



دانشگاه تهران
پردیس دانشکده‌های فنی

پردازش سیگنال‌های دیجیتال

گزارش پروژه ۲

سید علیرضا جاوید

۸۱۰۱۹۸۳۷۵

استاد

دکتر بدیعی

۱۸ دی ۱۴۰۱

فهرست مطالب

فهرست مطالب	
۱	آشنایی با اسپکتروگرام و تحلیل فرکانس-زمان
۳	فرکانس نمونه برداری و خواندن سیگنال صدا
۳	طیف توان حوزه فرکانس سیگنال
۳	معرفی اسپکتروگرام سیگنال در متلب
۵	بدست آوردن پیام فایل صوتی به کمک اسپکتروگرام
۶	تبدیل موجک
۸	تحلیل فرکانسی زبان و صدا
۸	فرکانس نمونه برداری و خواندن سیگنال صدا
۸	پهنای باند صحبت انسان
۸	طیف توان سیگنال فیلتر نشده
۹	طیف توان سیگنال فیلتر شده
۱۰	بررسی و تحلیل سیگنال ها در حوزه فرکانس

مقدمه

در این تمرین برخی از کاربردهای تبدیل فوریه در زندگی روزمره را بررسی می کنیم. در ابتدا با تبدیل فوریه زمان کوتاه ^۱ آشنا میشویم و سپس به کمک آن و آشنایی با مفهوم اسپکتروگرام یک پیام با تن های $dtmf$ را رمزگشایی می کنیم. سپس با تبدیل موجک ^۲ آشنا و ویژگی های آن را بررسی می کنیم. در بخش بعد نیز فرکانس های صدای مرد و زن را در زبان های عربی و فرانسوی تحلیل کرده و تفاوت های آن ها را بیان می کنیم.

Short Time Fourier Transform^۱
Transform Wavelet^۲

۱ آشنایی با اسپکتروگرام و تحلیل فرکانس-زمان

در این بخش به کمک داده های تمرین و اسپکتروگرام به رمزگشایی فایل صوتی داده شده می پردازیم.

۱.۱ فرکانس نمونه برداری و خواندن سیگنال صدا

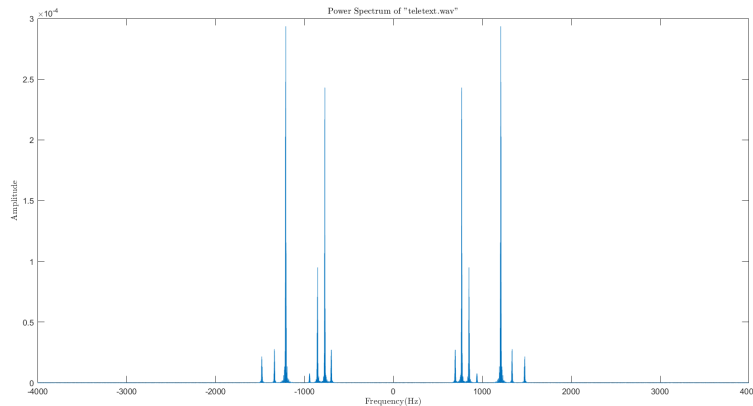
با پردازش فایل صوتی داده شده در متلب به کمک دستور *audioread* فرکانس نمونه برداری برابر $8kHz$ بدست می آید.

۲.۱ طیف توان حوزه فرکانس سیگنال

از راهنمایی داده شده در سوال برای بدست آوردن تبدیل فوریه پیوسته سیگنال استفاده می کنیم. همچنین با توجه به استفاده از *fft* برای محاسبه تبدیل فوریه پیوسته، در نهایت باید آنرا نرمالیزه کنیم. برای نرمالیزه کردن باید خروجی *fft* را بر تعداد سَمپل های سیگنال تقسیم کنیم. حالا برای محاسبه طیف توان یک سیگنال از رابطه زیر استفاده میکنیم.

$$PS_x(f) = |X(f)|^2$$

در شکل زیر می توان طیف توان به دست آمده سیگنال را در متلب مشاهده کرد.



شکل ۱: طیف توان سیگنال صوتی *teletext.wav*

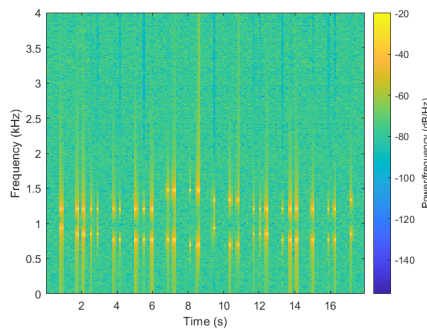
همانطور که با توجه به سیگنالینگ *dtmf* انتظار داشتیم، فرکانس های مشاهده شده در سیگنال تنها بازه 500 Hz تا 1.5 kHz را شامل می شود.

۳.۱ معرفی اسپکتروگرام سیگنال در متلب

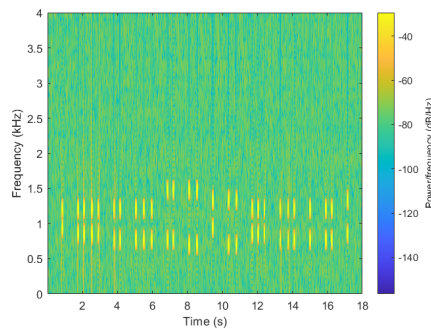
با استفاده از تابع *spectrogram* نشان میدهیم که در بازه های زمانی مختلف سیگنال چه فرکانس هایی دارد. ابتدا در خصوص ورودی و نحوه عملکرد این تابع توضیحاتی را بیان می کنیم.

spectrogram(x, windowsize, noverlap, nfft, fs)

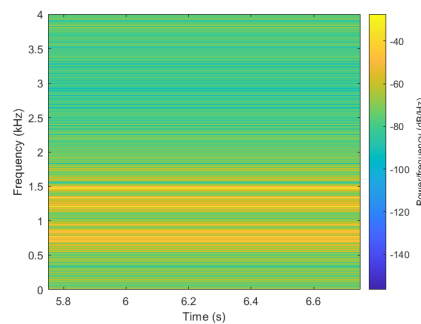
- ورودی اول x ، دریافت کننده سیگنال ورودی در حوزه زمان است.
 - این تابع بصورت چرخه ای و در هر چرخه fft یک پنجره از زمان را محاسبه کرده و سپس به سراغ پنجره بعدی میرود. هر کدام از این پنجره ها دارای یک سایز (تعداد عناصر) هستند که با window size مشخص می شود.
 - هر دو پنجره مجاور (برای مثال پنجره N و $N - 1$) می توانند با هم اشتراکی داشته باشند که با ورودی سوم یعنی noverlap مشخص می گردد.
 - ورودی چهارم یعنی nfft تعداد نقاط تبدیل فوریه fft هر پنجره را مشخص می کند.
 - ورودی آخر نیز فرکانس نمونه برداری سیگنال می باشد.
- برای بررسی حدی تغییرات آرگومان های ورودی بر نتیجه نهایی روی سیگنال به صورت نمونه پارامترهای زیر را بررسی می کنیم. مقادیر به صورت $\{\text{window size}, \text{noverlap}, \text{nfft}\}$ می باشد.



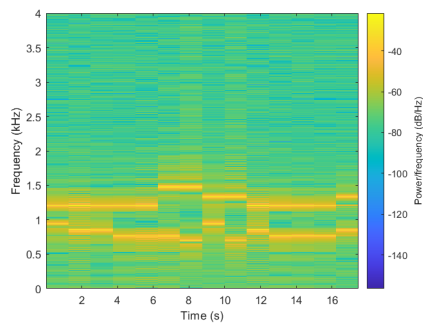
(ب) 1000, 10, 10000



(آ) 100, 10, 1000



(د) 100000, 10, 10000



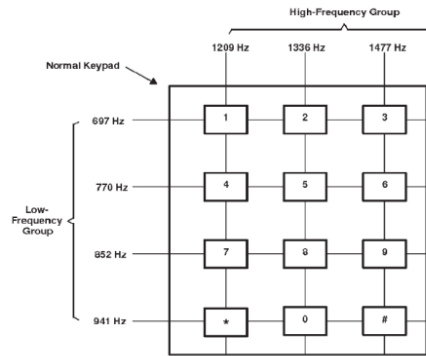
(ج) 10000, 10, 10000

شکل ۲: حالت های مختلف ورودی های spectrogram

همانطور که مشاهده می شود با بزرگتر شدن پنجره، رزولوشن زمانی ضعیف شده و اطلاعات کمتری راجع به زمان وقوع فرکانس بدست خواهیم آورد و به صورت حدی با افزایش بالای پنجره همانند تبدیل فوریه رزولوشن فرکانسی بالا و رزولوشن زمانی صفر خواهد داشت. کاهش سایز پنجره نیز به صورت عکس عمل خواهد کرد یعنی رزولوشن زمانی افزایش می یابد اما رزولوشن فرکانسی کاهش پیدا می کند. افزایش

نقاط fft به دقت بیشتر نمودار کمک می کند و تعیین مناسب noverlap می تواند دقت سیگنال رسم شده را با توجه به نمونه های مشترک افزایش دهد.

۴.۱ بدست آوردن پیام فایل صوتی به کمک اسپکتروگرام

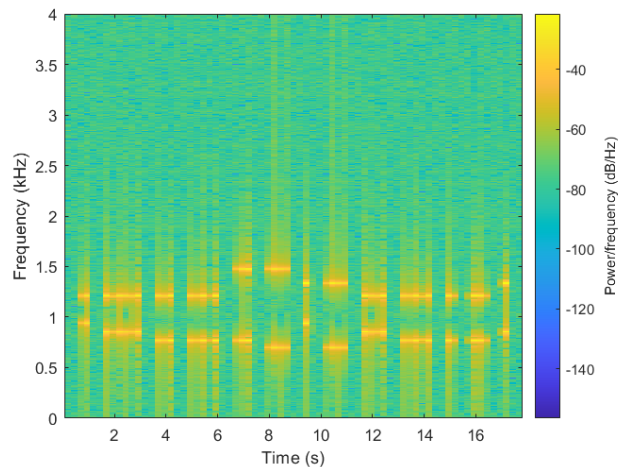


شکل ۳: فرکانس های dtmf

با توجه به تحلیل بخش قبل برای رمزگشایی پیام نیاز به رزولوشن فرکانسی مناسب به همراه رزولوشن زمانی بالا نیاز داریم. یک مقدار قابل قبول برای بررسی سیگنال به صورت

$$\{window\ size = 3000, \ noverlap = 1000, \ nfft = 1500\}$$

می باشد که شکل زیر را نتیجه می دهد.



شکل ۴: اسپکتروگرام سیگنال صوتی teletext.wav

با دقت به شکل ۴ و مقایسه با مقادیر فرکانسی شکل ۳ و توجه به فاصله زمانی 0.8 s می توان پیام سیگنال را به صورت زیر بیان کرد.

* 7777 44 444 66 33 0 22 777 444 4 44 8

همچنین تصویر زیر را برای کیبورد گوشی ۱۲ دکمه ای داریم.



شکل ۵: صفحه کلید گوشی ۱۲ دکمه ای

با توجه به شکل ۵ می توان پیام متنی را به صورت زیر بدست آورد.

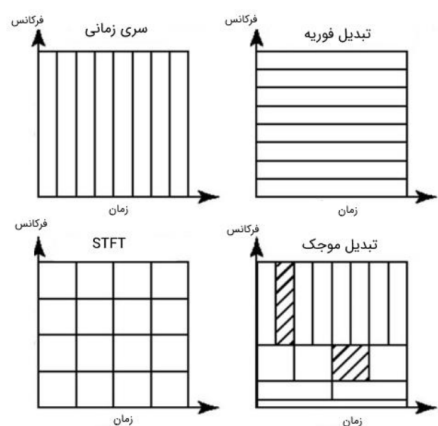
SHINE BRIGHT

۵.۱ تبدیل موجک

در این بخش یک پارامتر جدید به نام فاکتور مقیاس تعریف می کنیم. با توجه به اینکه واژه فرکانس بیشتر برای حوزه تبدیل فوریه به کار می رود از واژه مقیاس برای تبدیل موجک استفاده می کنیم. در بعضی حالت ها زمانی که بیان فرکانس از مقیاس بهتر باشد، با رابطه زیر مقیاس را با استفاده از فاکتور مقیاس به شبه فرکانس تبدیل می کنیم:

$$f_a = \frac{f_c}{a}$$

در این رابطه a بیانگر فاکتور مقیاس و طول موج موجک و f_c فرکانس مرکزی سیگنال مادر و f_a شبه فرکانس می باشد. پس هر چقدر که طول موج طولانی تر باشد، a بزرگتر بوده و f_a متناظر با فرکانس های پایین تر است و در نتیجه با مقیاس دهی به سیگنال موجک تحلیل فرکانس های کوچک تر بهتر شده و رزولوشن فرکانسی مناسب تر است. به طور عکس هرچه فاکتور مقیاس موجک تر تعیین شود، شبه فرکانس متناظر با فرکانس های بزرگتر بوده و در نتیجه با مقیاس دهی کوچکتر به موجک، تحلیل حوزه زمان دقیق تر بوده و رزولوشن زمانی بهتر می باشد. در شکل ۶ رزولوشن حوزه زمان و فرکانس در حوزه های مختلف نمایش داده شده است.



شکل ۶: رزولوشن زمانی و فرکانسی تبدیل های مختلف

همانطور که مشاهده میشود در تبدیل موجک به صورت مصالحه ای عمل میکند و در مشخصه هایی که بیشتر وابسته به زمان (فرکانس های بالا) هستند، رزولوشن فرکانسی ضعیف و رزولوشن زمانی خوب است و در مشخصه هایی که بیشتر وابسته به فرکانس (فرکانس های پایین) هستند، رزولوشن فرکانسی خوب و رزولوشن زمانی ضعیف می باشد.

۲ تحلیل فرکانسی زبان و صدا

در این بخش ۴ فایل صوتی که شامل گویش مرد و زن برای زبان های عربی و فرانسوی است را بررسی می کنیم.

۱.۲ فرکانس نمونه برداری و خواندن سیگنال صدا

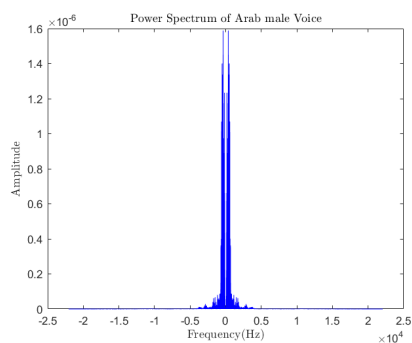
در ابتدا برای پردازش فایل های با فرمت *mp3* داده شده مشکلی وجود داشت که با تبدیل فایل ها به فرمت *wav* مشکل پردازش رفع شد. با پردازش فایل های صوتی در متلب به کمک دستور *audioread* فرکانس نمونه برداری برای همه سیگنال ها برابر 44.1 kHz بدست می آید.

۲.۲ پهنای باند صحبت انسان

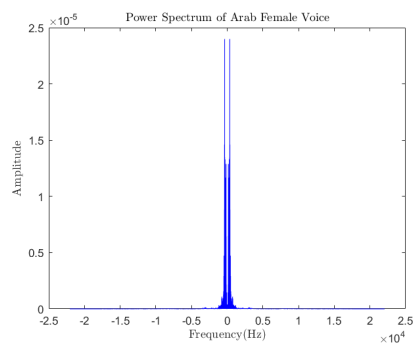
با توجه به این سایت می توان گفت فرکانس صحبت انسان به صورت کلی بین 200 Hz و 3500 Hz می باشد و به طور تقریبی پهنای باند صحبت انسان 3 kHz می باشد.

۳.۲ طیف توان سیگنال فیلتر نشده

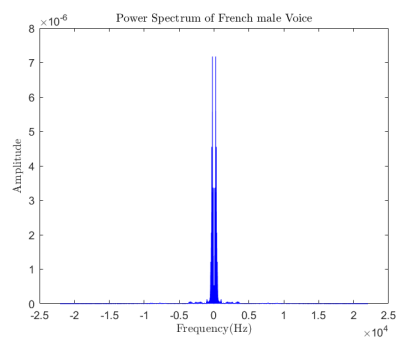
با کمک گرفتن از روش ارائه شده در بخش ۱ برای بدست آوردن تبدیل فوریه پیوسته و رسم طیف توان سیگنال، برای ۴ صدای داده شده نیز این عمل را تکرار کرده و طیف توان سیگنال را رسم می کنیم.



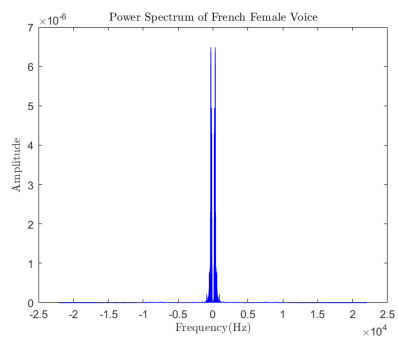
(ب) صدای مرد عرب



(آ) صدای زن عرب



(د) صدای مرد فرانسوی



(ج) صدای زن فرانسوی

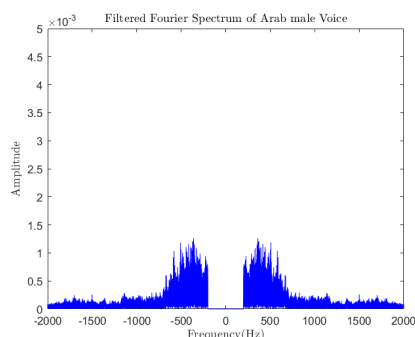
شکل ۷: طیف توان سیگنال های صوتی داده شده

۴.۲ طیف توان سیگنال فیلتر شده

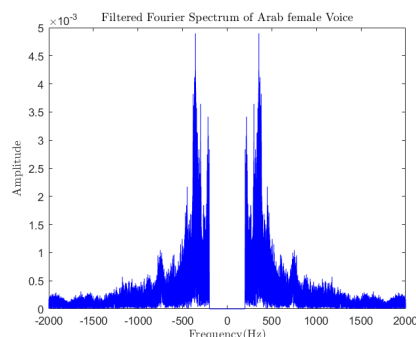
جالا یک فیلتر بندپس ایده آل به صورت زیر طراحی می کنیم:

```
1 bp_filter = zeros(length(PS_FRM), 1);
2 for i=1:length(bp_filter)
3     if (f(i) < 3500 && f(i) > 200) || (f(i) > -3500 && f(i) < -200)
4         bp_filter(i) = 1;
5     end
6 end
```

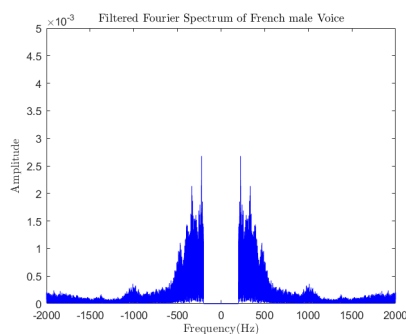
با عبور سیگنال ها از فیلتر band-pass طراحی شده داریم:



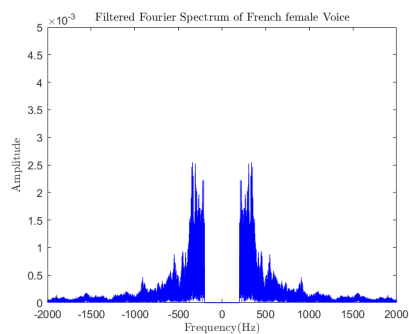
(ب) صدای مرد عرب



(آ) صدای زن عرب



(د) صدای مرد فرانسوی



(ج) صدای زن فرانسوی

شکل ۸: تبدیل فوریه سیگنال های صوتی فیلتر شده

برای بازسازی و پخش صدای سیگنال فیلتر شده در حوزه زمان زمان نیز از کد زیر استفاده می کنیم: (صوت زن عرب برای مثال نشان داده شده)

```
1 arab_f_filtered = ifft(fftshift(ARF_filtred)) * length(arab_f);
2 sound(arab_f_filtered,fs1)
```

۵.۲ بررسی و تحلیل سیگنال ها در حوزه فرکانس

به شکل ۸ با دقت بیشتری نگاه کنید. معمولاً صدای با فرکانس بیشتر (مانند صدای زنان) را زیر و صدای با فرکانس کمتر (مانند صدای مردان) را بم می نامند. سرعت ادای کلمات در زبان فرانسوی بیشتر است و در زبان عربی ادای کلمات معمولاً به صورت یک دست تر و آرام تر گفته می باشد. پهنای باند صدای زنان معمولاً بیشتر است و صدای زن عرب و فرانسوی به صورت کلی شدت بیشتری در فرکانس های بالا دارند. در صدای مرد عرب نويز شنيداري بيشترى در مقايسه با بقيه صدا ها وجود دارد. شدت صدای زن عرب در این نمونه ها از همه بیشتر و شدت صدای مرد عرب از همه کمتر می باشد که می تواند به دلیل تفاوت کیفیت صدای ضبط شده باشد.