

به نام خدا

پروژه درس سیگنال و سیستم

نام و نام خانوادگی و شماره دانشجویی اعضای گروه : امیرحسین دهقانپور 400101186 –  
عبدالله درستکار 400101164 – سید محسن حسینی 400101045

موضوع پروژه : PSOLA و تغییر صدا

## خواسته ها :

### 1- توضیحات کلی درباره الگوریتم ها و شیوه پیاده سازی ان ها :

- تبدیل فوریه کوتاه مدت

مفهوم اساسی : سیگنال را در حوزه زمان به تعدادی سیگنال با مدت زمان کمتر تقسیم کنید، سپس هر سیگنال را به حوزه فرکانس تبدیل کنید •. برای بازسازی قطعات سیگنال به تعداد هارمونیک کمتری نیاز دارد • به تعیین فاصله زمانی که در آن فرکانس های خاص رخ می دهد کمک میکند

معکوس STFT:

سیگنال اصلی را می توان از تبدیل توسط Inverse STFT بازیابی کرد .

پذیرفته شده ترین روش معکوس کردن STFT با استفاده از روش همپوشانی -افزودن (OLA) است که امکان تغییرات در طیف پیچیده STFT را نیز فراهم می کند

وضوح زمان و فرکانس :

STFT وضوح ثابتی دارد عرض تابع پنجره به نحوه نمایش سیگنال مربوط می شود

• تعیین می کند که آیا وضوح فرکانس خوبی وجود دارد (قطعات فرکانس نزدیک به همرا می توان از هم جدا کرد) یا وضوح زمانی خوب (زمانی که در آن فرکانس ها تغییر می کنند

• یک پنجره عریض (تبدیل باند پهن) وضوح فرکانس بهتری را ارائه می دهد اما وضوح زمانی ضعیفی را ارائه می دهد .

یک پنجره باریکتر (تبدیل باند باریک) وضوح زمانی خوب اما وضوح فرکانس ضعیف را ارائه می دهد.

یعنی نمی توانیم رزولوشن های خوبی هم در زمان و هم در فرکانس داشته باشیم

-برای توضیح این محدودیت، توجه داشته باشید که در تبدیل فوریه

• برای افزایش وضوح فرکانس پنجره، فاصله فرکانسی ضرایب (توالی در حوزه فرکانس) باید کاهش یابد.

- کاهش فرکانس (Nyquist حداکثر) باعث افزایش اندازه پنجره می شود - زیرا اکنون نمونه های کمتری در واحد زمان وجود دارد . بنابراین هر تلاشی برای افزایش وضوح فرکانس باعث بزرگتر شدن اندازه پنجره و در نتیجه کاهش وضوح زمانی می شود و بالعکس.

برنامه های کاربردی - :

پردازش سیگنال هر سیگنال غیر ثابت (سیگنال های صوتی، تحریکات زلزله، پاسخهای سازه ای به ارتعاشات محیطی و ...)

-در دینامیک سازه، STFT را می توان برای موارد زیر استفاده کرد.

حالت های غالب ارتعاش (و شکل ها و فرکانس های آنها) را در هر بازه زمانی تعیین کنید

• پایش سلامت و تشخیص آسیب از طریق مطالعه فرکانس های غالب

به عنوان مثال کاهش در فرکانس به طور کلی نشان دهنده آسیب هایی است که منجر به سازه های نرم تر می شود.

PSOLA (Pitch Synchronous Overlap and Add) یک تکنیک پردازش سیگنال دیجیتال است که برای پردازش گفتار و به طور خاص تر سنتز گفتار استفاده می شود. می توان از آن برای تغییر زیر و بم و مدت زمان سیگنال گفتار استفاده کرد. در حدود سال 1986 اختراع شد.

PSOLA با تقسیم شکل موج گفتار به بخش های کوچک همپوشانی کار می کند. برای تغییر گام سیگنال، بخش ها بیشتر از هم فاصله می گیرند (برای کاهش گام) یا نزدیک تر به هم (برای افزایش گام). برای تغییر مدت زمان سیگنال، بخش ها چندین بار تکرار می شوند (برای افزایش مدت زمان) یا برخی حذف می شوند (برای کاهش مدت زمان). سپس بخش ها با استفاده از تکنیک اضافه کردن همپوشانی ترکیب می شوند .

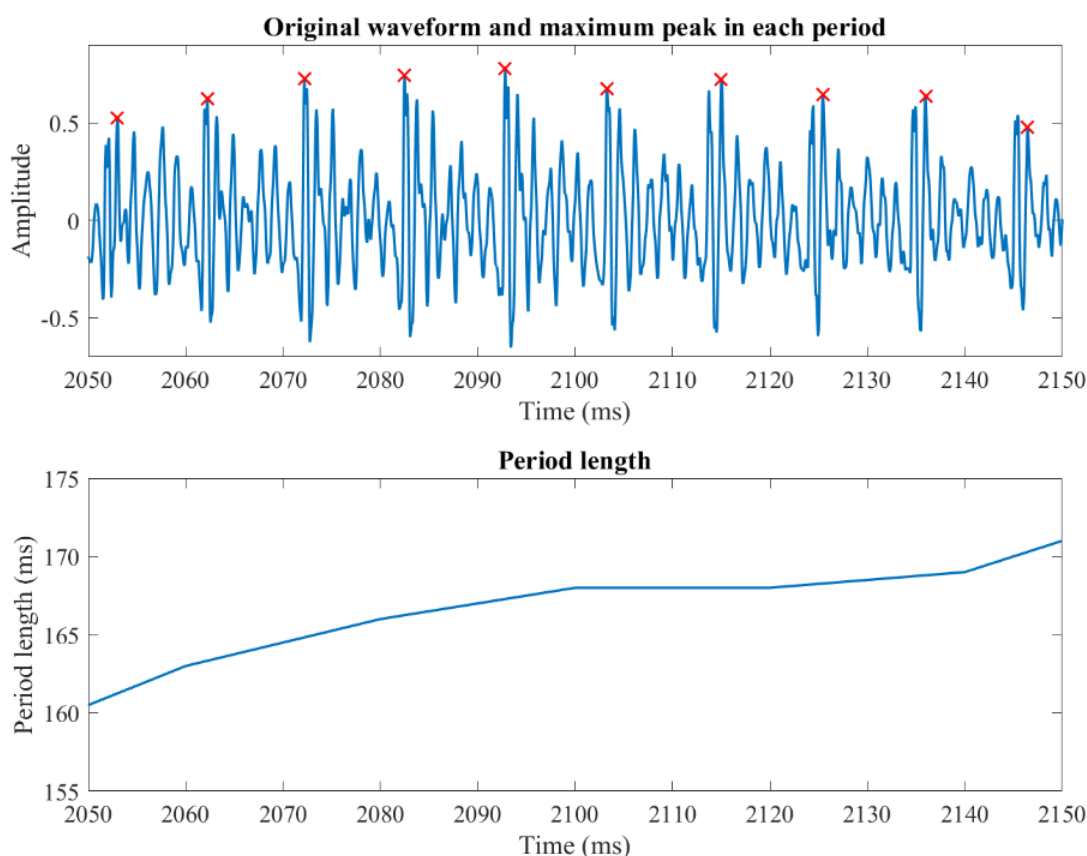
PSOLA می تواند برای تغییر عروض یک سیگنال گفتاری استفاده شود.

روش PSOLA یکی از روش های پردازش سیگنال صوتی است که برای تغییر فرکانس بندی یک صدای گفتاری به کار می رود. برای این کار، ابتدا باید فرکانس بندی صدای گفتاری را تخمین بزنیم. سپس با استفاده از روش PSOLA، سیگنال را به دوره های فرکانسی تجزیه می کنیم و با جابجایی دوره های فرکانسی، فرکانس بندی صدای گفتاری را تغییر می دهیم.

روش STFT (Short-Time Fourier Transform) نیز یکی از روش های پردازش سیگنال صوتی است که برای تحلیل فرکانسی یک سیگنال صوتی به کار می رود. در این روش، سیگنال صوتی به قطعات کوتاه تر تقسیم شده و برای هر قطعه، تبدیل فوریه انجام می شود. سپس طیف فرکانسی هر قطعه به صورت جداگانه محاسبه شده و به عنوان یک نمودار زمان-فرکانس نمایش داده می شود.

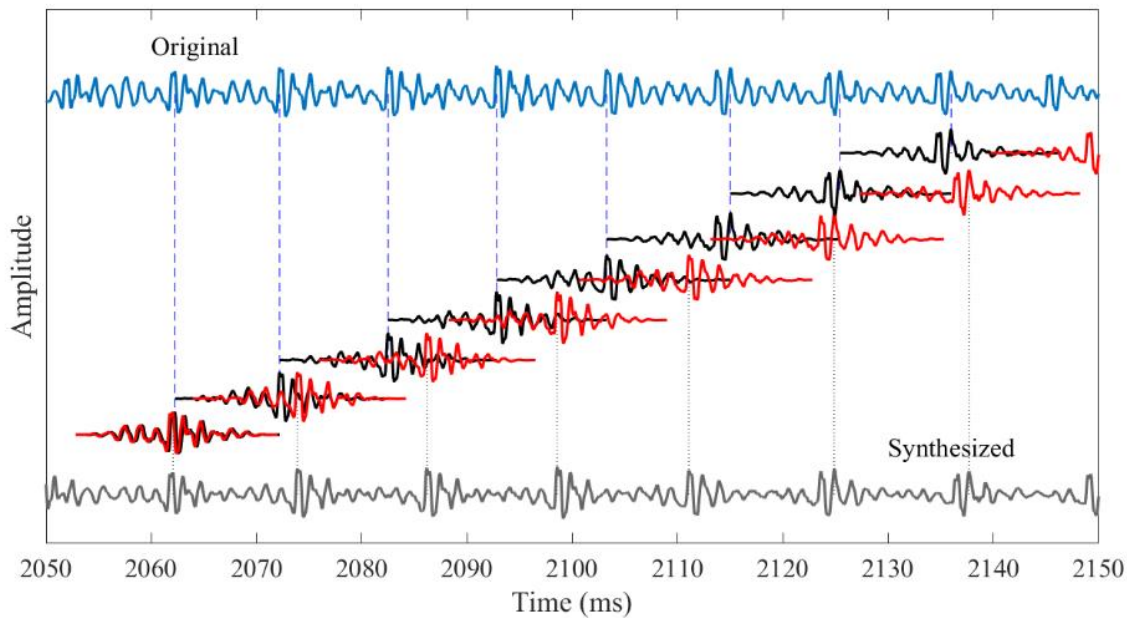
بسیاری از کاربردهای گفتاری به توانایی تغییر فرکانس اساسی نیاز دارند. برای یک کاربرد کلاسیک اما حاشیه ای، به عملکرد تنظیم خودکار که اغلب در پردازش پس از آواز خواندن استفاده می شود، فکر کنید. با چنین ابزارهایی می توان فرکانس اصلی صدای گوینده یا خواننده را بدون تغییر واج یا تن صدا تغییر داد. یکی از محبوب ترین ابزارهای توسعه یافته برای این منظور، هم پوشانی همزمان با زمین (PSOLA) است. همانطور که از نام آن پیداست، ارتباط نزدیکی با روش هم پوشانی-افزودن مورد استفاده در الگوریتم تبدیل فوریه کوتاه مدت دارد. این اجازه می دهد تا زیر و بمی صدای گفتار را بدون تغییر یا با تأثیر جزئی بر سایر ویژگی های سیگنال، مانند هویت صدا، تغییر دهید. علاوه بر تنظیم خودکار، یکی از کاربردهای مهم PSOLA سنتز گفتار است، که در آن ما می خواهیم بتوانیم گفتار را با هر خط خطی منطقی تولید کنیم. تبدیل صدا یک برنامه کاربردی دیگر است، که در آن هدف تبدیل گفتار یک شخص است، به طوری که مانند گفتار شخص دیگری به نظر برسد.

تصویری از فرآیند PSOLA. طول دوره و حداکثر اوج را در هر دوره بیابید.



ایده اصلی PSOLA این است که سیگنال را به دوره های زیر و بم جداگانه تجزیه کنیم، به طوری که ما بتوانیم دوره های گام را برای تغییر طول موثر آن دوره ها جابجا کنیم. یعنی فرکانس اساسی یک سیگنال به عنوان ساختار تناوبی سیگنال زمان بیان می شود. اگر سیگنال را به بخش مربوط به طول چنین ساختارهای تناوبی برش دهیم، می توانیم موقعیت آنها را به دلخواه تغییر دهیم و سپس آنها را دوباره به هم اضافه کنیم، مانند فرآیند همپوشانی-افزود. از آنجایی که همبستگی های کوتاه مدت در سیگنال تغییر نمی کند، یعنی سیگنال داخل پنجره ها/بخش ها تغییر نمی کند، پس پوشش طیفی سیگنال تغییر نمی کند.

پنجره سازی تجزیه و تحلیل PSOLA، تغییر زمان و پنجره سازی سنتز



برای نشان دادن این اصل، الگوریتم اصلی زیر را در نظر بگیرید:

1. کانتور فرکانس اساسی یک نمونه گفتار را تخمین بزنید.

2. دوره های زیر و بمی نمونه گفتار را پیدا کنید، برای مثال با شناسایی بزرگترین اوج در هر دوره.

3. پنجره های سیگنال گفتار را که دو دوره تناوب را پوشش می دهند استخراج کنید. عملکرد پنجره سازی را با بازسازی کامل اعمال کنید. (توجه داشته باشید: بازسازی کامل باید برای هر دوره اعمال شود، بنابراین ما برای هر دوره پنجره های نیمه بلند می سازیم. برعکس، قسمت های چپ و راست پنجره ها می توانند طول متفاوتی داشته باشند.)

4. پنجره ها را تغییر دهید تا با طول دوره گام مورد نظر مطابقت داشته باشد.

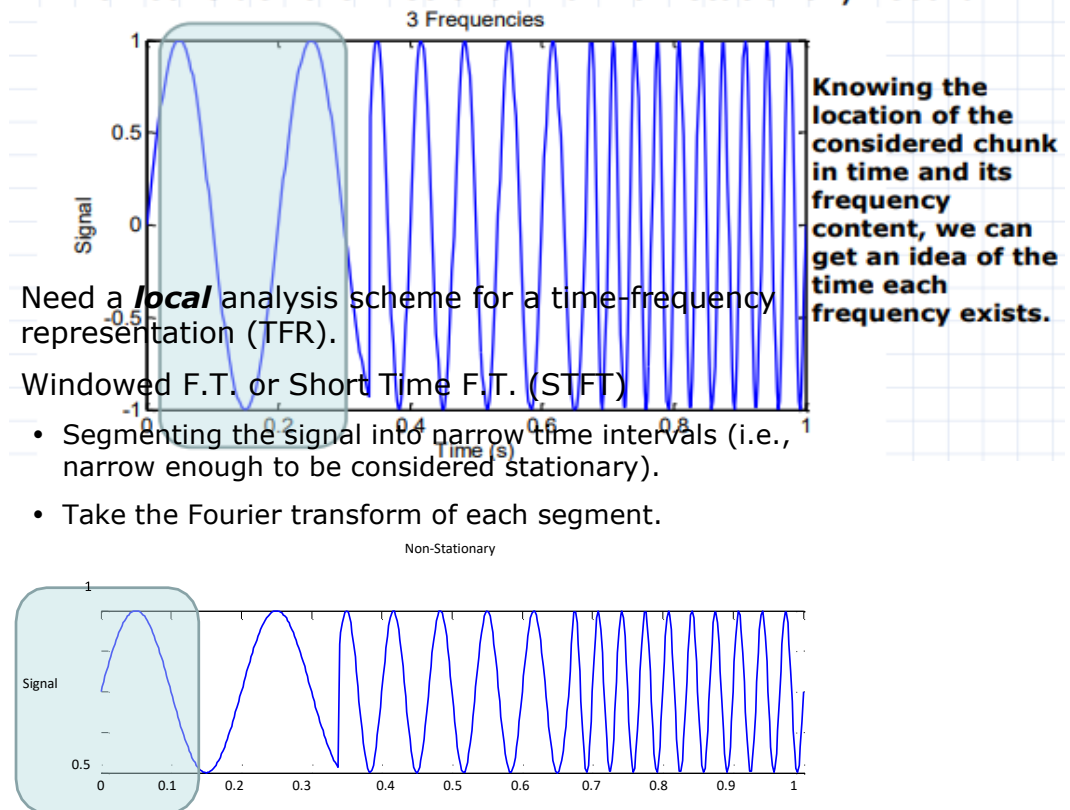
در مثال های صوتی سمت راست، طول دوره تناوب توسط یک ضریب ثابت تنظیم می شود تا فرکانس اصلی را افزایش یا کاهش دهد. توجه داشته باشید که پیاده سازی کاملاً تنظیم نشده است به طوری که صدای خروجی دارای برخی اعوجاج های شنیداری باشد.

## Short-Time Fourier Transform

- Basic Concept:
  - Break up the signal in time domain to a number of signals of shorter duration, then transform each signal to frequency domain
  - - Requires fewer number of harmonics to regenerate the signal chunks

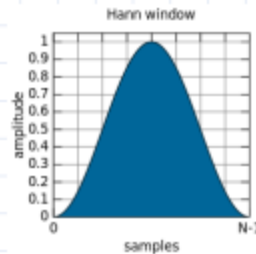
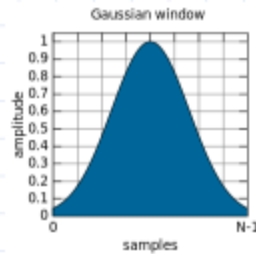
- Or consider the first chunk of non-stationary record

- Need a **local** analysis scheme for a time-frequency representation (TFR).
- Windowed F.T. or Short Time F.T. (STFT)
  - Segmenting the signal into narrow time intervals (i.e., narrow enough to be considered stationary).
  - Take the Fourier transform of each segment.



## Steps:

- Choose a window function of finite length
  - A window function is a function that is multiplied by the signal to keep a certain portion of it

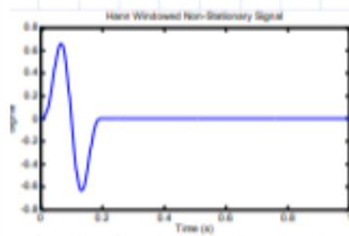


- Place the window on top of the signal at  $t=0$
- Truncate the signal using this window

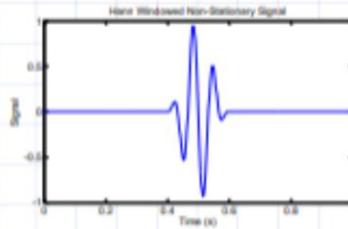
## • Steps (continued)

- Compute the FT of the truncated signal, save results.
  - For each time location where the window is centered, we obtain a different FT
    - Each FT provides the spectral information of a separate time-slice of the signal, providing **simultaneous** time and frequency information
- Incrementally slide the window to the right
- Repeat until window reaches the end of the signal

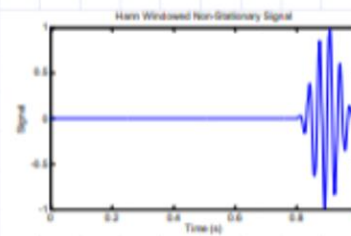
- What is being Fourier-transformed: **Window Length: 0.2s**



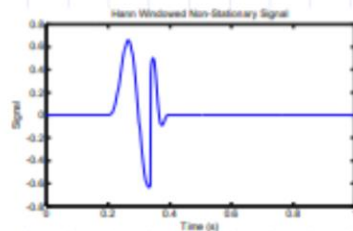
$\tau = 0.1s$



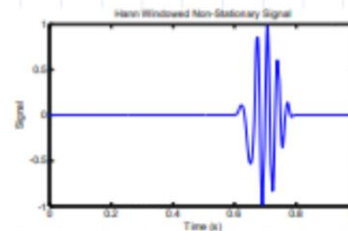
$\tau = 0.5s$



$\tau = 0.9s$

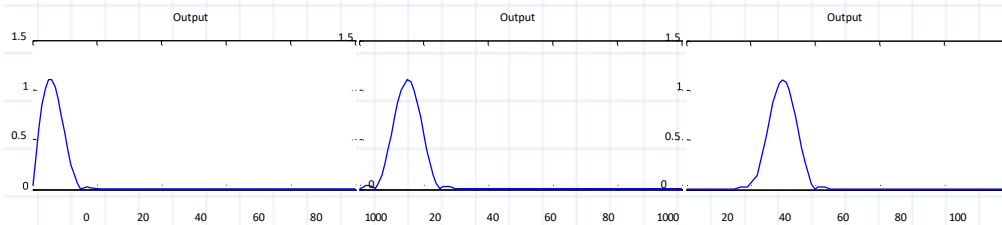


$\tau = 0.3s$



$\tau = 0.7s$

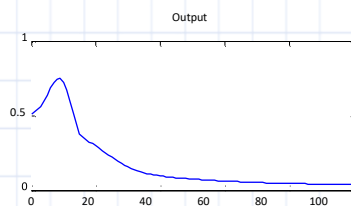
- Fourier transforms of windowed signal:



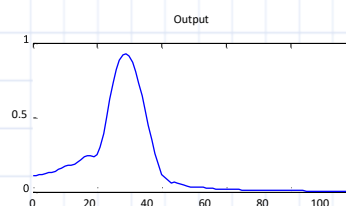
$\tau = 0.1s$

$\tau = 0.5s$

$\tau = 0.9s$



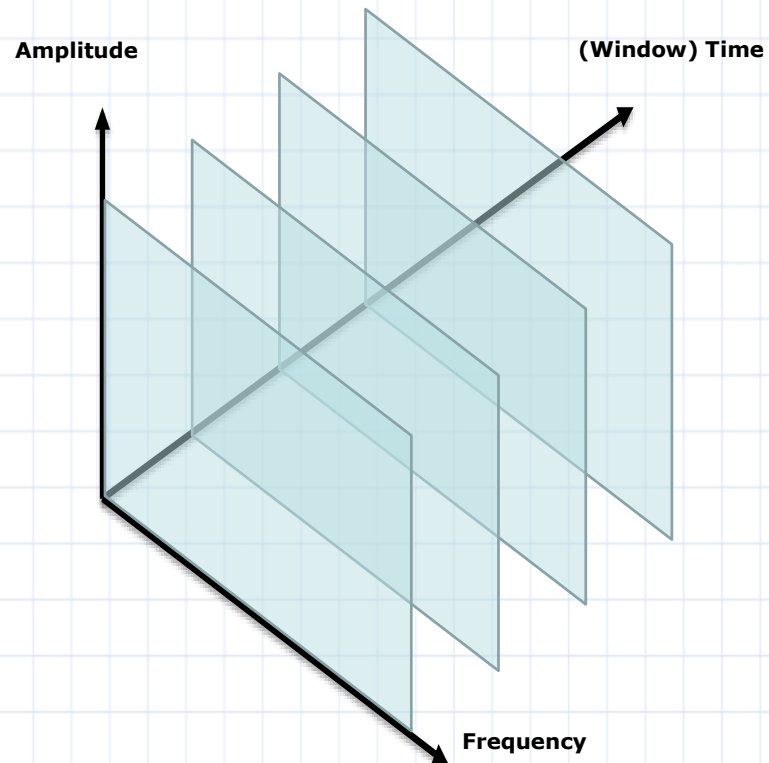
$\tau = 0.3s$



$\tau = 0.7s$



- Now we can plot the spectra next to each other to generate a ***surface***:



## 2- کم کردن گام صدا و افزایش گام صدا با استفاده از PSOLA :

### : TD\_PSOLA

1. The analysis step, where the original speech signal is first divided into separate but often overlapping short-term analysis signals (ST). Short term signals  $x_m(n)$  are obtained from the digital speech waveform  $x(n)$  by multiplying the signal by a sequence of the pitch-synchronous analysis window  $hm(n)$  as in Eq. 1:

$$X_m(n) = h_m(t_m - n)x(n), \quad (1)$$

where  $m$  is an index for the short-time signal

برای استفاده از این الگوریتم باید حواسمان باشد که ایندکس های سمپل در جای خود قرار گیرند و همچنین محتوای فرکانسی صدا ها برهم نریزد.

برای این منظور از تکه کد زیر استفاده می کنیم :

```
% Calculate the new length of the signal based on the pitch factor
new_length = round(numel(input_wave) / pitch_factor);
% Calculate the new length of the output signal based on the pitch factor and the length of the input signal

% Initialize the output signal
output_wave = zeros(new_length, 1);
% Create an empty array of zeros to store the modified output signal

% Implement the PSOLA method
for i = 1:new_length
    % Calculate the original index using the pitch factor
    original_index = min(round(i * pitch_factor), numel(input_wave));
    % Calculate the corresponding index in the original signal based on the pitch factor and the current index of the output signal

    % Copy the sample from the original signal to the modified signal
    output_wave(i) = input_wave(original_index);
    % Copy the sample from the original signal to the modified output signal at the current index
end
```

در این تکه کد برحسب Pitch\_Factor ایندکس های متفاوتی از همان اطلاعات اولیه را ذخیره میکنیم.

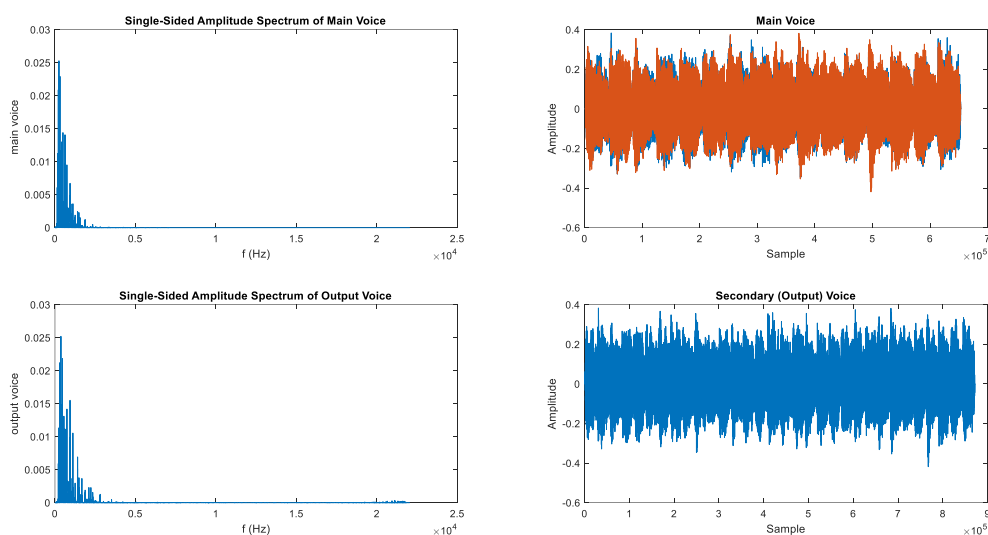
هرچه Pitch\_Factor بزرگتری داشته باشیم، صدا زیر تر (دارای فرکانس بیشتر) می شود و هر چه کوچک تر باشد(نباید از محدوده کمتر شود ) صدا بم تر می شود. (فرکانس کمتر و Pitch Period بیشتر)

برای اعمال PSOLA از 5 فایل سمپل استفاده می کنیم و PSOLA را یک بار برای افزایش گام صدا و بار دیگر برای کاهش گام صدا استفاده می کنیم.

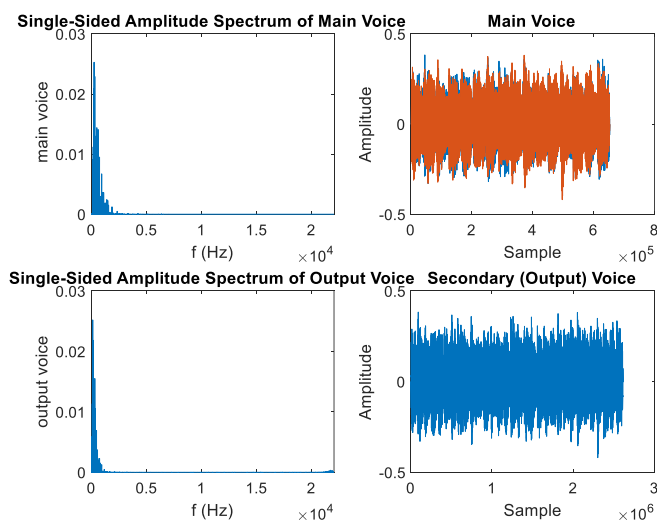
برای هر فایل سمپل، نمودار طیف فرکانسی آن و همچنین نمودار سیگنال صدا در حوزه زمان را رسم می کنیم. بنابراین داریم :

برای فایل اول (test1.mp3) داریم :

افزایش گام:

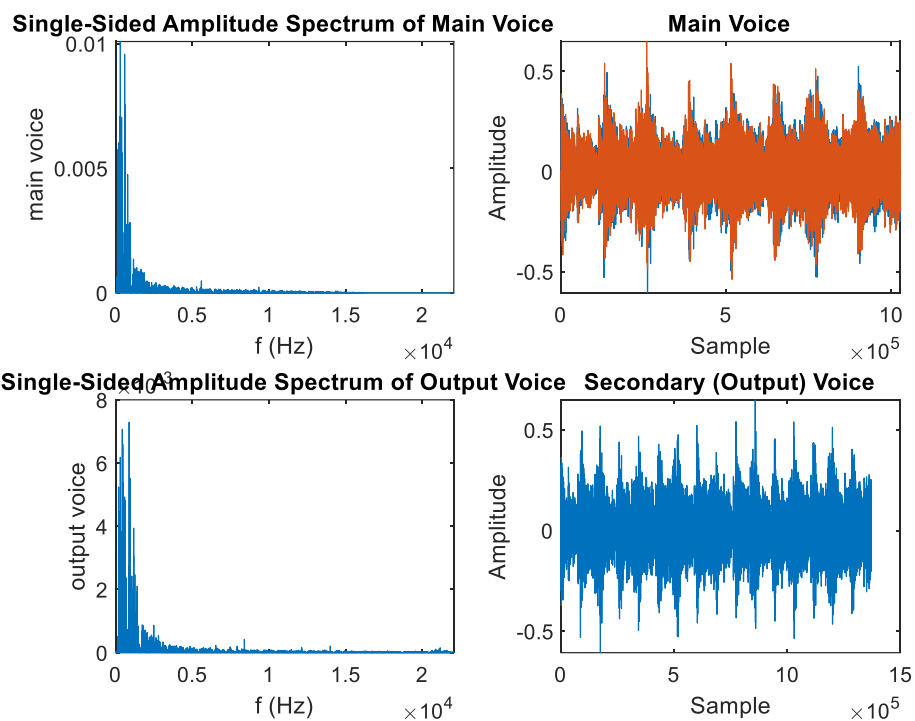


کاهش گام:

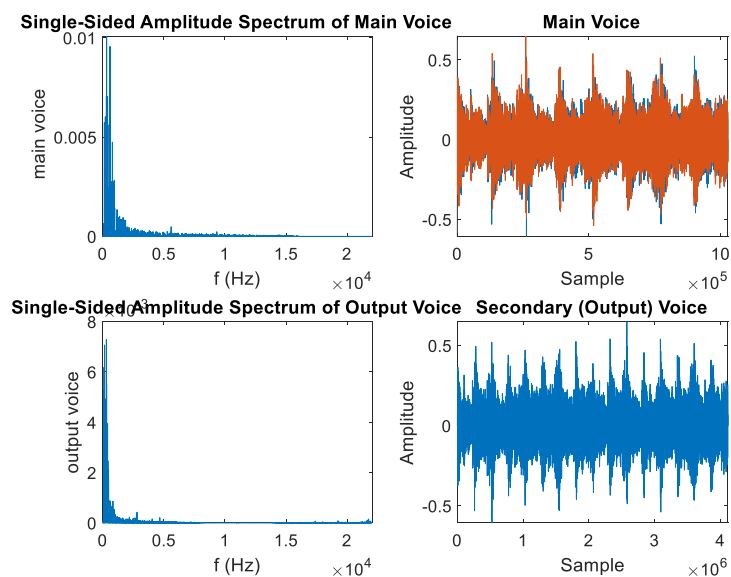


\*برای فایل دوم (test2.mp3) داریم :

افزایش گام:

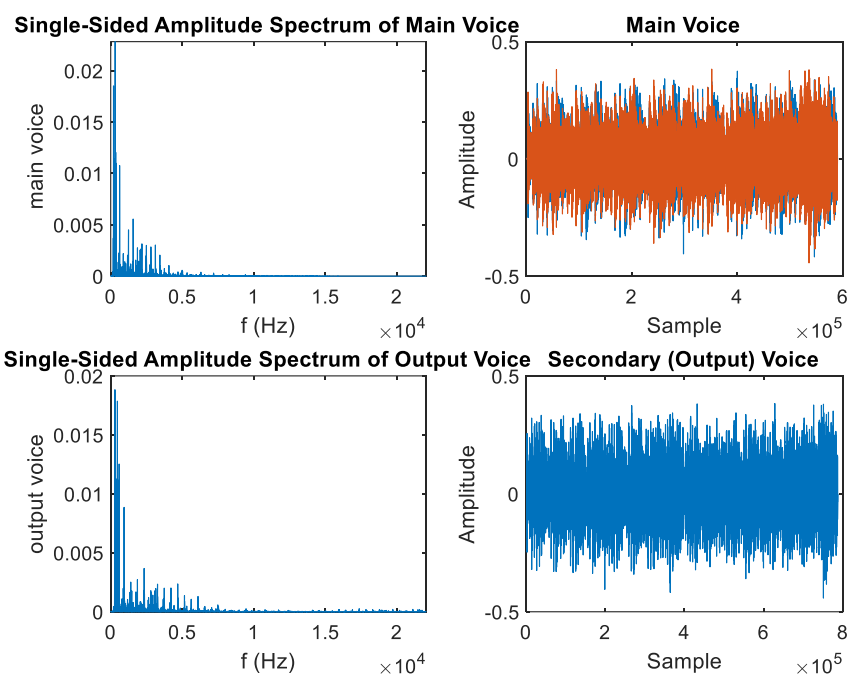


کاهش گام:

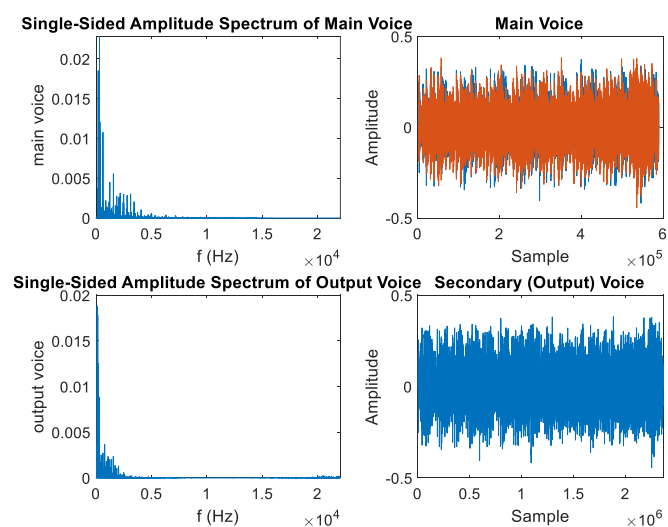


\*برای فایل سوم (test3.mp3) داریم :

افزایش گام:

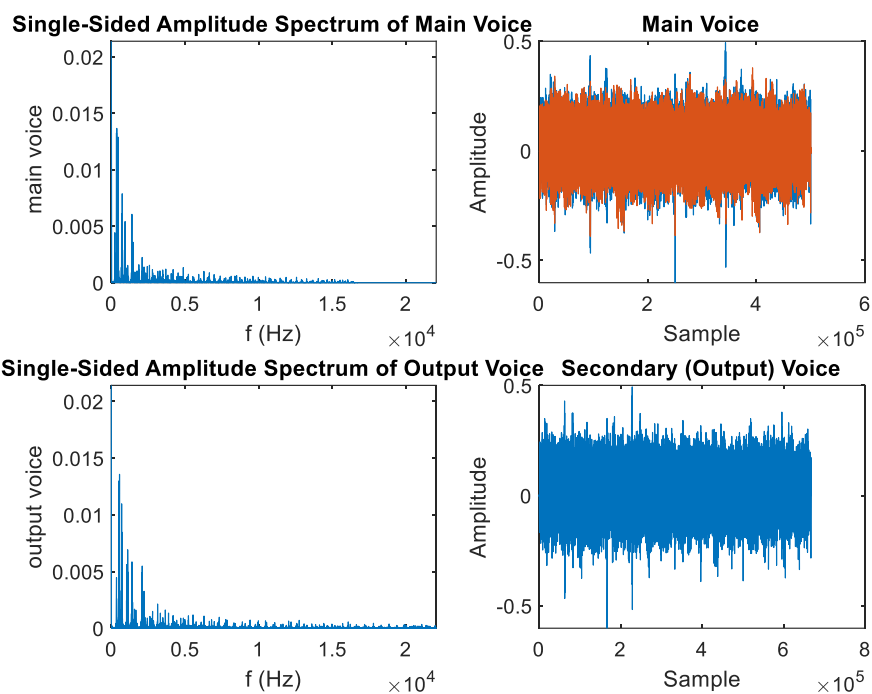


کاهش گام:

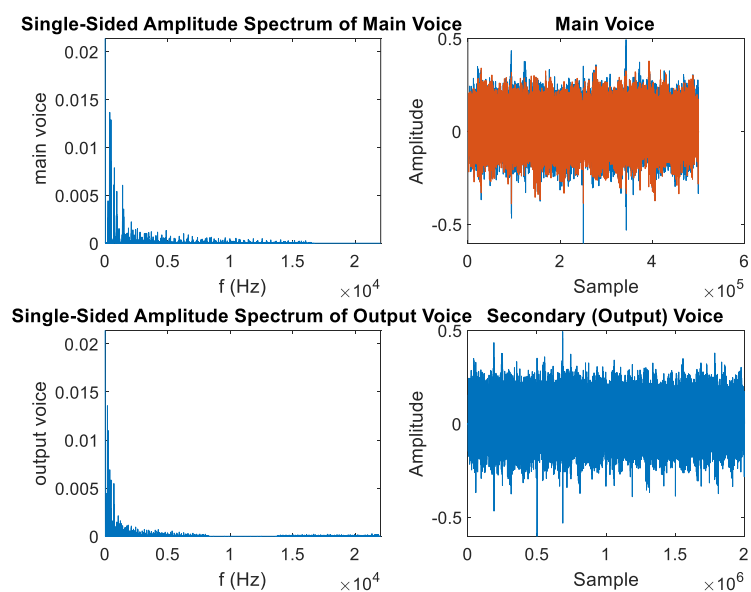


\*برای فایل چهارم (test4.mp3) داریم :

افزایش گام:

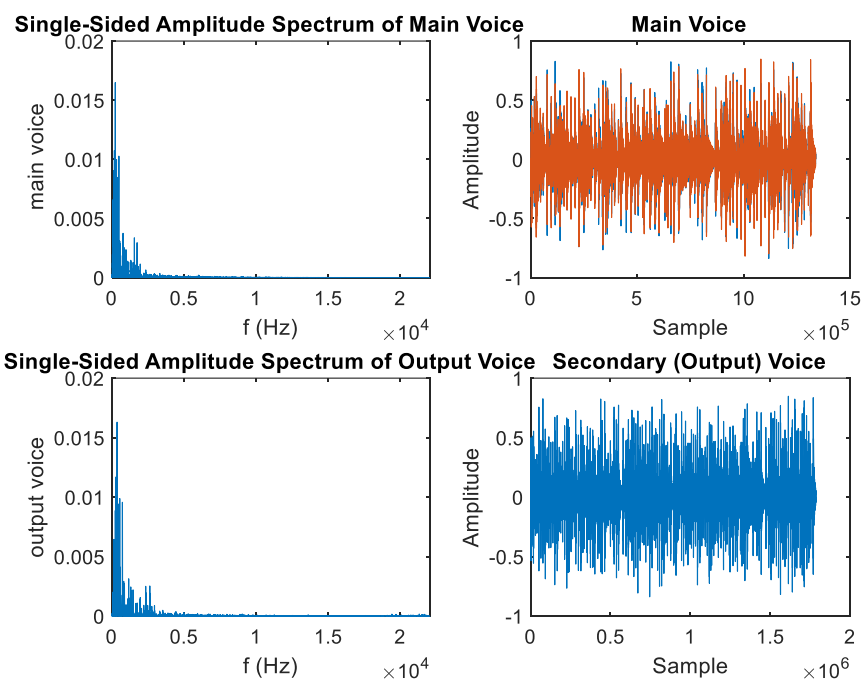


کاهش گام:

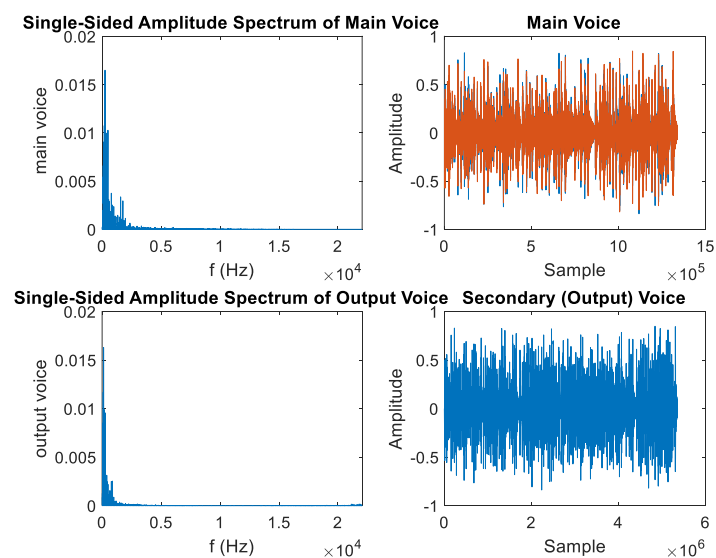


\*برای فایل پنجم (test5.mp3):

افزایش گام:



کاهش گام:



به طور کلی هر چه زمان موزیک بیشتر باشد باعث میشود در افزایش گام به ازای یک Pitch\_Factor ثابت، نسبت سمپل های موزیک افزایش گام یافته نسبت به اصلی بزرگتر باشد یا همین نسبت برای کاهش گام کمتر باشد که با استدلالی که در بالا گفتیم همخوانی دارد.

**\*خروجی های حاصل از فرایند افزایش گام صدا و کاهش گام صدا در فایل پیوست ذخیره شده است.**

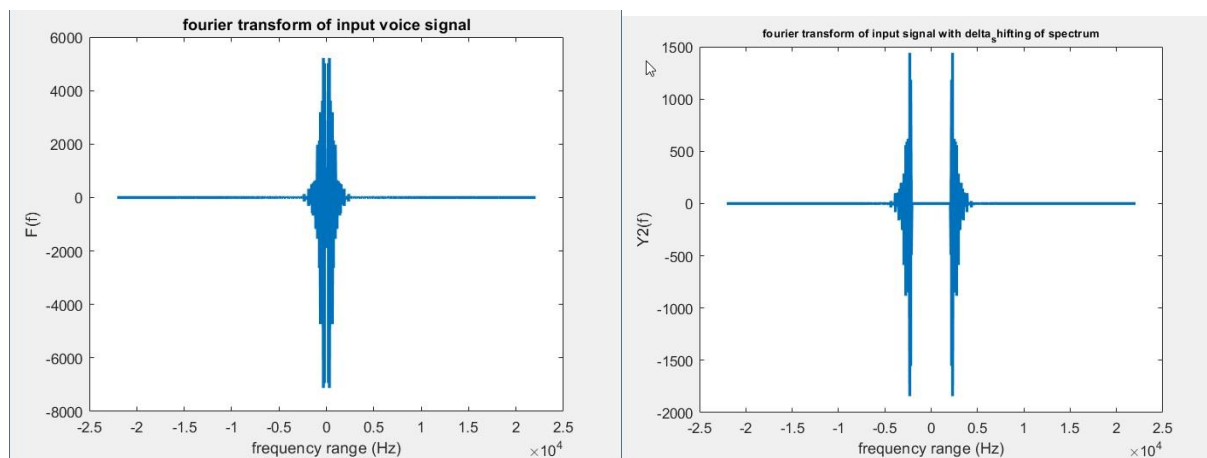
### 3- تغییر صدا با استفاده از شیفت فرکانسی :

فوریه می گیریم و سیگنال را به حوزه فرکانس انتقال می دهیم. حال برای شیفت فرکانسی سیگنال در حوزه فرکانس، سیگنال را در حوزه زمان در  $\cos(2\pi f_1 t)$  ضرب می کنیم. سپس سیگنال را از یک فیلتر HighPass عبور می دهیم، به طوریکه قسمتی از محتوای فرکانسی سیگنال اصلی را شامل شود و قسمت دیگری از محتوای فرکانسی حذف شود. پس از انتقال سیگنال از یک HighPass فیلتر، سیگنال را به حوزه زمان برمیگردانیم و سیگنال حاصل را در  $\cos(2\pi f_2 t)$  ضرب می کنیم. سپس سیگنال را از یک LowPass فیلتر عبور می دهیم. حال مشاهده می شود که طیف فرکانسی سیگنال حاصل به اندازه  $(f_1 - f_2)$  ، شیفت خورده طیف فرکانسی سیگنال اصلی می باشد. بنابراین با این الگوریتم توانستیم طیف فرکانسی سیگنال صوتی اولیه را به اندازه  $(\Delta = f_1 - f_2)$  شیفت دهیم.

برای اعمال این الگوریتم از 5 فایل سمپل استفاده می کنیم. برای هر کدام از فایل های سمپل طیف فرکانسی را به اندازه  $f=2000$  و  $f=5000$  و  $f=8000$  شیفت می دهیم. بنابراین داریم :

برای هر کدام از سمپل ها طیف فرکانسی سیگنال صوت اصلی و طیف فرکانسی شیفت خورده ان ها به اندازه  $f=2000$  و  $f=5000$  و  $f=8000$  و همچنین سیگنال صوت اصلی در حوزه زمان را رسم میکنیم. بنابراین داریم :

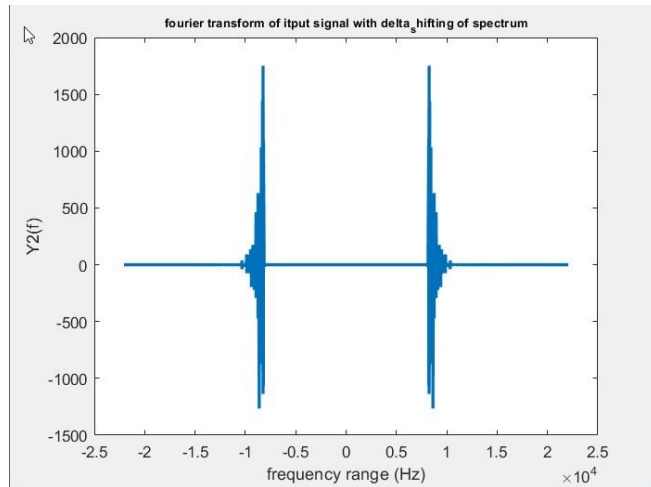
**برای فایل اول (test1.mp3) داریم :**



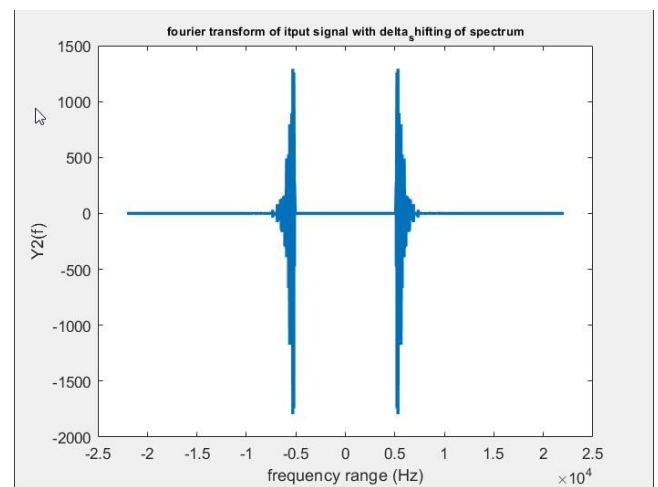
نمودار طیف فرکانسی سیگنال صوت اصلی

نمودار شیفت فرکانسی به اندازه  $f=2000$

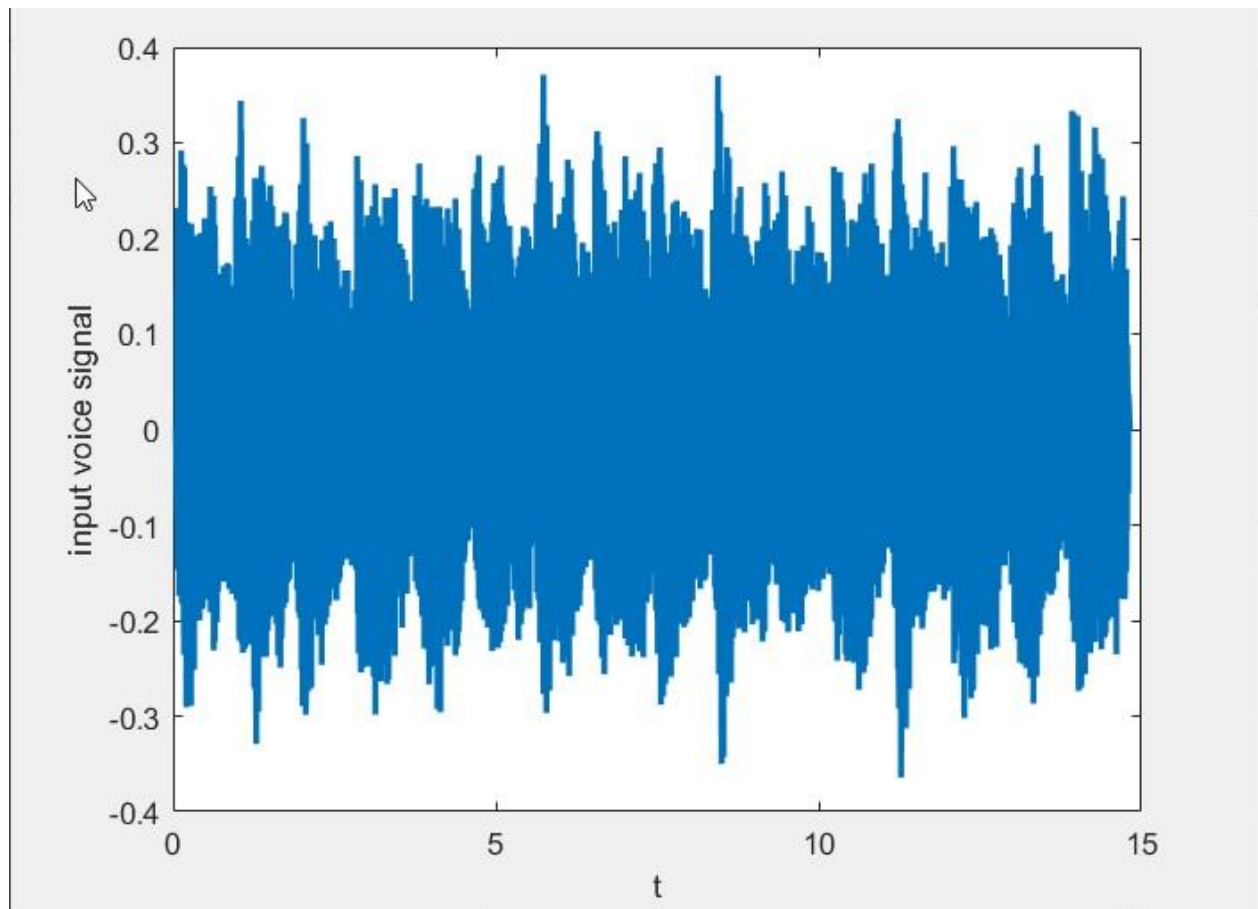




نمودار شیفت فرکانسی به اندازه  $f=8000$

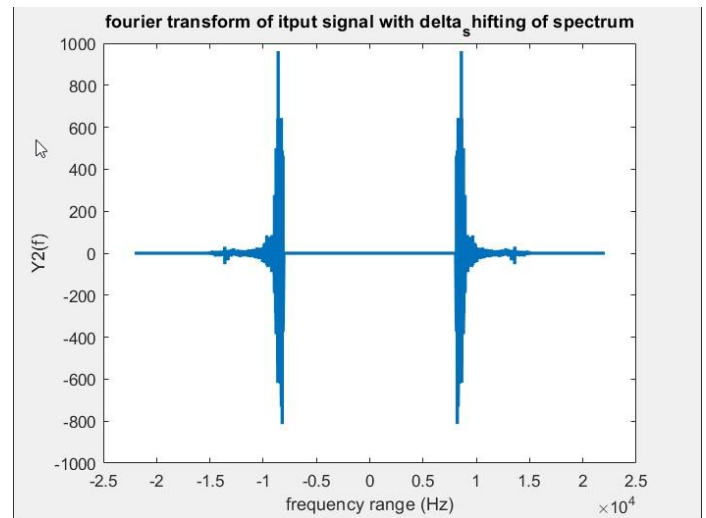
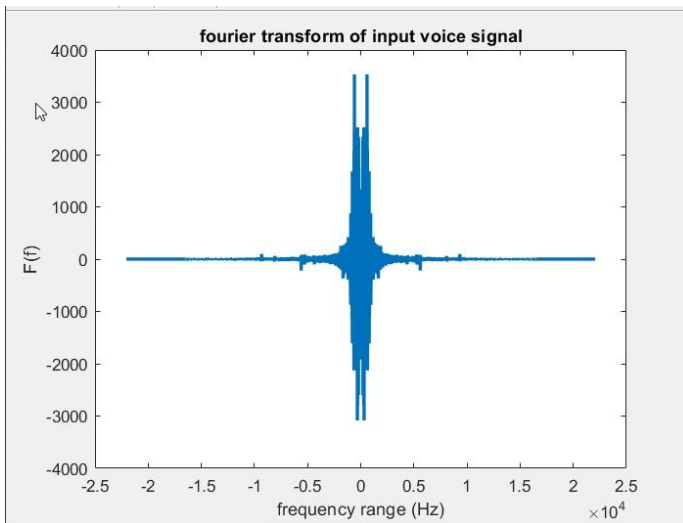


نمودار شیفت فرکانسی به اندازه  $f=5000$



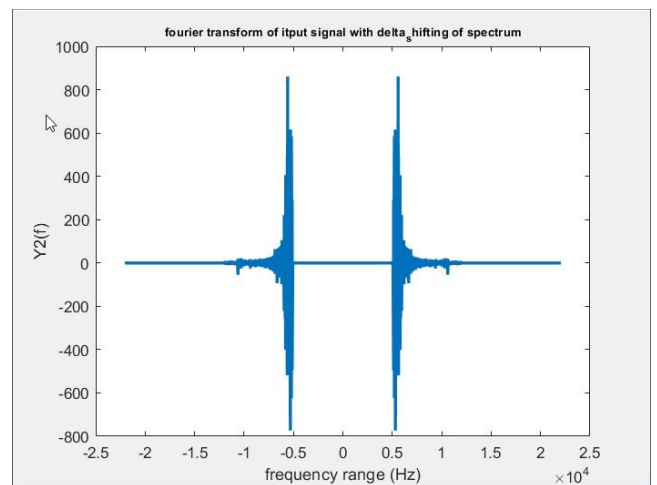
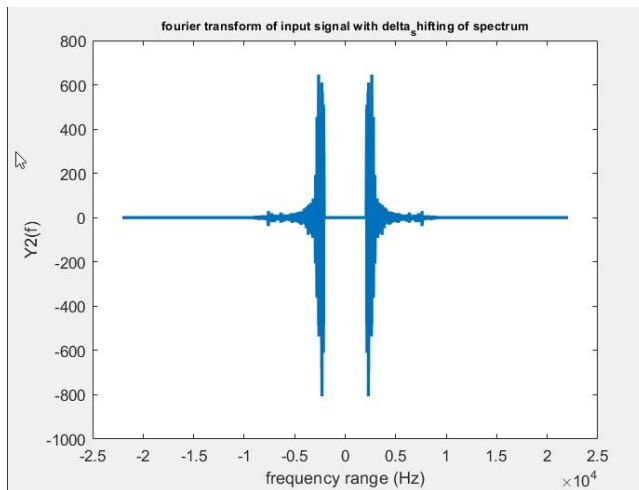
نمودار حوزه زمان سیگنال صوت اصلی

برای فایل دوم (test2.mp3) داریم :



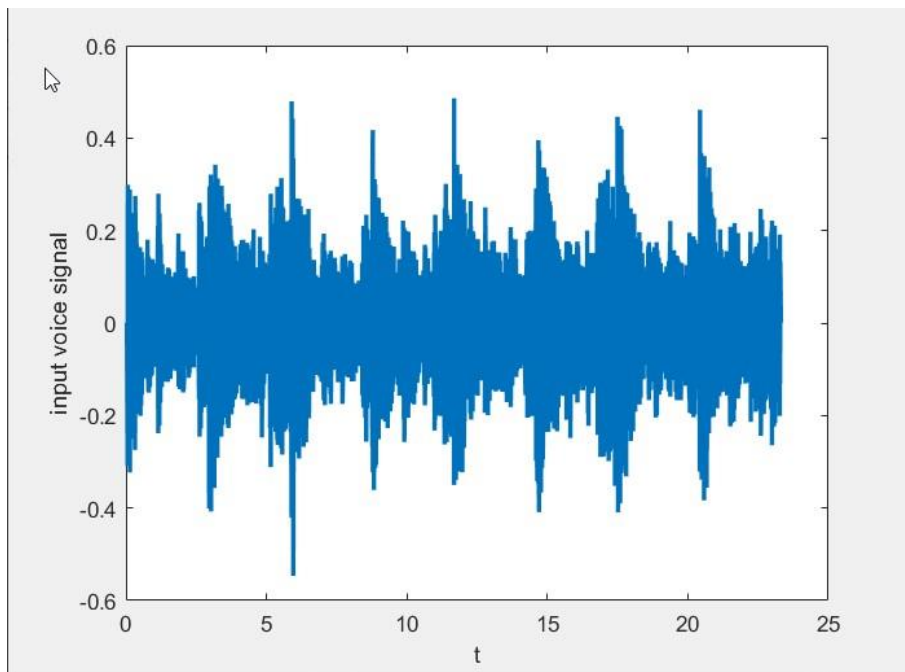
نمودار طیف فرکانسی سیگنال صوت اصلی

نمودار شیفت فرکانسی به اندازه  $f=8000$



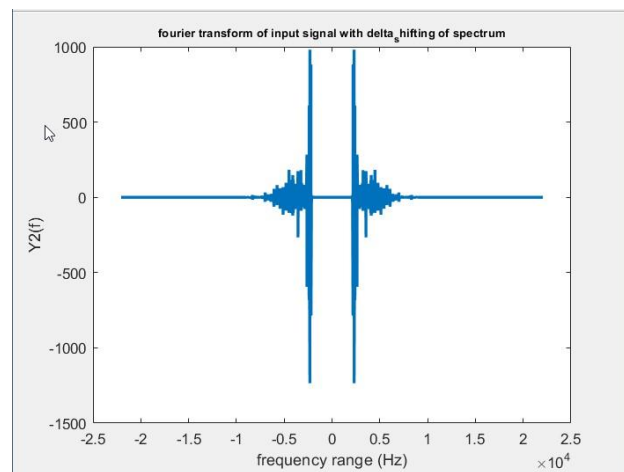
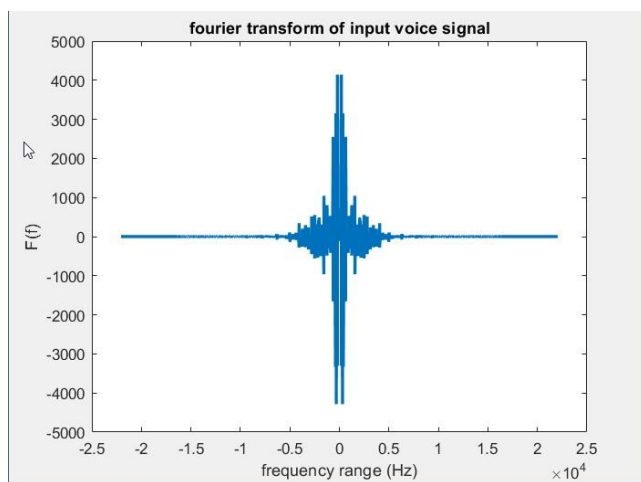
نمودار شیفت فرکانسی به اندازه  $f=2000$

نمودار شیفت فرکانسی به اندازه  $f=5000$



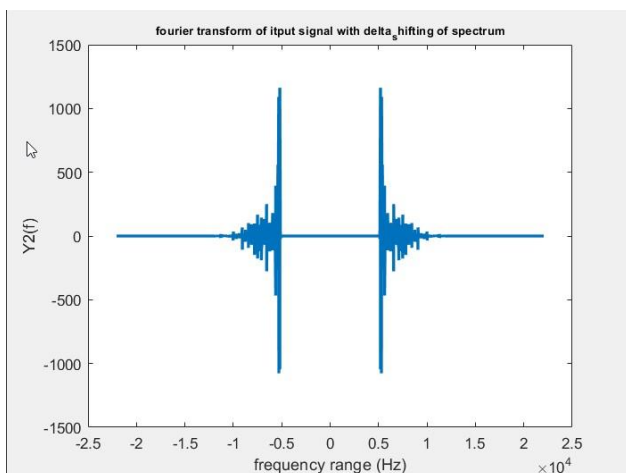
نمودار حوزه زمان سیگنال صوت اصلی

برای فایل سوم (test3.mp3) داریم :

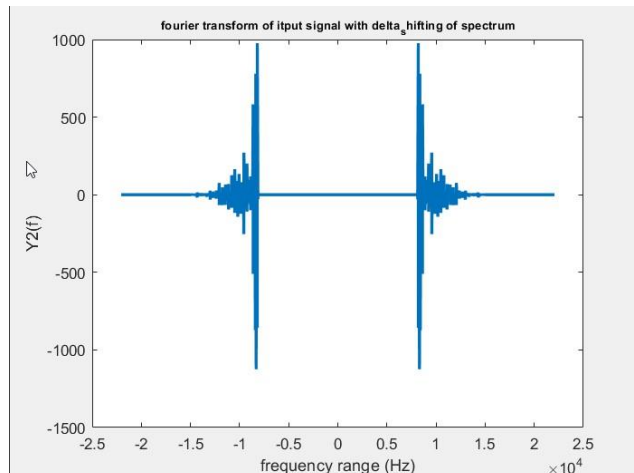


نمودار شیفتر فرکانسی به اندازه  $f=2000$

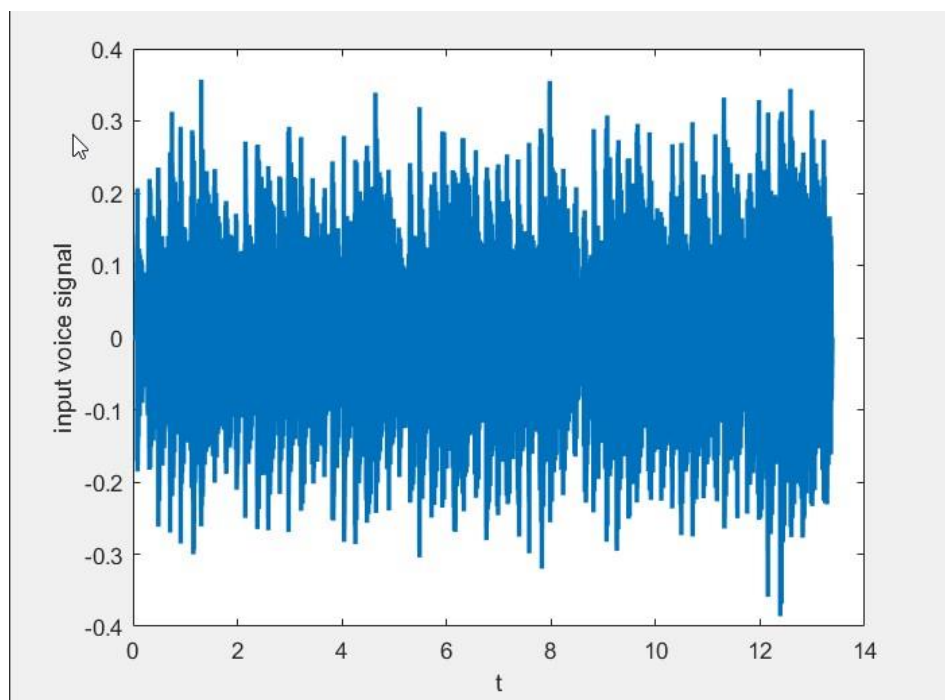
نمودار طیف فرکانسی سیگنال صوت اصلی



نمودار شیفت فرکانسی به اندازه  $f=5000$

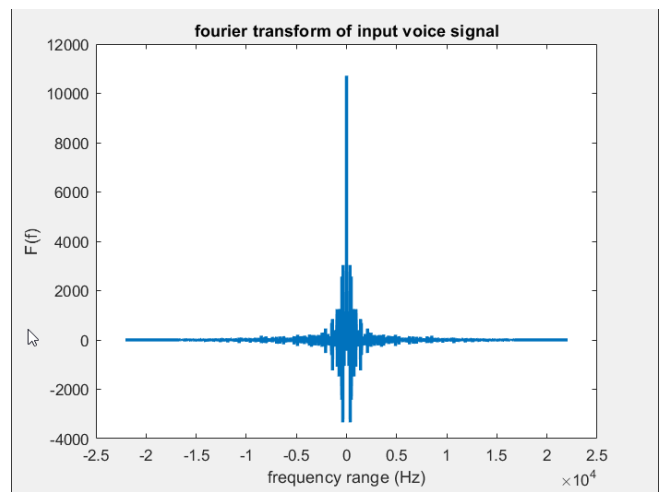
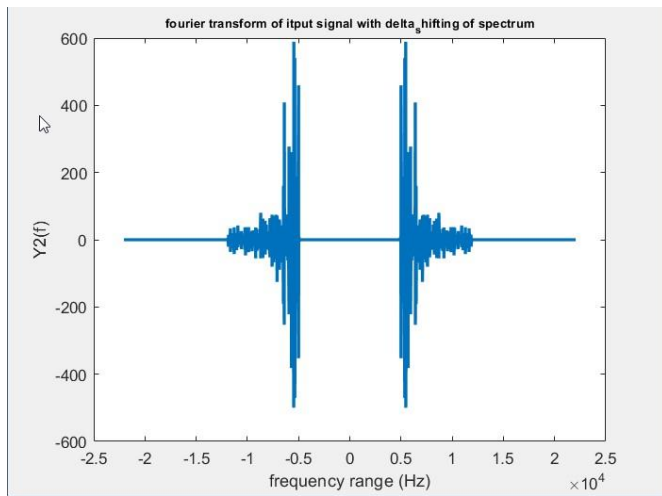


نمودار شیفت فرکانسی به اندازه  $f=8000$



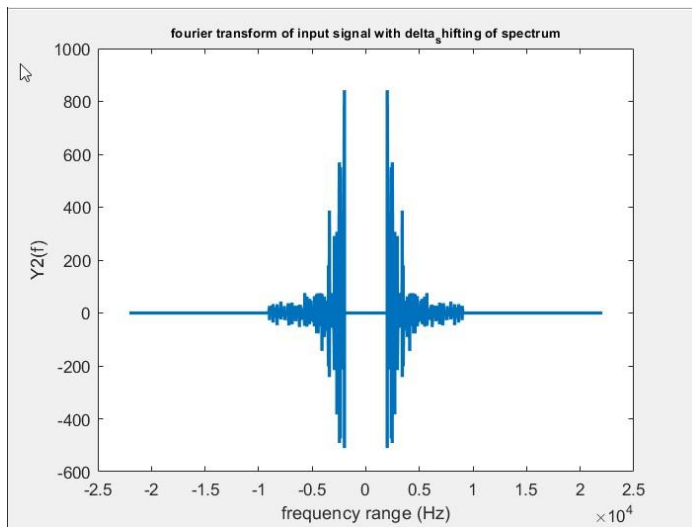
نمودار حوزه زمان سیگنال صوت اصلی

برای فایل چهارم (test4.mp3) داریم :

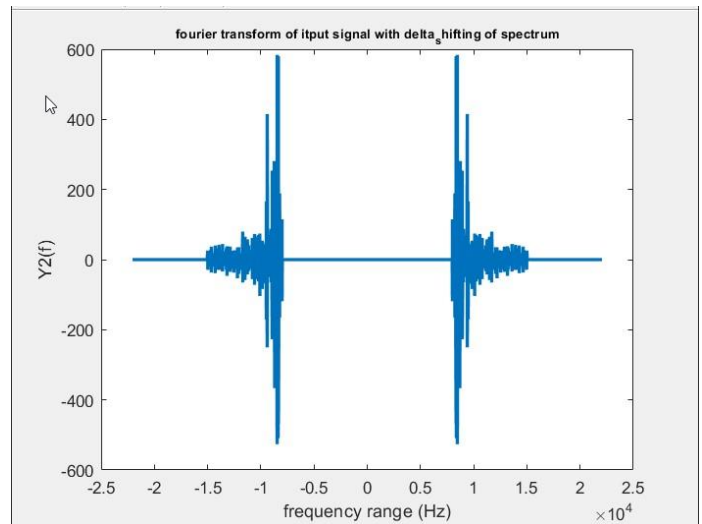


نمودار طیف فرکانسی سیگنال صوت اصلی

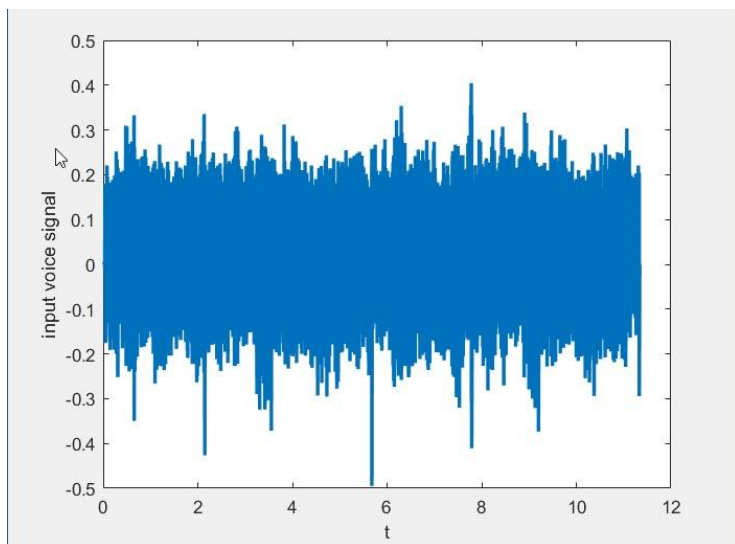
نمودار شیفت فرکانسی به اندازه  $f=5000$



نمودار شیفت فرکانسی به اندازه  $f=2000$

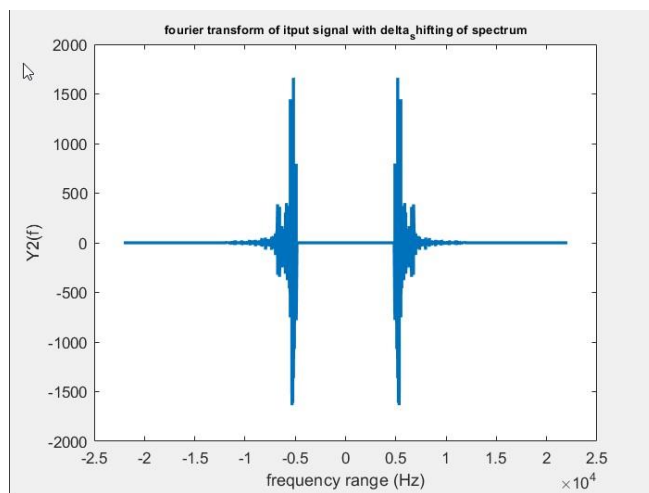
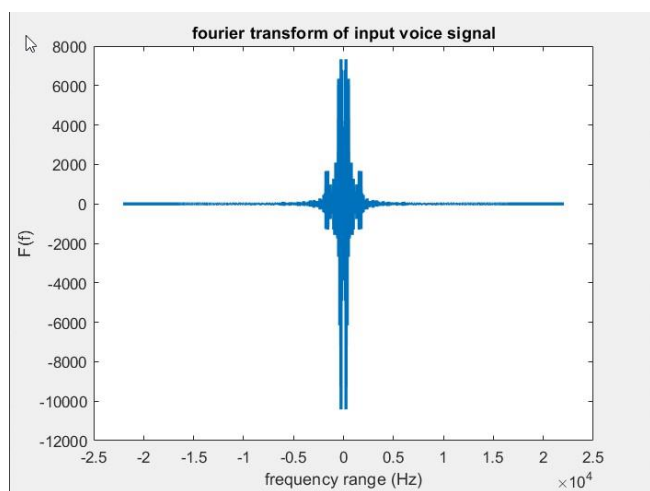


نمودار شیفت فرکانسی به اندازه  $f=8000$



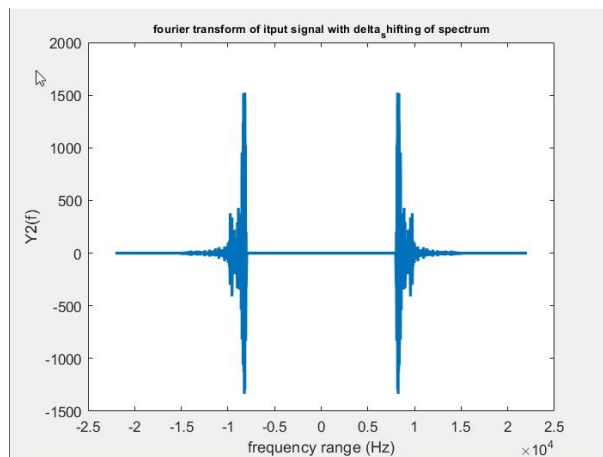
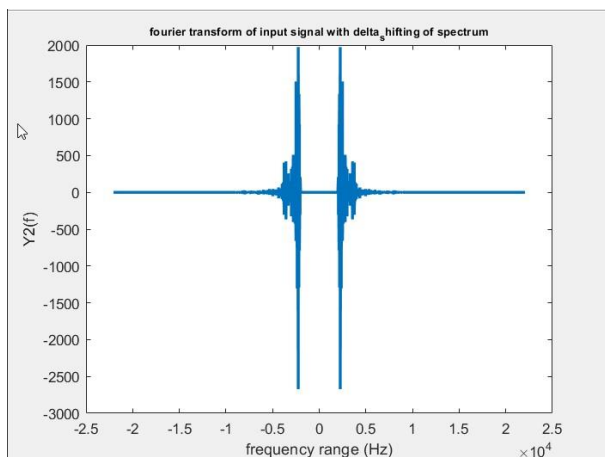
نمودار حوزه زمان سیگنال صوت اصلی

برای فایل پنجم (test5.mp3) داریم :



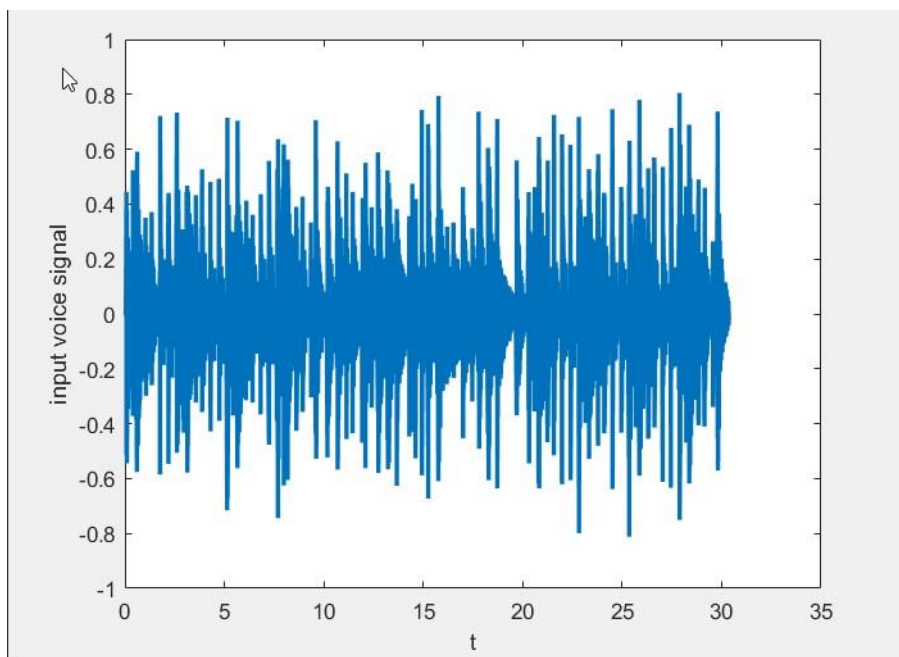
نمودار شیفت فرکانسی به اندازه  $f=5000$

نمودار طیف فرکانسی سیگنال اصلی



نمودار شیفت فرکانسی به اندازه  $f=8000$

نمودار شیفت فرکانسی به اندازه  $f=2000$



نمودار حوزه زمان سیگنال صوت اصلی

\*خروجی های حاصل از فرایند شیفت فرکانسی سیگنال صوت به ازای فرکانس های  $f=2000$  و  $f=5000$  و  $f=8000$  در فایل، پیوست شده است.