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

معکوسSTFT :

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:

STFTوضوح 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 تغییرمی کنند

• یک پنجره عریض (تبدیل باند پهن) وضوح فرکانس بهتری را ارائه می دهداما 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 - زیرااکنون 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 بخش‌های کوچک همپوشانی کار می‌کند. برای تغییر گام سیگنال، بخش‌ها بیشتر از هم فاصله می‌گیرند (برای کاهش گام) یا نزدیک‌تر 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 .

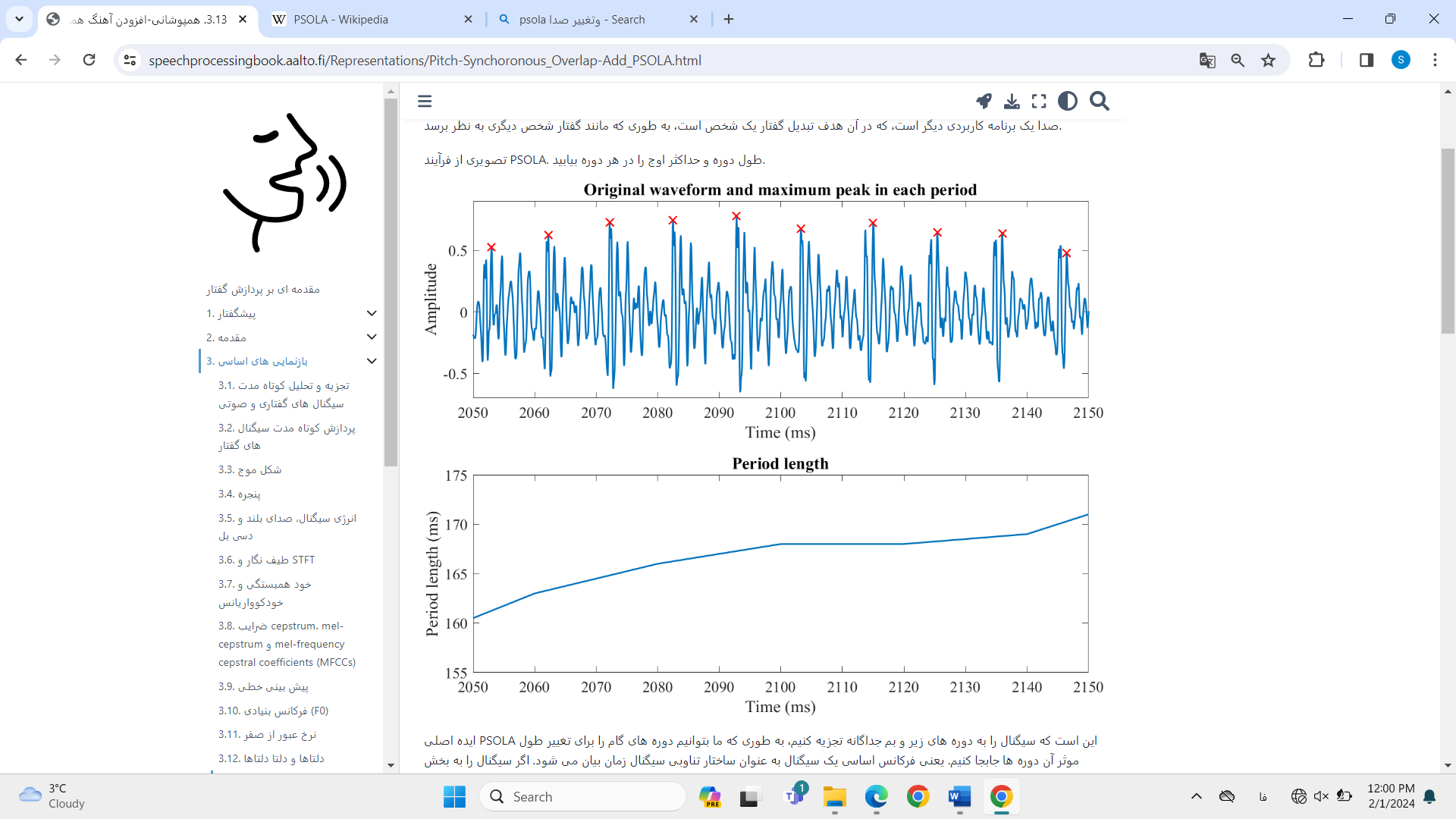
PSOLA می تواند برای تغییر عروض یک سیگنال گفتاری استفاده شود.

روش PSOLA یکی از روش‌های پردازش سیگنال صوتی است که برای تغییر فرکانس بندی یک صدای گفتاری به کار می‌رود.برای این کار، ابتدا باید فرکانس بندی صدای گفتاری را تخمین بزنیم. سپس با استفاده از روش PSOLA، سیگنال را به دوره‌های فرکانسی تجزیه می‌کنیم و با جابجایی دوره‌های فرکانسی، فرکانس بندی صدای گفتاری را تغییر می‌دهیم.

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

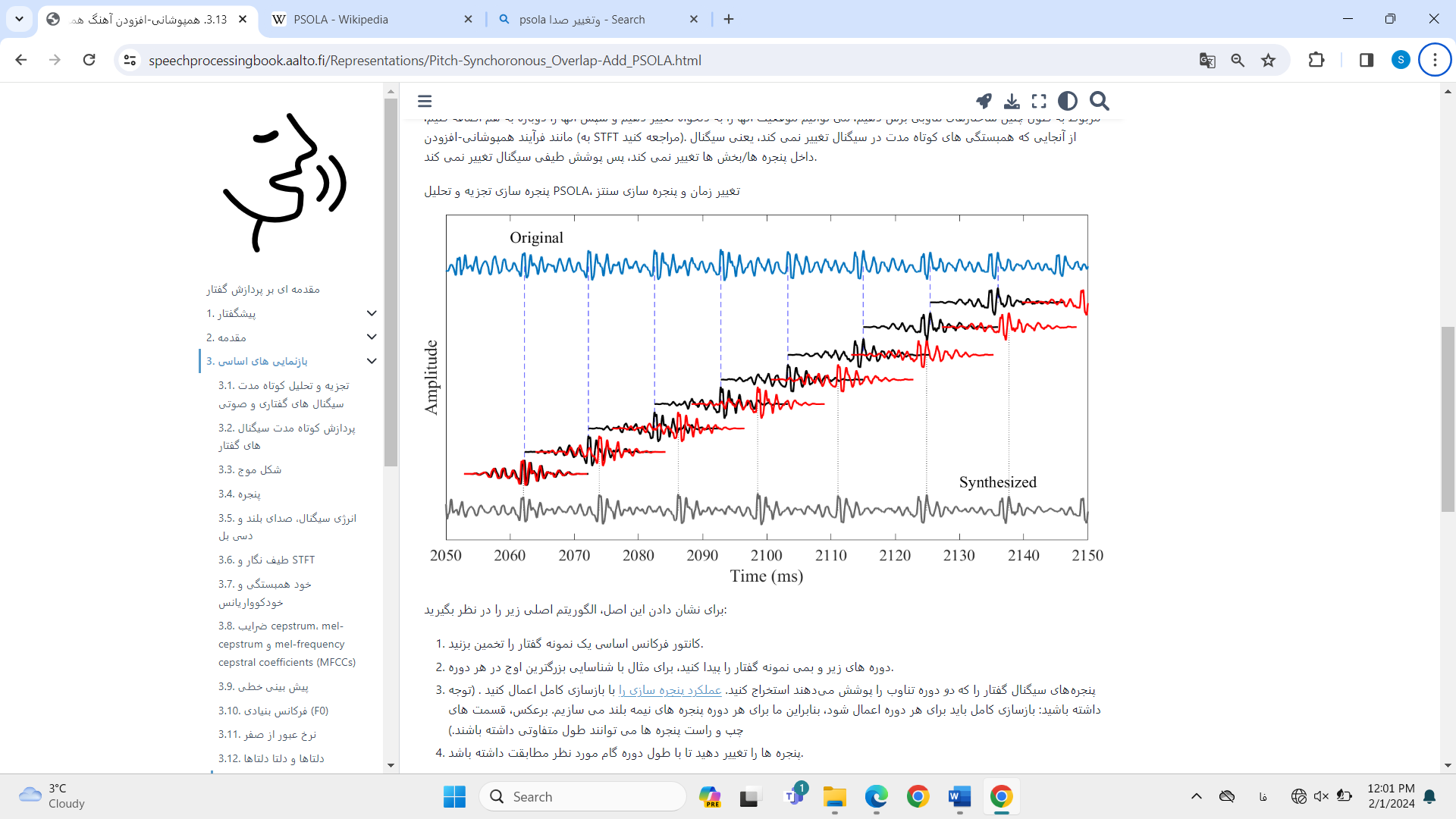
بسیاری از کاربردهای گفتاری به توانایی تغییر فرکانس اساسی نیاز دارند. برای یک کاربرد کلاسیک اما حاشیه ای، به عملکرد تنظیم خودکار که اغلب در پردازش پس از آواز خواندن استفاده می شود، فکر کنید. با چنین ابزارهایی می توان فرکانس اصلی صدای گوینده یا خواننده را بدون تغییر واج یا تن صدا تغییر داد. یکی از محبوب‌ترین ابزارهای توسعه‌یافته برای این منظور، هم‌پوشانی همزمان با زمین (PSOLA) است. همانطور که از نام آن پیداست، ارتباط نزدیکی با روش همپوشانی-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 صدادار، تغییر دهید. علاوه بر تنظیم خودکار، یکی از کاربردهای مهم PSOLA سنتز گفتار است ، که در آن ما می‌خواهیم 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.

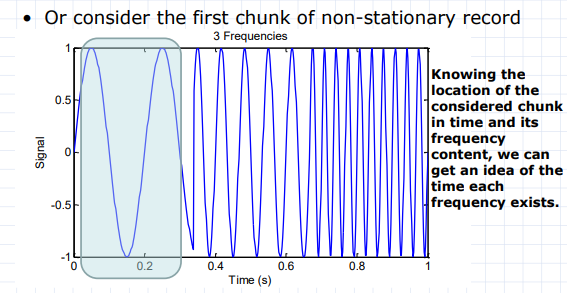
3. پنجره‌های سیگنال گفتار را که دو دوره تناوب را پوشش می‌دهند استخراج کنید. عملکرد پنجره سازی را با بازسازی کامل اعمال کنید . (توجه داشته باشید: بازسازی کامل باید برای هر دوره اعمال شود، بنابراین ما برای هر دوره پنجره های نیمه بلند می سازیم. برعکس، قسمت های چپ و راست پنجره ها می توانند طول متفاوتی داشته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

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

در مثال‌های صوتی سمت راست، طول دوره تناوب توسط یک ضریب ثابت تنظیم می‌شود 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



* 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.

Non-Stationary

1

Signal

0.5

0

-0.5

-1

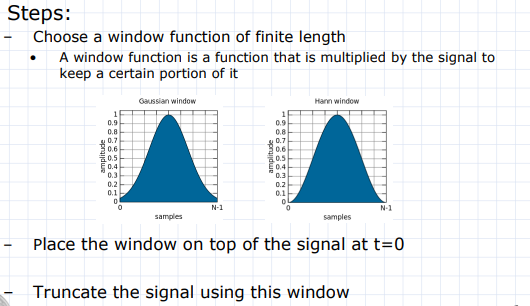
0 0.1 0.2 0.3

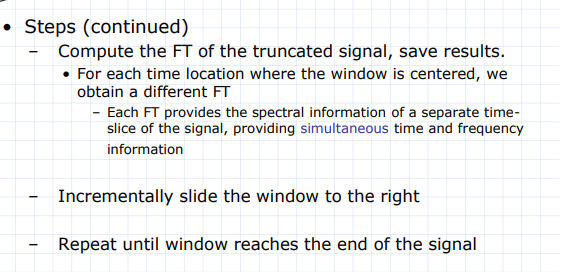
0.4 0.5 0.6

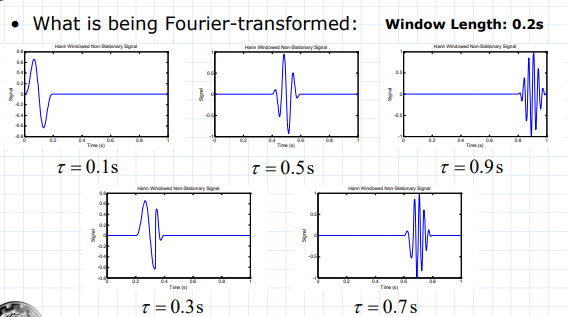
Time (s)

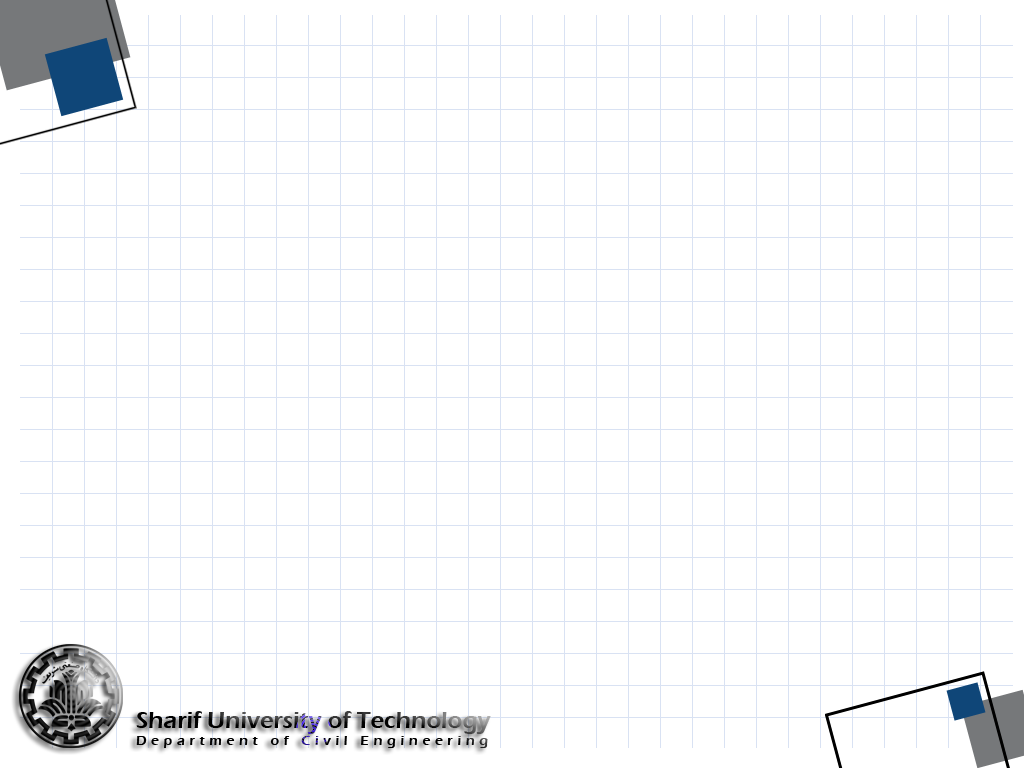
0.7 0.8 0.9 1

25









* Fourier transforms of windowed signal:

Output

Output

Output

1.5

1.5

1.5

1

1

0.5

0.5

0

0 0

0 20 40 60 80 1000 20 40 60 80 1000 20 40 60 80 100

 0.1s  0.5s  0.9 s

Output

Output

1

1

0.5

0.5

0

0

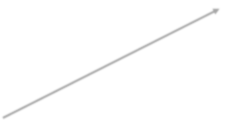
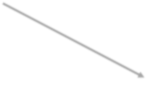
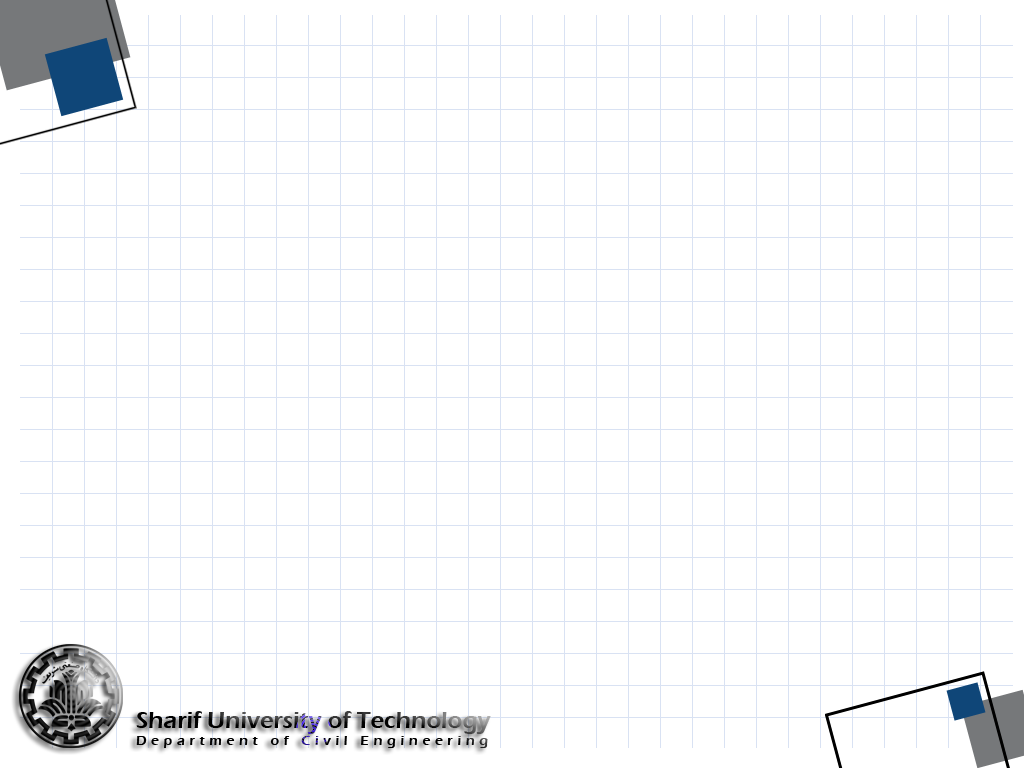
0 20 40 60 80 100

0 20 40 60 80 100

 0.3s

 0.7 s

29



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

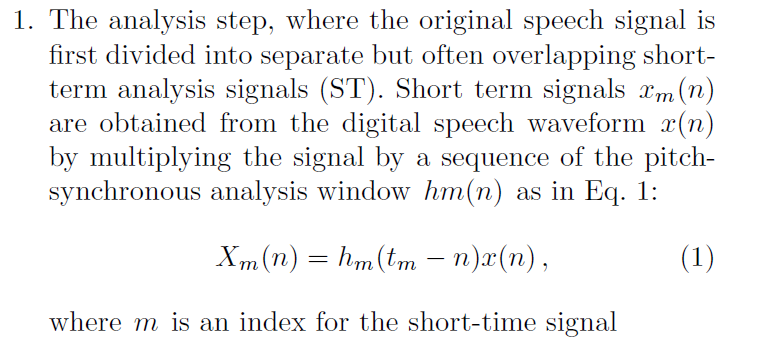
**Amplitude**

**(Window) Time**

**Frequency**

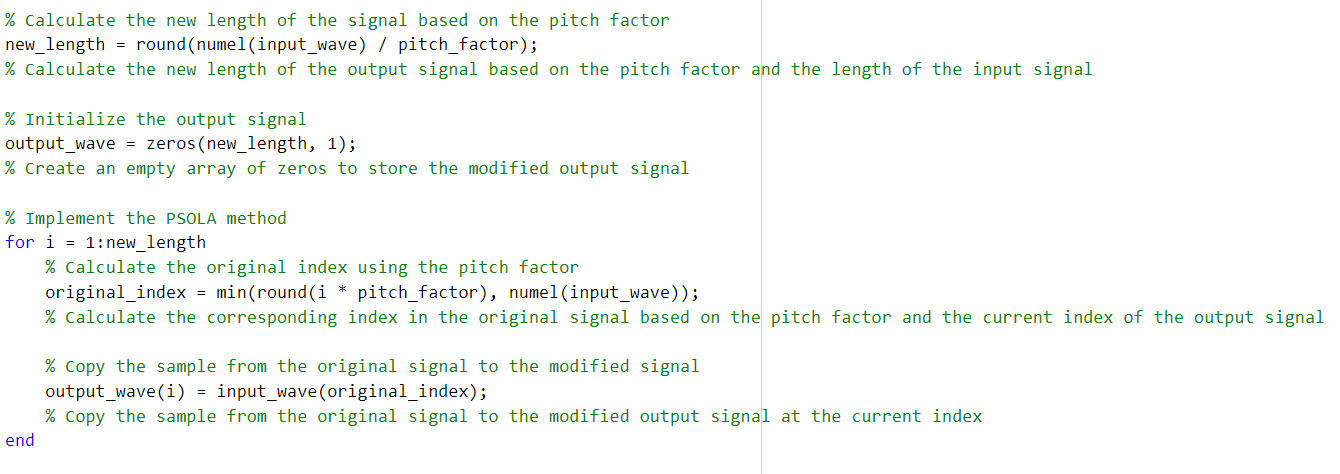
30

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

**TD\_PSOLA :**  

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 :



در این تکه کد برحسب Pitch\_Factor ایندکس های منتفاوتی 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 ان 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 موزیگ 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.

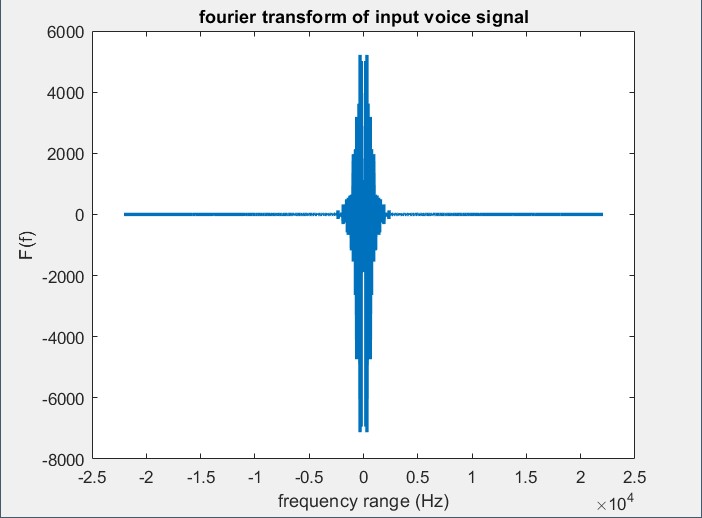
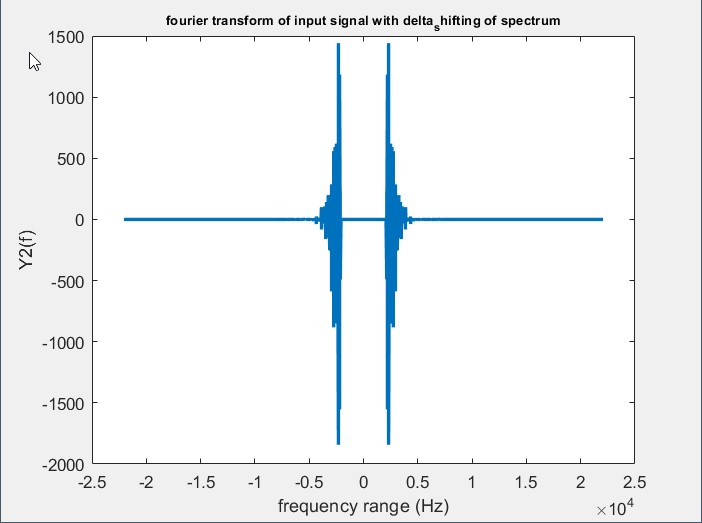
**\*خروجی های حاصل از فرایند افزایش گام صدا و کاهش کام صدا در فایل پیوست ذخیره شده است.**

**3-تغییر صدا با استفاده از شیفت فرکانسی** : در این الگوریتم ابتدا از سیگنال صوتی در حوزه زمان تبدیل 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\*f1\*t) ضرب می کنیم. سپس سیگنال را از یک فیلتر HighPass عبور می دهیم، به طوریکه قسمتی از محتواهای 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\*f2\*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 (f1-f2) ، شیفت خورده طیف فرکانسی سیگنال اصلی می باشد. بنابراین با این الگوریتم توانستیم طیف فرکانسی سیگنال صوتی اولیه را به اندازه =f1-f2 ) ∆ (شیفت دهیم.

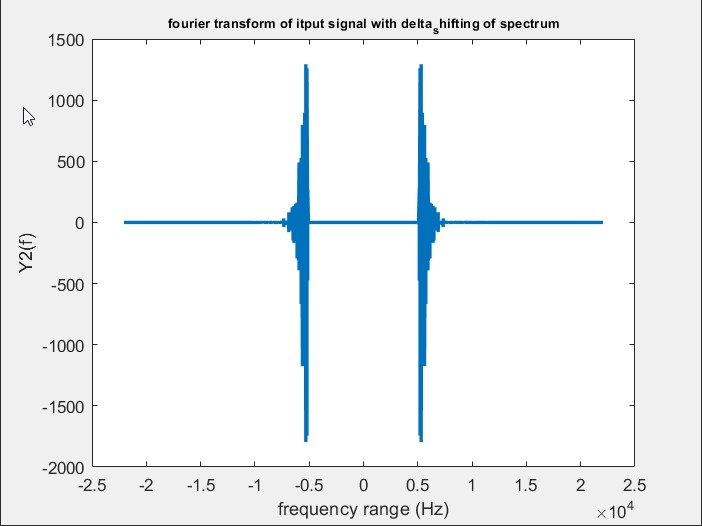
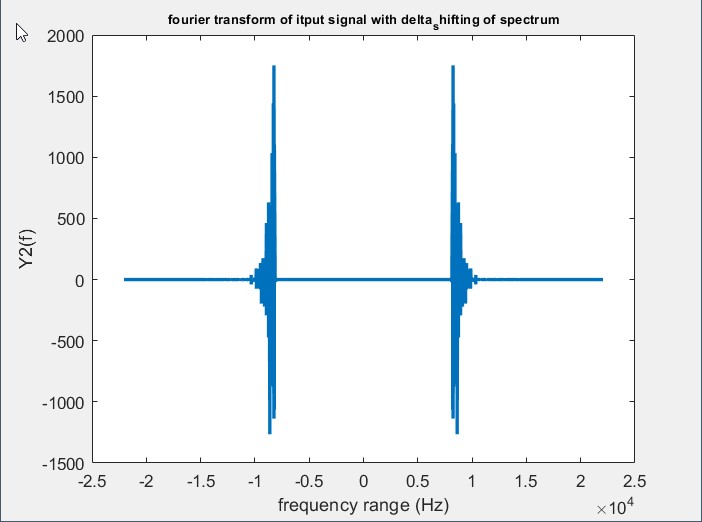
برای اعمال این الگوریتم از 5 فایل سمپل استفاده می کنیم. برای هر کدام از فایل های سمپل طیف فرکانسی را به اندازه f=2000 و f=5000 و f=8000 شیفت می دهیم. بنابراین داریم :

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 :

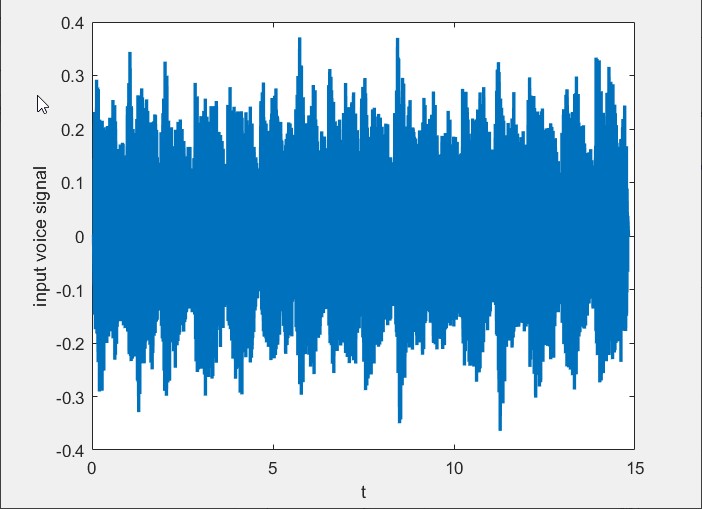
**برای فایل اول (test1.mp3) داریم :**

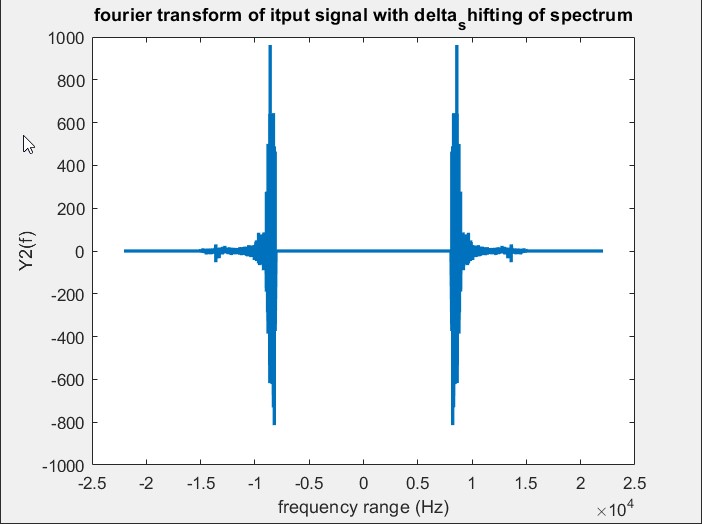
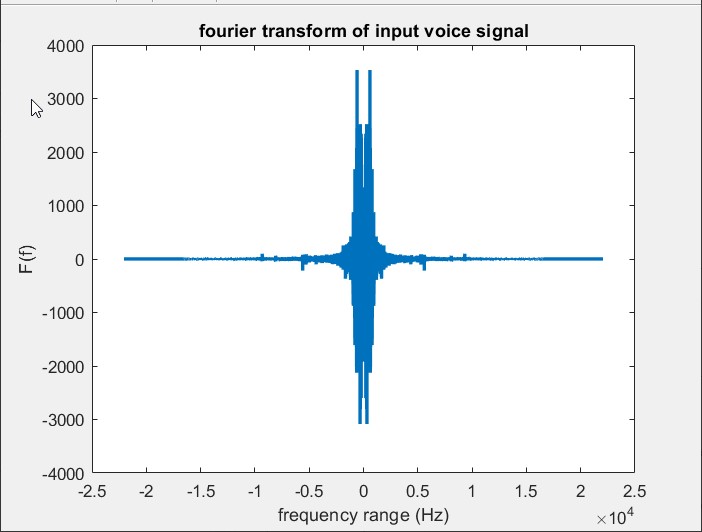
نمودار شیفت فرکانسی به اندازه f=2000 نمودار طیف فرکانسی سیگنال صوت اصلی



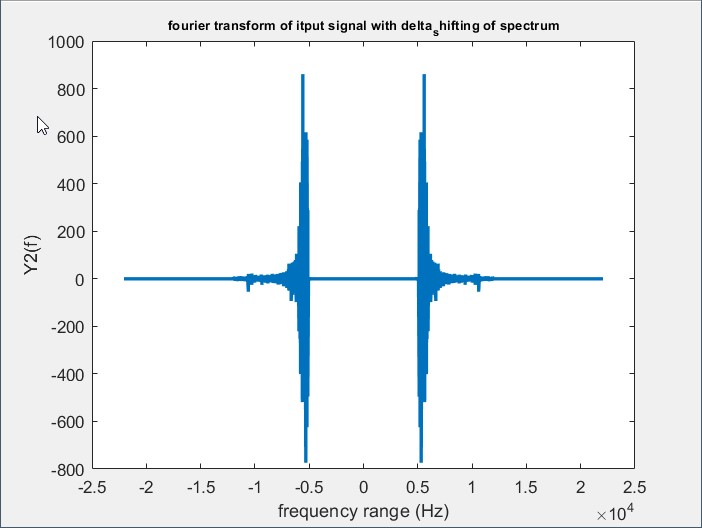
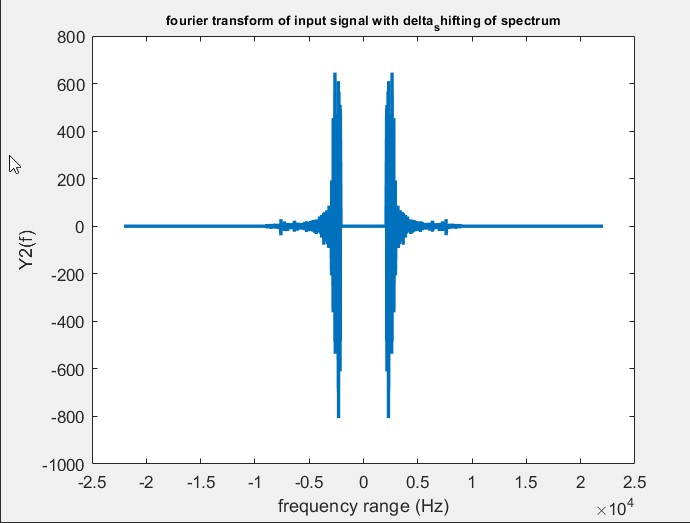
نمودار شیفت فرکانسی به اندازه f=5000 نمودار شیفت فرکانسی به اندازه f=8000



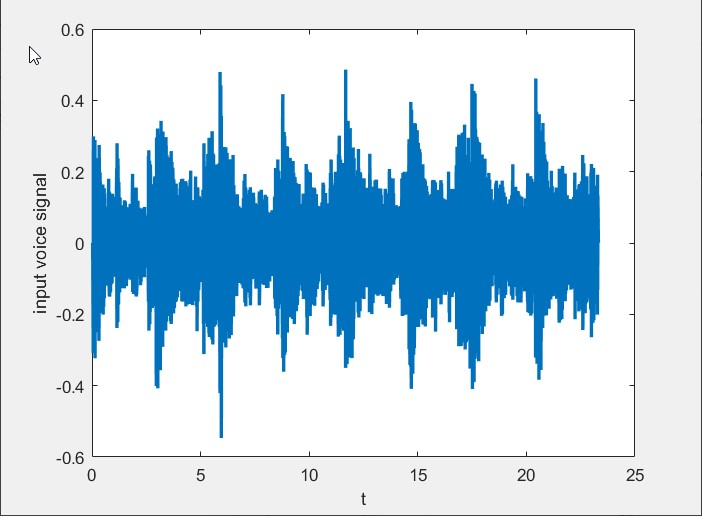
نمودار حوزه زمان سیگنال صوت اصلی

برای فایل دوم (test2.mp3) داریم :

نمودار شیفت فرکانسی به اندازه f=8000 نمودار طیف فرکانسی سیگنال صوت اصلی

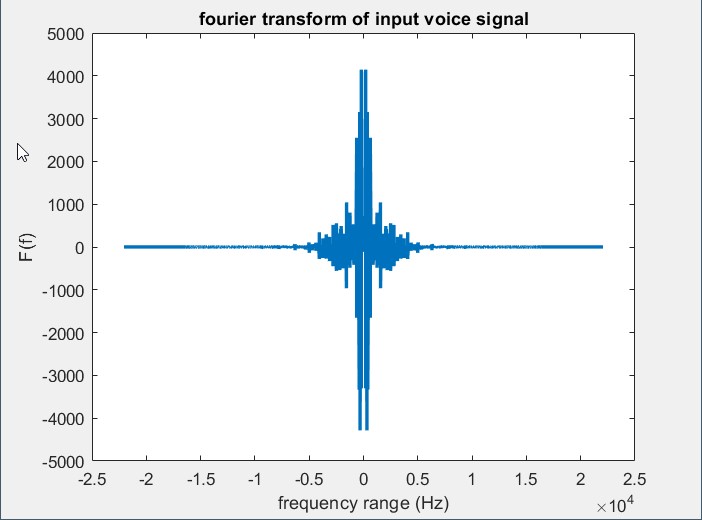
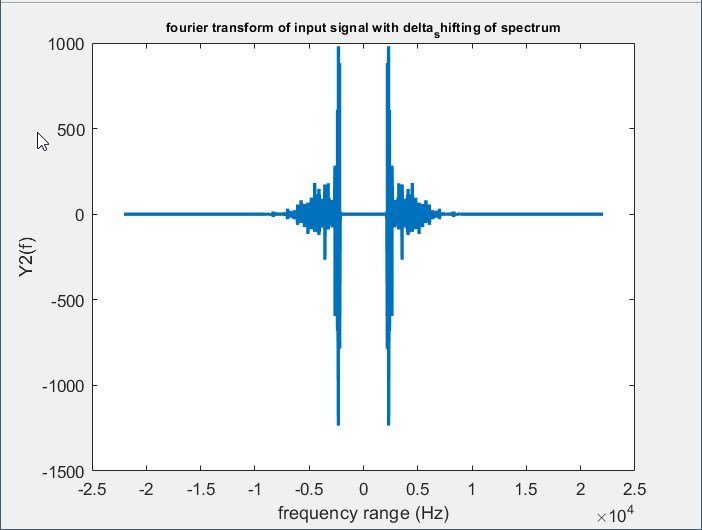


نمودار شیفت فرکانسی به اندازه f=5000 نمودار شیفت فرکانسی به اندازه f=2000



نمودار حوزه زمان سیگنال صوت اصلی

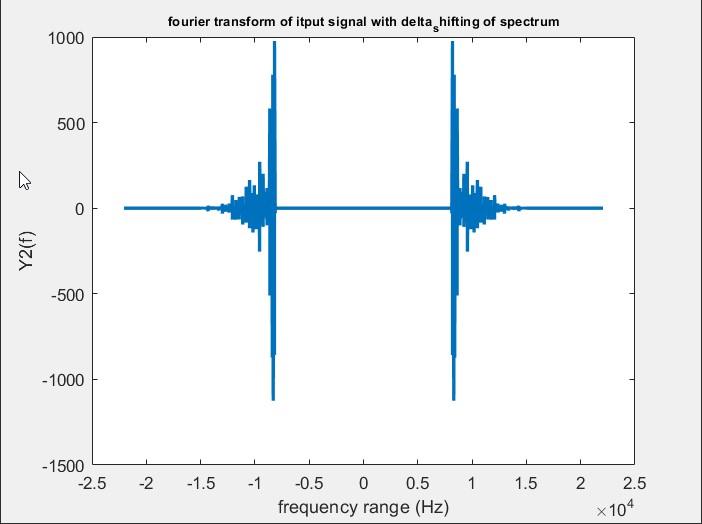
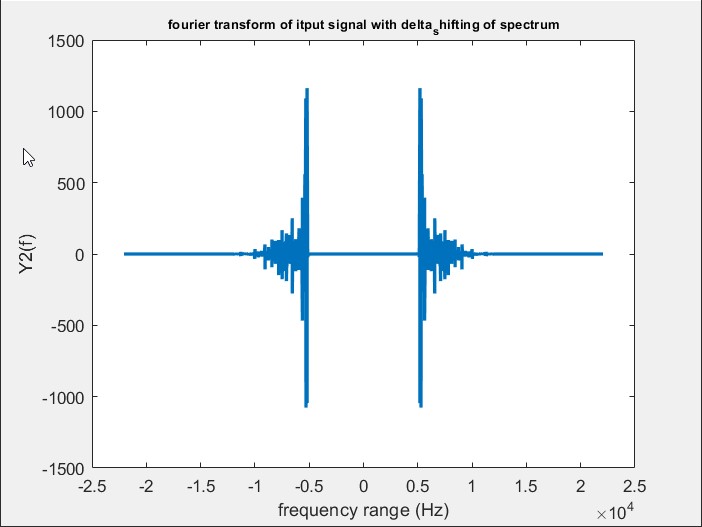
برای فایل سوم (test3.mp3) داریم :

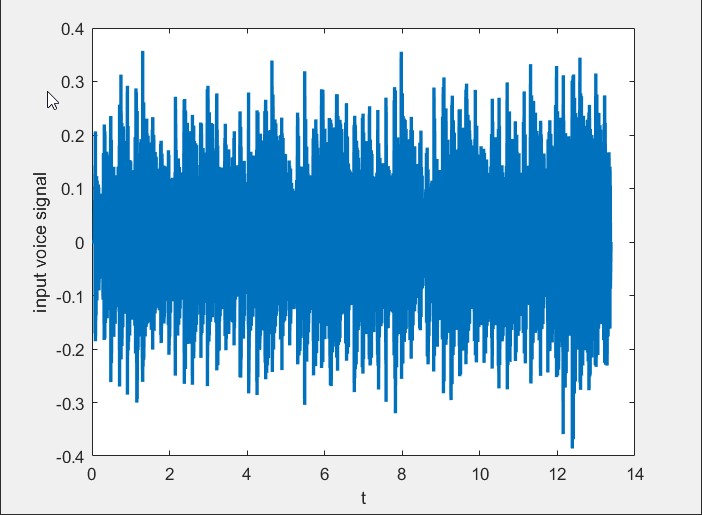


نمودار شیفت فرکانسی به اندازه f=2000

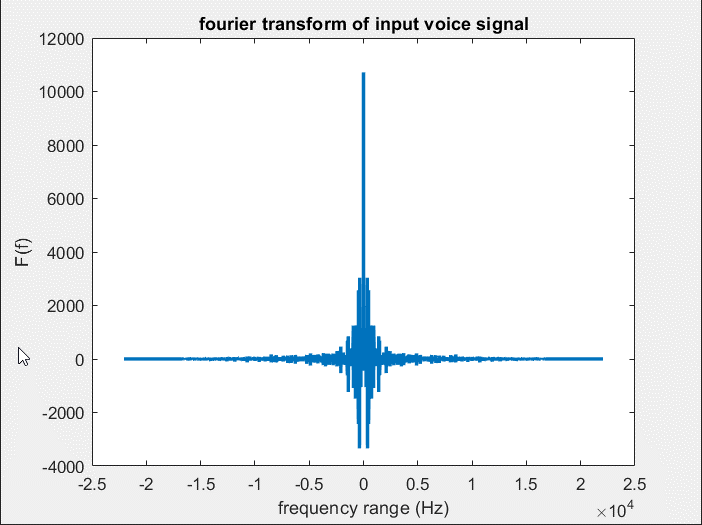
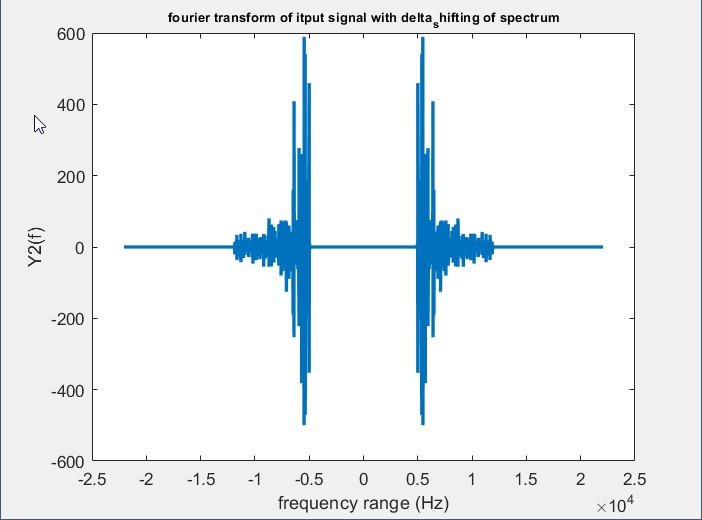
نمودار طیف فرکانسی سیگنال صوت اصلی

نمودار شیفت فرکانسی به اندازه f=5000 نمودار شیفت فرکانسی به اندازه f=8000

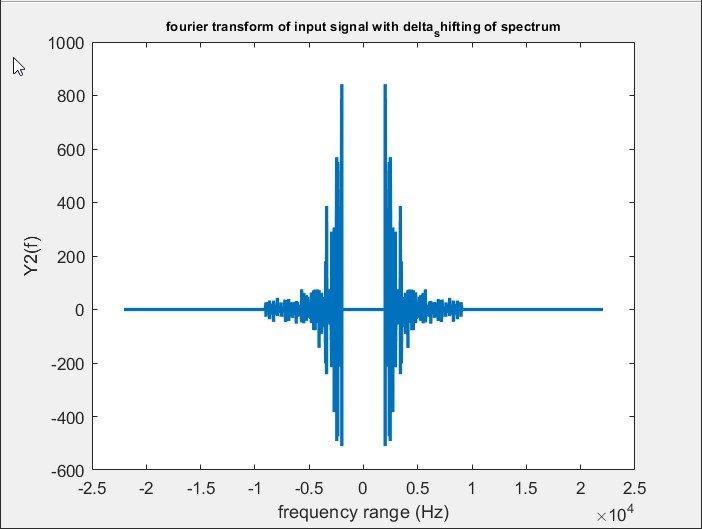
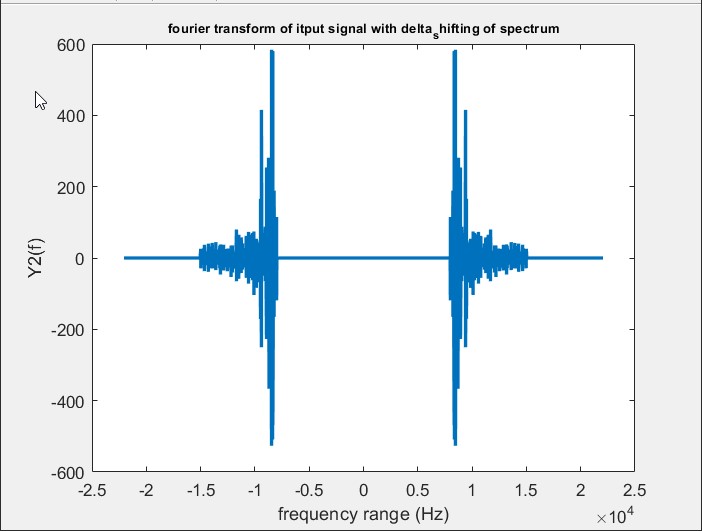




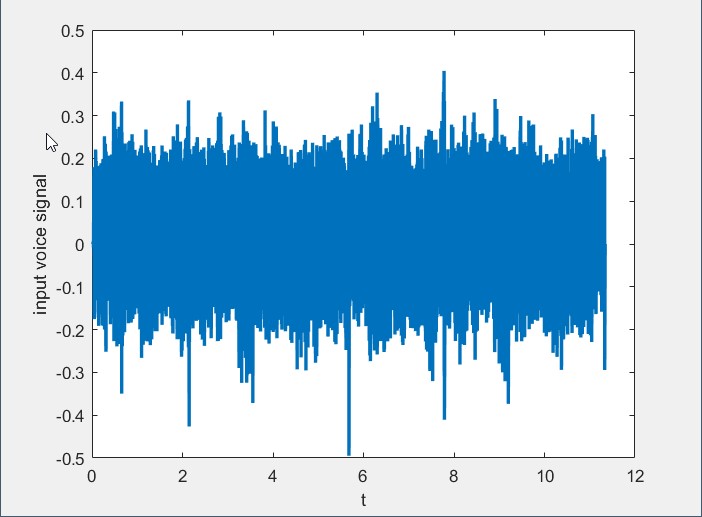
نمودار حوزه زمان سیگنال صوت اصلی

برای فایل چهارم (test4.mp3) داریم :

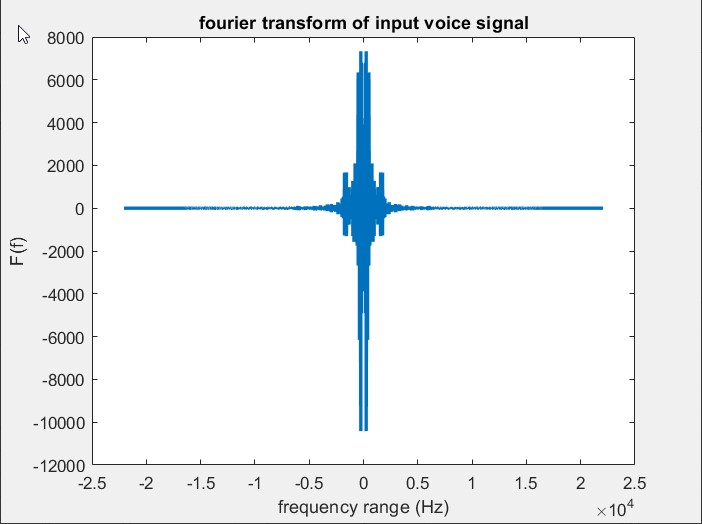
نمودار طیف فرکانسی سیگنال صوت اصلی

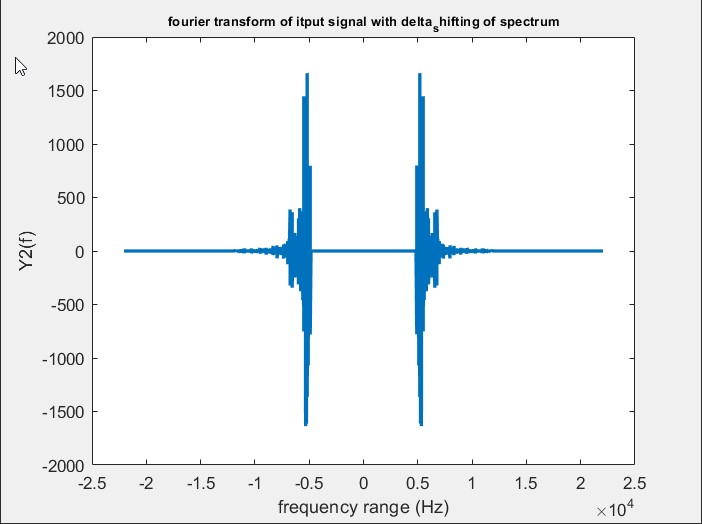
 نمودار شیفت فرکانسی به اندازه f=5000

نمودار شیفت فرکانسی به اندازه f=8000 نمودار شیفت فرکانسی به اندازه f=2000

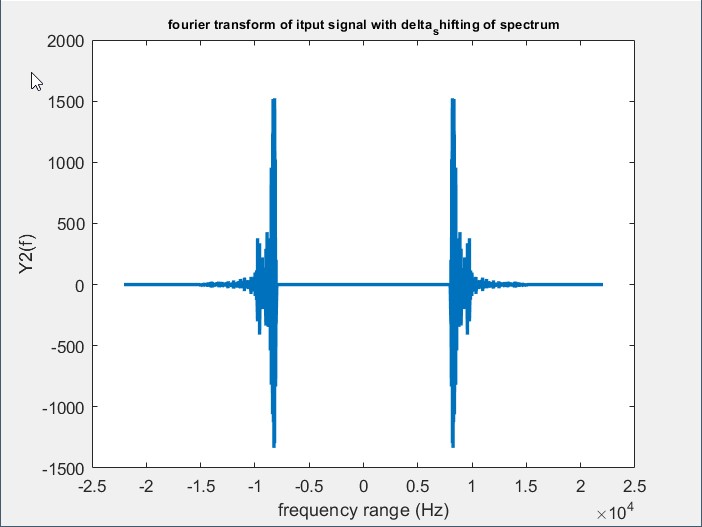
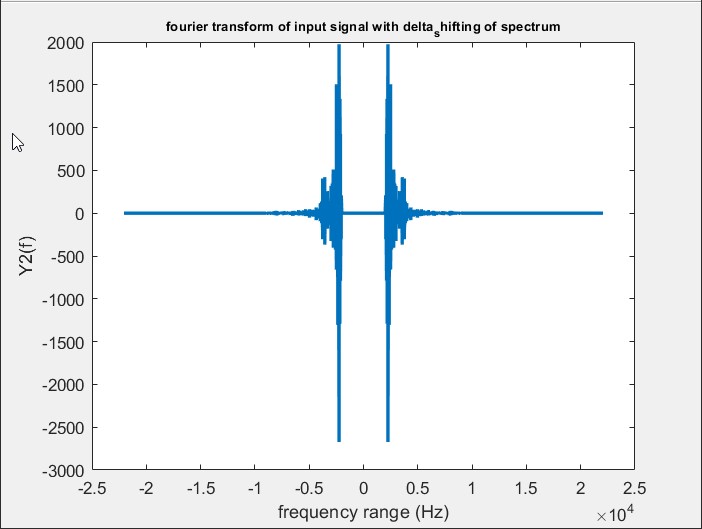


نمودار حوزه زمان سیگنال صوت اصلی

برای فایل پنجم (test5.mp3) داریم :

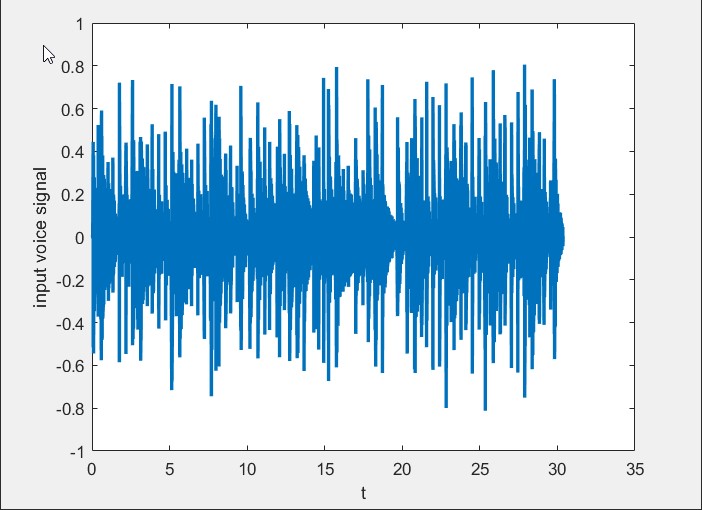
 نمودار شیفت فرکانسی به اندازه f=5000

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.**