

پروژه امتیازی درس سیگنال‌ها و سیستم‌ها

الگوریتم‌هایی برای پردازش صدا

علی فرجی

استاد درس دکتر مهدی راستی



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

دانشکده مهندسی کامپیوتر

بهمن ماه سال ۱۳۹۸

۱ تشخیص جنسیت از روی صدا

۱.۱ تجزیه و تحلیل طیفی به کمک تبدیل فوریه

تبدیل فوریه یکی از مفید ترین ابزار های ریاضی در زمینه های مختلف علوم و مهندسی است. تبدیل فوریه در پردازش سیگنال، فیزیک، ارتباطات، مباحث نوری ، پردازش صدا و تصویر و خیلی از زمینه های دیگر کاربرد دارد.

این تکنیک، یک تابع یا مجموعه ای از اطلاعات را از حوزه زمان یا نمونه (sample) به حوزه فرکانس منتقل می کند. این بدان معناست که تبدیل فوریه میتواند بخش فرکانسی را برای یک مجموعه از اطلاعات زمانی نمایش دهد.

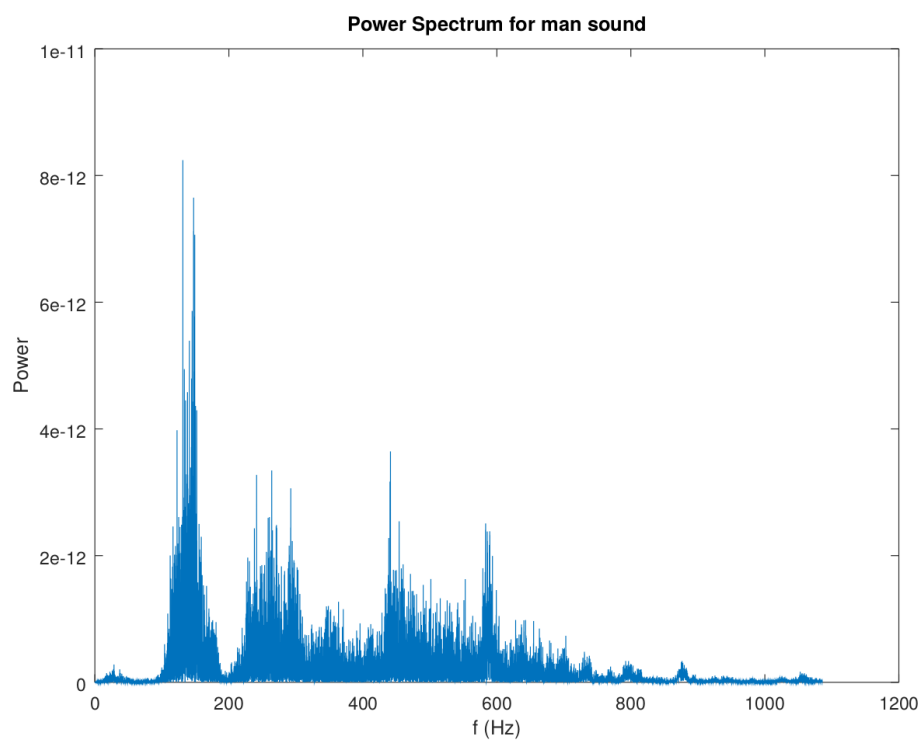
به کمک تبدیل فوریه و نمودار طیف فرکانسی میتوان فرکانس هایی که بیشترین تاثیر را در ساخت سیگنال دارند شناخت و مثلا در موسیقی نت ها و صدا ها را تشخیص داد و یا در طیف سنجی ها مثلا در علم شیمی به عناصر موجود پی برد.

۲.۱ بررسی دو مورد از صدا ها

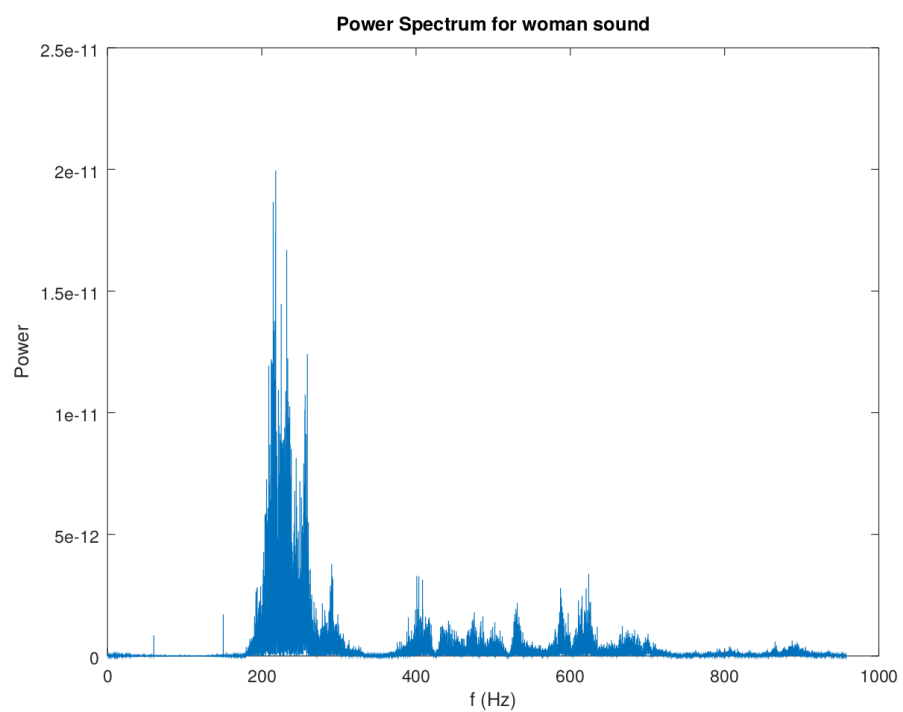
همان طور که در اول تعریف پروژه یک نمودار آورده شده است ما نیز با توجه به شکل های بدست آمده به آن پی میبریم مثلا برای مرد شکل ۱ حاصل که مطابق با نمودار است زیرا یک پیک بزرگ اطراف ۱۲۰ و یک پیک کمی کوچکتر اطراف ۲۲۰ دارد. کد نوشته شده در فایل Al.m قرار دارد.

همچنین مشاهده می شود (شکل ۲) که برای زن فرکانس های بدست آمده اطراف ۲۱۰ هرتز قدرت بیشتری دارند.

پس میتوان با استفاده از نقاط ماکسیمم های طیف فرکانس به جنسیت صاحب صدا پی برد.



شکل ۱: طیف توان برای صدای ضبط شده مرد شماره ۳



شکل ۲: طیف توان برای صدای ضبط شده زن شماره ۱۰

۳.۱ اوج صدا

میتوانیم تابع ماکزیمم را روی مقدار اندازه تبدیل فوریه صدا بزنییم تا مقدار اوج را به ما بدهد. تابع `getPeak` نوشته شد که مقدار پیک و فرکانس در آن پیک را برمیگرداند همچنین در فایل `A3.m` میتوانیم خروجی های آن را ببینیم.

۴.۱ تشخیص جنسیت

پس از پیدا کردن اوج و مقدار ایندکس آن میتوان فرکانس آن را حساب کرد سپس میتوان با توجه به مقدار این فرکانس زن یا مرد بودن را تشخیص داد. کدهای در فایل `A4.m` نوشته شدند که از تابعی که در بالا نوشتیم استفاده میکند. توجه کنید که بعد از اجرای دستور `A4` با تایپ `result` که محتویات خروجی را در درون آن می نویسد، میتوان برچسب ها و فایل مربوطه را مشاهده کرد.

```
>> result
```

```
result =
```

```
voices/v0.wav women
```

```
voices/v1.wav man
```

```
voices/v10.wav women
```

```
voices/v11.wav man
```

```
voices/v2.wav man
```

```
voices/v3.wav man
```

```
voices/v4.wav women
```

```
voices/v5.wav women
```

```
voices/v6.wav man
```

```
voices/v7.wav women
```

voices/v8.wav man

voices/v9.wav women

پیشنهاد برای افزایش دقت: ما چون فقط بر اساس یک ماکزیمم در حال تصمیم گیری هستیم خطای بالایی داریم بهتر است برای بهینه کردن و افزایش دقت مجموعه ای از ماکسیمم ها و مینیمم های محلی را پیدا کنیم و الگوی آنها را با الگوی اکستریمم های محلی هر جنسیت تطابق دهیم. البته چون ویس های ما زیاد نویز ندارند پس زیاد فرکانس ها تغییر نکرده اند و با همان پیک فرکانس میتوان به راحتی تشخیص داد. برای این کار میتوان از تابع findpeaks استفاده کرد که در نرم افزار octave در پکیج signal است.

۲ الگوریتم هایی برای بهبود صدا

۱.۲ معرفی مختصری از الگوریتم تفاضل طیف

الگوریتم تفاضل طیف یکی از قدیمی ترین و مشهورترین روش های بهسازی سیگنال گفتار است که برای بازیابی طیف توان سیگنال آلوده به نویز استفاده می شود. در این روش طیف توان نویز تخمین زده شده را از طیف توان سیگنال نویزی کم کرده و سیگنال بهسازی شده را تولید می کند.

این الگوریتم یک الگوریتم تک کاناله است روش های دیگر چند کاناله ای هم وجود دارند مثل استفاده از دو میکروفون یکی برای ضبط صدا و یکی برای ضبط نویز محیط که به کمک یک سیستم وفقی میتوانستیم صدای ضبط شده را بهبود ببخشیم.

در روش تفاضل طیفی ، تخمین اندازه طیف توان نویز با استفاده از نواحی سکوت و از طریق میانگین گیری اندازه طیف توان نویز در پنجره های مختلف این نواحی انجام می شود. اساسی ترین مشکل این روش وجود نویزی آزار دهنده بنام نویز موزیکال در سیگنال بهسازی

شده است که دلیل اصلی آن عدم دقت در تخمین طیف نویز است.

میدانیم که عواملی مانند طول فریم و... روی دقت تخمین طیف نویز موثر هستند و به طور خاص اثر آن بر روی نویز موزیکال ایجاد شده مورد مطالعه قرار می گیرد. مطالعات انجام شده نشان می دهند که هر چقدر طول فریم کمتر باشد و به عبارتی پنجره کوتاهتر باشد به علت *convolve* شدن طیف گفتار و طیف پنجره، طیف سیگنال گفتار نیز به این طریق خراب خواهد شد.

همه عوامل فوق ما را به این سمت سوق می دهند که طول فریم را بزرگ انتخاب کنیم. بزرگتر گرفتن طول فریم دقت تخمین طیف نویز را افزایش داده و بنابراین موجب کاهش نویز موزیکال و افزایش کیفیت سیگنال شنیداری بهسازی شده می گردد و ثانیاً اگر دقت تخمین طیف بیشتر باشد می توان در روش تفریق طیفی، ضریب تفریق را بزرگتر انتخاب کرده و مقدار بهبود SNR را افزایش داد.

هرچه میزان سکوت نویزی اول صدا بیشتر باشد تخمین بهتری زده خواهد شد. نسبت سیگنال به نویز (*SNR - Signal to Noise ratio*) معیاری برای نمایش میزان سیگنال مفید در مقابل سیگنال مزاحم (یا نویز) در سیستم های الکتریکی است. این عدد، نسبت توان سیگنال به توان نویز است، و آن را بر حسب دسی بل بیان می کنند.

۲.۲ پیاده سازی و نکات آن

ایده اصلی این الگوریتم برای حذف نویز صحبت در سال ۱۹۷۵ توسط ویس ارائه شد و نتیجه عملی آن در سال ۱۹۷۹ توسط بول منتشر گردید. ابتدائاً برای حذف نویز های سفید و ایسا طراحی شده بود و محدودیت عملیاتی زیادی داشت که از جمله آن نویز های موزیکال بودند. این روش بعد ها اصلاح شد و الگوریتم های پیچیده و روش های غیر خطی به کار گرفته شد تا علاوه بر حذف نویزهای واقعی محیط، نویز موزیکال را نیز کاهش دهد. ما در اینجا الگوریتم بول را پیاده سازی کرده ایم که در فایل SSA.m قرار دارد.

روش کار الگوریتم: ما از ناحیه ای که صدا ندارد و به اصطلاح سکوت است کمک میگیریم که نویز را تخمین بزنیم حال فرض میکنیم که این نویز از هر فریم به فریم دیگر زیاد تغییر نمیکند و در هر فریم طیف نویز را از صوت کم میکنیم.

توضیحات الگوریتم در سورس کد بصورت کامنت قرار داده شد.

توضیحات الگوریتم: این وب سایت

۳.۲ تست الگوریتم در مقابل نویزهای متفاوت

با اجرا کردن فایل B3.m میتوانیم قدرت الگوریتم را در مقابل ۴ نویز مشاهده کنیم.

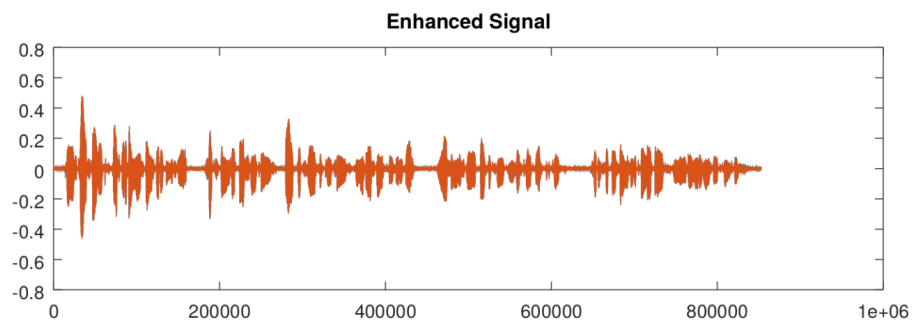
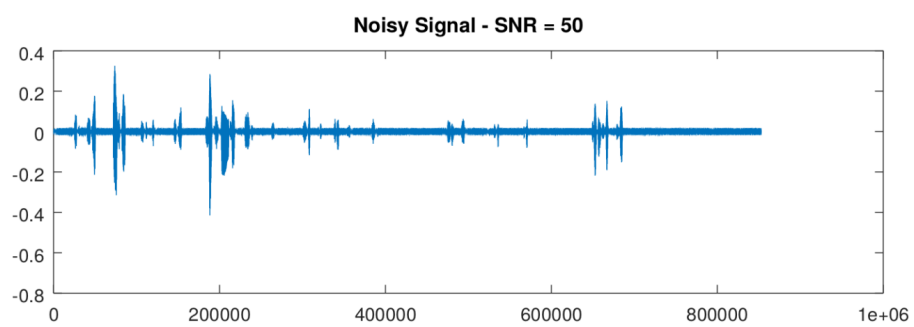
مشاهده میکنیم که این الگوریتم نویز گاوسی را به خوبی تشخیص و حذف میکند زیرا در میانگین گیری بهترین تخمین زده می شود برای همین است که این الگوریتم برای نویز های موزیکال خوب نیست زیرا میانگین نویز تخمین خوبی به ما نمی دهد که البته طبق مطالب گفته شده میتوان الگوریتم را بهتر کرد.

تمامی سیگنال های اولیه از سیگنال نویز دار بطور خوبی استخراج شده اند.

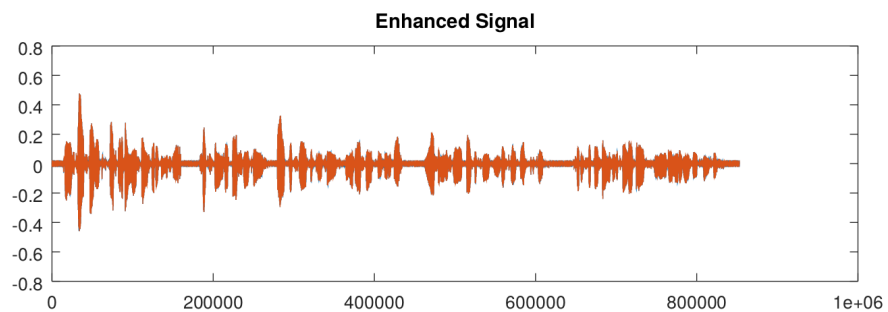
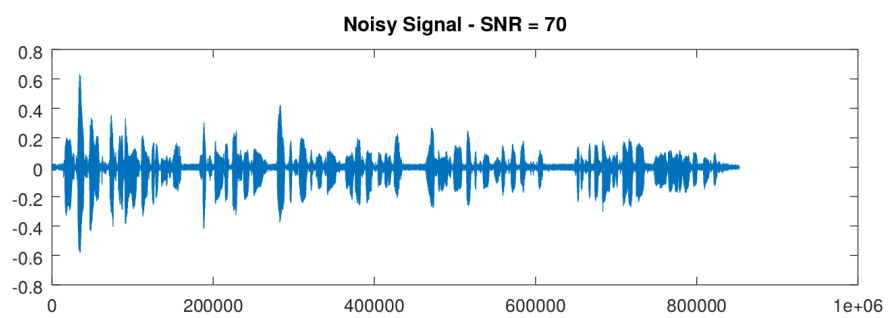
نمودار های خروجی در زیر آمده اند ولی برای دیدن نمودار های با کیفیت میتوانید از پوشه تصاویر کمک بگیرید که بصورت وکتور در پی دی اف قرار داده شده اند و با برای بررسی نمودار میتوانید کد را اجرا کرده و خروجی هر ۴ نمودار را روی صفحه ببینید.

اصلاحیه : بنده همه خروجی ها رو برای ویس ۳ انجام دادم (برای تست بود) در کد اصلاح شد و خروجی ها بررسی شد. برای دیدن خروجی فایل B3.m را اجرا کنید.

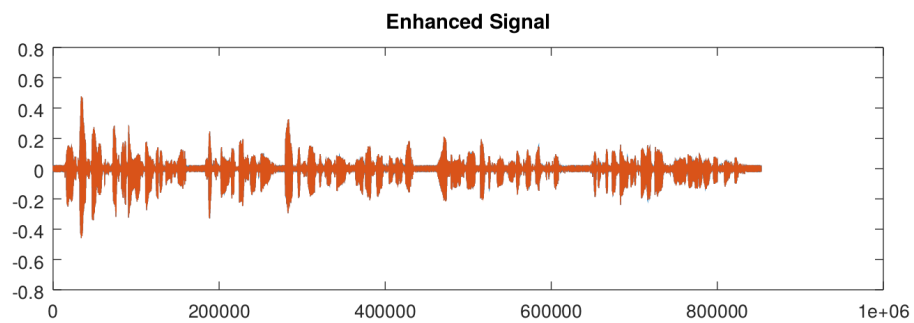
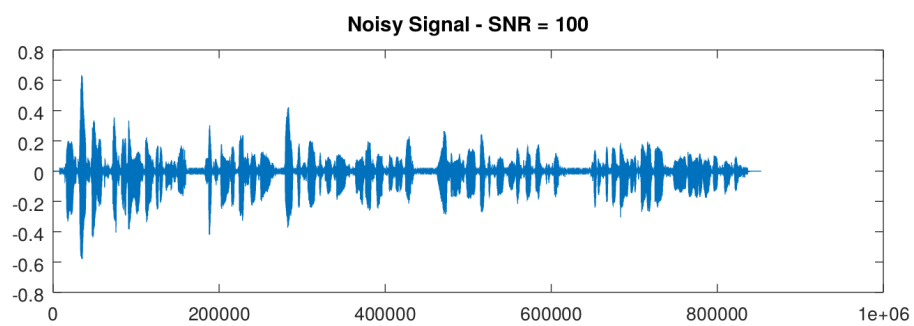
مشاهده می شود که تا حدودی سیگنال اولیه به دقت بازیابی میشود فقط در بعضی موارد که نویز زیاد است بعد از اجرای الگوریتم جاهایی که نیاز به افزایش یا همان بهبود نبود هم کمی افزایش داشته اند که خود عامل اضافی است.



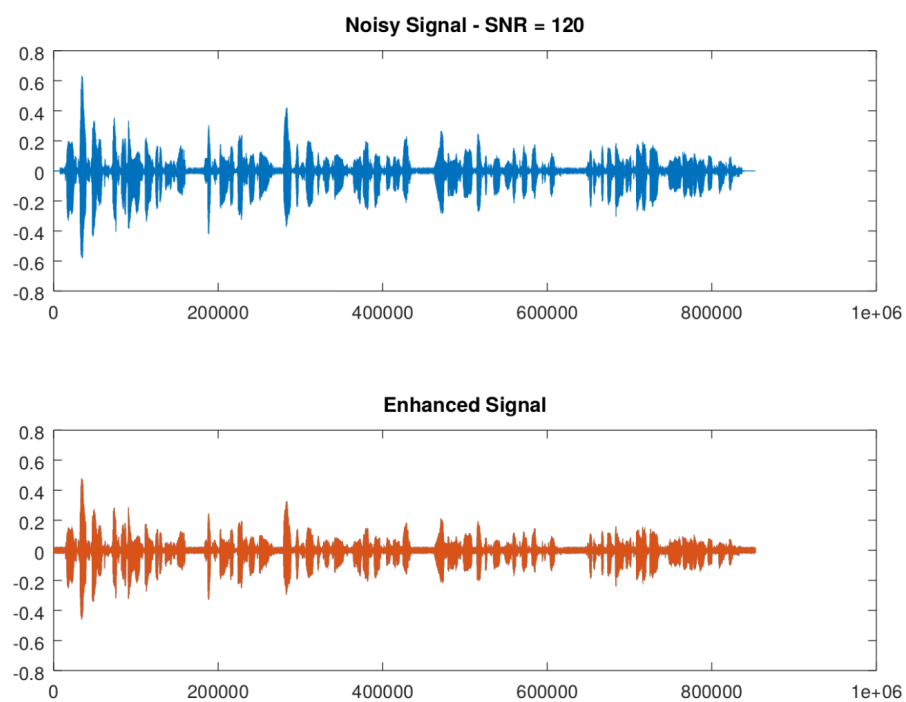
شکل ۳: خروجی برای نسبت سیگنال به نویز ۵۰



شکل ۴: خروجی برای نسبت سیگنال به نویز ۷۰



شکل ۵: خروجی برای نسبت سیگنال به نویز ۱۰۰



شکل ۶: خروجی برای نسبت سیگنال به نویز ۱۲۰