

پروژه امتیازی درس سیگنال‌ها و سیستم‌ها

نگار موقتیان، ۹۸۳۱۰۶۲

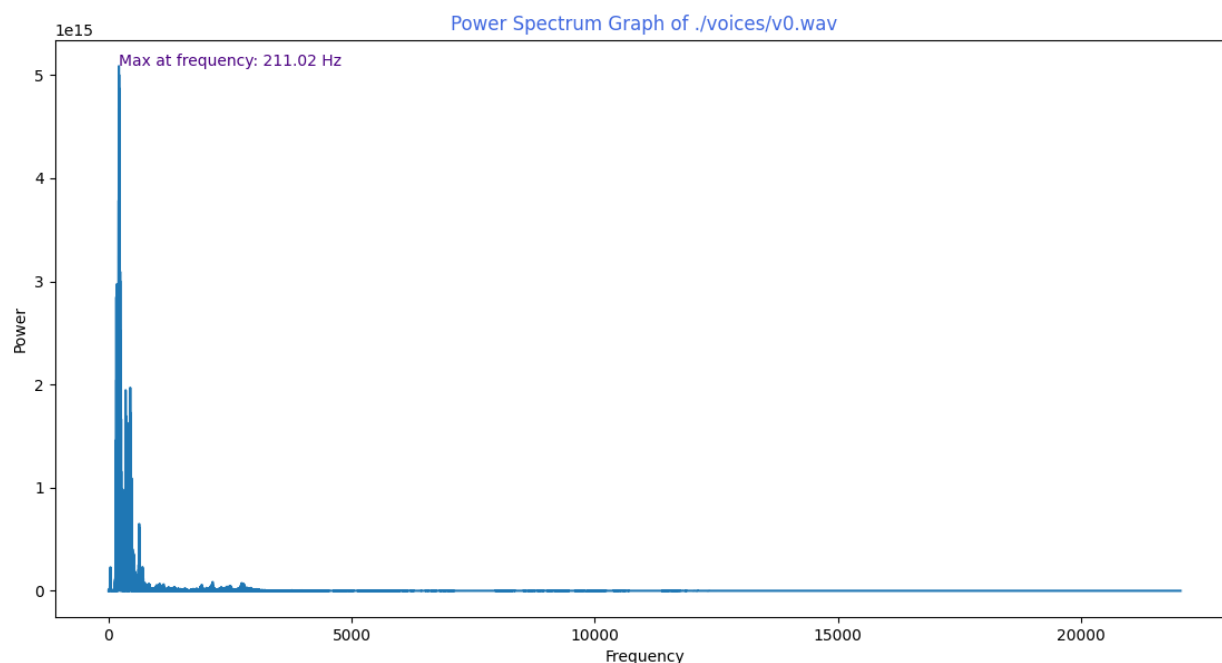
بخش اول – تشخیص صدا

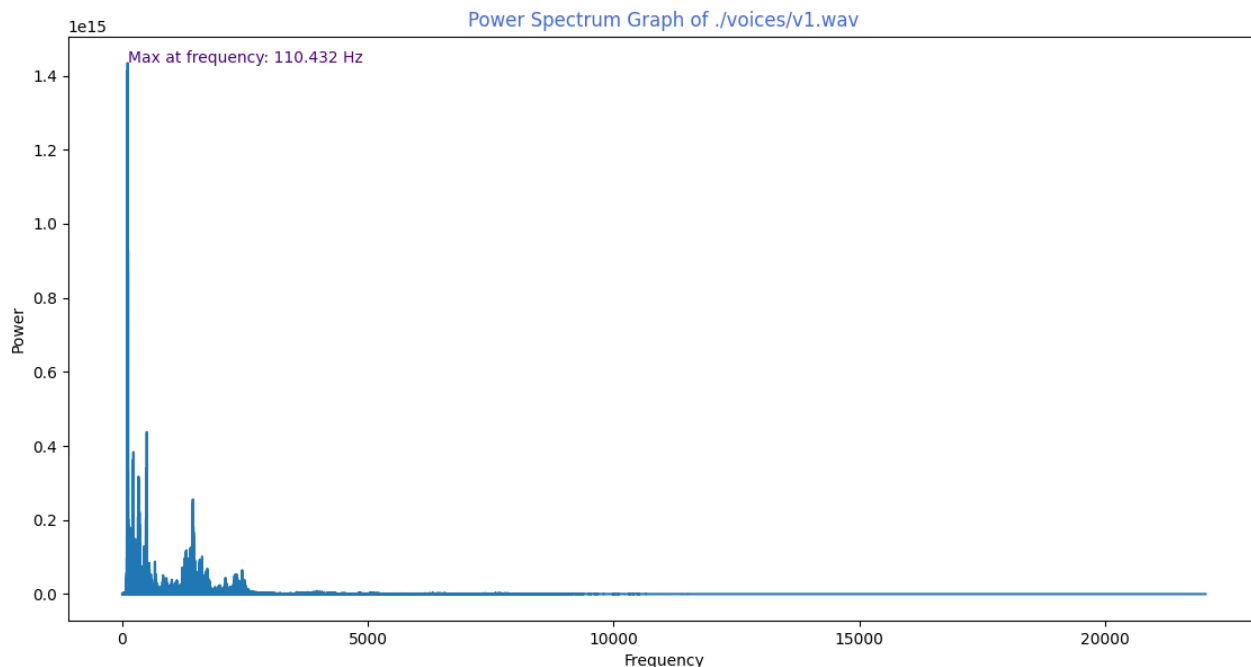
۱.

می‌دانیم اکثر سیگنال‌ها را با استفاده از تبدیل فوریه می‌توان به صورت جمعی از توابع سینوسی نوشت و می‌توان این توابع سینوسی را اجزاء تشکیل دهنده سیگنال دانست. همچنین می‌دانیم تبدیل فوریه هر سیگنال یک تابع یکتاست، بنابراین با بردن سیگنال به حوزه زمان در حقیقت داده یکسانی را از دو دید متفاوت بررسی می‌کنیم. بنابراین در تبدیل فوریه برای تجزیه و تحلیل طیفی، به جای تحلیل مستقیم سیگنال این اجزاء سازنده و نسبت آن‌ها با استفاده از تبدیل فوریه و در حوزه فرکانس بررسی می‌شوند. از این طریق می‌توان به اطلاعاتی دست یافت که استخراج آن‌ها در حوزه زمان دشوار است.

۲.

شکل‌های زیر نمودار power spectrum مربوط به یکی از صداهای زن و یکی از صداهای مرد را نشان می‌دهند.





همانطور که در این دو شکل مشاهده می‌شود، فرکانسی که به ازای آن توان مقدار ماکسیمم را دارد در صدای زنان (شکل اول - ۲۱۱ هرتز) از صدای مردان (شکل دوم - ۱۱۰ هرتز) به طور معمول بیش‌تر است. بنابراین از این طریق و با استفاده از تحلیل طیفی می‌توان میان صدای مردان و زنان تفکیک قائل شد.

۳.

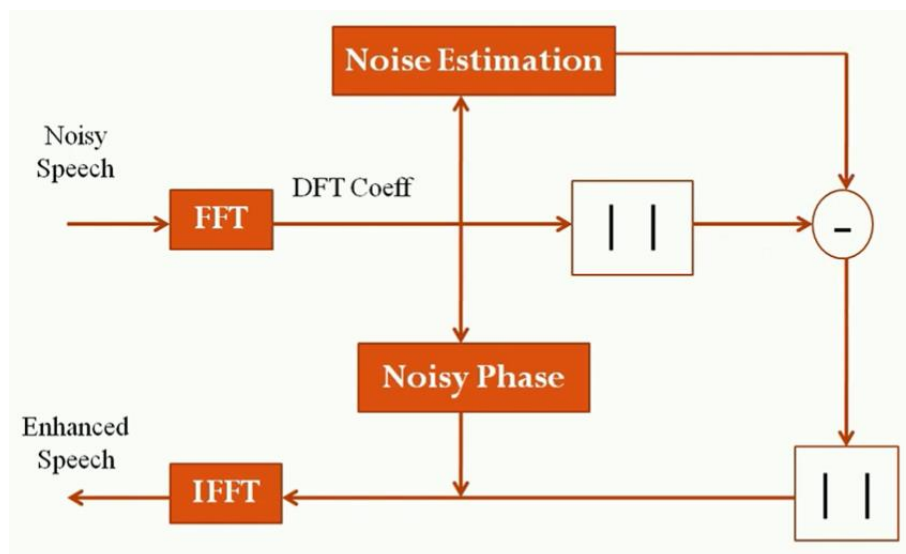
برای بهینه‌سازی زمان اجرا و کاهش پیچیدگی‌های این تابع می‌توان بازه فرکانسی مورد بررسی را محدود کرد. همانطور که در دستور کار نیز اشاره شده صدای معمول انسان میان ۵۰ تا ۳۰۰ هرتز متغیر است، بنابراین برای تشخیص جنسیت صدا نیازی نیست که طیف فرکانسی‌ای به گستردگی شکل بالا را بررسی کرده و تبدیل فوریه تابع را به ازای تمام این بازه محاسبه کنیم. بلکه کفایت برای مثال فرکانس‌هایی میان ۰ تا ۴۰۰ را مورد بررسی قرار دهیم.

همچنین برای افزایش دقت این برنامه می‌توان از یکسری داده training استفاده کرد که از پیش برچسب‌گذاری شده‌اند و از این طریق می‌توان بازه‌ای که برای صدای مردان و زنان مشخص شده را دقیق‌تر کرد.

بخش دوم - نویزگیری

۱.

Spectral subtraction روشی برای بازیابی توان و یا اندازه یک سیگنال نویزدار در حوزه فرکانس است. برای این کار ابتدا تخمینی از میانگین نویز زده شده و سپس این مقدار از سیگنال نویزدار کم می‌شود. روند کلی این روش در شکل زیر آمده است.



در ابتدا از سیگنال نویزدار تبدیل فوریه گرفته شده و اندازه و فاز آن جدا می‌شود. از طرف دیگر تخمینی از میانگین تبدیل فوریه نویز نیز زده می‌شود (در این روش فرض می‌شود نویز یکنواخت بوده و تبدیل فوریه آن در بازه‌های مورد بررسی تغییر زیادی ندارد). حال در صورتی که هدف ما بازیابی توان سیگنال باشد مربع میانگین نویز از مربع اندازه تبدیل فوریه سیگنال کم شده و از آن جذر گرفته می‌شود و اگر هدف بازیابی اندازه سیگنال باشد میانگین نویز از اندازه تبدیل فوریه سیگنال کم می‌شود (در برنامه نوشته شده از روش دوم استفاده شده است). در نهایت نیز فازی که از سیگنال نویزدار بدست آورده بودیم در این اندازه ضرب می‌شود و با گرفتن تبدیل فوریه معکوس از سیگنال حاصل، سیگنال بدون نویز تولید می‌شود.

الگوریتم پیاده‌سازی شده به طور کلی مانند توضیحات قسمت قبل است. در پیاده‌سازی این الگوریتم به جای FFT از STFT استفاده شده است. این تبدیل نوعی از تبدیل فوریه است که در آن به جای گرفتن تبدیل فوریه از کل سیگنال، سیگنال به بازه‌ها یا پنجره‌هایی (که با یکدیگر همپوشانی دارند) تقسیم شده و تبدیل فوریه در هر یک از این قسمت‌ها محاسبه و بررسی می‌شود. STFT به طور معمول در تحلیل سیگنال‌های صوتی استفاده می‌شود، زیرا در این سیگنال‌ها تغییرات زیادی در کل طول سیگنال داریم و لذا تبدیل فوریه را در بازه‌های کوچک‌تری بررسی می‌کنیم.

در برنامه نوشته شده این تبدیل با اندازه پنجره برابر با 30×10^{-3} و همپوشانی برابر با 50% از طریق تابع `audio_to_frames` اعمال شده است.

پس از آن با استفاده از تابع `estimate noise` میانگین نویز سیگنال در هر پنجره تخمین زده می‌شود. در این تابع بر روی سیگنال پیمایش شده و نویز هر پنجره طبق رابطه زیر از روی میزان نویز پنجره‌های قبلی محاسبه می‌شود.

$$y[n] = (1-\alpha)x[n] + \alpha y[n-1]$$

که در آن $y[n]$ مقدار تخمین زده شده و $x[n]$ خروجی بدست آمده از قسمت (با اعمال STFT) قبل است.

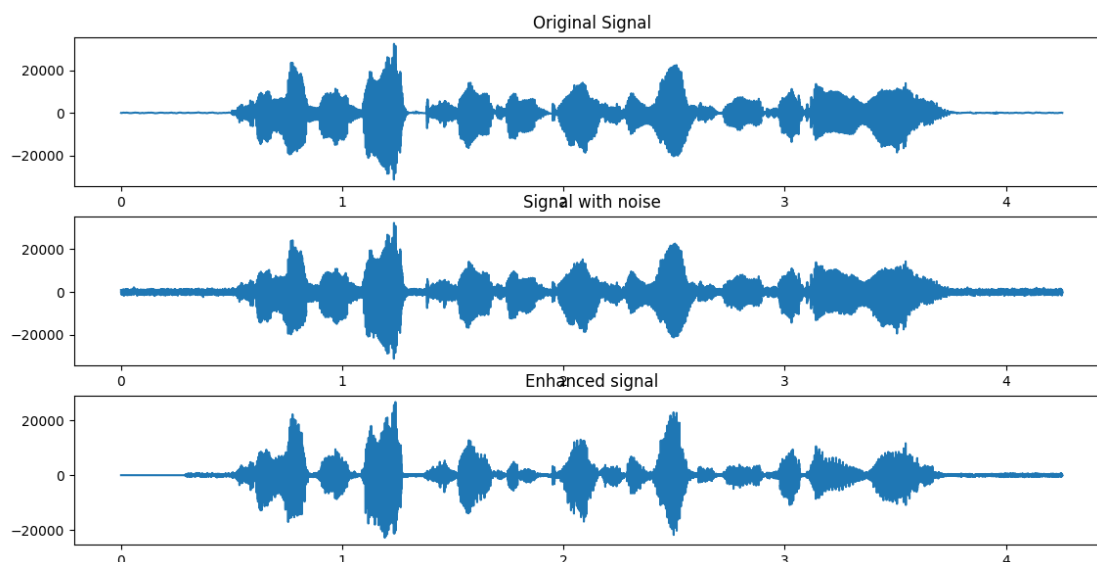
همچنین برای بدست آوردن مقدار آلفا در هر مرحله از رابطه زیر استفاده می‌شود.

$$\alpha(k) = \frac{1}{1 + e^{-a(\gamma(k)-T)}}$$

که در آن a برابر با ۲۵ و T برابر با ۱,۵ در نظر گرفته شده و گاما حاصل تقسیم مربع اندازه سیگنال نویزدار در هر مرحله بر روی مربع میانگین اندازه نویز تخمین زده شده در چند (در اینجا ۲۰) فریم قبلی است. همچنین مقدار نویز تخمین زده شده در ۲۰ فریم اول برابر با اندازه خود سیگنال قرار داده می‌شوند (فرض می‌شود که در ابتدای صوت تنها نویز داریم و شخص سکوت کرده است).

پس از تخمین نویز در تابع `spectral_subtraction` مرحله‌ای که در سوال ۱ توضیح داده شده بود طی شده و در نهایت با استفاده از تابع `frames_to_audio` توسط STFT معکوس به حوزه زمان برگردانده می‌شود.

شکل زیر مثالی از اجرا روند گفته شده می باشد.



همانطور که مشاهده می شود نویزها در شکل سوم کاهش داده شده اند. در چند فریم اول مقدار نویز برابر با مقدار خود سیگنال نویزدار قرار داده شده بود، لذا در ابتدای سیگنال خروجی هیچ نویزی نداریم. اما پس از آن مقدار نویز تخمین زده می شود و هنوز مقداری نویز داریم.

این روش برای نویزهایی با SNR ها مختلف کار می کند اما در صورتی که مقدار نویز بالا باشد به دلیل اینکه آن را تخمین می زنیم ممکن است شکل سیگنال اصلی پس از حذف نویز تغییر زیادی بکند و صدای فرد کیفیت اولیه را نداشته باشد.