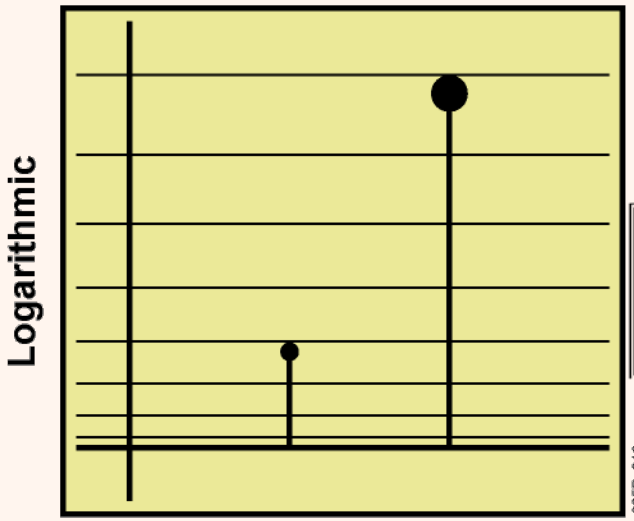
توجه: هرچند نسبت سیگنال به نویز و شدت صوت هر دو با dB بیان می‌شوند، باید در نظر داشت که دو مفهوم متفاوت هستند.

سانه‌ای

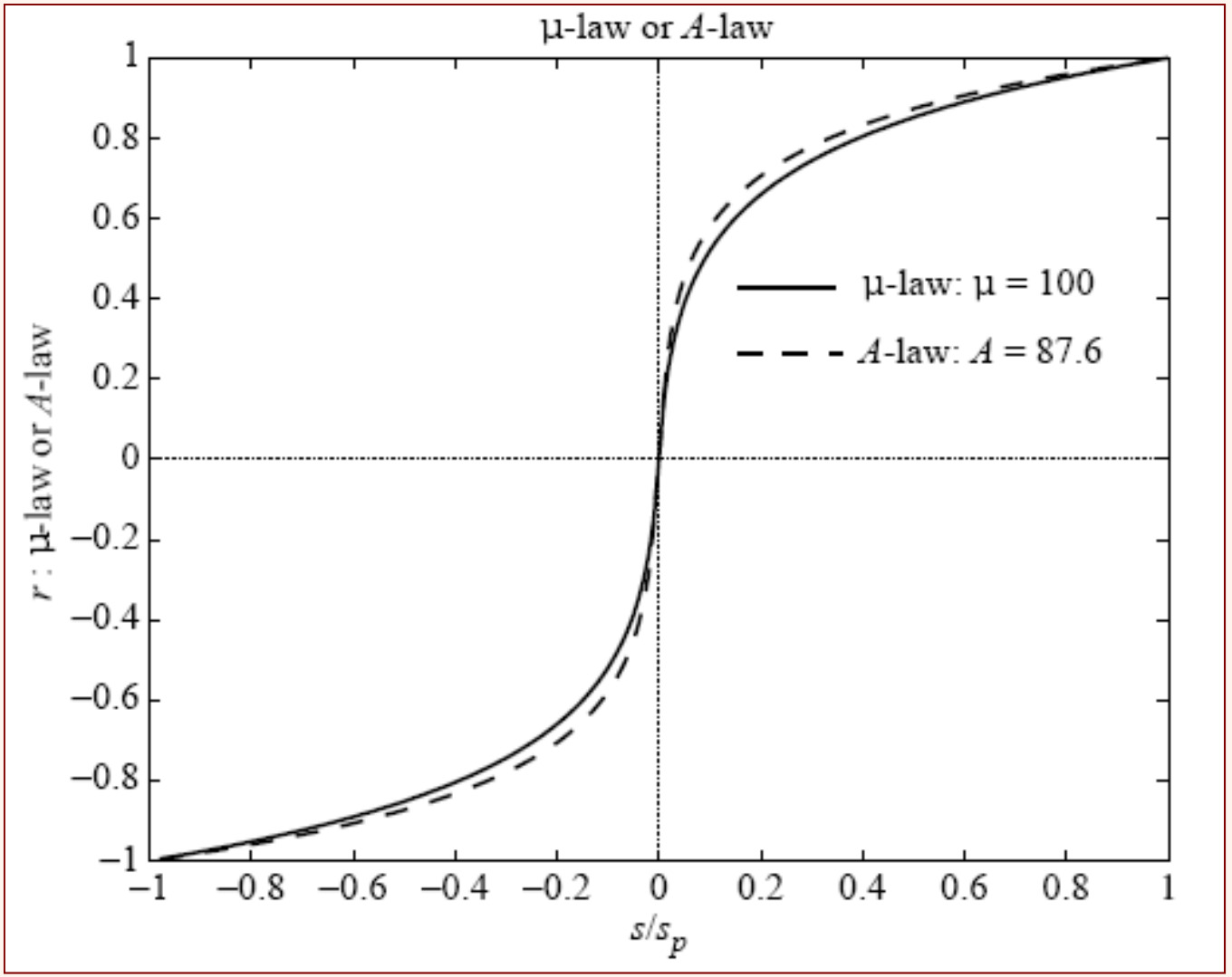
روش‌های چندی‌کردن

- **چندی‌کردن خطی** (Linear quantization)
  - SNR پایین برای سیگنال‌هایی با دامنه‌ی کم
  - SNR بالا برای سیگنال‌هایی با دامنه‌ی زیاد
- **چندی‌کردن (غیرخطی)غیریکنواخت** (Non-uniform quantization)
  - SNR یکسان برای انواع سیگنال‌ها





چندی کردن غیریکنواخت



- بهتر است روش چندی نمودن متناسب با عملکرد سیستم شنوایی انسان باشد.

A-law, is used in telephony in Europe

چندی نمودن غیرخطی

mozart.wav

mozartMUQ16.wav

mozartMUQ8.wav

mozartMUQ4.wav

چندی نمودن خطی

mozartQ16.wav

mozartQ8.wav

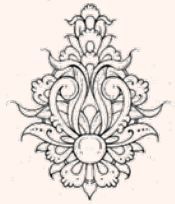
mozartQ4.wav

پیش‌پردازش

- قبل از نمونه‌برداری و تبدیل به سیگنال گسسته می‌باید برای سیگنال صوت فرکانس‌های ناخواسته را از میان برداریم.
- فرکانس‌های باقی‌مانده به کاربرد مورد نظر بستگی خواهند داشت.
- برای سیگنال صحبت اگر بازه‌ی مورد نظر از ۵۰ هرتز تا ۱۰ کیلو هرتز در نظر گرفته شود.
  - حذف فرکانس‌های بالا و پایین عموما به واسطه‌ی یک فیلتر میان‌گذر bandpass filter صورت می‌گیرد.

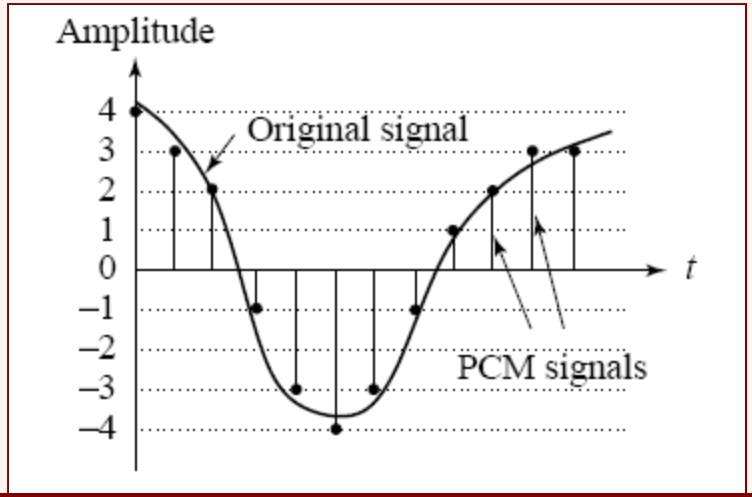
پیش‌پردازش (ادامه ...)

- برای سیگنال موسیقی بازه مورد نظر از ۲۰ هرتز تا بالاتر از ۲۰ کیلو هرتز است.
- برای استاندارد تلفنی چون سیگنال‌های تا ۴ کیلوهرتز را نیاز داریم می‌باید بیش‌تر از این فرکانس را از فایل ورودی جدا کنیم این مساله برای فرکانس های خیلی کوچک هم صادق است.

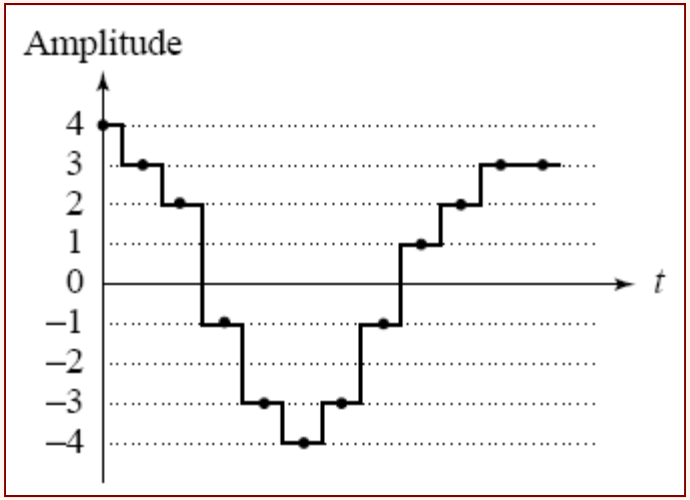


اعمال فیلتر پایین‌گذر

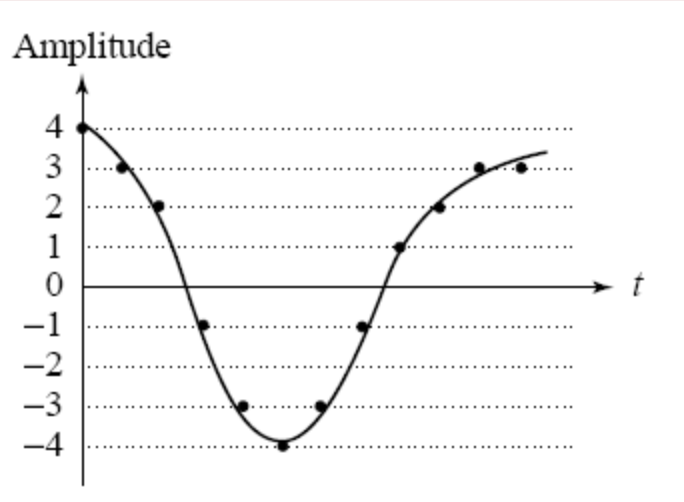
- معمولا پس از بازسازی به دلیل چندی شدن سیگنال نتیجه شده پله‌ای و دارای فرکانس‌های بالاست.
  - سیگنال بازسازی شده به دلیل بازسازی از سیگنال چندی شده یک سیگنال staircase است.
- سیگنال بازسازی شده سیگنالی پله‌ای است و می‌توان گفت از جهت تئوری به معنای حضور تعداد زیادی فرکانس‌بالاست.
- این مساله با آنالیز فوریه قابل توجیه است.
- فرکانس‌های بالای این دست سیگنال زائد هستند.
- پس سیگنال نتیجه شده را از یک فیلتر Lowpass عبور می‌دهیم تا نتیجه‌ای قابل قبول به لحاظ سیگنال صوت داشته باشیم.



Original analog signal
and its corresponding PCM signals

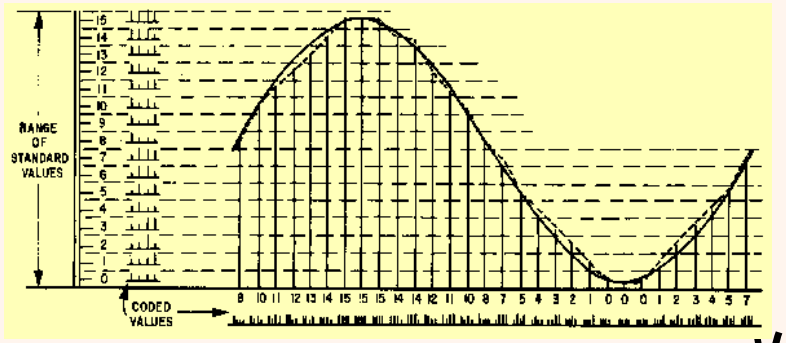


Decoded staircase signal



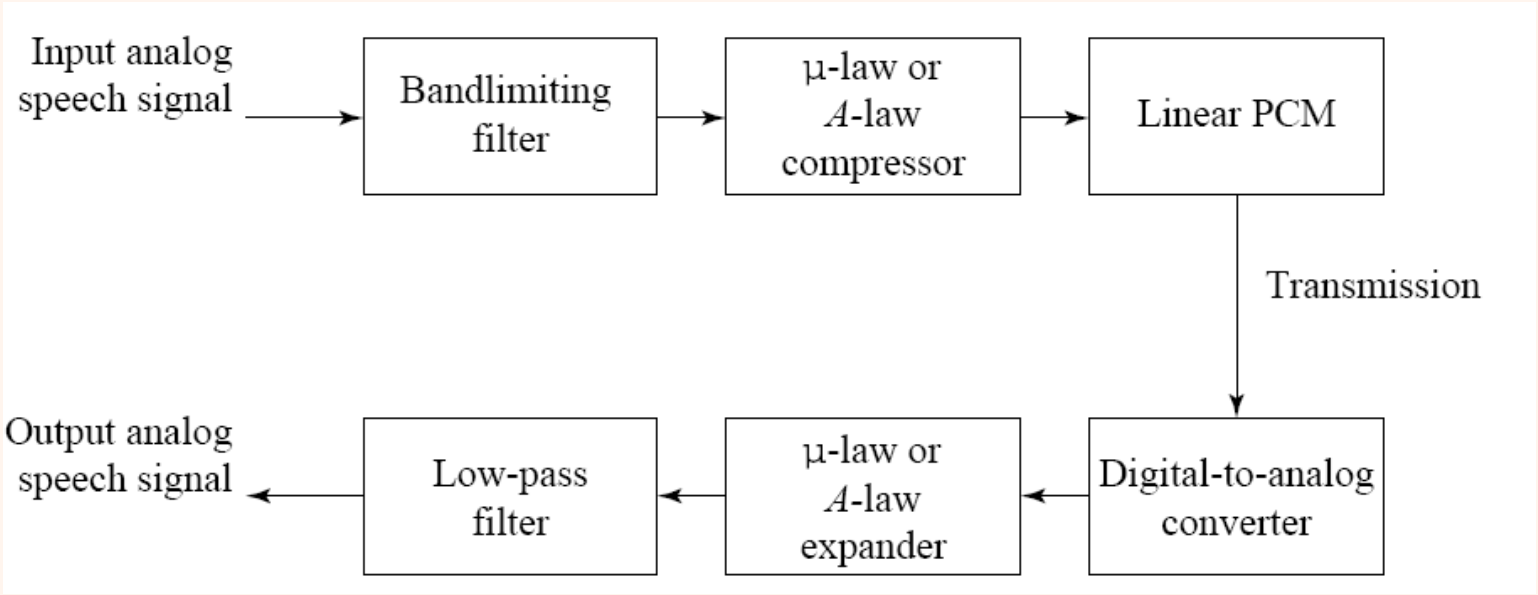
Reconstructed signal after low-pass filtering

مدولاسیون کد پالس



- استفاده از کدینگ PCM
  - تلفن‌های دیجیتال
  - سیستم‌های صوتی دیجیتال
  - صدای کامپیوتر
  - برخی پخش‌کننده‌های صوت دیجیتال
- در سی‌دی های صوتی فرکانس نمونه‌برداری ۴۴،۱کیلوهرتز و تعداد سطوح چندی ۶۵۵۳۶ (معادل ۱۶ بیت) است
- در استاندارد تلفن بالاترین فرکانس را ۴ کیلوهرتز در نظر می‌گیرند و بنابراین فرکانس نمونه‌برداری ۸ کیلو هرتز در نظر گرفته می‌شود.

PCM signal encoding and decoding



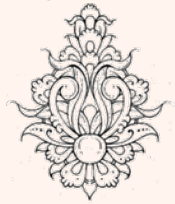
PCM Audio Data Rate and Data Size

Quality	Sampling Rate (KHz)	Bits per Sample	Data Rate Kbits/s Kbytes/s	Data Size in 1 minute 1 hour
Telephone	8	8 (Mono)	64Kbps 8	480KB 28.8MB
AM Radio	11.025	8 (Mono)	88.2Kbps 11.0	660KB 39.6MB
FM Radio	22.050	16 (Stereo)	705.6Kbps 88.2	5.3MB 317.5MB
CD	44.1	16 (Stereo)	1.41Mbps 176.4	10.6MB 635MB

با توجه به حجم بالای مورد نیاز به فشرده‌سازی سیگنال‌های صوتی بدیهی است.

روش‌های فشرده‌سازی

- به صورت کلی فراهم آوردن خروجی نمونه‌برداری و چندی شده‌ی یک سیگنال صدا را PCM یا همان (Pulse Code Modulation) می‌نامند.
- نسخه‌ای که از تفاوت سیگنال‌ها استفاده می‌کند را DPCM می‌نامند.
- نسخه‌ی منعطف آنرا نیز ADPCM گویند.



روش‌های فشرده‌سازی (ادامه...)

- تعداد ارزش‌هایی که پس از بازسازی سیگنال نمونه‌برداری و چندی‌شده به دست می‌آید را **سطوح بازسازی یا** (Reconstruction Level) می‌نامند.
- فشرده‌سازی

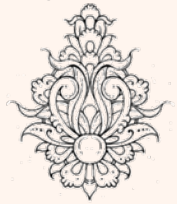
– تبدیل (به گونه‌ای که سیگنال تبدیل‌یافته آسان‌تر فشرده گردد).

– محدودیت برای تعداد سطوح سیگنال بازسازی شده (اتلاف داده‌ها)

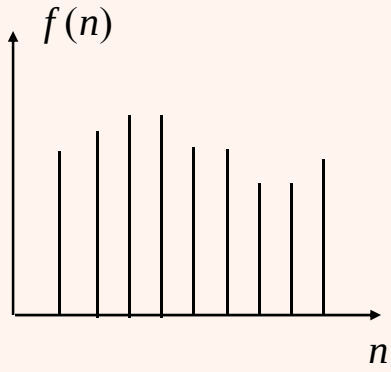
– کد نمودن (اختصاص تعداد بیت‌هایی برای کدکردن داده (طول ثابت یا متغیر همانند کدگذار هافمن))

Deferential Coding of Audio

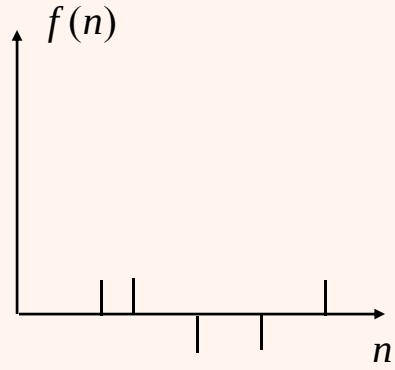
- سیگنال صوتی معمولا به فرم PCM ذخیره نشده به جای آن از فرمت DPCM که اختلاف دو سیگنال را ذخیره می‌کند استفاده می‌شود.
- در این حالت کدینگ با تعداد بیت‌های کم‌تر صورت می‌گیرد.
- اگر یک سیگنال وابسته به زمان داشته باشیم (که افزونگی زمانی دارد) با محاسبه‌ی تفاوت نمونه‌ی جاری از نمونه‌ی قبلی، سیگنالی خواهیم داشت که مقادیر آن در نزدیکی صفر متمرکز است.



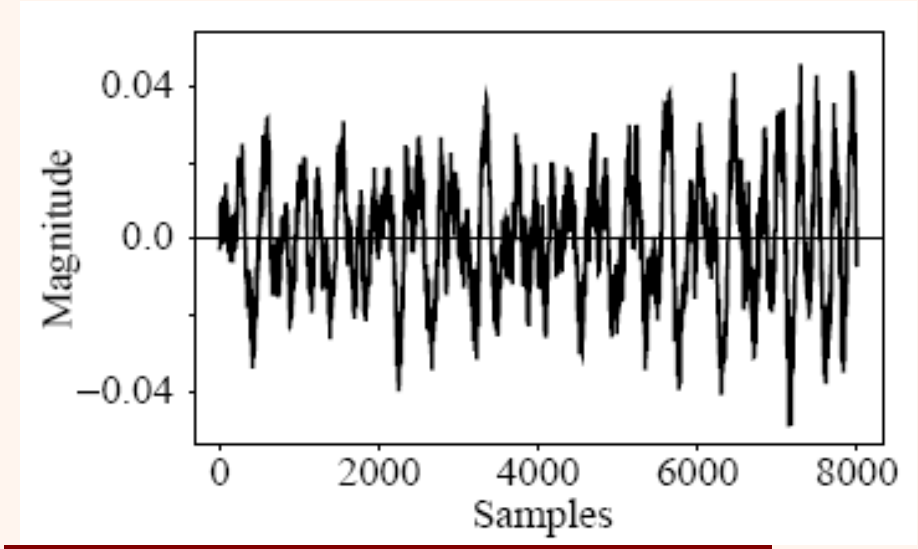
Differential Pulse Code Modulation (DPCM)



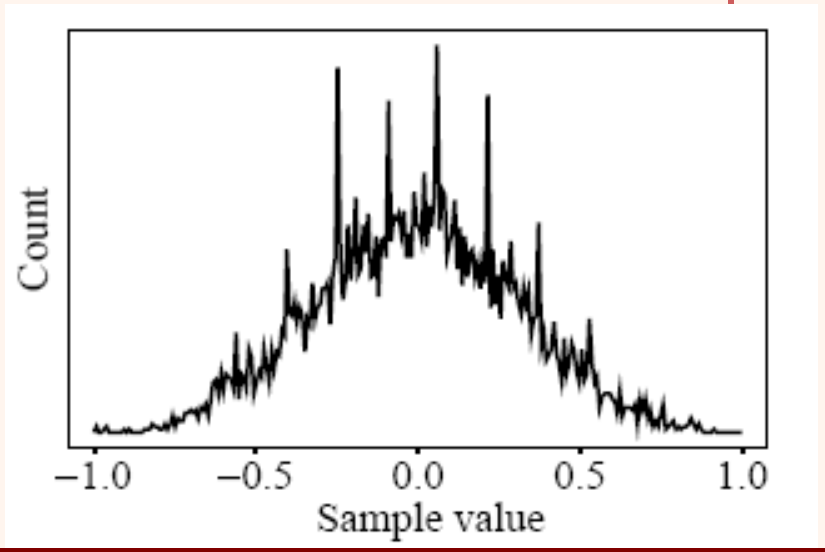
$$f(n) = 156,157,158,158,156,156,154,154,155$$



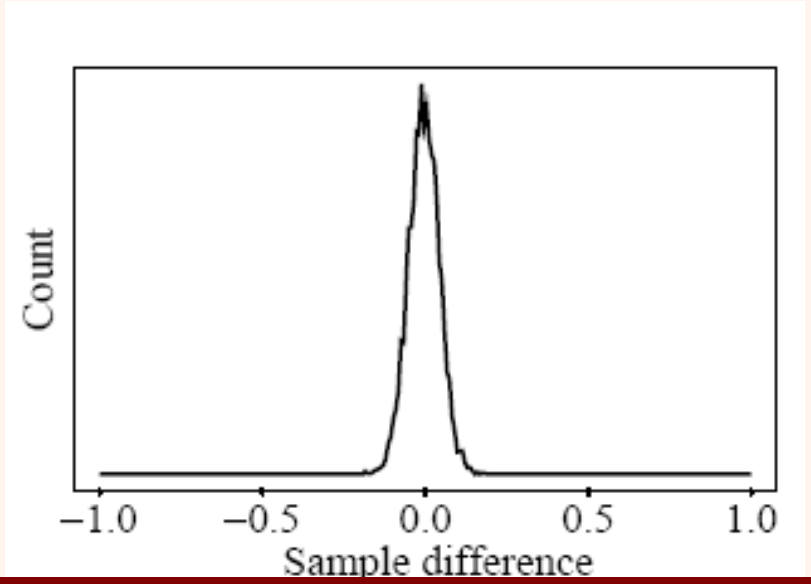
$$\delta f(n) = 156,1,1,0,-1,0,-1,0,1$$



Digital speech signal



Histogram of digital speech signal values



Histogram of digital speech signal differences

Li & Drew Hall 2003

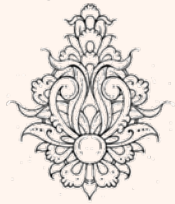
f_n – the original signal

$\hat{f}_n$ – the predicted signal

$\tilde{f}_n$ the quantized, reconstructed signal

- به واسطه‌ی کسر نمودن سیگنال پیش‌بینی شده از سیگنال اصلی خطای ناشی از پیش‌بینی را ارسال می‌کنیم:

$$e_n = f_n - \hat{f}_n$$



$$\hat{f}_n = \text{function_of}(\tilde{f}_{n-1}, \tilde{f}_{n-2}, \tilde{f}_{n-3}, \ldots),$$

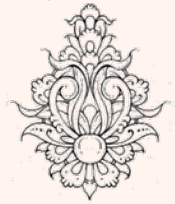
$$e_n = f_n - \hat{f}_n,$$

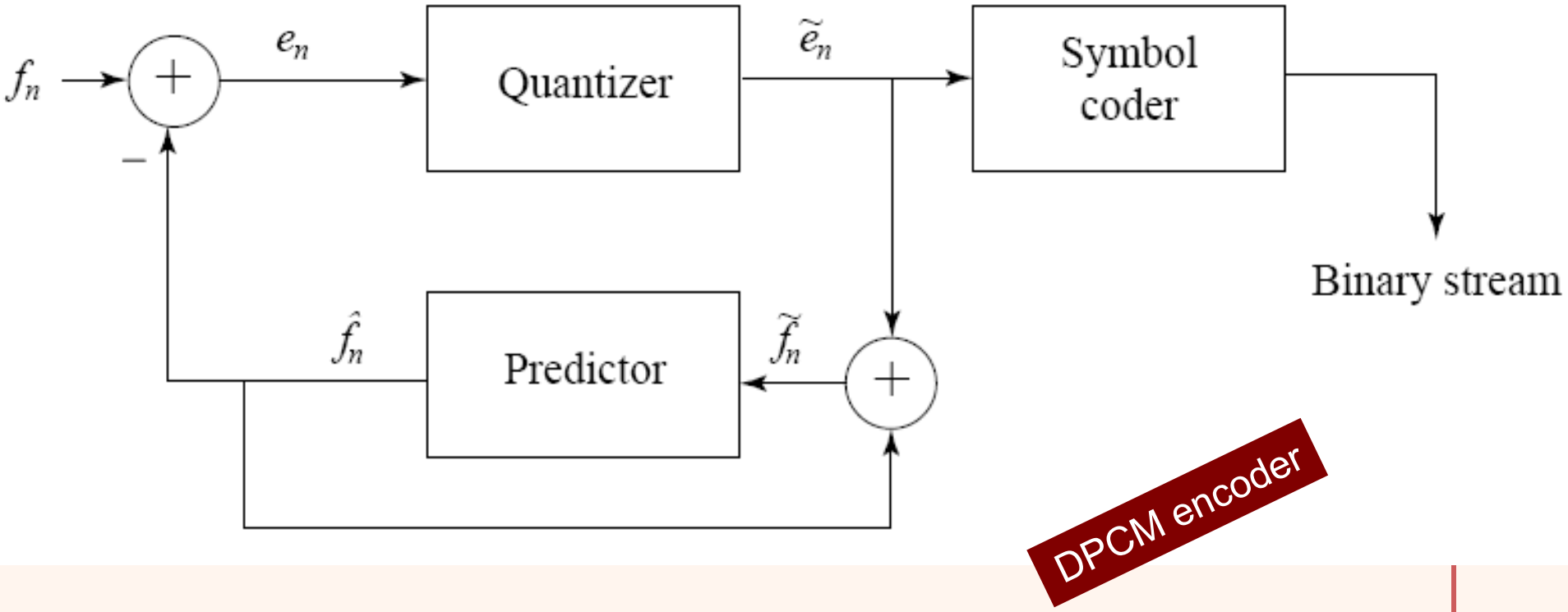
$$\tilde{e}_n = Q[e_n],$$

چندی کردن
خطای پیش‌بینی

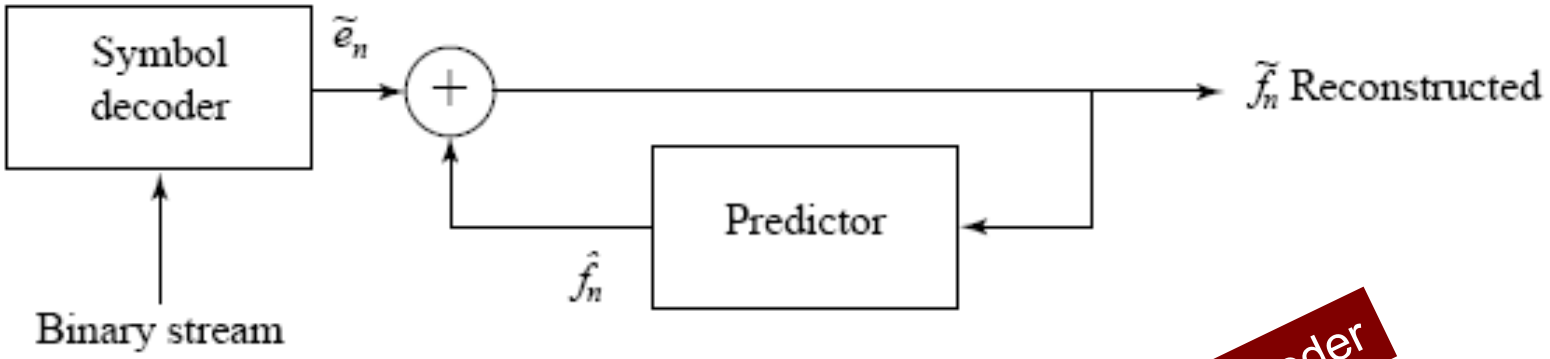
transmit $codeword(\tilde{e}_n)$,

reconstruct: $\tilde{f}_n = \hat{f}_n + \tilde{e}_n$.





DPCM encoder

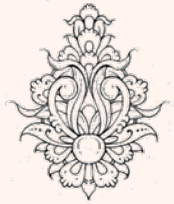


DPCM decoder

جدول چندی کردن

DPCM quantizer reconstruction levels.

e_n in range	Quantized to value
-255 .. -240	-248
-239 .. -224	-232
.	.
.	.
.	.
-31 .. -16	-24
-15 .. 0	-8
1 .. 16	8
17 .. 32	24
.	.
.	.
.	.
225 .. 240	232
241 .. 255	248



مثال

f_1	f_2	f_3	f_4	f_5
130	150	140	200	230

$(130+154)/2$

پیش‌بینی شده

$$\hat{f} = 130, \quad 130, \quad 142, \quad 144, \quad 167$$

$$e = \quad 0, \quad 20, \quad -2, \quad 56, \quad 63$$

$$\tilde{e} = \quad 0, \quad 24, \quad -8, \quad 56, \quad 56$$

$$\tilde{f} = 130, \quad 154, \quad 134, \quad 200, \quad 223$$

دوباره‌سازی شده