دانشکده‌ی مهندسی
کامپیوتر دانشگاه
اصفهان

استاد درس:
دکتر حمیدرضا برادران

دانشجو:
امیررضا صدیقین
۹۹۳۶۱۴۰۲۴

تکلیف دوم درس پردازش گفتار

۱۴۰۰/۰۹/۲۱

مراحل

۱. در این تمرین با دو فایل گفتاری که در تمرین اول کار کرده‌اید، کار می‌کنید.

۲. پیشنهاد ما کار با پایتون برای حل این تمرین است. استفاده از پایتون باعث می‌شود دست شما برای استفاده از کتابخانه‌های مختلف باز باشد و به علاوه تجربه تحلیل گفتار با استفاده پایتون در پروژه‌های غیر درسی نیز کمک کننده است. با این حال می‌توانید با Matlab نیز این تمرین را حل کنید.

۳. ابتدا فایل‌ها را باز کنید. سپس مراحل پیش‌پردازش بر روی فایل صوتی، یعنی مراحل فریم‌بندی، پیش‌تاکید و پنجره‌گذاری، را بر روی فایل‌ها با پارامترهای اولیه زیر اعمال کنید. توجه کنید که برای اعمال این مراحل، از کتابخانه‌های موجود(مثل librosa و ...) استفاده **نکنید** و خودتان این مراحل را بر روی داده‌های فایل صوتی اعمال کنید. (توجه شود که در اینجا منظور کتابخانه‌های مربوط به پردازش سیگنال است که مراحل پیش‌پردازش را خودشان انجام می‌دهند. استفاده از کتابخانه‌هایی مثل numpy و ... بلامانع است.)

a. پارامترهای اولیه:

T = frame length = 32 ms
H = hop size = 10 ms
α = pre-emphasiz = 0.97
window = Hamming

قطعه کد مربوط به این مرحله، در بخش ۳ (part 3) واقع در کدها آمده است. در این قسمت سه تابع pre_emphasis، framing، windowing نوشته شده است که به ترتیب عملیات پیش‌تاکید، فریم‌بندی و پنجره‌گذاری بر روی سیگنال یا سیگنال‌های ورودی انجام می‌دهند. سپس تابع preprocess نوشته شده است که پارامترهای اولیه را دریافت می‌کند و بر روی سیگنال ورودی عملیات پیش‌پردازش (به ترتیب پیش‌تاکید، فریم‌بندی و پنجره‌گذاری) را انجام می‌دهد و فریم‌های پیش‌پردازش شده را خروجی می‌دهد.

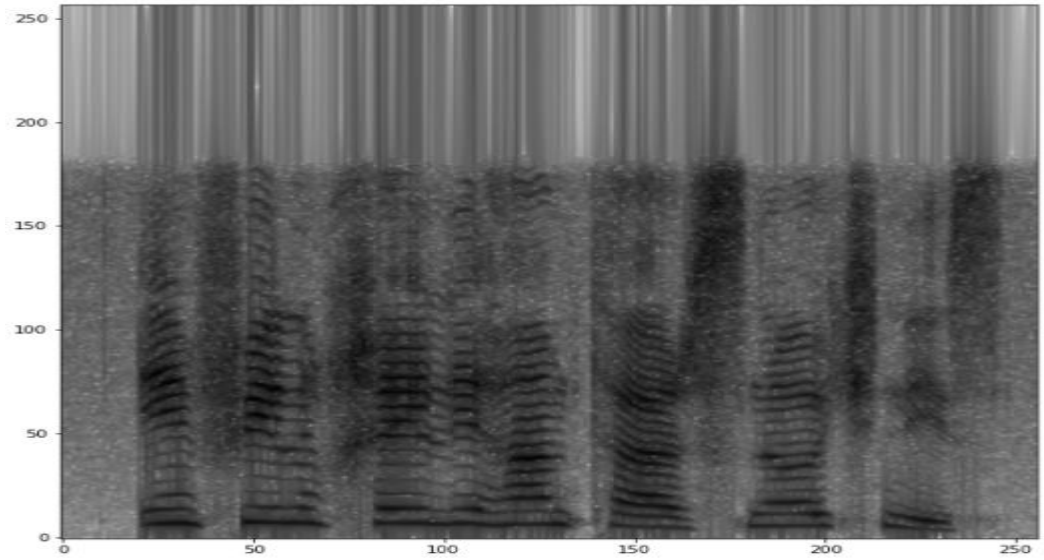
نکته) از تابع np.hamming برای پنجره‌گذاری hamming استفاده می‌شود.

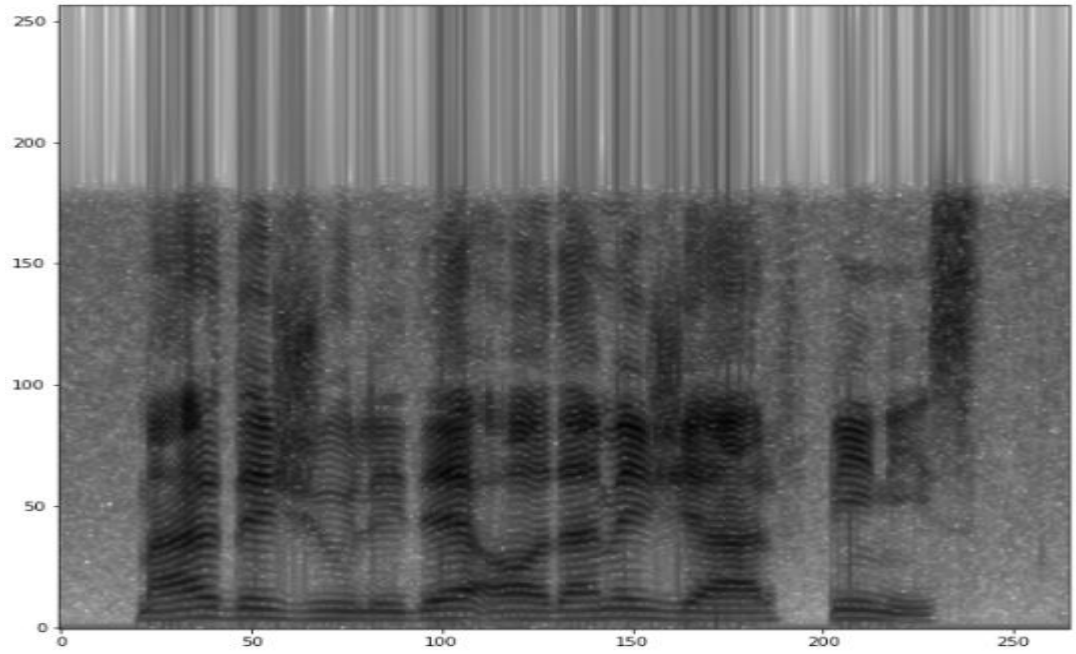
۴. تبدیل فوریه هر فریم را محاسبه کرده و دامنه طیف را ذخیره کنید. (تعداد نقاط fft را ۵۱۲ در نظر بگیرید.)

با استفاده از کلاس np.fft و تابع rfft تبدیل فوریه‌ی فریم‌ها را می‌گیریم و قدرمطلق آن‌ها را ذخیره می‌کنیم. قطعه کد این مرحله در بخش ۴ فایل jupyter آمده است.

۵. اسپکتروگرام سیگنال را از روی دامنه طیف در مرحله قبل بدست آورده و ترسیم کنید.

کد مربوط به این مرحله، در بخش ۵ فایل jupyter آمده است. نمودارهای اسپکتروگرام سیگنال‌های ۱ و۲ به ترتیب به صورت زیر می‌باشد. برای ترسیم اسپکتروگرام باید از فرمول $20\log(magnitud)$ استفاده شود و با به دست آمدن ماتریس مورد نظر و ترانهاده کردن آن، با کمک تابع imshow آن را ترسیم می‌کنیم.



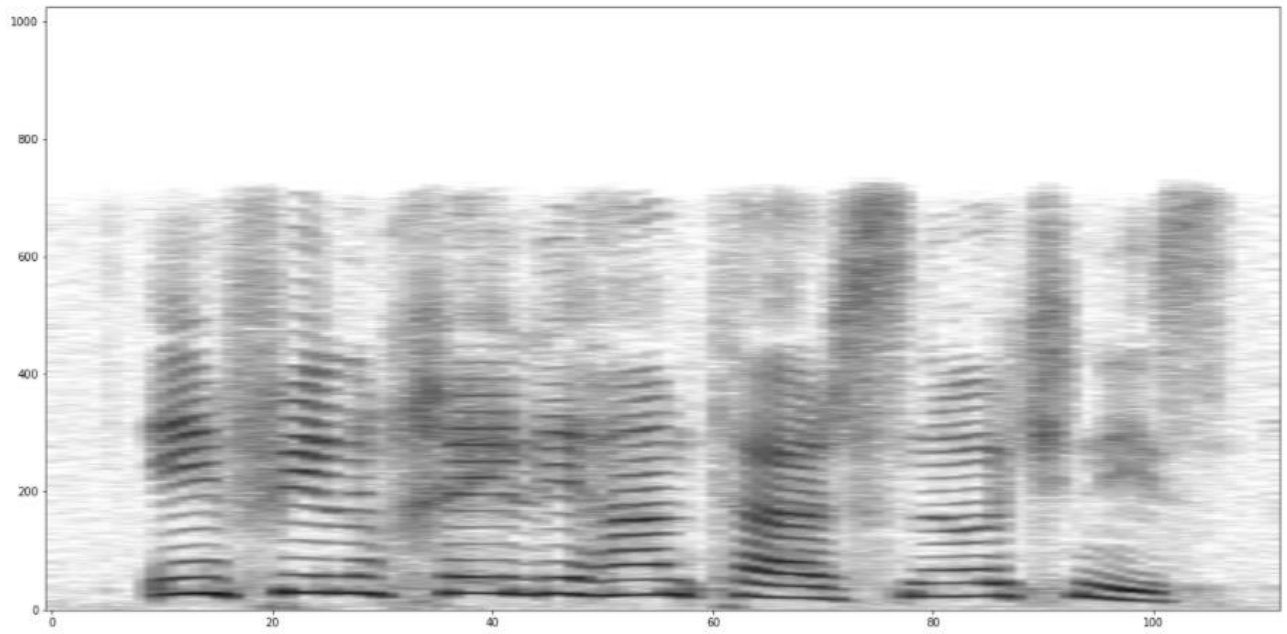


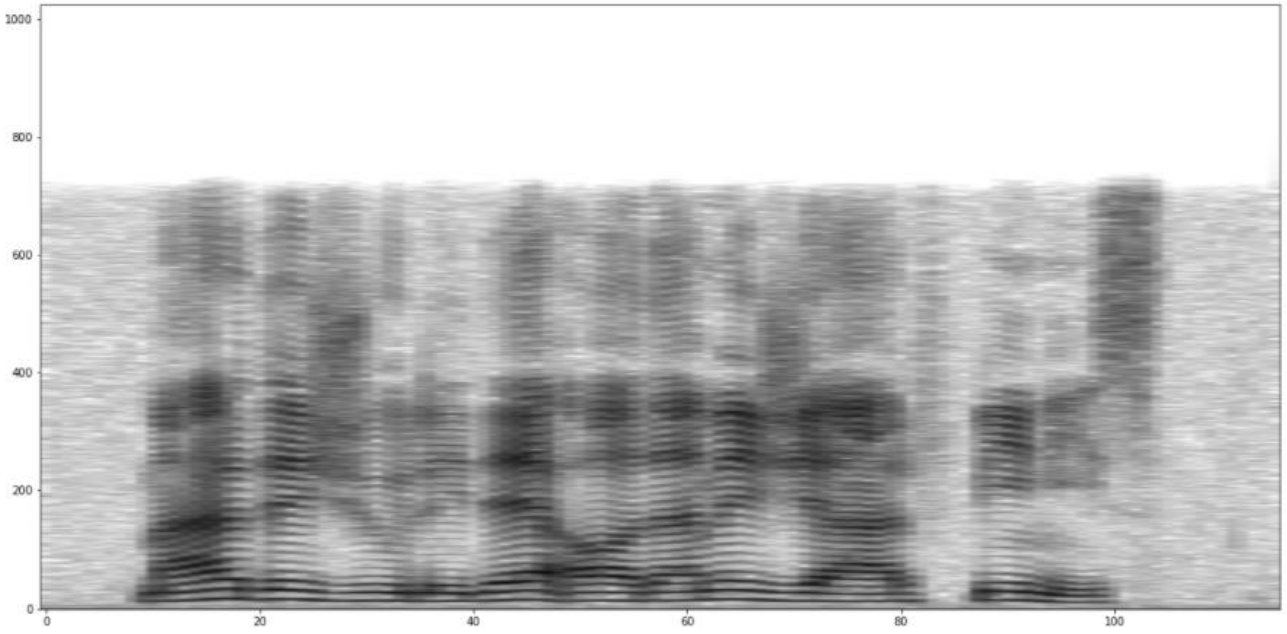
۶. یک بار هم اسپکتروگرام را به وسیله متدهای built-in رسم کنید و با نتیجه مرحله قبل مقایسه کنید. (در پایتون می‌توانید از کتابخانه librosa استفاده کنید و در Matlab از دستور spectrogram استفاده کنید.)

کد مربوط به این مرحله در بخش ۶ آمده است. تابعی تحت عنوان spectogrm_plt (به صورت lambda) نوشته شده است، که سیگنال خوانده شده گرفته و اسپکتروگرام آن را ترسیم می‌کند.

تصاویر به دست آمده، خیلی مشابه تصاویر قبلی می‌باشد و به مانند آن است که تمام اعداد ماتریس قبلی (آن که خودمان محاسبه کردیم.) چند برابر اعداد ماتریس فعلی است. (رنگ‌بندی قبلی پر رنگ تر است.) همچنین ابعاد ماتریس در دو بخش متفاوت است.

تصاویر جدید به ترتیب به صورت زیر است:





۷. یک فریم متناظر با یک واکه گفتاری را انتخاب کنید و مراحل زیر را روی آن انجام دهید:

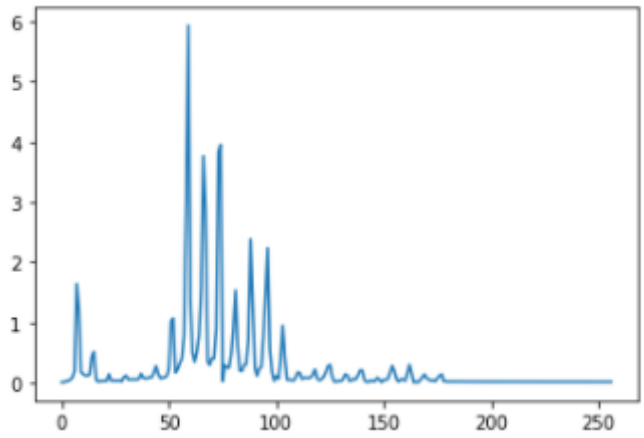
a. دامنه طیف آن را رسم کنید.

b. محل تقریبی فرمنت‌ها را از روی دامنه طیف مشخص کنید و مقدار فرمنت اول و دوم را بر اساس واکه انتخاب شده تحلیل کنید.

c. آیا تشخیص محل فرمنت‌ها از روی دامنه طیف ساده است؟ چه راهکار بهتری برای تشخیص فرمنت‌ها از روی دامنه طیف پیشنهاد می‌کنید؟

برای پیدا کردن یک واکه، فریم دارای بیشترین میانگین انرژی را دارا می‌باشد، انتخاب شده است. انرژی وابسته به مجذور magnitude است و اگر بیشترین میانگین مجذور magnitude فریم‌ها را بدست بیاوریم، یک واکه بدست آمده است.

کد مربوط به این مرحله در بخش ۷ آمده است که نشان می‌دهد فریم با اندیس ۵۲ یک واکه است. نمودار آن به صورت زیر است.



همچنین از اول تا فریم ۵۲ ام پخش شد که نشان می‌دهد برای واکه‌ی /ای/ است.

۸. یک فریم واکدار و یک فریم بی‌واک را انتخاب کنید.

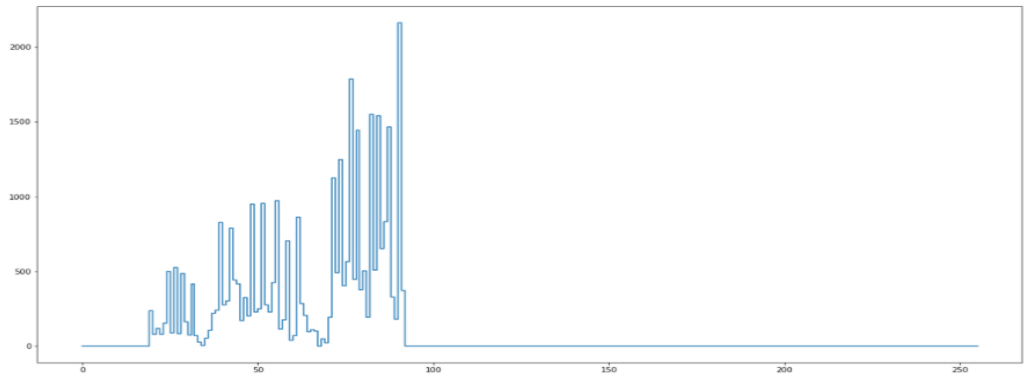
a. بدون اعمال مرحله اعمال پنجره Hamming (در مراحل پیش‌پردازش) دامنه طیف را برای این دو فریم رسم کنید. تفاوت این دو دامنه طیف را بررسی کنید و توضیح دهید.

b. بار دیگر بدون اعمال مرحله پیش تاکید، دامنه طیف را برای این دو فریم رسم کنید. تفاوت این دو دامنه طیف را بررسی کنید و توضیح دهید.

مقادیر میانگین pitch هر فریم را بدست می‌آوریم. بی واک به معنای صفر بودن این مقدار است و واکدار عدم صفر بودن. فریم با اندیس ۶۸ مربوط به یک بی‌واک و فریم با اندیس ۲۹ مربوط به یک واکدار است.

قطعه کد پیدا کردن pitch در قسمت PART 8 آمده است.

همچنین نمودار پله‌ای مقدار میانگین pitch بر حسب شماره‌ی فریم به صورت زیر است.



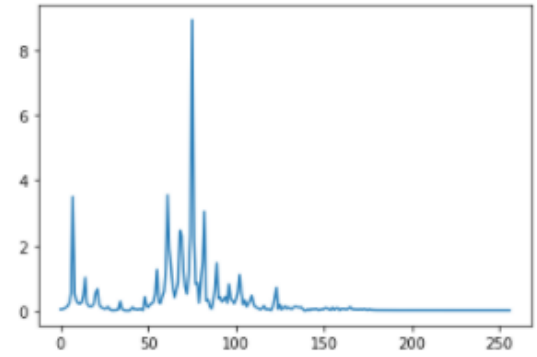
بخش a:

نمودار دامنه‌ی طیف فریم واکدار و بی واک بدون پنجره‌ی hamming به صورت زیر است. تفاوت آن‌ها مقدار دامنه‌ی طیف آن‌ها است.

واکدار

```
vac_dar_frame = preprocess(signal1,sample_rate1,frame_size,hop_size,pre_emphasis_value,None)[vac_dar_index]
mag=abs(np.fft.rfft(vac_dar_frame,512))
plt.plot(mag)
```

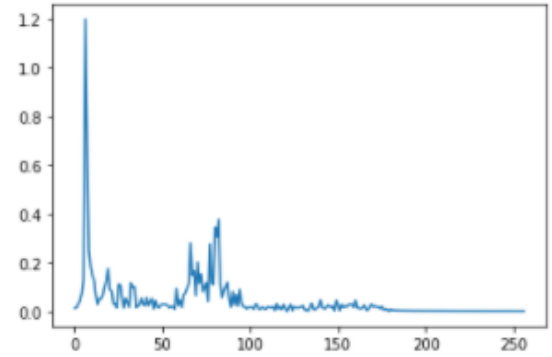
```
[<matplotlib.lines.Line2D at 0x2a3256e2af0>]
```



بی واک

```
bi_vac_frame = preprocess(signal1,sample_rate1,frame_size,hop_size,pre_emphasis_value,None)[bi_vac_index]
mag=abs(np.fft.rfft(bi_vac_frame,512))
plt.plot(mag)
```

```
[<matplotlib.lines.Line2D at 0x2a3257d3f70>]
```



بخش b:

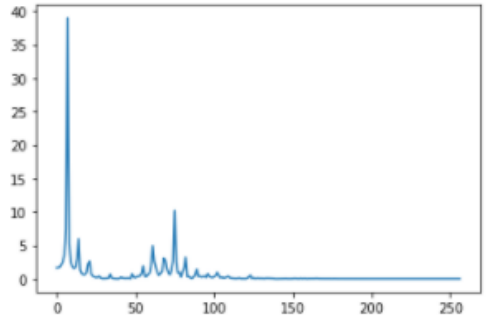
با تغییر پارامتر use_emphasis به مقدار False عملیات‌های بخش قبل را تکرار کردیم و اشکال زیر به دست آمده است.

به دلیل آن که epmphasis تغییرات را در نظر می‌گیرد، در جاهایی که تغییر بیشتر داشتیم نمودار زیاد می‌شود. همچنین باز هم تفاوت این دو نمودار در مقادیر دامنه طیف است.

واکدار

```
vac_dar_frame = preprocess(signal1,sample_rate1,frame_size,hop_size,pre_emphasis_value,None,False)[vac_dar_index]
mag=abs(np.fft.rfft(vac_dar_frame,512))
plt.plot(mag)
```

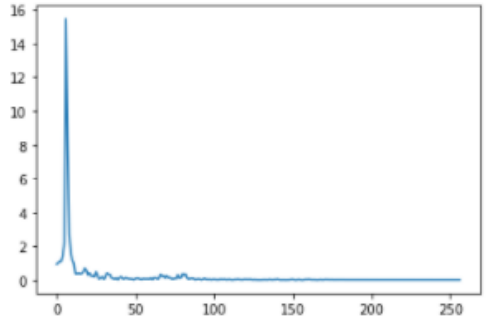
```
[<matplotlib.lines.Line2D at 0x2a3258798b0>]
```



بی واک

```
bi_vac_frame = preprocess(signal1,sample_rate1,frame_size,hop_size,pre_emphasis_value,None,False)[bi_vac_index]
mag=abs(np.fft.rfft(bi_vac_frame,512))
plt.plot(mag)
```

```
[<matplotlib.lines.Line2D at 0x2a3259eb580>]
```



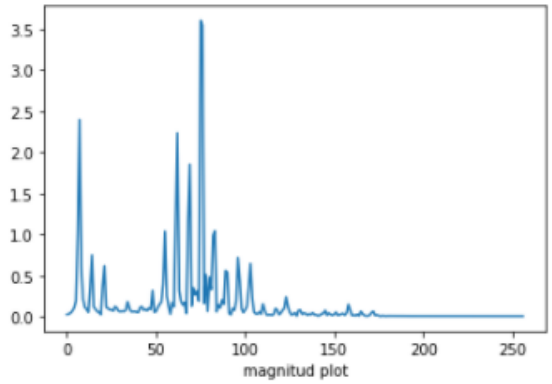
۹. تاثیر تغییر طول فریم بر دامنه طیف فریم را بررسی کنید. برای این کار دامنه طیف فریم را با طول فریم‌های 16 و 64 میلی‌ثانیه (با همان تنظیمات قبلی) رسم کرده و با دامنه طول فریم مرحله قبل (32 میلی‌ثانیه) مقایسه کنید.

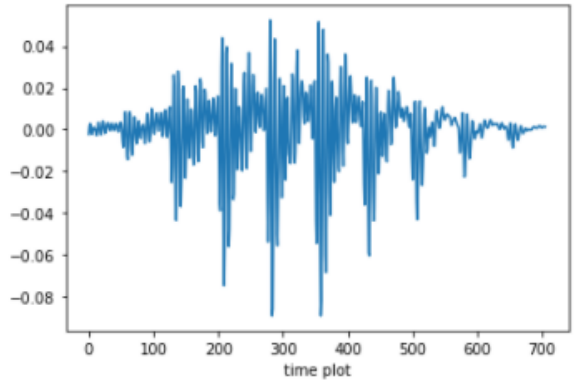
a. در کدام حالت رزولوشن زمانی بیشتر است و در کدام حالت رزولوشن فرکانسی؟

در این قسمت، یک تابع نوشته شده است که در آن پارامترهای preprocessing گرفته شده و برای فریم مثلا ۳۰ ام نمودارهای دامنه‌ی طیف و دامنه‌ی زمانی (خود فریم) کشیده می‌شود. برای سه حالت گفته شده نمودارها به صورت زیر است:

with frame_size = 32 ms

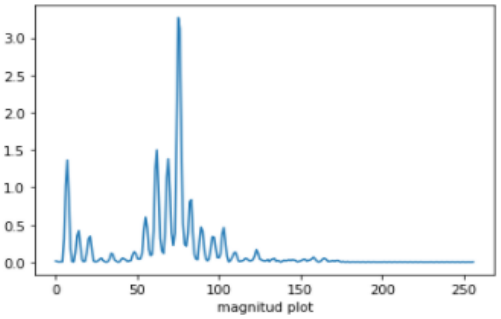
```
analysis_parameters(signal=signal1,sample_rate=sample_rate1
                    ,frame_size= 32/1000 ,hop_size=hop_size
                    ,pre_emphasis_value=pre_emphasis_value,window_type=np.hamming
                    , use_emphasis=True)
```

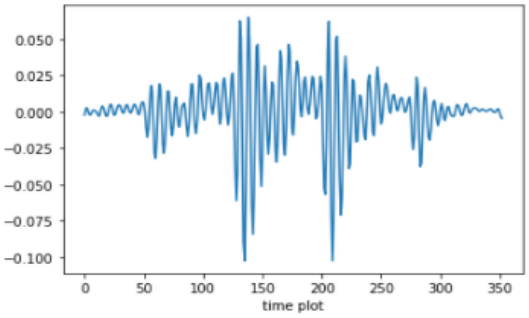




with frame_size = 16 ms

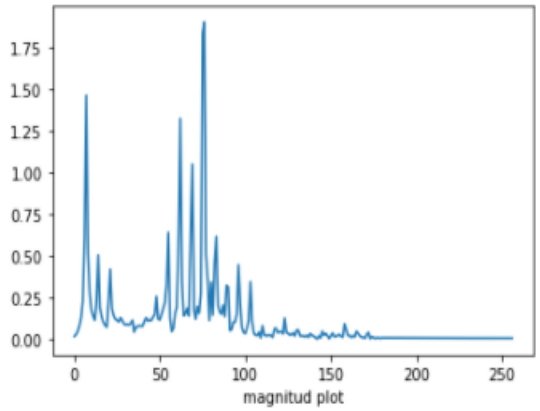
```
analysis_parameters(signal=signal1,sample_rate=sample_rate1
                    ,frame_size= 16/1000 ,hop_size=hop_size
                    ,pre_emphasis_value=pre_emphasis_value,window_type=np.hamming
                    , use_emphasis=True)
```

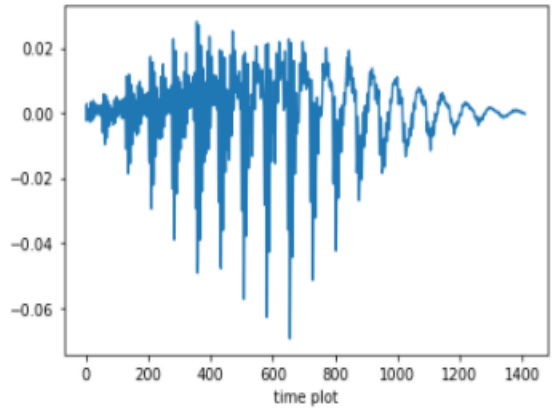




with fram_size = 64

```
analysis_parameters(signal=signal1,sample_rate=sample_rate1
                    ,frame_size= 64/1000 ,hop_size=hop_size
                    ,pre_emphasis_value=pre_emphasis_value,window_type=np.hamming
                    , use_emphasis=True)
```





در حالت frame-size=16 رزولوشن زمانی بیشتر است و در حالت frame-size=64 رزولوشن فرکانسی

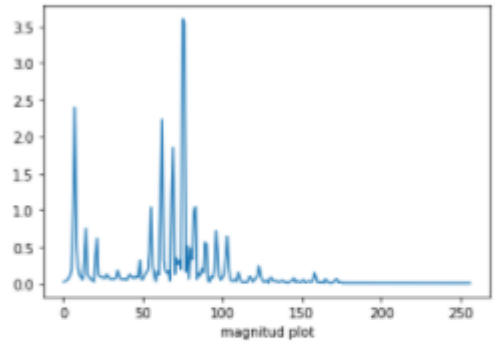
۱۰. تاثیر تغییر تعداد نقاط fft بر روی دامنه طیف فریم را بررسی کنید. برای این کار دامنه طیف فریم را با تعداد نقاط 256 و 1024 (با همان تنظیمات قبلی) رسم کنید و با دامنه طول فریم مرحله ۷ (512 نقطه) مقایسه کنید.

a. به نظر شما حداقل تعداد نقاط fft برای آنکه بتوان از روی طیف سیگنال اصلی را بازسازی کرد چقدر است؟

در این بخش تابعی نوشته شده است که تعداد نقاط fft و فریم را در ورودی می‌گیرد و نمودار طیف آن را ترسیم می‌کند که برای مقادیر بالا به صورت زیر است:

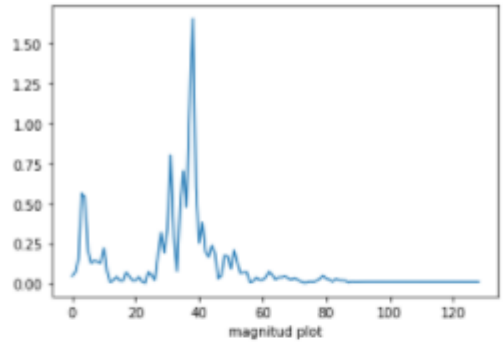
n_fft = 521

```
1 analysis_n_fft(n_fft=512,frame=frame)
```



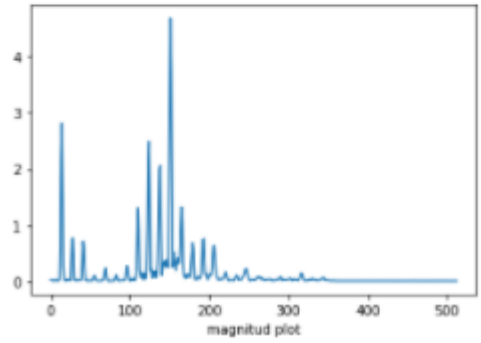
n_fft=256

```
1 analysis_n_fft(n_fft=256,frame=frame)
```



n_fft=1024

```
1 analysis_n_fft(n_fft=1024,frame=frame)
```



حداقل مقدار n_fft باید نصف مقدار sample_rate باشد تا مشکل aliasing رخ ندهد.

۱۱. یک مقاله در حوزه پردازش صوت پیدا کنید که در قسمت طیف فرکانسی یا نحوه تخمین فرمنت‌ها از روی طیف فرکانسی نوآوری داشته باشد. آن مقاله را بررسی کرده و به صورت خلاصه توضیح دهید.

مقاله‌ی Formant extraction from linear-prediction phase spectra را انتخاب کردم.

DOI : https://doi.org/10.1121/1.381864

در این مقاله گفته است که با استفاده از مشتق طیف فاز می‌توان فرمنت‌های دقیقی به دست آورد. کاربرد این روش از طریق نمونه‌هایی از طیف‌های پیش‌بینی خطی به‌دست‌آمده برای مدل‌های شبیه‌سازی‌شده و برای بخش‌های گفتار واقعی نشان داده شده است.