دانشگاه مازندران

دانشکده مهندسی برق و کامپیوتر

پایان نامه دوره کارشناسی ارشد در رشته برق- الکترونیک

موضوع:

غنی سازی سیگنال گفتار مبتنی بر روش تفریق طیفی و تجزیه مقادیر منفرد

استاد راهنما:

دکتر محمدرضا کرمی

اساتید مشاور:

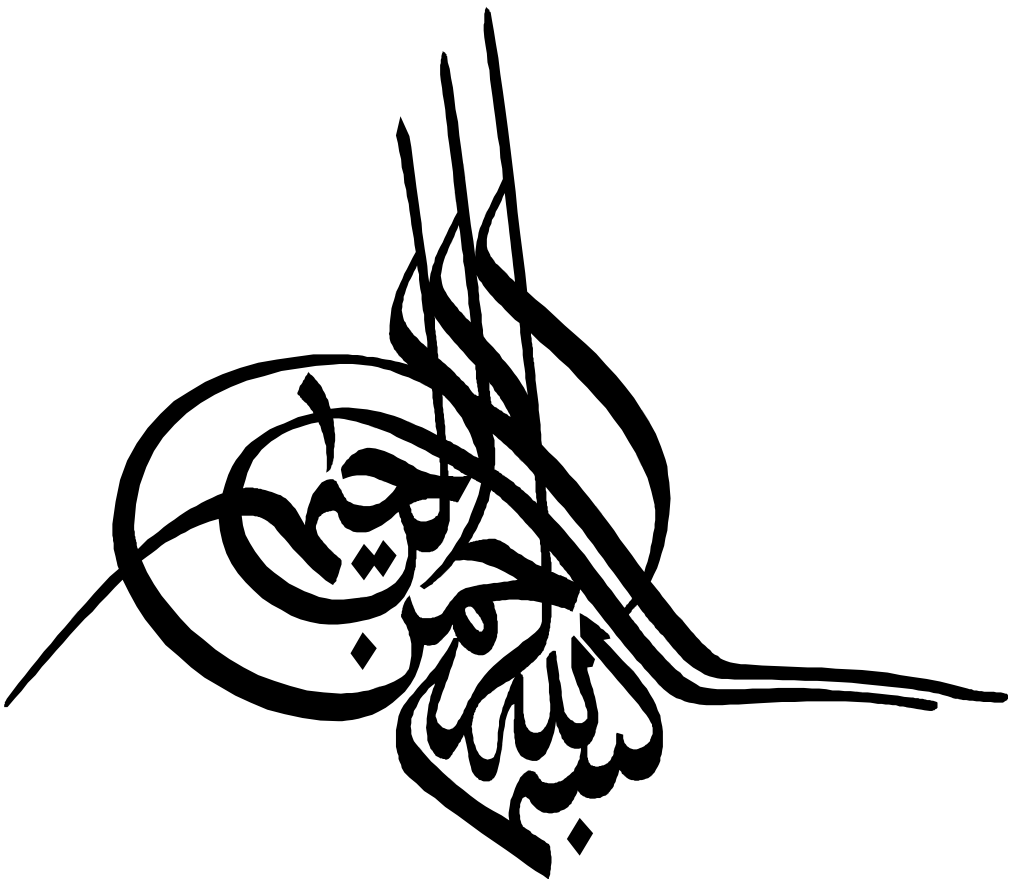
دکتر رضا قادری

دکتر حسین میارنعیمی

نام دانشجو:

مجتبی بندرآبادی

شهریور ۱۳۸۸


بسم الله الرحمن الرحيم

دانشگاه مازندران

دانشکده مهندسی برق و کامپیوتر

پایان نامه دوره کارشناسی ارشد در رشته برق- الکترونیک

موضوع:

غنی سازی سیگنال گفتار مبتنی بر روش تفریق طیفی و تجزیه مقادیر منفرد

استاد راهنما:

دکتر محمدرضا کرمی

اساتید مشاور:

دکتر رضا قادری- دکتر حسین میارنعیمی

اساتید داور:

دکتر محمد رضا ذهابی- دکتر عطاا... ابراهیم زاده

نام دانشجو:

مجتبی بندرآبادی

شهریور ۱۳۸۸

ت

سپاسگزاری

از استاد ارجمند جناب آقای **دکتر محمدرضا کرمی** به خاطر راهنمایی های بسیار مفیدشان در مطالعات، تحقیقات، نوشتن متون تخصصی و پیاده‌سازی الگوریتم ها کمال تشکر را دارم.

از اساتید گرامی، آقایان دکتر رضا قادری و دکتر حسین میارنعیمی به خاطر حمایت ها و راهنمایی های مفیدشان به عنوان اساتید مشاور در انجام این پایان نامه بسیار سپاسگزارم.

در پایان از خانواده و دوستان عزیز به خصوص آقایان مهندس آمارد افضلیان، دکتر جمال قاسمی و سرکار خانم پدیده پزشکی که با حمایت ها و پشتیبانی‌های خود در پیشبرد و تکمیل این پایان نامه مرا یاری نموده اند، کمال تشکر را دارم.

تقـدیم به

پدرم، که بزرگی رنج هایش را می‌شناسم و می‌دانم که توان جبران ذره‌ای از فداکاری‌هایش را نخواهم داشت.

مادرم، که به من خوب بودن و پاک بودن را آموخت و چگونه مهربان بودن و چگونه دوست‌داشتن را.

و ...

چکیده

در این پایان‌نامه ابتدا به بررسی خلاصه ای از چگونگی تولید گفتار در انسان پرداخته و سپس به بررسی اجمالی روش های عمده بهسازی گفتار و مسائل و مباحث موجود در این روش ها پرداخته شده است. در بررسی این روش ها سعی شده است تا ضمن بیان اصول و مبانی هر روش، نقاط ضعف و قوت آنها را ذکر شود. سپس در فصل دوم، روش مشهور تفریق طیفی[1] معرفی و بررسی شده است. در این فصل همچنین به روش های بهینه شده تفریق طیفی مثل تفریق طیفی تعمیم یافته، اصلاح مقادیر منفی، تفریق طیفی چند بانده و ... پرداخته شده است. در روش های مبتنی بر تفریق طیفی، روش پایه تفریق طیفی، تفریق طیفی تعمیم یافته، تفریق طیفی چند باند، پیاده-سازی و تحلیل و ارزیابی شدند. از مشکلات عمده در روش های مبتنی بر تفریق طیفی تولید نویز موزیکال می باشد که تلاش شد تا میزان تولید نویز موزیکال در روش بهبودیافته تفریق طیفی چند باند، کاهش یابد.

در فصل سوم به معرفی تجزیه مقادیر منفرد پرداخته می شود و توضیح کاملی در زمینه ریاضیات تجزیه مقادیر منفرد و خواص و قضایای مربوط به این تجزیه آورده شده است. همچنین روش های بدست آوردن مقادیر منفرد و بردارهای ویژه در این فصل توضیح داده شده است. سپس مروری بر کاربردهای این تجزیه به خصوص در پردازش سیگنال آورده شده است. در انتهای فصل نیز روش های تشخیص نواحی سکوت یا الگوریتم های VAD[2] بیان شده است.

در فصل چهارم، روش های پیشنهادی بهسازی گفتار بر اساس روش تفریق طیفی و تجزیه مقادیر منفرد ارایه شده اند. در این فصل ابتدا طریقه محاسبه نویز تخمینی با استفاده از تجزیه مقادیر منفرد توضیح داده شده است و سپس روش بهینه کردن نویز تخمینی با استفاده از الگوریتم های VAD و آشکار ساز حروف بی صدا و صدا دار بیان شده است. در تمامی روش های پیشنهادی حذف نویز مبتنی بر تفریق طیفی از تجزیه مقادیر منفرد به منظور تخمین نویز استفاده شده است. در روش های پیشنهادی سعی بر آن بوده است که حالت بهینه را با تغییر پارامتر ها

[1] Spectral subtraction
[2] Voice activity detector

بدست آوریم. در فصل پنجم نتایج حاصله از شبیه سازی روش های پیشنهادی آورده شده و با کارهای قبلی مقایسه و ارزیابی می شوند. در پایان، در فصل ششم به جمع بندی مطالب و موضوعات مورد بحث پرداخته شده است.

واژه های کلیدی:

غنی سازی سیگنال گفتار، تفریق طیفی، تجزیه مقادیر منفرد، آشکار ساز گفتار[3]، تبدیل موجک، طیف نگاره (اسپکتروگرام)، تست شنوائی.

[3] Voice activity detector

فهرست مطالب

عنوان **صفحه**

مقدمه ۱

فصل اول- مروری بر روش های عمده بهسازی گفتار

۱-۱- مبانی تولید سیگنال گفتار ۳

۱-۲- مفاهیم بهسازی گفتار ۷

۱-۳- کاربردهای عمده بهسازی گفتار ۸

۱-۴- طبقه بندی روش های بهسازی گفتار ۹

۱-۴-۱- کاهش نویز با استفاده از خاصیت تناوبی گفتار ۱۱

۱-۴-۲- بهسازی گفتار بر اساس مدل ۱۲

۱-۴-۳- تکنیک های مبتنی بر دامنه طیفی زمان-کوتاه ۱۳

۱-۴-۴- روش های مبتنی بر آستانه گذاری روی ضرایب موجک ۱۵

۱-۴-۵- حذف وفقی نویز ۱۶

فصل دوم- مروری بر روش های تفریق طیفی

۲-۱- تفریق طیفی ۱۹

۲-۲- تعمیم روش های تفریق طیفی ۲۱

۲-۳- مقادیر طیفی منفی ۲۳

۲-۴- اطلاحات روی تفریق طیفی ۲۴

۲-۴-۱- متوسط گیری اندازه ۲۴

۲-۴-۲- تعمیم تفریق طیفی ۲۵

۲-۴-۳- تفریق طیفی با استفاده از کف طیفی و تفریق بیش از حد ۲۵

۲-۴-۴- تفریق طیفی با تخمین گر MMSE STSA ۲۷

۲-۴-۵- تفریق طیفی بر مبنای ویژگی های ادراکی ۲۸

۲-۴-۶- روش های تفریق طیفی وابسته به فرکانس ۲۹

۲-۴-۷- تفریق طیفی چند باند ۳۰

۲-۴-۷-۱- پیش پردازش ۳۲

۲-۴-۷-۲- شرح تفریق طیفی چند باند ۳۴

فصل سوم- آشنائی با تجزیه مقادیر منفرد و الگوریتم های VAD

۳-۱- تجزیه مقادیر منفرد ۳۸
۳-۲- رابطه بین تجزیه مقادیر منفرد و تجزیه مقادیر ویژه ۳۹
۳-۳- مقادیر ویژه و بردارهای ویژه ۴۰
۳-۳-۱- روش یافتن مقادیر ویژه ۴۰
۳-۳-۲- مقادیر ویژه مختلط ۴۲
۳-۳-۳- مقادیر ویژه ماتریس های مثلثی و قطری ۴۳
۳-۳-۴- مقادیر ویژه ماتریس های هم خانواده با A ۴۴
۳-۳-۵- رابطه بین مقادیر ویژه ماتریس های AB و BA ۴۵
۳-۳-۶- تعیین چند جمله ای مشخصه یک ماتریس ۴۷
۳-۳-۶-۱- روش کریلف ۴۷
۳-۳-۶-۲- روش لوریبر ۴۹
۳-۳-۶-۳- روش ضرایب نامعین ۵۰
۳-۳-۷- روش های تبدیلی در یافتن مقادیر ویژه ۵۲
۳-۳-۷-۱- روش ژاکوبی ۵۲
۳-۳-۷-۲- روش گیونز ۵۴
۳-۳-۷-۳- روش هاوس هلدر ۵۵
۳-۳-۸- روش های مبتنی بر تجزیه ماتریس برای یافتن مقادیر ویژه ۵۷
۳-۳-۸-۱- روش LR برای یافتن مقادیر ویژه ۵۷
۳-۳-۸-۲- روش QR برای یافتن مقادیر ویژه ۶۰
۳-۴- کاربرد های تجزیه مقادیر منفرد ۶۲
۳-۴-۱- کاربرد تجزیه مقادیر منفرد در پردازش سیگنال ۶۳
۳-۵- الگوریتم های تشخیص نواحی سکوت و صدادار و بی صدا ۶۴
۳-۵-۱- تشخیص قسمت های سکوت با استفاده از قدرمطلق اندازه ۶۵
۳-۵-۲- روش نرخ عبور از صفر ۶۶
۳-۵-۳- روش های حوزه ی فرکانس ۶۷
۳-۵-۴- روش کدینگ پیشگویی خطی(LPC) ۶۸
۳-۵-۵- استفاده از تبدیل بسته موجک باز در تشخیص فریم های سکوت ۷۰

فصل چهارم- استفاده از تجزیه مقادیر منفرد در روش تفریق طیفی

۴-۱- استفاده از تجزیه مقادیر منفرد در روش تفریق طیفی (روش SVSS)..... ۷۴
۴-۱-۱- قسمت تخمین نویز..... ۷۵
۴-۱-۲- فیلتر کردن مقادیر منفرد..... ۷۶
۴-۱-۲-۱- تعیین آستانه در فریم سکوت..... ۸۰
۴-۱-۲-۲- تعیین آستانه در فریم بی صدا و صدا دار..... ۸۰
۴-۱-۳- قسمت تفریق طیفی..... ۸۵
۴-۲- تفریق طیفی با استفاده از SVD و الگوریتم های VADو V/UV detector..... ۸۵
۴-۳- تفریق طیفی چند بانده با استفاده از SVD و الگوریتم هایVADو V/UV detector (SVMBSS)..... ۸۶

فصل پنجم- نتایج پیاده سازی و ارزیابی

۵-۱- معیار های سنجش..... ۸۹
۵-۱-۱- نسبت سیگنال به نویز..... ۸۹
۵-۱-۲- اسپکتروگرام (طیف نگاره)..... ۸۹
۵-۱-۲-۱- طراحی اسپکتروگرام..... ۹۲
۵-۱-۲-۲- نحوه توزیع مقادیر در اسپکتروگرام..... ۹۵
۵-۱-۳- تست شنوائی..... ۹۷
۵-۲- بررسی و مقایسه نتایج حاصله..... ۹۷
۵-۲-۱- تست شنوائی..... ۱۰۴

فصل ششم- جمع‌بندی و نتیجه‌گیری

۶-۱- جمع‌بندی و نتیجه‌گیری ۱۰۸

فهرست جداول

عنوان **صفحه**

جدول ۱-۱- تکنیک های مختلف غنی سازی گفتار..... ۱۰
جدول ۴-۱- مقادیر بهینه درصد انرژی $\bar{\Sigma}_2$..... ۸۳
جدول ۵-۱- نتایج غنی سازی سیگنال گفتار آلوده به نویز سفید گوسی توسط روش های PSS، SVPSS، MBSS و SVMBSS..... ۱۰۳
جدول ۵-۲- نمره کیفیت شنیدن نویز..... ۱۰۵
جدول ۵-۳- نتایج تست شنوائی غنی سازی سیگنال گفتار آلوده به نویز سفید گوسی توسط چهار روش PSS، SVPSS، MBSS وSVMBSS..... ۱۰۷

فهرست اشکال

عنوان **صفحه**

شکل ۱-۱- برش عرضی سیستم صوتی انسان..... ۴
شکل ۱-۲- مدل زمان گسسته عمومی برای تولید سیگنال گفتار..... ۵
شکل ۱-۳- حرف بی صدا s..... ۶
شکل ۱-۴- حروف صدادار e..... ۶
شکل ۱-۵- تبدیل فوریه حرف بی صدای s..... ۶
شکل ۱-۶- تبدیل فوریه حرف صدا دار e..... ۷
شکل ۱-۷- بلوک دیاگرام یک سیستم بهسازی مبتنی بر خاصیت تناوبی گفتار..... ۱۱
شکل ۱-۸- بلوک دیاگرام یک سیستم غنی ساز مبتنی بر دامنه طیفی زمان - کوتاه..... ۱۳
شکل ۱-۹- بلوک دیاگرام الگوریتم حذف نویز در حوزه موجک..... ۱۵
شکل ۱-۱۰- دیاگرام کلی سیستم حذف کننده وفقی نویز..... ۱۷
شکل ۲-۱- بلوک دیاگرام تفریق طیفی توان (PSS)..... ۲۳
شکل ۲-۲- بلوک دیاگرام تفریق طیفی تعمیم یافته (GSS)..... ۲۳
شکل ۲-۳- نمودار تعیین a از روی SNR (با فرض $a_0 = 4$)..... ۲۶
شکل ۲-۴- SNR های مقطعی برای ۴ باند فرکانسی گفتار نویزی..... ۳۲
شکل ۳-۱- نمودار اندازه یک سیگنال صوتی بر حسب زمان..... ۶۶
شکل ۳-۲- بلوک دیاگرام آنالیز کپسترال..... ۶۸
شکل ۳-۳- میزان خطای خروجی فیلتر معکوس LPC به ازای حروف صدا دار و بی صدا..... ۶۹
شکل ۳-۴- مرز تصمیم برای جداسازی قسمت های صدادار سیگنال..... ۷۲
شکل ۴-۱- بلوک دیاگرام کلی روش ارائه شده SVD-Spectral Subtraction (SVSS)..... ۷۵
شکل ۴-۲- فریم بی صدا در حوزه زمان..... ۷۷
شکل ۴-۳- فریم صدا دار در حوزه زمان..... ۷۸
شکل ۴-۴- خود همبستگی در فریم سکوت آغشته به نویز..... ۷۸
شکل ۴-۵- خود همبستگی در فریم بی صدا..... ۷۹
شکل ۴-۶- خود همبستگی در فریم صدا دار..... ۷۹
شکل ۴-۷- مقایسه مقادیر منفرد فریم های بی صدا و صدا دار ۷۹
شکل ۴-۸- مقادیر منفرد فریم سکوت نویزی..... ۸۱
شکل ۴-۹- مقادیر منفرد فریم صدا دار (الف) تمیز (ب) نویزی با SNR=10 db ۸۲
شکل ۴-۱۰- مقادیر منفرد فریم بی صدا (الف) تمیز (ب) نویزی با SNR=10 db..... ۸۲

شکل ۴-۱۱- بلوک دیاگرام روش بدست آوردن مقادیر بهینه آستانه..... ۸۳
شکل ۵-۱- هیستوگرام مقادیر فرکانسی..... ۹۶
شکل ۵-۲- نتایج غنی سازی سیگنال گفتار آغشته به نویز سفید گوسی بصورت حالت گذرا..... ۹۹
شکل ۵-۳- نتایج غنی سازی سیگنال گفتار آغشته به نویز سفید گوسی بصورت طیف نگاره..... ۱۰۰
شکل ۵-۴- نتایج غنی سازی سیگنال گفتار آغشته به نویز سفید گوسی بصورت حالت گذرا..... ۱۰۱
شکل ۵-۵- نتایج غنی سازی سیگنال گفتار آغشته به نویز سفید گوسی بصورت طیف نگاره..... ۱۰۲
شکل ۵-۶- متوسط غنی سازی ده سیگنال گفتار مختلف آلوده به نویز سفید گوسی توسط چهار روش PSS، SVPSS، MBSS وSVMBSS..... ۱۰۴
شکل ۵-۷- نتایج تست شنوائی غنی سازی سیگنال گفتار آلوده به نویز سفید گوسی توسط چهار روش PSS، SVPSS، MBSS وSVMBSS..... ۱۰۶

مقدمه:

با رشد روز افزون استفاده از سیستم های گفتاری در کاربردهای عملی روزمره، نیاز بـه حفـظ کیفیـت گفتار به عنوان امری اجتناب ناپذیر مطرح شده است. شرایط ایده آل و عـاری از نـویز کـه در کارهـا و شـبیه سازی های آزمایشگاهی در نظر گرفته می شود، در بسیاری از کاربردهای واقعی به طور جدی نقـض گردیـده و برقراری آنها زیر سئوال رفته است. به عنوان مثـال اسـتفاده از تلفـن همـراه، سـمعک، سیسـتم بازشناسـی گفتار[1] و یا هر وسیله ارتباط گفتاری در یک محیط نویزی ، همگی از مواردی است که در آن حفـظ کیفیـت گفتار و بالا نگه داشتن کارایی سیستم، از اهمیت خاصی برخوردار است. بدون استفاده از روش های بهسـازی گفتار[2]، عملکرد این سیستم ها به شدت تنزل یافته و چه بسا که غیر قابل قبول می شود. از ایـن رو، مبحـث بهسازی گفتار به عنوان یکی از ضرورت های کاربردی و عملی، از زمینه های فعال تحقیقـاتی در سـال هـای اخیر بوده است.

فرضیات پایان نامه:

کلیه روش های شرح داده شده در پایان‌نامه و نیز کلیه روابط حاکم بر آنها، برمبنای سه فرض اساسی ذیل استوار می‌باشند.

- نویز به سیگنال اضافه می شود.
- فقط یک کانال در دسترس است.
- سیگنال گفتار و نویز ناهمبسته هستند.

به این ترتیب آغشتگی سیگنال گفتار به نویز را می توان به شکل زیر مدل کرد:

$$y(n) = s(n) + Gd(n)$$

[1] Speech recognition system
[2] Speech Enhancement

که s(n) بیانگر سیگنال اصلی و تمیز صوت ، d(n) نویز آلوده کننده و G ضریب کنترل کننده انرژی نویز (و لذا کنترل کننده SNR[3]) است. در این پایان‌نامه، این ضریب با $d(n)$ ادغام شده و یک سیگنال نویزی بصورت ذیل نمایش داده شده است:

$$y(n) = s(n) + d(n)$$

[3] Signal-to-noise ratio

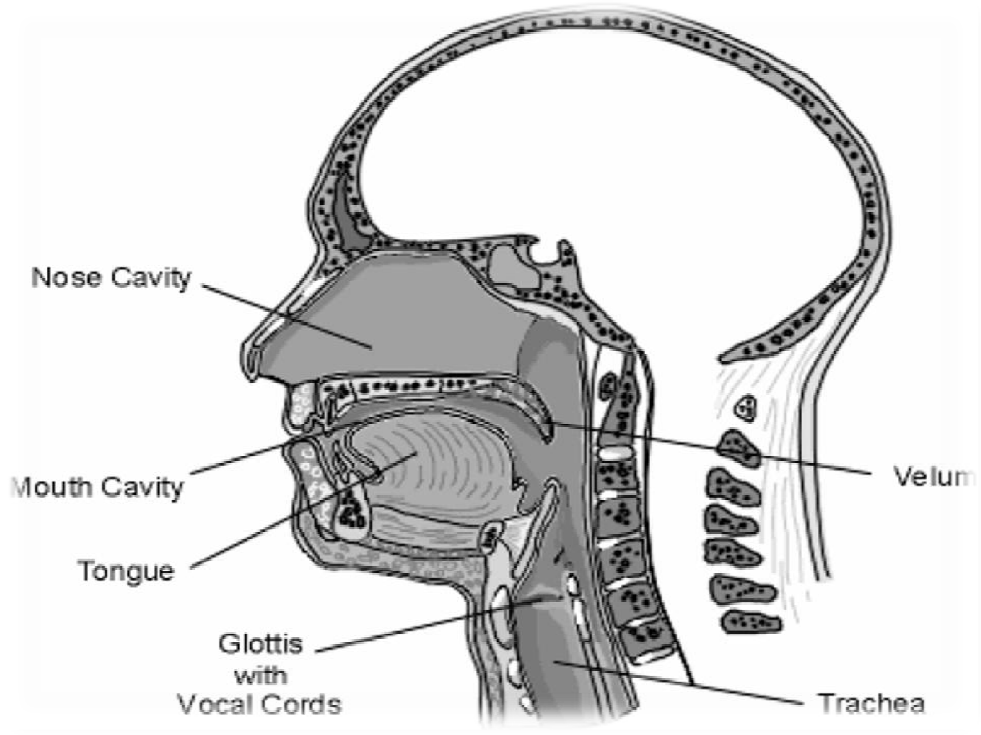
فصل اول

مروری بر روش های عمده بهسازی گفتار

در این فصل ابتدا به بررسی خلاصه ای از چگونگی تولید گفتار در انسان پرداخته و سپس به بررسی اجمالی روش های عمده بهسازی گفتار و مسائل و مباحث موجود در این روش ها می پردازیم. در بررسی این روش ها سعی شده است تا ضمن بیان اصول و مبانی هر روش، نقاط ضعف و قوت آنها را ذکر کنیم.

1-1- مبانی تولید سیگنال گفتار

سیگنال گفتار جهت برقراری ارتباط بین گوینده و شنونده مورد استفاده قرار می گیرد. گوینده ازطریق یکسری فرآیندهای عصبی و حرکات ماهیچه ای موج سیگنال گفتار را تولید می کند و شنونده از طریق سیستم شنیداری آن را دریافت کرده و سپس از پردازش آن را به سیگنال های عصبی بر می گرداند شکل (1-1) برش عرضی از سیستم تولید سیگنال گفتار را نشان می دهد. همانطورکه در شکل دیده می شود اجزای اصلی سیستم شامل شش ها، نای (لوله هوا)، حنجره (که رکن اصلی در تولید صوت است.)، حفره حلقی (گلو)، حفره دهانی (دهان) و حفره دماغی (بینی) می باشد.



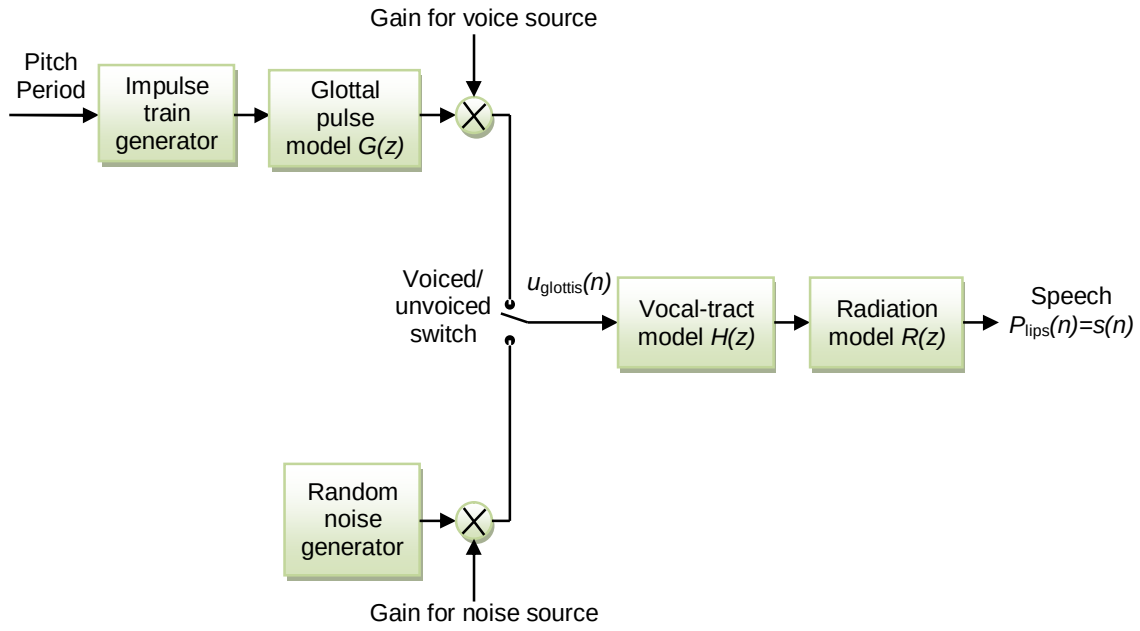
شکل1-1- برش عرضی سیستم صوتی انسان

درمباحث تکنیکی حفره های حلقی و دهانی را با هم در نظر گرفته و به آن حفره صوتی می گویند. بنابراین حفره صوتی از انتهای حنجره شروع و تا لب ها ادامه دارد.

برای تحلیل مهندسی سیستم تولید سیگنال گفتار، باید تولید سیگنال گفتاررابه شکل یک فیلتر مدل کرد. باتوجه به شکل(1-1) یک مدل خطی زمان گسسته در شکل (1-2) نشان داده شده است. در این مدل فرایند تولید سیگنال گفتار بر مبنای خصوصیات سیگنال خروجی نمایش داده شده است.

گفتار به دو دسته کلی حروف صدادار[1] و حروف بی صدا[2] تقسیم می گردد. همانطور که در شکل (1-2) دیده می شود منبع تحریک درتولید گفتار بی صدا ، نویز با طیف صاف می باشد که با یک تولید کننده نویز به شکل تصادفی[3] مدل می شود. همچنین هنگام تولید گفتار صدادار، تحریک از تخمین پریود گام[4] محلی استفاده می کند تا تولید کننده قطار ضربه را تنظیم کند تا فیلتر شکل دهنده پالس حنجره $G(Z)$

[1] Voice
[2] Unvoice
[3] Random Noise
[4] Pitch



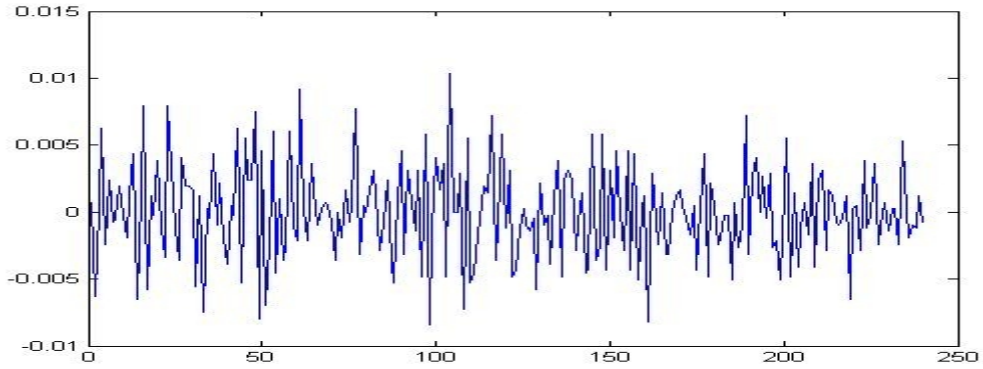
شکل1-2- مدل زمان گسسته عمومی برای تولید سیگنال گفتار

را تغذیه کند. تابع تبدیل $H(Z)$ نیز به شکل زیر مدل می شود.

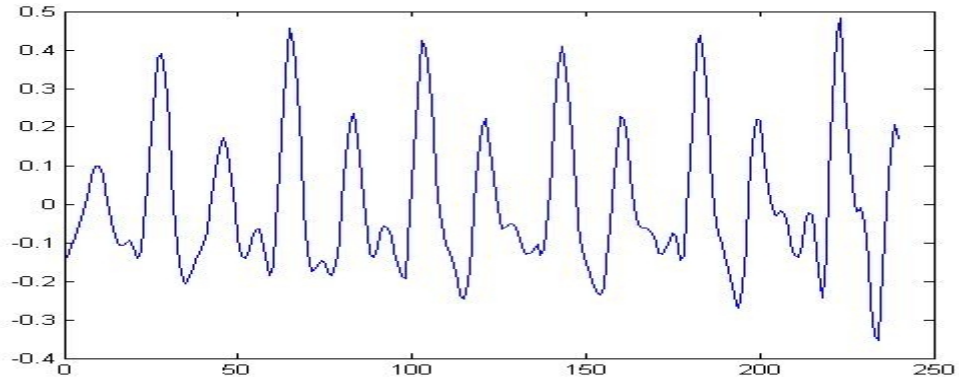
$$H(Z)=\frac{H_0}{1-\sum_{K=1}^{N} b_k Z^{-k}}=\frac{H_0}{\prod_{K=1}^{N}\left(1-P_k Z^{-1}\right)} \qquad (1\text{-}1)$$

که H_0 بیانگر ترم گین سراسری است و P_K مکان قطب موهومی برای مدل N لوله است. همانطور که ذکر شد، حروف صدادار شامل یک گرفتگی کوچک در حفره صوتی می باشد در حالیکه حروف بی صدا از یک گرفتگی قابل ملاحظه در حفره صوتی ایجاد می شود. لذا از لحاظ دامنه کوچکتر از حروف صدادار است و ماهیت نویزی دارد. برخی از تفاوت های حروف بی صدا و صدادار در شکل های (1-3) و (1-4) دیده می شود. حالت پریودیک حروف صدادار در شکل (1-4) به وضوح دیده می شود. برای مطالعه فرکانس های موجود در گفتار از تبدیل فوریه استفاده می شود. در شکل (1-5) تبدیل فوریه حرف بی صدای /s/ و در شکل (1-6) تبدیل فوریه حرف صدادار /e/ از کلمه seven نشان داده شده است. توجه کنید که در هر دوی این اصوات مخصوصا در حرف صدادار /e/ بعضی نواحی فرکانسی تقویت شده که به آنها نقاط تشدید می گوییم و در نواحی دیگر فرکانس ها تضعیف شده اند که به آنها غیر تشدید می گوییم. نواحی تشدید

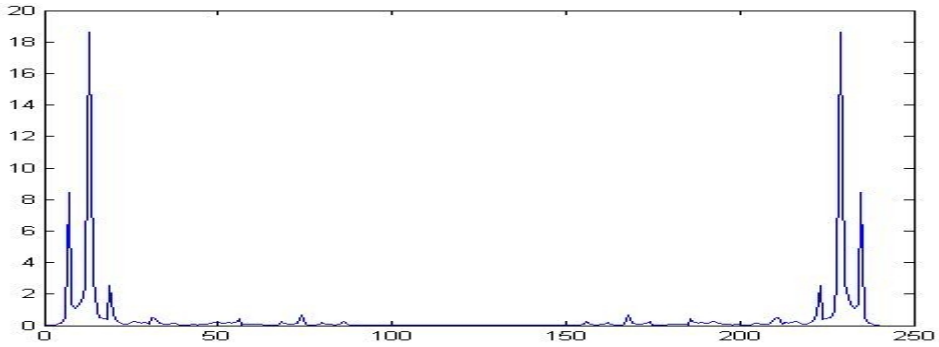
(فرکانس های تشدید) بستگی به شکل حفره صوتی دارند. در پردازش صوت بـه ایـن فرکـانس هـا اصـطلاحا فورمنت[5] می گویند .



شکل1-3- حرف بی صدا s

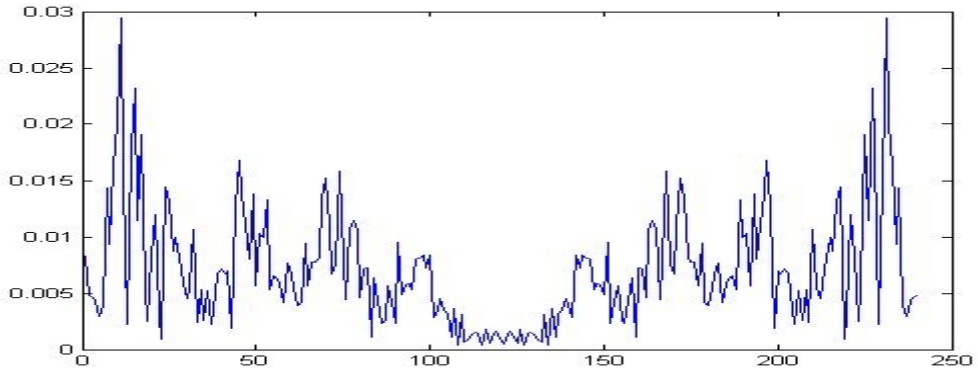


شکل1-4- حروف صدادار e



شکل 1-5- تبدیل فوریه حرف بی صدای s

[5] Formant



شکل 1-6- تبدیل فوریه حرف صدا دار e

همانطور که ذکر شد گفتار به دو دسته کلی صدا دار و بی صدا تقسیم بندی می شود. با این حال علاوه بر این دو دسته، گفتار تولیدی، بر حسب تحریک اعمال شده، به چند طبقه جز ئی تر قابل تقسیم بندی است. صوت ممکن است همزمان صدادار و بی صدا باشد. به این گونه اصوات، مرکب[6] می گویند. به عنوان مثال صدای حرف z در جمله three zebras از این نوع است. همچنین در بعضی حروف مثل t ابتدا یک گرفتگی کامل ایجاد می شود و در ادامه هوای جمع شده به یکباره از حفره دهانی خارج می شود. به اینگونه اصوات اصطلاحا انفجاری[7] می گویند.

1-2- مفاهیم بهسازی گفتار :

در قالب یک تعریف کلی، موضوع بهسازی گفتار عبارت است از تلاش برای بهبود عملکرد سیستم‌های ارتباط گفتاری در مواردی که سیگنال گفتار تحت تأثیر نویز، انعکاس، و سایر عوامل تخریبی واقع گردیده است. نیاز به بهسازی گفتار از آنجا ناشی می‌شود که سیگنال گفتار :

- یا از منبعی تولید می‌شود که در محیط نویزی قرار دارد.
- یا تحت تأثیر کانال انتشار و در اثر نویز یا انعکاس[8] دچار تخریب می‌شود.

[6] Mixed
[7] Plosive
[8] Echo