به نام خدا

دانشگاه صنعتی اصفهان

دانشکده مهندسی برق و کامپیوتر

گزارش آزمایش شماره 3

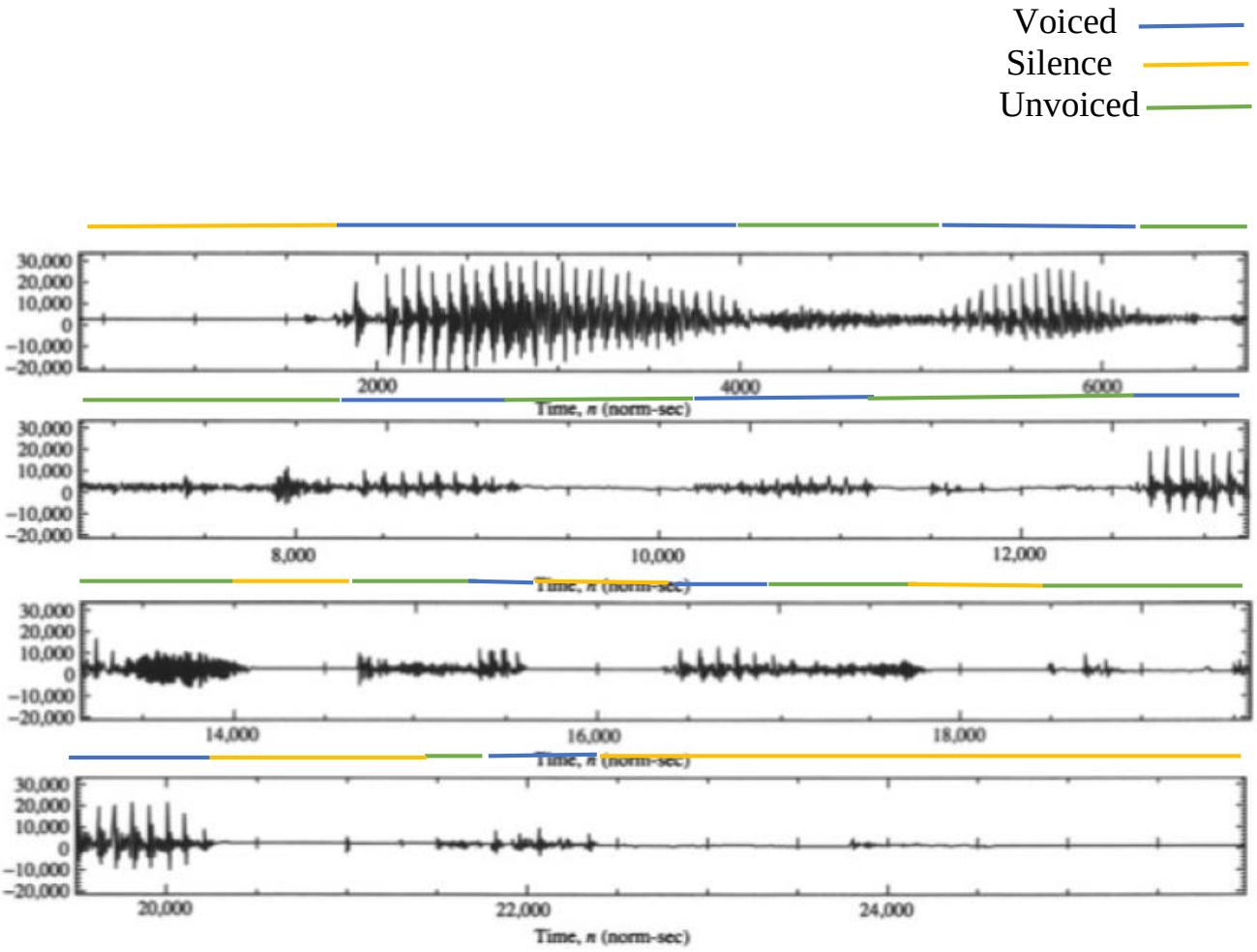
آزمایشگاه پردازش سیگنال های دیجیتال

عنوان:

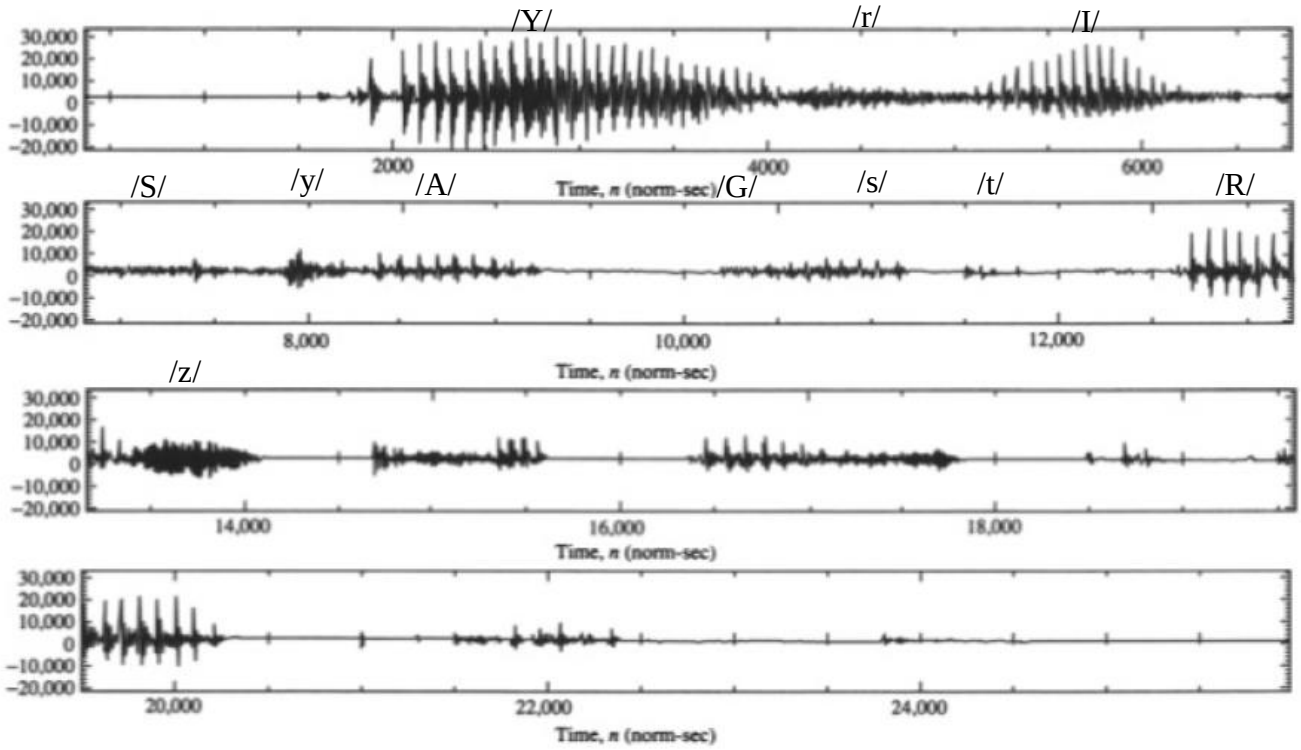
آشنایی با سیگنال صحبت

سوال 1)

الف) قسمت های مختلف را به شکل زیر برچسب‌گذاری می‌کنیم :



ب)



ج) برای هر کدام از واکه ها خواهیم داشت :

/Y/ : 2100-2160 → n = 2160-2100 = 60 → n = 60 samples → $f_0 = \frac{8000}{60} = 133.33\ Hz$

/I/ : 5490-5560 → n = 5560-5490 = 70 → n = 70 samples → $f_0 = \frac{8000}{70} = 114.28\ Hz$

/A/ : 8500-8590 → n = 8590-8500 = 90 → n = 90 samples → $f_0 = \frac{8000}{90} = 88.89\ Hz$

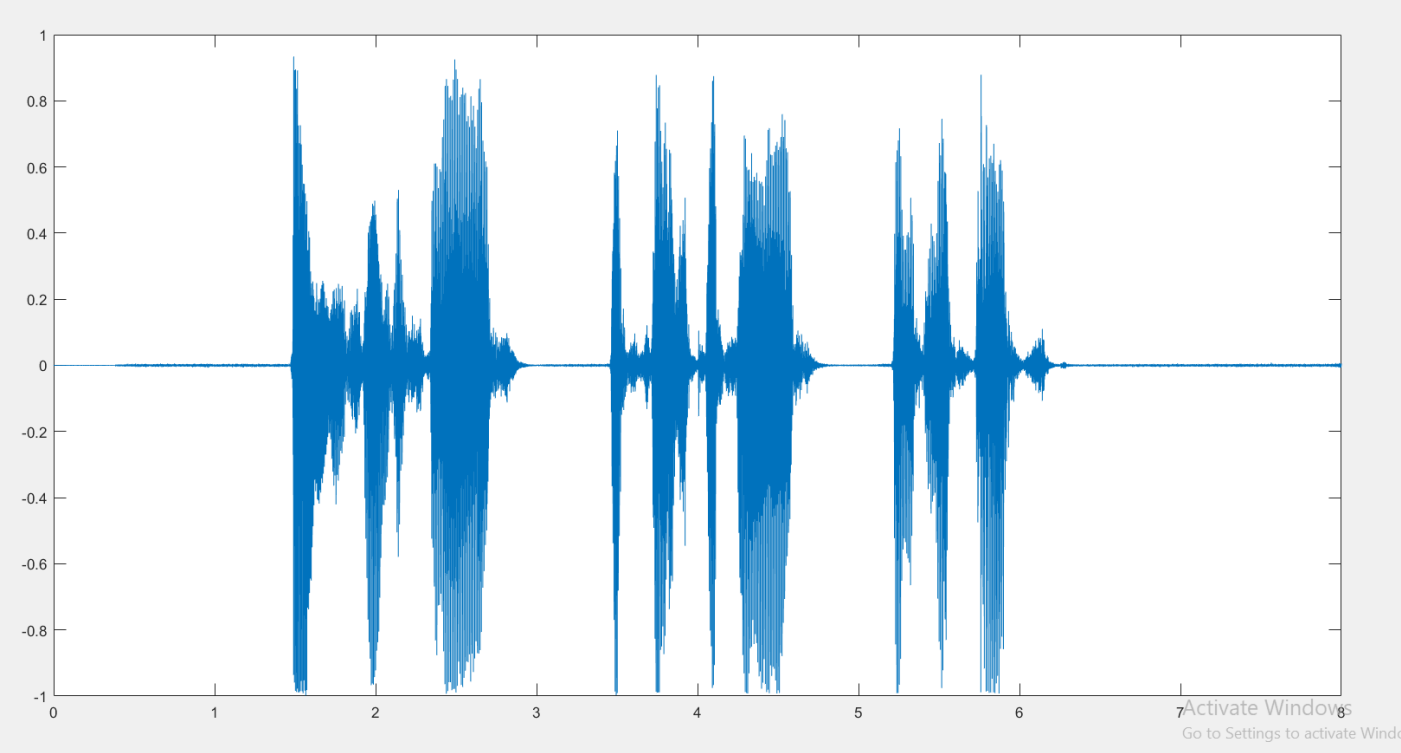
/R/ : 12850-12890 → n = 12890-12850 = 40 → n = 40 samples → $f_0 = \frac{8000}{40} = 200\ Hz$

$$f_{0,avg} = \frac{133.33 + 114.28 + 88.89 + 200}{4} = 134.125\ Hz$$

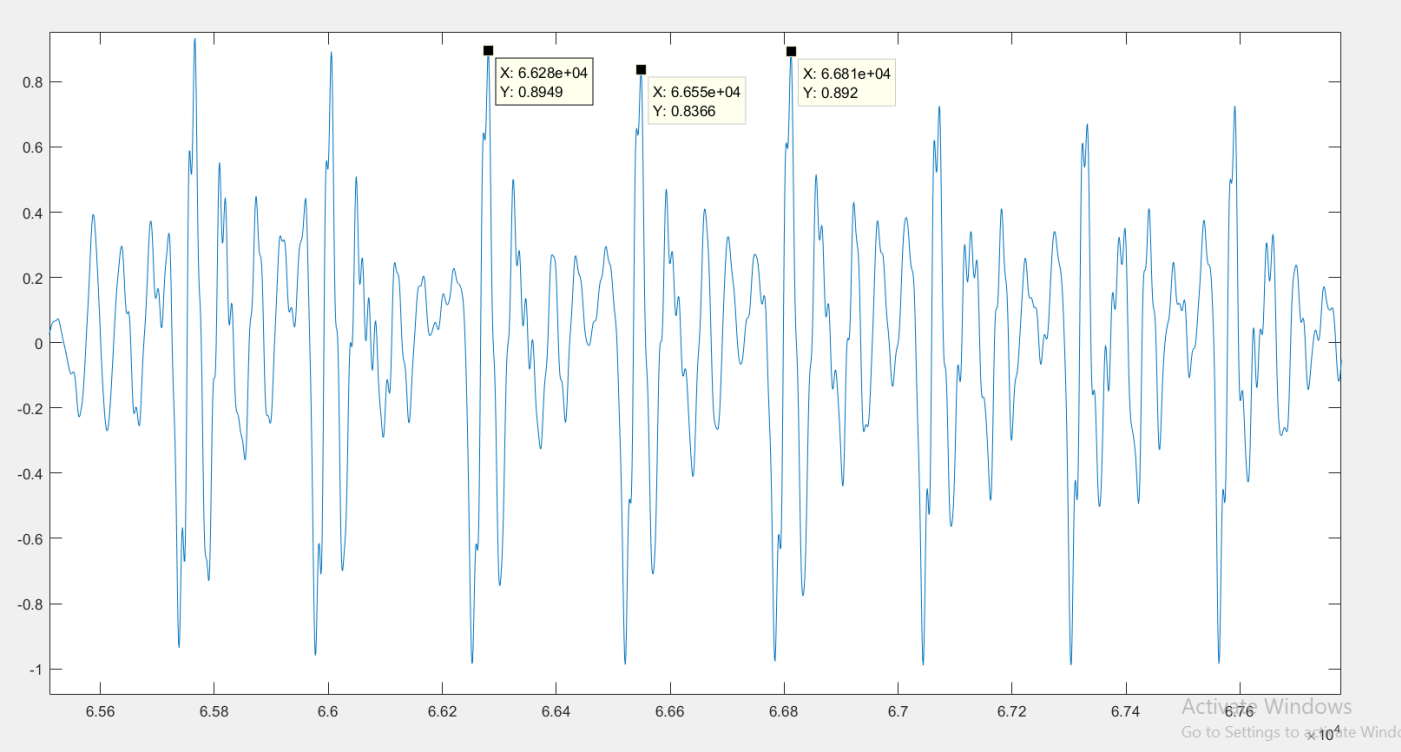
د) با توجه به عدد بدست آمده می‌توان نتیجه گرفت گوینده مرد بوده است.

ه) محاسبات بالا بر اساس همین خواسته مسئله بوده لذا می‌توان بازه مورد نظر برای صدا را 200 – 88.89 خواهیم داشت که بسیار به بازه استاندارد مربوط به صدای مرد نزدیک است.

و) شکل سیگنال تلفظ شده توسط خودمان :

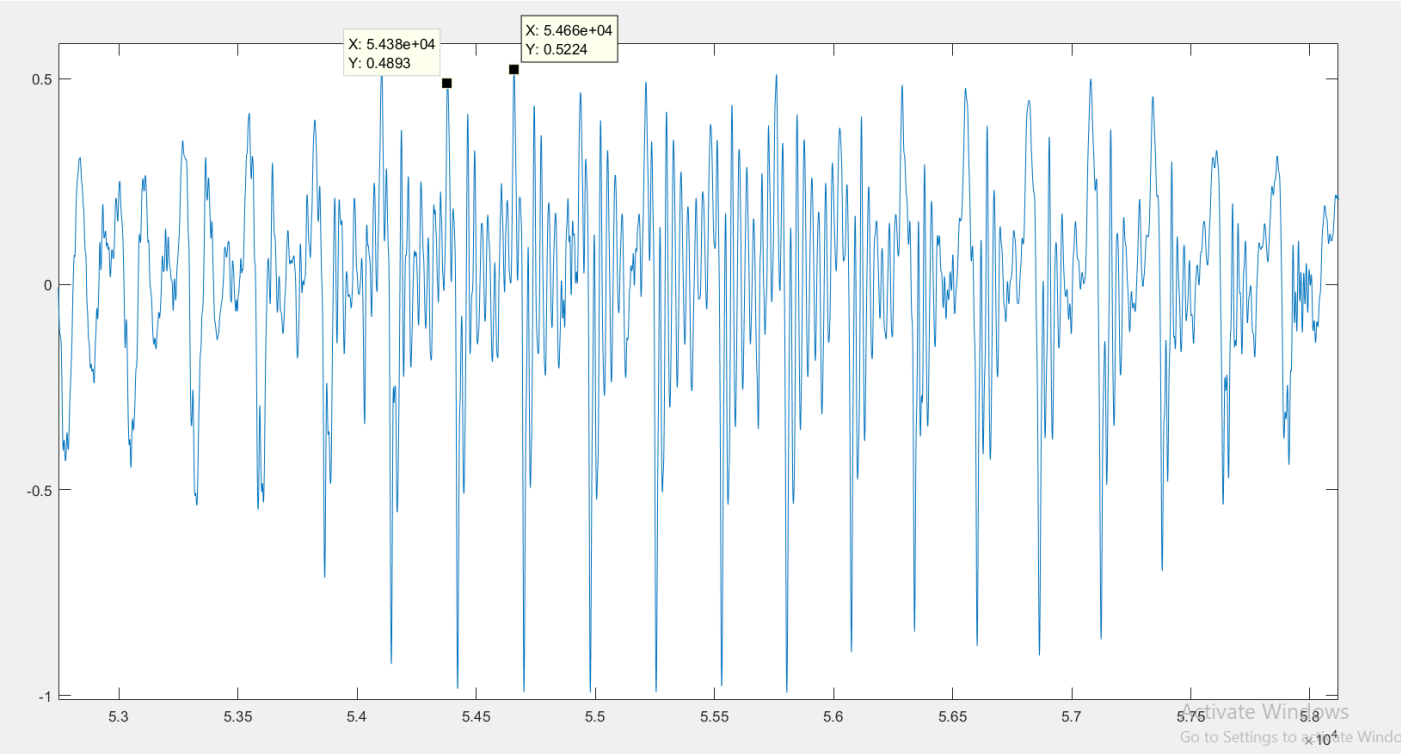


واکه /Y/ :



$$f_0 = \frac{F_s}{samples\ per\ period} = \frac{44100}{(66810 - 66550)} = \frac{44100}{260} = 169.615$$

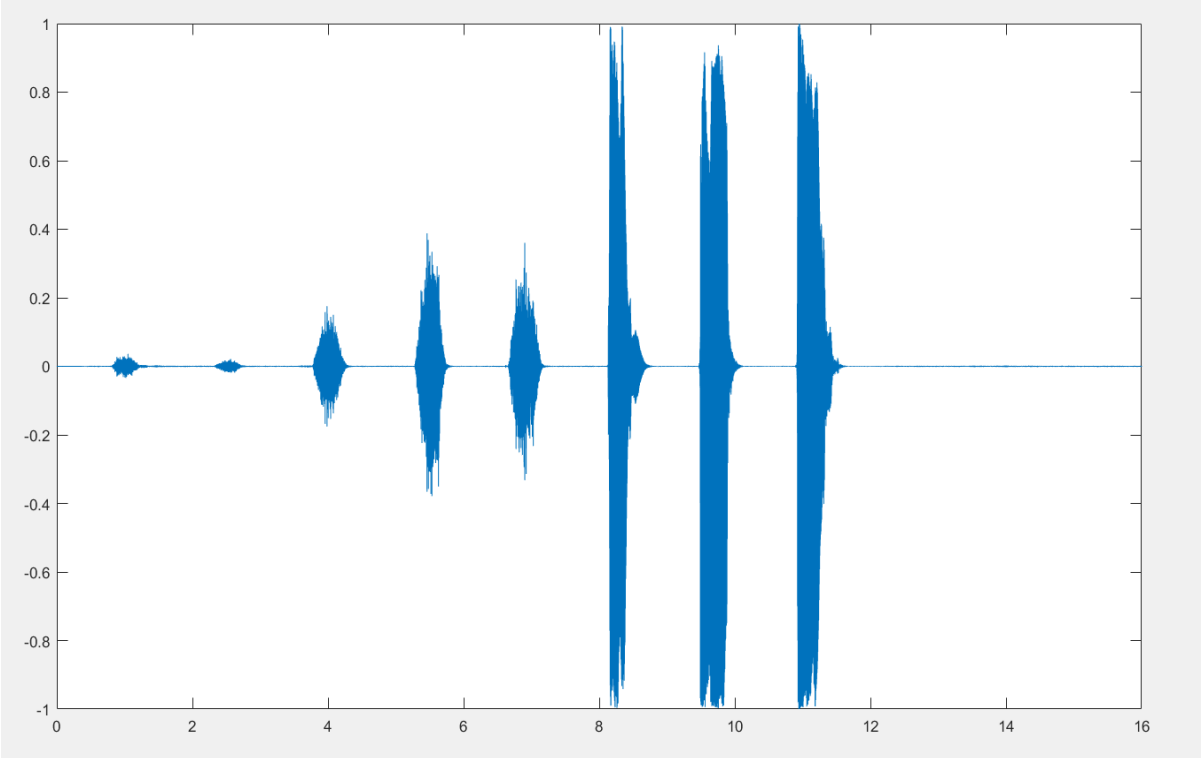
واکه /I/ :



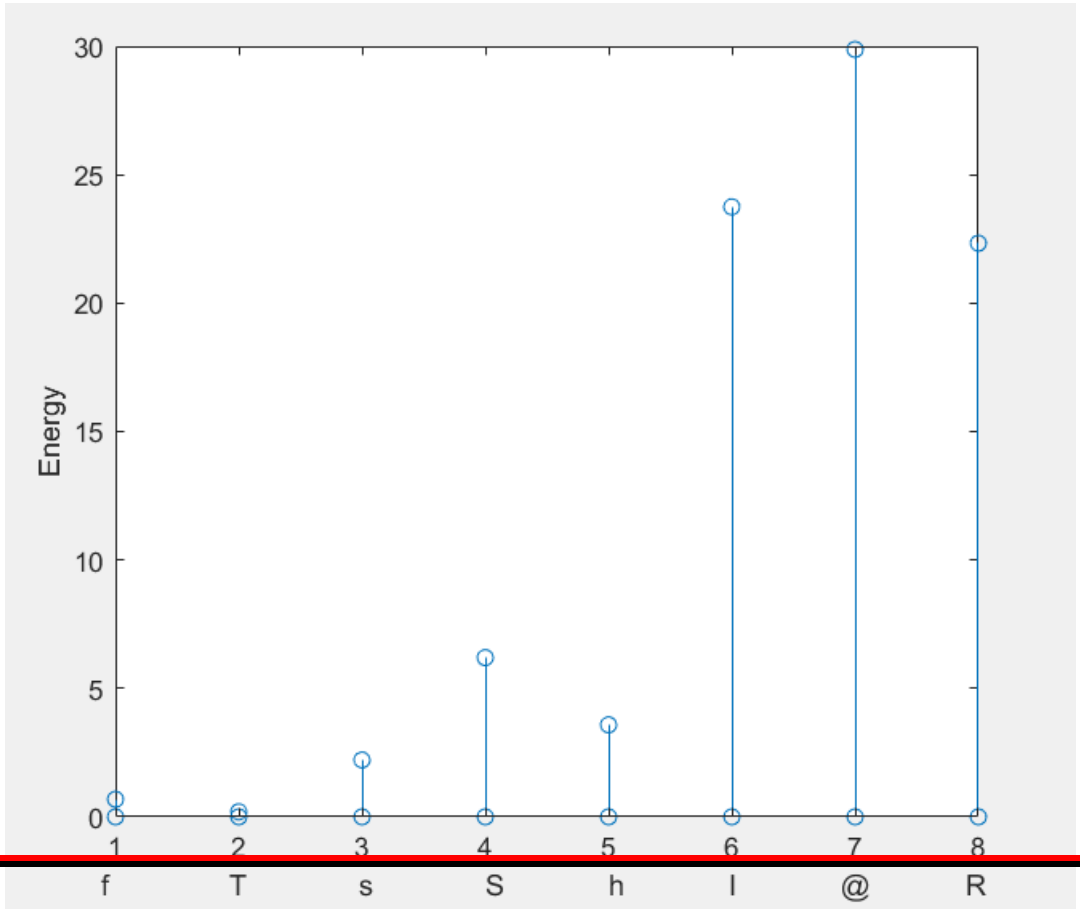
$$f_0 = \frac{F_s}{samples\ per\ period} = \frac{44100}{(54660 - 54380)} = \frac{44100}{280} = 157.5$$

همان طور که مشاهده می‌شود میانگین pitch ها در دو حالت تقریبا یکسان بدست می‌آید چون هر دو گوینده مرد می‌باشند.

سوال 2) شکل حروف ضبط شده در حوزه زمان :



انرژی حروف تلفظ شده را در شکل زیر مشاهده می‌کنیم :

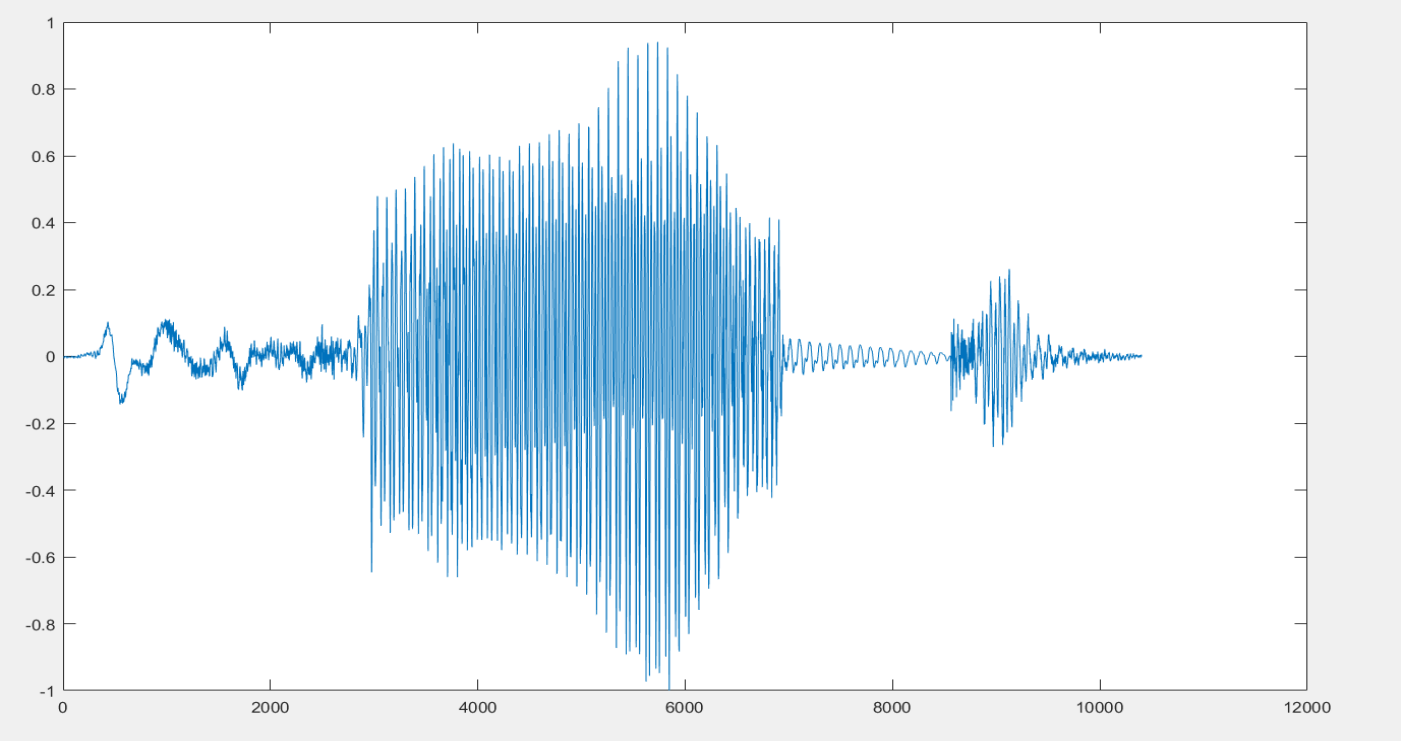


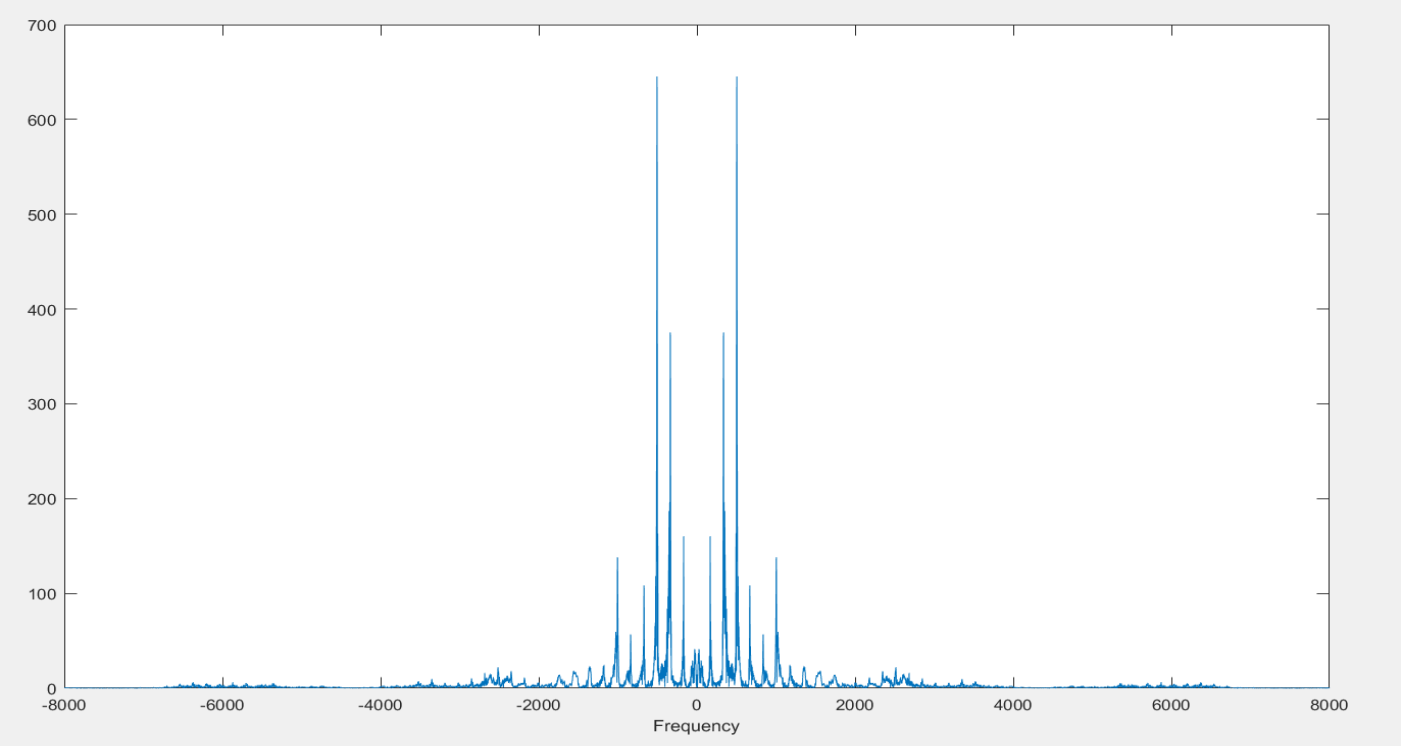
مطابق انتظار انرژی حروف سایشی نسبت به واکه ها همواره کمتر است چون که منشا تولید واکه قطار ضربه اما حروف سایشی معمولا با هوای داخل دهان (نویز) تولید می شوند که در این بین هم آن هایی که با لب و دندان (قسمت جلویی دهان مثل f,th) نیز انرژی به مراتب کمتری نسبت به آنهایی که با قسمت میانی دهان مثل sh,h تلفظ می‌شوند ؛ دارند.

لازم به ذکر است فایل Q2.m در کدهای متلب مربوط به ضبط صدا و پخش صوت ضبط شده به همراه شکل سیگنال در حوزه زمان می‌باشد و فایل Q2_energy.m به محاسبه انرژی هر کدام از صداها پرداخته است.

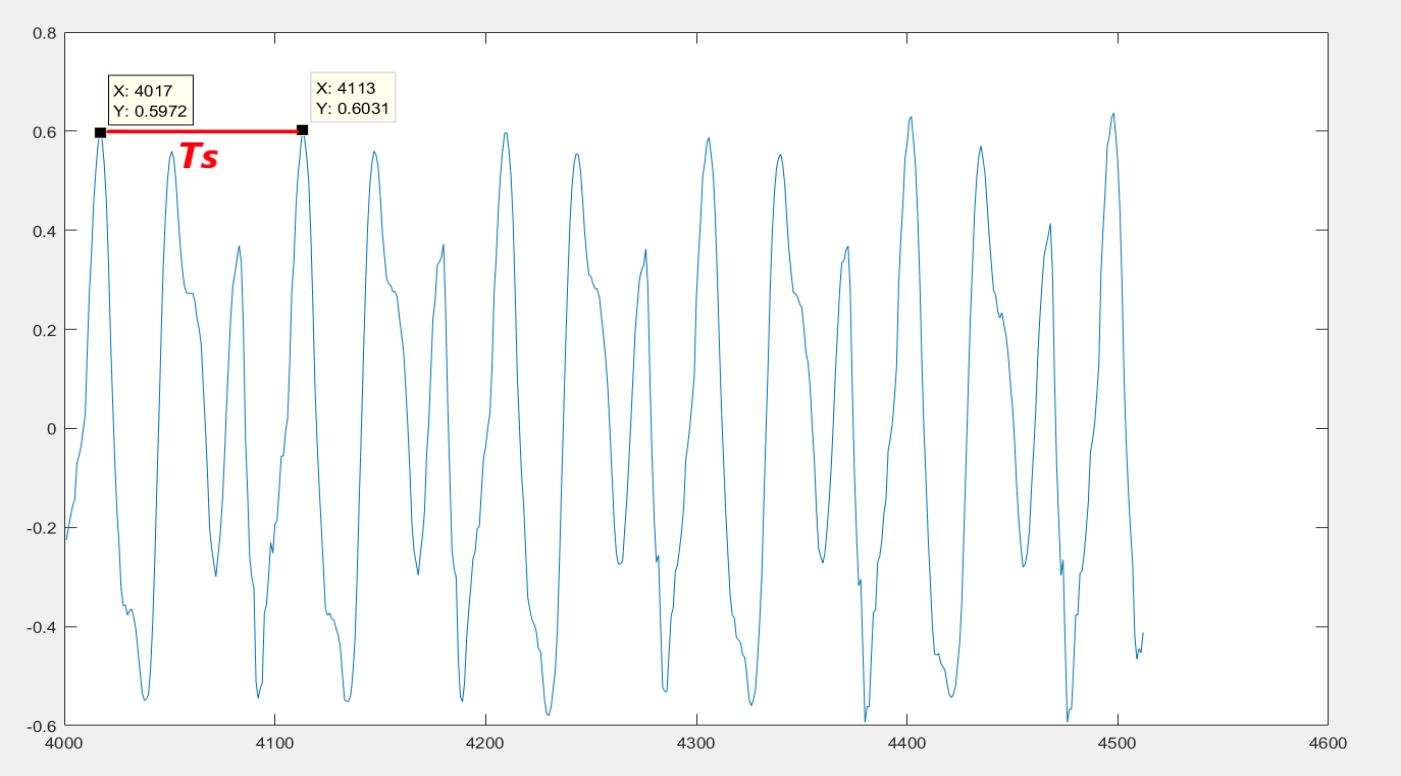
سوال 3) در این قسمت برای نمونه سیگنال n01oo.wav را به طور کامل با رسم شکل و غیره بررسی می کنیم سپس برای بقیه سیگنال های این فایل نیز مشابه این کارها صورت گرفته و در جدولی بیان خواهد شد.

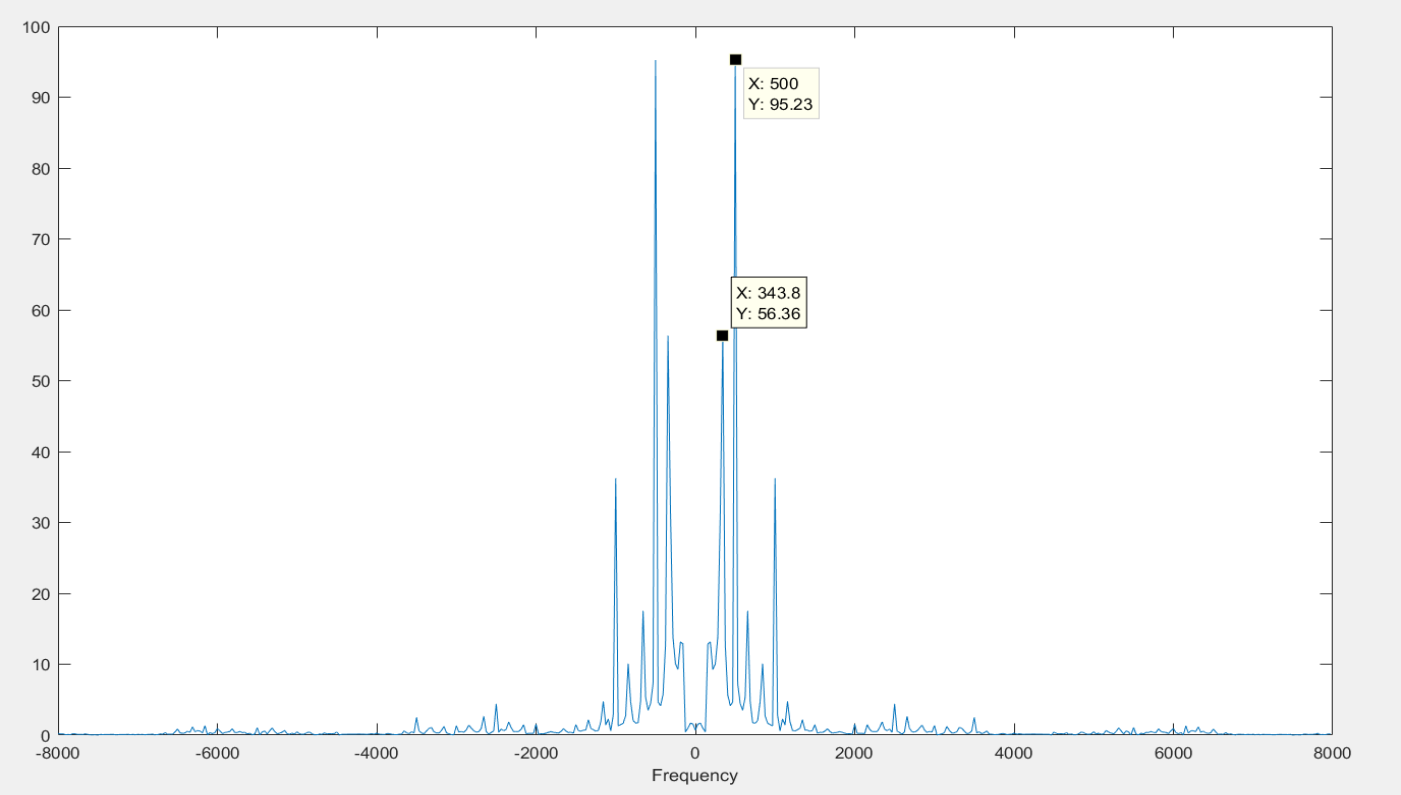
شکل سیگنال مورد نظر در حوزه زمان و فرکانس :





حال برای آنکه قسمت واکه مربوطه (**oo**\) را مشاهده کنیم کافی است بازه انتخابی را بین 4000 تا 4512 امین نمونه قرار دهیم (مطابق صورت سوال می‌بایست بازه ای به طول 512 از این سیگنال را انتخاب نمود.)





مطابق انتظار اگر پنجره 512 نقطه ای را به صورت کامل در واکه مربوطه انتخاب کنیم سیگنال حوزه زمانی پریودیک می باشد و به کمک همین دوره تناوب یا دو پیک متوالی در حوزه فرکانس می‌توان pitch را محاسبه نمود. برای مثال در مورد بالا مقدار pitch از روی تبدیل فوریه سیگنال برابر با 500 – 343.8 = 156.2 است و به کمک سیگنال حوزه زمان خواهیم داشت :

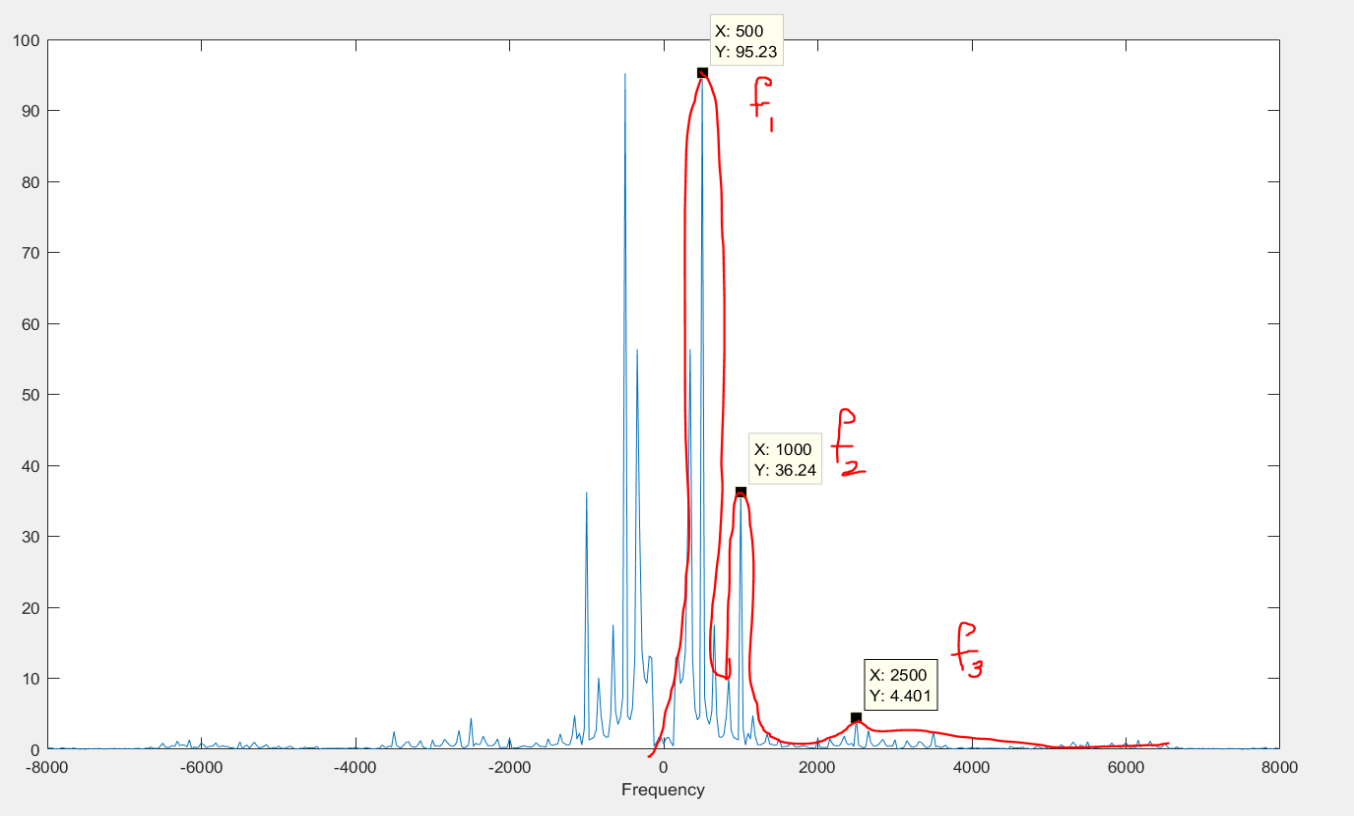
$$4113 - 4017 = 96\ Samples$$

$$\mathrm{Hz} f_0 = \frac{16000}{96} = 166.67$$

با توجه به اعداد به دست آمده و رنج pitch صدای زن (pitch = 120-500Hz) و مرد (pitch = 50-250Hz) می‌توان نتیجه گرفت که گوینده مرد می باشد.

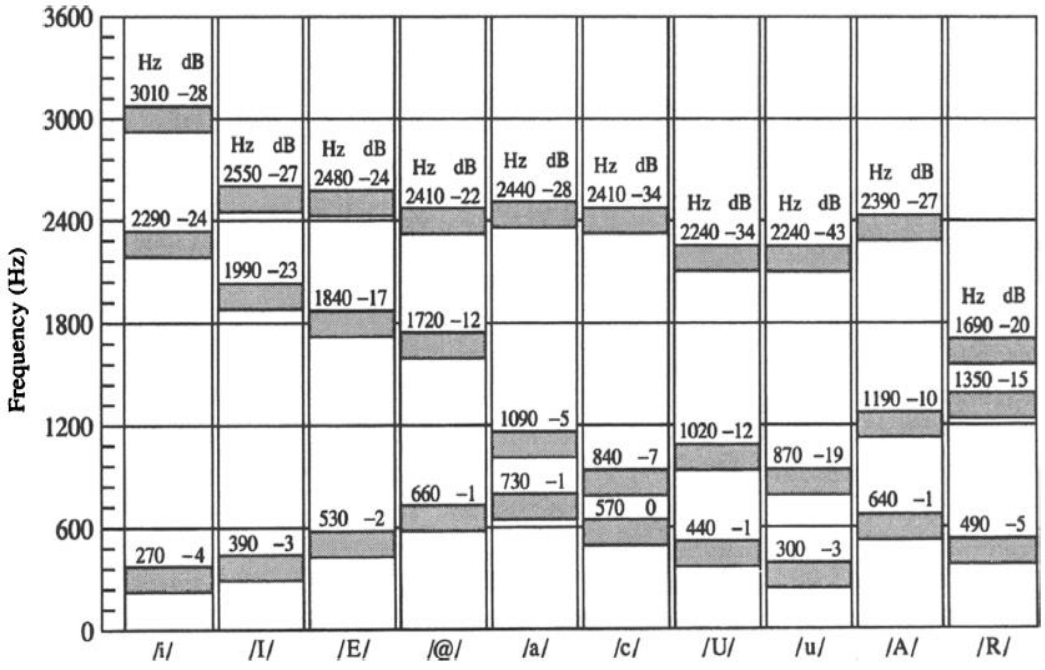
سیگنال ها در حوزه زمان مشخصه ندارند و فقط متناوب هستند.

هم چنین در شکل زیر فرمنت مربوط به این واکه مشخص شده است :



مطابق شکل بالا $f_3 = 2500$ و $f_2 = 1000$ و $f_1 = 500$. این اعداد بسیار به واکه \oo استاندارد نیز نزدیک اند.

جدول مربوط به فرمنت های تمام واکه ها :



روند فوق را برای بقیه واکه های oo\ نیز تکرار می کنیم. نتایج مربوط به این قسمت در جدول زیر قابل مشاهده است :

اسم فایل	جنسیت	pitch	فرمنت
n02oo	Male	109.35	437.5-1094
n06oo	Male	125	468.8-1188
n13oo	Male	125	437.5-1125
n14oo	Male	125	437.5-1344
t13oo	Female	250	531.3-1063
t14oo	Female	250	468.8-1438
t15oo	Female	250	687.5-1594
y01oo	Female	250	718.8-1688
y05oo	Female	218.7	687.5-1594
y10oo	Female	187.5	625-1281
y25oo	Female	250	468.8-1406
d05oo	Female	250	500 - ×
d11oo	Female	218.8	593.8-1563
d12oo	Female	218.7	437.5-1531
d13oo	Female	250	468.8-1656
d20oo	Female	218.7	437.5-×
d21oo	Female	250	468.8-×
d25oo	Female	218.8	625-1719
d26oo	Female	187.5	562.5-1313

اسم فایل	فرمنت	حدس به کمک جدول
1	843.8-1469	/a/
2	500-1000	/U/
3	500-2781	/i/
4	500-1281	هیچ کدام (با توجه به واکه های موجود)
5	406.3-1031	/U
6	531.3-2156	/E
7	375-781.3	/c/
8	437.5-875	/c/
9	500-1219	/U/
10	437.5-2844	/i/
11	437.5-1536-1938	/R/
12	468.8-2625	/i/
13	343.8-1156	/U/
14	625-1375	/a/
15	437.5-1344	/U/
16	468.8-1063	/U/
17	250-2031	/i/
18	531.3-1813-2469	/I/
19	468.8-1250	/U/
20	468.8-2125-2719	/I/
21	750-1625	/@/
23	656.3-937.5	/a/
24	812.5-1219	/a/
25	687.5-2031	/E/
26	437.5-1563	/U/
27	468.8-937.5-1656	/a/
28	687.5-2063	هیچ کدام (با توجه به واکه های موجود)
29	437.5-1125	/u/
30	281.3-3000	/i/
31	500-2688	/E/
32	406.3-1594	/R/
33	500-2781	/E/
34	718.8-2375	/E/
35	468.8-937.5-2781	/a/
36	437.5-1563	/@/
37	681.3-1875	/@/
38	437.5-×	/U/
39	750-1781	/a/

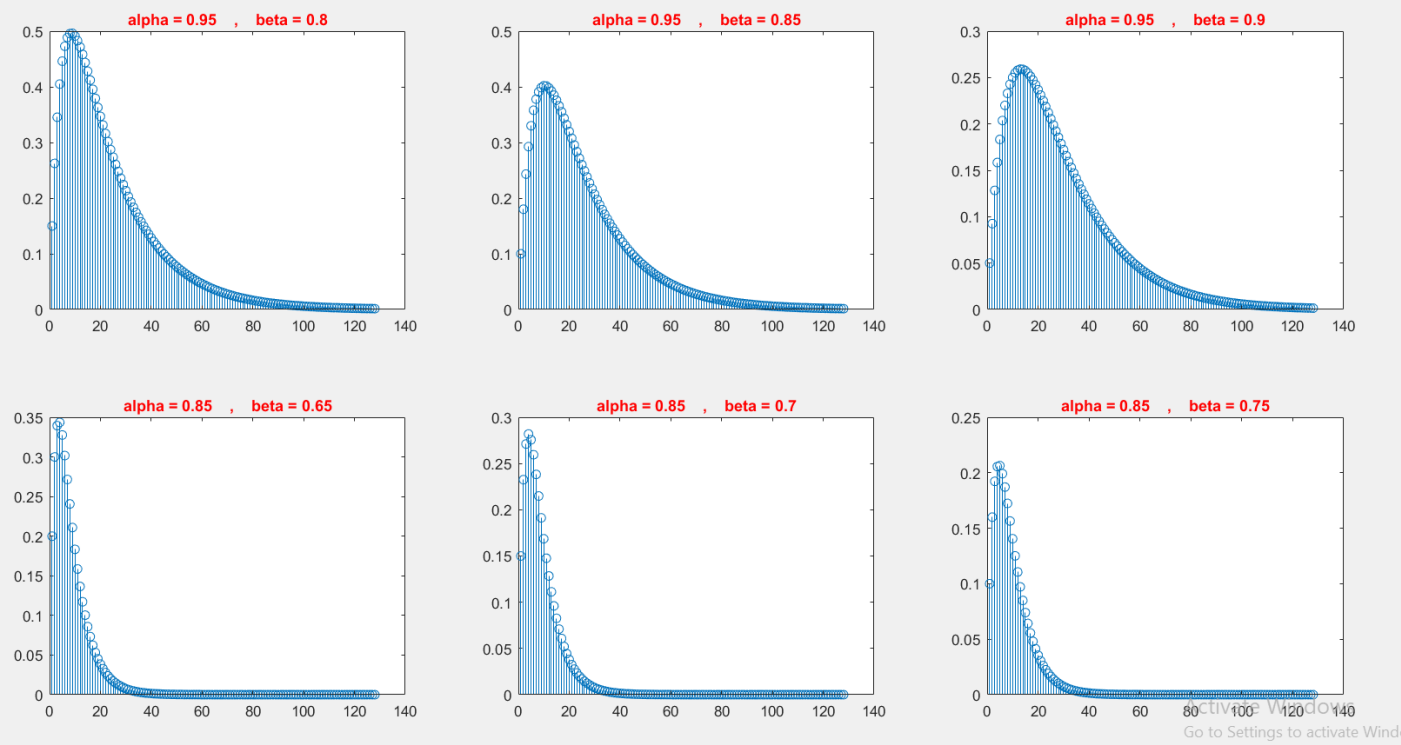
سوال 4)

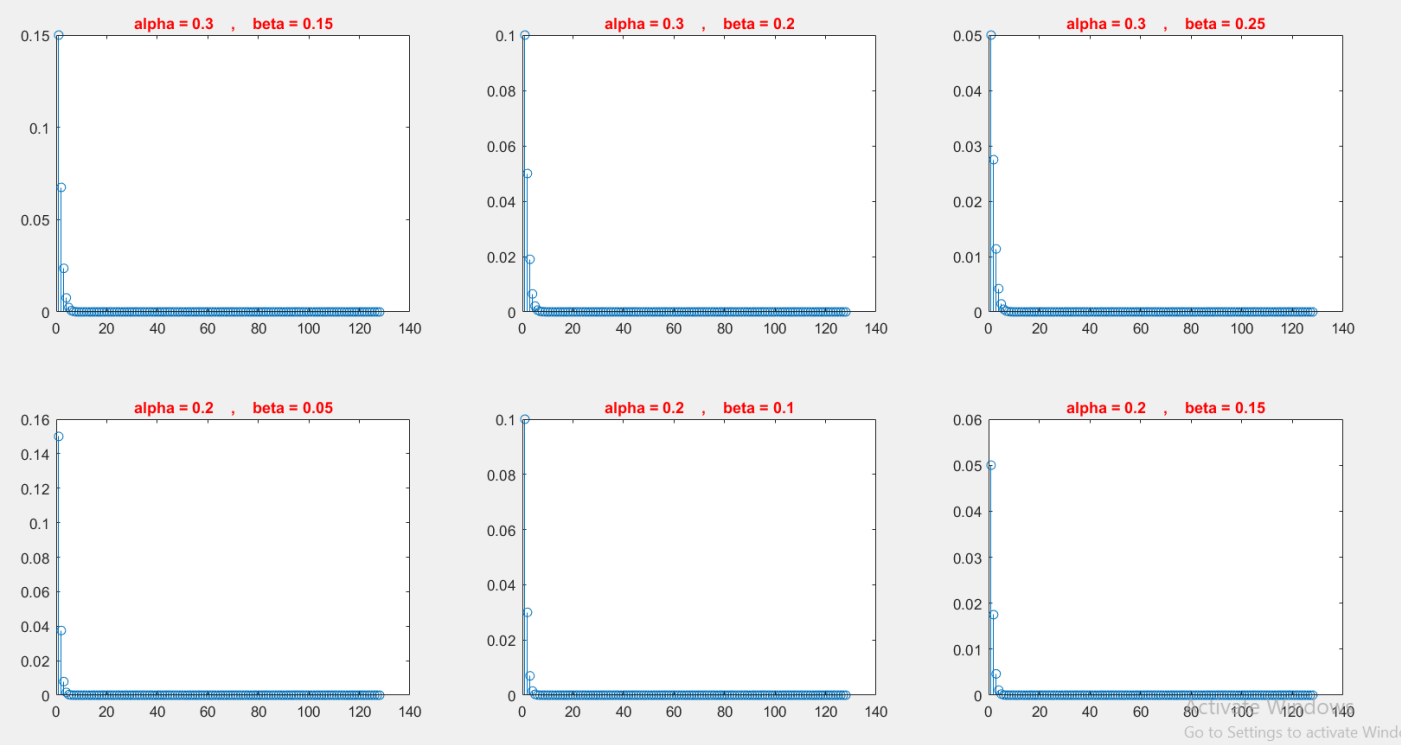
a) در تمامی شکل های زیر سیگنال حوزه زمان با در نظرگرفتن فرض‌های مسئله ؛ میرا می‌شود.

مطابق شکل های بالا برای آن که در لحظه n = 64 سیگنال زمانی برابر صفر شود می بایست آلفا کوچکتر از 0.95 باشد.

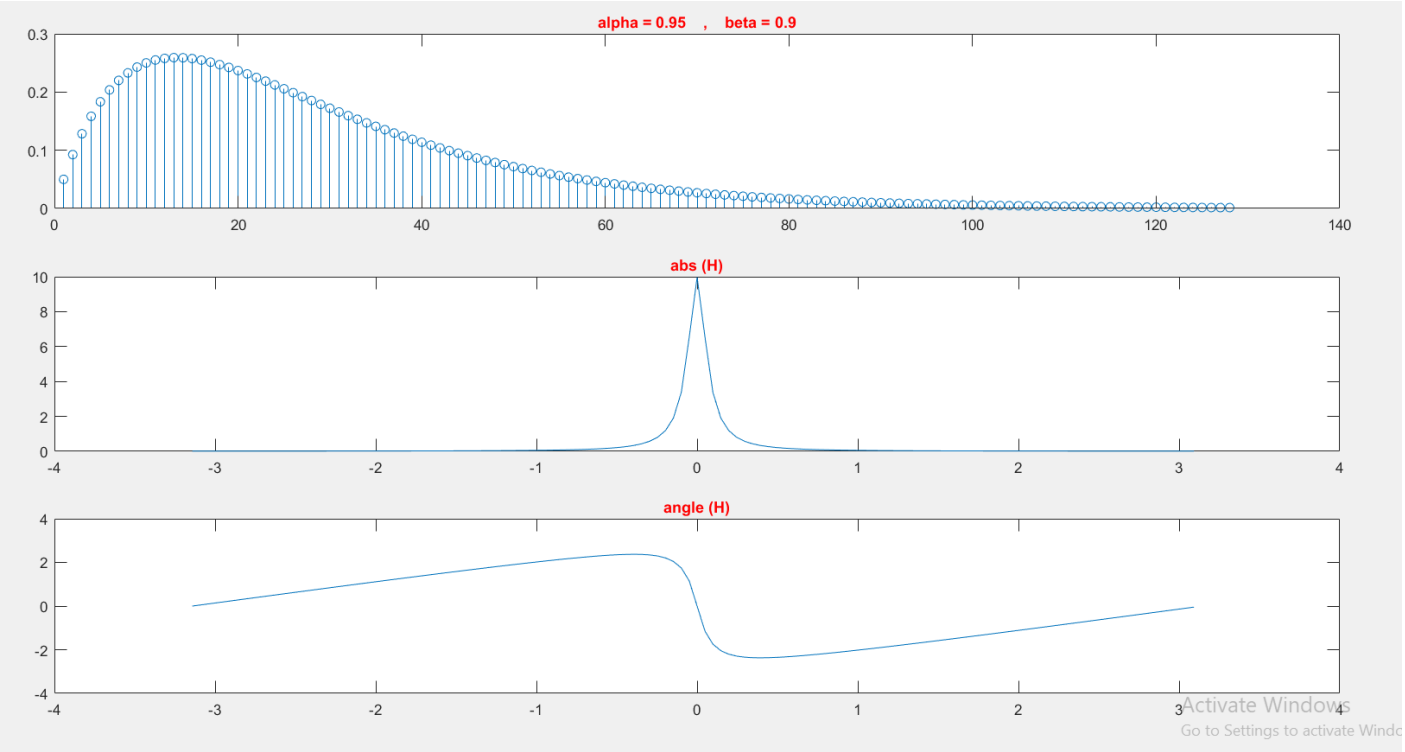
اما با توجه به صورت سوال که اشاره به نزدیکی مقدار آلفا به عدد 1 دارد ؛ مقادیر کوچکتر از 0.3 برای بتا مناسب است.

پیاده سازی glottal shaping filter [$g(n) = (\alpha^n - \beta^n)\, u(n)$] در حوزه زمان :

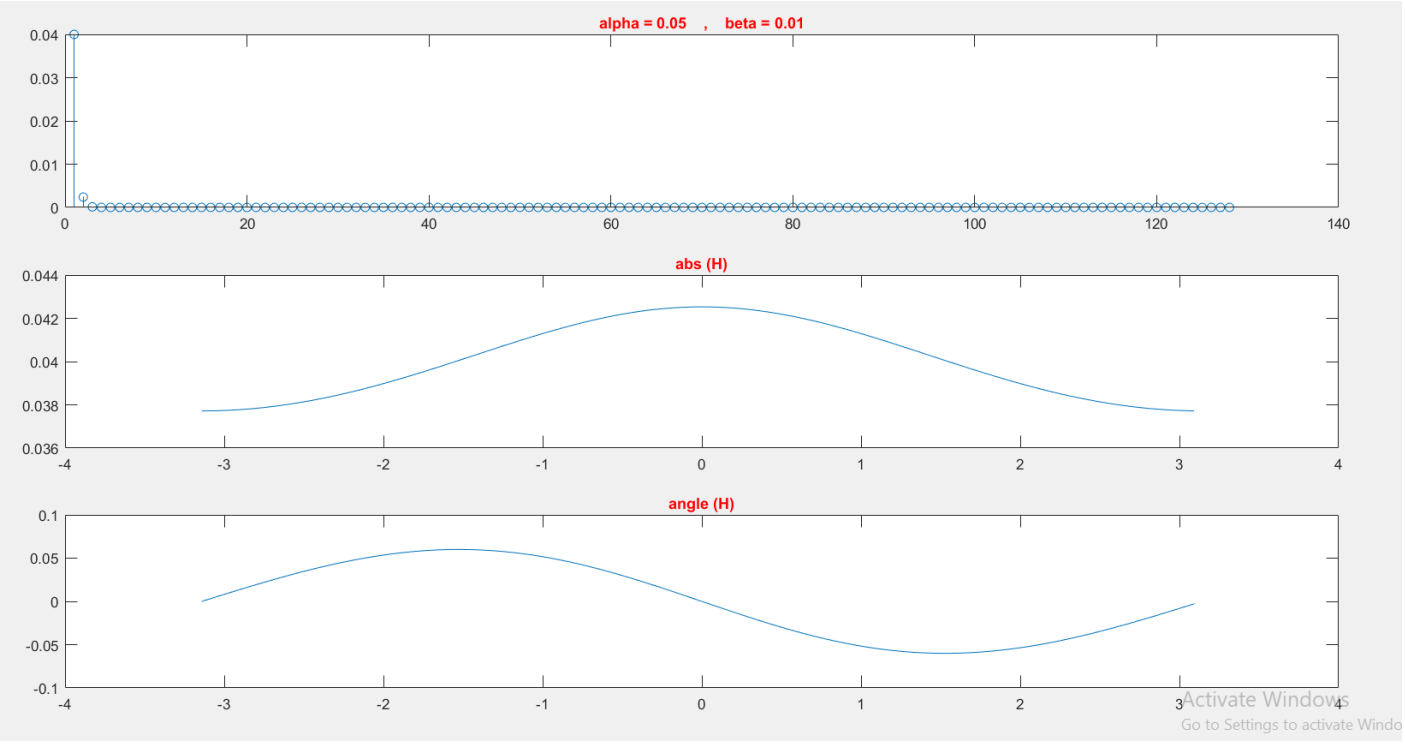




b) پیاده سازی glottal shaping filter [$g(n) = (\alpha^n - \beta^n)\, u(n)$] در حوزه فرکانس :



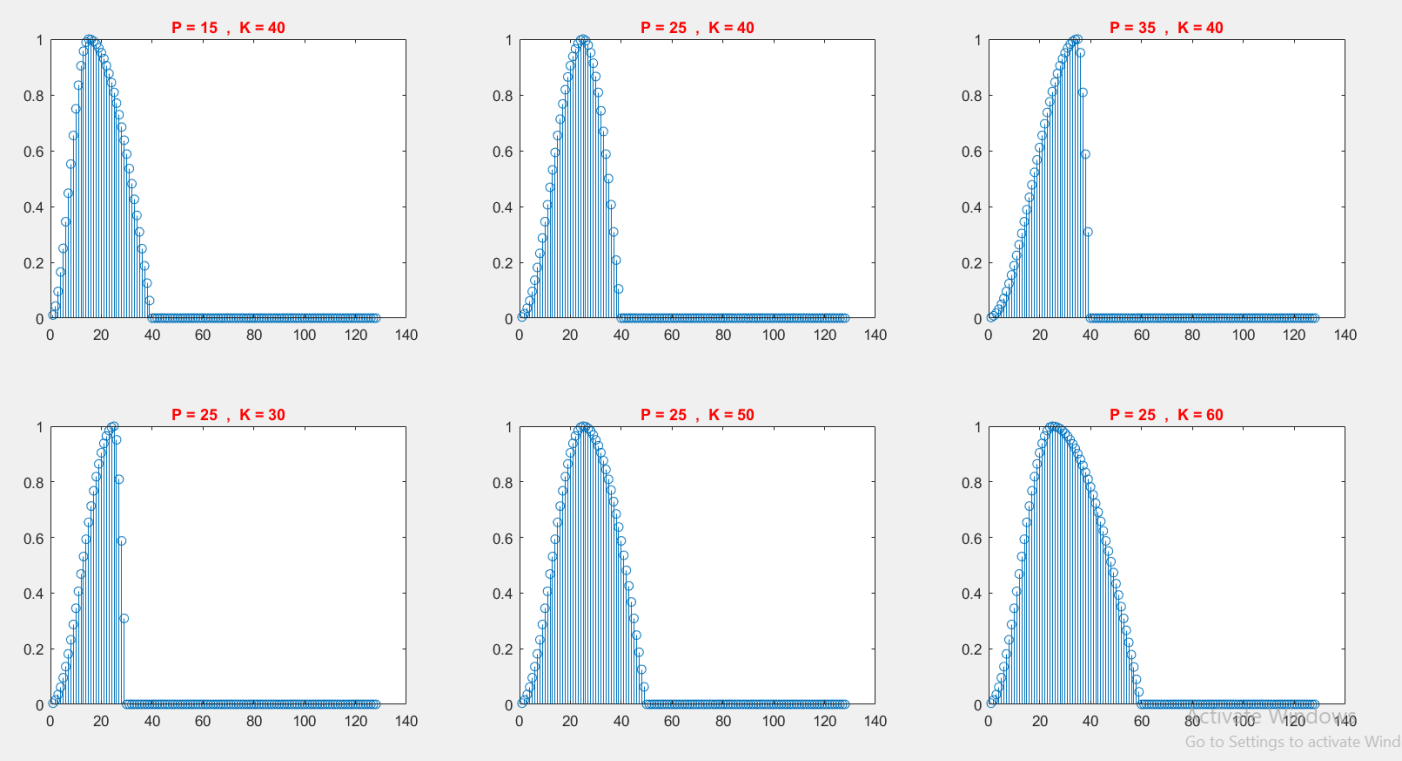
$\alpha = 0.95$ و $\beta = 0.9$

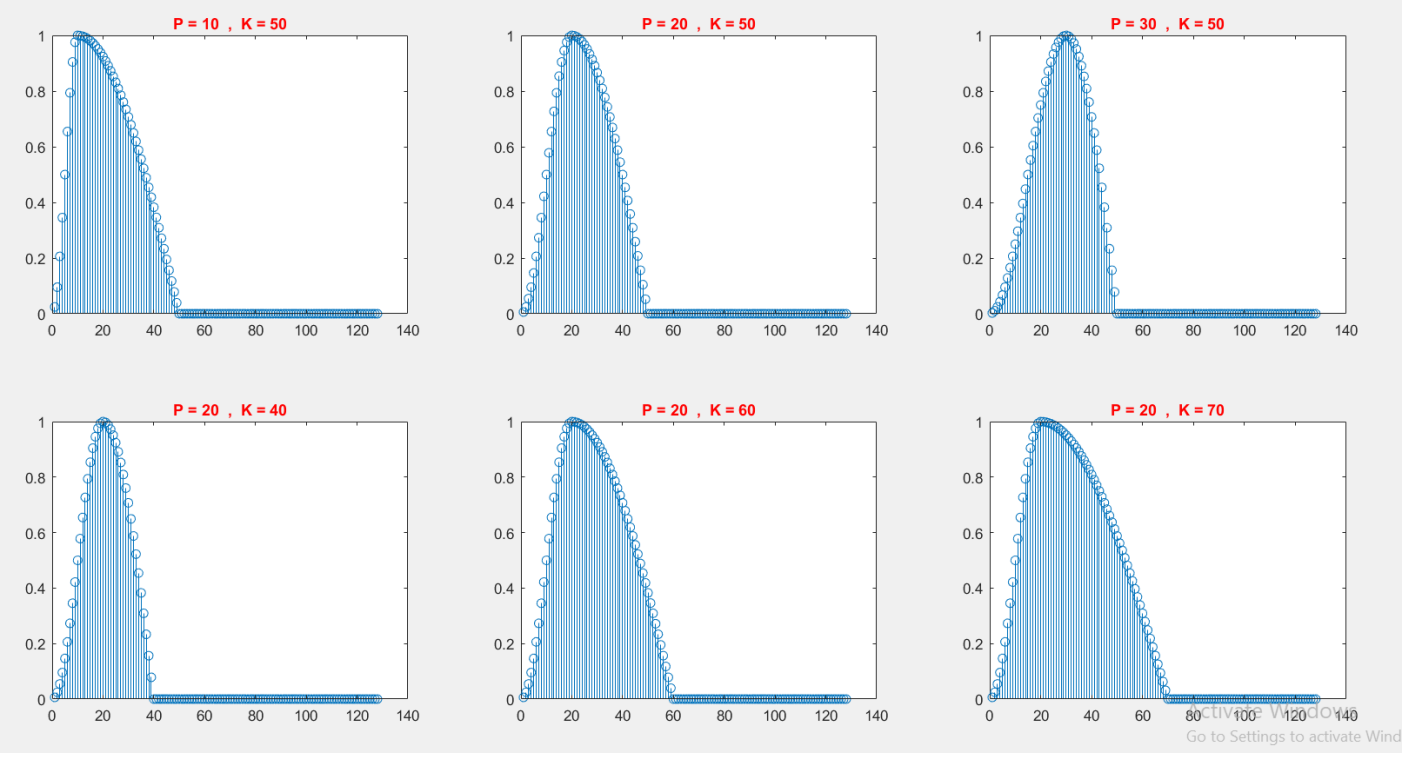


$\alpha = 0.05$ و $\beta = 0.01$

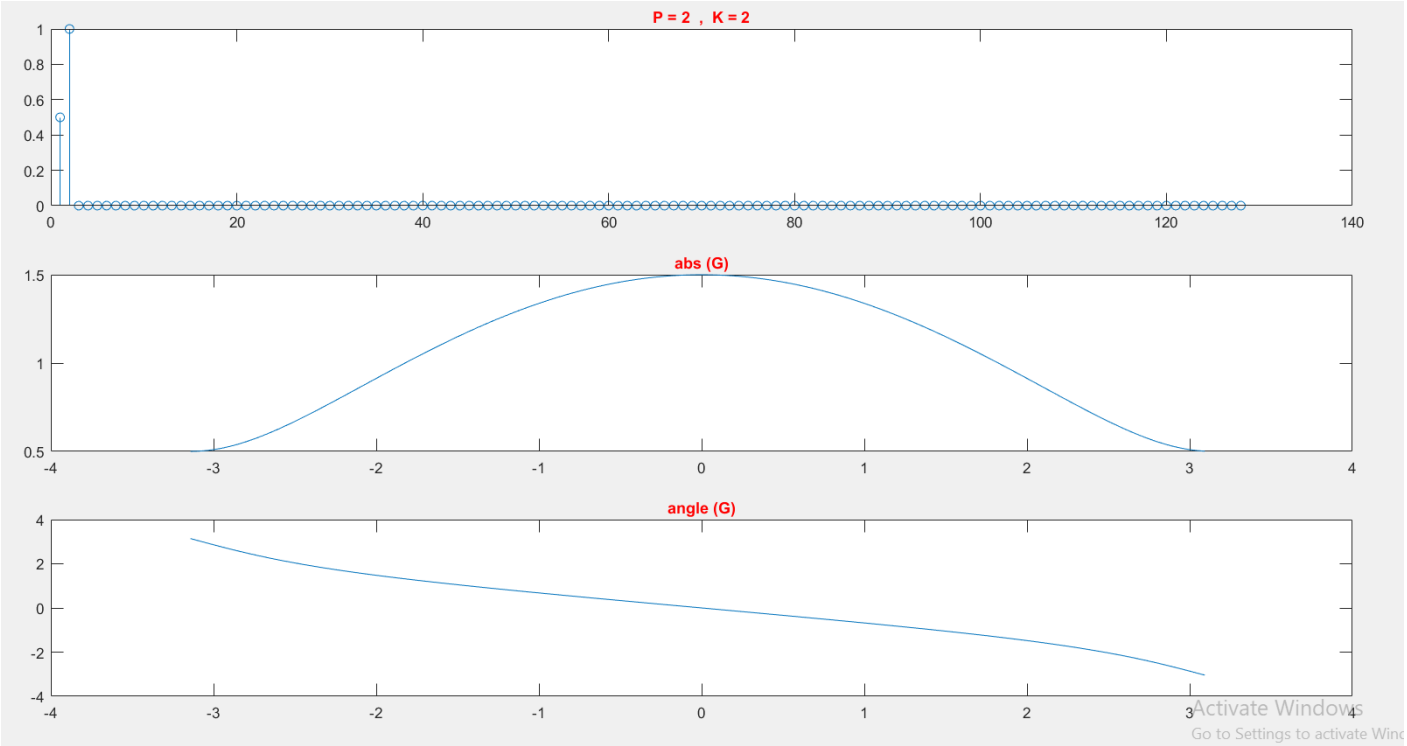
c) از آنجا که $P <= K$ و طبق ضابطه سوال برای مقادیر بزرگتر از K حاصل تابع برابر صفر می‌باشد لذا کافی است $K < 65$ انتخاب شود تا در لحظه $n = 64$ سیگنال زمانی برابر صفر شود.

پیاده سازی Rosenberg Pulse در حوزه زمان :

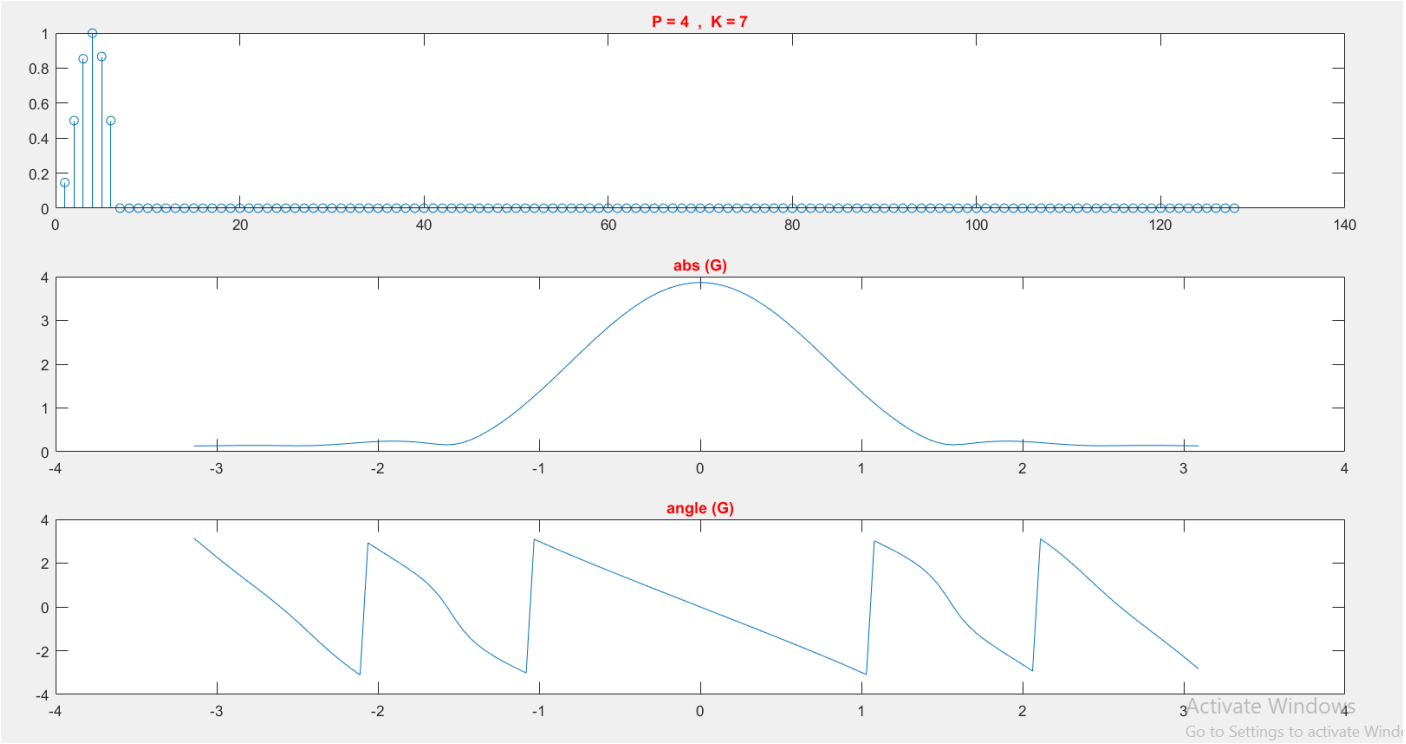




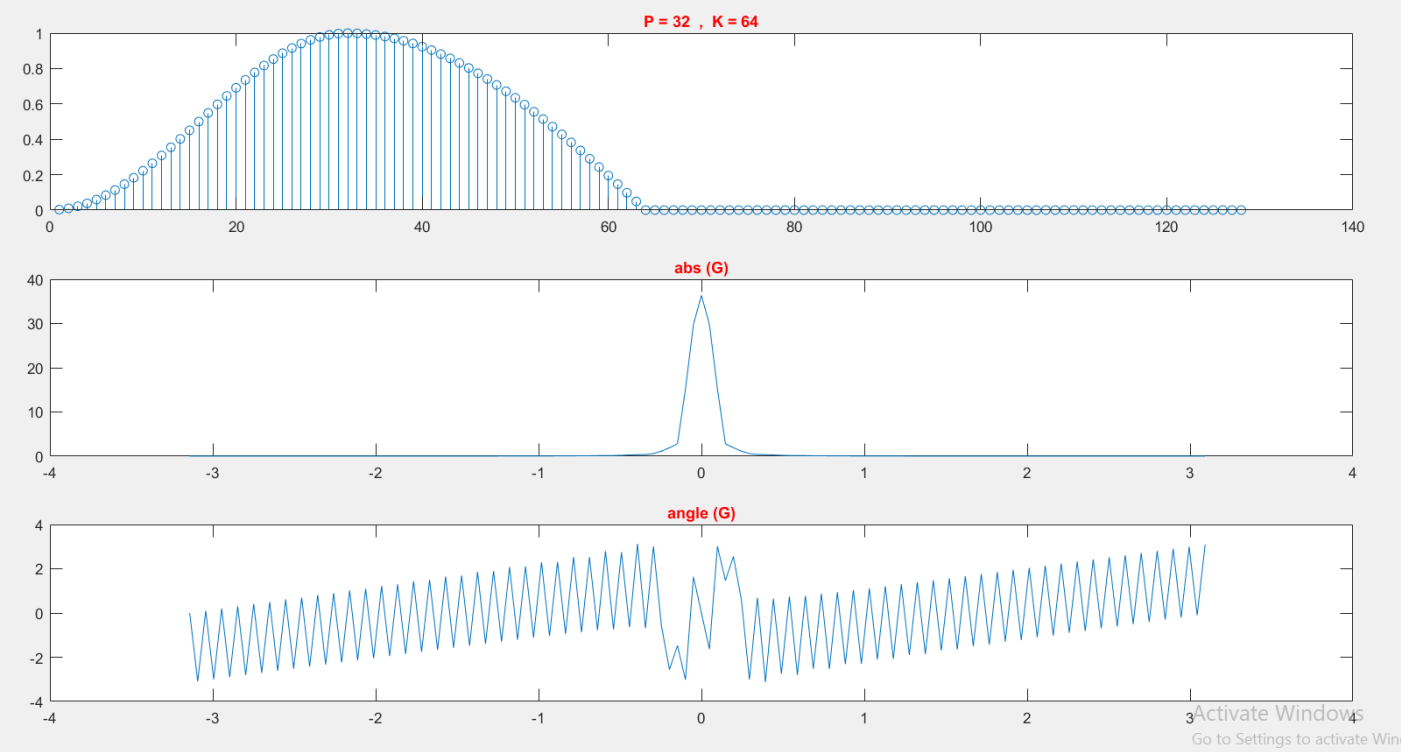
d) پیاده سازی Rosenberg Pulse در حوزه فرکانس :



$K = 2$ و $P = 2$



$K = 7$ و $P = 4$



$K = 64$ و $P = 32$

مقایسه ویژگی های زمانی و فرکانسی دو فیلتر بالا :

glottal shaping filter فیلتر IIR است درحالی که Rosenberg filter فیلتر FIR می‌باشد.

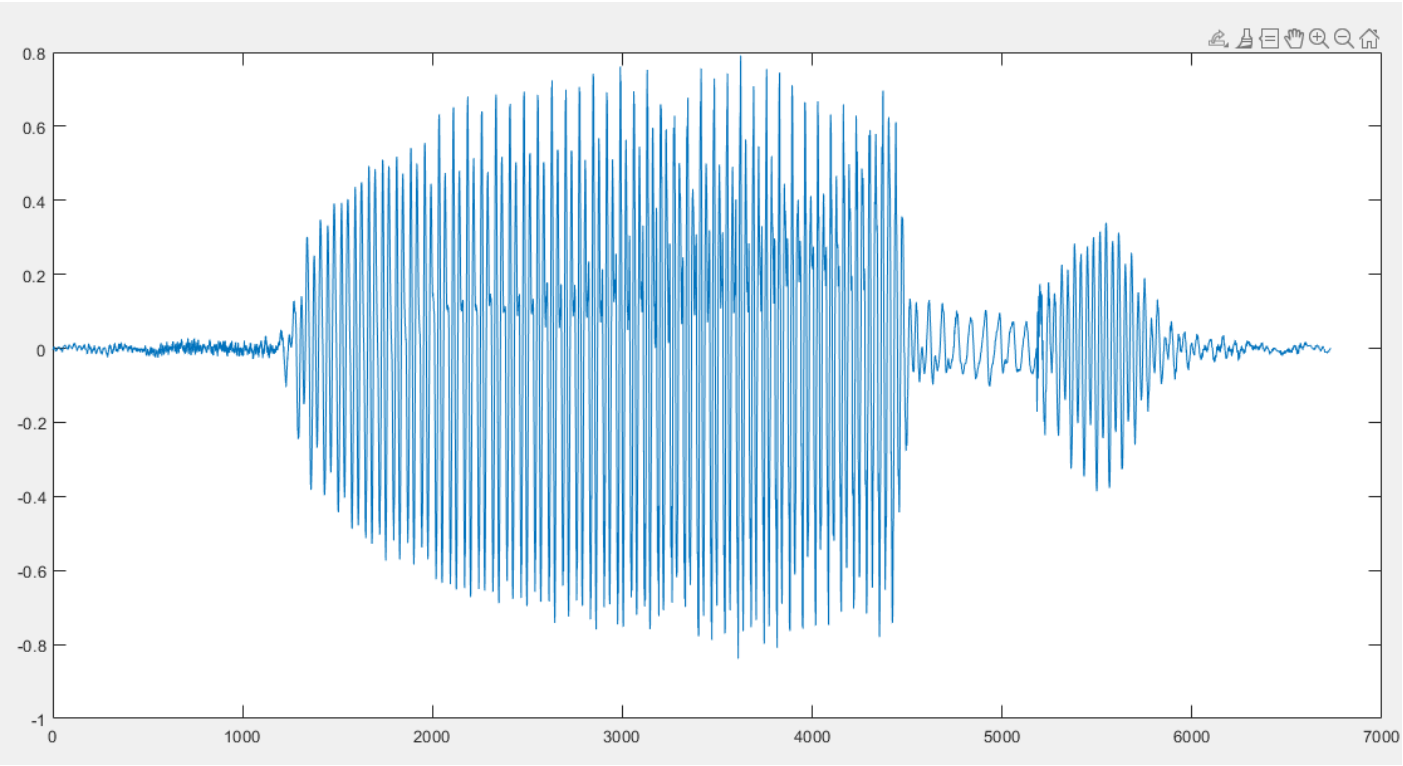
به کمک دو پارامتر β و α می‌توان glottal shaping filter طراحی کرد که در آنها با افزایش n مقدار سیگنال به سمت صفر میل می‌کند اما برای آن که در لحظه n = 64 خروجی صفر گردد نیاز است با توجه به فرض سوال مقدار آلفا نزدیک 1 انتخاب گردد و به تبع آن مقدار بتا کوچکتر از 0.3 باشد در حالی‌که برای Rosenberg filter کافی است K <= 64 باشد. می‌دانیم هر چه سیگنال در حوزه زمان زودتر میرا شود ؛ در حوزه فرکانس آرام تر به سمت صفر میل می‌کند و از آنجا که ما به دنبال آن هستیم که فیلتر حنجره طول کمی داشته باشد تا بتوانیم فرضیه‌ای مطرح شده در درس برای صرف نظر از این فیلتر را برآورده سازد.

سوال 5)

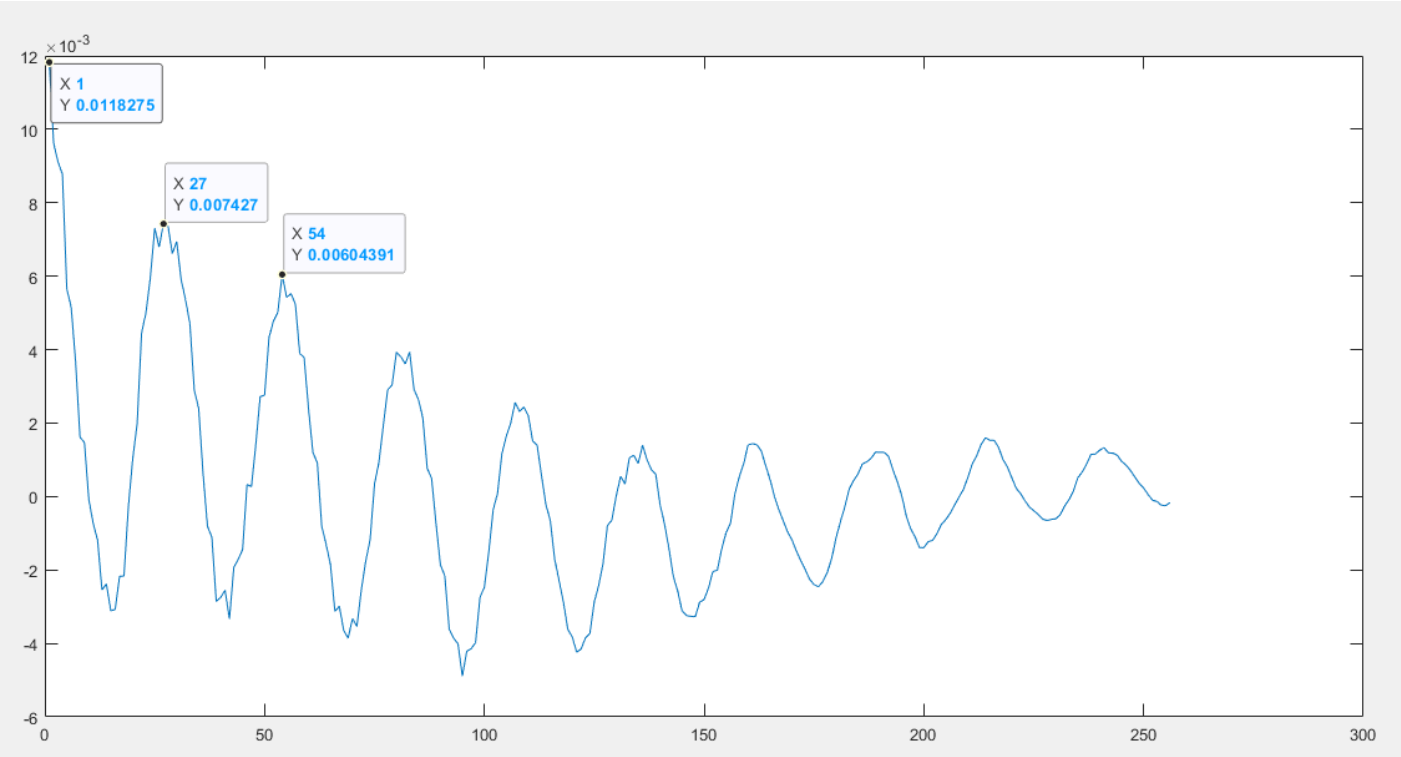
قسمت اول و دوم : محاسبه short – time autocorrelation و DTF

سیگنال صوت "Y0500.wav"

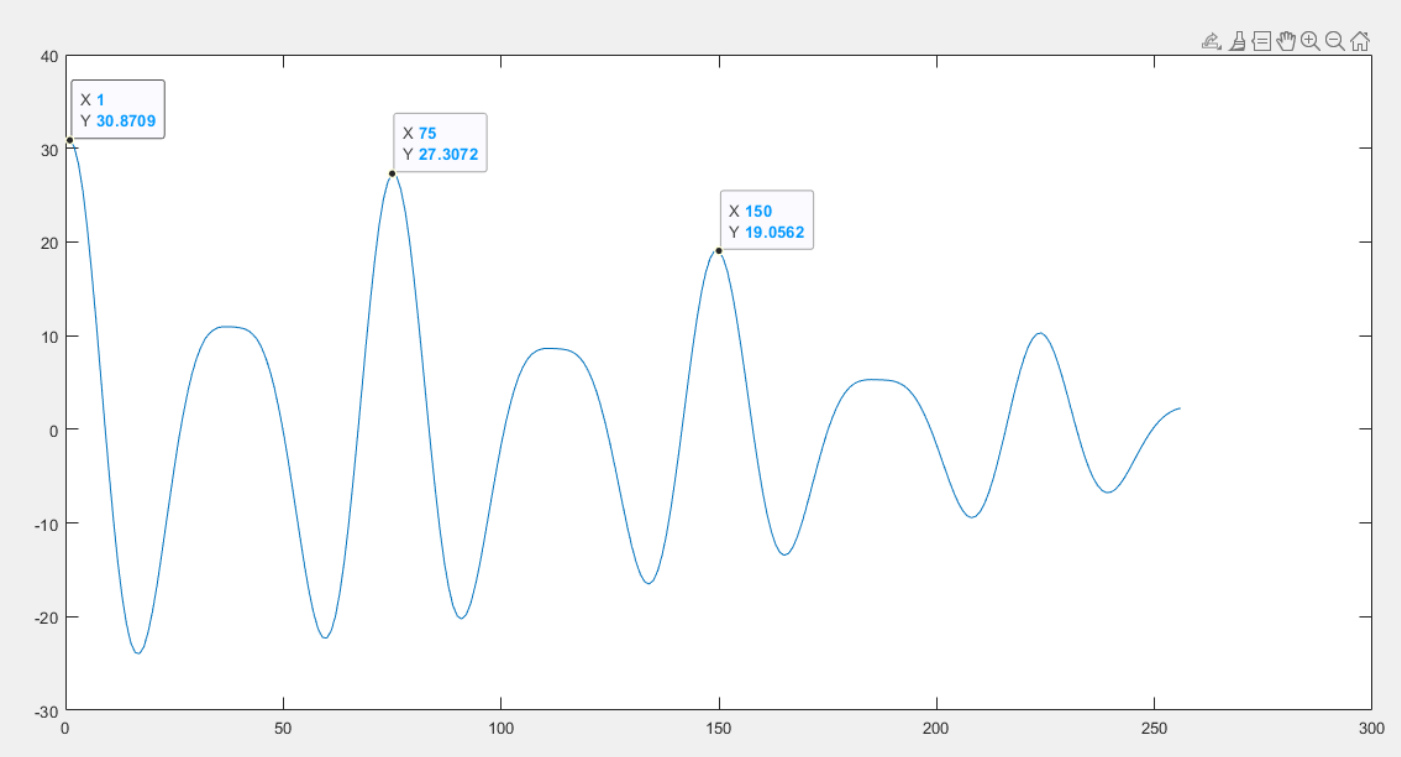
کل سیگنال مورد نظر در حوزه ی زمان



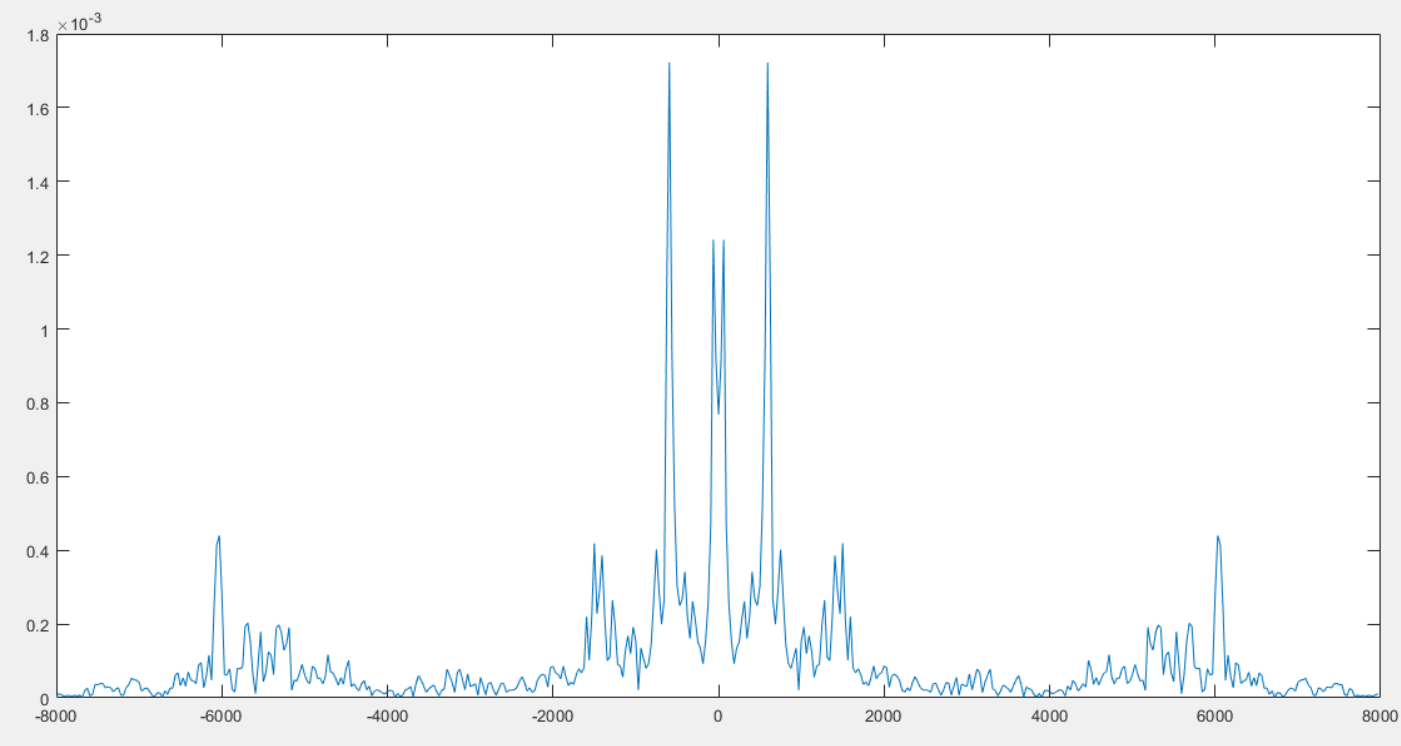
خود همبستگی سیگنال در پنجره ی اول



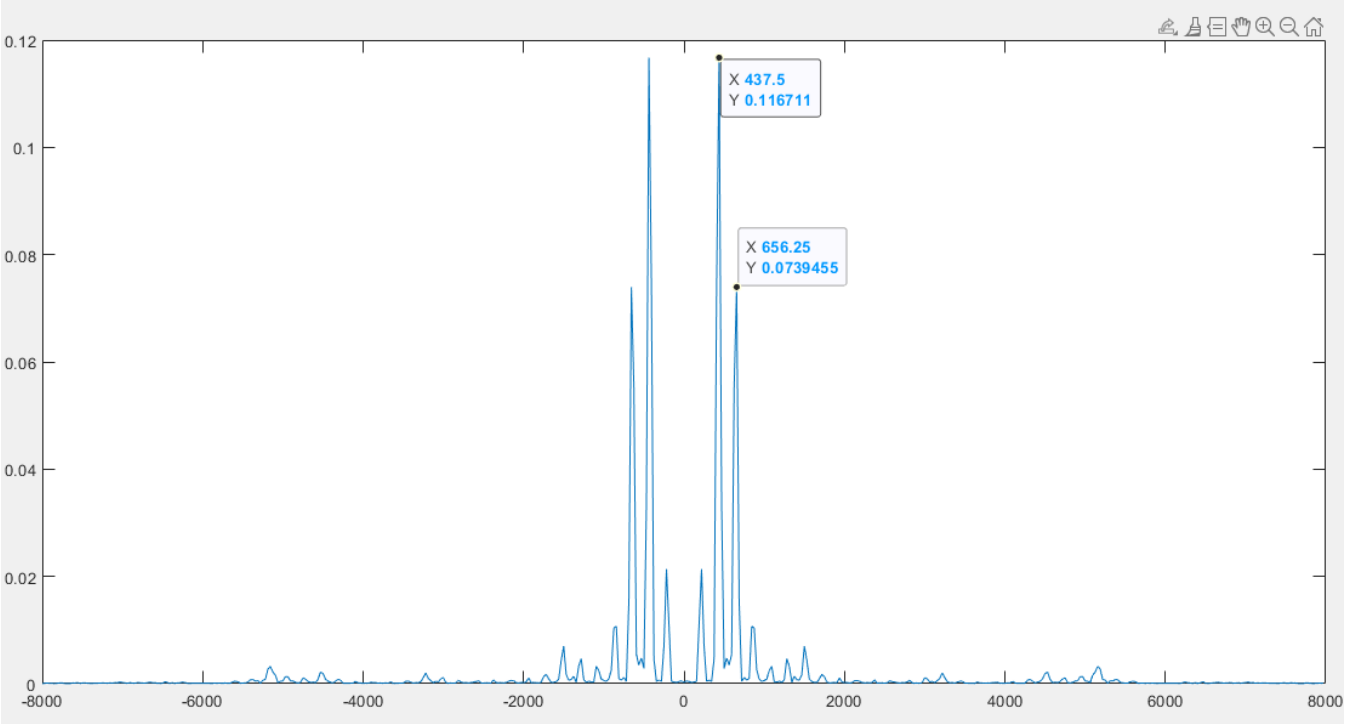
خود همبستگی سیگنال در پنجره پنجم



تبدیل فوریه 512 نقطه ای سیگنال موردنظر در پنجره اول



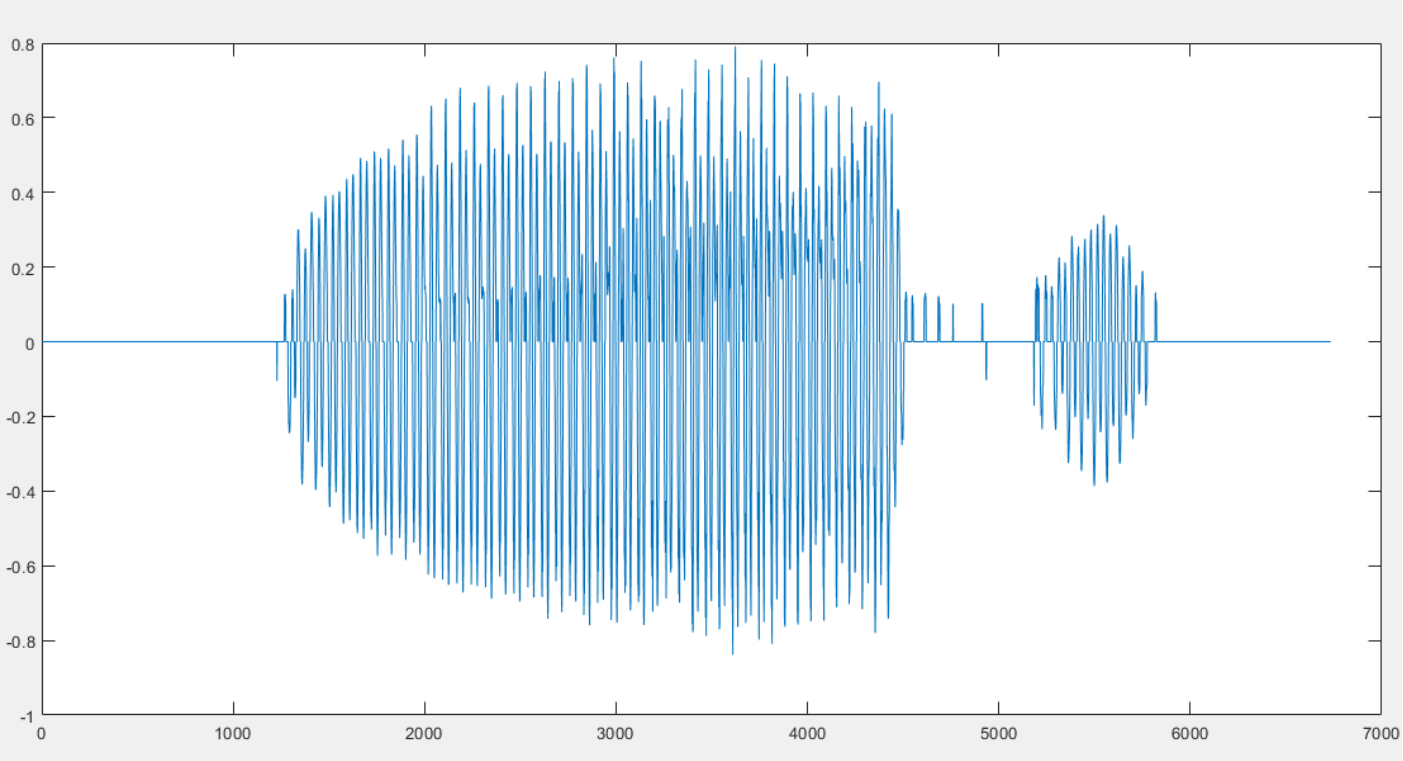
تبدیل فوریه 512 نقطه ای سیگنال موردنظر در پنجره پنجم



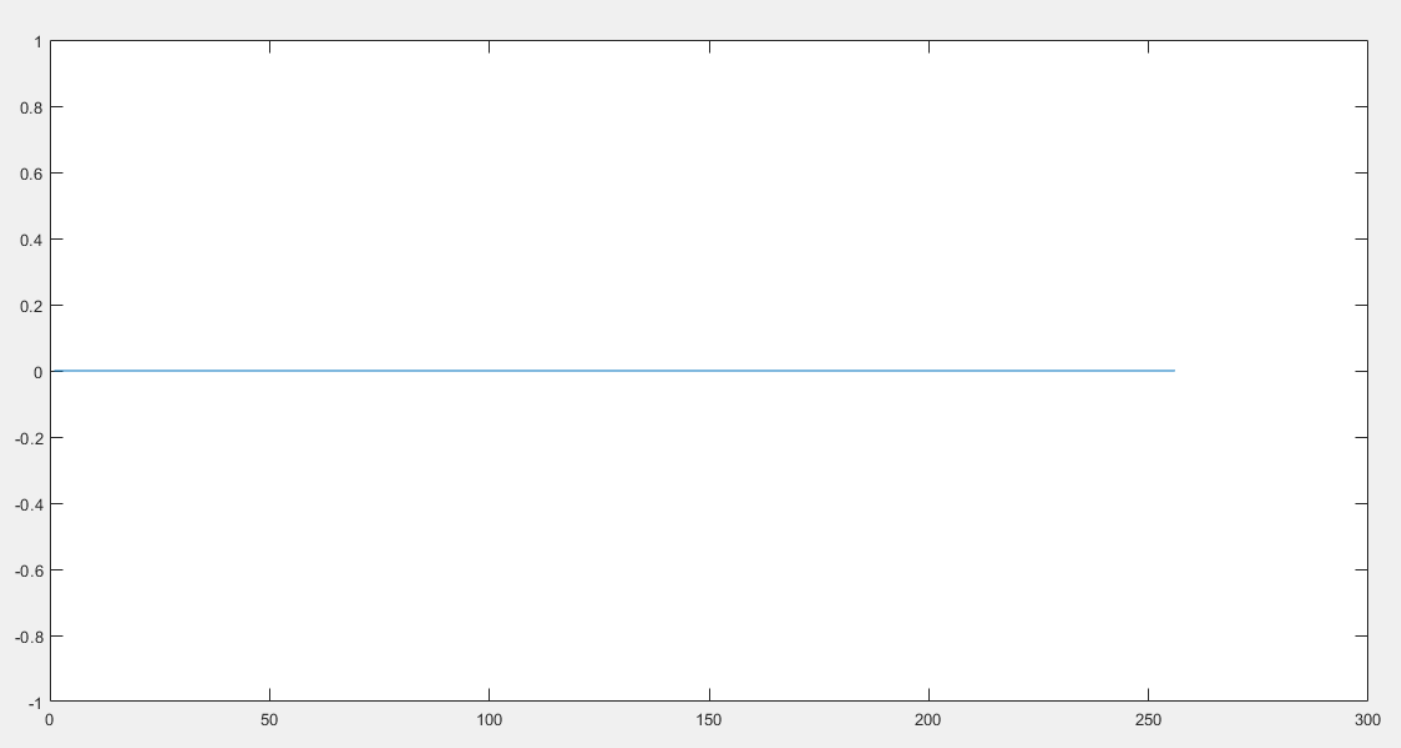
در شکل دیدیم که وقتی ما صدای واکه داریم ؛ تابع خود همبستگی ما انرژی بساری دارد و فاصله ی مبدا تا ماکسیمم مطلق pitch ما رو تعیین می‌کند.

قسمت سوم : وقتی که تابع غیر خطی cutter اضافه می کنیم

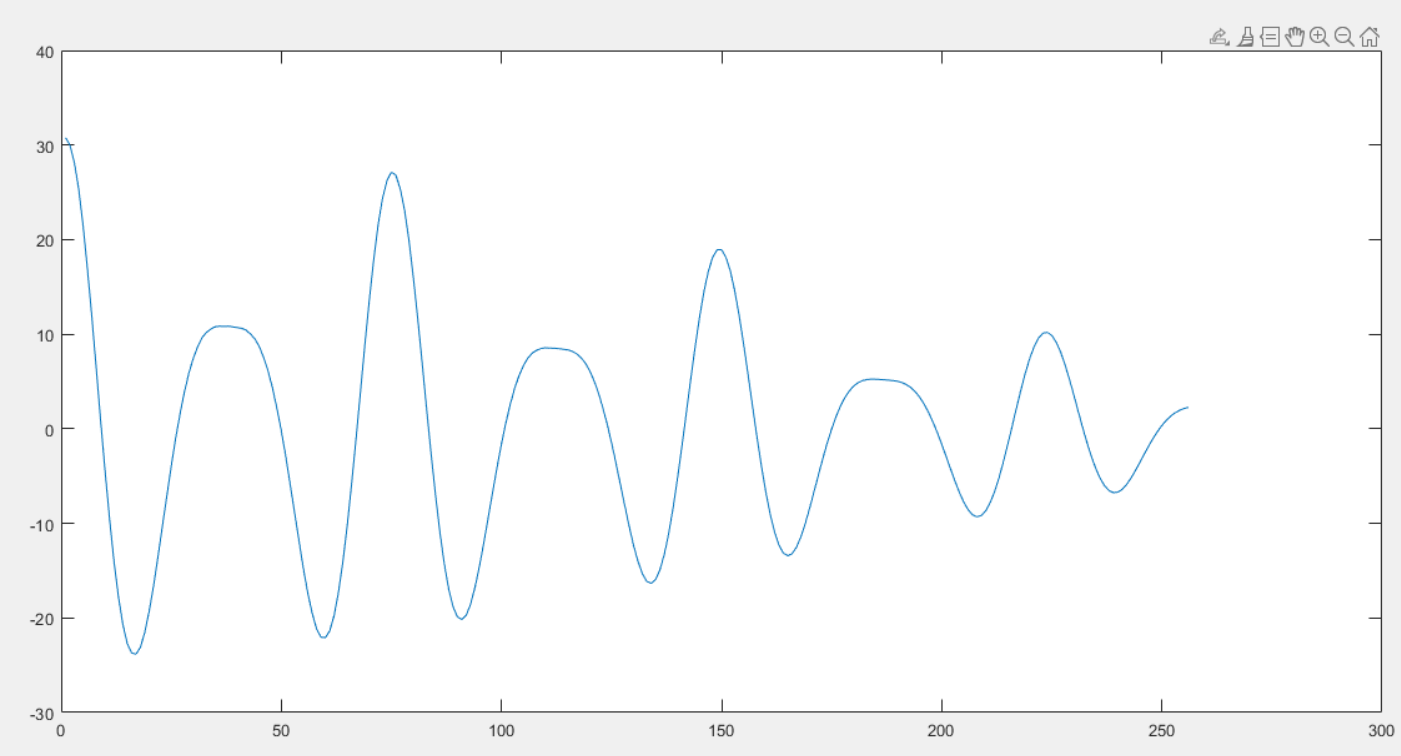
کل سیگنال مورد نظر در حوزه ی زمان



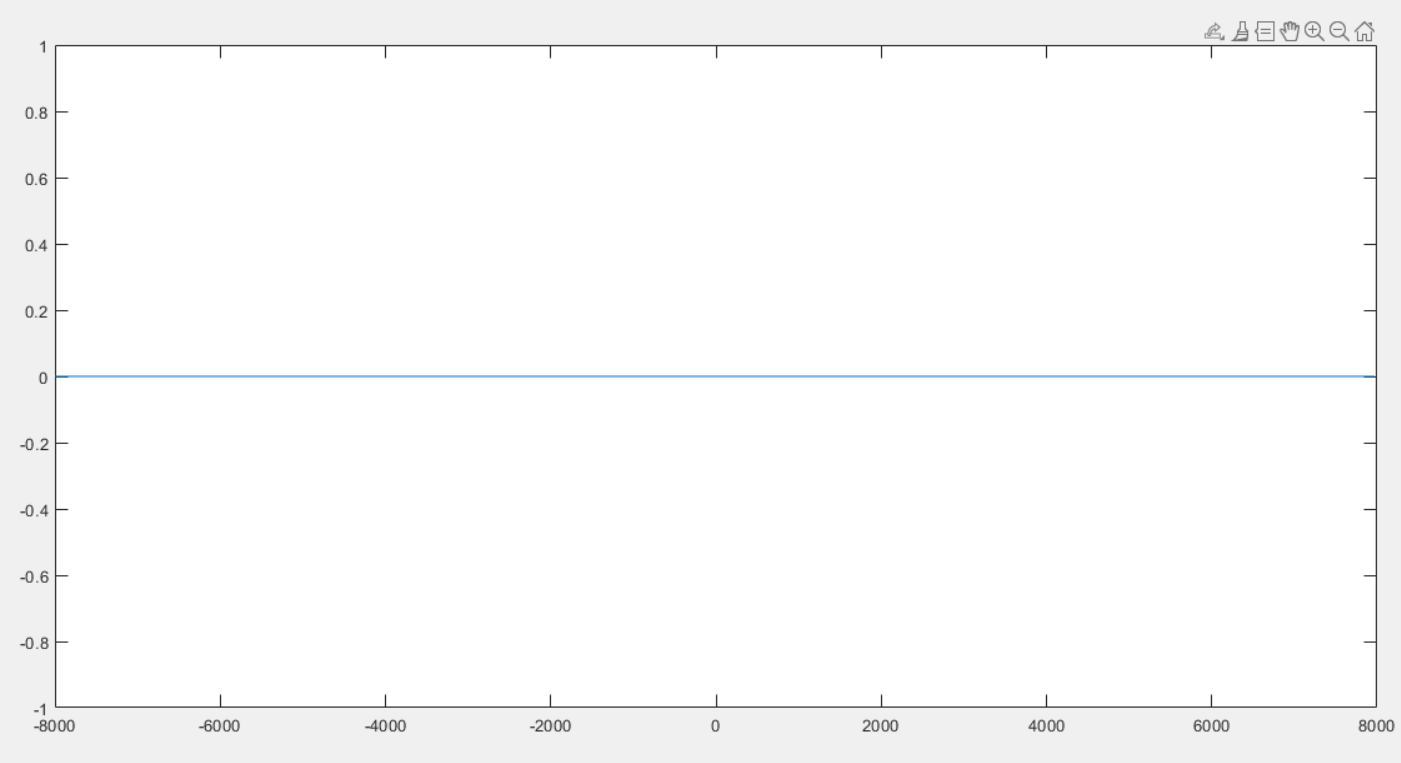
خودهمبسگتی سیگنال در پنجره ی اول



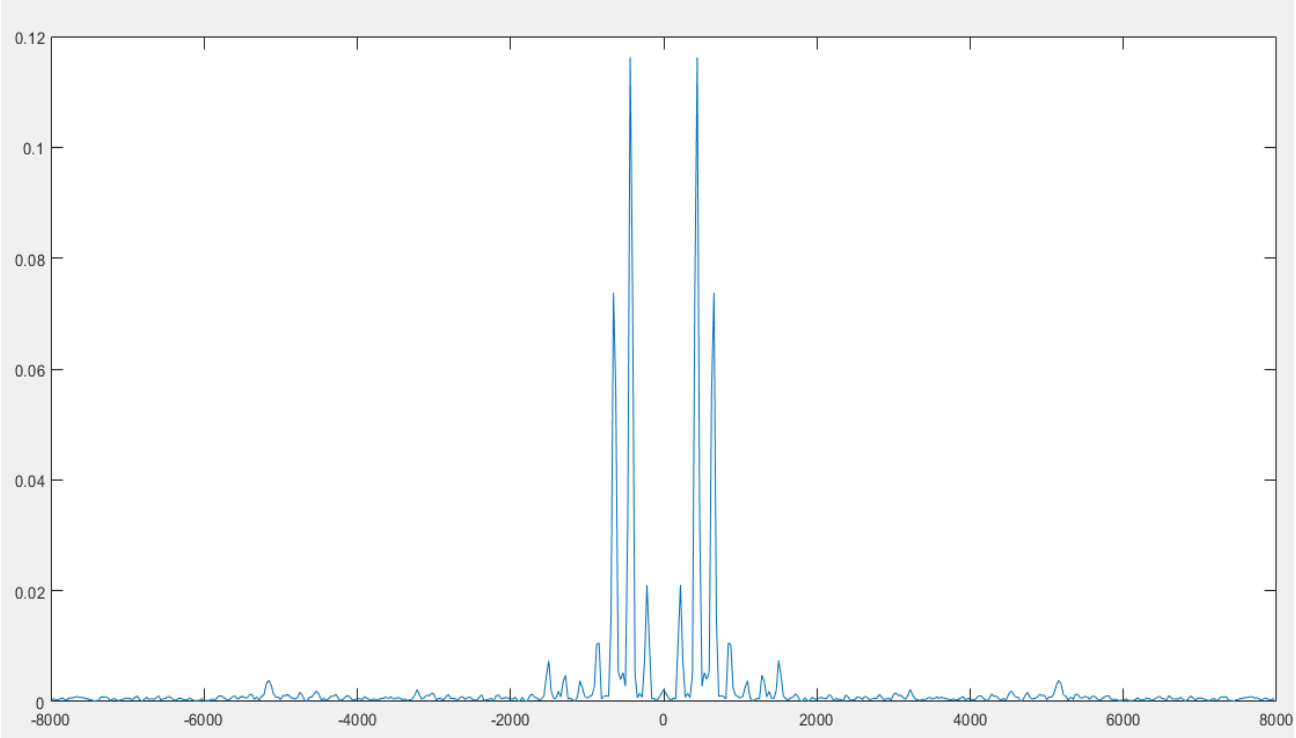
خودهمبسگتی سیگنال در پنجره پنجم



تبدیل فوریه 512 نقطه ای سیگنال موردنظر در پنجره اول



تبدیل فوریه 512 نقطه ای سیگنال موردنظر در پنجره پنجم

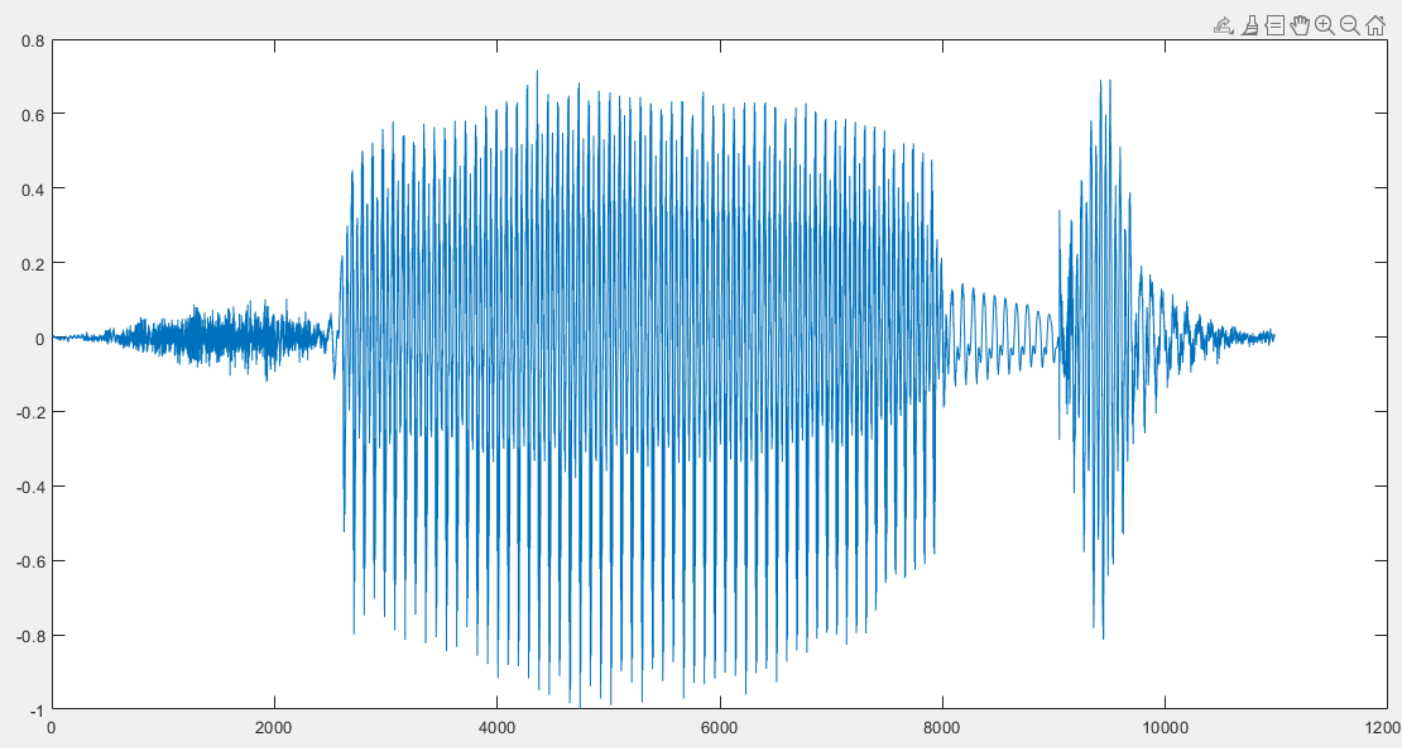


تابع غیر خطی cutter تابع غیر خطی خوبی است که البته می توان ان را با تقریب خوبی همانند خطی دانست که سبب گردیده صدای هایی که انرژی فوق العاده کمی نسبت به واکه ها دارند حذف گردند .

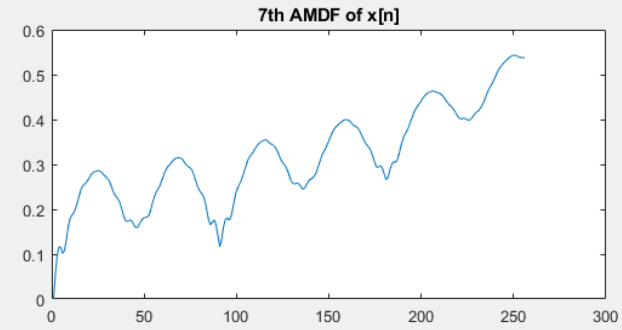
سوال 6)

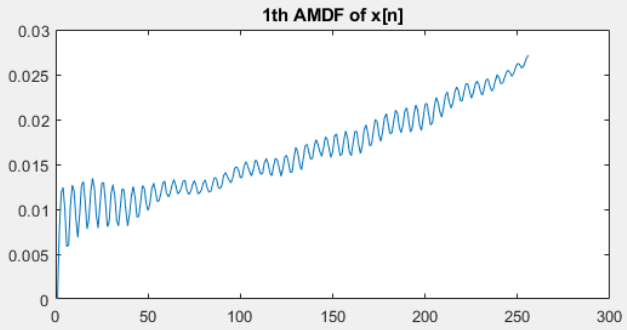
فایل صوتی مورد نظر "m01iy.wav" می باشد

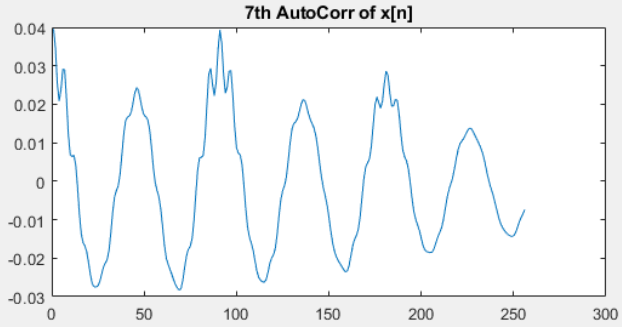
سیگنال مورد نظر در حوزه زمان

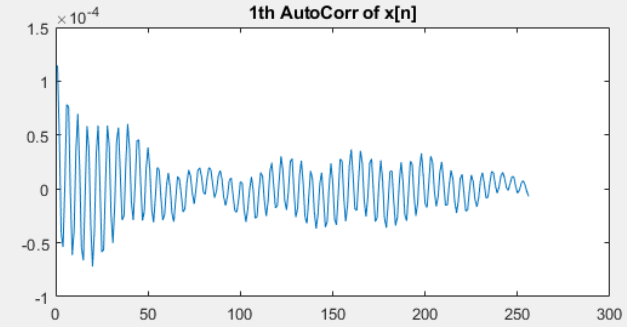


Bartlett

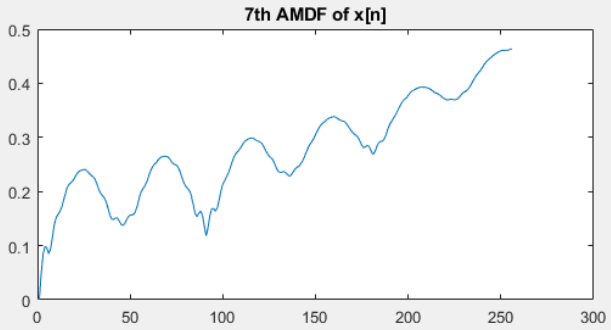


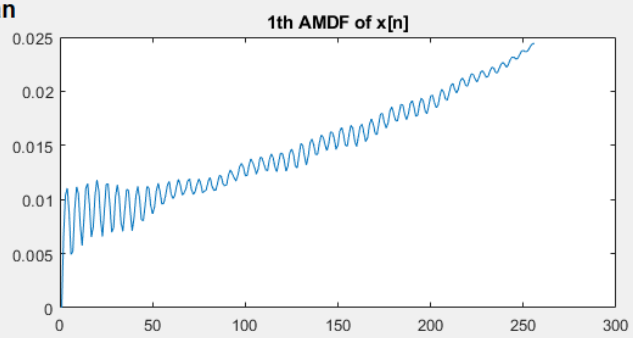


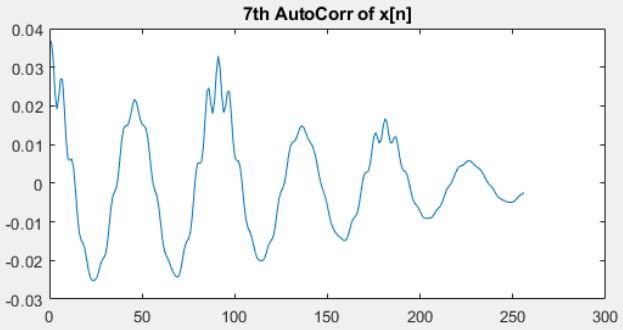


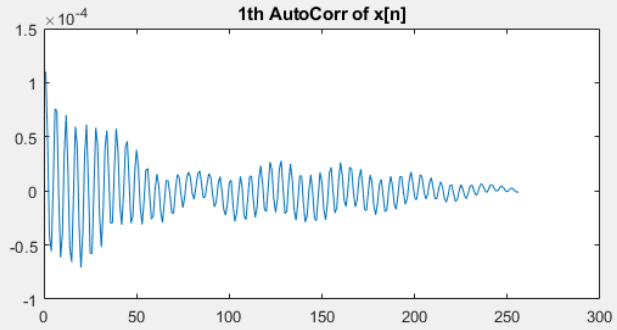


Blackman

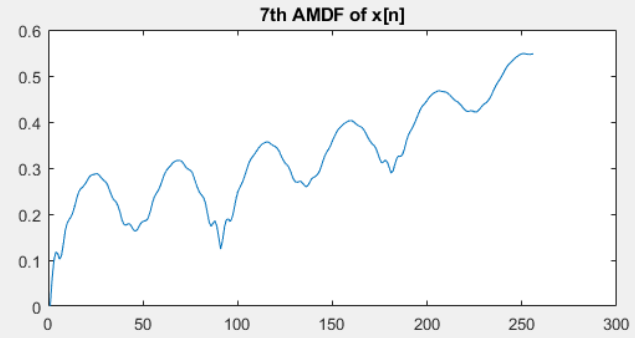


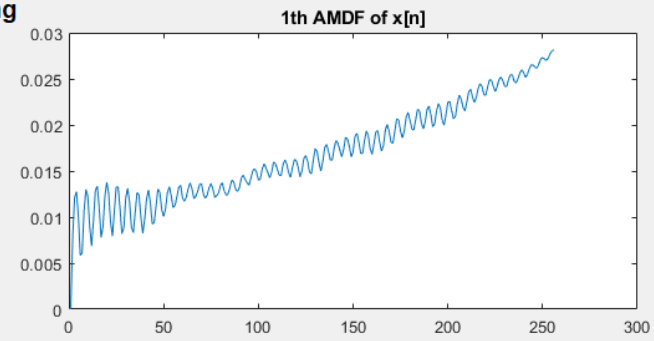


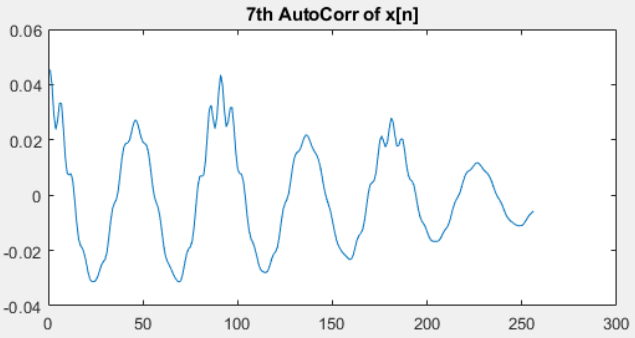


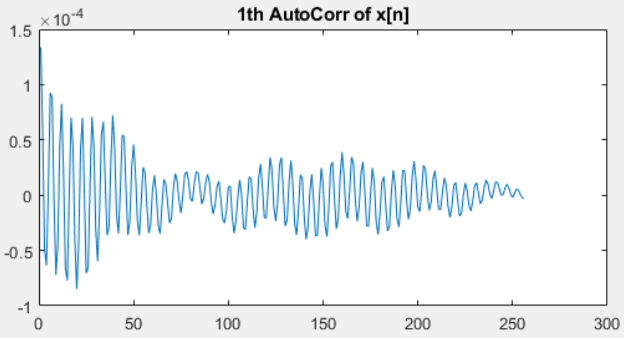


Hanning

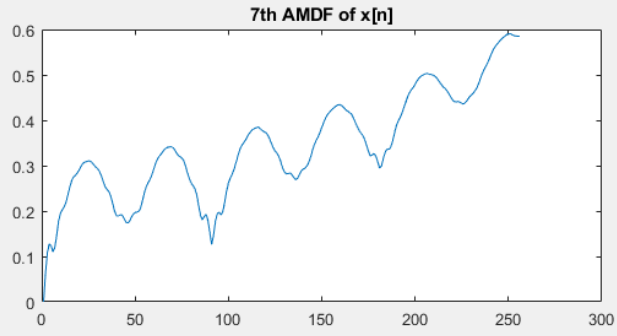


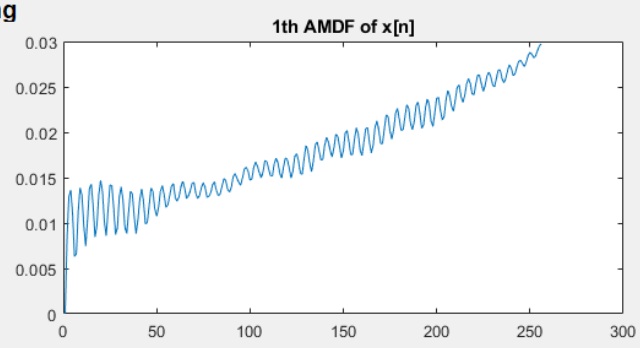


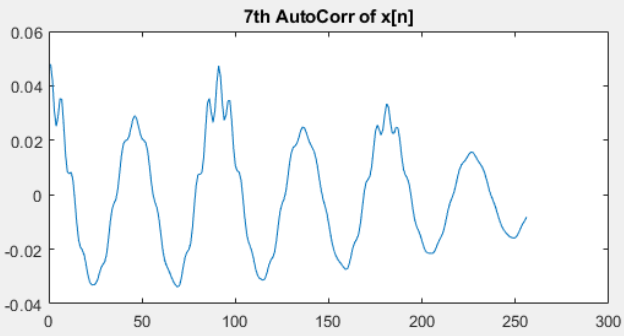


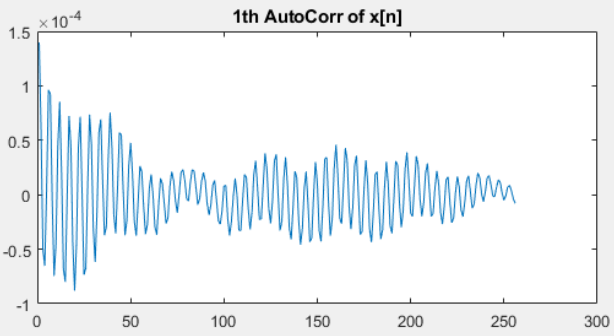


Hamming

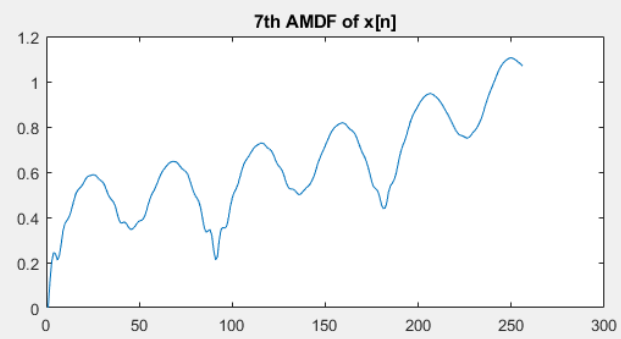


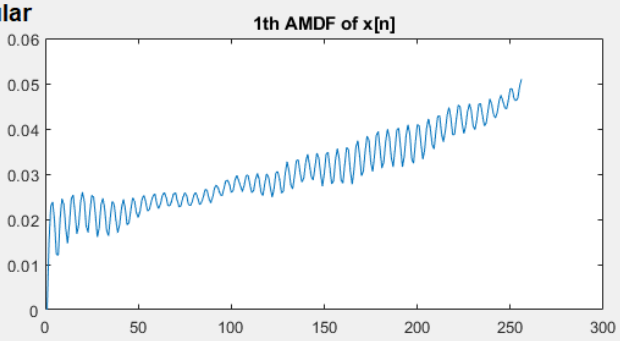


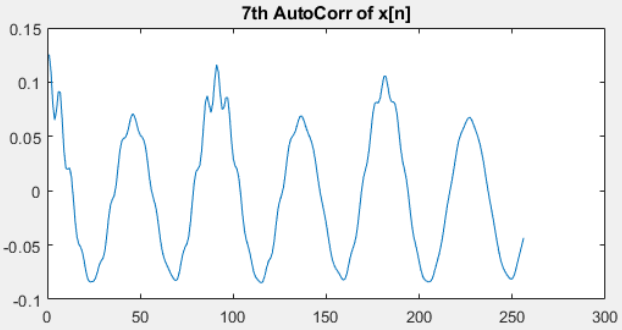


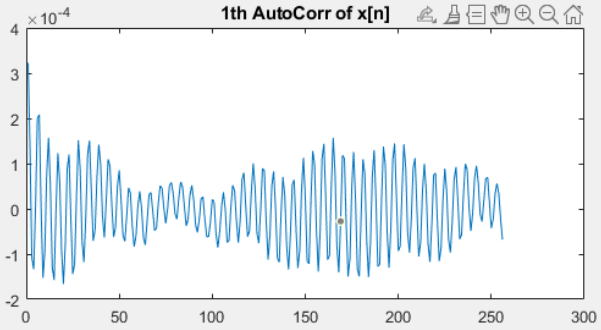


Rectangular

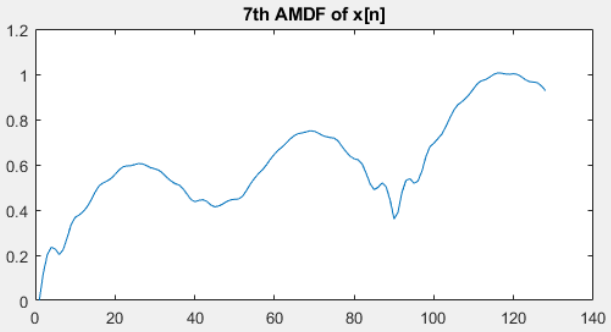


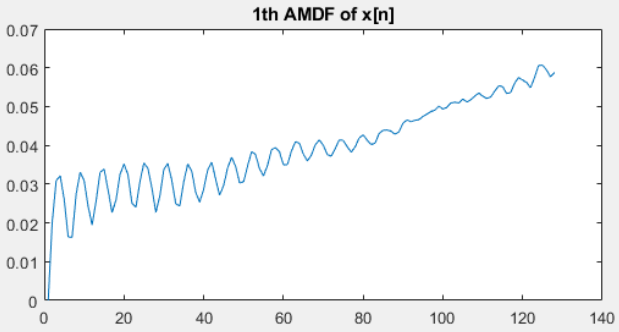


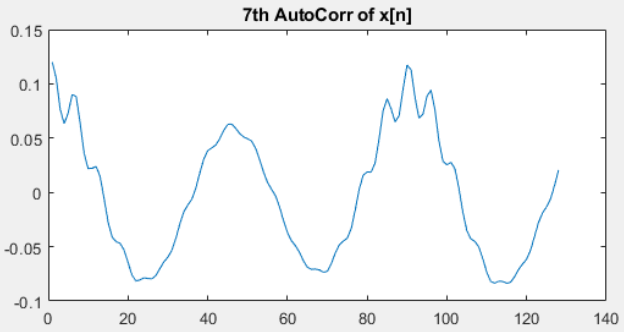


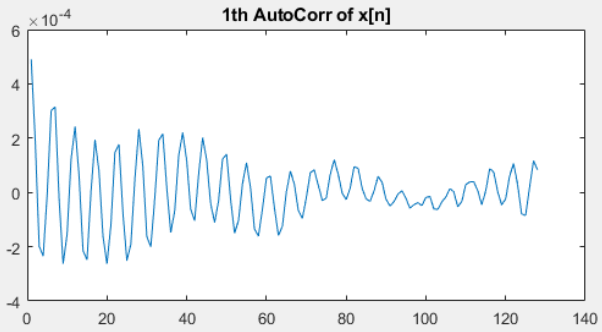


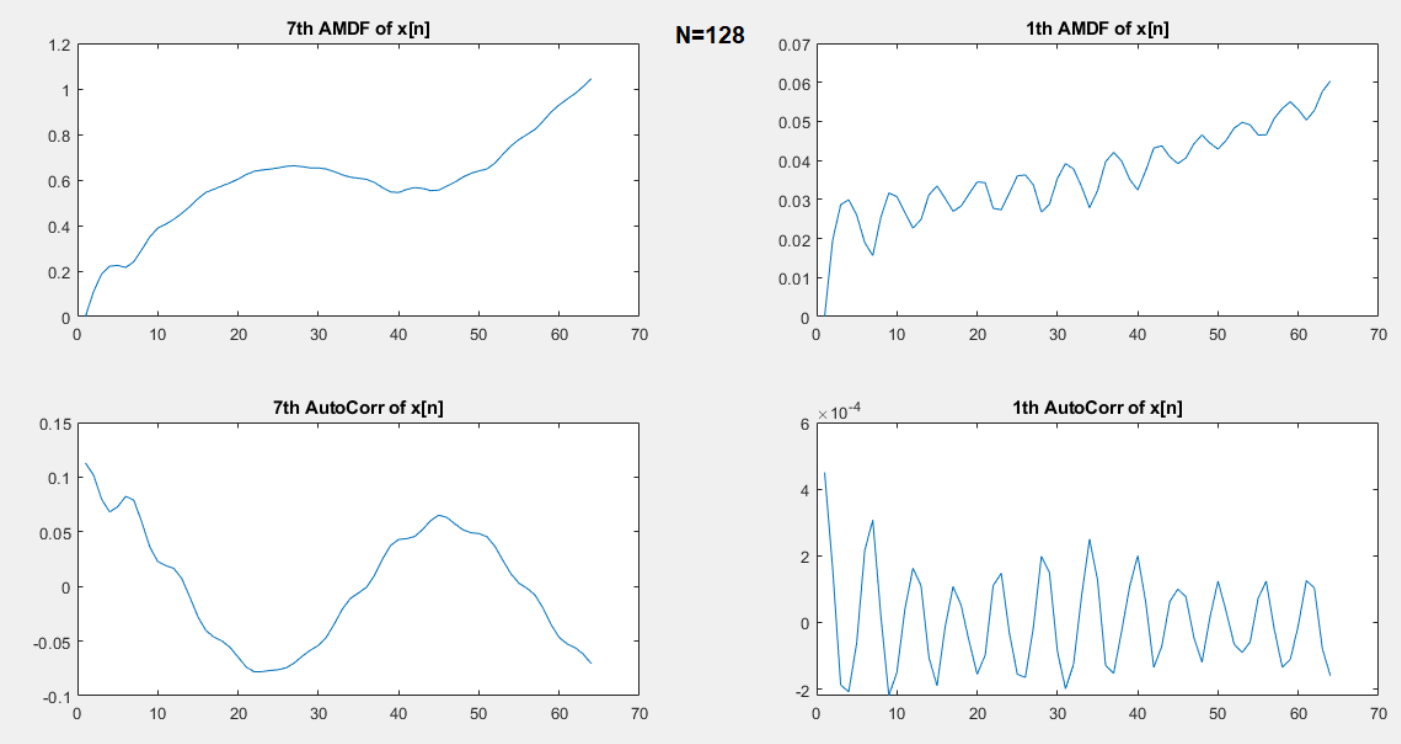
N=256











فرمول های پنجره زمانی

Rectangular

$$w[n] = \begin{cases} 1, & 0 \le n \le M, \\ 0, & \text{otherwise} \end{cases} \tag{7.47a}$$

Bartlett (triangular)

$$w[n] = \begin{cases} 2n/M, & 0 \le n \le M/2, \\ 2 - 2n/M, & M/2 < n \le M, \\ 0, & \text{otherwise} \end{cases} \tag{7.47b}$$

Hanning

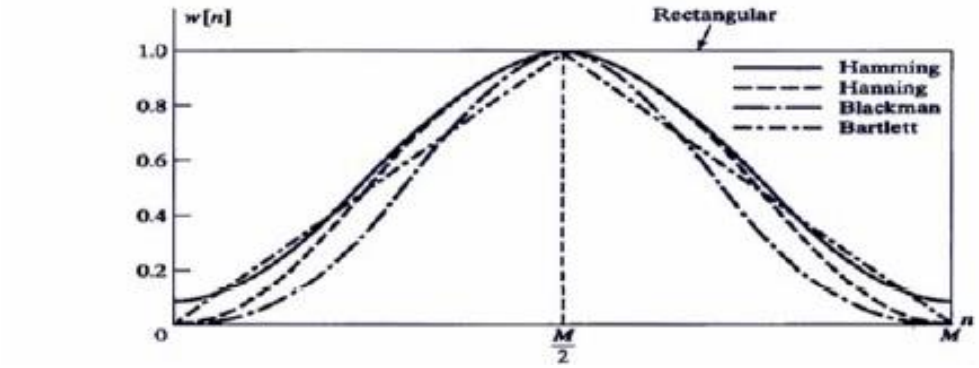
$$w[n] = \begin{cases} 0.5 - 0.5\cos(2\pi n/M), & 0 \le n \le M, \\ 0, & \text{otherwise} \end{cases} \tag{7.47c}$$

Hamming

$$w[n] = \begin{cases} 0.54 - 0.46\cos(2\pi n/M), & 0 \le n \le M, \\ 0, & \text{otherwise} \end{cases} \tag{7.47d}$$

Blackman

$$w[n] = \begin{cases} 0.42 - 0.5\cos(2\pi n/M) + 0.08\cos(4\pi n/M), & 0 \le n \le M, \\ 0, & \text{otherwise} \end{cases} \tag{7.47e}$$



در شکل های بالا از پنجره های هایی که در درس DSP تحت عناوین rectangular وhamming و hanning و blackman و bartlett استفاده گردیده است.

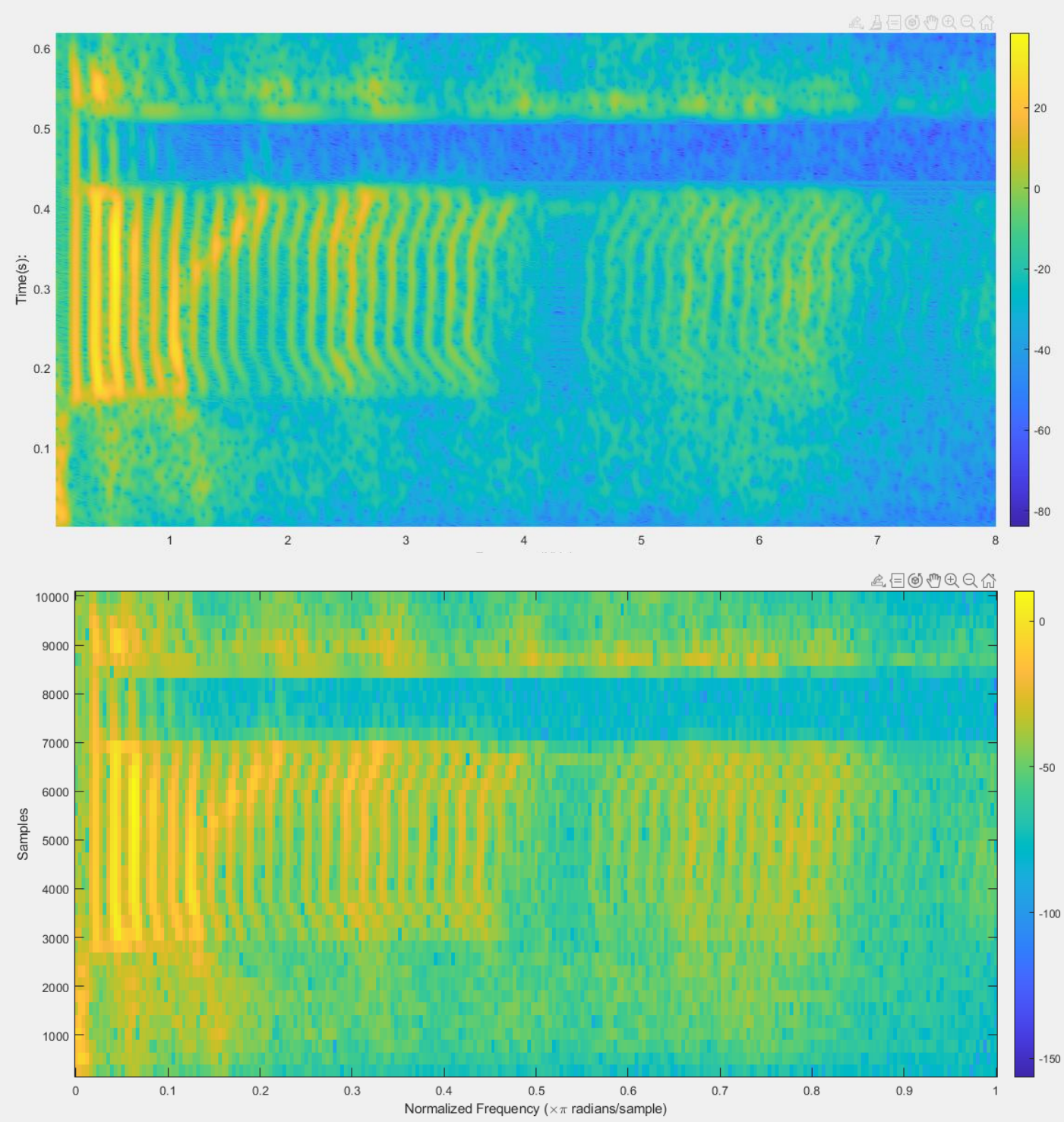
طول پنجره ها هم 512 و 256 و 128 است.

همان طور که شکل ها مشاهده می‌شود ؛ نوع پنجره اختلاف زیادی در تشخیص درست pitch ایجاد نمی کند.
اما در مورد طول پنجره این اختلاف در تشخیص درست با طول 128 دیده شد. البته طبیعی و منطقی به نظر می‌رسد که هر چه طول پنجره کم گردد از دقت ایده آل فاصله می‌گیریم

سوال 7)

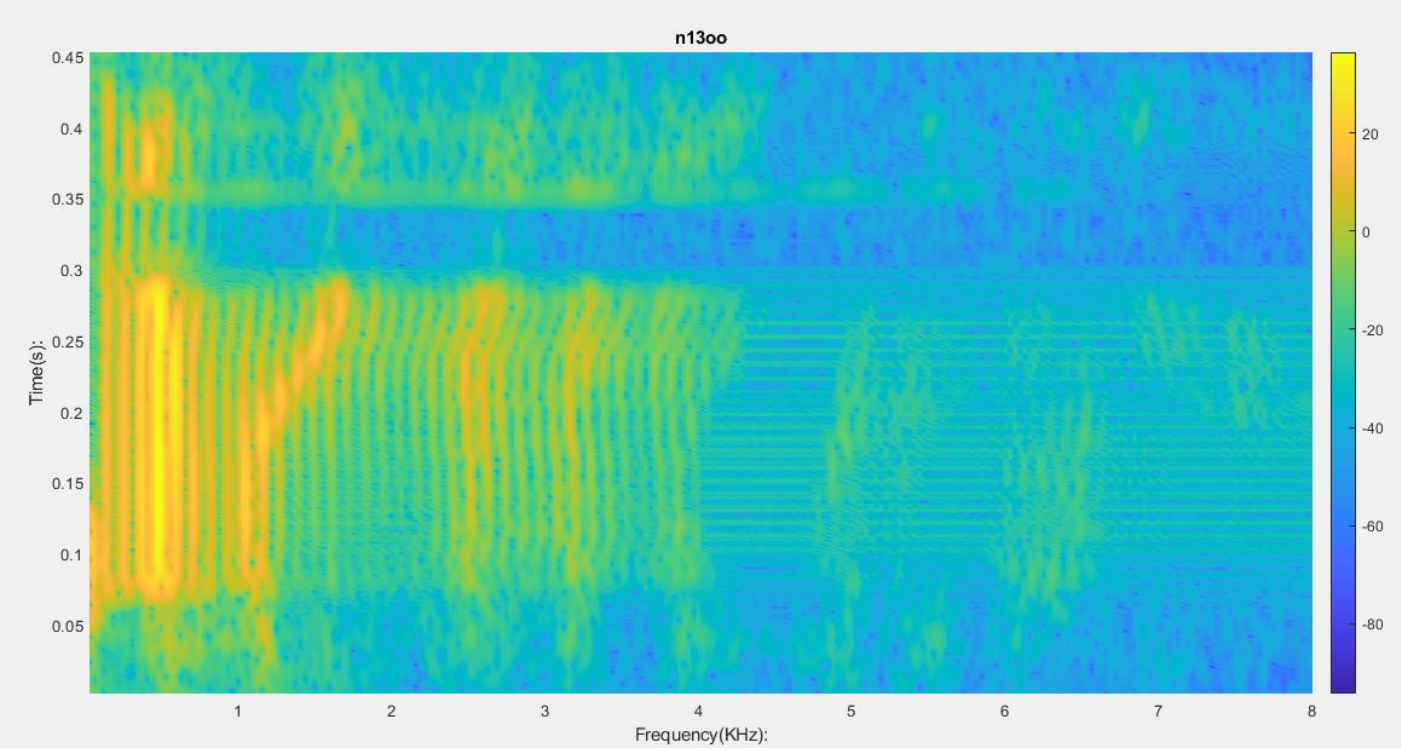
فایل صوتی مورد نظر "y05oo.wav" می باشد

قسمت اول

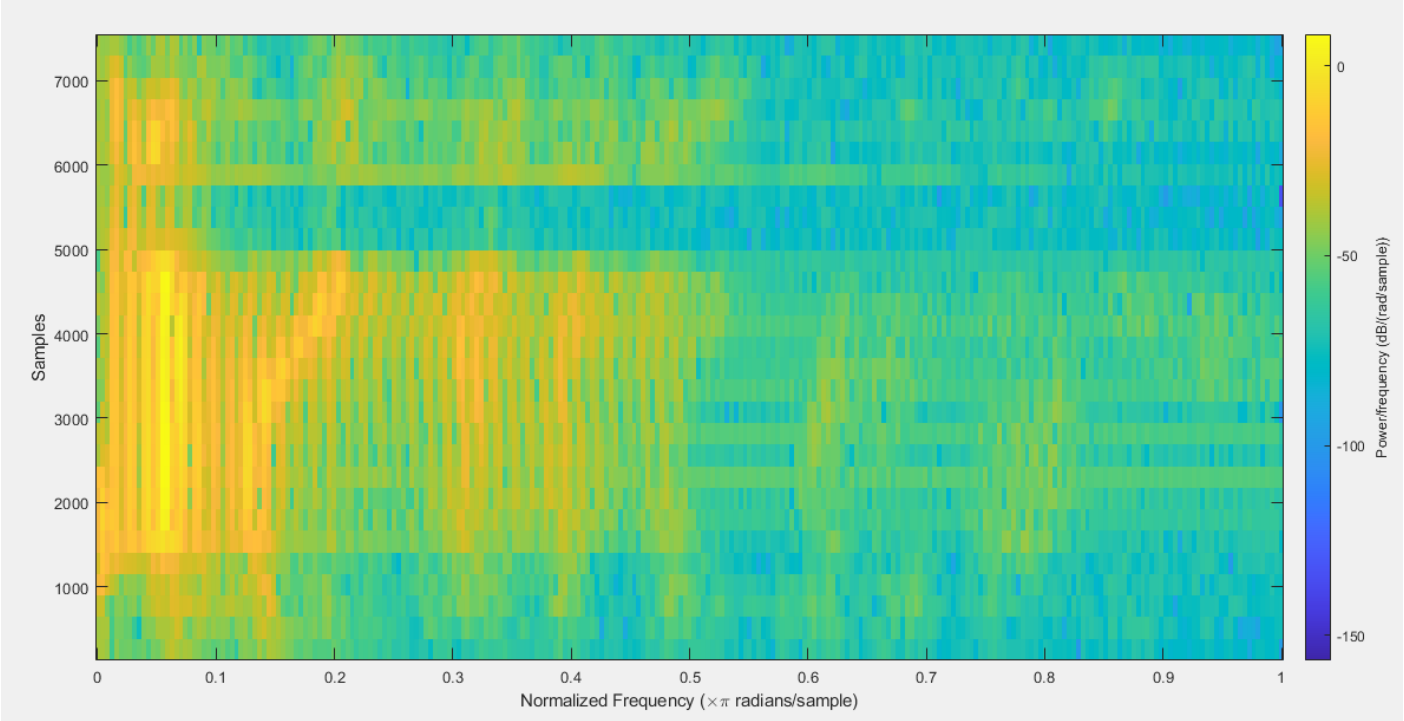


قسمت ب)

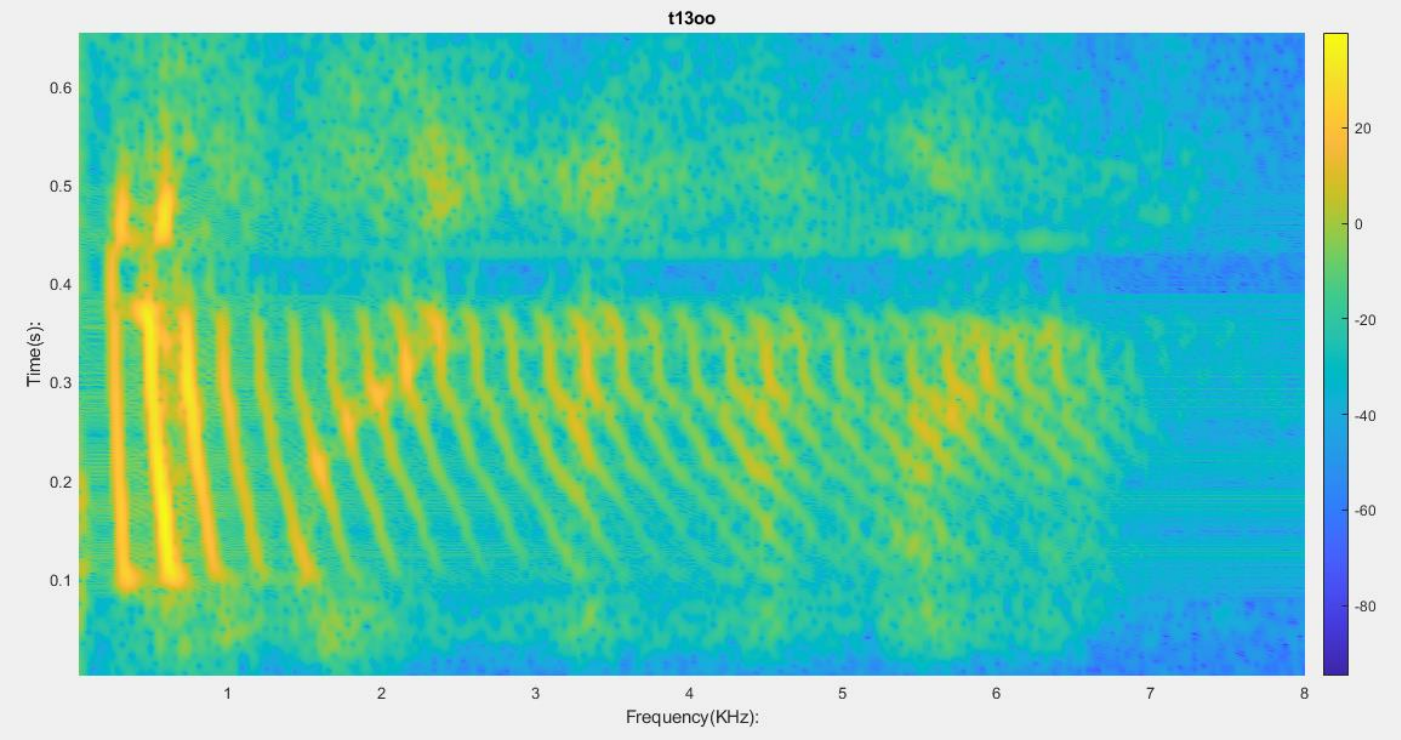
شکل رسم شده توسط خودمان :



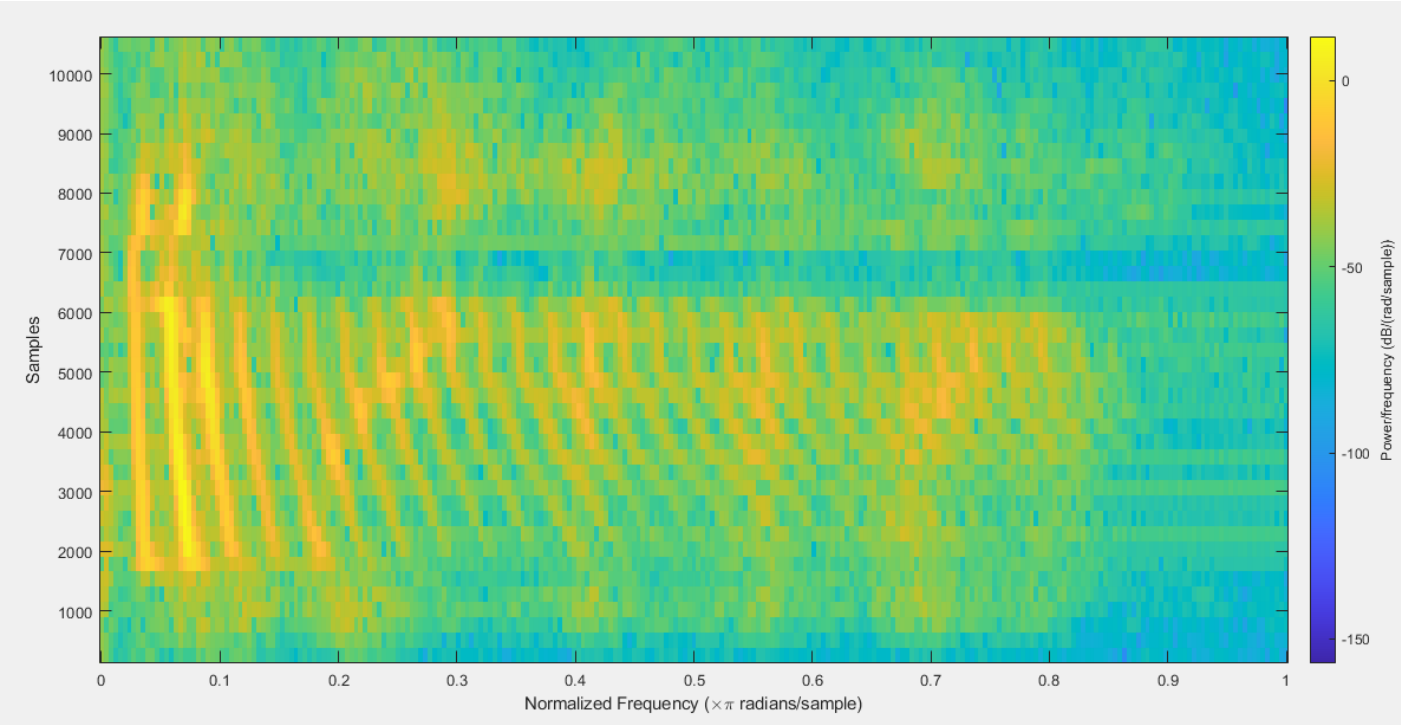
تابع Spectrogram متلب :



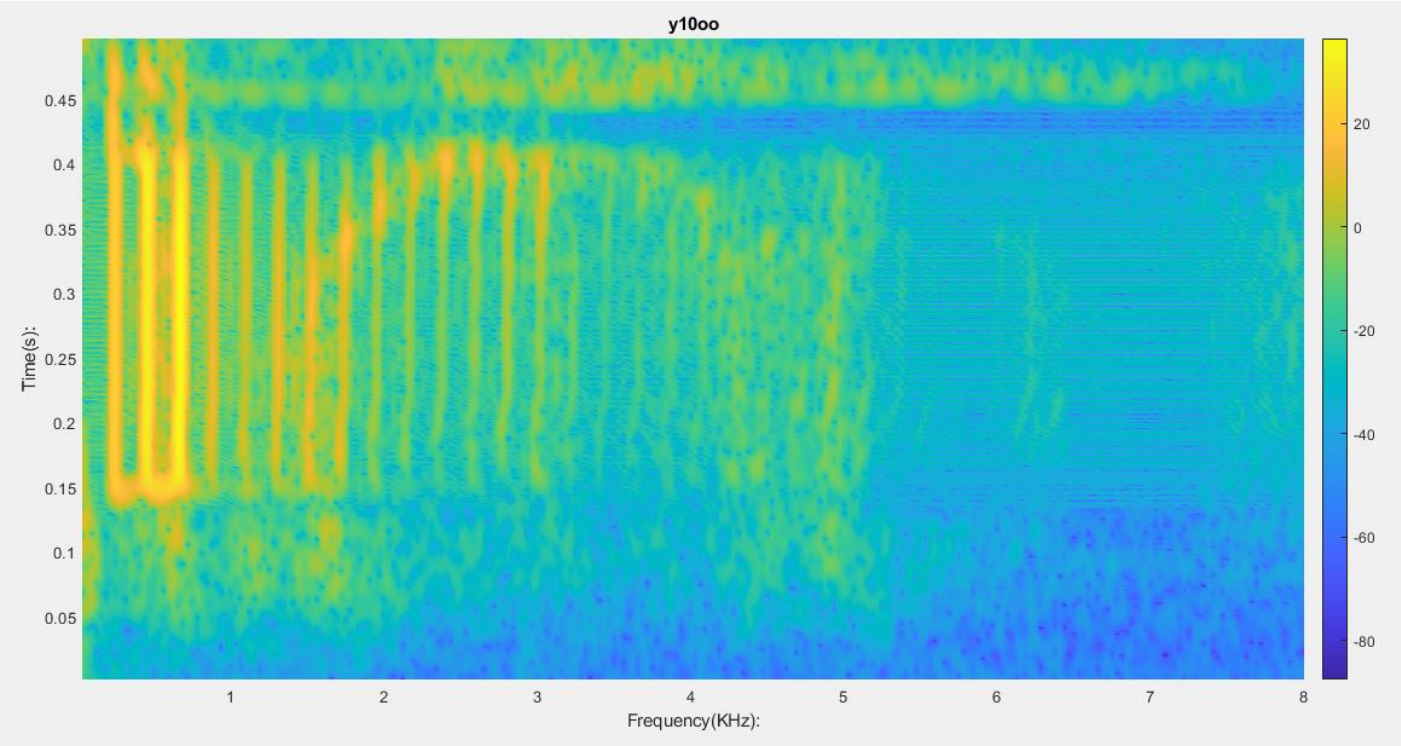
شکل رسم شده توسط خودمان :



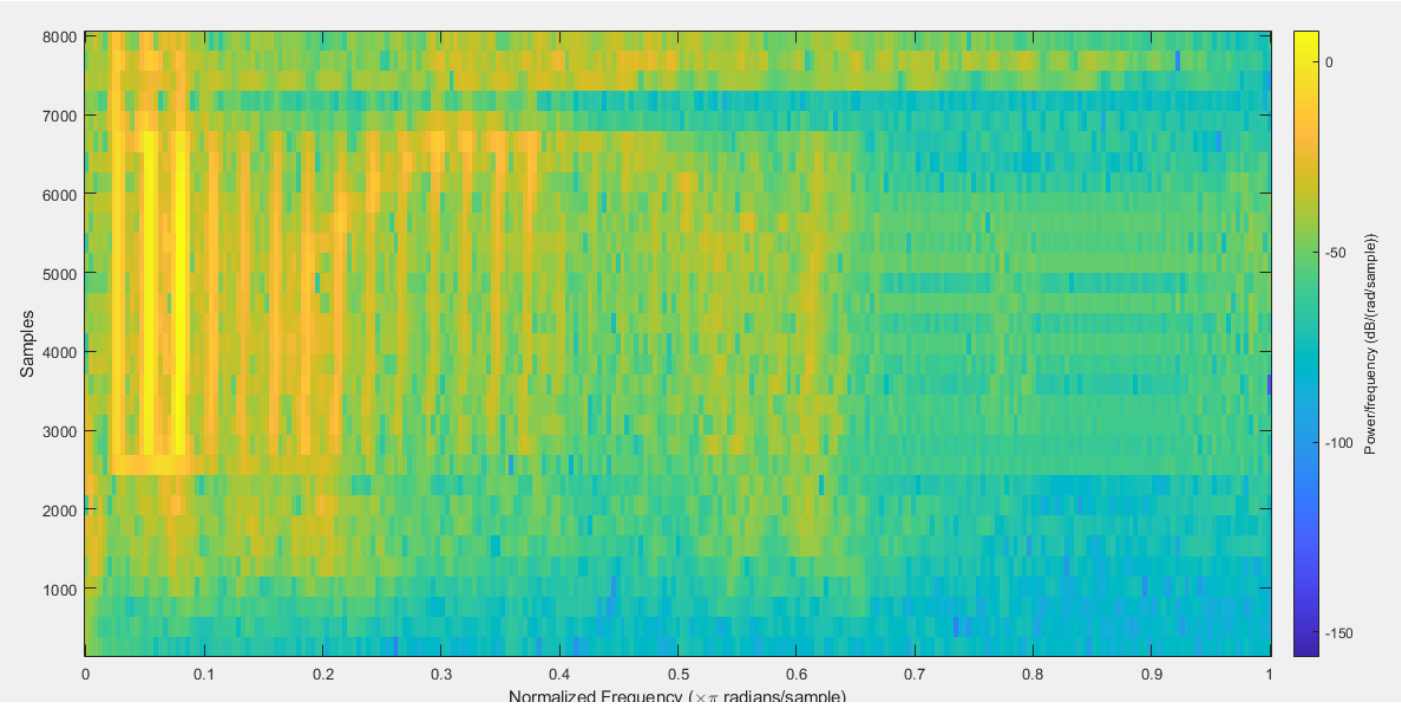
تابع Spectrogram متلب :



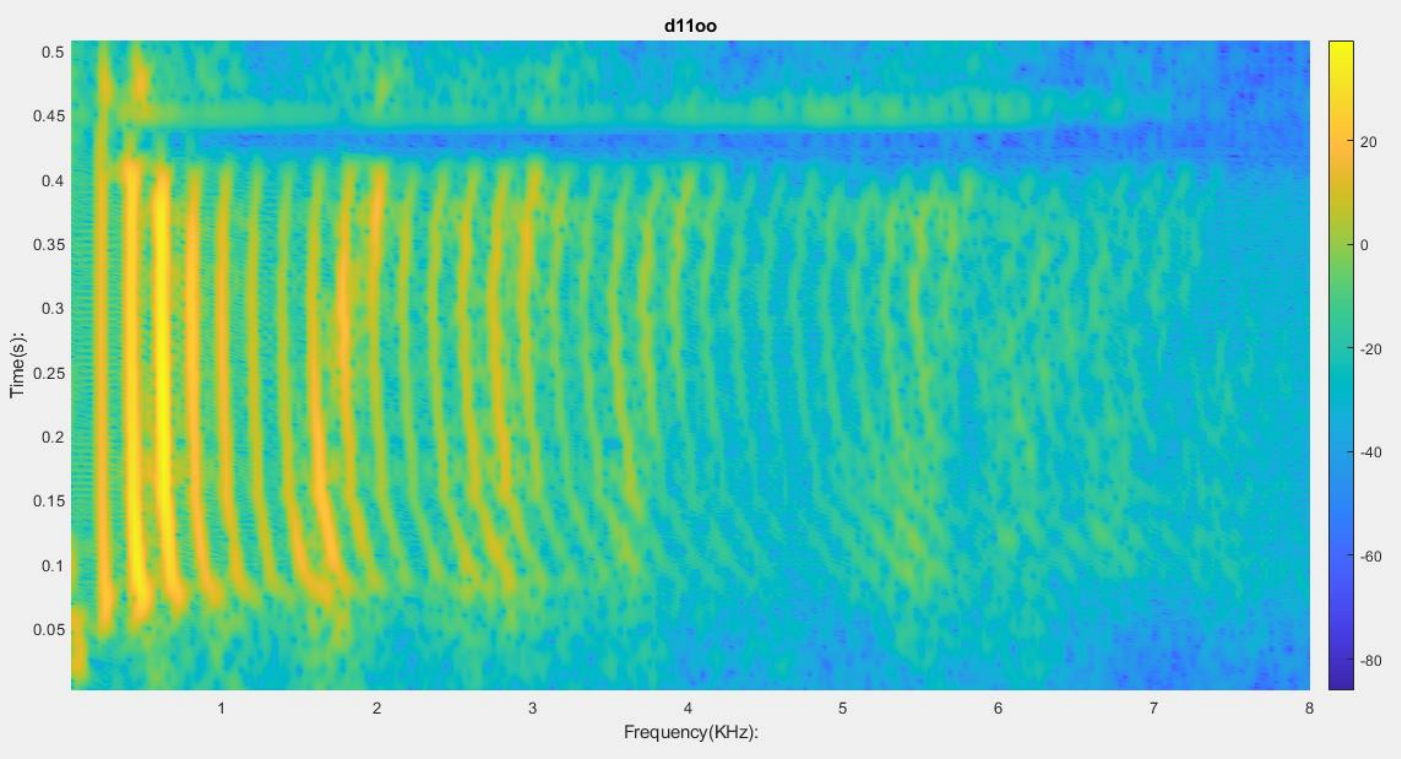
شکل رسم شده توسط خودمان :



تابع Spectrogram متلب :



شکل رسم شده توسط خودمان



تابع Spectrogram متلب