
الله
بسم الله الرحمن الرحيم

دانشکده: برق و رباتیک

گروه: الکترونیک

پایان نامه دوره کارشناسی ارشد مهندسی برق- الکترونیک

کاربرد مدل آماری مخلوط لاپلاس در بهسازی گفتار

زینب محمدپوری

اساتید راهنما:

جناب آقای دکتر حسین مروی

جناب آقای دکتر امید رضا معروضی

استاد مشاور:

جناب آقای دکترعلی سلیمانی

زمستان ۱۳۸۹

تقدیم به

پدر و مادرم مهربانم

که هرچه دارم از وجود پرمهر آنهاست

و همسرم

که همواره حامی و مشوق من بودند.

تشکر و قدردانی

خداوند متعال را به جهت الطاف بی پایان و خاصه اش که همـوار شـامل حـال من کرده بسیار شاکرم.

پس از آن بر خود لازم می دانم از زحمات و راهنمایی هـای اسـاتید گرانقـدرم جناب آقای دکتر مروی، دکتر معروضی و دکتر سلیمانی تشکر نمایم.

در ضمن از جناب آقای هادی ویسی به جهت راهنمایی ها و کمک هایشـان و جناب آقای امیرحسین آفریدون هم به جهت زحماتشان برای این پایـان نامـه، کمال تشکر را دارم.

چکیده

بهسازی گفتار یا Speech enhancement بیانگر گروه بزرگی از روش هاست کـه بـا انجـام پـردازش هایی روی سیگنال های نویزی، نهایتا منجر به بهبود کیفیت و قابلیت فهم گفتار می شود.

از مهم ترین روش های بهسازی گفتار، روش های آماری هستند که به لحاظ کارایی بالاتر نسبت به سایر روش ها و ایجاد اعوجاج کمتر در سیگنال نهایی بیشتر مورد توجه می باشد. در این روش یک مدل آماری برای گفتار و نویز فرض می شود و پردازش ها بر مبنای این مدل صورت می گیرد. این مدل آماری برای ضرایب فوریه یا پارامترهای دیگر هر قاب سیگنال گفتار یا نویز فرض می شود.

در این پایان نامه، روش آماری جدیدی برای تخمین سیگنال تمیز از روی سیگنال نویزی در حضور نویز جمع شونده و مستقل از سیگنال تمیز، در حوزه فرکانس ارائه شده است. تا کنون از مدل های گوسی، لاپلاس، گاما و مخلوط گوسی ، برای مدل کردن طیف سیگنال گفتار استفاده شده؛ اما در این پایان نامه، توزیع مخلوط لاپلاس برای مدل کردن بخش های حقیقی و موهومی طیف گفتار پیشنهاد شده است. پارامترهای این مدل (میانگین ها، واریانس ها و ضرایب) به صورت برون خطی و با استفاده از الگوریتم EM و سیگنال صحبت بدون نویز از پایگاه داده TIMIT محاسبه شده است. سپس رابطه تخمین گرMMSE یا کمترین میانگین مربعات خطا، با توجه به توزیع مخلوط لاپلاس به دست آورده شده است. واریانس نویز به صورت درون خطی و با استفاده از روش ردیابی کمینه ها به دست آمده است.

در نهایت برای ارزیابی و عملکرد روش پیشنهادی از سه معیار سیگنال به نویز قطعه‌ای، PESQ و Log Likellihood Ratio (LLR) استفاده شده؛ و روش پیشنهادی با روشهایی که مدل گوسی و لاپلاس را برای سیگنال تمیز فرض کرده اند، مقایسه شده است. بررسی نتایج بیانگر عملکرد قابل قبول روش پیشنهادی است.

کلمات کلیدی:

بهسازی گفتار، توزیع مخلوط لاپلاس، EM الگوریتم، تخمین گر MMSE، آمارگان کمینه

فهرست مقالات پذیرفته شده

۱- The Application Of Laplacian Mixture Model In Speech Enhancement: ICSAP CONFERENCE ۲۰۱۱, Singapore

۲- Speech Spectral Estimation Based on MMSE Estimator and Laplacian Mixture Model: ICECE CONFERENCE ۲۰۱۱, Dubai

۳- Speech Enhancement Using Mixture Of Laplacian Based MMSE Estimator: International Conference on Comunication System, Zahedan, ۲۰۱۰

فهرست مطالب

عنوان صفحه

۳- فصل سوم: مروری بر روش های بهسازی گفتار تک کاناله

۴- فصل چهارم: بهسازی گفتار با روش های آماری

۵- فصل پنجم: روش پیشنهادی بر مبنای توزیع مخلوطی از لاپلاس ها

فهرست شکل ها

عنوان صفحه

فهرست جدول ها

عنوان صفحه

فصل اول

مقدمه

۱-۱- مقدمه ای بر بهسازی گفتار

همان طور که می دانیم مهمترین وسیله ارتباطی افراد از طریق صدا و گفتار می‌باشد. به همین جهت پردازش گفتار، به جهت کاربردهای فراوان در زندگی بشر، از سالیان قبل مد نظر دانشمندان و محققین بوده است. یکی از اساسی ترین بخش های پردازش گفتار، بهسازی گفتار و تلاش برای بهبود کیفیت عملکرد سیستم های ارتباط گفتاری می‌باشد.

شاید تا به حال برای شما پیش آمده باشد که از یک نوار کاست قدیمی که حاوی سخنرانی یا اطلاعات مورد نظر شماست، استفاده کرده باشید و احتمالا به علت گذشت زمان یا مشکلات زمان ضبط، صداهای روی آن به سختی شنیده شود. قطعا شما دوست دارید که به طریقی کیفیت صداهای ضبط شده روی نوار کاست را بالا ببرید. یا حالتی را در نظر بگیرید که در داخل خودرو با تلفن همراه صحبت می کنید. قطعا صداهای مزاحم نیز مانند صدای باد، موتور ماشین و سرنشینان همراه با صدای شما وارد میکروفون می‌شوند.

این موارد تنها مثال های ساده ای از کاربردهای فراوان روش های کاهش یا حذف اثرات نویز بر سیگنال صوت است که به روش های بهسازی گفتار معروف هستند.

در قالب یک تعریف کلی، بهسازی گفتار عبارت است از تلاش برای بهبود کیفیت عملکرد سیستم های ارتباط گفتاری در مواردی که سیگنال گفتار تحت تاثیر نویز، انعکاس[۱] و سایر عوامل تخریبی واقع گردیده است.

نیاز به بهسازی گفتار از آنجا ناشی می شود که سیگنال گفتار معمولا، یا از منبعی تولید می‌شود که در محیط نویزی قرار دارد، یا تحت تاثیر کانال انتشار و در اثر نویز یا انعکاس دچار تخریب می‌شود، یا در گیرنده به نویز آلوده می‌شود و یا ممکن است منشا نویز ترکیبی از سه حالت فوق باشد [۶۶].

---

۱ - Echo

۱-۲- کاربردهای سیستم های بهسازی گفتار

سیستم های بهسازی گفتار کاربردهای وسیع و متفاوتی در حوزه های مختلف دارند که به طور کلی می‌توان آنها را به صورت زیر طبقه بندی کرد:

- تشخیص خودکار گفتار[۱] و تشخیص هویت گوینده[۲] : در سیستم های تشخیص خودکار گفتار و هویت گوینده ، عملکرد سیستم در محیط هایی با سیگنال به نویز پایین مناسب نمی‌باشد. ثابت شده که نرخ تشخیص با به کارگیری الگوریتم های بهسازی گفتار بهبود می یابد [۵۲].
- سیستم های تلفن: که در آنها گفتار اصلی به وسیله نویز زمینه یا نویز موجود در مسیر مخابره و نیز در اثر انعکاس صدای طرفین مکالمه خراب می‌شود.
- تلفن های عمومی: که در محیط های پر سر و صدا و شلوغ واقع اند.
- سیستم های مخابرات هوا به زمین: که در آنها نویز اتاق خلبان، پیام ارسال شده از سوی خلبان را خراب می کند.
- سمعک: که به عنوان یک تقویت کننده، هم سیگنال گفتار و هم نویز موجود در محیط را تقویت نموده و موجب ناراحتی کاربر می‌شود.
- حذف نویز در گفتارهای ضبط شده: که در آنها صدای ضبط شده اشخاص آغشته به نویز است و جهت بهبود کیفیت آنها، باید نویز را حذف کرد.
- کدینگ اطلاعات: که اثر نویز بر گفتار به شدت عملکرد آنها را تحت تاثیر قرار می‌دهد[۶۶].

۱-۳- تاریخچه روش های بهسازی گفتار

با توجه به کاربرد های مهم و فراوانی که برای روش های بهسازی گفتار بیان شد، و اهمیت آن، روش های مختلفی برای این کار پیشنهاد شده اند.

---

۱ - Automatic Speech Recognition
۲ - Speaker Identification

از مهمترین و قدیمی ترین روش های بهسازی گفتار می توان به فیلتر وینر[۱] [۶۰] و مشتقات آن اشاره کرد. در این روش ها، که برای اولین بار در سال ۱۹۴۹ وارد حوزه بهسازی گفتار گردید، سعی می‌شود که تخمین سیگنال تمیز بر اساس یک معیار بهینه سازی باشد. یکی دیگر از روش های قدیمی و بسیار معروف بهسازی گفتار، تفریق طیفی [۲] [۴] و [۶]، است که در سال ۱۹۷۵ پایه گذاری شد و در سال ۱۹۷۹ با پردازش در حوزه فرکانس مطرح شد. پس از آن بهبودهای زیادی بر روی این روش صورت گرفت. فیلتر کالمن[۳] هم در سال ۱۹۸۷با مقاله Paliwal & Basu [۴۷] وارد این حوزه گردید.

روش های آماری محض[۴] به عنوان یکی از مهم ترین روش های بهسازی گفتار، با کارهای Lim & Oppenheim [۳۵] و McAullay & Malpass [۴۵] و مقاله معروف Malah & Ephraim [۲۲] پیشنهاد شدند و بسیار مورد توجه محققین قرار گرفتند. روش های آماری مبتنی بر مدل[۵] نیز با کارهای Ephraim در [۲۴] و [۱۸] مطرح شدند و امروزه به عنوان یکی مهم ترین زمینه های علمی و کاربردی در بهسازی گفتار مورد توجه می‌باشند. می توان ادعا کرد که در کل روش های آماری چه از لحاظ کیفی و چه از لحاظ کمی نتایج بهتری نسبت به اغلب روشهای دیگر ارائه می‌دهند.

روش مهم دیگر بهسازی گفتار که مبتنی بر تئوری جبر خطی می‌باشد، زیر فضای سیگنال[۶] است که در ابتدا در [۱۴] و [۱۶] و [۲۵] و در سال ۱۹۹۱ارائه گردید و سپس بهبودهای زیادی روی آن صورت پذیرفت. اخیرا روش های مبتنی بر کپسترال[۷]، تبدیل موجک[۸] و تبدیل فوریه کسینوسی[۹] نیز نیز وارد حوزه بهسازی گفتار شده اند.

علاوه بر این روش ها، روش های دیگری نیز برای بهسازی گفتار پیشنهاد شده اند که در آنها از دو یا

---

۱ -Wiener Filter
۲ - Spectral Subtraction
۳ - Kalman Filter
۴ - Pure Statistical Method
۵ - Statistical Model-Based
۶ - Signal Subspace
۷ - Cepstral
۸ - Wavelet
۹ -Discerete Cosine Transform

چند میکروفن برای ضبط سیگنال نویزی استفاده می‌شود. از مهم ترین این روش ها می توان به روش فیلتر وفقی دوکاناله که در سال ۲۰۰۳ مطرح شد، اشاره کرد.

تا امروز مقالات و روش های بسیاری در زمینه بهسازی گفتار ارائه شده است، اما تخمین سیگنال تمیز از روی سیگنال نویزی هنوز یک مساله حل نشده در پردازش گفتار است و چالش های متعددی را در خود نگه داشته است. یکی از دلایل ساده برای اثبات این ادعا این است که علی رغم قدمت بیشتر بهسازی گفتار نسبت به برخی کاربردهای دیگر پردازش گفتار مثل بازشناسی گفتار، هنوز روش قطعی و غالبی در این زمینه وجود ندارد. از دلایل اصلی مشکل بودن حل این مساله، نامشخص بودن جواب دقیق برخی از سوالات کلیدی و ابهامات در این حوزه است. از جمله این سوالات عبارتند از:

- آیا می‌توان یک رابطه مشخص را برای معیارهای کیفیت و قابلیت فهم، به صورتی که انسان درک می کند، به دست آورد؟ پاسخ به این سوال، به نوبه خود منجر به یافتن معیار بهینه سازی معنی داری می شود که با ادراک انسان متناسب است. با داشتن روابط مشخص برای این دو، می توان یکی از آنها را بدون داشتن اثر منفی بر دیگری بهبود بخشید و می توان به این سوال نیز پاسخ داد که چه روش هایی برای بهسازی گفتار در کاربرد کمک شنیداری مناسب است؟ کدام یک برای کاربرد بازشناسی گفتار و کدام یک برای کدینگ؟
- چه پردازش هایی در سطوح مختلف سیستم شنوایی انسان انجام می‌گیرد و چگونه می‌توان آنها را مدل کرد؟ پاسخ این سوال نیز منجر به روشن شدن شبهات دیگری مانند نحوه ترکیب تکنیک های پردازش سیگنال با روش های ادراکی انسان می‌شود.

۱-۴- طبقه بندی روش های بهسازی گفتار

با توجه به تعدد روش های بهسازی گفتار، آنها را می‌توان به طرق مختلفی طبقه بندی کرد [۲۸] که در این فصل فقط آنها را نام می بریم. این طبقه بندی در فصل بعد توضیح داده خواهد شد.

۱- طبقه بندی بر اساس تعداد کانال ورودی

۲- طبقه بندی براساس حوزه فعالیت

۳- طبقه بندی بر اساس پارامتری و غیر پارامتری بودن

۱-۵- اهداف پایان نامه

در حال حاضر، رویکرد غالب در بهسازی گفتار تک کاناله، روش های آماری هستند که در کل کارایی بهتری در مقایسه با روش های دیگر دارند. این روش ها را که از دانش پیشین سیگنال تمیز و نویز بهره می‌گیرند، می‌توان در دو دسته آماری محض و آماری مبتنی بر مدل تقسیم بندی کرد. در روش های آماری مبتنی بر مدل، سعی می‌شود با استفاده از روش های مدل سازی، توزیع های گفتار تمیز و نویز به کمک داده های آموزشی به دست آید. سپس از این توزیع ها در تخمین سیگنال تمیز از روی سیگنال نویزی استفاده می‌شود. در مقابل، در روش های آماری محض، یک شکل توزیع خاص برای سیگنال و نویز فرض می‌شود و پارامترهای این توزیع ها با استفاده از سیگنال نویزی محاسبه می‌شوند. سپس با استفاده از این توزیع ها، روابط تخمین سیگنال تمیز از روی سیگنال نویزی به دست می آیند. در این پایان نامه از روش های آماری به منظور بهسازی استفاده شده است. اگرچه روش پیشنهادی را نمی‌توان به طور کامل متعلق به یکی از دو حوزه روش های آماری محض و آماری مبتنی بر مدل دانست، ولی شباهت آن به روش های آماری محض بیشتر می‌باشد. در این روش یک توزیع خاص برای سیگنال تمیز فرض می‌شود. ولی پارامترهای آن به جای استفاده از سیگنال نویزی، با کمک داده های آموزشی به دست می‌آیند. همچنین برای نویز نیز یک توزیع مشخص فرض می‌شود که پارامترهای آن از روی سیگنال نویزی محاسبه می‌شوند. هدف از انجام این پایان نامه و جهت گیری های اصلی آن را می‌توان در جملات زیر خلاصه کرد:

- انجام بهسازی گفتار به روش آماری محض، با به کارگیری توزیع مخلوط لاپلاس[۱] برای طیف طیف

گفتار تمیز، به منظور تخمین سیگنال تمیز از روی سیگنال نویزی در حضور نویز جمع شونده

---

[۱] -Laplacian Mixture Model

- بهبود روش های آماری محض مبتنی بر توزیع گوسی[۱] و لاپلاس[۲]
- ارزیابی سیستم بهسازی با معیارهای سیگنال به نویز قطعه ای[۳] ، LLR[۴] و PESQ[۵]

۱- ۶- ساختار پایان نامه

این تحقیق شامل شش فصل می‌باشد: تعاریف اولیه وکلیاتی درباره بهسازی گفتار در فصل جاری مطرح شد. در فصل دوم مختصرا در مورد بهسازی گفتار، عوامل موثر در طراحی سیستم های بهسازی گفتار و چگونگی ارزیابی سیستم های بهسازی گفتار بحث می گردد. فصل سوم نیز مروری بر مهم ترین روش های بهسازی تک کاناله بهسازی گفتار از جمله تفریق طیفی، زیرفضای سیگنال، فیلتر کالمن و روش های آماری می‌باشد.

در فصل چهارم نیز به بررسی روش های آماری، علی الخصوص روش آماری محض به سبب ارتباطی که با روش پیشنهادی دارد، پرداخته شده است. در این فصل تخمین گرهای بیشترین شباهت (ML)[۶] و روش بیز[۷] مرور شده و روش های بهسازی گفتار مرتبط با آنها معرفی می‌گردد. فصل پنجم هم به بیان روش پیشنهادی و نتایج آن می‌پردازد. این فصل شامل پنج بخش اساسی می‌باشد. این بخش ها عبارتند از: محاسبه روابط تخمین گر MMSE با فرض توزیع مخلوط لاپلاس برای طیف سیگنال گفتار تمیز، بسط EM[۸] الگوریتم برای محاسبه پارامترهای مخلوط لاپلاس، محاسبه واریانس نویز، ارزیابی نتایج روش پیشنهادی و مقایسه روش پیشنهادی با سه روش معروف بهسازی گفتار. فصل ششم نیز راجع به جمع بندی، نتیجه گیری و پیشنهادات برای کارهای آینده می باشد.

---

۱ - Gaussian
۲ - Laplacian
۳ -Segmental Signal to Noise
۴ -Log Likellihood Ratio
۵ - Perceptual Evaluation of Speech Quality
۶ - Maximum Likellihood
۷ - Bayesian
۸ - Expectation Maximization

فصل دوم

بهسازی گفتار

۲-۱- مقدمه

همان طور که قبلا بیان شد، بهسازی گفتار تلاش برای بهبود کیفیت عملکرد سیستم های ارتباط گفتاری در مواردی که سیگنال گفتار تحت تاثیر نویز، انعکاس و سایر عوامل تخریبی واقع گردیده است، می‌باشد. در این فصل به بررسی مفاهیم مرتبط با بهسازی گفتار پرداخته می‌شود و انواع طبقه بندی های موجود برای بهسازی گفتار بیان خواهد شد. همچنین مطالبی نیز راجع به روش های ارزیابی سیستم های بهسازی گفتار آورده شده است. منابع و مشخصات نویزهای صوتی نیز به عنوان یکی از عوامل مهم در شناخت ضرورت بهسازی گفتار، بررسی شده است.

۲-۲-بهسازی گفتار

بهسازی گفتار[۱] به روش هایی منتهی می‌شود که با انجام پردازش هایی روی سیگنال های نویزی، نهایتا منجر به بهبود کیفیت[۲] و قابلیت فهم[۳] گفتار می‌شود [۵۹].

کیفیت و قابلیت فهم دو معیار مهم ارزیابی یک سیگنال صوتی برای انسان می‌باشند. کیفیت یک معیار ذهنی[۴] ارزیابی سیگنال است که وضوح مطالب بیان شده و سطح نویز زمینه موجود در آن را اندازه می‌گیرد. می‌توان این معیار را با میزان خوشایند بودن سیگنال برای شنونده یا میزان تلاش شنونده برای درک مطلب بیان شده در سیگنال بیان کرد. در مقابل قابلیت فهم یک معیار عینی[۵] است که درصد کلماتی (که می توانند بی معنی باشند) که به درستی توسط شنونده تشخیص داده

---

۱ - Speech enhancement
۲-Quality
۳-Intelligibility
۴ - Subjective
۵ - Objective

شده را بیان می کند. این دو معیار از هم مستقل هستند، یعنی یک سیگنال صوتی ممکن است کیفیت بالایی داشته باشد، در حالیکه قابلیت فهم آن پایین باشد یا بالعکس. به عنوان مثال صدای تولید شده توسط ماشین دارای کیفیت بسیار پایین است، اما قابلیت فهم بالایی دارد [۲۱] .

بهبود کیفیت و افزایش قابلیت فهم سیگنال صوتی به صورت توام، از اهداف بهسازی گفتار است. اما تقریبا تمامی الگوریتم های بهسازی تک کاناله فقط کیفیت را بهبود می بخشند، در حالیکه قابلیت فهم را نه تنها افزایش نمی‌دهند، بلکه به علت اعوجاجی[۱] که در سیگنال ایجاد می‌کنند، آن را کاهش نیز می‌دهند. این مشکل به علت تاثیر گذاری روش بهسازی بر بخش های نویزگونه انرژی پایین سیگنال گفتار مانند صامت ها[۲] به وجود می‌آید. زیرا این روش ها معمولا صامت ها را نویز تشخیص داده و حذف می‌کنند. در حالیکه صامت ها نقش موثرتری نسبت به مصوت ها[۳] در فهم گفتار دارند [۲۰]. از این رو بیشتر شنونده ها اطلاعات بیشتری را از سیگنال نویزی در مقایسه با سیگنال بهسازی شده استخراج می‌کنند.

در برخی از کاربردها افزایش کیفیت سیگنال مورد توجه است. به ویژه زمانی که شنونده برای زمانی طولانی در معرض نویز شدید قرار دارد. مانند محیط کارخانه یا کابین خلبان هلی کوپتر، در مقابل در برخی کاربردها مانند سیستم های نظامی، قابلیت فهم سیگنال از کیفیت آن مهم تر است.

علی رغم اینکه روش های مختلفی برای این دو معیار وجود دارند، اما هنور یک روش کمی دقیق که بتواند این دو معیار را بدون نیاز به شنونده انسانی محاسبه نماید، وجود ندارد. همین مساله یکی از مهم ترین دلایل مشکل بودن کار بهسازی گفتار است. چرا که تا به حال روش ریاضی دقیقی که با ادراک انسان سازگار باشد، برای این معیارها و در نتیجه محاسبه دقیق اعوجاج سیگنال، پیشنهاد نشده است. یکی از نکات کلیدی در طراحی یک سیستم بهسازی گفتار این است که روش بهسازی قادر به حذف نویز زمینه (افزایش کیفیت سیگنال) باشد، بدون اینکه سیگنال اصلی را دچار اعوجاج کند. این بدین معنی است که این سیستم ها در سیگنال بهسازی شده دو نوع اثر از خود به جای می

---

۱ - Distortion
۲ -consonant
۳ -vowel

گذارند، یکی حذف نویز زمینه و دیگری اعوجاج سیگنال. در صورت دسترسی به معیاری دقیق برای فرموله کردن کیفیت و قابلیت فهم، می توان کیفیت را بدون ایجاد اعوجاج در سیگنال، افزایش داد.

۲-۳- طبقه بندی روش های بهسازی گفتار

در طول دهه های اخیر و با پیشرفت سخت افزاری و نرم افزاری روش های مختلف پردازش سیگنال گفتار، طیف وسیعی از رویکردها و روش ها برای بهسازی گفتار ارائه گردیده است. با توجه به تعدد روش های بهسازی گفتار، آنها را به روش های متفاوتی طبقه بندی می‌کنند.

یکی از روش های تقسیم بندی تکنیک های بهسازی گفتار بر پایه تعداد کانال (میکروفون) ورودی می باشد. بر این اساس دو خانواده تک کاناله[۱] و چند کاناله[۲] برای گروه بندی روش های مختلف موجود می‌باشد. در روش تک کاناله تنها یک میکروفون ورودی در دسترس بوده و کلیه اطلاعات لازم باید از همین سیگنال استخراج شود. اساس کار این خانواده از روش ها، مبتنی بر ایستان[۳] بودن نویز می‌باشد. فرضی که در مواجهه با نویزهای غیرایستان[۴] زیر سئوال رفته و عملکرد روش را دچار مشکل می‌نماید. در روش های چندکاناله، دو یا چند گیرنده (میکروفن) در ورودی سیستم بهسازی گفتار مورد استفاده واقع می‌شود. افزایش تعداد میکروفون ها و یا کانال های ورودی، قدرت روش را در پاکسازی سیگنال نویزی بالا می‌برد و کار بهسازی را ساده تر می‌کند، اما در مقابل بر هزینه و پیچیدگی پیاده سازی سیستم نیز می‌افزاید.

از روش های تک کاناله می توان به تفریق طیفی، فیلتر وینر، زیرفضای سیگنال، فیلتر کالمن، روشهای مبتنی بر تبدیل ویولت[۵]، روشهای مبتنی برکپسترال[۶] و روش های آماری اشاره کرد. از روش های چند کاناله نیز می توان فیلتر وفقی[۱] (بیشتر به صورت دوکاناله)، جداسازی سیگنال[۲] (قابل

(قابل

۱- Single Channel
۲- Multi Channel
۳-Stationary
٤-Non-Stationary
٥ - Wavlate
٦ - Cepstral

استفاده هم به صورت تک کاناله و هم به صورت چند کاناله) و روش های تشکیل پرتو[۳] را نام برد.

روش های بهسازی گفتار را می توان براساس حوزه فعالیت نیز دسته بندی کرد که بر این اساس آنها را به دو دسته کلی حوزه زمان (مانند فیلترکالمن و فیلتروفقی) و حوزه فرکانس یا طیف[۴] (مانند تفریق طیفی و فیلتر وینر) تقسیم می‌کنند. هرچند برخی از روش ها مانند روش های مبتنی بر تبدیل موجک یا کپسترال به حوزه طیف نزدیکترند؛ اما نمی‌توان آنها را به طور مستقیم به یکی از دو دسته فوق نسبت داد.

بیشتر روش های بهسازی گفتار در حوزه ی فرکانس و روی طیف عمل می‌کنند. زیرا در حوزه طیف قابلیت جداسازی نویز و سیگنال بالاست و می‌توان روش های بهینه یا ابتکاری را به خوبی پیاده سازی کرد. همچنین، مستقل بودن مولفه های طیف امکان کار با هرکدام از آنها را به صورت مستقل فراهم می‌کند که این به نوبه خود منجر به ساده تر شدن روابط ریاضی وکاهش محاسبات می‌گردد. انتقال سیگنال به حوزه طیف در بیشتر مواقع به کمک تبدیل فوریه زمان کوتاه [۵]صورت می‌گیرد که منجر به تجزیه سیگنال به دو بخش دامنه[۶] و فاز[۷] یا دو بخش حقیقی و موهومی می‌شود. اغلب روش های بهسازی گفتارکه در حوزه طیف عمل می‌کنند، فقط اندازه دامنه یا مجذور اندازه طیف را پردازش می‌کنند و از فاز سیگنال نویزی به عنوان فاز تخمینی سیگنال تخمین زده شده استفاده می‌کنند. یکی از دلایل اصلی این مساله حساسیت کمتر سیستم شنوایی انسان به فاز در مقایسه با دامنه است. برخی روش های دیگر نیز به تخمین بخش های حقیقی و موهومی سیگنال، به صورت جداگانه می پردازند. سپس این نتایج را با هم ترکیب می‌کنند تا به تخمین نهایی طیف سیگنال تمیز برسند. این روش ها اگر چه از روش های تخمین دامنه و مربع دامنه پیچیده تر به نظر می-

---

[۱]- Adaptive Filter
[۲] -Signal Sepration
[۳] -Beam Forming
[۴] -Spectrum
[۵] -Short time Fourier Transform
[۶]- Amplitude
[۷]- Phase

رسند، اما نتایج آنها بر حسب سیگنال به نویز[۱] بهتر از این دو روش است[۹]. دراغلب روش های بهسازی گفتار، سیگنال نهایی، پس از اعمال تبدیل فوریه معکوس بر روی طیف سیگنال بهسازی شده، از روش هم پوشانی - جمع[۲] بدست می آید.

یکی دیگر از راه های دسته بندی کردن روشهای بهسازی گفتار تفکیک این روش ها براساس پارامتری یا غیر پارامتری بودن آنهاست. روش هایی مانند تفریق طیفی و زیر فضای سیگنال به دلیل کار بر روی خود سیگنال، روش های غیر پارامتری محسوب می شوند و روش های مبتنی برمدل مخفی مارکوف[۳] (HMM) و مدل مخلوط گوسی[۴] (GMM) به دلیل استفاده از مدل های آماری پارامتری برای سیگنال و نویز در ردیف روش های پارامتری قرار می‌گیرند [۲۸].

با وجود طبقه بندی های فوق، به دلیل تشابه زیاد روش های بهسازی گفتاری در ایده ها و پایه های بنیادی، تقسیم بندی آنها در قالب چند روش کاملا مجزا، کار ساده ای نیست. به گونه ای که می‌توان گفت برخی از روش ها، حالت کلی یا تعمیم یافته برخی روش های دیگر بوده و یا این که برخی از روش ها صرف نظر از استفاده از تبدیل هایی متفاوت در نمایش سیگنال (مانند فوریه و موجک) دارای اساس مشترکی هستند.

۲-۴-عوامل موثر در طراحی سیستم های بهسازی گفتار

علی رغم وجود روش های مختلف در بهسازی گفتار، باید توجه داشت که طراحی یک سیستم بهسازی گفتار به عوامل مختلفی وابسته است . از جمله این عوامل می توان به موارد زیر اشاره کرد:

- کاربرد مورد نظر: دربرخی کاربرد ها افزایش کیفیت مساله اصلی است مانند وسایل کمک شنوایی، ولی در برخی از آنها قابلیت فهم و درک (مانند کاربرد های نظامی و باز شناسی گفتار) مهم تر است.

---

۱ SNR
۲ - Overlap and Add
۳ - Hidden Markov Model
۴ -Gaussian Mixture Model

- نوع نویز و مشخصات آن: اینکه نویز ایستان یا غیر ایستان باشد، از جنس گفتار (مانند نویز همهمه)، جنس نویز زمینه (صدای ماشین ، پنکه و باد) و یا از نوع پژواک اتاق و نویز کانال باشد، یا اینکه نویز دارای پهنای باند بالا یا کم باشد، در طراحی سیستم بهسازی گفتار موثر است.
- رابطه نویز و گفتار تمیز: اثر نویز برگفتار می تواند جمع شونده[۱] (مثل نویزسفید)، کانوالو-شونده[۲] (مثل نویز کانال تلفن یا نویز پژواک[۳]) ویاضرب شونده[۴] (نویز محوسازی) باشد.
- نویز و گفتار می توانند همبسته[۵] یا غیر همبسته[۶] باشند.
- تعداد میکروفون ها: هر چه تعداد میکروفون ها بیشتر باشد، بهسازی گفتار راحت تر است[۵۹].

۲-۵- ارزیابی سیستم های بهسازی گفتار

مشابه سایر سیستم ها، برای ارزیابی کارایی روش های بهسازی گفتار و مقایسه تکنیک های مختلف با هم، نیازمند داشتن ابزارهایی هستیم که بتواند عملکرد یک سیستم بهسازی گفتار را به صورت کمی نشان دهند. هر چند یک انسان به عنوان شنونده، ممکن است معیارهای متفاوتی را برای سنجیدن کیفیت یک سیگنال در نظر داشته باشد، اما با توجه به محدودیت هایی که در فرموله کردن این معیارها وجود دارند، به روش های ساده تر و قابل اندازه گیری توجه می‌گردد.

ارزیابی یک سیستم بهسازی گفتار می تواند وابسته به کاربردی که از این سیستم مورد نظر است، باشد. مثلا اگر هدف از بهسازی گفتار، بهبود کارایی یک سیستم بازشناسی گفتار است، دقت یا صحت سیستم بازشناسی می‌تواند بهترین معیار ارزیابی برای روش های بهسازی گفتار باشد.

در یک نگاه کلی می توان تکنیک های ارزیابی کیفیت گفتار را به دو دسته ذهنی و عینی تقسیم بندی کرد.

معیارهای ذهنی، براساس نظر شنونده های انسانی بوده که به صورت شهودی رتبه ای را به سیگنال

---

۱ - Additive
۲ - Convolve
۳ - Reverberation
۴ -Multiply
۵ - correlated
۶ - Uncorrelated

گفتار در یک بازه ی امتیازی از قبل تعیین شده اختصاص می‌دهند. از معایب این معیارها این است که، نتایج یک آزمایش ذهنی تحت تاثیر عوامل زیادی از جمله محتوای آزمایش و علایق انجام دهنده آزمایش است. به علاوه این روش ها نیاز به دسترسی به تعداد افراد زیادی دارد که فرآیند ارزیابی را وقت گیر و پرهزینه می سازد.

در روش های عینی از تکنیک های محاسباتی جهت مقایسه سیگنال اصلی و سیگنال پردازش شده استفاده می‌کنند. اکثر روش های عینی اندازه گیری کمی کیفیت سیگنال را از روی مدل های سیستم شنوایی انسان انجام می‌دهند. امروزه به این روش ها در ارزیابی (علی رغم ارجحیت ارزیابی ذهنی) توجه بیشتری می‌شود[۳۷].

روش های عینی و ذهنی زیادی برای ارزیابی سیگنال گفتار ارایه شده است [۲۷] و [۳۷] که برخی از مهم ترین آنها در جدول (۲-۱) آورده شده است. از میان این روش ها محاسبه سیگنال به نویز، MOS[۱] و PESQ جهت ارزیابی کیفیت و روش DRT[۲] در ارزیابی قابلیت فهم پرکاربردتر می باشند. در ادامه این بخش معیارهای SNR ، PESQ و LLR ، با توجه به اینکه در فصل ۵ برای ارزیابی روش پیشنهادی استفاده شده اند، بررسی خواهند شد.

۲-۵-۱- معیار سیگنال به نویز (SNR)

محاسبه سیگنال به نویز از جمله قدیمی ترین و معمول ترین روش های ارزیابی عینی کیفیت سیگنال است که علاوه بر SNR کلی[۳]، انواع مختلفی ازآن مانند SNR قطعه ای و SNR قطعه ای وزن دار[۴] معرفی شده است. با فرض جمع شدن نویز d(n) با سیگنال گفتار تمیز s(n)، می‌توان سیگ-سیگ- نال نویزی را به صورت y(n) =s(n)+d(n)، در نظر گرفت. اگر $\hat{s}(n)$ را سیگنال بهسازی شده در نظر بگیریم، آنگاه رابطه $e(n)= s(n) - \hat{s}(n)$ =، معادل خطای بین سیگنال اصلی و سیگنال تخمین زده شده برای نمونه nام است. با استفاده از این فرضیات، SNR کلی برابر با نسبت انرژی

۱- Mean Opinion Score
۲- Diagnosis Rhythm Test
۳- Global SNR
۴- Frequency Weighted Segmental SNR

سیگنال اصلی به انرژی خطا تعریف می شود که رابطه آن بر حسب دسی بل به صورت زیر است:

$$SNR = ۱۰ \log_{۱۰} \frac{\Sigma_{n=۱}^{N} s^{۲}(n)}{\Sigma_{n=۱}^{N} [s(n)-\hat{s}(n)]^{۲}} \quad (۱-۲)$$

که درآن N تعداد کل نمونه های سیگنال است.

جدول(۲-۱) : مهم ترین روشهای ارزیابی سیستم های بهسازی گفتار [۵۹].

معیارهای عینی	
اسم روش	نوع ارزیابی
Articulation Index (AI)	قابلیت فهم
Copmposite Measure	کیفیت
Frequency Weighted Segmental SNR	کیفیت
Itakura Log-Likelihood Ratio(LLR)	کیفیت
Itakura Saito(IS)	کیفیت
Log Area Ratio (LAR)	کیفیت
Perceptual Evaluation of Speech Quality (PESQ)	کیفیت
Segmental SNR	کیفیت
Signal to Noise Ratio (SNR)	کیفیت
Weighted-Spectral Slop(WSS) (WSS)	کیفیت

معیارهای ذهنی	
اسم روش	نوع ارزیابی
Diagnosis Rhythm Test (DRT)	قابلیت فهم
Hearing in Noise Test (HINT)	قابلیت فهم
Modified Rhythm Test (MRT)	قابلیت فهم
Speech Perception in Noise (SPIN)	قابلیت فهم
Speech Reception Test (SRT)	قابلیت فهم
Diagnostic Accessibility Measure (DAM)	کیفیت
Isometric Absolute Jodgment (IAJ)	کیفیت
ITU-T P.۸۳۵	کیفیت

کیفیت	Mean Opinion Score (MOS)
کیفیت	Paired Accessibility Rating (PAR)
کیفیت	Parametric Absolute Jodgment (PAJ)
کیفیت	Quality Acceptance Rating Test (QUART)

بدیهی است که در این رابطه، به سیگنال تمیز نیاز داریم. بنابراین این رابطه بیشتر در شبیه سازی ها و برای زمانی که هر دو سیگنال تمیز و نویزی در دسترس هستند، کاربرد دارد. مهم ترین مزیت این

روش سادگی آن از نظر ریاضی است اما به علت عدم همبستگی با معیارهای ذهنی و یکسان فرض کردن کل سیگنال، در بسیاری از موارد تخمین مناسبی از کیفیت ارائه نمی دهد. می دانیم که در عمل، انرژی سیگنال گفتار با زمان تغییر می کند و بعضی از مولفه های سیگنال نقش بیشتری در تامین کیفیت ایفا می کنند. یکی از حالت هایی که می تواند به طور نادرست منجر به SNR بالا شود، وقتی است که سیگنال گفتار دارای قسمت های گفتاری متمرکزی باشد، در حالی که نویز بر روی قسمت های گفتار با انرژی کم اثر ادراکی شدیدی داشته باشد. از این رو اگر فرض کنیم که اعوجاج ناشی از نویز نوسانات انرژی کمی دارد، اندازه گیری می تواند قاب[1] به قاب تغییر کند.

این مسائل مبنای ارائه روش SNR قطعه ای است که SNR را در قاب های کوچک تر محاسبه می کنند و میانگین این مقادیر را به عنوان ارزیابی نهایی ارائه می کنند. رابطه این معیار به صورت زیر است که در آن M تعداد قاب ها و L طول هر قاب است.

$$\mathrm{SNR}_{seg} = \frac{10}{M}\sum_{f=1}^{M} \log_{10}\left[\frac{\sum_{n=Lf+1}^{Lf+L} s^2(n)}{\sum_{n=Lf+1}^{Lf+L}[s(n)-\hat{s}(n)]^2}\right] \qquad (۲-۲)$$

در برخی از موارد که قاب های سکوت نیز در سیگنال وجود دارند ممکن است این مقدار منفی شود. برای رفع این مشکل می توان بازه های سکوت را تشخیص داد و آنها را در محاسبه لحاظ نکرد. به علاوه، با در نظر گرفتن یک آستانه پایین و جایگزین کردن مقادیر کوچکتر از آن با مقدار آستانه، می توان از منفی شدن جلوگیری کرد. همچنین گوش انسان بین قاب هایی با مقادیر SNR بالاتر از ۳۵dB تمایز چندانی نمی گذارد، بنابراین یک آستانه بالا برای متعادل کردن SNR هایی که بسیار بالا هستند، به کار می رود. این دو آستانه که معمولا ۰ dB و ۳۵ dB است، باعث می شوند تا جواب نهایی در بازه قابل قبولی باقی بمانند. توسعه رابطه SNR به حوزه فرکانس با نام SNR قطعه ای وزندار، به صورت زیر است که در آن K تعداد کل باندهای فرکانسی و W_k هر کدام از آنهاست. M

[1] - frame

تعداد قاب ها، $F(f,k)$ و $\hat{F}(f,k)$ دامنه فیلتر بانک برای سیگنال تمیز و تخمینی در فرکانس kام از قاب f است [۳۷].

$$SR_{fw-seg} = \frac{10}{M}\sum_{f=1}^{M}\frac{\sum_{k=1}^{K} W_k \log_{10}\left[\frac{F^2(f,k)}{[F(f,k)-\hat{F}(f,k)]^2}\right]}{\sum_{n=1}^{K} W_k} \qquad (۳-۲)$$

۲-۵-۲- معیار PESQ

برخی از معیارهای ارزیابی مورد استفاده در بهسازی گفتار از حوزه های دیگر پردازش گفتار، از جمله سنتز[۱] و کدینگ گفتار گرفته شده اند که روش PESQ یکی از آنهاست. در سال ۲۰۰۰ مسابقه ای توسط ITU[۲] برای معرفی یک معیار عینی جدید ترتیب داده شد. هدف اصلی از تعریف این معیار عینی جدید قابلیت اطمینان بالا در شبکه های مختلف و شرایط متفاوت بود و در این میان PESQ معرفی شد. به طور خلاصه، PESQ یک معیار عینی برای تخمین یا پیش بینی امتیاز یک آزمایش واقعی (معیار ذهنی MOS) است. در این آزمایش واقعی، کیفیت سیگنال از طریق گوش دادن افراد مختلف به سیگنال صوتی و امتیاز دادن به آن توسط افراد، سنجیده می شود. ضریب همبستگی بین این روش و روش های ذهنی ۹۳.۵٪ است. علی رغم این همبستگی بالا بر این نکته که این معیار نباید به جای روش های ذهنی به کار گرفته شود، تاکید شده است. (امتیاز این روش به وسیله پایگاه داده بزرگی از معیار های ذهنی کالیبره شده است.)

در این روش سیگنال بهسازی شده با سیگنال اصلی مقایسه می شود و نمره ای بین ۰.۵- تا ۴.۵ داده می شود. هر چقدر که این معیار بیشتر باشد به معنای کیفیت بالاتر سیگنال صوتی است. در اولین گام برای محاسبه امتیاز PESQ ، ابتدا تاخیر بین دو سیگنال با مکانیزم هایی که در [ITU ۲۰۰۰] تشریح شده است، حذف می شود. سپس دو سیگنال با استفاده از یک مدل ادراکی با هم مقایسه می شوند. در این مقایسه سعی شده نمایشی که سیستم شنوایی انسان از سیگنال ایجاد می کند، مورد

---

[۱] - Speech Synthesis
[۲] - International Telecomunication Union

استفاده قرار گیرد که برای رسیدن به این هدف چند گام مانند تطبیق زمانی، تطبیق سطوح سیگنال ها با سطوح شنوایی، تبدیل زمان- فرکانس و سنجش شدت سیگنال در نظر گرفته شده است.

به صورت خلاصه می توان گفت که الگوریتم PESQ به این ترتیب عمل می کند. ابتدا هر دو سیگنال (اصلی و اعوجاج یافته) به یک سطح استاندارد شنوایی نگاشته شده و با فیلتری میان گذر با فرکانس های قطع ۳۰۰ و ۳۰۰۰ هرتز، فیلتر می شوند. در مرحله بعد دو سیگنال از لحاظ زمانی همزمان شده و به این ترتیب تاخیرهای زمانی (که در کار کدینگ به خاطر عبور از شبکه معمول تر است) حذف می شود. پس از این مرحله، سیگنال ها توسط یک تبدیل شنوایی پردازش شده و میزان بلندی آنها به دست می آید. تفاوت بلندی دو سیگنال محاسبه شده و در طول زمان و فرکانس میانگین گیری می شوند تا تخمینی از معیار MOS به دست آید. از لحاظ پیچیدگی محاسباتی PESQ پیچیده ترین معیار عینی است [۳۷].

۲-۵-۳- معیار LLR

پس از بررسی دو معیار بالا به بررسی LLR(Log Likellihood Ratio) می پردازیم که به صورت رابطه (۲-۴) تعریف می شود:

$$LLR=\log\left(\frac{a_s^T R_x a_s}{a_x^T R_x a_x}\right) \qquad (۴-۲)$$

که درآن a_x و a_s به ترتیب ضرایب LPC[۱] سیگنال تمیز و سیگنال بهسازی شده وR_x ماتریس همبستگی[۲] سیگنال تمیز است. معیار LLR برای همه پنجره های یک جمله حساب می شود. در واقع این مقدار برای هر قاب از سیگنال محاسبه شده و در نهایت از مقادیر به دست آمده برای قاب ها میانگین گرفته می‌شود [۳۱].

۲-۶- منابع و مشخصات نویزهای صوتی

---

۱ - Linear Predictive Coding
۲ -Autocorrelation

نویز یا صداهای مزاحم تقریبا در همه موارد همراه سیگنال صوت می‌باشند. زمانی که شما از خیابان عبور می‌کنید (صدای خودروها و ساخت و ساز و ...)، وقتی داخل خودرو نشسته اید و در حرکت هستید (صدای موتور ماشین، صدای باد و...) زمانی که به یک فروشگاه یا رستوران می‌روید (صدای همهمه دیگران)، وقتی که به دفتر محل کار خود می‌رسید (صدای کولر، زنگ تلفن و ...) و به طور خلاصه در همه حالات، شما با انواعی از نویزها، که هر کدام دارای ویژگی های خاصی هستند، سروکار دارید. به طور کلی نویزها را می توان به دو دسته ایستان و غیر ایستان تقسیم نمود. نویز ایستان، نویزی است که مشخصات آماری آن با زمان تغییر نمی‌کند. مانند صدای کولر. در مقابل نویز غیر ایستان، نویزی است که مشخصه های زمانی و طیفی آن متغیر باشد. مانند نویز فرودگاه که ترکیبی از نویز همهمه و سایر صداهاست.

اثر نویز برگفتار می تواند جمع شونده[۱] (مثل نویزسفید)، کانوالوشونده[۲] (مثل نویز کانال تلفن یا نویز پژواک) و یا ضرب شونده[۳] (مانند نویز محوسازی[۴]) باشد. جنس نویز می تواند شبیه به گفتار باشد و با گفتار همبستگی داشته باشد، یا فقط یک نویز زمینه (مانند صدای کولر و پنکه) بوده و غیر همبسته با گفتار باشد.

با توجه به تنوع موجود در شکل و مشخصات نویزها، واضح است که قبل از هرکاری نیاز به شناسایی دقیق نویزها از نظر مشخصات زمانی و طیفی و درک تفاوت بین آنها هستیم. از ویژگی های مهمی که انواع نویز را از هم متمایز می‌کند شکل طیف و توزیع انرژی آنها درحوزه فرکانس است. به عنوان مثال انرژی نویز ناشی از باد بیشتر در فرکانس های پایین حدود ۵۰۰ هرتز متمرکز است. اما توزیع انرژی نویز رستوران، پهنای باند فرکانسی وسیع تری را پوشش می‌دهد.

یکی دیگر از عواملی که در طراحی سیستم بهسازی گفتار به آن نیاز داریم، شناسایی میزان SNR درکاربردها و شرایط مختلف است.

---

۱ - additive
۲ - convolutive
۳ - Multiplicative
۴ - Fading Noise

یک مطالعه جامع در هشت محیط مختلف نشان داده SNR درمحیط های کم نویز مثل بیمارستان و کلاس بین ۵ تا ۱۵دسی بل و برای محیط های دارای نویز زیاد مثل قطار حتی به ۰dB هم می رسد. از این رو می توان گفت فاصله SNR موثر برای عملکرد سیستم های بهسازی گفتار۱۵-۰ dB است. همچنین فاصله بین فرستنده (گوینده) و گیرنده (شنونده) یکی دیگر از عوامل موثر در کاهش نسبت سیگنال به نویز است. مثلا در کلاس درس کیفیت صدای معلم برای افراد ردیف جلو بالاتر از افراد ردیف های آخر است [۵۹].

فصل سوم

مروری بر روش های بهسازی گفتار تک کاناله

۳-۱- مقدمه

بهسازی گفتار، به جهت اهمیت و کاربردهای مهم و فراوانی که دارد، یکی از شاخه های مورد توجه پردازش گفتار می باشد و تا کنون روش های متعددی برای آن پیشنهاد شده است. در این فصل مرور مختصری بر مهم ترین روش های بهسازی گفتار تک کاناله خواهیم داشت. روش های زیادی برای این مساله ارائه شده اند که می توان آنها را در دو گروه اصلی روش های آماری و غیرآماری تقسیم کرد. علی رغم وجود ماهیت تصادفی و آماری سیگنال، منظور از روش های غیرآماری، روش هایی مانند تفریق طیفی و زیر فضای سیگنال است که در آنها هیچ فرض خاصی در مورد شکل توزیع سیگنال یا طیف آن و اطلاعات پیشین مرتبط انجام نمی شود. در مقابل در روش های آماری فرض می شود که سیگنال یا طیف آن از توزیع خاصی پیروی می کنند. روش های آماری که خود به دو گونه آماری محض و مبتنی بر مدل تقسیم می شوند، در این فصل به طور خلاصه بررسی می شوند. در این فصل روش های تفریق طیفی، زیرفضای سیگنال، فیلتر کالمن و روش آماری، به عنوان مهم ترین و معروف ترین روش های بهسازی گفتار، بررسی می شود.

۳-۲- تفریق طیفی

تفریق طیفی از قدیمی ترین و رایج ترین الگوریتم های بهسازی گفتار است که به خاطر سادگی و پایین بودن محاسبات آن هنوز هم مورد توجه محققین قرار دارد. ایده این روش بسیار واضح و ساده است . اگر فرض شود که سیگنال گفتار تمیز $s(n)$ در اثر جمع شدن با نویز مستقل $d(n)$ به سیگنال نویزی $y(n)$ تبدیل شود (۳-۱)، آنگاه با اعمال تبدیل فوریه زمان کوتاه، رابطه (۳-۱)، به صورت رابطه (۳-۲) خواهد شد.

$$y(n) = s(n) + d(n) \qquad (۳-۱)$$

$$Y(\omega_k) = S(\omega_k) + D(\omega_k) \Rightarrow Y_k e^{j\theta_{y(k)}} = S_k e^{j\theta_{s(k)}} + D_k e^{j\theta_{d(k)}} \qquad (۳-۲)$$

که در آن $\omega_k = 2k\pi/N$ k=۰,۱,۲,…,N و N طول هر قاب سیگنال گفتار است. در این رابطه Y_k، S_k و D_k به ترتیب دامنه طیف سیگنال نویزی، تمیز و نویز هستند و $\theta_{y(k)}$، $\theta_{s(k)}$ و $\theta_{d(k)}$ به ترتیب فاز سیگنال های نویزی، تمیز و نویز هستند.

در این روش برای تخمین دامنه طیف سیگنال تمیز از روی سیگنال نویزی می توان به سادگی دامنه طیف نویز را از دامنه طیف سیگنال نویزی پیدا کرد. یعنی:

$$\hat{S}_K = Y_K + D_K \qquad (۳-۳)$$

نهایتا در تخمین نهایی می توان از فاز سیگنال نویزی به عنوان فاز سیگنال تمیز استفاده کرد.

$$\hat{S}(\omega_k) = S_k e^{j\theta_{s(k)}} = [Y_K + D_K]e^{j\theta_{y(k)}} \qquad (۴-۳)$$

همانگونه که از رابطه ی (۳-۴) مشخص است برای پیاده سازی به تخمینی ازدامنه ی طیف نویز نیاز داریم. روشهای مختلفی برای این کار وجود دارند که در تفریق طیفی معمولا از VAD[۱] یا آشکار کننده ی گفتار از غیرگفتار استفاده می شود. در این روش، فرض اساسی ایستان بودن موضعی نویز در فاصله میان دو سکوت ضروری است. زیرا فرض می شود مشخصات نویز در فاصله دو سکوت ثابت بوده و تخمین طیف نویز در این فاصله به روز نمی شود.

با وجود سادگی روش تفریق طیفی عمل تفریق باید با دقت انجام شود تا باعث ایجاد اعوجاج در سیگنال تخمینی نشود. در شرایطی به علت تخمین بیش از حد نویز ممکن است مقداری از اطلاعات سیگنال گفتار حذف شود. از طرف دیگر با تخمین کم نویز سیگنال بهسازی شده دارای نویز خواهد بود. یکی از مشکلاتی که معمولا در عمل در تفریق به وجود می آید منفی شدن برخی از مولفه های طیف تخمینی است که این معمولا به علت تخمین نامناسب طیف نویز ایجاد می شود. معمول ترین راه حل رفع این مشکل استفاده از یکسوسازهای نیم موج است که درآن مانند رابطه زیر مقادیر منفی با صفر یا یک مقدار مثبت کوچک جایگزین می گردد.

اگر: $|Y(\omega)| > |\hat{D}(\omega)|$

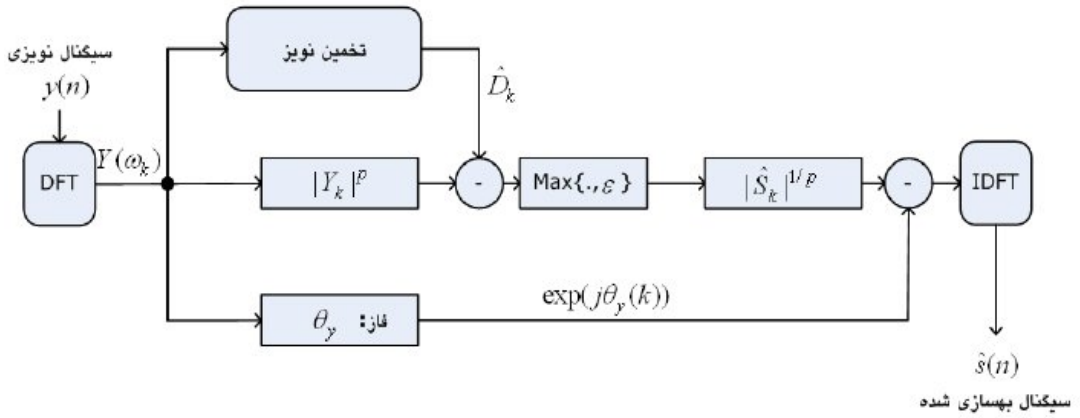
---

[۱] - Voice Activity Detection

آنگاه: $|\hat{S}(\omega)|=|Y(\omega)|-|\hat{D}(\omega)|$

در غیر این صورت: $|\hat{S}_i(\omega)|=0$

شکل (۳-۱) نمای کلی الگوریتم تفریق طیفی را نشان می دهد. همانگونه که اشاره شد روش های خانواده تفریق طیفی دارای مزایای سادگی و محاسبات پایین هستند. اما در مقابل دارای کاستی هایی نیز هستند. عمده ترین عیب این روش ایجاد اعوجاج در سیگنال نهایی و تولید نویزی موسوم به نویز موزیکال[۱] یا موسیقی است [۶] و [۴]. بسیاری از بهبودهای پیشنهاد شده روی این خانواده از روش ها برای حذف این اعوجاج ها و نویزهای ناخواسته بوده است. دلیل اصلی ایجاد این نویزها را می توان در منفی شدن طیف تخمینی که خود نتیجه ی تخمین نادرست طیف نویز است جستجوکرد. روش یکسوسازی نیم موج به علت غیر خطی بودن باعث ایجاد قله های کوچک و جدا از هم که توزیع آنها نیز تصادفی است در طیف سیگنال و در فرکانس های متفاوت می شود. این صداهای اضافی همان نویزهای موسیقی هستند که اثر آنها بر شنونده گاه از نویز اولیه هم آزاردهنده تر است. یکی دیگر از کاستی های این روش که در مقابل نویز موسیقی از اهمیت کمتری برخوردار



شکل (۳-۱): نمای کلی الگوریتم تفریق طیفی [۵۹].

جدول(۳-۱):مروری برروش تفریق طیفی.

[۱] - Musical Noise

فرضیات اساسی	نویز ایستان یا ایستان زمان کوتاه، مستقل و جمع شونده با گفتار گفتار ایستان زمان کوتاه
مزایا	ساده و دارای محاسبات پایین
معایب	ایجاد نویز زمینه (نویز موسیقی) در سیگنال بهسازی شده

است، استفاده از فاز سیگنال نویزی به عنوان فاز سیگنال بهسازی شده است. این جایگزینی در برخی از موارد می تواند باعث تغییرات زیاد و مداوم درکیفیت سیگنال شود. علت این جایگزینی کم اثر بودن فاز سیگنال نویزی بر روی کیفیت سیگنال گفتار و عدم حساسیت زیاد سیستم شنوایی انسان به آن است [۴۸]. اما فقط در صورت بالا بودن SNR سیگنال گفتار (بالاتر از ۸dB) می توان فاز سیگنال نویزی را بدون از دست دادن کیفیت به جای فاز سیگنال تمیز استفاده کرد [۵۸]. جدول (۳-۱) اطلاعات کلی راجع به روش تفریق طیفی ارئه می دهد.

۳-۳- زیرفضای سیگنال

اغلب روش های بهسازی براساس تئوری پردازش سیگنال و تخمین آماری هستند. اما روش های زیرفضای سیگنال مبتنی بر تئوری جبرخطی هستند. در واقع این روش بر این اصل استوار است که می توان سیگنال نویزی را به زیر فضاهایی تجزیه کرد که در آنها بخش های سیگنال تمیز و نویز متمایز می شوند و با صفرکردن بخش های نویزی در این زیر فضا می توان تخمینی از سیگنال تمیز بدست آورد. تجزیه سیگنال نویزی به دو زیر فضای سیگنال تمیز و نویز با استفاده از روش های شناخته شده تجزیه ماتریس های متعامد در جبرخطی مثل تجزیه مقادیر تکین[۱] (SVD) یا تجزیه براساس بردارهای ویژه[۲] و مقادیر ویژه[۳] امکان پذیر است. روش تجزیه بر اساس مقادیر ویژه و

[۱] - Singular Value Decomposition
[۲]- Eigen Vector
[۳]- Eigen Value

بردارهای ویژه به اسم تبدیل کارهونن- لاف[۱]، تبدیل هتلینگ[۲] و تحلیل اجزای اصلی[۳] نیز معروف است [۵۹]. جزئیات بیشتر درباره این روش در [۵۹] آمده است. مزایا و معایب این روش به طور خلاصه در جدول (۳-۲) آورده شده است.

جدول(۳-۲): مروری برروش زیر فضای سیگنال .

فرضیات اساسی	نویزو گفتار مستقل بوده و اطلاعات آماری درجه دوم آنها در دسترس است
مزایا	ساده و دارای محاسبات پایین
معایب	عدم کارایی بالا

۳-۴- فیلتر کالمن

فیلتر کالمن را می توان از دسته روش های پارامتری به حساب آورد، که در این بخش به طور خلاصه مرور می گردد. در این روش با استفاده از مدل AR[۴]، اطلاعات سیگنال گفتاراستخراج شده و مورد استفاده قرار می گیرد.

اولین استفاده از فیلترکالمن برای بهسازی گفتار را می توان در مقاله paliwal در سال ۱۹۸۷یافت [۴۷] که در آن فیلتر برای حذف نویز سفیدی که با گفتار جمع شده بود به کار رفت. پس از آن نیز روش های متفاوتی جهت بهبود و تعمیم فیلتر کالمن براساس ایده اصلی آن ارائه شد، که در اینجا فقط به بیان ایده اصلی اکتفا می کنیم.

فیلتر کالمن در واقع روشی برای کاهش مربعات خطا به صورت تطبیقی است که با استفاده از یک روش بازگشتی برای جداسازی سیگنال و نویز مورد استفاده قرار می گیرد. این الگوریتم از معادلات فضای حالت[۵] برای مدل کردن فرآیند تولید سیگنال گفتار و از معادلات مشاهده برای مدل کردن

---

۱ - Karhunen- Loeve Transform(KLT)
۲ - Hotelling Transform
۳ -Principal Component Analysis (PCA)
۴ - Auto Regressive
۵ - State Space

سیگنال آغشته به نویز استفاده می کند. برای این کار ابتدا یک مدل AR از مرتبه P به صورت رابطه (۳-۵) برای بردار گفتار تمیز درنظر می گیریم کـه در آن $\{\alpha_i, i = ۱,۲,\dots, P\}$ ضـرایب پـیش بینـی خطی و w(n) نویز سفید گوسی با میانگین صفر و واریانس $\sigma_w^۲$ است.

$$s(n) = \sum_{i=۱}^{p} \alpha_i s(n-i) + w(n) \qquad (۳-۵)$$

اگر فرض کنیم:

$$S(n) \triangleq [s(n), s(n-۱) \dots s(n-p+۱)]^T \qquad (۳-۶)$$

رابطه (۳-۵) را می توان به صورت زیر نوشت :

$$S(n) = FS(n-۱) + gw(n) \qquad (۳-۷)$$

که در آن:

$$F = \begin{bmatrix} \alpha_۱ & \alpha_۲ & \cdots & \alpha_p - ۱ & \alpha_p \\ ۱ & ۰ & \cdots & ۰ & ۰ \\ ۰ & ۱ & \ddots & ۰ & ۰ \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ ۰ & ۰ & \cdots & ۱ & ۰ \end{bmatrix}$$

$$g = [۰, ۰, \dots\dots, ۱]' \qquad (۳-۸)$$

همانگونه که قبلا بیان شد، داریم:

$$y(n) = s(n) + d(n) \qquad (۳-۹)$$

که در آن $y(n)$ مشاهدات یا همان سیگنال نویزی است. رابطه فوق را می توان به صورت زیر نوشت:

$$y(n) = h^T S(n) + d(n) \qquad (۳-۱۰)$$

که نمایش فضای حالت سیگنال نویزی است و در آن h=g می باشد. در این معادلات $S(n)$ بردار p×۱ بعدی سیگنال تمیز، معادل بردار حالت در لحظه n می باشد. F ماتریس p×p بعدی است که آن را ماتریس انتقال می نامند و واسط بین حالت n و n-۱ است. همچنین d(n) نیز بردار p×۱ بعدی نویز گوسی با میانگین صفر است [۴۷].

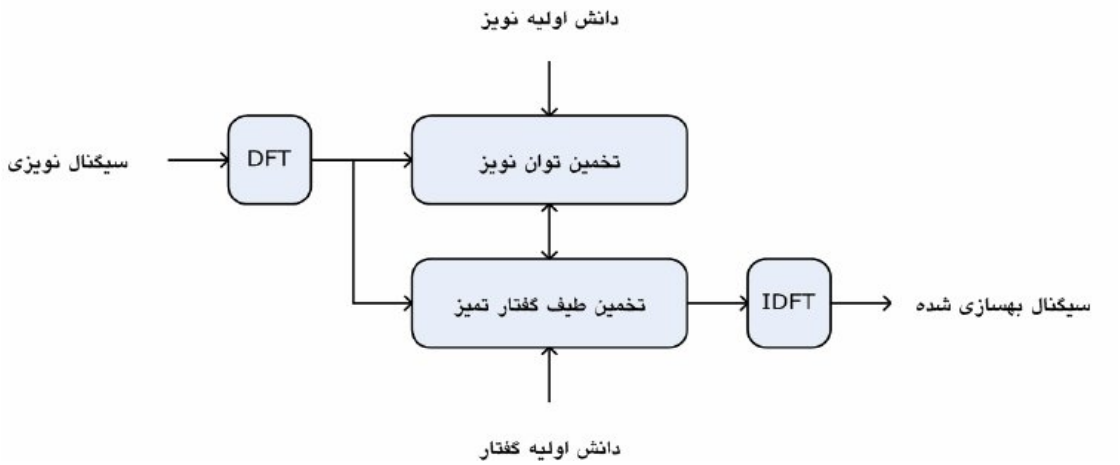
۳-۵- روش های آماری

روش های آماری به تکنیک هایی گفته می شود که در آنها یک مدل آماری برای سیگنال گفتار تمیز و نویز در نظر گرفته می شود و پردازش براساس این مدل ها صورت می گیرد. این روش به دو دسته روش های آماری محض و روش های مبتنی بر مدل تقسیم می شود: روش های آماری محض که در آن یک مدل آماری خاص (مثلا توزیع گاوسی) برای ضرایب فوریه یا پارامترهای دیگر هر قاب سیگنال گفتار و نویز فرض می شود و سپس روابط بهسازی گفتار متناسب با این مدل فرض شده به دست آید. این روش ها منجر به روابط و فیلترهایی می شوند که در سطح قاب عمل کرده و اطلاعات سیگنال گفتار و نویز را در همین سطح مدل سازی می کند. از این رو در برخی منابع به این روش ها، روش های مبتنی بر مدل کوتاه مدت یا روش های آماری بدون آموزش گفته می شود.

دسته دیگر از روش های آماری روش های مبتنی بر مدل یا روش های long term یا روش آموزش یافته هستند. در این روش ها به کمک داده های آموزشی مدل یا مدل هایی ایجاد می شود که خصوصیات مورد نظر را (که معمولا تابع توزیع احتمال و پارامترهای آن است) در خود دارد. در این روش با استفاده از داده های آموزشی، خصوصیات بلند مدت و مشترک بین قاب های مختلف فرا گرفته می شود.

بهسازی گفتار با روش های آماری محض توسط Lim and Oppenheim [۳۵] و McAullay and Malpass [۴۳] و مقاله معروف Ephraim andMalah [۲۱] شروع شده و امروزه به عنوان یکی از مهم ترین زمینه های علمی و کاربردی در بهسازی گفتارمورد توجه است. در خانواده روش های آماری مبتنی بر مدل و به طور خاص HMM نیز می توان به کارهای Ephraim [۱۷] و [۲۳] و [۱۹]اشاره کرد که بعد از آن استفاده از این رویکرد در بهسازی گفتار گسترش زیادی داشته است. شکل (۳-۲) نمای کلی روش های آماری را نمایش می دهد. همانطور که از شکل نیز مشخص است، در روش های آماری، علاوه بر توزیع های آماری که بخش جدانشدنی الگوریتم هستند، نیاز به یک

معیار اعوجاج یا تخمین گر نیز وجود دارد که بر اساس کمینه کردن اعوجاج یا بهینه کردن رابطه مربوط به تخمین گر، روابط تخمین سیگنال تمیز به دست می آید. از معروفترین معیارهای مورد استفاده در روش های آماری، می توان به تخمین گرهای کمترین میانگین مربعات خطا (MMSE)، بیشترین شباهت (ML) و بیشترین احتمال پسین (MAP) اشاره کرد که روابط مربوط به آنها در فصل بعد خواهد آمد.



شکل (۳-۲) نمای کلی روش های آماری [۵۹] .

امروزه به جرات می توان ادعا کرد که روش های آماری چه از لحاظ کیفی و چه از لحاظ کمی نتایج بهتری نسبت به اغلب روش های دیگر ارائه می دهند. جدول (۳-۳) خلاصه روش های آماری را بیان می کند. در فصل بعد توضیحات بیشتری درباره این روش بیان خواهد شد.

جدول(۳-۳): خلاصه روش های آماری.

فرضیات اساسی	نویز و گفتار مستقل وجود دانش قبلی از سیگنال تمیز و نویز مثلا نوع توزیع آنها
مزایا	کارایی بالاتر نسبت به سایر روش ها ایجاد نویز موسیقی کم اعوجاج کمتر در سیگنال
معایب	محاسبات و پیچیدگی زیاد

فصل چهارم

بهسازی گفتار با روش های آماری

۴-۱- مقدمه

از مدل های آماری در کاربردهای مختلف پردازش گفتار مانند بازشناسی گفتار، بهسازی گفتار و کدینگ به صورت موفقیت آمیزی استفاده شده است. رویکرد آماری در حل مساله بهسازی گفتار را می توان جزو روش های پارامتری به حساب آورد که در آن مدل های آماری، که معمولا توابع چگالی احتمال هستند، مجموعه ای از پارامترها را تشکیل می دهند. با استخراج توابع چگالی احتمال، رفتار تصادفی سیگنال مدل شده و می توان در کنار معیارهای اعوجاج، سیستم های بهینه برای بهسازی گفتار را به دست آورد. راه حل رایج برای مدل سازی آماری این است که سیگنال گفتار یک فرآیند تصادفی فرض شود که به صورت محلی و در سطح قاب های ۲۰ تا ۴۰ میلی ثانیه ای ایستان هستند. بر این اساس گفتار باید قاب به قاب پردازش شود. این روش ها به دو دسته روش های آماری محض و آماری مبتنی بر مدل تقسیم می شوند که در این فصل بررسی خواهند شد. همچنین تخمین گرهای بیشترین شباهت و روش بیز نیز مرور شده و روش های بهسازی مرتبط با آنها معرفی گردیده است.

۴-۲- روش های آماری محض

همان طور که قبلا گفته شد، در روش های آماری محض یک مدل آماری خاص (مثلا توزیع گوسی) برای ضرایب فوریه یا پارامترهای هر قاب سیگنال گفتار و نویز فرض می شود و سپس روابط بهسازی گفتار متناسب با این مدل فرض شده به دست می آید. هرچند می توان مدل های آماری را درحوزه زمان و در سطح نمونه های سیگنال نیز بدست آورد، اما این مدل ها معمولا در حوزه تبدیل فوریه فرض می شوند. در این روش ها ابتدا توزیع های آماری خاصی مثلا گوسی برای سیگنال بدون نویز و نویز در نظر گرفته می شود. سپس با استفاده از یک معیار تخمین گر، روابط تخمین سیگنال تمیز

بدست می آید.

قبلا گفته شد که یکی از اجزای روش های بهسازی گفتار آماری، تخمین گرها هستند. مهم ترین تخمین گرهای مورد استفاده در این حوزه، تخمین گر بیشترین شباهت، کمترین میانگین مربعات خطا و بیشترین احتمال پسین هستند که به طور مفصل توضیح داده خواهند شد.

مهم ترین روش های آماری محض عبارت اند از:

- فیلتر وینر با فرض گوسی بودن سیگنال تمیز و نویز و تخمین گر MMSE[۱] [۲] و [۶۰]
- تخمین دامنه طیف با MMSE [۲۲]
- تخمین لگاریتم دامنه طیف و توان دوم و Pام دامنه طیف سیگنال تمیز بر اساس تخمین MMSE [۲۳] و [۱] و [۶۳]
- تخمین دامنه ی طیف با تخمین گر ML [۴۵]
- تخمین طیف سیگنال تمیزبه روش MAP[۲] [۶۲] و [۳۸] و [۶۱]

۴-۳- تخمین گر بیشترین شباهت

بیشترین شباهت یکی از پرکاربردترین روش های تخمین آماری است که در کاربردهای مختلفی برای تخمین پارامترهای مورد نظر به کار می رود. در این روش فرض می شودکه L نمونه داده از y(n) داریم که به پارامتر نامعلوم δ وابسته اند و تابع چگالی احتمال y که با $p(y;\delta)$ نشان داده می شود در دسترس باشد. هدف این تخمین یافتن مقدار پارامتر δ به گونه ای است، که تابع چگالی احتمال y بیشینه شود:

$$\delta_{ML} = \operatorname{argmax} p(y;\delta) \qquad (۱-۴)$$

ابتدا فرض می شود متغیر نامعلوم معین[۳] و غیر تصادفی است. این مساله تخمین بیشترین شباهت را از تخمین بیزی که در آن فرض می شود پارامتر نامعلوم یک متغیر تصادفی[۱] است، متفاوت می کند.

---

۱ - Minimum Meam Square Error
۲ - Maximum A Posteriori
۳ -Deterministic

برای حل مساله فوق ابتدا یک شکل توزیع خاص مثلا گوسی برای تابع چگالی احتمال که به آن تابع شباهت[۲] نیز گفته می شود، فرض می شود و از آن نسبت به پارامتر نامعلوم مشتق گرفته مساوی صفر قرار داده می شود. می توان برای ساده تر شدن محاسبات از لگاریتم تابع شباهت[۳] استفاده کرد. مشابه قبل، ابتدا فرض می شود که سیگنال گفتار تمیز $s(n)$ در اثر جمع شدن با نویز $d(n)$ سیگنال نویزی $y(n)$ را طبق رابطه (۴-۲) تولید کرده است. نمایش این رابطه در حوزه فرکانس، به صورت رابطه (۴-۳) می باشد که در آن $\omega_k = ۲k\pi/N$ k=۰,۱,۲,…,N و N طول هر قاب گفتار است. در این رابطه $Y_{k'}$، S_k و $D_{k'}$ به ترتیب دامنه طیف سیگنال نویزی، تمیز و نویز و $\theta_{y(k)}$، $\theta_{s(k)}$ و $\theta_{d(k)}$ به ترتیب فاز سیگنال های نویزی، تمیز و نویز هستند.

تخمین گر بیشترین شباهت اولین بار توسط Mcaulay & Malpass در بهسازی گفتار به کار گرفته شد [۴۵]. آنها فرض کردند مشاهدات، سیگنال نویزی باشد. همچنین، دامنه و فاز سیگنال تمیز، پارامترهای نامعلوم و معین بوده و ضرایب فوریه نویز، گوسی با میانگین صفر و واریانس (در هر دو بخش حقیقی و موهومی) $\sigma_d(k)/۲$ باشند. بر اساس این فرضیات، تابع توزیع ضرایب فوریه سیگنال نویزی $Y(\omega_k)$ ، نیز گوسی، با میانگین $s_k e^{j\theta_s(k)}$ و واریانس $\sigma_d(k)$ خواهد بود. بنابراین رابطه (۴-۴) را برای تابع احتمال مورد نظر می توان نوشت.

$$y(n) = s(n) + d(n) \qquad (۴-۲)$$

$$Y(\omega_k) = S(\omega_k) + D(\omega_k) \Rightarrow Y_{k'} e^{j\theta_{y(k)}} = S_{k'} e^{j\theta_{s(k)}} + D_{k'} e^{j\theta_{d(k)}} \qquad (۴-۳)$$

$$p\left(Y(\omega_k); s_k, \theta_s(k)\right) = \frac{۱}{\pi\sigma_d(k)} \exp\left[-\frac{|Y(\omega_k) - s_k e^{\theta_s(k)}|^۲}{\sigma_d(k)}\right]$$

$$= \frac{۱}{\pi\sigma_d(k)} exp\left[-\frac{Y_k^۲ - ۲S_k Re\{Y(\omega_k) e^{-j\theta_s(k)}\} + X_k^۲}{\sigma_d(k)}\right] \qquad (۴-۴)$$

بیشینه کردن این تابع که از دو متغیر نامعلوم دامنه و فاز تشکیل شده است، کار پیچیده ای است. از این رو ابتدا تابع احتمال فوق را با فرض توزیع یکنواخت[۴] برای فاز و استفاده از رابطه انتگرالی (۴-۵)

---

۱ -Random
۲ -Likellihood
۳ -Likellihood Function
۴ - Uniform Distribution

تبدیل به تابعی با یک متغیر نامعلوم به صورت رابطه (۴-۶) می نمایند.

$$p_m(Y(\omega_k);s_k) = \int_0^{2\pi} p\left(Y(\omega_k);s_k,\theta_s(k)\right)p\left(\theta_s(k)\right)d\theta_s(k) \qquad (۴-۵)$$

$$p_m(Y(\omega_k);s_k) = \frac{1}{\pi\sigma_d(k)}exp\left[-\frac{Y_k^2+x_k^2}{\sigma_d(k)}\right]\int_0^{2\pi} exp\left[\frac{2s_k Re\left[Y(\omega_k)e^{-j\theta_s(k)}\right]}{\sigma_d(k)}\right]d\theta_s(k) \qquad (۴-۶)$$

در این رابطه $Re(.)$ بیانگر بخش حقیقی طیف است و انتگرال فوق تابع بسل تغییریافته[۱] نوع اول است که می توان آن را با تابع نمایی تقریب زد. حال از این تابع که فقط بر حسب متغیر s_k است، مشتق گرفته و برابر صفرقرار داده می شود تا تخمین بیشترین شباهت دامنه به صورت رابطه (۴-۷) به دست آید. سپس با قرار دادن مقدار فاز سیگنال نویزی به جای فاز سیگنال تمیز؛ تخمین طیف سیگنال تمیز، مطابق رابطه (۴-۸) به دست می آید.

$$\hat{s}_k = \frac{1}{2}\left[Y_k + \sqrt{Y_k^2 - \sigma_d(k)}\right] \qquad (۴-۷)$$

$$\hat{S}(\omega_k) = \hat{s}_k e^{j\theta_y(k)} = \hat{s}_k\frac{Y(\omega_k)}{Y_k} = \left[\frac{1}{2}+\frac{1}{2}\sqrt{\frac{Y_k^2-\lambda_d(k)}{Y_k^2}}\right]Y(\omega_k)$$

$$= \left[\frac{1}{2}+\frac{1}{2}\sqrt{\frac{\gamma_k-1}{\gamma_k}}\right]Y(\omega_k) = G_{ML}(\Upsilon_k)Y(\omega_k) \qquad (۴-۸)$$

در این روابط $\Upsilon_k = Y_k^2/\sigma_d(k)$ سیگنال به نویز پسین[۲] و $G_{ML}(\Upsilon_k)$ را اصطلاحا تابع بهره[۳] تخمین ML می نامند که تابعی از سیگنال به نویز پسین است. سیگنال به نویز پسین γ_k معادل SNR مشاهده شده جز k ام طیف است، زمانی که سیگنال نویز به سیگنال افزوده شده باشد.

۴-۴- تخمین گر بیز

در تخمین گر بیشترین شباهت فرض شد که پارامتر نامعلوم مورد نظر δ، یک متغیر معین است. در حالیکه در تخمین گر بیز، فرض بر این است که پارامتر نامعلوم یک متغیر تصادفی است. از آنجا که در فرآیند تخمین از قانون بیز استفاده می شود، آن را تخمین گر بیز می نامند. در این روش،

---

۱ - Modified Bessel Function
۲ - A Posteriori SNR
۳ - Gain Function

همچنین فرض می شود که یک دانش اولیه (پیشین[1]) مانند توزیع متغیر نامعلوم، در دسترس است. با توجه به توانایی این تخمین گر در استفاده از این دانش ، در کاربردهایی که تخمین مناسبی از توزیع وجود داشته باشد، این روش کارایی بهتری نسبت به روش بیشترین شباهت دارد. این تخمین گر شامل روشهای کمترین مربعات خطا (MMSE) و بیشترین احتمال پسین (MAP) است [۳]. علاوه بر روش کمترین مربعات خطا چند روش برگرفته از آن مانند Log-MMSE یا MMSE با توان های بیش از یک نیز وجود دارند.

۴-۵- تخمین گر MMSE

فرض کنید که $\hat{X}$ تخمینی از متغیر X باشد که این تخمین بر اساس متغیردیگری مانندY که مقدار آن مشخص است (مشاهدات)، به دست آمده باشد. میانگین مربعات خطا (MSE)، همانگونه که از اسمش پیداست، برای متغیر X و تخمین آن $\hat{X}$، به صورت رابطه (۴-۹) تعریف می شود. در این رابطه $p(Y,X)$ توزیع توام[2] X و Y و $p(X|Y)$ تابع چگالی احتمال X بعد از مشاهده Y است.

$$e = E\{|\hat{X} - X|^2\} = \iint (\hat{X} - X)^2 \, p(Y,X) dY dx$$
$$= \int \left(\int (\hat{X} - X)^2 p(X|Y)\, dx \right) p(Y)\, dY \qquad (۴-۹)$$

هدف تخمین گر کم ترین میانگین مربعات خطا (MMSE) کمینه کردن رابطه (۴-۹) است. برای انجام این کار کافیست که عبارت انتگرال داخلی کمینه شود. با مشتق گیری از این رابطه نسبت به $\hat{X}$ و برابر صفر قرار دادن آن، رابطه (۴-۱۰) به عنوان تخمین MMSE نتیجه می شود [۴۸].

$$\min_{\hat{X}} e = E\{|\hat{X} - X|^2\} \Rightarrow \hat{X} = E(X|Y) \qquad (۴-۱۰)$$

در بهسازی گفتار متغیر شناخته شده (مشاهدات)، طیف گفتار نویزی $Y = [Y(\omega_0), \ldots, Y(\omega_{N-1})]$ و پارامتر نامعلوم همان طیف گفتار تمیز $s(\omega_k)$ است. بنابراین تخمین بهینه اندازه ی طیف و تخمین بهینه طیف سیگنال تمیز با MMSE به ترتیب به صورت روابط (۴-۱۱) و (۴-۱۲) خواهد بود.

---

[1] - A Priori
[2] - Joint Distribution

$$\hat{s}_k = E[s_k|Y] = E[s_k|Y(\omega_0), Y(\omega_1), \ldots, Y(\omega_{N-1})] \quad (۱۱-۴)$$

$$\hat{s}(\omega_k) = E[s_k|Y] = E[s(\omega_k)|Y(\omega_0), Y(\omega_1), \ldots, Y(\omega_{N-1})] \quad (۱۲-۴)$$

۴-۵-۱- فیلتر وینر

فیلتر وینر یکی از روش های شناخته شده در زمینه فیلتر کردن آماری سیگنال است و در زمینه های مختلفی مانند پردازش سیگنال دیجیتال، پیش بینی خطی و . . . کاربرد دارد. این روش که ابتدا توسط نوربرت وینر (۱۹۴۹) فرموله شده است، به حل مساله زیر می پردازد.

فرض کنید که یک سیگنال ورودی مانند s(n) از یک سیستم خطی نامتغیر با زمان عبور کرده و سیگنال خروجی y(n) را تولید کند. هدف این است که فیلتر خطی نامتغیر با زمان h به گونه ای طراحی شود که معکوس سیستم فوق را شبیه سازی کند و با اعمال آن روی y(n)، سیگنال $\hat{s}(n)$ را که تخمینی از s(n) است، نتیجه دهد. یک راه حل ممکن برای این مساله این است که خطای بین s(n) و $\hat{s}(n)$ یعنی $e(n) = s(n) - \hat{s}(n)$، را در نظر گرفته شده و سعی کنیم آن را کمینه نماییم. این روش ابتدا توسط وینر برای حل این مساله در حالت پیوسته ، مورد استفاده قرار گرفته است. فیلتر می تواند در حوزه زمان (ضرایب فیلتر) و یا حوزه فرکانس به دست آید. در هر دو حوزه میانگین مربعات خطا محاسبه شده و با مشتق گیری از آن و برابر صفر قرار دادن، کمینه می شود. یکی از کاربرد های فیلتر وینر، در بهسازی گفتار است. در این کاربرد سیگنال تمیز s(n) با عبور از یک سیستم توسط نویز جمع شونده d(n) به سیگنال نویزی y(n) تبدیل می شود. فیلتر وینر h در حوزه زمان، به صورت h.y= $\hat{s}(n)$ روی سیگنال نویزی اعمال شده و تخمین سیگنال تمیز را نتیجه می‌دهد. در اینجا $h = [h_0, h_1, \ldots, h_{M-1}]$ بردار ضرایب فیلتر و $y = [y_0, y_1, \ldots, y_{M-1}]$ بردار سیگنال نویزی برای M نمونه پشت سر هم است. برای استخراج ضرایب h_k این فیلتر، در حوزه زمان از $j = E\{e^2(n)\}$ نسبت به h_k مشتق گرفته و برابر صفر قرار داده می شود، یعنی:

$$\frac{\partial j}{\partial h_k} = \frac{\partial E\{e^2(n)\}}{\partial h_k} = \frac{\partial E\{[s(n) - h.y]^2\}}{\partial h_k}$$

$$= \frac{\partial\left(E\{s^{2}(n)\} - s^{2}(n) - 2h.E\{y.s(n)\} + h.E\{y.y^{T}\}h^{T}\right)}{\partial h_{k}}$$

$$= -2E\{e(n)y(n-k)\} = 0, k = 0,1,2,\ldots,M-1 \qquad (۱۳-۴)$$

در این رابطه $E\{y.s(n)\} = E\{[y_0, y_1, \ldots, y_{M-1}].s(n)\}$ که یک بردار $M\times 1$ است، همبستگی بین سیگنال تمیز و سیگنال نویزی را نشان می دهد و با r_{ys} نشان داده می شود. همچنین $E\{y.y^{T}\}$ یک ماتریس $M\times M$ است که ماتریس خودهمبستگی سیگنال نویزی است و با R_{yy} نشان داده می شود. رابطه فوق شرط لازم و کافی برای کمینه شدن تابع هزینه $j = E\{e^{2}(n)\}$ است. به صورت مشابه، مشتق گیری از شکل برداری- ماتریسی رابطه فوق منجر به دستگاه معادلات زیر (۱۴-۴) می گردد که شامل M معادله و M مجهول (همان ضرایب فیلتر) است. حل این معادلات که به معادلات Wiener&Hopf معروف اند، ضرایب بهینه فیلتر را نتیجه می دهد.

$$\frac{\partial j}{\partial h} = -2r_{ys} + 2h.R_{rr} = 0 \Rightarrow r_{ys} = 2h.R_{rr} \Rightarrow h = R_{yy}^{-1}r_{ys}$$

$$\sum_{k=0}^{M} h_k r_{yy}(m-k) = r_{yd}(-m), m = 0,1,\ldots,M-1 \qquad (۱۴-۴)$$

فیلتر وینر را می توان در حوزه فرکانس نیز استخراج کرد. در بهسازی گفتار استفاده از این حوزه رایج تر بوده و عمده مطالعات تکمیلی نیز در این حوزه بوده است. می دانیم که عبور سیگنال y(n) از فیلتر خطی نامتغیر با زمان h(n) منجر به خروجی $\hat{s}(n) = h(n)^{*} - y(n)$ می شود و در حوزه فرکانس داریم: $\hat{s}(\omega_k) = H(\omega_k)Y(\omega_k)$ و $E(\omega_k) = s(\omega_k) - s(\omega_k)$. در این حالت کمینه کردن میانگین مربعات خطا منجر به رابطه (۱۵-۴) می شود.

$$\frac{\partial\left(E\{|s(\omega_k) - H(\omega_k)Y(\omega_k)|^{2}\}\right)}{\partial H(\omega_k)} = H(\omega_k)p_{yy}(\omega_k) - p_{ys}(\omega_k) = 0$$

$$\Rightarrow H(\omega_k) = \frac{p_{ys}(\omega_k)}{p_{yy}(\omega_k)} \qquad (۱۵-۴)$$

که در آن:

$$p_{yy}(\omega_k) = E\{|Y(\omega_k)|^{2}\}E\{[S(\omega_k) + D(\omega_k)][S(\omega_k) + D(\omega_k)]^{*}\}$$

$$= E\{S(\omega_k) + S^*(\omega_k)\} + E\{S(\omega_k) + D^*(\omega_k)\}$$
$$+E\{S^*(\omega_k) + D(\omega_k)\} + E\{D(\omega_k) + D^*(\omega_k)\}$$
$$= p_{ss}(\omega_k) + p_{dd}(\omega_k) \qquad (۱۶-۴)$$

و

$$p_{ys}(\omega_k) = E\{|Y(\omega_k)S^*(\omega_k)|\} = E\{[S(\omega_k) + D(\omega_k)]S^*(\omega_k)\}$$
$$= E\{S(\omega_k)S^*(\omega_k)\} + E\{S^*(\omega_k)D(\omega_k)\} = p_{ss}(\omega_k) \qquad (۱۷-۴)$$

همچنین با فرض مستقل بودن نویز و سیگنال تمیز داریم:

$$E\{S(\omega_k)D^*(\omega_k)\} = E\{S^*(\omega_k)D(\omega_k)\} = ۰ \qquad (۱۸-۴)$$

در نتیجه فیلتر وینر نهایی به صورت زیرخواهد بود که درآن ξ_k را سیگنال به نویز پیشین[۱] برای فرکانس K ام می گویند که به صورت رابطه (۴-۲۰) تعریف می شود. سیگنال به نویز پیشین ξ_k، مقدار SNR واقعی جزk ام طیف است که بعدا روش های محاسبه آن توضیح داده خواهد شد.

$$H(\omega_k) = \frac{p_{ys}(\omega_k)}{p_{yy}(\omega_k)} = \frac{p_{ss}(\omega_k)}{p_{ss}(\omega_k)+p_{dd}(\omega_k)} = \frac{\xi_k}{\xi_k+۱} \qquad (۱۹-۴)$$

$$\xi_k = p_{ss}(\omega_k)/p_{nn}(\omega_k) = E\{|S(\omega_k)|^۲\}/E\{|D(\omega_k)|^۲\} \qquad (۲۰-۴)$$

۴-۵-۲- تخمین اندازه طیف (STSA-MMSE[۲])

در بخش قبل، روش وینر به عنوان یک تخمین گر خطی بهینه برای تخمین طیف مختلط سیگنال مرور شد. در این بخش یک تخمین بهینه برای دامنه طیف براساس معیار MMSE به دست می آید که به آن تخمین کمترین میانگین مربعات خطای دامنه طیف زمان کوتاه (STSA-MMSE) نیز می گویند. در این روش فرض می شودکه ضرایب فوریه ی مختلط گفتار دارای توزیع گوسی با میانگین صفر و واریانس متغیرهستند. به علاوه فرض شده این ضرایب مستقل و در نتیجه ناهمبسته هستند. فرض گوسی بودن ضرایب براساس قضیه حد مرکزی[۳] است. این قضیه بیان می کند که توزیع

---

[۱] -A Priori SNR
[۲] -Minimum Mean Square Error Estimation of Short-Time Spectral Amplitude
[۳] - Central Limit Theorem

مجموع N متغیر تصادفی، وقتی که N بزرگ باشد به سمت توزیع گوسی میل می کند و هر ضریب فوریه در یک قاب با طول N، برابر مجموع وزندار N متغیر تصادفی آن قاب است. پس توزیع آن گوسی فرض می شود. اما این فرض هنگامی صحیح است که طول قاب زیاد باشد.

همانگونه که قبلا بیان شد رابطه تخمین گر MMSE برای تخمین دامنه به صورت رابطه (۴-۱۱) می باشد. با توجه به مستقل بودن ضرایب فوریه سیگنال تمیز و نویز این رابطه به صورت رابطه (۴-۲۱) در می آید.

$$\hat{S}_k = E\left(S_k|Y(\omega_0)Y(\omega_1)\dots Y(\omega_{N-1})\right) = E[S_k|Y(\omega_k)] \qquad (۴-۲۱)$$

گام اول در محاسبه امید ریاضی بیان شده در این رابطه، تعیین تابع توزیع پسین S_k یا همان $p(S_k|Y(\omega_k))$ است. برای این کار قانون بیز به کار گرفته می شود که منجر به رابطه (۴-۲۲) می شود.

$$p(S_k|Y(\omega_k)) = \frac{p(Y(\omega_k)|S_k)p(S_k)}{\int_0^\infty p(Y(\omega_k)|S_k)p(S_k)dS_k} \qquad (۴-۲۲)$$

با استفاده از رابطه (۴-۲۲)، امید ریاضی بیان شده به صورت رابطه (۴-۲۳) محاسبه می شود.

$$\hat{S}_k = E(S_k|Y(\omega_k)) = \int_0^\infty S_k p(S_k|Y(\omega_k))dS_k$$

$$= \frac{\int_0^\infty S_k p(Y(\omega_k)|S_k)p(S_k)dS_k}{\int_0^\infty p(Y(\omega_k)|S_k)p(S_k)dS_k} = \frac{\int_0^\infty S_k p(Y(\omega_k)|S_k)p(S_k)dS_k}{\int_0^\infty p(Y(\omega_k)|S_k)p(S_k)dS_k} \qquad (۴-۲۳)$$

از طرفی می توان نوشت:

$$p(Y(\omega_k)|S_k)p(S_k) = \int_0^{2\pi} p(Y(\omega_k)|S_k,\theta_s(k))p(S_k,\theta_s(k))d\theta_s(k) \qquad (۴-۲۴)$$

که در آن $\theta_s(k)$ متغیر تصادفی، معادل فاز $S(\omega_k)$ است. بنابراین داریم:

$$\hat{S}_k = \frac{\int_0^\infty \int_0^{2\pi} S_k p(Y(\omega_k)|S_k,\theta_s(k))p(S_k,\theta_s(k))d\theta_s(k)dS_k}{\int_0^\infty \int_0^{2\pi} p(Y(\omega_k)|S_k,\theta_s(k))p(S_k,\theta_s(k))d\theta_s(k)dS_k} \qquad (۴-۲۵)$$

حال برای حل این رابطه باید $p(Y(\omega_k)|S_k,\theta_s(k))$ و $p(S_k,\theta_s(k))$ محاسبه شوند. به همین منظور از مدل آماری گوسی برای سیگنال و نویز استفاده می شود. با توجه به این مدل می توان گفت که $Y(\omega_k)$ نیز دارای توزیع گوسی است، زیرا از جمع دو متغیر گوسی با میانگین صفر تشکیل

شده است. به همین ترتیب می توان نتیجه گرفت که $p(Y(\omega_k)|S_k,\theta_s(k))$ نیز گوسی است. بنابراین داریم:

$$(Y(\omega_k)|S_k,\theta_s(k)) = \frac{1}{\pi\sigma_d(k)} exp\left[-\frac{|Y(\omega_k)-S(\omega_k)|^2}{\sigma_d(k)}\right] \quad (۲۶-۴)$$

که در آن $\sigma_d(k)$ بیانگر واریانس نویز در مولفه kام طیف است. از طرفی اندازه دامنه و فاز یک متغیر تصادفی گوسی مستقل از هم بوده [۴۹] و می توان نوشت:

$$p(S_k,\theta_s(k)) = p(S_k)p(\theta_k(k)) \quad (۲۷-۴)$$

به علاوه، اندازه یک متغیر تصادفی گوسی دارای توزیع رایلی[۱] و توزیع فاز آن در بازه $(-\pi\ ,\ \pi)$ یکنواخت است [۴۹]. بنابراین:

$$p(S_k,\theta_s(k)) = \frac{S_k}{\pi\sigma_d(k)} exp\left[-\frac{S_k^2}{\sigma_s(k)}\right] \quad (۲۸-۴)$$

که در آن $\sigma_s(k)$ بیانگر واریانس سیگنال تمیز در مولفه kام طیف است. با جایگزینی روابط (۴-۲۸) و (۴-۲۶) در رابطه (۴-۲۵)، در نهایت تخمین اندازه طیف به صورت زیر به دست می آید [۲۲]:

$$\hat{S}_k = \frac{\sqrt{v_k}}{\gamma_k}\Gamma(1.5)\Phi(-0.5,1;-v_k)Y_k$$
$$= \frac{\sqrt{\pi}}{2}\frac{\sqrt{v_k}}{\gamma_k}\exp\left(-\frac{v_k}{2}\right)\left[(1+v_k)I_0\frac{v_k}{2}+v_k\left(\frac{v_k}{2}\right)\right]Y_k = G(\xi_k,\gamma_k)\,Y_k \quad (۲۹-۴)$$

که در آن:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt \quad (۳۰-۴)$$

$$\Phi(a,b;x) = 1+\frac{a}{b}\frac{x}{1!}+\frac{a(a+1)}{b(b+1)}\frac{x}{2!}+\frac{a(a+1)(a+2)}{b(b+1)(b+2)}\frac{x}{3!}+\cdots \quad (۳۱-۴)$$

$$\sigma_k = \frac{\sigma_s(k)\sigma_d(k)}{\sigma_s(k)+\sigma_d(k)} = \frac{\sigma_s(k)}{1+\xi_k} \quad (۳۲-۴)$$

$$v_k = \frac{\xi_k}{\xi_k+1}\gamma_k \quad (۳۳-۴)$$

$$\xi_k = \frac{\sigma_s(k)}{\sigma_d(k)} \quad (۳۴-۴)$$

$$\gamma_k = \frac{Y_k^2}{\theta_d(k)} \quad (۳۵-۴)$$

---

[۱] - Rayleigh

و $G(\xi_k,\gamma_k)$ تابع بهره تخمین است که به دو معیار سیگنال به نویز پسین و پیشین وابسته است. تحلیل تابع بهره که وابسته به دو پارامتر سیگنال به نویز پیشین و پسین است، در درک بیشتر این تخمین گر و ارزیابی کارایی آن نقش موثری دارد. وابستگی تابع بهره MMSE به این دو پارامتر یکی از عوامل موثر در کاهش نویز موزیکال این روش درمقایسه با سایر روش ها و از جمله فیلتر وینر است. نکات مهم این تخمین عبارت اند از: نقش سیگنال به نویز پیشین در حذف نویز پررنگتر است. همچنین سیگنال به نویزهای بالا MMSE دامنه، حساسیت زیادی به تغییرسیگنال به نویزپیشین به میزان اندک ندارد. MMSE به فرو-تخمین سیگنال به نویز پیشین بیشتر از فرا تخمین آن حساسیت دارد [۲۲]. مقایسه عینی انجام شده در [۳۰] نیز کارایی بهتر تخمین دامنه با MMSE را در مقایسه با سایر روش ها (تفریق طیفی، زیر فضای سیگنال و فیلتر وینر) نشان می دهد.

۴-۵-۳- تخمین فاز با کمترین میانگین مربعات خطا

همانگونه که می دانیم برای تخمین سیگنال تمیز از روی سیگنال نویزی، علاوه بر تخمین دامنه به تخمین فاز نیز نیاز است. با داشتن فاز سیگنال تمیز، نهایتا طیف کامل سیگنال تخمینی به صورت $\hat{S}(\omega_k) = \hat{S}_k B_k \exp\left(j\theta_y(k)\right)$ به دست می آید و از روی آن سیگنال در حوزه طیف تخمین زده شده و سپس در حوزه زمان محاسبه می شود. برای تخمین فاز، Ephraim & Malah درمقاله معروف خود [۲۲] دو راه حل پیشنهاد دادند که در ادامه مرور خواهند شد.

در روش اول مشابه مراحلی که برای دامنه انجام شد، تخمین MMSE برای بخش موهومی $\hat{S}(\omega_k)$ که به صورت $\exp\left(j\theta_s(k)\right) = E\left\{\exp\left(j\theta_s(k)\right)|Y(\omega_k)\right\}$ است، انجام می شود. با توجه به مدل فرض شده، تخمین بخش نمایی طیف به صورت رابطه (۴-۳۶) نتیجه می شود.

$$\exp\left(j\theta_s(k)\right) = \frac{\sqrt{\pi}}{2}\sqrt{v_k}\exp\left(-\frac{v_k}{2}\right)\left[I_0\frac{v_k}{2} + I_1\left(\frac{v_k}{2}\right)\right]\exp\left(j\theta_y(k)\right)$$
$$\triangleq B_k\exp\left(j\theta_y(k)\right) \qquad (۴-۳۶)$$

حال با ترکیب این تخمین برای فاز و تخمین دامنه که در بخش قبل به دست آمد، رابطه تخمین

طیف سیگنال تمیز به صورت $\hat{S}(\omega_k) = \hat{S}_k B_k \exp\left(j\theta_y(k)\right)$ به دست می آید.

با توجه به اینکه تخمین دامنه و فاز به صورت مستقل از هم به دست آمده اند، تخمین رابطه فوق با دامنه جدید $\hat{S}_k B_k$ الزاما بهینه نخواهد بود. چرا که $|B_k|$ برابر یک نیست و درنتیجه این مساله بهینه بودن $\hat{S}_k$ را تحت تاثیر قرار می دهد. از این رو راه حل دومی نیز که در همان مقاله ارائه شده، یک تخمین گرMMSE است که درآن یک شرط به مساله اضافه شده است. دراین حالت جواب با این محدودیت که قدرمطلق قسمت موهومی برابر یک باشد، به دست آورده شده است، یعنی صورت مساله به شکل رابطه زیر درآمده است:

$$\min_{\exp(j\hat{\theta}_s(k))} E\left\{\left|\exp\left(j\hat{\theta}_s(k)\right) - \exp(j\theta_s(k))\right|^2\right\} \text{ و}$$

$$\left|\exp\left(j\hat{\theta}_s(k)\right)\right| = 1 \qquad (۳۷-۴)$$

حل این مساله با کمک گرفتن از ضرایب لاگرانژ منجر به پاسخ زیر می شود.

$$\exp\left(j\hat{\theta}_s(k)\right) = \exp(j\theta_s(k)) \qquad (۳۸-۴)$$

این رابطه بیان می کند که فاز بهینه برای سیگنال از نظر MMSE همان فاز سیگنال نویزی است که استفاده از آن در بیشتر سیستم های بهسازی گفتار به جای فاز سیگنال تمیز، امری رایج است.

۴-۵-۴- محاسبه سیگنال به نویز پیشین

همانگونه که در روابط قبلی آورده شد، تخمین بهینه MMSE برای طیف سیگنال منجر به استفاده ازمقادیر واریانس نویز و واریانس سیگنال تمیز جهت محاسبه سیگنال به نویز پیشین و پسین است. در عمل که فقط به سیگنال نویزی دسترسی داریم، می توان با فرض ایستان بودن نویز، واریانس آن را از روی بخش های غیرگفتار سیگنال با استفاده از VAD یا تخمین گرهای نویز محاسبه کرد. اما تخمین واریانس سیگنال تمیز جهت محاسبه سیگنال به نویز پیشین، با توجه به در دسترس نبودن سیگنال تمیز، کار آسانی نیست. برای تخمین این نسبت روش های زیادی ارائه شده است که عمده آنها مبتنی بر دو روش ارائه شده در [۲۲] هستند. این دو روش یکی مبتنی بر روش بیشترین

شباهت و دیگری براساس رابطه این پارامتر با سیگنال به نویز پسین با نام تصمیم گرا [۱] ارائه شده است که اغلب کارهای بعدی توسعه و بهبود یکی از این دو روش است.

۴-۵-۴-۱- روش بیشترین شباهت

دراین رویکرد فرض می شود واریانس سیگنال تمیز یک پارامتر نامعلوم و معین است و سعی می شود با استفاده از روش بیشترین شباهت تخمین زده شود. برای این کار فرض می شودکه تخمین $\sigma_s(k)$ در قاب mام به کمک L نمونه قبلی مشاهده شده از اندازه دامنه سیگنال نویزی یعنی $Y_k(m) \triangleq \{Y_k(m), Y_k(m-1), Y_k(m), \dots, Y_k(m-L+1)\}$ انجام شود.

بافرض استقلال میان L نمونه و گوسی بودن توزیع، تابع شباهت زیر به دست می آید.

$$p\left(Y_k(m); \sigma_s(k), \sigma_s(k)\right) = \prod_{i=0}^{L-1} \frac{1}{\pi[\sigma_s(k)+\sigma_s(k)]} \exp\left(-\frac{Y_k^2(m-i)}{\sigma_s(k)+\sigma_s(k)}\right) \qquad (۳۹-۴)$$

که مشتق گیری از آن نسبت به $\sigma_s(k)$ و یافتن نقطه بیشینه تابع فوق منجر به تخمین زیر می شود.

$$\hat{\sigma}_s(k,m) = \begin{cases} \frac{1}{L}\sum_{i=0}^{L-1} Y_k^2(m-i) - \lambda_d(k,m) & \text{if non-negativ} \\ 0 & \text{else} \end{cases}$$

$$= \max\left(\frac{1}{L}\sum_{i=0}^{L-1} Y_k^2(m-i) - \lambda_d(k,m), 0\right) \qquad (۴۰-۴)$$

باتوجه به رابطه (۴۰-۴) سیگنال به نویز پیشین به صورت زیر محاسبه می شود:

$$\hat{\xi}_k(m) = \max\left(\frac{1}{L}\sum_{i=0}^{L-1} \gamma_k^2(m-i) - 1, 0\right) \qquad (۴۱-۴)$$

که در این رابطه $\gamma_k(m)$ همان سیگنال به نویز پسین برای قاب mام است و عملگر max باعث می شود که مقدار تخمین همیشه غیرمنفی باشد. در عمل میانگین L نمونه از سیگنال به نویز پسین در قاب m را به صورت رابطه (۴۰-۴۲) محاسبه می کنند که درآن پارامتر های $0 \leq \alpha < 1$ و $\beta \geq 1$ به ترتیب، ضرایب ثابت هموارسازی و همبستگی هستند.

۱- Decision Directed

$$\bar{\gamma}_k(m) = \alpha\bar{\gamma}_k(m-1) + (1-\alpha)\frac{\gamma_k(m)}{\beta} \qquad (۴۲-۴)$$

استفاده از این رابطه منجر به تغییر رابطه $\hat{\xi}_k(m)$ به صورت $\hat{\xi}_k(m) = \max(\bar{\gamma}_k(m) - 1{,}0)$ می شود.

۴-۵-۴-۲- روش تصمیم گرا

پرکاربردترین روش تخمین سیگنال به نویز پیشین روش تصمیم گراست که اولین بار در [۲۲] ارائه گردید. دراین روش از رابطه بین سیگنال به نویز پیشین و پسین جهت تخمین این پارامتر استفاده می شود. با توجه به تعریف سیگنال به نویز پیشین در رابطه (۴-۳۴) داریم:

$$\hat{\xi}_k(m) = \frac{\sigma_s(k,m)}{\sigma_d(k,m)} = \frac{E\{Y_k^2(m)\}}{\sigma_d(k,m)} = \frac{E\{Y_k^2(m) - D_k^2(m)\}}{\sigma_d(k,m)}$$

$$= \frac{E\{Y_k^2(m)\}}{\sigma_d(k,m)} - E\{\gamma_k(m)\} - 1 \qquad (۴۳-۴)$$

از رابطه فوق می توان رابطه (۴-۴۴) را نتیجه گرفت و بر اساس آن رابطه نهایی تخمین $\hat{\xi}_k$ را به صورت (۴-۴۵) نوشت که درآن $0 < a < 1$ ضریب وزن جایگزین شده به جای $\frac{1}{2}$ در رابطه (۴-۴۴) است و $\hat{S}_k^2(m-1)$ تخمین اندازه به دست آمده برای قاب قبلی است.

$$\hat{\xi}_k(m) = E\left\{\frac{1}{2}\frac{\hat{S}_k^2(m)}{\sigma_d(k,m)} + \frac{1}{2}[\gamma_k(m) - 1]\right\} \qquad (۴۴-۴)$$

این رابطه را تخمین تصمیم گرا می نامند چون بر اساس تخمین اندازه دامنه در قاب قبلی، کار تخمین را انجام می دهد. این رابطه بیان می کند که تخمین سیگنال به نویز پیشین برابر با جمع وزندار سیگنال به نویز پیشین قاب قبل و سیگنال به نویز پسین قاب جاری است. برای مقداردهی اولیه به رابطه (۴-۴۵) در قاب اول که m=۰ است، توصیه می شود که از رابطه $\hat{\xi}_k(m) = a + (1-a)\,max\{\gamma_k(0) - 1{,}0\}$ استفاده شود که در آن مقدار ۹۸ .۰= a منجر به بهترین نتیجه می شود[۲۲].

$$\hat{\xi}_k(m) = a\frac{\hat{S}_k^2(m-1)}{\sigma_d(km-1)} + (1-a)\max\{\gamma_k(m)-1,0\} \qquad (۴۵-۴)$$

نتایج مقایسه روش MMSE که در آن از روش تصمیم گرا (a =۰. ۹۸) جهت تخمین سیگنال به نویز پیشین استفاده شده، با روش های تفریق طیفی و فیلتر وینر درحضور نویز سفید نشان می دهد که در صورت استفاده از این روش سیگنال بهسازی شده دارای نویز موزیکال نخواهد بود. در مقابل اگر از روش بیشترین شباهت جهت تخمین سیگنال به نویز پیشین استفاده شود، سیگنال خروجی در روش MMSE دارای نویز موزیکال خواهد بود. همچنین درصورت استفاده از روش بیشترین شباهت جهت تخمین سیگنال به نویز پیشین در روش های وینر و MMSE ، این دو روش تقریبا مشابه عمل می کنند. در صورتی که اگر از روش تصمیم گرا در روش وینر استفاده شود، سیگنال خروجی اعوجاج بیشتری نسبت به MMSE دارد.

نتایج نشان می دهد که کم کردن مقدار a منجر به کاهش اعوجاج در سیگنال بهسازی شده اما افزایش نویز زمینه باقیمانده در سیگنال می شود که بیانگر نقش این پارامتر در کنترل میزان اعوجاج و نویز زمینه است [۵۹]. به علاوه مقایسه روش MMSE و روش بیشترین شباهت نشان می دهد که روش بیشترین شباهت نویز موزیکال بیشتری را ایجاد می کند.

۴-۵-۵- تخمین MMSE با توزیع های غیر گوسی

یکی از فرض های اساسی برای بدست آوردن تخمین های MMSE، گوسی بودن طیف سیگنال بود که بیان شد این فرض براساس قضیه حد مرکزی است. از طرفی این فرض هنگامی صحیح است که طول قاب زیاد باشد. ولی در کاربردهای پردازشی طول هر قاب کوتاه (بین ۲۰ تا۴۰ میلی ثانیه) است. بر همین اساس، می توان انتظار داشت استفاده از توزیع های غیرگوسی منجر به نتایج بهتری شوند. این نکته که ابتدا در [۵۰] مطرح شد، اخیرا بسیار مورد توجه محققین قرار گرفته است. در کارهای مختلفی از توزیع های گاما [۷]، رایلی [۳] ، لاپلاس [۴۳] و مخلوط گوسی [۲۶] به جای گوسی در مدل کردن بخش های حقیقی و موهومی ضرایب فوریه سیگنال استفاده شده است.

همچنین در برخی از مقالات، برای نویز هم از توزیع های غیر گوسی استفاده شده است.

در این بخش یک تخمین گر MMSE، با فرض گاما بودن ضرایب سری فوریه سیگنال تمیز و گوسی بودن ضرایب فوریه نویز به دست آورده شده است. ابتدا فرض می شود که $Y_R(k) = Re\{Y(\omega_k)\}$ قسمت حقیقی طیف سیگنال نویزی و $Y_I(k) = Im\{Y(\omega_k)\}$ قسمت حقیقی طیف سیگنال نویزی باشد، یعنی $Y=Y_R + jY_I$ (برای سادگی بیشتر از اندیس k صرف نظر می شود). بر همین اساس طیف مختلط سیگنال تمیز نیز به صورت $S=S_R + jS_I$ خواهد بود. با فرض گاما بودن توزیع بخش های حقیقی و موهومی طیف سیگنال تمیز، خواهیم داشت:

$$p(S_R) = \frac{\sqrt[4]{2.5}}{2\sqrt{\pi\sigma_s}} |S_R|^{-0.5} \exp\left(-\frac{\sqrt{2.5}|S_R|}{\sqrt{\sigma_s}}\right) \quad (۴۶-۴)$$

$$P(S_I) = \frac{\sqrt[4]{2.5}}{2\sqrt{\pi\sigma_s}} |S_I|^{-0.5} exp\left(-\frac{\sqrt{2.5}|S_I|}{\sqrt{\sigma_s}}\right) \quad (۴۷-۴)$$

که در این روابط σ_s /۲ برابر واریانس بخش های حقیقی و موهومی طیف سیگنال تمیز است. برای به دست آوردن تخمین MMSE طیف سیگنال تمیز از روی سیگنال نویزی، فرض می شود که بخش های حقیقی و موهومی سیگنال مستقل هستند. با این فرض، تخمین هر کدام از این دو بخش به صورت مجزا به دست آمده و سپس با هم ترکیب می شوند. یعنی:

$$E[s(\omega_k)|Y(\omega_k)] = E[S|Y] = E[S_R|Y_R] + jE[S_I|Y_I] \quad (۴۸-۴)$$

حال فرض می شود که نویز دارای توزیع گوسی باشد. در این صورت برای بخش حقیقی طیف داریم

$$\hat{S}_R = E[S_R|Y_R] = \frac{\int_{-\infty}^{+\infty} S_R p(Y_R|S_R) p(S_R) dS_R}{p(Y_R)} \quad (۴۹-۴)$$

برای محاسبه این رابطه باید $p(Y_R|S_R)$، $p(Y_R)$ و $p(S_R)$ (که همان توزیع گاماست) محاسبه شوند. با توجه به اینکه بخش حقیقی طیف سیگنال نویزی برابر مجموع بخش های حقیقی طیف سیگنال تمیز و نویز است، برای $p(Y_R|S_R)$ داریم:

$$p(Y_R|S_R) = p_D(Y_R - S_R) = \frac{1}{\sqrt{\pi\sigma_d}} \exp\left(-\frac{(Y_R - S_R)^2}{\sigma_d}\right) \quad (۵۰-۴)$$

این رابطه با توجه به گوسی بودن توزیع طیف نویز به دست آمد. محاسبه $p(Y_R)$ نیز با استفاده از رابطه (۴-۵۱) قابل انجام است.

$$p(Y_R) = \int_{-\infty}^{+\infty} p(S_R|Y_R)\, p(S_R) dS_R \quad (۵۱-۴)$$

بنابراین تخمین بخش حقیقی طیف به صورت رابطه (۴-۵۲) به دست خواهد آمد.

$$\hat{S}_R = \frac{\sqrt[4]{1.5}}{2\pi\sqrt{\sigma_d \sigma_s p(Y_R)}} \int_{-\infty}^{+\infty} S_R |S_R|^{-0.5} exp\left(-\frac{(Y_R - S_R)^2}{\sqrt{\sigma_d}} - \frac{\sqrt[4]{1.5|S_R|}}{\sqrt{\sigma_s}}\right) dS_R$$
$$= \frac{\sqrt{\sigma_d Z_2}}{2\sqrt{2Z_1}} \quad (۵۲-۴)$$

که درآن :

$$C_p(z) = \frac{\exp\left(-\frac{z^2}{2}\right)}{\Gamma(-p)} \int_0^{\infty} x^{-p-1} \exp\left(-xz - \frac{x^2}{2}\right) dx \quad (۵۳-۴)$$

$$G_1 = \frac{\sqrt{1.5\sigma_d}}{2\sqrt{\sigma_s}} + \frac{Y_R}{\sqrt{\sigma_d}} = \frac{\sqrt{1.5}}{2\sqrt{\xi}} + \frac{Y_R}{\sqrt{\sigma_d}} \quad (۵۴-۴)$$

$$G_2 = \frac{\sqrt{1.5\sigma_d}}{2\sqrt{\sigma_s}} + \frac{Y_R}{\sqrt{\sigma_d}} = \frac{\sqrt{1.5}}{2\sqrt{\xi}} + \frac{Y_R}{\sqrt{\sigma_d}} \quad (۵۵-۴)$$

$$Z_1 = \exp\left(\frac{G_2^2}{2}\right) C_{-0.5}(\sqrt{2}G_2) - \exp\left(\frac{G_1^2}{2}\right) C_{-0.5}(\sqrt{2}G_1) \quad (۵۶-۴)$$

$$Z_2 = \exp\left(\frac{G_2^2}{2}\right) C_{-1.5}(\sqrt{2}G_2) - \exp\left(\frac{G_1^2}{2}\right) C_{-1.5}(\sqrt{2}G_1) \quad (۵۷-۴)$$

به صورت مشابهی می توان بخش موهومی را به دست آورد که منجر به تخمینی مشابه تخمین بخش حقیقی می شود. ترکیب این دو تخمین به صورت $\hat{S}_I = \hat{S}_R + j\hat{S}_I$، منجر به تخمین نهایی طیف که یک تخمین مختلط است، می شود. مقایسه این تخمین و فیلتر وینر بیانگر عملکرد مشابه آنها از نظر اعوجاج سیگنال در SNRهای بالا و متفاوت بودن عملکرد آنها در SNRهای پایین است.

۴-۶- تخمین گر بیشترین احتمال پسین MAP

علاوه بر تخمین گرهای ML و MMSE، تخمین گر MAP یکی از روش های تخمین در بهسازی گفتار است [۳۸] و [۶۱] و [۶۲]. درMAP هدف بیشینه کردن احتمال پسین $p(s_k|Y(\omega_k))$ است. تخمین MAP را می توان به عنوان یک راه حل جایگزین برای MMSE در مواردی که تخمین بیشینه احتمال پسین از محاسبه میانگین آن ساده تر است به کاربرد. یکی دیگر از مزایای تخمین MAP این است که می توان تخمین دامنه و فاز را به صورت همزمان انجام داد.

در ادامه تخمین MAP به صورت همزمان برای دامنه و فاز آورده خواهد شد. تخمین همزمان دامنه و فاز به روش MAP منجر به رابطه (۴-۵۸) می شود که در بخش آخر آن، با توجه به مستقل بودن $p(Y(\omega_k))$ از s_k و θ_s، از آن صرفنظر شده است. بافرض گوسی بودن توزیع ها رابطه (۴-۵۹) نتیجه می شود.

$$(\hat{s}_k, \hat{\theta}_s) = \underset{s_k,\theta_s}{\operatorname{argmax}}\, p(s_k,\theta_s|Y(\omega_k)) = \underset{s_k,\theta_s}{\operatorname{argmax}} \frac{p(Y(\omega_k)|s_k,\theta_s)p(s_k,\theta_s)}{p(Y(\omega_k))} = \underset{s_k,\theta_s}{\operatorname{argmax}}\, P(Y(\omega_k)|s_k,\theta_s)p(s_k,\theta_s) \quad (۴-۵۸)$$

$$p(Y(\omega_k)|s_k,\theta_s)p(s_k,\theta_s) = \frac{s_k}{\pi^2 \sigma_s(k)\sigma_d(k)} \exp\left(-\frac{|Y(\omega_k)-s_k e^{j\theta_s}|^2}{\sigma_s(k)} - \frac{s_k^2}{\sigma_s(k)}\right) \quad (۴-۵۹)$$

با مشتق گیری از این رابطه (یا از لگاریتم آن) نسبت به θ_s و برابر صفر قرار دادن آن، تخمین نتیجه می شود که همان تخمین نتیجه شده از MMSE است و همچنین مشتق گیری از $\hat{\theta}_s = \theta_y$ رابطه فوق نسبت به s_k و یافتن نقطه بیشینه آن، رابطه (۴-۶۰) برای تخمین دامنه نتیجه می شود:

$$\hat{S}_R = \frac{\xi_k + \sqrt{\xi_k^2 + 2(1+\xi_k)\frac{\xi_k}{\gamma_k}}}{2(1+\xi_k)} Y_k = G_{MAP} Y_k \quad (۴-۶۰)$$

۴-۷- به کارگیری عدم قطعیت گفتار (SPU)[1] در بهسازی گفتار

در روش های بیان شده در بخش های قبلی، یکی از فرض های اصلی همه روش ها این است که گفتار در همه زمان ها در سیگنال مورد بررسی وجود دارد. اما در عمل، در بسیاری از بخش های سیگنال گفتار، به علت وجود بخش هایی همچون سکوت و توقف گفتار وجود ندارد و یا حداقل به میزان کمی وجود دارد. به علاوه، حتی در بخش های گفتاری سیگنال نیز ممکن است در برخی از باندهای فرکانسی گفتار وجود نداشته باشد. در بهسازی گفتار می توان برای هر باند فرکانسی دو حالت حضور و عدم حضور گفتار را متصور بود و براساس آن کار بهسازی گفتار را انجام داد. در اینجا از مساله فوق به این صورت که گفتار ممکن است در بعضی از زمان ها و در برخی از فرکانس ها وجود نداشته باشد، استفاده شده است. برای فرموله کردن بیان فوق، یک مدل دو حالته برای

[1] - Speech Present Uncertainty

سیگنال گفتار فرض می شود که در آن هر حالت بیانگر یکی از دو فرضیه حضور یا عدم حضور گفتار است. به عبارتی داریم:

$$H_0^k: \text{Speech absent } |Y(\omega_k)| = |D(\omega_k)| \quad (۶۱-۴)$$

$$H_1^k: \text{Speech present } |Y(\omega_k)| = |S(\omega_k) + D(\omega_k)| \quad (۶۲-۴)$$

در این رابطه H_0^k بیانگر فرضیه تهی[۱] است که به معنی عدم حضور گفتاردر فرکانس k ام است و H_1^k بیانگر فرضیه وجود گفتار در سیگنال نویزی در فرکانس k ام است. حال برای استفاده از این مدل دودویی در روش های تخمین بیان شده، بایستی از جمع وزن دار دو فرضیه استفاده کرد [۲۲].

۴-۷-۱- تخمین گر MMSE با SPU

برای روشن تر شدن موضوع به عنوان مثال تخمین گر MMSE با SPU به طور خیلی خلاصه توضیح داده شده است. برای تخمین گر MMSE که $\hat{s}_k = E\{S_k|Y_k\}$ است، داریم:

$$\hat{s}_k = p(H_0^k|Y_k)E\{S_k|Y_k,H_0^k\} + p(H_1^k|Y_k)E\{S_k|Y_k,H_1^k\} \quad (۶۳-۴)$$

در این رابطه $E\{S_k|Y_k,H_1^k\}$ بیانگر متوسط دامنه سیگنال گفتار تمیز به شرط داشتن دامنه طیف و عدم حضور گفتار است که می توان آن را صفر فرض کرد. با این فرض تخمین گر جدید MMSE به صورت رابطه (۴-۶۴) خواهد بود:

$$\hat{s}_k = p(H_1^k|Y_k)E\{S_k|Y_k,H_1^k\} \quad (۶۴-۴)$$

۴-۸- روش های تخمین نویز در بهسازی گفتار

یکی از نیازهای کلیه روش های بهسازی گفتارتک کاناله، تخمین نویز یا مشخصات آن است که کارایی روش بهسازی گفتار را به شدت تحت تاثیر قرار می دهد. در روش های بیان شده تا به حال فرض شد که تخمینی از نویز در دسترس است. در این قسمت روش های تخمین نویز بررسی

---

[۱] - Null Hypothesis

خواهند شد.

ساده ترین روش تخمین نویز VAD است. در این روش با استفاده از انرژی، نرخ عبور از صفر[۱]، متناوب بودن و . . . برای هر قاب، در مورد گفتار یا غیر گفتار بودن آن تصمیم گیری می شود. تصمیم گیری اغلب براساس مقایسه این مقادیر با مقادیرآستانه است. روش های مختلفی برای VAD تا به حال پیشنهاد شده اند که مرور کاملی بر آنها در [۳۳] آورده شده است. برخی از روش ها مبتنی بر انرژی و نرخ عبور از صفر هستند [۳۲] و برخی دیگر بر اساس ضرایب کپسترال[۲۸]، معیار فاصله LPC ایتاکورا [۵۱] و متناوب بودن گفتار [۵۷] عمل می کنند. در این روش مشخصات نویزدر بخش های سکوت و غیرگفتاری سیگنال به روز می شود. این روش درSNRهای بالا و برای نویزهای ایستان مناسب است. به علاوه برخی از آنها نیاز به برخی پارامترها مانند مقادیر آستانه و برخی تنظیمات دیگر دارند که ممکن است از محیطی به محیط دیگر تغییر کند. سیستم های بهسازی گفتار برای داشتن عملکرد مطلوب در کاربردهای واقعی باید از توانایی تخمین مشخصات نویز در شرایط غیر ایستان و دنبال کردن این مشخصات برای همه زمان ها حتی، در شرایطی که گفتار وجود داشته باشد، برخوردار باشند. از این رو روش های تخمین نویز به عنوان یکی از بخش های مهم در سیستم های بهسازی گفتار مطرح شده اند.

در ادامه سه روش از مهم ترین روش های تخمین نویز توضیح داده می شود. در همه این روش ها فرض می شود که توان نویز نسبت به گفتار تغییرات کمتری دارد و ایستان تر است.

۴-۸-۱- روش میانگین گیری بازگشتی[۲]

مشاهده شده است که علاوه بر بخش های سکوت در سیگنال گفتار نویزی، بخش های دیگری نیز وجود دارد که در زمان حضور گفتار نیز انرژی آنها بسیار کم و نزدیک به انرژی نویز است. از این بخش ها می توان به بخش های فرکانس پایین (به ویژه زیر ۲ کیلو هرتز) و بخش های فرکانس

---

۱ - Zero Crossing

۲- time-recursive averaging

بالا(به ویژه بالای ۴ کیلوهرتز) در مصوت ها اشاره کرد. به طور کلی می توان گفت برای هر نوع نویزی، باند های فرکانسی خاصی از سیگنال که احتمال حضور گفتار در آنها کمتر است و یا دارای SNR پایین هستند، برای به روز کردن مشخصات نویز مناسب تر هستند. براساس این مشاهده روش میانگین گیری بازگشتی، طیف نویز را به صورت میانگین وزندار مقادیر قبلی تخمین نویز و مقدار جاری طیف سیگنال نویزی، طبق رابطه (۴-۶۵) به روز می کند. وزن این رابطه متغیر با فرکانس و زمان بوده و مقدار آن برای هر باند فرکانسی به SNR آن باند و یا به احتمال حضور گفتار در آن باند مرتبط است.

$$\hat{\sigma}_d^2(f,\omega_k) = \alpha(f,\omega_k)\hat{\sigma}_d^2(f-1,\omega_k) + (1-\alpha(f,\omega_k))|Y(f,\omega_k)|^2 \qquad (۴-۶۵)$$

در این رابطه $\hat{\sigma}_d^2(f,\omega_k)$ توان نویز تخمین زده شده در فرکانس ω_k وقاب f ، $|Y(f,\omega_k|^2$ مجذور اندازه طیف سیگنال نویزی و $\alpha(f,\omega_k)$ ضریب هموارسازی وابسته به زمان و فرکانس است.

تمامی روشهای میانگین گیری بازگشتی از رابطه فوق پیروی می کنند و تفاوت آنها با یکدیگر در نحوه محاسبه ضریب $\alpha(f,\omega_k)$ است. برخی از محققان مقدار این ضریب را بر اساس تخمین SNR هر باند فرکانسی محاسبه کرده [۳۶] (روش های مبتنی بر SNR) و برخی دیگر مقدار ضریب هموارسازی را ثابت گرفته و در عوض تنها در برخی شرایط مقدار $\hat{\sigma}_d^2(f,\omega_k)$ را به روز کرده اند. (روش های مبتنی بر تصمیم گیری) [۲۹] و [۵۳]. همچنین تعدادی دیگر از محققین مقدار آن را بر حسب احتمال حضور/عدم حضور گفتار در باند های فرکانسی حساب کرده اند [۱۱] و [۳۹] .

۴-۸-۲- روش مبتنی برهیستوگرام[۱]

هیستوگرام باندهای فرکانسی سیگنال نشان می دهند که بیشترین تکرار (مقدار بیشینه هیستوگرام) در یک باند فرکانسی خاص، متناسب با سطح نویز در آن باند است. در مواردی هیستوگرام انرژی طیف دارای دو قله است که یکی از آنها کم انرژی و مربوط به بخش های سکوت و قسمت های کم انرژی سیگنال نویزی است و دومی انرژی بالاست و مربوط به بخش های صدادار سیگنال است . در

[۱] - Histogram-Based

بیشتر موارد قله کم انرژی بالاتر از قله با انرژی بالاست(تکرار بیشتردرهیستوگرام)، هرچند که این مساله به نوع نویز و طول سیگنال مورد بحث نیز وابسته است. این مشاهده اساس روش های مبتنی بر هیستوگرام است که در بیشتر آنها تخمین نویز از روی هیستوگرام مقادیر قبلی طیف انجام می شود. بدین معنی که برای هر قاب در هر فرکانس، هیستوگرام توان طیف از روی چند صد میلی ثانیه قبلی سیگنال (مثلا با استفاده از یک پنجره ۴۰۰ میلی ثانیه ای) محاسبه می شود و بر اساس بیشترین تکرار موجود در هیستوگرام، نویز تخمین زده می شود. همچنین گاهی برای هموارسازی طیف نویز تخمین زده شده از یک رابطه بازگشتی استفاده می شود. با توجه به موارد فوق می توان الگوریتم تخمین نویز به روش هیستوگرام را به صورت زیر خلاصه کرد:

- برای هر قاب f از گفتار نویزی مراحل زیر را انجام دهید:
- توان طیف سیگنال نویزی، $|Y(f,\omega_k|^2$، را محاسبه کنید.
- توان طیف سیگنال نویزی را به صورت زیر هموار کنید.

$$X(f,\omega_k) = \alpha X(f,\omega_k) + (1-\alpha)|Y(f,\omega_k|^2 \qquad (۶۶-۴)$$

- هیستوگرام D نمونه از توان طیف های قاب های قبلی، یعنی $\{X(f-1,\omega_k), X(f-2,\omega_k), \dots, X(f-1,\omega_k)\}$ را محاسبه نمایید.
- در هیستوگرام از R بین (Bin) استفاده شود.
- با فرض اینکه $c = [c_1, c_2, \dots, c_R]$ تعداد تکرار هر بین در هیستوگرام بوده و $s = [s_1, s_2, \dots, s_R]$ مراکز متناسب آنها باشد، رابطه زیر را حساب کنید:

$$c_{max} = \arg\max_{1<i<R} c_i \qquad (۶۷-۴)$$

- حال با استفاده از c_{max}، تخمین نویز $H_{max}(f,\omega_k)$ را به صورت $s_{max} = H_{max}(f,\omega_k)$ به دست آورید که s_{max} مرکز بین متناسب با c_{max} در هیستوگرام است.
- توان طیف تخمینی نویز را با رابطه زیر هموار کنید.

$$\hat{\sigma}_d^2(f,\omega_k) = \beta\hat{\sigma}_d^2(f-1,\omega_k) + (1-\beta)H_{max}(f,\omega_k) \qquad (۶۸-۴)$$

در این الگوریتم $\hat{\sigma}_d^2(f,\omega_k)$ توان طیف تخمین زده شده برای نویز و α و β به ترتیب ضرایب هموار سازی توان طیف سیگنال نویزی و توان طیف نویز هستند. نتایج بررسی این روش نشان می دهد که اگر هموار سازی روی توان طیف سیگنال نویزی اعمال نشود، کارایی این روش افت کرده و دچار فرو تخمین می شود. طول پنجره مورد استفاده جهت محاسبه هیستوگرام اهمیت زیادی دارد. هرچه طول پنجره کمتر باشد، تخمین نویز دچار فرا تخمین می شود، چرا که در این حالت رخ دادن یک بخش دارای انرژی بالا در گفتار ممکن است قله هیستوگرام را از نویز به گفتار عوض کند. این مشکل به ویژه در فرکانس های پایین که انرژی سیگنال بالاست، برجسته تر است. در فرکانس های بالا که انرژی کمتری در آنها متمرکز است، استفاده از پنجره با طول کوچک چندان مطرح نیست. برای حل مشکل فرا تخمین می توان به دو روش عمل کرد. روش اول این است که D بزرگ باشد. اما بزرگ گرفتن D باعث می شود که این روش نتواند با سرعت کافی تغییرات موجود در نویز را دنبال کند. راه حل دوم این است که از گذاشتن قاب با مقدار توان بالا در محاسبه هیستوگرام ممانعت شود. یعنی قاب های حاوی بخش گفتاری سیگنال فیلتر شده و کنار گذاشته شوند. برای تحقق این امر می توان از سیگنال به نویز پسین استفاده کرد و تنها از قاب هایی که مقدار این پارامتر برای آنها از مقداری کمتر است، در محاسبه هیستوگرام استفاده شود [۵۳].

۴-۸-۳- روش ردیابی کمینه ها[۱]

توان سیگنال نویزی حتی در بخش های گفتاری، در برخی از باندهای فرکانسی به سطح توان نویز کاهش می یابد. از این رو می توان با دنبال کردن مقادیر کمینه طیف سیگنال نویزی در تمام باندهای فرکانسی تخمینی از سطح نویز به دست آورد. این ایده مبنای روش ردیابی کمینه ها برای تخمین نویز است که خود به دو دسته تقسیم می شود:

۱- آمارگان کمینه[۲] است که ابتدا توسط مارتین در ۱۹۹۴ارائه گردید [۴۰] و سپس بهبودهایی روی

---

[۱] - Minimal Tracking

[۲]- Minimum Statistic

آن داده شد [۴۱]. در این روش کمینه توان طیف در یک پنجره محدود (به طول قسمت آنالیز) دنبال می شود. در روش آمارگان کمینه بعد از محاسبه پریودوگرام[۱] سیگنال نویزی ، توان این سیگنال با توجه به میزان نوسان بالای پریودوگرام، به صورت رابطه (۴-۶۹) محاسبه می شود. در این رابطه $0 \le \alpha \le 1$ یک ضریب هموار سازی است. با ردیابی کمینه $p(f, \omega_k)$ بر روی یک پنجره با طول محدود، تخمینی از توان نویز به دست می آید. طول پنجره مورد نظر در [۴۰] بین ۰.۸ تا ۱.۴ ثانیه پیشنهاد شده است. شکل (۴-۱) پریودوگرام، پریودوگرام هموار شده و توان تخمینی نویز را برای یک سیگنال نویزی و برای فرکانس k=۲۵ نشان می دهد.

$$p(f, \omega_k) = \alpha p(f - ۱, \omega_k) + (۱ - \alpha)|Y(f, \omega_k)|^۲ \qquad (۴-۶۹)$$

یافتن کمینه مقدار $p(f, \omega_k)$ بر روی D مقدار پشت سر هم این تابع، به صورت رابطه زیر می باشد.

$$p_{min}(f, \omega_k) = min\{p(f, \omega_k), p(f - ۱, \omega_k), \dots, p(f - D + ۱, \omega_k)\} \qquad (۴-۷۰)$$

تخمین نویز با این روش و مقایسه آن با طیف واقعی نویزی نشان می دهد که مقدار نویز به سمت مقادیر پایین تر بایاس شده و در برخی موارد کمتر از مقدار واقعی تخمین زده می شود. برای به دست آوردن مقدار بایاس و حذف اثر فروتخمین نویز می توان از این واقعیت استفاده کرد که در توزیع های احتمال مرتبط به تخمین پریودوگرام، کمینه مجموعه ای از متغیرهای تصادفی از مقدار میانگین این متغیرها کمتر است و در نتیجه تخمین نویز با روش فوق منجر به بایاس شدن آن به سمت مقادیر پایین می شود. بنابراین برای جبران اثر بایاس شدن می توان مقدار میانگین کمینه ها را بدست آورد و با ضرب معکوس آن در تخمین توزیع توان نویز به صورت زیر این اثر را از بین برد.

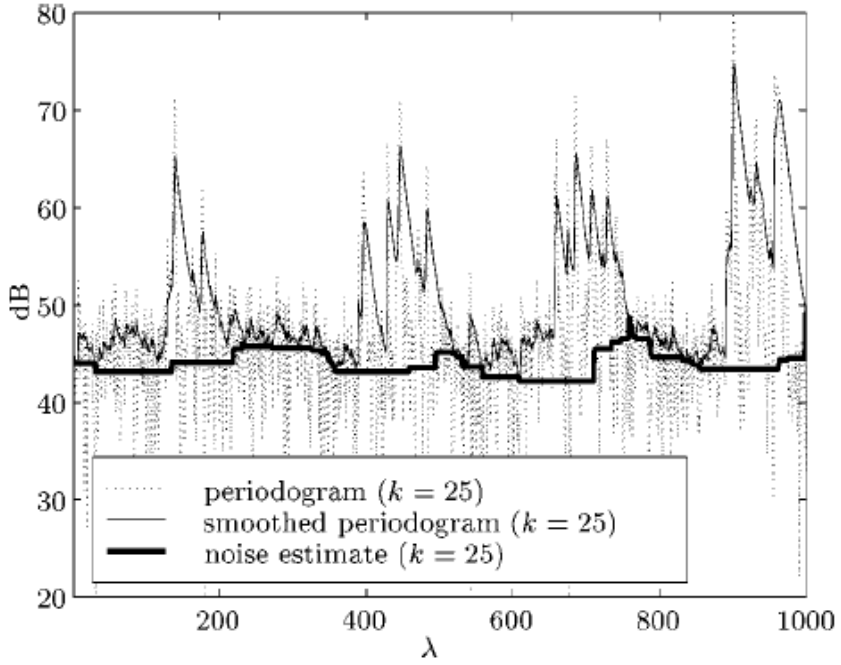
$$\hat{\sigma}_d^۲(f, \omega_k) = B_{min}(f, \omega_k).p_{min}(f, \omega_k) \qquad (۴-۷۱)$$

که $B_{min}(f, \omega_k)$ ضریب جبران سازی بایاس است که برابر با معکوس میانگین $p_{min}(f, \omega_k)$ است.

همچنین بررسی مقادیر مختلف α بیانگر این واقعیت است که مقادیر بالا برای α (مثلا $\alpha = 0.9$) منجر به هموارتر شدن پریودوگرام و حذف فرورفتگی های مختلف موجود آن می شود. این مساله به

---

[۱] - Periodogram



شکل (۴-۱): پریودوگرام $|Y(\lambda,k)|^2$، پریودوگرام هموار شده $P(\lambda,k)$ و تخمین نویز برای فرکانس k=۲۵ [۴۱].

نوبه خود منجر به وسیع تر شدن قله های پریودوگرام و نزدیک شدن تخمین نویز به این قله ها شده و تخمین نویز را دچار تخمین رو به بالا می کند. در واقع برای دنبال کردن نویز بهترآن است که برای بخش های حاوی گفتار، مقدار این پارامتر نزدیک به صفر و در نتیجه از $p(f,\omega_k) \approx Y|(f,\omega_k)|^2$ که همان حالت هموار نشده پریودوگرام است، استفاده شود. پس با به کارگیری ضریب هموارسازی α به صورت متغیر با زمان و فرکانس، تخمین نویز با روش آمارگان کمینه بهبود می یابد. برای تغییر ضریب ثابت هموارسازی α به $\alpha(f,\omega_k)$ ، از کمینه کردن خطای بین طیف تخمینی و طیف واقعی به رابطه (۴-۷۲) می رسیم.

$$\hat{\alpha}(f,\omega_k) = \frac{\alpha_{max}-\alpha_c}{1+\left[p(f-1,\omega_k)/\hat{\sigma}_d^2(f-1,\omega_k)-1\right]^2} \qquad (۴-۷۲)$$

که درآن:

$$\alpha_c(f,\omega_k) = \frac{2}{1+\left[\sum_{i=0}^{M-1} p(f-1,\omega_k)/\sum_{i=0}^{M-1}|Yf-1,\omega_k|^2-1\right]} \qquad (۴-۷۳)$$

و α_{max} که در [۴۱] برابر با ۰.۹۶ قرار داده شده است برای محدود کردن سقف ضریب هموارسازی است. جزئیات بیشتر درباره این روش در [۴۱] آمده است.

۲- در این روش مینیمم به صورت پیوسته و بدون نیاز به پنجره دنبال می شود.[۱] یکی از معایب روش قبل عدم توانایی دنبال کردن تغییرات سریع طیف نویز است. به همین دلیل روشی متفاوت در [۳۷] پیشنهاد شد. در این روش تخمین نویز به صورت پیوسته در هر فرکانس، با یک رابطه هموارسازی غیرخطی، به روز می شود. در این روش نیز برای دنبال کردن مینیمم توان نویزی از پریودوگرام هموار شده استفاده می شود.

$$P(f,\omega_k) = \alpha p(f-1,\omega_k) + (1-\alpha)|Y(f,\omega_k)|^2 \qquad (۴-۷۴)$$

که در آن α فاکتور هموارسازی و در بازه $0.7 \le \alpha \le 0.9$ می باشد.

رابطه غیر خطی که برای تخمین نویز بر پایه دنبال کردن مینیمم طیف توان سیگنال نویزی، در هر فرکانس استفاده می شود، به صورت زیر است:

$$\begin{aligned}
&If \quad p_{min}(f-1,\omega_k) < p(f,\omega_k) \\
&p_{min}(f,\omega_k) = \lambda p_{min}(f-1,\omega_k) + \frac{1-\lambda}{1-\beta}(p(f,\omega_k) - \beta\, p(f-1,\omega_k)) \\
&\text{Else} \\
&\qquad p_{min}(f,\omega_k) = p(f,\omega_k) \\
&\text{End}
\end{aligned} \qquad (۴-۷۵)$$

که در آن $p_{min}(f,\omega_k)$ تخمین نویز است. مزیت این روش نسبت به روش قبل پیچیدگی محاسباتی کمتر است.

۴-۹- روش های آماری مبتنی بر مدل

در این روش ها به کمک داده های آموزشی مدل یا مدل هایی ایجاد می شودکه خصوصیات مورد نظر را (که معمولا تابع توزیع احتمال پارامترهاست) در بردارد. در این روش با استفاده از داده های آموزشی، خصوصیات بلندمدت و مشترک بین قاب های مختلف آموزش داده می شود.

استفاده از رویکرد مدل سازی آماری در پردازش گفتار به ویژه بازشناسی گفتار بسیار رایج و موفق بوده است. به کارگیری این مفهوم در بهسازی گفتار را می توان به کار [۴۵] نسبت داد که در آن فرض شده که هر قاب سیگنال در یکی از دو حالت سکوت و غیر سکوت قرار می گیرد. استفاده از

---

[۱] - Continuos Spectral Minimum Tracking

چند حالت مختلف (کلاس) برای سیگنال در بهسازی گفتار قبلا در [۱۷] ارائه شده بود که در آن گفتار به پنج کلاس fricative، stop، vowel، glids و nasal تقسیم شده و سیگنال نویز به یکی از این کلاس ها نسبت داده می شود. سپس برای هر کدام از حالات از تخمین گر خاص آن کلاس استفاده شده است. امروزه ایده حالات و مدل های فوق به صورت کلی تری تعمیم داده شده است که از روش های پرکاربرد آنها می توان به مدل مخفی مارکوف HMM ، مدل مخلوط گوسی GMM و چندی سازی بردار VQ[۱] اشاره کرد.

مدل سازی در این روش را می توان شامل دو فاز آموزش و آزمون دانست. در فاز آموزش با استفاده از روش های HMM ، GMM یا VQ مدل هایی برای گفتار و نویز بر اساس داده های آموزشی به دست می آید. در این فاز پارامترهای مدل ها به کمک روش های آموزش متناسب از روی داده های آموزشی استخراج می شوند. این پارامترها در واقع همان توزیع های گفتار و نویز هستند. فاز استفاده در این روش ها شامل به کارگیری توزیع های به دست آمده در فاز آموزش برای تخمین سیگنال تمیز از روی سیگنال نویزی است. برای این کار از تکنیک های تخمین که رایج ترین آنها MMSE و MAP است، استفاده می شود. همچنین برای سادگی بیشتر در آموزش و استفاده از مدل ها، مدل سازی روی روش های پارامتری کوتاه مدت مانند مدل AR انجام می گیرد.

می توان HMM را حالت کلی تر روش های GMM و VQ در نظر گرفت. روش GMM را می توان یک HMM یک حالته در نظرگرفت که تنها تفاوت آن با HMM حذف فرض مارکوفی بودن سیگنال است. استفاده از مدل مارکوف چند حالته برای مدل کردن گفتار مناسب تر است چرا که به نظر می رسد گفتار ذاتا دارای چند حالت است و این حالات مختلف ساختار غیر ایستانی گفتار و وابستگی زمانی و فرکانسی بین قاب های مختلف سیگنال را بهتر نمایش می دهند. از روش VQ نیز هر چند به صورت مستقل برای کاربردهای ذکر شده استفاده می شود، اما در بیشتر موارد به عنوان روشی برای مقداردهی اولیه آموزش استفاده شده است.

---

[۱] - Vector Quanzisation

فصل پنجم

روش پیشنهادی بر مبنای توزیع مخلوطی از لاپلاس ها

۵-۱- مقدمه

پس از توضیحاتی که در فصل های قبل درباره اصول روش های بهسازی گفتار و معرفی روش های متداول آماری بیان شد، در این فصل روش پیشنهادی مورد بررسی قرار خواهد گرفت. این روش که در حوزه روش های آماری جای می گیرد، مبتنی بر توزیع مخلوط لاپلاس است. روش پیشنهادی را می توان به سه بخش کلی زیر تقسیم کرد:

۱) به دست آوردن رابطه تخمین گر MMSE با فرض توزیع مخلوط لاپلاس برای سیگنال تمیز.

۲) الگوریتم امید ریاضی بیشینه (EM) برای تخمین پارامترهای توزیع مخلوط لاپلاس.

۳) تخمین پارامترهای نویز

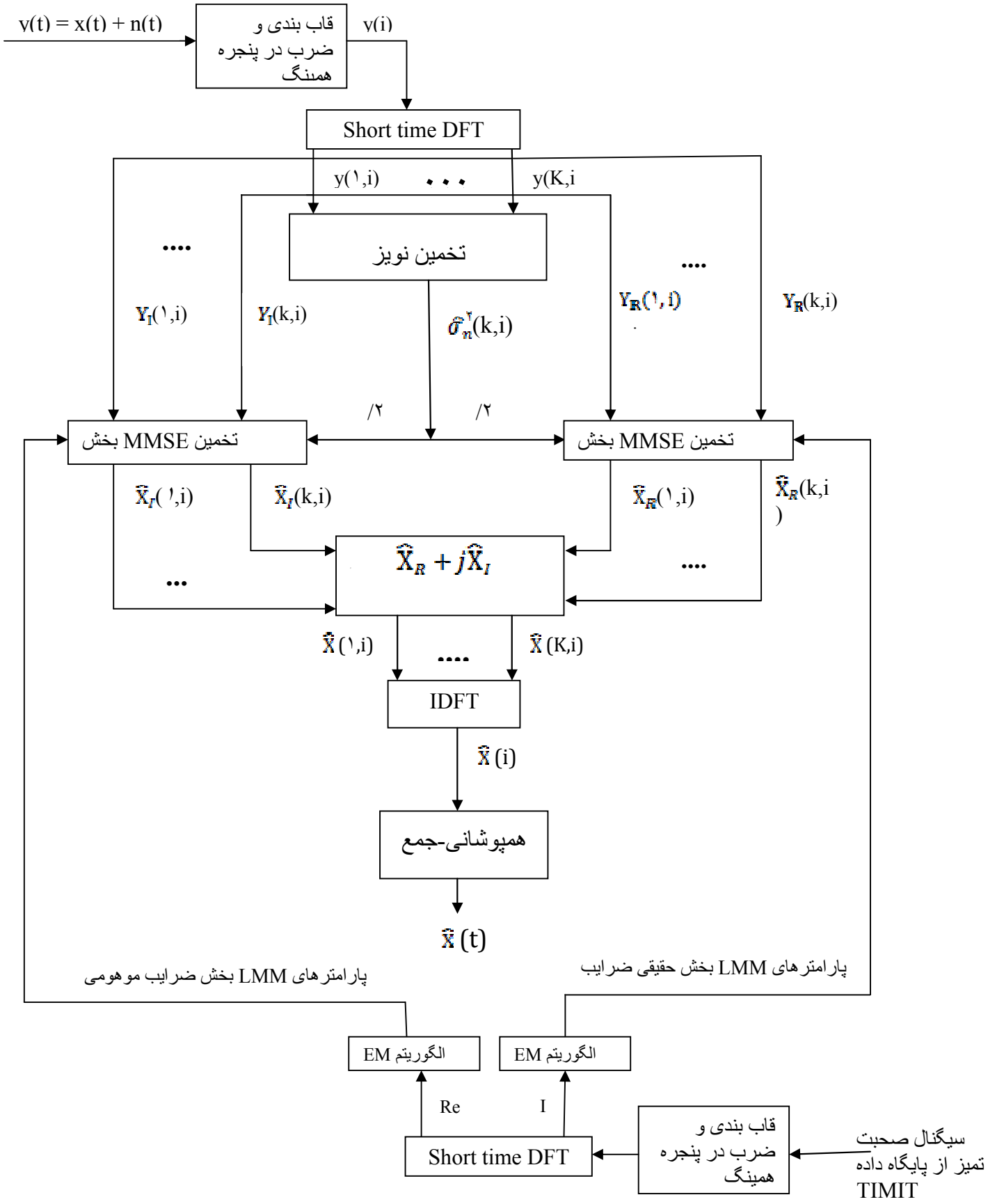
بلوک دیاگرام روش پیشنهادی در شکل (۵-۱) آورده شده است. کلیه بخش های این بلوک در ادامه فصل و بعد از بخش معرفی پایگاه داده، توضیح داده خواهند شد.

در انتهای فصل نیز به بررسی نتایج ارزیابی روش پیشنهادی و مقایسه آن با سه روش دیگر بهسازی گفتار براساس معیار های ارزیابی بیان شده، پرداخته می شود.

۵-۲- معرفی پایگاه داده

در این پایان نامه از پایگاه داده TIMIT[۱] استفاده شده است. علت استفاده از این پایگاه داده معتبر و پر استفاده بودن آن است. در واقع در بیشتر مقالات معتبر در زمینه بهسازی گفتار از این پایگاه داده استفاده شده است.این پایگاه داده مجموعه ای از جملات است، که برای مطالعات در حوزه گفتار، علی‌الخصوص تشخیص خودکار گفتار ایجاد شده است. TIMIT شامل ۶۳۰۰ جمله می‌باشد. این ۶۳۰۰ جمله در واقع ده جمله است، که به وسیله ۶۳۰ گوینده از هشت منطقه از ایالات متحده

[۱] - Texas Instruments and Massachusette Institute of Technology



شکل(۵-۱): بلوک دیاگرام روش پیشنهادی .

آمریکا، با لهجه های متفاوت، ادا شده است. این افراد، کسانی بودند که از بچگی در آن منطقه زندگی کرده اند و کاملا بر لهجه آن منطقه مسلط می باشند.

جملات این پایگاه داده به دو دسته آموزشی و تست تقسیم شده اند که بر اساس نام و جنسیت گوینده، نامگذاری شده اند. پایگاه داده TIMIT نتیجه تلاش مشترک MIT[۱] و TI[۲] است [۶۵].

۵-۳- تخمین طیف سیگنال تمیز

در فصل گذشته روش های بهسازی آماری توضیح داده شدند. روش پیشنهادی در این پایان نامه در حوزه روش های آماری است که در آن از تخمین گر MMSE استفاده شده است. همچنین قبلا تخمین طیف سیگنال تمیز هم با فرض گوسی بودن مولفه های طیف سیگنال تمیز و سیگنال نویز (فیلتر وینر) و هم با فرض غیر گوسی بودن مولفه های طیف سیگنال تمیز (گاما) و گوسی بودن طیف سیگنال نویز بررسی شد. در روش پیشنهادی نیز طیف سیگنال تمیز با فرض توزیع مخلوطی از لاپلاس ها برای سیگنال تمیز و توزیع گوسی برای نویز، به دست خواهد آمد.

۵-۳-۱- فرضیات

قبل از هر چیز لازم است فرضیات اساسی به کار رفته در روش پیشنهادی مطرح شود. در روش پیشنهادی فرض می شود که مولفه های طیف سگینال تمیز مستقل از هم بوده [۴۲] و تابع چگالی احتمال آنها مخلوطی از لاپلاس ها با میانگین غیر صفر می باشد. فرض دیگر این است که نویز از سیگنال تمیز مستقل بوده [۴۲] و مولفه های طیف آن از توزیع گوسی با میانگین صفر پیروی می کنند. فرض بعدی که به طور عمده ای مسئله تخمین را ساده تر می کند، مستقل بودن بخش حقیقی و موهومی طیف سیگنال است [۴۴].

برای به دست آوردن روابط تخمین گر MMSE، مساله به صورت زیر مطرح می شود. ابتدا فرض می شود که سیگنال گفتار تمیز $s(n)$ در اثر جمع شدن با نویز $d(n)$ سیگنال نویزی $y(n)$ را طبق

---

۱ - Massachusette Institute of Technology
۲ - Texas Instruments

رابطه (۵-۱) تولید کرده است.

$$y(i)=s(i)+n(i) \qquad (۱-۵)$$

سپس سیگنال ها به وسیله تبدیل فوریه زمان کوتاه به حوزه فرکانس برده می شود، یعنی سیگنال به قاب هایی که با هم همپوشانی[۱] دارند تقسیم شده و سپس قاب ها در یک پنجره مثل همینگ[۲] یا هنینگ[۳] ضرب شده و از آنها تبدیل فوریه گرفته می شود. علت این که سیگنال به قاب های کوچک تقسیم می شود، این است که سیگنال صوت عملا یک سیگنال غیر ایستان است، ولی در این قاب های کوچک ایستان فرض می شود. در این پایان نامه طول قاب ها ۳۲ میلی ثانیه و با همپوشانی ۵۰٪ انتخاب شده و از پنجره همینگ برای ضرب در قاب های سیگنال، استفاده می شود [۲۲] و [۳۷]. پنجره همینگ و طول قاب ۳۲ میلی ثانیه مواردی است که در معتبرترین منابع و مراجع بهسازی گفتار پیشنهاد شده است. رابطه (۵-۱) به صورت زیر در حوزه فرکانس بازنویسی می شود:

$$Y(m,k) = S(m,k) + N(m,k) \qquad (۲-۵)$$

که درآن m شماره قاب و k اندیس متناظر با فرکانس $2k\pi/L$ می باشد. (L تعداد نمونه های یک قاب است). برای راحتی نمایش تبدیل فوریه زمان کوتاه سیگنال نویزی، سیگنال تمیز و نویز را در قاب mام و فرکانس kام، به ترتیب با S ,Y و N نمایش می دهیم. Y_R و Y_I ، S_R و S_I ، N_R و N_I به ترتیب نمایش دهنده بخش حقیقی و موهومی سیگنال نویزی، سیگنال تمیز و نویز هستند.

۵-۳-۲- مدل کردن سیگنال تمیز با توزیع مخلوط لاپلاس

در این بخش، برای واضح تر شدن مطالب، ابتدا توضیح لاپلاس بررسی شده است.

توزیع لاپلاس به صورت زیر نمایش داده می شود [۶۴]:

$$L(x.c.m) = ce^{-\sqrt{2}c|x-m|} \qquad (۳-۵)$$

---

۱ - overlapp
۲ - Hamming
۳ - Hanning

که درآن $\frac{1}{c^2}$ بیانگر واریانس و m بیانگر میانگین توزیع می باشد. شکل (۵-۲) توزیع لاپلاس را به ازای میانگین و واریانس های متفاوت نشان می دهد.

همانطور که قبلا ذکر شد، تابع چگالی احتمال سیگنال صحبت در حوزه زمان بهتر است با توزیع گاما یا لاپلاس به جای توزیع گوسی مدل شود. در حوزه تبدیل فوریه زمان کوتاه (سایز پنجره کمتر از ۱۰۰ میلی ثانیه) نیز توزیع های گاما و لاپلاس برای مدل کردن بخش حقیقی و موهومی تابع چگالی احتمال ضرایب طیف سیگنال صحبت مناسب تر می باشند [۴۳] .

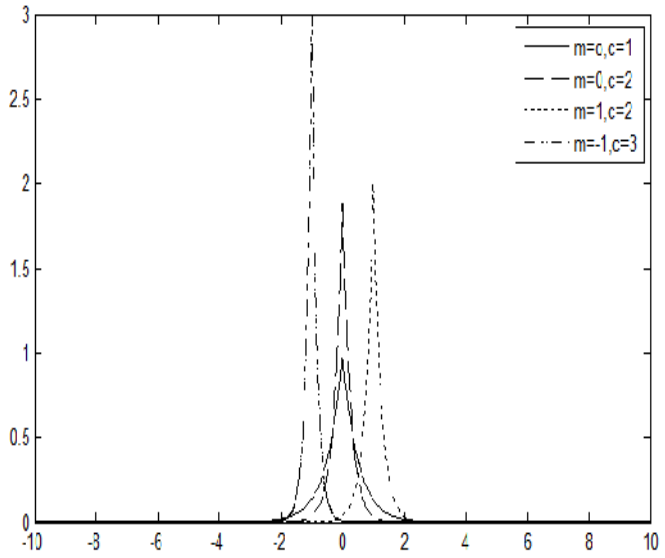
۵-۳-۲-۱- توزیع مخلوط لاپلاس (LMM)

توزیع مخلوطی از لاپلاس ها (LMM) به صورت زیر تعریف می شود [۴۶]:

$$p(x) = \sum_{i=1}^{N} \alpha_i \; L\left(x, c_i, m_i\right) = \sum_{i=1}^{N} \alpha_i \; c_i e^{-\sqrt{2}c_i|x-m_i|} \qquad (۴-۵)$$

که درآن α_i ، m_i و c_i به ترتیب ضریب، میانگین، و واریانس هر کدام از لاپلاس ها و N تعداد لاپلاس هاست و داریم : $\sum_{i=1}^{N} \alpha_i = 1$. متداول ترین روش برای تخمین این پارامترها EM الگوریتم است که در بخش بعدی توضیح داده خواهد شد.

در شکل های (۵-۳) و (۵-۴) برای نمونه هیستوگرام [۱] بخش حقیقی و موهومی تبدیل فوریه زمان



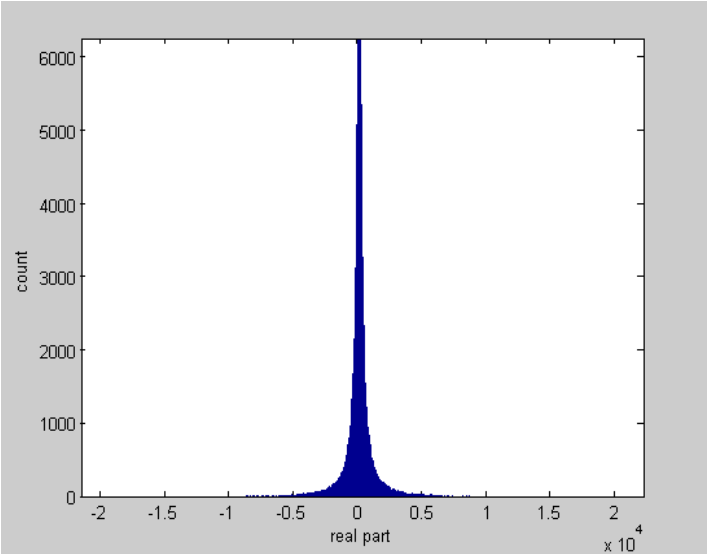
[۱] - Histogram

شکل (۵-۲): توزیع لاپلاس به ازای واریانس و میانگین های متفاوت .

کوتاه یک میلیون داده بدون نویز که از پایگاه داده TIMIT استخراج شده اند، رسم شده است.

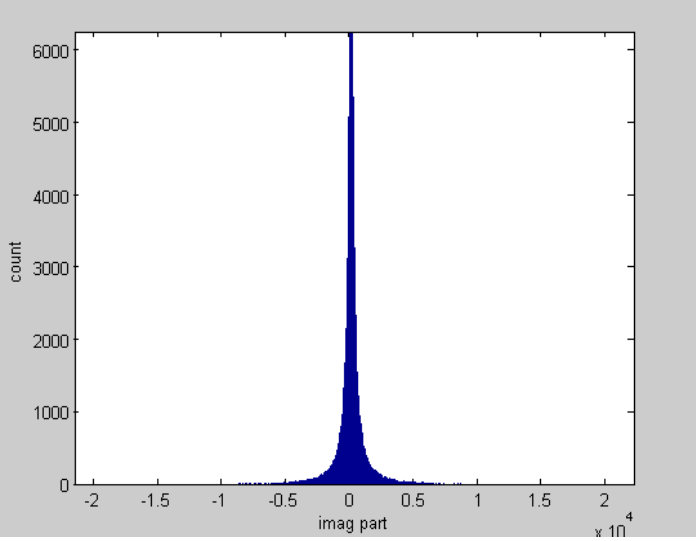
همانطور که مشخص است، هیستو گرام ها شباهت زیادی به توزیع لاپلاس دارند. ولی با بررسی دقیق تر نمودارها و مقادیر آنها، و مقایسه آنها با نمودار توزیع لاپلاس که در شکل (۵-۲) آورده شد، مشخص می شود که توزیع داده ها را نمی توان تنها با یک لاپلاس با میانگین صفر تقریب زد. زیرا هیستوگرام ها حول مبدا متقارن نیستند، ولی توزیع لاپلاس با میانگین صفر حول مبدا متقارن است. از طرفی اگر از یک لاپلاس با میانگین غیر صفر هم استفاده شود، دنباله های[۱] توزیع داده ها به درستی مدل نمی شود. همچنین مجموع چند لاپلاس با میانگین صفر نیز باز حول مبدا متقارن بوده و برای تقریب مناسب نیست. به همین جهت به نظر می رسد، مناسب ترین گزینه برای مدل کردن طیف سیگنال تمیز توزیع مخلوطی از چند لاپلاس با میانگین های غیر صفر باشد.

در شکل (۵-۵) هیستوگرام بخش حقیقی تبدیل فوریه زمان کوتاه سیگنال تمیز به ترتیب با یک لاپلاس با میانگین غیر صفر، مخلوطی از شش لاپلاس با میانگین غیر صفر و مخلوطی از شش گوسی با میانگین غیرصفر مدل شده است. (تصویر برای وضوح بیشتر در حالت بزرگنمایی است) با توجه به



[۱] - tail

شکل (۵-۳): هیستوگرام بخش حقیقی تبدیل فوریه زمان کوتا سیگنال تمیز .



شکل (۵-۴): هیستوگرام بخش موهومی تبدیل فوریه زمان کوتا سیگنال تمیز .

شکل مشاهده می شود که توزیع مخلوطی از لاپلاس ها تقریب بهتری برای هیستوگرام داده ها، نسبت به سایرین می باشد. *لذا در این پایان نامه، توزیع مخلوطی از لاپلاس ها با میانگین غیر صفر برای مدل کردن بخش حقیقی و موهومی پیشنهاد می شود.*

لازم به ذکر است که، پیش از این از توزیع مخلوط لاپلاس در پردازش تصویر و حوزه های دیگر پردازش صوت، مثل تشخیص گفتار، استفاده شده است. ولی در زمینه بهسازی گفتار، تحقیقی در این زمینه مشاهده نشده است.

با توجه به توضیحات قبل، پیشنهاد می شود که مطابق روابط (۵-۵) و (۵-۶) بخش حقیقی و موهو- می طیف سیگنال تمیز با LMM مدل شوند. یعنی:

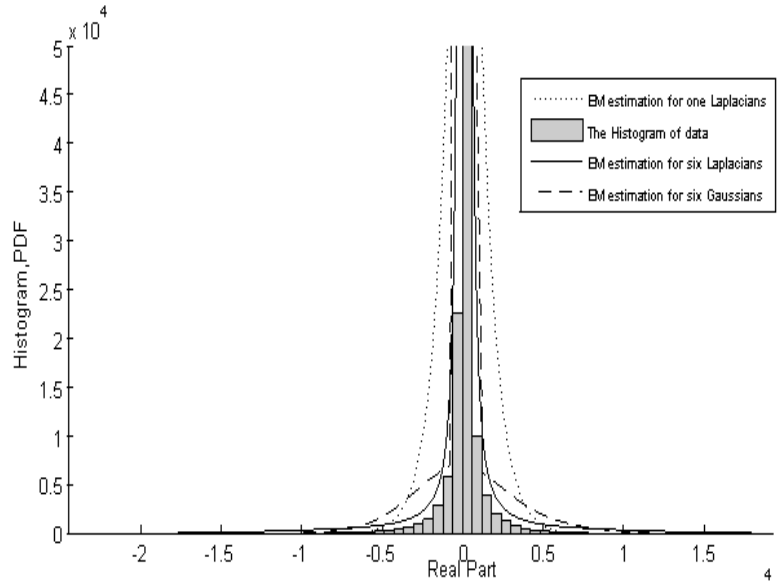
$$p(S_R) = \sum_{i=1}^{N} \alpha_i \; c_i e^{-\sqrt{2}c_i|S_R - m_i|} \qquad (۵-۵)$$

$$p(S_I) = \sum_{j=1}^{N} \alpha_j \; c_j e^{-\sqrt{2}c_j|S_I - m_j|} \qquad (۵-۶)$$

همچنین برای نویز، توزیع گوسی با میانگین صفر به صورت روابط (۵-۷) و (۵-۸) پیشنهاد می شود.

$$P(N_R) = \frac{1}{\sqrt{\pi}\sigma_n} \exp\left(\frac{-(N_R)^2}{\sigma_n^2}\right) \qquad (۵-۷)$$

$$P(N_I) = \frac{1}{\sqrt{\pi}\sigma_n} \exp\left(\frac{-(N_I)^2}{\sigma_n^2}\right) \qquad (۵-۸)$$



شکل (۵-۵): نقطه چین: تخمین توزیع بخش حقیقی تبدیل فوریه زمان کوتاه سیگنال تمیز با یک لاپلاس، خط یکپارچه : تخمین توزیع بخش حقیقی تبدیل فوریه زمان کوتاه سیگنال تمیز با مخلوطی از شش لاپلاس، خط چین: تخمین توزیع بخش حقیقی تبدیل فوریه زمان کوتاه سیگنال تمیز با مخلوطی از شش گوسی.

۵-۳-۳- تخمین طیف سیگنال تمیز با تخمین گر MMSE

در فصل گذشته روابط تخمین گر MMSE برای تخمین طیف سیگنال تمیز، با فرض توزیع گاما برای مولفه های طیف، به دست آمد. تخمین سیگنال تمیز از روی سیگنال نویزی با تخمین گر MMSE همانطور که قبلا توضیح داده شد، با فرض مستقل بودن بخش حقیقی و موهومی طیف سیگنال[۴۳]، به صورت زیر خواهد بود:

$$E(S|Y) = E(S_R|Y_R) + jE(S_I|Y_I) \qquad (۹-۵)$$

تخمین هر کدام از بخش های حقیقی و موهومی به صورت جداگانه و ترکیب آنها، تخمین نهایی طیف را خواهد داد. تخمین MMSE بخش حقیقی طیف به صورت رابطه (۵-۱۰) می باشد.

$$\widehat{S_R} = E[S_R|Y_R] = \int_{-\infty}^{+\infty} S_R P(S_R|Y_R) dS_R = \frac{\int_{-\infty}^{+\infty} S_R p(Y_R|S_R) p(S_R) dS_R}{p(Y_R)} \qquad (۱۰-۵)$$

که برای محاسبه آن، ابتدا باید $P(Y_R|S_R)$ و $P(Y_R)$ محاسبه شوند. $P(S_R)$ نیز توزیع مخلوط لاپلاس است که پارامترهای آن باید محاسبه شوند. با توجه به اینکه بخش حقیقی طیف سیگنال نویزی برابر

مجموع بخش های حقیقی طیف سیگنال تمیز و نویز است و با فرض گوسی بودن توزیع نویز $P(Y_R|S_R)$ به صورت زیر در نظر گرفته می شود:

$$P(Y_R|S_R) = \frac{1}{\sqrt{\pi}\sigma_n} \exp\left(\frac{-(Y_R-S_R)^2}{\sigma_n^2}\right) \quad (۱۱-۵)$$

با جایگذاری رابطه (۵-۱۱) به جای $P(Y_R|S_R)$ و (۵-۵) برای $P(S_R)$ در رابطه (۵-۱۰)، این رابطه به صورت رابطه (۵-۱۲) درخواهد آمد.

$$E(S_R|Y_R) = \frac{1}{\sqrt{\pi}\sigma_n p(Y_R)} \int_{-\infty}^{+\infty} S_R \sum_{i=1}^{N} \alpha_i c_i \exp\left(\frac{-(Y_R-S_R)^2}{\sigma_n^2}\right) \exp\left(-2c_i|s_R-m_i|\right) ds_R \quad (۱۲-۵)$$

برای محاسبه این انتگرال، از نتایج انتگرال های حل شده در [۴۳] استفاده شده است. این انتگرال ها با تغییر متغیر و اندکی تغییرات و دستکاری به صورت انتگرال (۵-۱۲) در می آیند. سپس از جواب های این انتگرال ها برای حل رابطه (۵-۱۲) استفاده می شود. روابط (۵-۱۳) و (۵-۱۴) نشان دهنده این انتگرال هاست.

$$\int_{-\infty}^{+\infty} s_R \exp\left(\frac{-(Y_R-S_R)^2}{\sigma_n^2}\right) \exp\left(-2c|s_R|\right) dS_R = \frac{\sqrt{\pi}\sigma_n^2 \exp(c^2\sigma_n^2)}{2}$$
$$\left[L_{R1}(Y_R)\exp(2cY_R)\operatorname{erfc}(L_{R1}(Y_R)) - L_{R2}(Y_R)\exp(-2cY_R)\operatorname{erfc}(L_{R2}(Y_R))\right] \quad (۱۳-۵)$$

$$\int_{-\infty}^{+\infty} \exp\left(\frac{-(Y_R-S_R)^2}{\sigma_n^2}\right) \exp\left(-2c|s_R|\right) dS_R = \frac{\sqrt{\pi}\sigma_n \exp(c^2\sigma_n^2)}{2}$$
$$\left[\exp(2cY_R)\operatorname{erfc}\left(L_{R1}(Y_R) + \exp(-2cY_R)\operatorname{erfc}(L_{R2}(Y_R))\right)\right] \quad (۱۴-۵)$$

دراین روابط erfc متمم تابع خطا (complementary error function) است که از رابطه زیر محاسبه می شود:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt = 1 - \operatorname{erfc}(x) \quad (۱۵-۵)$$

و $L_{R1}(Y_R)$ و $L_{R2}(Y_R)$ به صورت زیر تعریف می شوند:

$$L_{R1}(Y_R) = c\sigma_n + \frac{Y_R}{\sigma_n} \quad (۱۶-۵)$$

$$L_{R_2}(Y_R) = c\sigma_n - \frac{Y_R}{\sigma_n} \qquad (۱۷-۵)$$

با عوض کردن جای انتگرال و سیگما در رابطه (۵-۱۲) و تغییر متغیر در آن به صورت زیر :

$$S_R - m_i = t_i \Longrightarrow S_R = t_i + m_i \quad , \quad ds_R = dt_i \qquad (۱۸-۵)$$

و با استفاده از روابط (۵-۱۳) تا (۵-۱۷)، رابطه (۵-۱۲) به صورت رابطه (۵-۱۹) در می‌آید.

$$E(S_R|Y_R) = \frac{1}{2p(Y_R)} \sum_{i=1}^{N} \alpha_i c_i \exp(c_i^2 \sigma_n^2) \{\sigma_n [L_{R_1}(Y_R - m_i) \exp(2c_i(Y_R - m_i)) \operatorname{erfc}(L_{R_1}(Y_R - m_i)) - L_{R_2}(Y_R - m_i) \exp(-2C_i(Y_R - m_i)) \operatorname{erfc}(L_{R_2}(Y_R - m_i))] + m_i [\exp(2c_i(Y_R - m_i)) \operatorname{erfc}(L_{R_1}(Y_R - m_i)) + \exp(-2c_i(Y_R - m_i)) \operatorname{erfc}(L_{R_2}(Y_R - m_i))]\} \qquad (۱۹-۵)$$

از رابطه (۵-۱۹) مشخص می شود که، $P(Y_R)$ نیز باید محاسبه شود. از آنجایی‌که $P(Y_R)$ برابر با رابطه $\int_{-\infty}^{+\infty} P(S_R)\, P(Y_R|S_R) dS_R$ است، با توجه به روابط (۵-۱۳) تا (۵-۱۷) و تغییر متغیر قبل، رابطه (۵-۲۰) برای محاسبه آن به دست می آید.

$$P(Y_R) = \frac{1}{2} \sum_{i=1}^{N} \alpha_i c_i \exp(c_i^2 \sigma_n^2) [\exp(2c_i(Y_R - m_i)) \operatorname{erfc}(L_{R_1}(Y_R - m_i)) + \exp(-2c_i(Y_R - m_i)) \operatorname{erfc}(L_{R_2}(Y_R - m_i))] \qquad (۲۰-۵)$$

با جایگذاری $P(Y_R)$ در رابطه (۵-۱۷) تخمین نهایی بخش حقیقی طیف به صورت رابطه (۵-۲۱) به دست خواهد آمد.

$$\widehat{S_R} = E(S_R|Y_R) = \qquad (۲۱-۵)$$

$$\frac{\sum_{i=1}^{N} \alpha_i c_i \exp(c_i^2 \sigma_n^2) \left[\begin{array}{c} \sigma_n \left(L_{R_1}(Y_R - m_i) \exp(2c_i(Y_R - M_i)) \operatorname{erfc}(L_{R_1}(Y_R - m_i)) - L_{R_2}(Y_R - m_i) \exp(-2C_i(Y_R - m_i)) \operatorname{erfc}(L_{R_2}(Y_R - m_i))\right) + \\ m_i \left(\exp(2c_i(Y_R - m_i)) \operatorname{erfc}(L_{R_1}(Y_R - m_i)) + \exp(-2c_i(Y_R - m_i)) \operatorname{erfc}(L_{R_2}(Y_R - m_i))\right) \end{array}\right]}{\sum_{i=1}^{N} \alpha_i c_i \exp(c_i^2 \sigma_n^2) [\exp(2c_i(Y_R - m_i)) \operatorname{erfc}(L_{R_1}(Y_R - m_i)) + \exp(-2c_i(Y_R - m_i)) \operatorname{erfc}(L_{R_2}(Y_R - m_i))]}$$

برای بخش موهومی طیف سیگنال هم به طریق مشابه عمل شده و رابطه (۵-۲۲) به دست می آید.

$$\widehat{S_I} = E(S_I|Y_I) = \qquad (۲۲-۵)$$

$$\frac{\sum_{i=1}^{N} \alpha_i c_i \exp(c_i^2 \sigma_n^2) \left[\begin{array}{c} \sigma_n \left(L_{R_1}(Y_I - m_i) \exp(2c_i(Y_I - M_i)) erfc(L_{R_1}(Y_I - m_i)) - L_{R_2}(Y_I - m_i) \exp(-2c_i(Y_I - m_i)) erfc(L_{R_2}(Y_I - m_i))\right) + \\ m_i \left(\exp(2c_i(Y_I - m_i)) erfc(L_{R_1}(Y_I - m_i)) + \exp(-2c_i(Y_I - m_i)) erfc(L_{R_2}(Y_I - m_i))\right) \end{array}\right]}{\sum_{i=1}^{N} \alpha_i c_i \exp(c_i^2 \sigma_n^2) [\exp(2c_i(Y_I - m_i)) erfc(L_{R_1}(Y_I - m_i)) + \exp(-2c_i(Y_I - m_i)) erfc(L_{R_2}(Y_I - m_i))]}$$

دو رابطه (۵-۲۱) و (۵-۲۲) مطابق رابطه (۵-۹) با هم ترکیب شده و تخمین نهایی طیف را خواهند داد. سپس با تبدیل معکوس فوریه زمان کوتاه گرفتن از این تخمین، سیگنال بهسازی شده قاب بندی شده (ماتریسی که سطرها یا ستون های آن قاب های سیگنال هستند) در حوزه زمان به دست خواهد آمد. با اعمال روش همپوشانی-جمع بر روی این ماتریس به سیگنال اصلی خواهیم رسید. در روش همپوشانی – جمع، همپوشانی بین قاب ها حذف کرده و آنها را با هم ادغام می کنند.

۵-۴- تخمین پارامترهای توزیع سیگنال تمیز با EM الگوریتم

در قسمت قبلی گفته شد که برای تخمین سیگنال تمیز از روی سیگنال نویزی با روش MMSE، توزیع سیگنال تمیز مورد نیاز است. توزیع سیگنال تمیز نیز که LMM فرض شد، دارای پارامترهای نامعلومی مانند میانگین، واریانس و ضریب است که باید به دست آورده شوند. برای به دست آوردن این پارامترها از EM الگوریتم استفاده شده که در این بخش به معرفی آن پرداخته شده است. در واقع ابتدا، به وسیله داده های آموزشی مناسب و با EM الگوریتم پارامترهای LMM به دست می آید و سپس این پارامترها در رابطه های (۵-۲۱) و (۵-۲۲) جایگذاری می شود.

به طور کلی EM (امید ریاضی بیشینه) الگوریتم در مسائلی به کار می رود که می خواهیم مجموعه ای از پارامترهای θ را تخمین بزنیم که مبتنی بر یک توزیع احتمالی اند و تنها بخشی از داده هایی که توسط این توزیع احتمالی تولید شده اند را داریم نه کل آنها را. ارائه کامل این الگوریتم، در مقاله اولیه Dempster;Larid and Rubin(۱۹۷۷) [۳۴] و در کتاب های Little and Rubin (۱۹۸۷) و McLachlan and Krishman(۱۹۹۷) یافت می شود. در این بخش به تشریح روش بیشترین شباهت و نحوه ارتباط آن با EM الگوریتم برای مسئله تخمین پارامترهای نامعلوم می پردازیم. سپس به تشریح تخمین پارامترهای تابع چگالی احتمال مخلوطی از لاپلاس ها خواهیم پرداخت.

۵-۴-۱- بیشترین شباهت

درباره این روش در فصل های قبل توضیح مختصری داده شد که در اینجا برای بررسی نحوه ارتباط آن با EM الگوریتم دوباره تکرار می شود. فرض می شود N داده با توزیع فرضی مشخصی با پارامترهای θ که نامعلوم و معین است، موجود باشند و هدف به دست آوردن پارامترهای توزیع فرضی مشخص باشد. همچنین فرض می شود که بردار داده ها مستقل و به طور یکسان توزیع شده[۱] (i. i. d) با توزیع p باشند. در نتیجه تابع چگالی داده ها به صورت زیر می باشد:

$$p(x|\theta) = \Pi_{i=1}^{N} p(X_i|\theta) = \mathcal{L}(\theta|x) \qquad (۲۳-۵)$$

که به آن تابع شباهت نیز می گویند. حال θ به گونه ای به دست می آید که تابع شباهت را طبق رابطه (۵-۲۴) بیشینه کند.

$$\theta^* = \text{argmax}\mathcal{L}(\theta|x) \qquad (۲۴-۵)$$

معمولا به علت سادگی بیشتر از لگاریتم تابع شباهت در محاسبات استفاده می شود. درجه پیچیدگی مساله تخمین بستگی به فرم $p(x|\theta)$ دارد. مثلا اگر این تابع یک گوسی و پارامترهای مجهول میانگین و واریانس آن باشد، به راحتی از لگاریتم تابع شباهت نسبت به پارامترهای مجهول مشتق گرفته و مساوی صفر قرار داده می شوند و پارامترها با حل این معادله به دست می آیند. ولی در اغلب مسایل به دست آوردن پارامترها با مساوی صفر قرار دادن مشتق، غیر ممکن است و باید از تکنیک های دیگری استفاده شود [۵].

۵-۴-۲- EM الگوریتم

EM الگوریتم یک روش کلی برای به دست آوردن تخمین بیشترین شباهت پارامترهای توزیعی است که داده های مربوط به آن توزیع کامل نیستند. یکی از کاربردهای مهم دیگر این الگوریتم در مواقعی است که بهینه کردن تابع شباهت یا لگاریتم آن، کار مشکلی باشد، اما بتوان شکل این تابع را

---

[۱]- independent and identically distributed.

با فرض وجود داده های دیگری ساده تر کرد. به این داده ها، داده های گمشده[۱] یا مشاهده نشده گفته می شود. فرض می شود، کل داده ها اجتماع داده های مشاهده شده[۲] (که به آنها، داده های ناقص[۳] نیز گفته می شود) و داده های گمشده باشند. داده های مشاهده شده با X نمایش داده می شوند و فرض می شود که این داده ها از توزیع خاصی پیروی می کنند. داده های گمشده فرضی نیز با Y نشان داده شده و کل داده ها هم با Z=(X,Y) نمایش داده می شوند. تابع توزیع توام X و Y که به آن تابع توزیع داده های کامل[۴] نیز گفته می شود، به صورت زیر تعریف می شود:

$$p(z|\theta)= p\left(x,y\middle|\theta\right)= \left(y\middle|x,\theta\right)p(x|\theta) \qquad (۲۵-۵)$$

حال تابع شباهت داده های کامل به صورت $L(\theta|z) = L(\theta|x,y) = p(x,y|\theta)$ و تابع شباهت داده های ناقص (مشاهده شده) به صورت $L(\theta|x) = p(x|\theta)$ تعریف می شوند. در این روش سعی می شود داده های مشاهده نشده به صورتی تعریف شوند که شکل تابع شباهت داده های کامل ساده تر از تابع شباهت داده های ناقص باشد. این نکته قابل توجه است که، تابع شباهت داده های کامل یک متغیر تصادفی است. زیرا داده های ناقص و پارامترها ی نامعلوم، معین و داده های مشاهده نشده ماهیتی تصادفی دارند. EM الگوریتم ابتدا، امید ریاضی لگاریتم تابع شباهت داده های کامل را که به صورت رابطه (۵-۲۶) تعریف می شود، محاسبه می کند. این مرحله، مرحله E (محاسبه امید ریاضی) الگوریتم نامیده می شود.

$$Q(\theta,\theta^{(i-1)}) = E\left[\log P(x,y|\theta)\middle|x,\theta^{(i-1)}\right] \qquad (۲۶-۵)$$

که درآن $\theta^{(i-1)}$ پارامترهای جاری یا پارامترهایی هستند که از مرحله قبل تخمین محاسبه شده اند و برای محاسبه امید ریاضی و پارامترهای جدید استفاده می شوند، و θ پارامترهایی است که با بهینه شدن Q به دست می آیند. در واقع مقدار این امید ریاضی[۵] با استفاده از داده های ناقص و

---

۱ - -missing value
۲- observed
۳ - incomplete
۴ - complete
۵ - Expectation

پارامترهای جاری محاسبه می شود. سمت راست رابطه (۵-۲۶) را می توان به صورت رابطه (۵-۲۷) بازنویسی کرد:

$$E\left[\log P(x,y|\theta)\middle|x,\theta^{(i-1)}\right] = \int_{y\in Y} \log P(x,y|\theta) f\left(y\middle|x,\theta^{i-1}\right)dy \qquad (۲۷-۵)$$

که درآن $f\left(y\middle|x,\theta^{i-1}\right)$ تابع توزیع مرزی داده های مشاهده نشده می باشد که هم به پارامترهای جاری و هم به داده های مشاهده شده بستگی دارد. در بهترین حالت این توزیع، تابع ساده ای از پارامترهای جاری و شاید داده های مشاهده شد می باشد و در بدترین حالت ممکن است فراهم کردن آن کار بسیار پیچیده ای باشد. □ هم نشان دهنده فضایی است که y می تواند مقادیر خود را روی آن اختیار کند. سمت راست رابطه (۵-۲۷) یک تابع معین است که می توان از آن مشتق گرفت. در مرحله بعدی EM الگوریتم، $Q\left(\theta,\theta^{(i-1)}\right)$ نسبت به θ بهینه می شود. یعنی از تابع $Q\left(\theta,\theta^{(i-1)}\right)$ نسبت به θ مشتق گرفته می شود و این مشتق مساوی با صفر قرار داده می شود. به این مرحله M یا مرحله بیشینه کردن[۱] می گویند.

خلاصه EM الگوریتم را می توان بدین صورت بیان کرد: EM الگوریتم یک روش بیشینه کردن لگاریتم تابع شباهت داده های ناقص، با استفاده از ماکزیمم کردن امید ریاضی لگاریتم تابع شباهت داده های کامل، به صورت تکراری می باشد. الگوریتم با یک مقداردهی اولیه θ^{g} در فضای θ آغاز شده و امید ریاضی $L\left(\theta\middle|x,y\right)$ با کمک داده های مشاهده شده وθ^{g} به عنوان پارامترهای توزیع داده های مشاهده شده، محاسبه می شود. این مرحله، مرحله E (محاسبه امید ریاضی) الگوریتم نامیده می شود. در مرحله بعدی، امید ریاضی محاسبه شده نسبت به θ بیشینه شده و θ جدید محاسبه می گردد. به این مرحله M یا مرحله بیشینه کردن می گویند. سپس با θ جدید امید ریاضی لگاریتم تابع شباهت دوباره محاسبه می شود. این دو مرحله اینقدر تکرار می شوند که مقدار پارامترها دردو تکرار متوالی برابر شوند. اما در عمل به خاطر کم شدن حجم محاسبات، الگوریتم را

---

[۱] - maximization

وقتی مقدار پارامترهای حساب شده در دو مرحله به مقدار کوچکی اختلاف داشته باشند، متوقف می کنند.

در EM الگوریتم:

- در هر تکرار EM الگوریتم، امید ریاضی لگاریتم تابع شباهت داده های کامل و لگاریتم تابع شباهت داده های ناقص حتما افزایش می یابد.
- EM الگوریتم، به جواب تخمین روش بیشترین شباهت برای بیشیته کردن تابع شباهت داده های ناقص همگرا می شود.
- در EM الگوریتم همگرایی به یک ماکزیمم محلی تضمین شده است.
- استخراج پارامترهای مدل از تابع شباهت داده های کامل به مراتب راحت از استخراج آنها از تابع شباهت داده های ناقص می باشد [۳۴].

۵-۴-۳- محاسبه پارامترهای توابع توزیع مخلوط

به دست آوردن پارامترهای مخلوطی از چگالی ها یکی از مهمترین و گسترده ترین کاربرد های EM الگوریتم در شناسایی الگو می باشد. در این قسمت به دست آوردن پارامترهای توابع توزیع مخلوط را برای حالت کلی به دست می آوریم و سپس آن را برای چگالی مخلوطی از لاپلاس ها بسط خواهیم داد.

مدل احتمالاتی زیر را در نظر بگیرید:

$$(X|\theta) = \sum_{i=1}^{M} \alpha_i p_i (X|\theta_i) \qquad (۲۸-۵)$$

که در آن p_iها توابع چگالی احتمالی هستند که با ضرایب α_i با هم جمع شده اند و $(\alpha_1, \dots, \alpha_M, \theta_1, \dots, \theta_M)$ پارامترهایی هستند که با توجه به قید $\sum_{i=1}^{M} \alpha_i = 1$ محاسبه خواهند شد.

لگاریتم تابع شباهت داده های ناقص، به صورت زیر تعریف می شود:

$$\log\big(\mathcal{L}(\theta|x)\big) = \log \prod_{i=1}^{n} p(X_i|\theta) = \sum_{i=1}^{N} \log\Big(\sum_{j=1}^{M} \alpha_i p_{i(X_i|\theta_j)}\Big) \qquad (۲۹-۵)$$

که بهینه کردن آن مشکل می باشد، زیرا مجموعی از چندین لگاریتم است. اگر داده های مشاهده نشده را به صورت $y = \{y_i\}_{i=1}^{N}$ فرض کنیم، که مقادیر آن بیان کننده این است که کدام داده به وسیله کدام جز[1] از چگالی احتمال مخلوط تولید می شود، تابع شباهت تا حد زیادی ساده خواهد شد. در واقع برای هر i داریم : $y_i \in 1, \dots\dots, M$ و اگر iامین نمونه با kامین جز مخلوط تولید شد، خواهیم داشت: $y_i = k$ در نتیجه رابطه (۵-۲۹) به صورت زیر در خواهد آمد:

$$\log(\mathcal{L}(\theta|x,y)) = \log(P(x,y|\theta)) =$$

$$\sum_{i=1}^{N} \log\left(P((x_i)|y_i)P(y)\right) = \sum_{i=1}^{N} \log\left(\alpha_{y_i} P_{y_i}(x_i|\theta_{y_i})\right) \qquad (۵-۳۰)$$

که شکل خاصی از توابع چگالی مخلوط را می دهد و با انواع روش ها می توان آن را بهینه کرد ولی مشکل اصلی این است که مقدار y نامشخص است و اگر فرض شود یک بردار تصادفی است، روابط ساده تر خواهند شد. در ابتدا باید توزیع داده های مشاهده نشده بیان شوند. فرض می شود $\theta^g = (\alpha_1^g, \dots, \alpha_M^g, \theta_1^g, \dots, \theta_M^g)$ پارامترهای مناسب برای تابع شباهت $\mathcal{L}\left(\theta^g|x,y\right)$ است. با داشتن این پارامترها به راحتی می توان $P_j\left(x_i\middle|\theta_j^g\right)$ را برای هر i و j حساب کرد. اگر α_j احتمال پیشین جز jام مخلوط لاپلاس تعریف شود، با استفاده از قاعده بیزین می توان تابع چگالی احتمال داده های مشاهده نشده را به صورت رابطه (۵-۳۱) بیان کرد.

$$P((y_i)|\theta^g, x_i) = \frac{\alpha_{y_i}^g p_{y_i}(x_i|\theta_{y_i}^g)}{p(x_i|\theta^g)} = \frac{\alpha_{y_i}^g p_{y_i}(x_i|\theta_{y_i}^g)}{\sum_{k=1}^{M} \alpha_k^g p_k(x_i|\theta_k^g)} \qquad (۵-۳۱)$$

و

$$p(y|x,\theta^g) = \prod_{i=1}^{N} p(y_j|x_j,\theta^g) \qquad (۵-۳۲)$$

در حالی که $y = (y_1, \dots\dots, y_N)$ تعدادی از داده های مشاهده نشده است که به طور مستقل بیرون کشیده شده اند.

حال می توان معادله (۵-۲۶) را به صورت زیر باز نویسی کرد:

$$Q(\theta, \theta^g) =$$

---

[1] - Component

$$\sum_{y\in Y} \log\left(\mathcal{L}(\theta|x,y)\right)p(y|x,\theta^g) = \sum_{y\in Y}\sum_{i=1}^{N} \log\left(\alpha_{y_i} p_{y_i}(x_i|\theta_{y_i})\right)\prod_{j=1}^{N} p(y_j|x_j,\theta^g)$$

$$= \sum_{y_1}^{M}\sum_{y_2}^{M}\dots\sum_{y_N=1}^{M}\sum_{i=1}^{N}\sum_{t=1}^{M} \delta_{t,y_i}\log(\alpha_l p_l(x_i|\theta_l))\prod_{j=1}^{N} p(y_j|x_j,\theta^g) =$$

$$\Sigma_{l=1}^{M}\Sigma_{i=1}^{N}\log(\alpha_l p_l(x_i|\theta_l))\,\Sigma_{y_1=1}^{M}\Sigma_{y_2=1}^{M}\dots\Sigma_{y_N=1}^{M}\delta_{l,y_i}\,\Pi_{j=1}^{M} p(y_i|x_j,\theta^g) \qquad (۳۳-۵)$$

و برای $l \in ۱, \dots, M$ داریم:

$$\sum_{y_1=1}^{M}\sum_{y_2=1}^{M}\dots\sum_{y_N=1}^{M}\delta_{l,y_i}\prod_{j=1}^{M} p(y_i|x_j,\theta^g)$$

$$= \left(\sum_{y_1=1}^{M}\dots\sum_{y_{i-1}=1}^{M}\sum_{y_{i+1}=1}^{M}\dots\sum_{y_N=1}^{M}\prod_{j=1,j\neq i}^{N} p(y_i|x_j,\theta^g)\right)p(l|x_i,\theta^g)$$

$$= \Pi_{j=1,j\neq i}^{N}\left(\Sigma_{y_j=1}^{M} p(y_j|x_j,\theta^g)\right)p(l|x_i,\theta^g) = p(l|x_i,\theta^g) \qquad (۳۴-۵)$$

و چون $\Sigma_{i=1}^{M} p(i|x_j,\theta^g) = ۱$، با استفاده از رابطه (۵-۳۴)، رابطه (۵-۳۳) را می توان به صورت رابطه (۵-۳۵) بازنویسی کرد. برای بیشینه کردن این عبارت، باید عبارات شامل α_l و θ_l به طور مستقل بیشینه شوند.

$$Q(\theta,\theta^g) = \sum_{l=1}^{M}\sum_{i=1}^{N}\log(\alpha_l p_l(x_i|\theta_l))\,p(l|x_i,\theta^g)$$

$$= \Sigma_{l=1}^{M}\Sigma_{i=1}^{N}\log(\alpha_l)p(l|x_i,\theta^g) + \Sigma_{l=1}^{M}\Sigma_{i=1}^{N}\log(p_l(x_i|\theta_l))\,p(l|x_i,\theta^g) \qquad (۳۵-۵)$$

برای به دست آوردن α_l تحت قید $\Sigma_l \alpha_l = ۱$ ، با کمک گرفتن از ضرایب لاگرانژ، معادله زیر به دست خواهد آمد:

$$\frac{\partial}{\partial\alpha_l}\sum_{l=1}^{M}\sum_{i=1}^{N}\log(\alpha_l)p(l|x_i,\theta^g) + \lambda\sum_l \alpha_l - ۱ = ۰$$

$$\Sigma_{i=1}^{N}\, p(l|x_i,\theta^g) + \lambda = ۰ \Longrightarrow \qquad (۳۶-۵)$$

که از حل معادله قبل جواب زیر برای محاسبه ضرایب لاپلاس ها به دست خواهد آمد[۵] :

$$\alpha_l = \frac{۱}{N}\Sigma_{i=1}^{N}\, p(l|x_i,\theta^g) \qquad (۳۷-۵)$$

۵-۴-۴- محاسبه پارامترهای LMM

در این قسمت با توجه به روابط (۵-۳۵) و (۵-۳۷) که در بخش قبل، برای یک توزیع دلخواه به دست آورده شد، تلاش می شود که پارامترهای توزیع LMM که در رابطه (۵-۳۸) آورده شده است، به دست آید.

$$p(\theta) = \sum_{i=1}^{N} \alpha_i \; L\left(\theta, c_i, \theta_i\right) = \sum_{i=1}^{N} \alpha_i \; c_i e^{-2c_i|\theta_n - \theta_i|} \qquad (۵-۳۸)$$

اگر تعداد لاپلاس ها N و تعداد کل داده ها T باشد،رابطه (۵-۳۷) برای محاسبه α_i ها یا همان ضرایب هر جز از مخلوط لاپلاس به صورت رابطه (۵-۳۹) به دست خواهد آمد که درآن $p(i|\theta_n)$ یا احتمال پیشین جز iام به صورت رابطه (۵-۴۰) تعریف می شود.

$$\alpha_i = \frac{1}{T}\sum_{n=1}^{T} \; p(i|\theta_n) \qquad (۵-۳۹)$$

$$p(1|\theta_n) \; = \frac{\alpha_i c_i e^{-2c_i|\theta_n - \theta_i|}}{\sum_{i=1}^{N} \alpha_i \; c_i e^{-2c_i|\theta_n - \theta_i|}} \qquad (۵-۴۰)$$

پس از آن رابطه (۵-۳۵) برای به دست آوردن θ_iها و c_iها به صورت رابطه (۵-۴۱) تشکیل می شود و چون بخش اول رابطه مذکور به این دو پارامتر وابسته نبوده و در مشتق گیری نسبت به این دو پارامتر نقشی ندارد، فقط بخش دوم آن آورده شده است.

$$J(c_i, \theta_i) = \sum_{n=1}^{T} \; \sum_{i=1}^{N} (\log c_i - 2c_i|\theta_n - \theta_i|) \; p(i|\theta_n) \qquad (۵-۴۱)$$

با مشتق گرفتن از رابطه فوق نسبت به c_i و مساوی صفر قرار دادن آن، رابطه (۵-۴۲) برای محاسبه c_i ، برای هر جز از مخلوط لاپلاس به دست خواهد آمد. (واریانس هر جز $\frac{1}{c_i^2}$ می باشد.)

$$c_i = \frac{\sum_{n=1}^{T} \; p(i|\theta_n)}{2\sum_{n=1}^{T} \; |\theta_n - \theta_i| \; p(i|\theta_n)} \qquad (۵-۴۲)$$

همچنین با مشتق گرفتن از رابطه (۵-۴۱) نسبت به θ_i و مساوی صفر قرار دادن آن، رابطه (۵-۴۳) برای محاسبه میانگین هر جز به دست خواهد آمد. مشاهده می شود که در این رابطه به خاطر وجود

تابع علامت (sgn)، به دست آوردن مقدار دقیق برای θ_i ممکن نیست. اما می توان مقدار جدید آن را با کمک تخمین قبلی $|\theta_n - \theta_i|$ به دست آورد. در نتیجه رابطه نهایی (۵-۴۴) برای تقریب میانگین هر جز به دست آمد [۴۶].

$$\frac{\partial J}{\partial \theta_i} = \sum_{n=1}^{T} \left(-2c_i \frac{\partial}{\partial \theta_i} |\theta_n - \theta_i|\right) p(i|\theta_n) =$$

$$\sum_{n=1}^{T} -2c_i \, sgn(\theta_n - \theta_i) \; p(i|\theta_n) = 0 \qquad \Rightarrow$$

$$\sum_{n=1}^{T} sgn(\theta_n - \theta_i) \; p(i|\theta_n) = \sum_{n=1}^{T} \frac{\theta_n - \theta_i}{|\theta_n - \theta_i|} \; p(i|\theta_n) = 0 \qquad (۴۳-۵)$$

$$\sum_{n=1}^{T} \frac{\theta_n}{|\theta_n - \theta_i|} p(i|\theta_n) = \sum_{n=1}^{T} \frac{\theta_i}{|\theta_n - \theta_i|} p(i|\theta_n)$$

$$\Rightarrow \theta_i = \frac{\sum_{n=1}^{T} \frac{\theta_n}{|\theta_n - \theta_i|} p(i|\theta_n)}{\sum_{n=1}^{T} \frac{1}{|\theta_n - \theta_i|} p(i|\theta_n)} \qquad (۴۴-۵)$$

۵-۴-۵- پیاده سازی EM الگوریتم

همانطور که قبلا گفته شد، برای مدل کردن سیگنال بخش حقیقی و موهومی طیف سیگنال تمیز توزیع LMM پیشنهاد شد. همچنین گفته شد که برای محاسبه پارامترهای LMM با استفاده از داده های آموزشی از EM الگوریتم استفاده می شود.

برای پیاده‌سازی EM الگوریتم، از داده های آموزشی برای تخمین پارامترهای توزیع مورد نظر استفاده می‌شود. هر چه تعداد این داده‌ها بیشتر باشد، مدل دقیق تر و بهتری ایجاد می‌شود. در این پایان نامه از حدود ده دقیقه سیگنال تمیز از پایگاه داده TIMIT (۲۰۰ جمله از ۱۰۰ مرد و ۱۰۰ زن) استفاده شده است. این داده ها با نرخ دو کاهش داده شده است. سپس از داده ها، تبدیل فوریه زمان کوتاه گرفته و بخش حقیقی و موهومی طیف از کل قاب ها به عنوان داده آموزشی به طور جداگانه به الگوریتم داده شده تا پارامترهای LMM برای بخش ها حقیقی و موهومی به طور جداگانه محاسبه شود. در واقع EM الگوریتم دو بار اجرا می شود. یکبار با داده های بخش حقیقی و یکبار هم با داده های بخش موهومی و در نهایت دو LMM به دست خواهد آمد.

مقدار دهی اولیه EM الگوریتم برای به دست آوردن پارامترهای LMM را بدین طریق انجام شده است:

- مقدار اولیه α_i ها (ضرایب لاپلاس ها)، برابر ۱/N در نظر گرفته شده است.
- مقدار اولیه θ_i و c_i را هم به صورت تصادفی انتخاب شده است. بدین صورت که از تابع rand
- نرم افزار MATLAB استفاده شده و خروجی آن برای θ_i در میانگین کل داده ها و برای c_i در واریانس داده ها ضرب شده است. به ازای هر N و برای پارامترهای LMM دو مقدار دهی به صورت تصادفی به طریقی که توضیح داده شد، انجام شده است و سپس با توجه به معیاری که در رابطه (۵-۴۵) بیان شده، [۴۴] جواب بهتر انتخاب شده است.

$$I_{KL} = \sum_x p_H(x) \log\left(\frac{p_H(x)}{p(x)}\right) \qquad (۴۵-۵)$$

در این رابطه $p_H(x)$ هیستوگرام داده ها و $p(x)$ مقدار توزیع مخلوط لاپلاس به دست آمده با شرایط اولیه مذکور، به ازای x است. هر چه این معیار کمتر باشد، بیانگر این است که توزیع مخلوط لاپلاس به دست آمده، با خطای کمتری بر توزیع داده ها برازنده شده است.

همانطور که قبلا گفته شد، EM الگوریتم یک الگوریتم تکراری است که شرط توقف آن می تواند تعداد تکرار خاص یا عدم تغییر محسوس پارامترها باشد.

در این پایان نامه، شرط توقف تعداد تکرار ۴۵ در نظر گرفته شده است. این مقدار با توجه به عدم تغییرات محسوس در مقادیر پارامترهای LMM ، در آزمایشات متعددی که با Nهای متفاوت انجام شد، به دست آمد.

به عنوان مثال، جدول (۵-۱) ، نشان دهنده پارامترهای LMM و I_{KL}، برای بخش حقیقی تبدیل فوریه زمان کوتاه داده های TIMIT به ازای N=۳ و بعد از ۱۰، ۳۰، ۴۰ و ۵۰ تکرار EM الگوریتم

است. K بیانگر تعداد تکرا ر الگوریتم می باشد. قبلا گفته شد که کمتر بودن مقدار I_{KL} بدین معنی است که توزیع به دست آمده با تقریب بهتری PDF[1] داده ها را نمایش می دهد. با توجه به جدول، کم شدن مقدار I_{KL} در هر تکرار بیانگر این است که در هر مرحله تکرار الگوریتم توزیع به دست آمده

جدول (۵-۱): پارامترهای LMM و I_{KL}، برای بخش حقیقی تبدیل فوریه زمان کوتاه داده های TIMIT به ازای N=۳ و بعد از ۱۰، ۳۰، ۴۰ و ۵۰ تکرار EM الگوریتم .

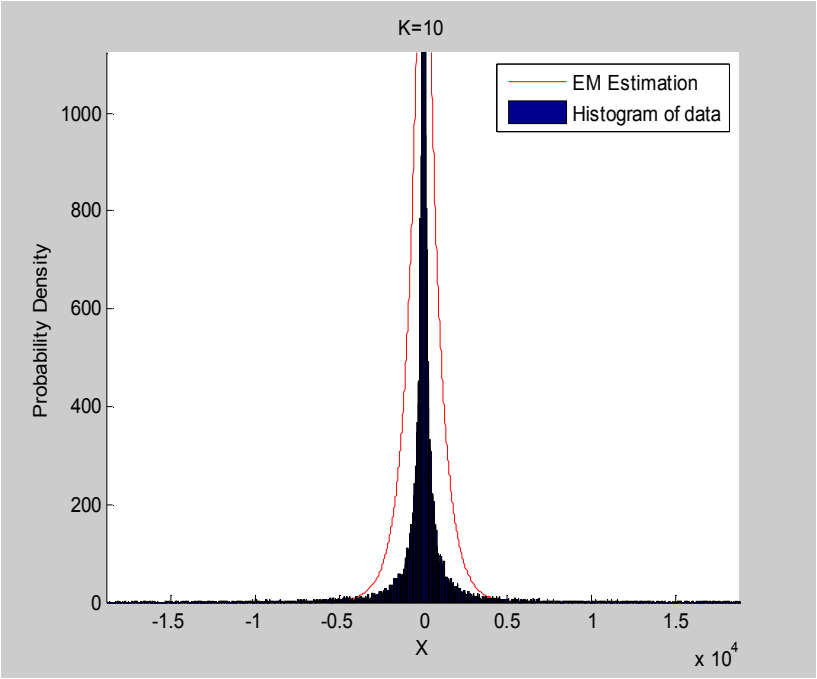
K	α_i	c_i	θ_i	I_{KL}
۱۰	۰.۰۸۲ ۰۴۶۲۱ ۰.۴۵۵۹	۰.۰۰۰۱ ۰.۰۱۱۹ ۰.۰۰۱	۴۵.۴۴۶۹ ۱.۲۸۲۳ ۸.۸۷۵۹	۳۳۶۱۴۰
۳۰	۰.۰۸۳۷ ۰ ۰.۴۲۳۱ ۰.۴۹۳۲	۰.۰۰۰۱ ۰.۰۱۴۱ ۰.۰۰۱	۱۵۱.۳۲۱ ۰.۰۶۷۷ ۳.۳۵۱۶	۳۳۵۹۵۰
۴۰	۰.۰۸۵۴ ۰.۴۱۸۷ ۰.۴۹۵۹	۰.۰۰۰۱ ۰.۰۱۴۳ ۰.۰۰۱۱	۱۶۴.۲۸۷۷ -۰.۲۳۲ ۲.۹۸۸۱	۳۳۵۹۶۰
۵۰	۰.۰۸۶۲ ۰.۴۱۶۸ ۰.۴۹۷	۰.۰۰۰۱ ۰.۰۱۴۴ ۰.۰۰۱۱	۱۶۴.۸۱۱۲ -۰.۲۴۳۶ ۲.۹۱۵۷	۳۳۵۹۶۰

بهتر بر PDF داده ها برازانده می شود و عدم تغییر محسوس آن در دو تکرار، نشان دهنده این است که توزیع های به دست آمده در دو تکرار زیاد متفاوت نیستند. همچنین با توجه به جدول، اختلاف همه پارامترهای LMM برای K =۴۰ و K =۵۰ بسیار کم است. (این اختلاف با افزایش N کمتر هم می شود) در حالیکه اختلاف این پارامترها برای K =۱۰ و K =۳۰ بیشتر است. نکته دیگر این است که با افزایش K زمان اجرای برنامه و پیچیدگی محاسباتی بیشتر می شود، به همین خاطر میزان تکرارها ۴۵ در نظر گرفته شده است.

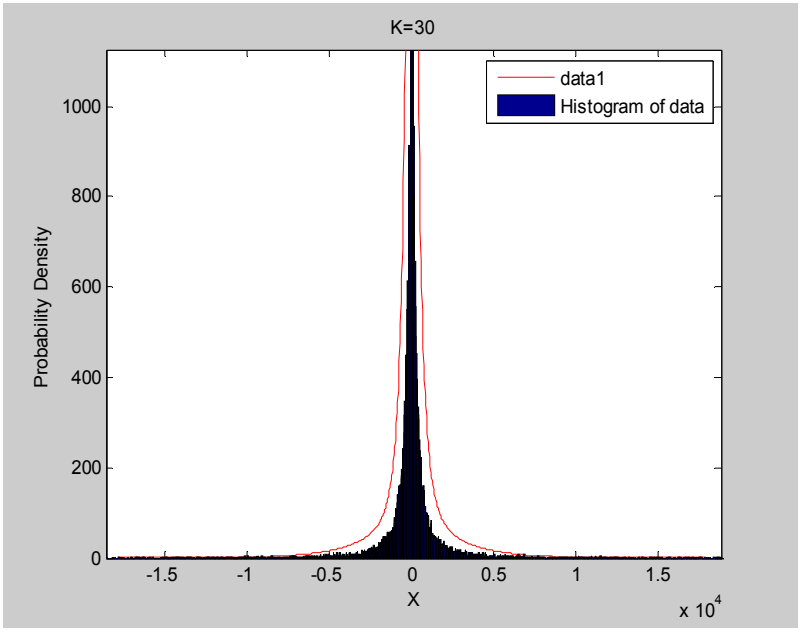
شکل های (۵-۶) نیز نتیجه EM الگوریتم را برای مثال قبل نمایش می دهند. همان طور که از نمودارها مشخص است، بین تخمین EM الگوریتم برای K =۱۰ و K =۳۰ و K =۴۰ تفاوت نسبتا زیادی وجود دارد، در حالی که اختلاف برای K =۴۰ و K =۵۰ بسیار کم است.

[1] - Probability Density Function

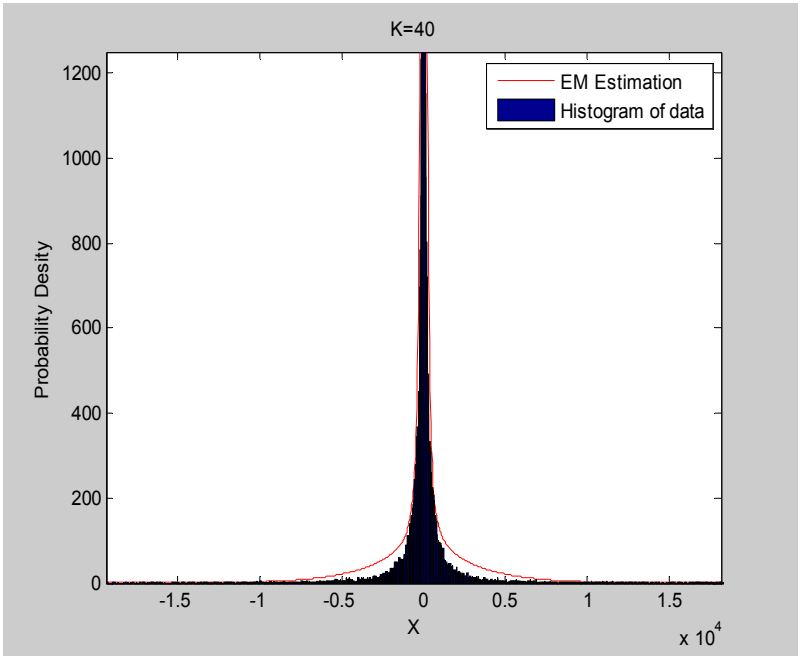
شکل های (۵-۷) و (۵-۸) نیز تخمین EM الگوریتم را برای بخش حقیقی و موهومی طیف سیگنال تمیز، به ازای N های متفاوت نشان می دهند. با توجه به نمودارها، هر چه N بزرگتر باشد، تخمین EM الگوریتم به هیستوگرام داده ها نزدیکتر است. برای اینکه با یک معیار دقیق و عددی نشان دهیم که با افزایش N منحنی تخمین زده شده بهتر بر PDF داده ها برازانده می شود، از معیار I_{KL} که قبلا معرفی شد، استفاده نمودیم. مشاهده می شود که با افزایش N این مقدار کم می شود. مثلا برای N=۱ مقدار این معیار برابر ۳۵۵۱۴۰ و برای N=۱۰ این معیار برابر با ۳۳۵۵۲۰ است. قبلا ذکر شد که



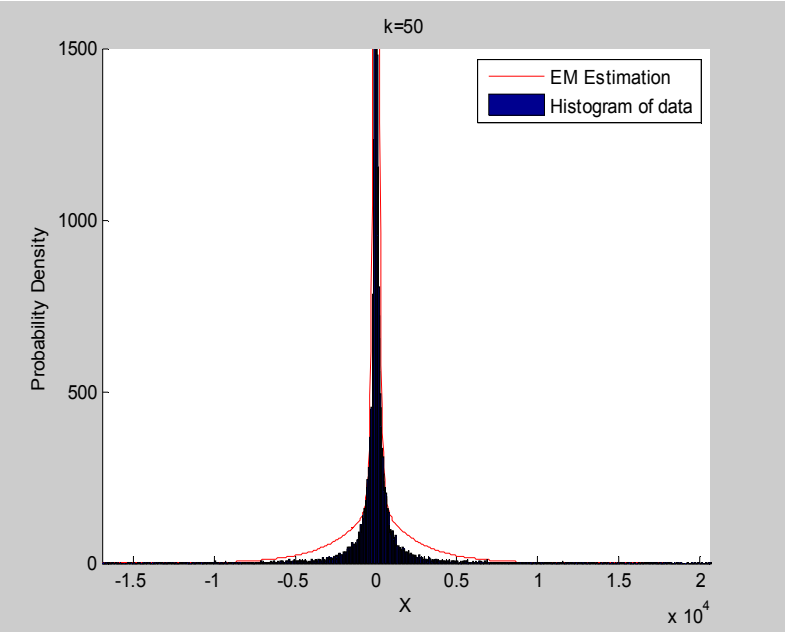
(الف)


K=30
data1
Histogram of data
Probability Density
1000
800
600
400
200
0
-1.5
-1
-0.5
0
0.5
1
1.5
X
x 10^4

(ب)

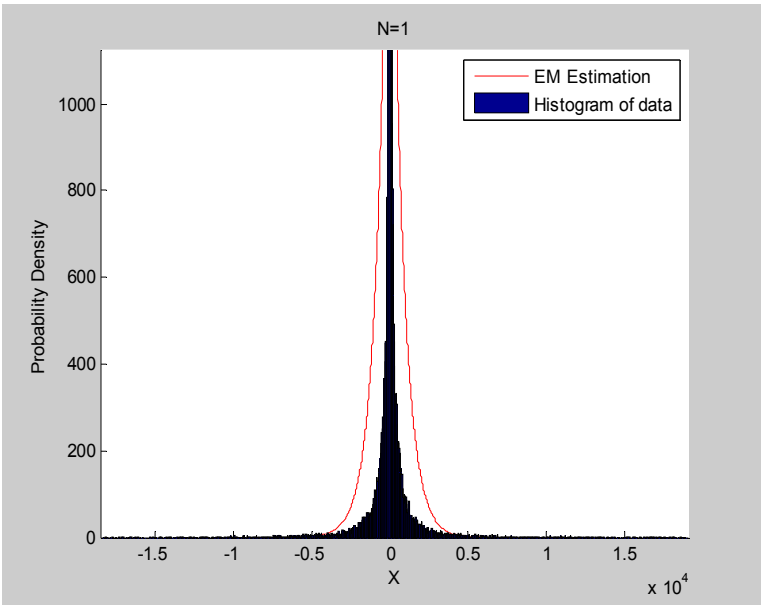

K=40
EM Estimation
Histogram of data
Probability Desity
1200
1000
800
600
400
200
0
-1.5
-1
-0.5
0
0.5
1
1.5
X
x 10^4

(ج)

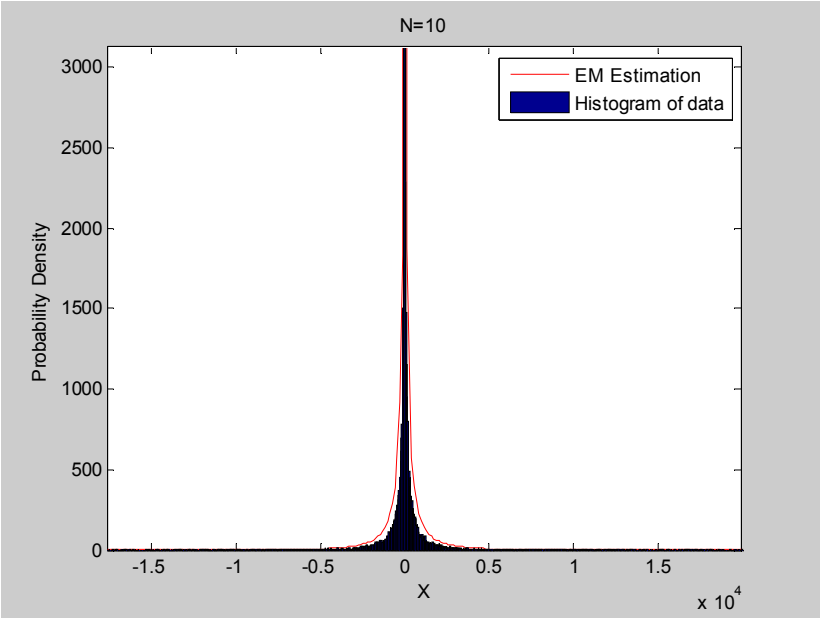


(چ)

شکل (۵-۶): تخمین EM الگوریتم برای بخش حقیقی طیف سیگنال تمیز با N=۳، (الف) پس از ۱۰ بار تکرار، (ب) پس از ۳۰ بار تکرار، (ج) پس از ۴۰ بار تکرار، (چ) پس از ۵۰ بار تکرار .

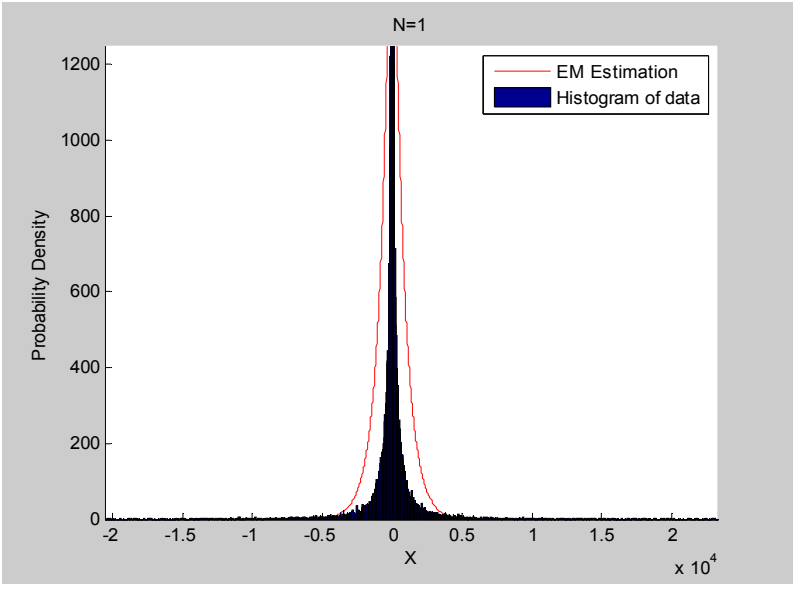


(الف)

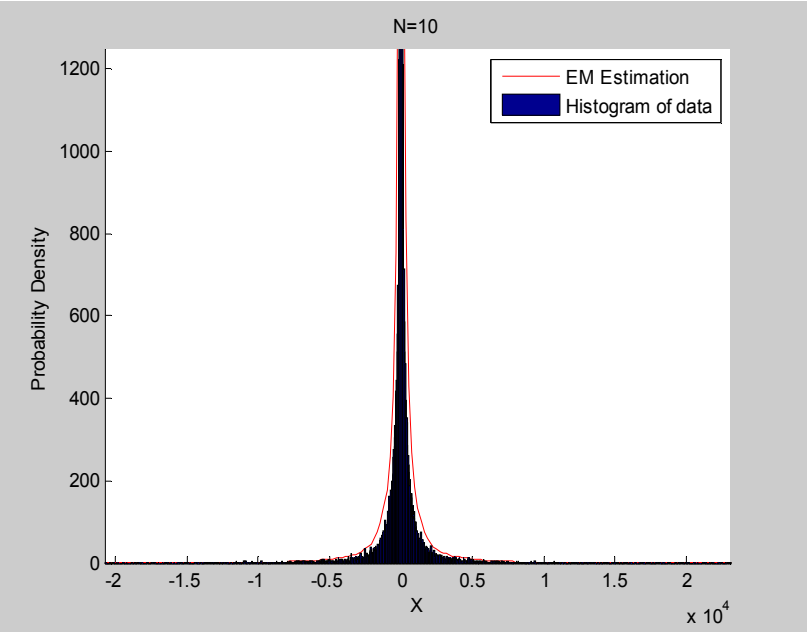


(ب)

شکل (۵-۷): تخمین EM الگوریتم برای بخش حقیقی طیف سیگنال تمیز: (الف) به ازای یک لاپلاس، (ب) به ازای ده لاپلاس .



(الف)



(ب)

شکل (۵-۸): تخمین EM الگوریتم برای بخش موهومی طیف سیگنال تمیز : (الف) به ازای یک لاپلاس، (ب) به ازای ده لاپلاس .

کمتر بودن این معیار به معنای بهتر بودن تقریب است که نشان می دهد N=۱۰ نسبت به N=۱ تقریب مناسب تری از سیگنال ارائه می دهد. نکته دیگر این است که با افزایش N تغییرات این معیار کم می شود. یعنی اختلاف این معیار به ازای دو N متفاوت کم و کمتر می شود. این نشان می دهد که با افزایش N اختلاف بین توزیع های به دست آمده از EM الگوریتم کم می شود. به عنوان مثال مقدار این معیار برای N=۱ برابر ۳۵۵۱۴۰، برای N=۱۰ برابر ۳۳۵۵۲۰ و برای N=۲۰ برابر ۳۲۴۷۰ است. واضح است اختلاف این معیار برای دو حالت N=۱۰ و N=۲۰ کمتر از اختلاف این معیار برای دو حالت N=۱ و N=۱۰ می باشد.

۵-۵- تخمین نویز

همان طور که در رابطه تخمین MMSE مشاهده شد، برای به دست آوردن تخمینی از سیگنال تمیز، علاوه بر پارامترهای LMM، به واریانس نویز نیز نیاز است. با توجه به اینکه فقط سیگنال

نویزی و واریانس آن در دسترس است، لازم است که واریانس نویز به طریقی از آن استخراج گردد. روش های تخمین نویز در فصل های قبلی توضیح داده شد. در این روش ها فرض می‌شود که نویز غیرایستان است. ولی تغییرات توان آن به آرامی و کندتر از توان سیگنال صحبت تمیز صورت می پذیرد.

برای تخمین نویز از روش Continuos Spectral Minimum Tracking (ردیابی‌کمینه ها، روش دوم) که در فصل قبل توضیح داده شد، استفاده گردیده است. همانطور که در فصل قبل گفته شد، در این روش مینیمم توان سیگنال نویزی به صورت پیوسته و بدون نیاز به پنجره دنبال می شود. در این روش تخمین نویز، با مقایسه طیف هموار شده سیگنال نویزی در هر فرکانس با تخمین نویز در همان فرکانس و در قاب قبلی، با یک رابطه هموارسازی غیرخطی، به روز می شود. در این روش برای دنبال کردن مینیمم توان سیگنال نویزی به جای پریودوگرام سیگنال نویزی، از پریودوگرام هموار شده آن استفاده می شود.

$$P(m,k) = \alpha p(m-1,k) + (1-\alpha)|Y(m,k)|^2 \qquad (۴۶-۵)$$

که در آن α فاکتور هموارسازی و در بازه $0.7 \le \alpha \le 0.9$ و k فرکانس و m شماره قاب می باشد.

رابطه غیر خطی که برای تخمین نویز بر پایه دنبال کردن مینیمم طیف توان سیگنال نویزی، در هر فرکانس استفاده می شود، به صورت زیر است:

If $p_{min}(m-1,k) < p(m,k)$

$$p_{min}(m,k) = \lambda p_{min}(m-1,k) + \frac{1-\lambda}{1-\beta}(p(m,k) - \beta\, p(m-1,k))$$

Else

$$p_{min}(m,k) = p(m,k)$$

End $\qquad (۴۷-۵)$

که در آن $p_{min}(m,k)$ تخمین نویز و $p(m,k)$ پریودوگرام هموار شده سیگنال نویزی می باشد.

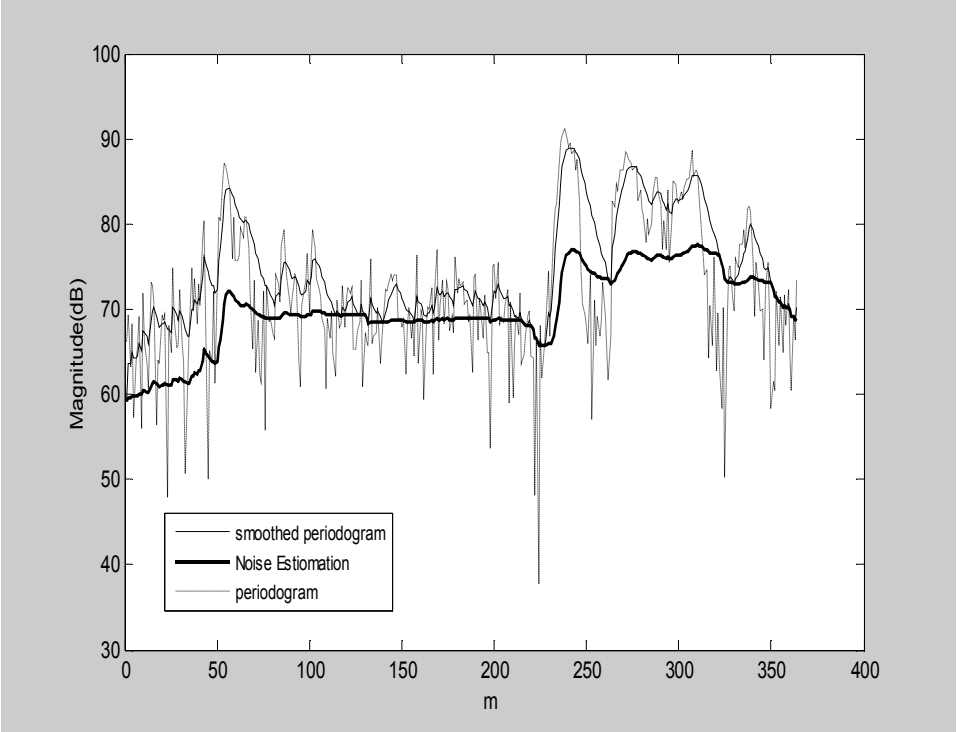
برای پیاده سازی این روش با توجه به نتایج روش پیشنهادی، مقادیر زیر را پیشنهاد شده است: (این مقادیر بهترین جواب ها را در بخش ارزیابی روش پیشنهادی داشته است.)

$\alpha = ۰.۷$ و $\beta = ۰.۹۶$ $\lambda = ۰.۹۹۹$

شکل (۵-۹) نمایش دهنده پریودوگرام سیگنال نویزی که با نویز F۱۶ و در سیگنال به نویز ۵ دسی بل, آغشته شده است، پریودوگرام هموار شده و نویز تخمین زده شده به روش فوق، در فرکانس k=۱۰، برای کلیه قاب ها می باشد.

۵-۶- ارزیابی روش پیشنهادی

پس از پیاده سازی روش پیشنهادی، برای بررسی و سنجش عملکرد روش، باید آن را با معیارهای مشخص ارزیابی کرد. در این پایان نامه، برای ارزیابی روش پیشنهادی ده جمله از پایگاه داده TIMIT که به وسیله پنج زن و پنج مرد گفته شده بودند، انتخاب شده و در آزمایش های مختلف، سه نویز سفید گوسی، جنگنده F۱۶ و همهمه در سه سطح سیگنال به نویز ۰، ۵ و ۱۰ به سیگنال اضافه گردید. (نویز سفید گوسی به وسیله نرم افزار MATLAB تولید شده است) سپس سیگنال نویزی به قاب هایی به طول ۳۲ میلی ثانیه با همپوشانی ۵۰٪ تقسیم و در پنجره همینگ ضرب شد و از آن تبدیل فوریه زمان کوتاه گرفته شد. لازم به ذکر است که علاوه بر پنجره همینگ به طول ۳۲



شکل (۵-۹): خط چین: پریودوگرام سیگنال نویزی که با نویز F۱۶ و در سیگنال به نویز ۵ دسی بل, آغشته شده است.
خط یکپارچه نازک: پریودوگرام هموارشده و خط یکپارچه کلفت: تخمین نویز .

میلی ثانیه، از پنجره هنینگ به طول ۳۲ میلی ثانیه و پنجره همینگ به طول ۱۶ میلی ثانیه نیز برای پیاده سازی روش پیشنهادی استفاده شد که سیگنال بهسازی شده با آنها بر حسب معیارهای ارزیابی که بعدا توضیح داده می شوند، نسبت به همینگ به طول ۳۲ میلی ثانیه قابل قبول نبودند. چون نتایج بهسازی گفتار با آنها خوب نبودند و فقط برای برخی از نویزها و سیگنال به نویزها سنجیده شده بودند، در این بخش از آوردن آنها خودداری شده است. پس از اعمال روش پیشنهادی بر روی طیف زمان کوتاه سیگنال نویزی، از آن تبدیل فوریه زمان کوتاه معکوس گرفته شده و سیگنال بهسازی شده به حوزه زمان انتقال داده می شود. ولی سیگنال به دست آمده به صورت قاب هایی است که باهم همپوشانی دارند. با اعمال روش همپوشانی-جمع بر روی آن، سیگنال بهسازی شده نهایی به دست می آید. حال، برا ی ارزیابی عینی عملکرد روش پیشنهادی، از سه معیار سیگنال به نویز قطعه ای ، PESQ و LLR(Log Likelihood ratio) که در فصل دو توضیح داده شد، استفاده می شود. علت انتخاب سیگنال به نویز قطعه ای به عنوان معیارهای ارزیابی این بود که سیگنال به نویز پارامتر مهمی در نشان دادن موفقیت روش در حذف نویز است به همین جهت در بیشتر روش هایی که در زمینه بهسازی گفتار ارائه می شود، از آن به عنوان معیار ارزیابی استفاده می شود. ولی این معیار همبستگی پایینی (حدود۰.۳۱) با معیار های ارزیابی ذهنی دارد [۳۱].

همچنین در این پایان نامه استفاده از معیارهای ذهنی یا تست های شنیداری مانند MOS مقدور نبود. زیرا، در این معیارها فرد به سیگنال بهسازی شده گوش داده و به آن امتیازی در بازه مشخص می دهد. استاندارد تست های شنوایی این است که سیگنال صوتی که فرد به آن گوش می دهد و به آن امتیاز می دهد، باید به زبان مادری آن فرد گفته شده باشد. از طرفی، جملات گفته شده در پایگاه داده استفاده شده در این پایان نامه، به زبان انگلیسی است و چون پیدا کردن افرادی که زبان مادریشان انگلیسی باشد، مشکل بود، از دو معیار عینی استفاده شد که همبستگی بالایی با معیارهای

ذهنی داشته باشند. این دو معیار عبارتند از: PESQ و LLR که بیانگر کیفیت سیگنال بهسازی شده هستند و همبستگی بالایی با معیار های ارزیابی ذهنی دارند [۳۱]. همچنین در مقالات معتبر بسیاری از این دو معیار جهت ارزیابی سیگنال بهسازی شده، استفاده شده است.

در ابتدا نتیجه سه معیار ارزیابی، با نویزهای ذکر شده و در سه سطح سیگنال به نویز بیان شده، به ازای تغییر تعداد اجزای LMM که با N نشان داده شده است و سپس با توجه به مقادیر به دست آمده از این آزمایشات، مقدار مناسب N انتخاب گردیده است.

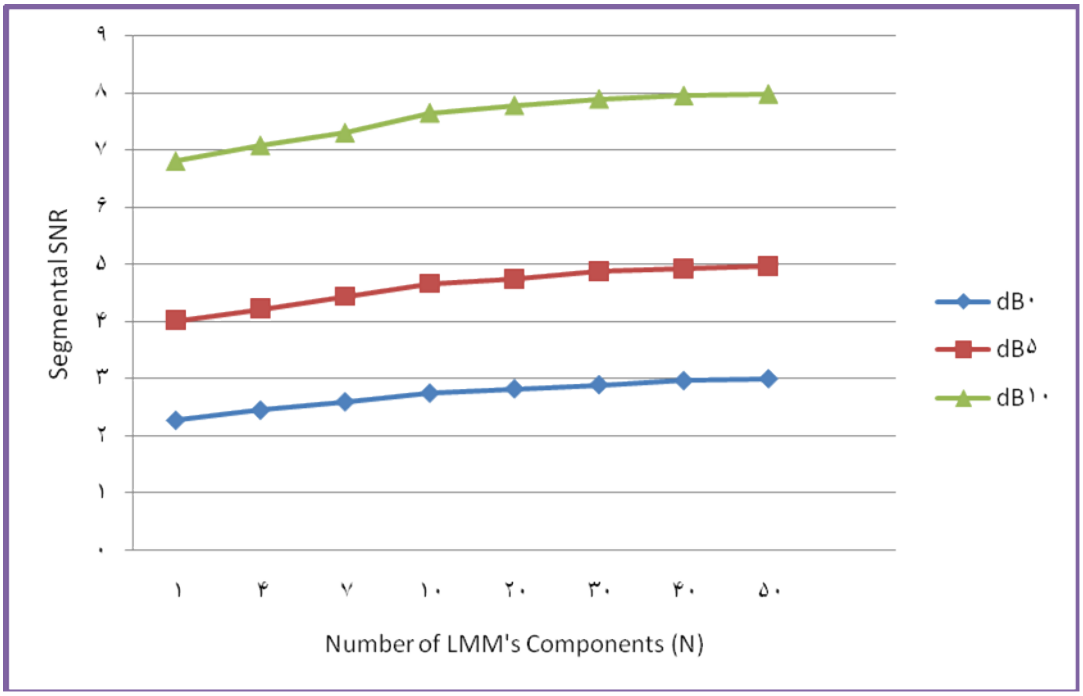
۵-۶-۱- معیار ارزیابی سیگنال به نویز قطعه ای

اولین معیاری که مورد بررسی قرار داده شده است، سیگنال به نویز قطعه ای می باشد. مقادیر این معیار هم در نمودار و هم در جدول ذکر شده است. در سه شکل (۵-۱۰)، (۵-۱۱) و (۵-۱۲) نمودارهای سیگنال به نویز قطعه ای برای سه نویز سفید، همهمه و F۱۶ در سه سطح سیگنال به نویز ۰، ۵ و ۱۰ دسی بل به ازای مقادیر مختلف تعداد اجزای مخلوط لاپلاس یعنی N رسم شده است. مقادیر آستانه برای محاسبه سیگنال به نویز قطعه‌ای ۱۰dB- و ۳۵dB در نظر گرفته شده است [۳۱]. لازم به ذکر است که در این حالت محاسبه سیگنال به نویز قطعه ای ورودی نیز الزامی است. زیرا به ازای یک سیگنال به نویز ورودی، اگر سیگنال به نویزهای قطعه ای ورودی متفاوت باشد، سیگنال به نویزهای قطعه ای خروجی نیز متفاوت خواهند بود. این مطلب درباره دو معیار دیگر نیز صادق است. یعنی در یک سیگنال به نویز معین، هر چه سیگنال به نویز قطعه ای ورودی بیشتر باشد، نتایج به ازای این دو معیار ارزیابی نیز بهتر خواهند بود.

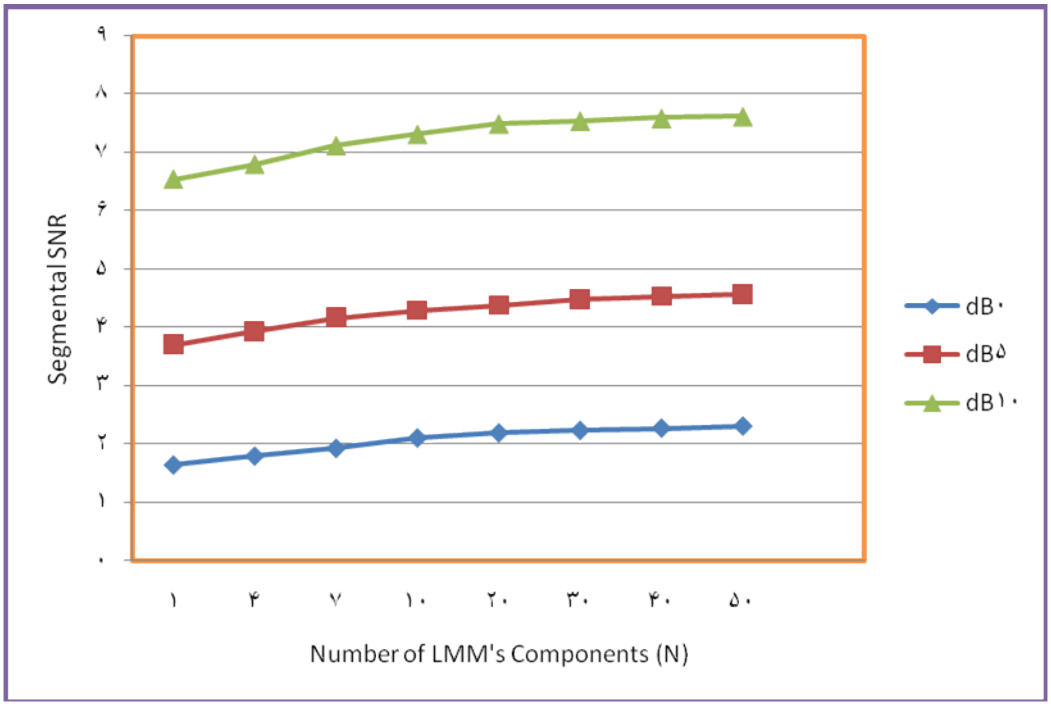
با بررسی این سه نمودار مشخص می شود که میزان سیگنال به نویز قطعه ای سیگنال بهسازی شده با افزایش N، افزایش می یابد، یعنی با افزایش N میزان حذف نویز روش پیشنهادی بیشتر می شود، که منطقی به نظر می رسد. زیرا هر چه N بزرگتر شود مدلی که برای سیگنال در نظر گرفته شده است، با دقت بیشتری به توزیع واقعی سیگنال برازنده می شود. مثلا برای حالتی که طیف سیگنال

تمیز با مخلوطی از N =۱ یا N =۴ لاپلاس تقریب زده می شود، نتایج خوب نیستند. زیرا این دو حالت تقریب خوبی از توزیع سیگنال نمی باشند.

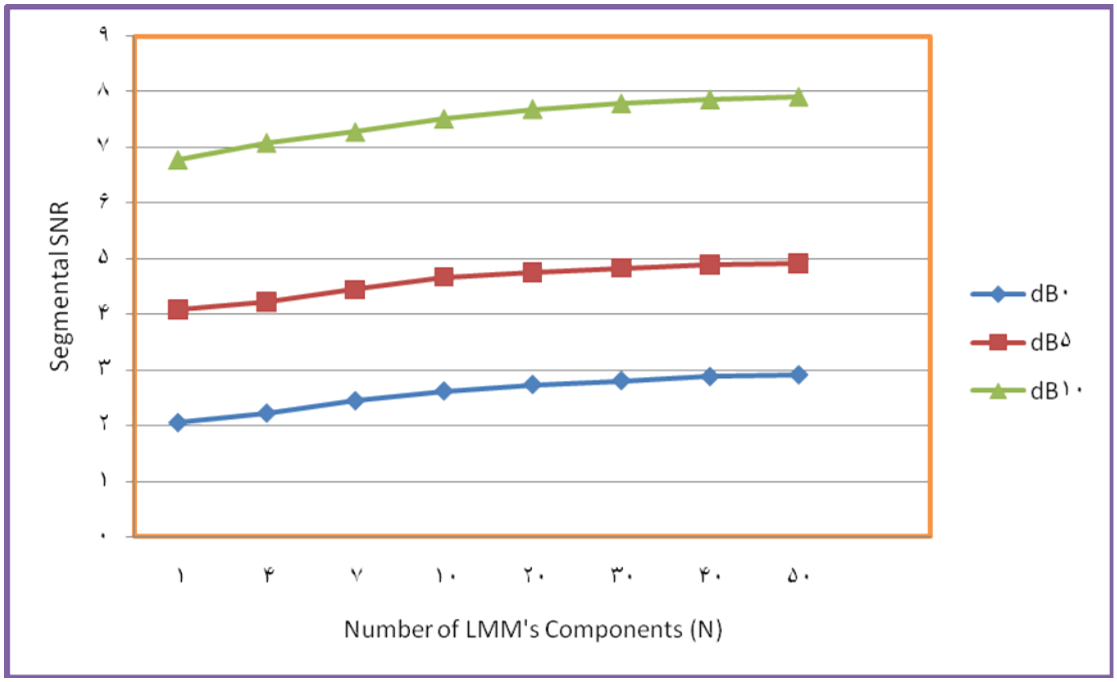
نکته دیگری که از نمودارها مشخص می شود، این است که به ازای تغییر N، وقتی Nها کوچک هستند، اختلاف مقادیر سیگنال به نویز قطعه ای متناظر با آنها، برای هر سه نویز و سیگنال به نویز بیشتر است. علت این مطلب این است که به ازای مقادیر کوچک N، با تغییرات N، تقریب های متناظر با این N ها از سیگنال، با هم اختلاف زیادی دارند (این مطلب در بخش پیاده سازی EM الگوریتم بیان شد)، ولی با زیاد شدن N، همان طور که قبلا بیان شد، اختلاف بین تقریب ها کمتر می‌شود، و در نتیجه اختلاف نتایج نیز کم و کمتر می‌شود. با توجه به نمودارها، N =۳۰ پاسخ خوبی داشته و پس از آن به ازای تغییرات N، نتایج تغییرات زیادی ندارند. این نشان می دهد که تقریب توزیع سیگنال تمیز با توزیع LMM از حدود N =۳۰ به بعد، به واقعیت نزدیک و نزدیک تر می‌شود. منطقی است که با افزایش N نتایج بهتر شود. ولی پیچیدگی محاسباتی نیز به میزان قابل توجهی افزایش یافته و اجرای الگوریتم بسیار زمان بر می شود. به همین دلیل افزایش N در ۵۰ متوقف شده است. زیرا با افزایش N نتایج تغییر چشم گیری نداشتند در حالی که افزایش N به خاطر زمان بر شدن و حجم حافظه اشغالی مقرون به صرفه نبود.



شکل (۵-۱۰): نمودار سیگنال به نویز قطعه ای سیگنال بهسازی شده، پس از اعمال روش پیشنهادی روی سیگنال آغشته به نویز سفید .



شکل (۵-۱۱): نمودار سیگنال به نویز قطعه ای سیگنال بهسازی شده،پس از اعمال روش پیشنهادی روی سیگنال آغشته به نویز F۱۶ .



شکل (۵-۱۲): نمودار سیگنال به نویز قطعه ای سیگنال بهسازی شده، پس از اعمال روش پیشنهادی روی سیگنال آغشته به نویز همهمه .

جدول (۵-۲) : مقایسه سیگنال به نویزهای قطعه ای سیگنال بهسازی شده، برای سه نویز همهمه، سفید و F۱۶، در سه سیگنال به نویز ۰، ۵، ۱۰ و N های متفاوت .

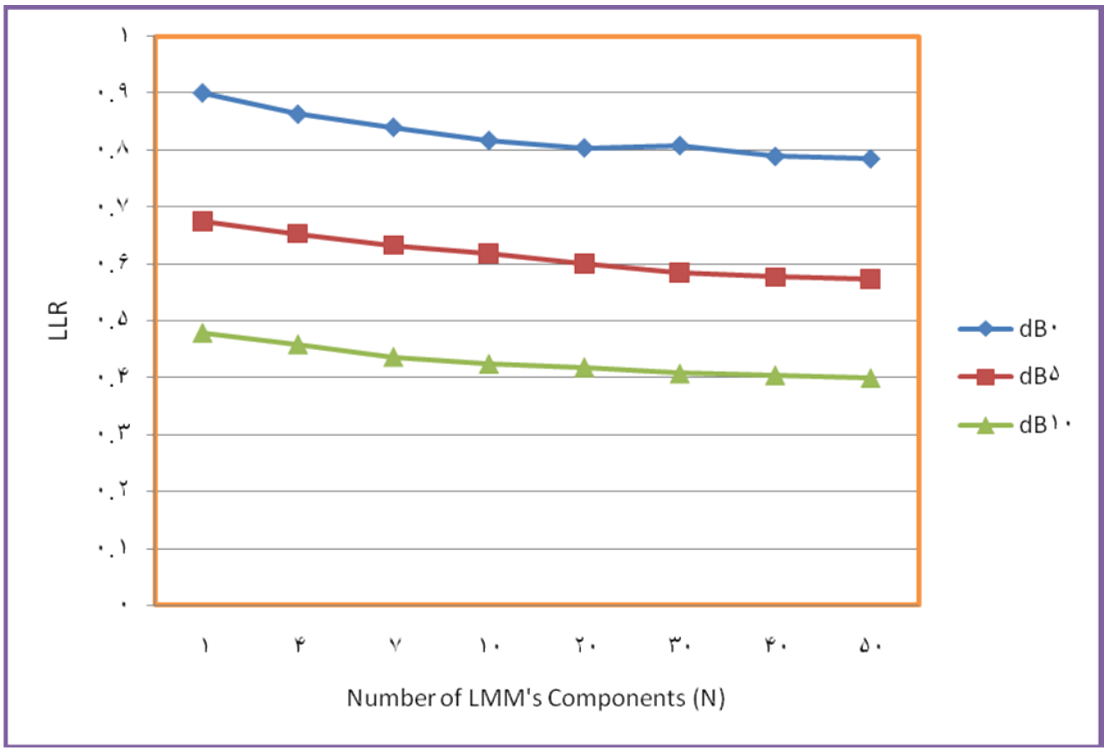
Noise type	White noise			F۱۶ noise			Babble noise		
	-۱۲.۱/۰dB	-۷.۴۵/۵dB	-۲.۲۲/۱۰dB	-۱۲.۴/۰dB	-۷.۸۳/۵dB	-۲.۳/۱۰dB	-۱۱.۹/۰dB	-۷.۱/۵dB	-۲.۰۱/۱۰dB
N=۱	۲.۲۷۹	۴.۰۲۵	۶.۸۲	۱.۶۴	۳.۷۰۲	۶.۵۴۱	۲.۰۵۵	۴.۰۸	۶.۷۸۳
N=۴	۲.۴۵۶	۴.۲۳۶	۷.۰۹۵	۱.۷۹۳	۳.۹۳۲	۶.۸	۲.۲۲۵	۴.۲۱۶	۷.۰۸۹
N=۷	۲.۶۰۴	۴.۴۵۱	۷.۳۱۸	۱.۹۲۶	۴.۱۶۹	۷.۱۲۱	۲.۴۴۷	۴.۴۴۲	۷.۲۸۲
N=۱۰	۲.۷۵۸	۴.۶۷۴	۷.۶۶۵	۲.۱۰۵	۴.۲۸۶	۷.۳۱۵	۲.۶۲۲	۴.۶۶۸	۷.۵۱۷
N=۲۰	۲.۸۳۲	۴.۷۵۵	۷.۷۹۴	۲.۱۹۲	۴.۳۷۹	۷.۴۹۸	۲.۷۳۸	۴.۷۴۴	۷.۶۹۲
N=۳۰	۲.۹۰۴	۴.۸۸۱	۷.۹۰۶	۲.۲۳۵	۴.۴۷۴	۷.۵۴۶	۲.۸۰۷	۴.۸۲۷	۷.۷۹۳
N=۴۰	۲.۹۷۶	۴.۹۲۷	۷.۹۶۴	۲.۲۷۵	۴.۵۳	۷.۵۹۱	۲.۸۷۹	۴.۸۹۱	۷.۸۶۲
N=۵۰	۳.۰۱	۴.۹۸۱	۷.۹۹۳	۲.۳۱	۴.۵۷۱	۷.۶۲۳	۲.۹۱۲	۴.۹۱۵	۷.۹۱۲

نکته دیگری که می‌توان در این سه نمودار ملاحظه کرد، این است که با تغییر N بیشترین تغییرات سیگنال به نویز قطعه ای برای هر سه نویز، مربوط به حالت سیگنال به نویز ۱۰dB و سپس ۵dB و پس از آن ۰dB است. این امر بیانگر این است که با افزایش سیگنال به نویز عملکرد روش پیشنهادی بهتر می‌شود. بهترین جواب ها در سه سیگنال به نویز ۰ و ۵ و ۱۰ مربوط به نویز سفید است. پس از آن نویز همهمه و در نهایت نویز F۱۶ قرار می‌گیرد. با مشاهده و بررسی طیف توان این سه نویز مشخص شد که تغییرات توان نویز سفید از دو نویز دیگر کمتر است. این تغییرات برای نویز همهمه کمتر از نویز F۱۶ می‌باشد. همچنین در بخش تخمین نویز بیان شد که فرض اساسی روش‌های تخمین نویز، کند بودن تغییرات توان نویز است. احتمالا به این دلیل نتایج برای نویز سفید از دو نویز دیگر بهتراست و نویز همهمه نیز نتایج بهتری از نویز F۱۶ دارد.

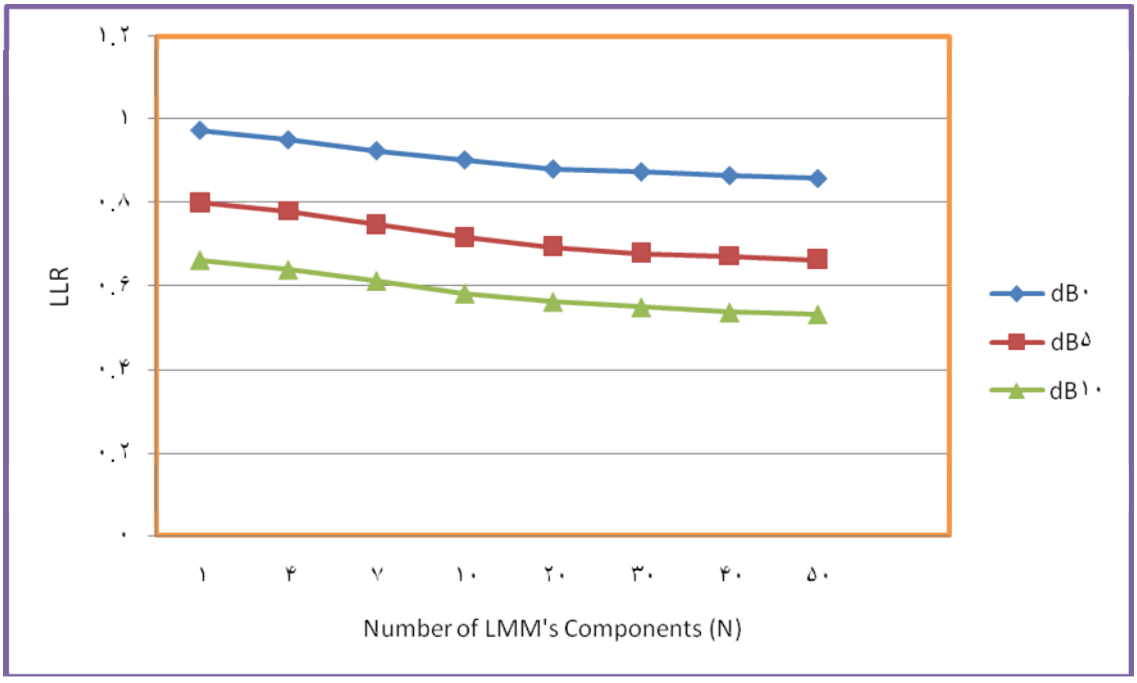
برای اینکه مقادیر رسم شده در نمودارهای بالا به طور دقیق تر مشاهد شود، در جدول (۵-۲) نیز نمایش داده شده اند.

۵-۶-۲- معیار ارزیابی LLR

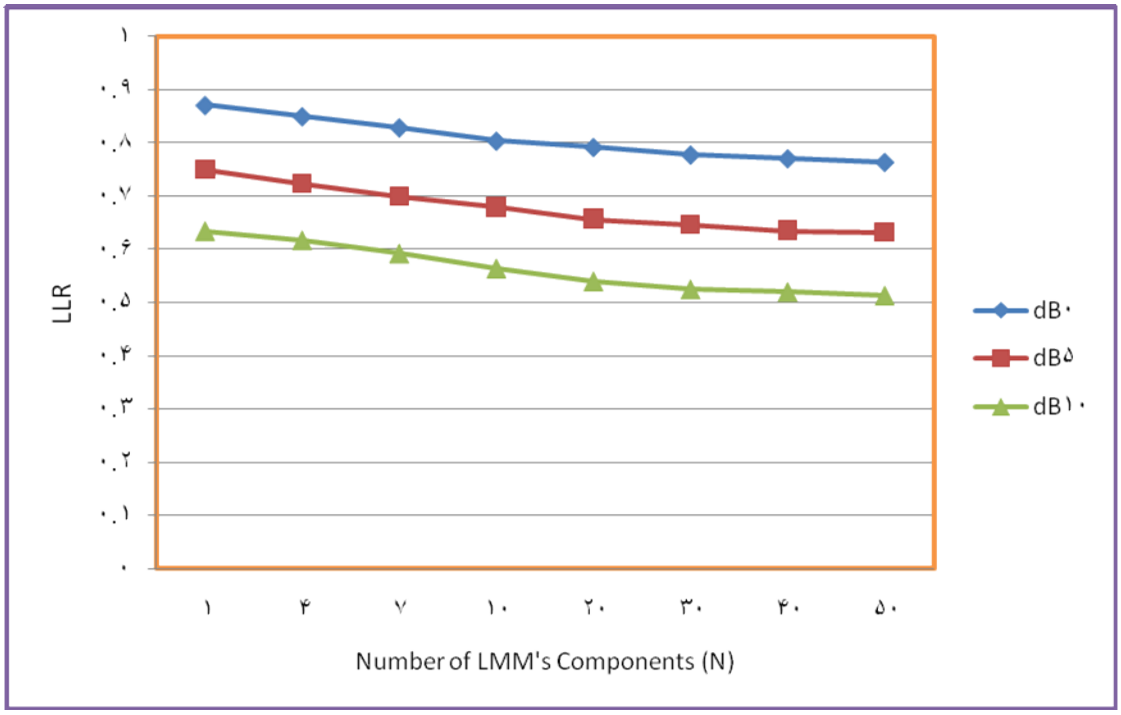
پس از بررسی معیار سیگنال به نویز قطعه ای، معیار LLR برای سه نویز مذکور و سه سطح سیگنال به نویز قبل بررسی شده است. مقادیر معیار LLR برای این سه نویز و سه سطح سیگنال به نویز، در نمودارهای (۵-۱۳)، (۵-۱۴) و (۵-۱۵) رسم شده است. قبلا بیان شد که کمتر بودن مقدار LLR به معنی بهتر بودن کیفیت سیگنال خروجی است. از نمودارها مشخص است که مقدار LLR با افزایش N کم می شود، یعنی کیفیت سیگنال خروجی با افزایش N زیاد می شود که همان نتیجه ای است که در بخش قبل نیز به دست آمد. در این بخش نیز بهترین جواب ها مربوط به حذف نویز سفید و سپس همهمه می باشد. به ازای یک N مشخص، در هر سه نویز، کمترین LLR مربوط به سیگنال به نویز ۱۰ است، که نشان می دهد در سیگنال به نویزهای بالاتر، که دامنه سیگنال خیلی از دامنه نویز بزرگتر است، کیفیت سیگنال خروجی بالاتر است. در این بخش نیز مشاهده می شود که مانند بخش قبل، تغییرات نتایج در ابتدای نمودارها که مربوط به مقادیر پایین N است، زیاد می باشد. اما این



شکل (۵-۱۳): نمودار LLR برای سیگنال بهسازی شده، پس از اعمال روش پیشنهادی روی سیگنال آغشته به نویز سفید .



شکل (۵-۱۴) : نمودار LLR برای سیگنال بهسازی شده، پس از اعمال روش پیشنهادی روی سیگنال آغشته به نویز F۱۶ .



شکل (۵-۱۵) : نمودار LLR برای سیگنال بهسازی شده، پس از اعمال روش پیشنهادی روی سیگنال آغشته به نویز همهمه .

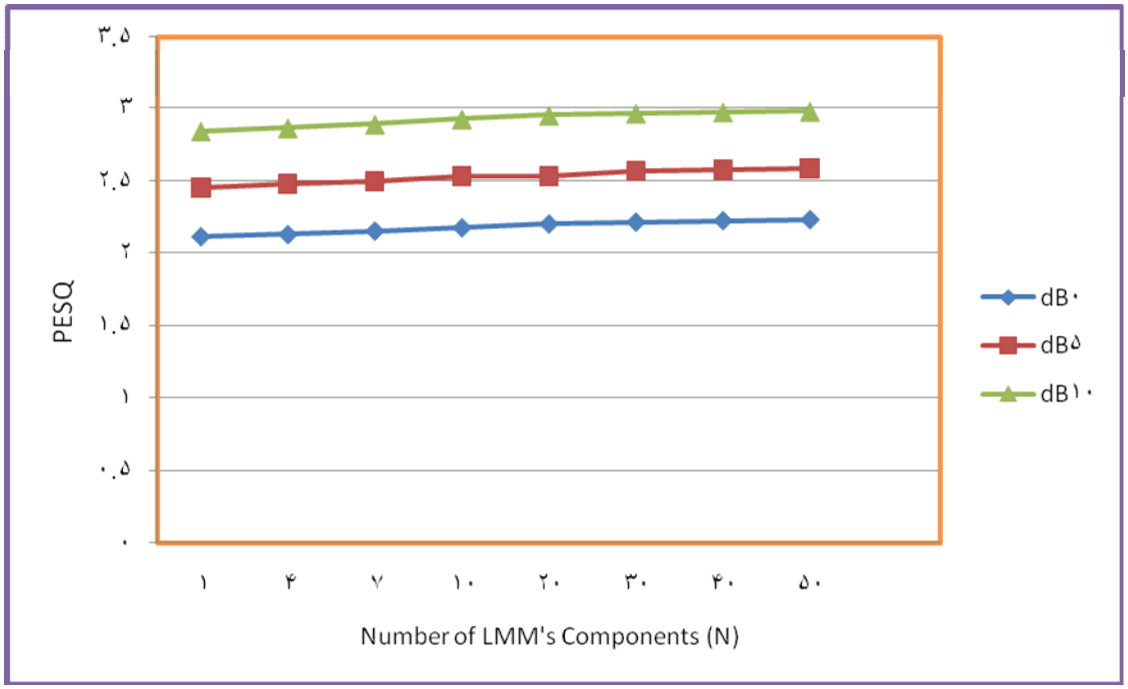
جدول(۵-۳) : مقایسه LLR سیگنال بهسازی شده، برای سه نویز همهمه، سفید و F۱۶ در سه سیگنال به نویز ۰، ۵، ۱۰ و N های متفاوت .

Noise type	White noise			F۱۶ noise			Babble noise		
	۰dB	۵dB	۱۰dB	۰dB	۵dB	۱۰dB	۰dB	۵dB	۱۰dB
N=۱	۰.۹	۰.۶۷۵	۰.۴۷۹	۰.۹۹۳	۰.۷۸۲	۰.۶۹۱	۰.۸۹۱	۰.۷۳	۰.۶۳۴
N=۴	۰.۸۶۳	۰.۶۵۳	۰.۴۵۹	۰.۹۷	۰.۷۵۸	۰.۶۶۸	۰.۸۷	۰.۷۰۴	۰.۶۱۷
N=۷	۰.۸۴	۰.۶۳۳	۰.۴۳۷	۰.۹۴۴	۰.۷۲۷	۰.۶۴۲	۰.۸۴۹	۰.۶۸۳	۰.۵۹۲
N=۱۰	۰.۸۱۷	۰.۶۱۸	۰.۴۲۵	۰.۹۲۲	۰.۶۹۶	۰.۶۱	۰.۸۲۴	۰.۶۵۹	۰.۵۶۴
N=۲۰	۰.۸۰۴	۰.۶۰۱	۰.۴۱۹	۰.۹	۰.۶۷۴	۰.۵۹۲	۰.۸۱۸	۰.۶۳۷	۰.۵۴
N=۳۰	۰.۸۰۸	۰.۵۸۵	۰.۴۰۸	۰.۸۸۳	۰.۶۵۹	۰.۵۷۹	۰.۸۰۹	۰.۶۲۶	۰.۵۲۵
N=۴۰	۰.۷۸۹	۰.۵۷۷	۰.۴۰۵	۰.۸۷۲	۰.۶۵۱	۰.۵۶۷	۰.۸	۰.۶۱۵	۰.۵۲
N=۵۰	۰.۷۸۵	۰.۵۷۳	۰.۴	۰.۸۶۱	۰.۶۴۳	۰.۵۶۲	۰.۷۹۴	۰.۶۱۲	۰.۵۱۳

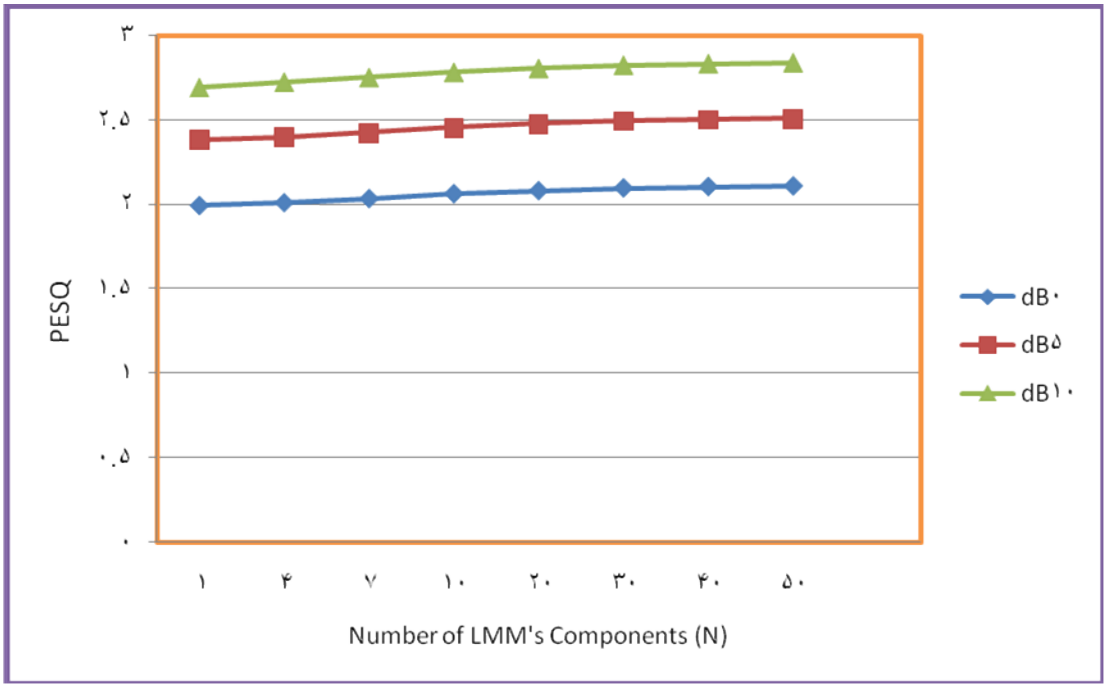
تغییرات در انتهای نمودار و از حدود N = ۳۰ به بالا بسیار کم می شود که علت این امر در قسمت قبل توضیح داده شد. به همین خاطر افزایش N در N = ۵۰ متوقف شده است. جدول (۵-۳) نیز مقادیر عددی LLR را برای سه نویز مذکور، در سه سیگنال به نویز آورده شده، نشان می دهد.

۵-۶-۳- معیار ارزیابی PESQ

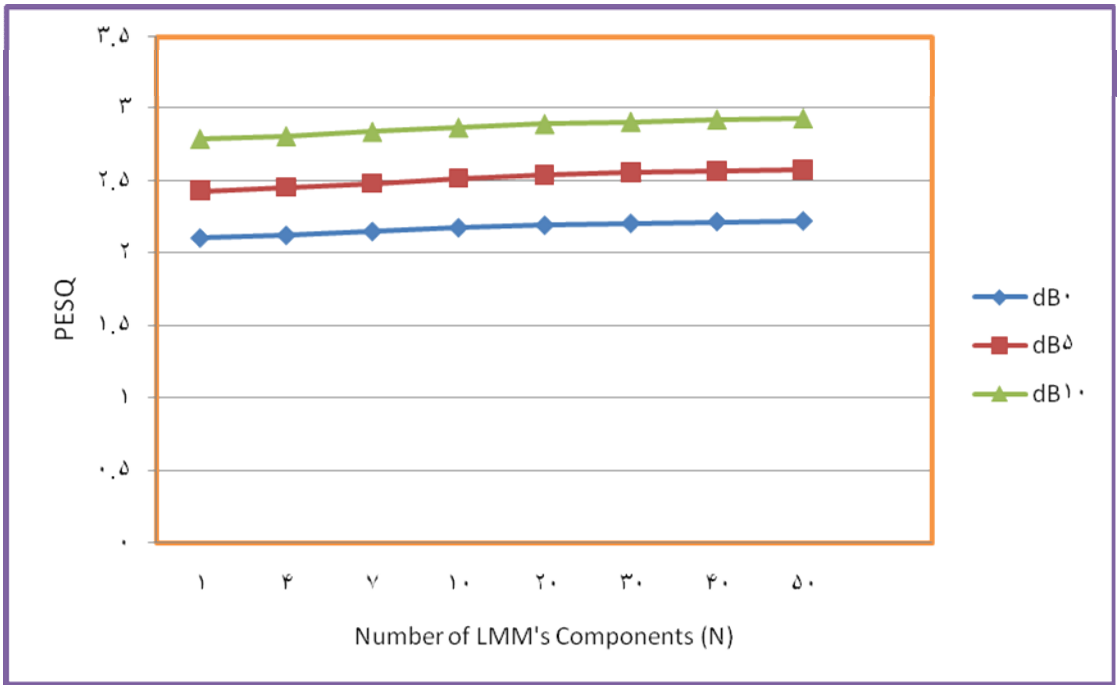
همانطور که در فصل های قبل گفته شده PESQ برای پیش بینی میزان امتیاز معیار MOS طراحی شده و عددی بین ۵. - تا ۴.۵ را می دهد [۱۰]. هر چه این عدد بزرگتر باشد بیانگر کیفیت بالاتر سیگنال صوتی است. در نمودارهای (۵-۱۶)، (۵-۱۷) و (۵-۱۸) مقادیر این معیار برای سه نویز قبل و سه سیگنال به نویز مذکور آورده شده است. با توجه به نمودارها مشخص می شود که با افزایش N مقادیر PESQ افزایش یافته و در واقع نتایج بهبود می یابند. بالاترین مقادیر PESQ برای هر سه نویز، در سیگنال به نویز ۱۰ اتفاق افتاده است. بر اساس این معیار نیز بهترین نتایج برای نویز سفید و پس از آن همهمه و F۱۶ است. روند تغییرات نمودارها نیز مانند نمودارهای دو بخش قبل است. یعنی در



شکل (۵-۱۶): نمودار PESQ برای سیگنال بهسازی شده، پس از اعمال روش پیشنهادی روی سیگنال آغشته به نویز سفید .



شکل (۵-۱۷): نمودار PESQ برای سیگنال بهسازی شده، پس از اعمال روش پیشنهادی روی سیگنال آغشته به نویزF۱۶ .



شکل (۵-۱۸): نمودار PESQ برای سیگنال بهسازی شده، پس از اعمال روش پیشنهادی روی سیگنال آغشته به نویز همهمه .

جدول(۵-۴): مقایسه PESQ سیگنال بهسازی شده، برای سه نویز همهمه، سفید و F۱۶ در سه سیگنال به نویز ۰، ۵، ۱۰ و N های متفاوت.

Noise Type	white noise			F۱۶ noise			Babble noise		
	۰dB	۵dB	۱۰dB	۰dB	۵dB	۱۰dB	۰dB	۵dB	۱۰dB
N=۱	۲.۱۱۷	۲.۴۵۳	۲.۸۴۵	۱.۹۹۴	۲.۳۸۴	۲.۶۹۴	۲.۱۰۸	۲.۴۳۱	۲.۷۹۲
N=۴	۲.۱۳۱	۲.۴۷۸	۲.۸۶۷	۲.۰۱۱	۲.۴۰۲	۲.۷۲۶	۲.۱۲۶	۲.۴۵۵	۲.۸۱۱
N=۷	۲.۱۵۵	۲.۵	۲.۸۹۳	۲.۰۳۵	۲.۴۲۵	۲.۷۵۳	۲.۱۵۲	۲.۴۸۳	۲.۸۴
N=۱۰	۲.۱۷۸	۲.۵۲۹	۲.۹۲۸	۲.۰۶۴	۲.۴۵۶	۲.۷۸۴	۲.۱۷۸	۲.۵۱۶	۲.۸۶۹
N=۲۰	۲.۲۰۴	۲.۵۳	۲.۹۵۴	۲.۰۸۲	۲.۴۷۸	۲.۸۰۷	۲.۱۹۵	۲.۵۳۹	۲.۸۹۳
N=۳۰	۲.۲۱۷	۲.۵۶۸	۲.۹۶۸	۲.۰۹۸	۲.۴۹۷	۲.۸۲۶	۲.۲۰۸	۲.۵۵۵	۲.۹۰۸
N=۴۰	۲.۲۲۷	۲.۵۷۷	۲.۹۷۶	۲.۱۰۶	۲.۵۰۵	۲.۸۳۵	۲.۲۲	۲.۵۶۸	۲.۹۲۳
N=۵۰	۲.۲۳۴	۲.۵۸۴	۲.۹۸۱	۲.۱۱	۲.۵۱	۲.۸۴۱	۲.۲۲۵	۲.۵۷۶	۲.۹۳

بخش انتهایی نمودارها و پس از N =۳۰ تغییر چشمگیری در نتایج دیده نمی شود. افزایش N نیز به دلایلی که قبلا گفته شد، در N = ۵۰ متوقف شده است. جدول (۵-۴) مقادیر PESQ را برای هر سه نویز و سیگنال به نویز مذکور، نمایش می دهد.

۵-۷- مقایسه روش پیشنهادی با روش های دیگر

برای بررسی بهبود یا عدم بهبود روش پیشنهادی، سه روش معروف دیگر بهسازی گفتار پیاده سازی شدند و روش پیشنهادی با این سه روش مقایسه گردیده است. نتایج این مقایسه ها در این بخش آورده شده است.

روش هایی که برای مقایسه پیاده سازی شده اند، عبارت اند از:

- تخمین MMSE دامنه مبتنی بر توزیع گوسی [۲۲] که با MMSE نمایش داده شده است.
  تخمین LogMMSE دامنه مبتنی بر توزیع گوسی [۲۳] که با Log-MMSE نمایش داده شده است. (در LogMMSE یک تخمین بهینه با کمینه کردن میزان خطای مربعات لگاریتم دامنه به دست می آید.)
- تخمین MMSE طیف مبتنی بر توزیع لاپلاس [۴۳] که با Lap-MMSE نمایش داده شده است.
- تخمین گر پیشنهادی نیز با LMM-MMSE نمایش داده شده است.

می توان گفت که این سه روش، موارد مناسبی برای مقایسه می باشند. زیرا هر سه روش آماری محض هستند و در هر سه روش از تخمین گر MMSE استفاده شده است. دو روش از این سه روش بر پایه فرض توزیع گوسی برای سیگنال تمیز و روش دیگر بر پایه فرض توزیع لاپلاس برای سیگنال تمیز می باشند. در واقع هدف از مقایسه روش پیشنهادی با این سه روش، بررسی مناسب یا نامناسب بودن فرض LMM برای توزیع سیگنال تمیز نسبت به توزیع های لاپلاس و گوسی است.

برای مقایسه این چهار روش از دو نویز F۱۶ و همهمه و سه سطح سیگنال به نویز قبل استفاده شده است . لازم به ذکر است که برای پیاده سازی سه روش دیگر نیز از همان ده جمله از پایگاه داده TIMIT استفاده شد که برای پیاده سازی روش پیشنهادی مورد استفاده قرار داده شد. برای مقایسه نتایج، از مقادیر روش پیشنهادی در N=۵۰ استفاده شده است.

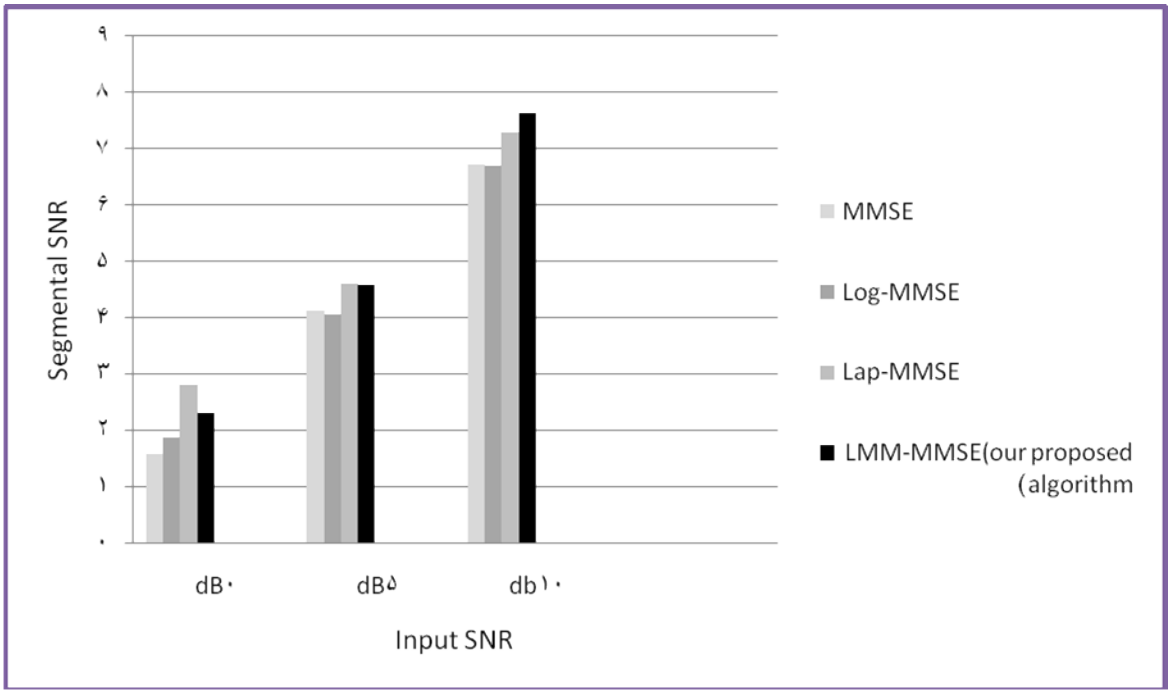
۵-۷-۱- معیار سیگنال به نویز قطعه ای

شکل های (۵-۱۹) و (۵-۲۰) نمودار سیگنال به نویز قطعه ای را برای چهار روش مذکور، در سه سطح سیگنال به نویز ۰و ۵ و ۱۰ و برای نویزهای F۱۶ و همهمه نشان می دهد.

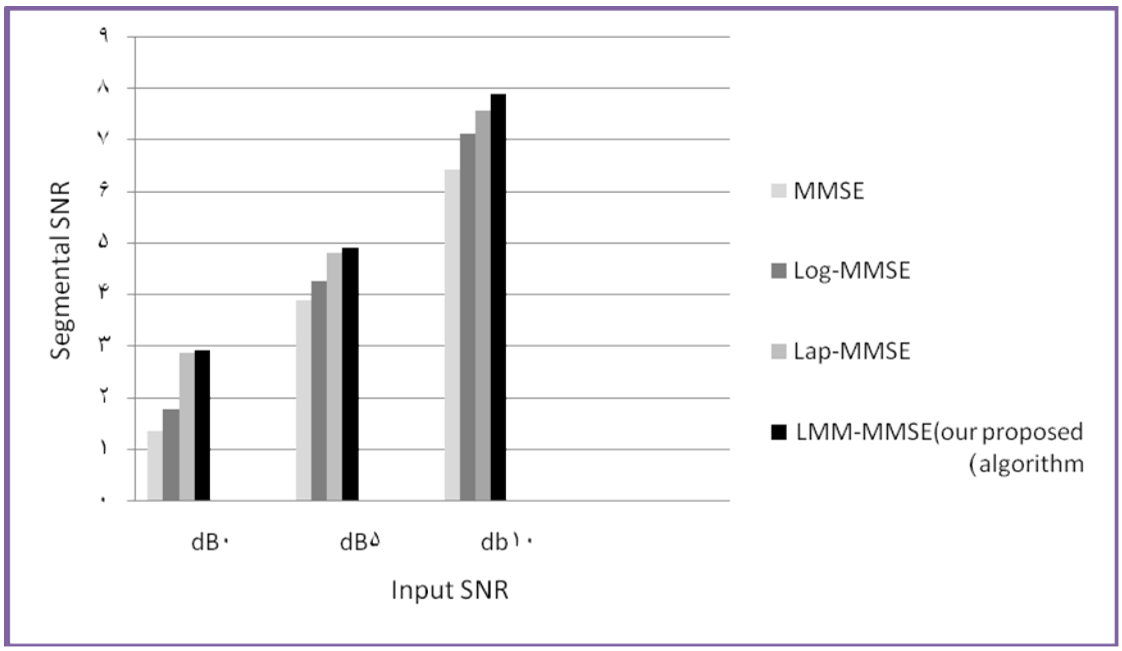
با توجه به نمودار شکل (۵-۱۹)، برای نویز F۱۶، در سیگنال به نویز صفر و پنج عملکرد روش پیشنهادی از دو روش MMSE و Log-MMS بهتر می باشد ولی عملکرد آن از روش Lap-MMSE بهتر نیست. در حالی که در سیگنال به نویز ۱۰dB روش پیشنهادی از سه روش دیگر موفق تر عمل

می کند. در واقع برای نویز F۱۶ و در سیگنال به نویزهای پایین روش Lap-MMSE که در آن سیگنال با یک لاپلاس مدل می شود و پارامترهای لاپلاس، از روی سیگنال نویزی تعیین می شود، از روش پیشنهادی بهتر عمل می کند. ولی با افزایش سیگنال به نویز عملکرد روش پیشنهادی بهتر می شود و از این روش پیشی می گیرد.

با توجه به نمودار (۵-۱۹)، دو روش MMSE و Log-MMSE که در آن سیگنال با توزیع گوسی مدل می شود، مقادیر سیگنال به نویز پایین تری از دو روش دیگر دارند. این مطلب بیانگر این است که مدل کردن سیگنال با توزیع لاپلاس و مخلوطی از لاپلاس از مدل کردن سیگنال با توزیع گوسی مناسب ترمی باشد. نکته دیگر که از جدول (۵-۵) مشخص می شود، این است که در همه سیگنال به نویزها، روش پیشنهادی بیش از یک دسی بل بهبود روی سیگنال به نویز قطعه ای، نسبت به این دو روش MMSE و Log-MMSE ایجاد می کند که مقدار قابل توجهی است.



شکل(۵-۱۹): مقایسه روش پیشنهادی با روش های MMSE, Log-MMSE, Lap-MMSE, بر حسب سیگنال به نویز قطعه ای و برای نویز F۱۶ .



شکل(۵-۲۰): مقایسه روش پیشنهادی با روش های MMSE, Log-MMSE, Lap-MMSE بر حسب سیگنال به نویز قطعه ای و برای نویز همهمه.

جدول(۵-۵): مقایسه روش پیشنهادی با روش های MMSE, Log-MMSE, Lap-MMSE بر حسب سیگنال به نویز قطعه ای و برای نویز همهمه و F۱۶ .

Noises	Babble noise			F۱۶ noise		
Estimators	-۱۱.۹/۰dB	-۷.۱/۵dB	-	-۱۲.۴/۰dB	-۷.۸۳/۵dB	-۲.۳/۱۰dB
MMSE [۲۲]	۱.۳٤۱	۳.۸۹۲	٦.٤۲	۱.۵۷	٤.۱۳۲	٦.۷۳۱
Log-MMSE [۲۳]	۱.۷۷۳	٤.۲۵۱	۷.۱۲۳	۱.۸۷۳	٤.۰۵۲	٦.٦۹
Lap-MMSE [٤۳]	۲.۸۷٤	٤.۸۰۸	۷.۵۸۲	۲.۸۱	٤.٦۱۳	۷.۲۸۸
LMM-MMSE (Proposed method)	۲.۹۱۲	٤.۹۱۵	۷.۹۱۲	۲.۳۱	٤.۵۷۱	۷.٦۲۳

برای نویز همهمه نیز با توجه به نمودار شکل (۵-۲۰)، روش پیشنهادی در همه سیگنال به نویزها از سه روش پیشی گرفته است. البته با توجه به نمودار، با افزایش سیگنال به نویز اختلاف روش پیشنهادی با سه روش دیگر زیاد می شود یعنی عملکرد آن بهتر می شود. جدول (۵-۵) نیز نشان دهنده مقادیر نمودارهای شکل (۵-۱۹) و (۵-۲۰) می باشد.

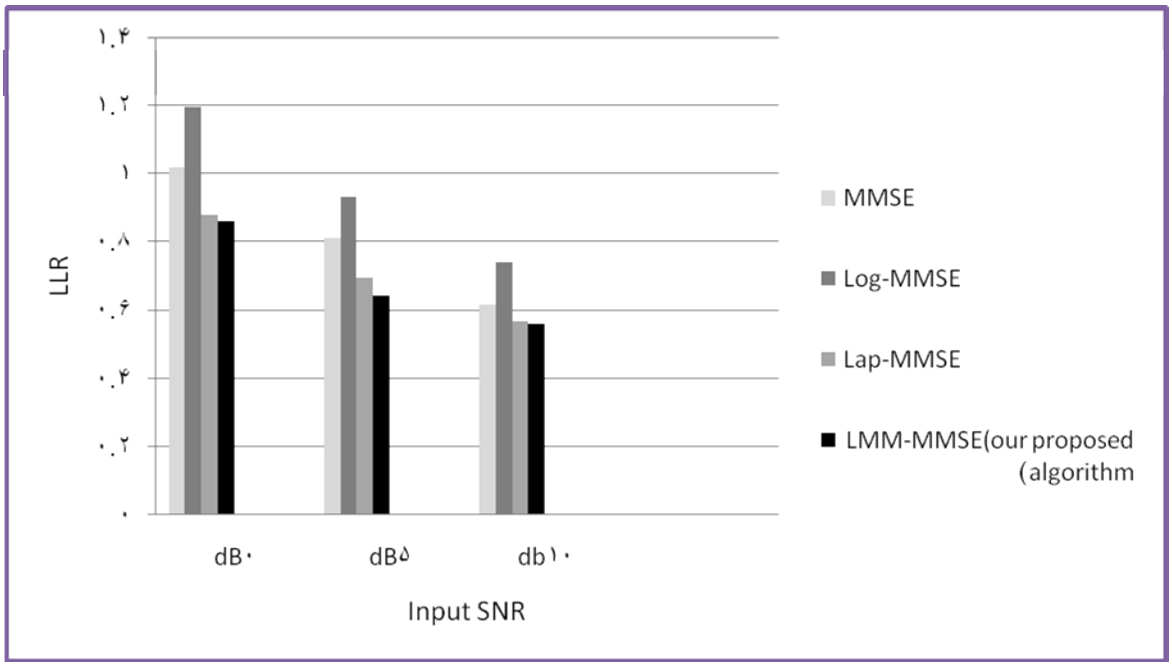
۵-۷-۲- معیار LLR

در این بخش معیار LLR جهت ارزیابی روش پیشنهادی با سایر روش های ذکر شده در بخش قبل به کار می رود. نتایج حاصل از این مقایسه، در شکل های (۵-۲۱) و (۵-۲۲) آورده شده است. با توجه به نمودار، در همه سیگنال به نویزها و برای هر دو نویز، روش پیشنهادی دارای LLR کمتر و در نتیجه سیگنال بهسازی شده با کیفیت بالاتر می باشد. در واقع مشخص می شود که LMM گزینه مناسبی برای مدل کردن سیگنال تمیز می باشد. با توجه به نمودارها، اختلاف روش پیشنهادی با دو روش مبتنی بر توزیع گوسی زیاد می باشد؛ اما اختلاف آن با روش مبتنی بر توزیع لاپلاس زیاد نیست. جدول (۵-۶) مقادیر عددی دو نمودار شکل (۵-۲۱) و (۵-۲۲) را نمایش می دهد.

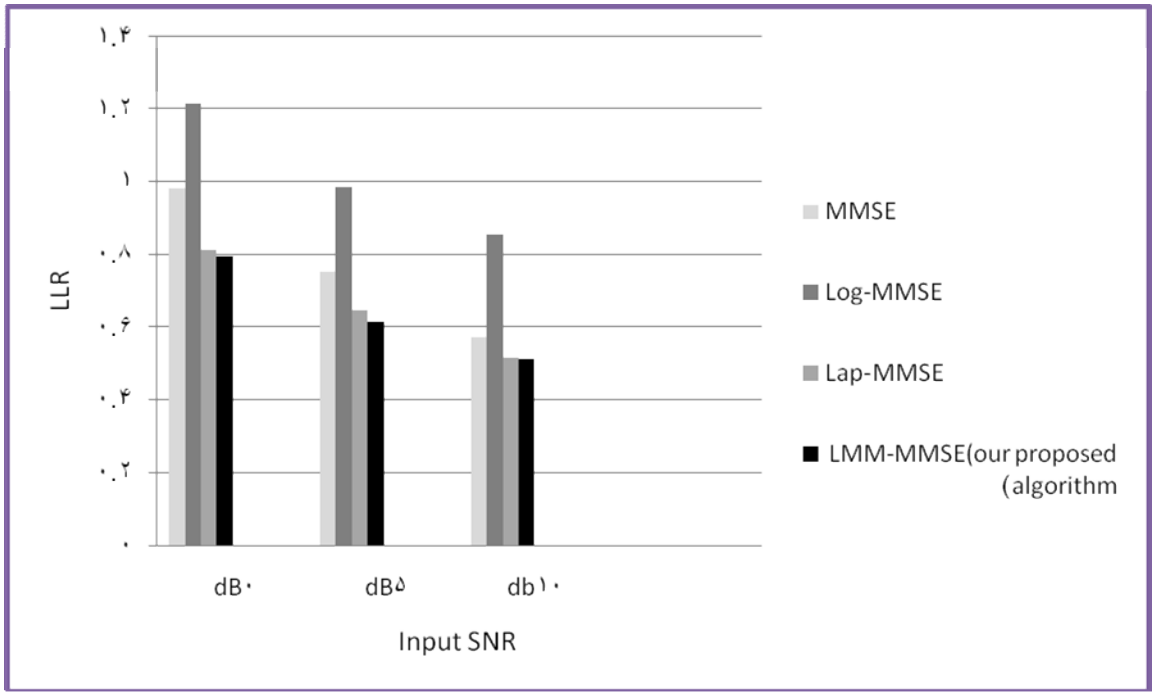
۵-۷-۳- معیار PESQ

در این بخش معیار PESQ جهت مقایسه کارایی روش پیشنهادی با سه روش دیگر به کار می رود. نتایج این مقایسه، در نمودارهای (۵-۲۳) و (۵-۲۴) آورده شده است. با توجه به نمودار شکل (۵-۲۳) در سیگنال به نویز صفر و برای نویز F۱۶ روش پیشنهادی از دو روش مبتنی بر توزیع گوسی بهتر ولی از روش مبتنی بر توزیع لاپلاس بدتر می باشد. (مقادیر بالاتر بیانگر سیگنال بهسازی شده با کیفیت بالاتر می باشند) اما برای سیگنال به نویز های دیگر، کارایی روش پیشنهادی از هر سه روش دیگر بهتر بوده و با افزایش سیگنال به نویز، اختلاف مقادیر PESQ برای این روش با روش های دیگر، بیشتر هم می گردد، یعنی عملکرد روش بهتر می شود.

برای نویز همهمه در همه سیگنال به نویزها، روش پیشنهادی، بر اساس این معیار، نسبت به سه روش دیگر، بهتر عمل می کند. در این حالت نیز با افزایش سیگنال به نویز اختلاف مقادیر PESQ برای روش پیشنهادی و بقیه روش ها بیشتر می شود. جدول (۵-۶) مقادیر نمودار های شکل (۵-۲۳) و (۵-۲۴) را نشان می دهد.



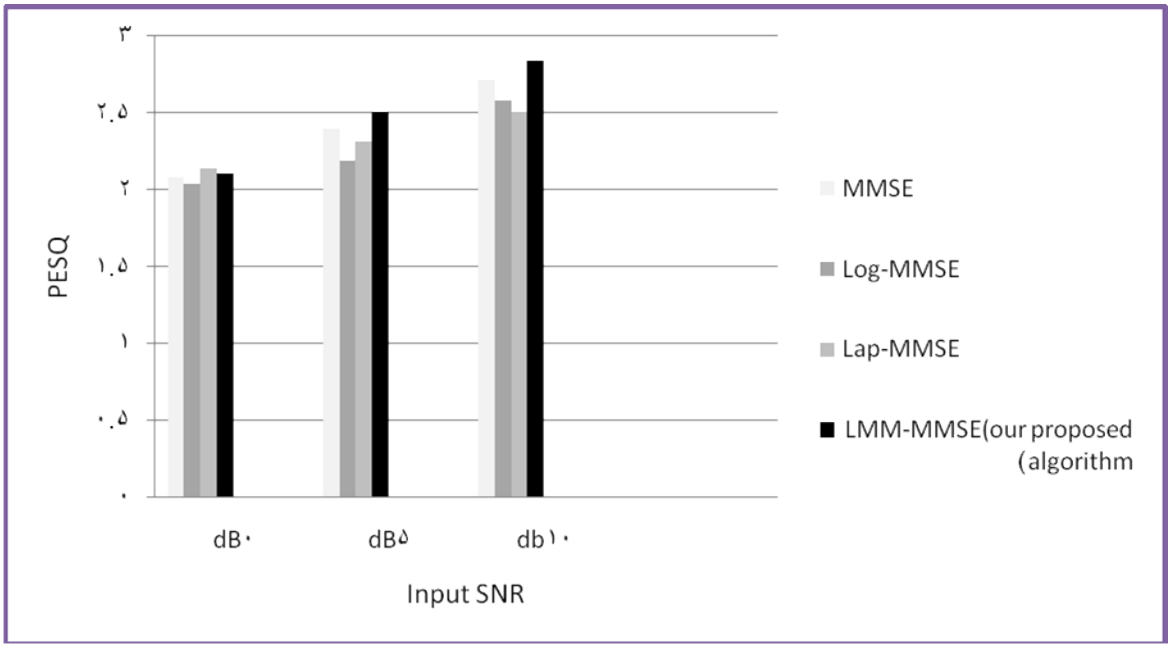
شکل(۵-۲۱): مقایسه روش پیشنهادی با روش های Lap-MMSE,Log-MMSE,MMSE, بر حسب LLR و برای نویز F۱۶.



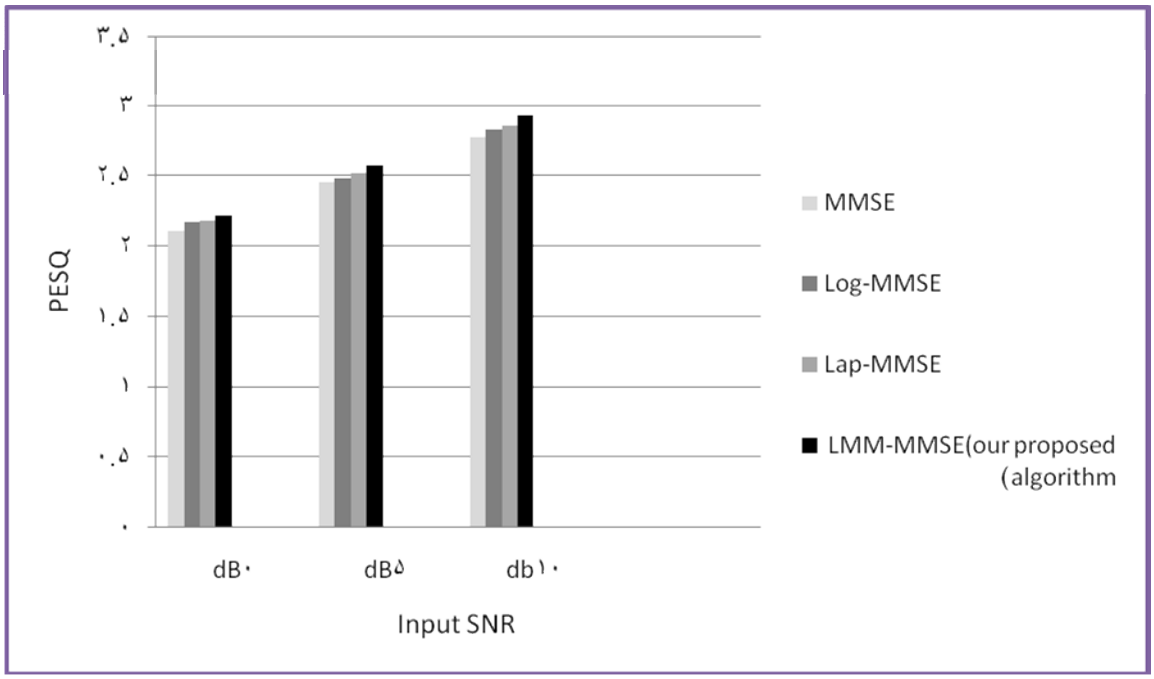
شکل(۵-۲۲): مقایسه روش پیشنهادی با روش های Lap-MMSE,Log-MMSE,MMSE, بر حسب LLR و برای نویز همهمه.

جدول(۵-۶): مقایسه روش پیشنهادی با روش های Lap-MMSE,Log-MMSE,MMSE, بر حسب LLR و برای نویز همهمه و F۱۶.

Noises / Estimators	Babble noise			F۱۶ noise		
	۰ dB	۵ dB	۱۰ dB	۰ dB	۵ dB	۱۰ dB
MMSE [۲۲]	۰.۹۸۱	۰.۷۵۱	۰.۵۷۲	۱.۰۲۱	۰.۸۱۲	۰.٦۱۹
Log-MMSE [۲۳]	۱.۲۱۳	۰.۹۸٤	۰.۸۵۲	۱.۱۹٦	۰.۹۳۵	۰.۷٤۳
Lap-MMSE[٤۳]	۰.۸۱	۰.٦٤٤	۰.۵۱٦	۰.۸۷۹	۰.٦۹٦	۰.۵۷
LMM-MMSE (Proposed method)	۰۷۹۳	۰.٦۱۲	۰.۵۱۳	۰.۸٦۱	۰.٦٤۳	۰.۵٦۲



شکل(۵-۲۳): مقایسه روش پیشنهادی با روش های Lap-MMSE,Log-MMSE,MMSE, بر حسب PESQ و برای نویز F۱۶.



شکل(۵-۲۴): مقایسه روش پیشنهادی با روش های Lap-MMSE,Log-MMSE,MMSE, بر حسب PESQ و برای نویز همهمه.

جدول(۵-۷): مقایسه روش پیشنهادی با روش های Lap-MMSE,Log-MMSE,MMSE, بر حسب PESQ و برای نویز همهمه و F۱۶.

Noises / Estimators	Babble noise			F-۱۶ cockpit nois		
	۰ dB	۵ dB	۱۰ dB	۰ dB	۵ dB	۱۰ dB
MMSE [۲۲]	۲.۱۱٤	۲.٤٥٦	۲.۷۸۱	۲.۰۸۱	۲.۳۹٦	۲.۷۱۲
Log-MMSE [۲۳]	۲.۱۷۲	۲.٤۸۷	۲.۸۳۳	۲.۰۳٥	۲.۱۹	۲.٥۸٥
Lap-MMSE[٤۳]	۲.۱۸۳	۲.٥۲	۲.۸٥۷	۲.۱۳٦	۲.۳۱٤	۲.٥۰۸
LMM-MMSE (Proposed method)	۲.۲۲٥	۲.٥۷٦	۲.۹۳	۲.۱۱	۲.٥۱	۲.۸٤۱

با توجه به مجموعه نمودارها و جداول بالا نکاتی مشخص می شوند که آنها را می توان به طور خلاصه و به صورت زیر بیان کرد:

۱- با افزایش تعداد اجزای مخلوط لاپلاس N، عملکرد روش پیشنهادی بهبود می یابد. که به دلیل دقیق تر شدن تقریب LMM از طیف سیگنال می باشد.

۲- مقدار بهینه برای N با توجه به مقادیر نمودارها و جداول، ۵۰ انتخاب می گردد. بهینه بودن این مقدار از این جهت مطرح می شود که باید بین تقریب دقیق تر - نتایج بهتر و پیچیدگی - زمان اجرای برنامه - حجم حافظه اشغالی کمتر، مصالحه شود.

۲- چهار روشی که با هم مقایسه شدند، برای حذف نویز همهمه عملکرد بهتری نسبت به حذف نویز F۱۶ دارند.

۳- روش پیشنهادی در بیشتر موارد، علی الخصوص در سیگنال به نویزهای بالا و برای نویز همهمه، از جهت حذف نویز و کیفیت سیگنال بهسازی شده، کارایی بهتری از سه روش دیگر دارد. بنابراین می توان نتیجه گرفت که مدل کردن سیگنال تمیز با LMM مناسب تر از مدل کردن سیگنال با لاپلاس تنها یا گوسی تنها می باشد.

۴- با افزایش سیگنال به نویز، عملکرد روش پیشنهادی بهتر می شود. ولی این بدان معنا نیست که روش پیشنهادی در سیگنال به نویزهای پایین عملکرد قابل قبولی نداشته باشد.

۵- به ازای Nهای کوچک مثل ۱ و ۴، روش پیشنهادی برای هر دو نویز و هر سه سیگنال به نویز، کارایی مناسبی نسبت به روش مبتنی بر لاپلاس ندارد. زیرا در روش پیشنهادی به ازای Nهای کو-چک، تقریب خوبی از سیگنال نداشته و همچنین پارامترهای توزیع سیگنال تمیز به صورت برون خطی[۱] تخمین زده می شوند. در حالی که در روش مبتنی بر لاپلاس، شاید یک لاپلاس برای تخمین طیف سیگنال تمیز مناسب نباشد ولی پارامترهای آن به صورت درون خطی[۲]، تخمین زده می شوند.

---

۱ - offline
۲ - online

فصل ششم

نتیجه گیری و پیشنهادات

۶-۱- مقدمه

در این فصل ابتدا یک جمع بندی کلی در مورد کارهای انجام شده در این پایان نامه آمده است و در ادامه پیشنهاداتی برای کارهای آینده ارائه خواهد شد.

۶-۲- نتیجه گیری و جمع بندی

در این پایان نامه به بررسی بهسازی گفتار، مفاهیم مربوط به آن و روش های مختلف بهسازی گفتار، علی الخصوص روش آماری محض پرداخته شد. سپس با ارائه مدل آماری مخلوط لاپلاس برای سیگنال گفتار تمیز و با استفاده از تخمین گر MMSE روش جدیدی برای تخمین سیگنال تمیز از روی سیگنال نویزی (نویز با سیگنال جمع شده و مستقل از آن است) ارائه شد.

در این پایان نامه، برای تخمین پارامترهای سیگنال تمیز، از EM الگوریتم و برای تخمین پارامترهای نویز از روش مبتنی بر ردیابی کمینه ها استفاده شد.

برای پیاده سازی روش پیشنهادی، سیگنال گفتار تمیز با سه نویز سفید، همهمه و F۱۶ و در سه سطح سیگنال به نویزهای ۰، ۵ و ۱۰ جمع گردید و سپس از سه معیار سیگنال به نویز قطعه ای، LLR و PESQ برای ارزیابی روش استفاده شد. در نهایت نیز عملکرد روش پیشنهادی به ازای افزایش تعداد اجزای مخلوط لاپلاس بررسی گردید، و نتایج این روش با سه روش معروف دیگر بهسازی گفتار (MMSE, Log-MMSE, Lap-MMSE) مقایسه شد.

بررسی نمودارها و جداول موجود در فصل پنجم، نشان می دهد که به ازای افزایش تعداد اجزای مخلوط لاپلاس N، عملکرد روش پیشنهادی، با توجه به مقادیر سه معیار ارزیابی ذکر شده، برای هر سه نویز و سیگنال به نویز بهتر می شود. دلیل این امر را می توان اینگونه بیان کرد که هرچه N بزرگتر شود LMM متناظر با آن با خطای کمتری بر تابع چگالی احتمال داده ها برازنده می شود و به بیان ساده تر مدل به واقعیت نزدیکتر می شود. مثلا برای حالت های N=۱ یا N=۴ که تقریب

خوبی از سیگنال موجود نمی باشد، نتایج هر سه معیار ارزیابی (در مقایسه با مقادیر سه روش بهسازی دیگر) مناسب و قابل قبول نمی باشد.

نکته دیگری که از مشاهده نمودارها و جداول نتیجه می شود، این است که به ازای N های کوچک، با تغییرN ، اختلاف نتایج نسبتا زیاد است. ولی با زیاد شدن N اختلاف نتایج کم و کمتر می شود. علت این امر این است که وقتی N کوچک است، به ازای تغییرات N تقریب های متناظر با Nها از طیف سیگنال تمیز اختلاف زیادی با هم دارند، ولی با افزایش N اختلاف بین تقریب های متناظر با Nهای متفاوت، کم می شود. (این مطلب با استفاده از معیار I_{KL} در فصل پنج نشان داده شد) با توجه به نمودارها، N =۳۰ پاسخ خوبی داشته و پس از آن به ازای تغییرات N، نتایج تغییرات زیادی ندارند. این نشان می دهد که تقریب توزیع سیگنال تمیز با توزیع LMM از حدود N =۳۰ به بعد، به واقعیت نزدیک و نزدیک تر می شود. به همین جهت افزایش N در ۵۰ توقف می یابد، زیرا در حالیکه با افزایش N پیچیدگی محاسباتی و زمان اجرای برنامه به شدت افزایش می یابد، نتایج تغییرات زیادی ندارد. در واقع، مقدار بهینه برای N با توجه به مقادیر نمودارها و جداول،۵۰ در نظر گرفته شد. بهینه بودن این مقدار از این جهت مطرح می شود که باید بین تقریب دقیق تر - نتایج بهتر و پیچیدگی - زمان اجرای برنامه - حجم حافظه اشغالی کمتر، مصالحه شود.

مطلب مهم دیگری که می توان به عنوان نتیجه به آن اشاره کرد، این است که؛ بهترین جواب ها براساس هر سه معیار، در سه سیگنال به نویز ۰ و ۵ و ۱۰ دسی بل، مربوط به نویز سفید گوسی است. پس از آن نویز همهمه و در نهایت نویز F۱۶ قرار می گیرد. با مشاهده و بررسی طیف توان این سه نویز مشاهده شد که تغییرات توان نویز سفید از دو نویز دیگر کمتر است. همچنین این تغییرات برای نویز همهمه کمتر از نویز F۱۶ می باشد. در بخش تخمین نویز بیان شد که فرض اساسی روش های تخمین نویز، کند بودن تغییرات توان نویز است. احتمالا به این دلایل نتایج برای نویز سفید از دو نویز دیگر بهتر است و نویز همهمه نیز نتایج بهتری از نویز F۱۶ دارد.

در مقایسه روش پیشنهادی با سه روش Lap-MMSE , Log-MMSE ,MMSE ، برای سیگنال های جمع شده با نویزهای F۱۶ و همهمه در سه سطح سیگنال به نویزهای ۰و ۵ و۱۰دسی بل، براساس سه معیار سیگنال به نویز قطعه ای، LLR و PESQ، این نتیجه حاصل شد که روش پیشنهادی برای حذف نویز همهمه در هر سه سطح سیگنال به نویز ذکر شده، از سه روش دیگر بهتر عمل می کند. اختلاف نتایج روش پیشنهادی و دو روش مبتنی بر توزیع گوسی خیلی زیاد است در حالیکه اختلاف نتایج روش پیشنهادی و روش Lap-MMSE زیاد نمی باشد. برای حذف نویز F۱۶، روش پیشنهادی در همه سیگنال به نویزها از دو روش مبتنی بر توزیع گوسی موفق تر است ولی در مقایسه با روش مبتنی بر توزیع لاپلاس، این روش در سیگنال به نویزهای بالاتر (۵ و ۱۰ دسی بل) بهتر عمل می کند و در سیگنال به نویز ۰ دسی بل روش مبتنی بر لاپلاس نتایج بهتری دارد. برای هر دو نویز، با افزایش سیگنال به نویز، اختلاف مقادیر به دست آمده برای روش پیشنهادی با روش های دیگر بیشتر می شود که نشانگر بهبود عملکرد روش پیشنهادی با افزایش سیگنال به نویز است.

نکته آخر این است که با مقایسه جدول های مربوط به بخش ارزیابی روش پیشنهادی و جدول های بخش مقایسه روش پیشنهادی با روش های دیگر، مشخص می شود که به ازای Nهای کوچک مثل ۱ و ۴ و ۷ مقادیر روش پیشنهادی برای هر دو نویز و هر سه سطح سیگنال به نویز، کارایی مناسبی نسبت به روش مبتنی بر توزیع لاپلاس ندارد. زیرا در روش پیشنهادی به ازای Nهای کوچک، هم تقریب خوبی از سیگنال نداریم و هم پارامترهای توزیع سیگنال تمیز به صورت برون خطی تخمین زده می شوند. در حالی که در روش مبتنی بر لاپلاس، شاید یک لاپلاس برای تخمین طیف سیگنال تمیز مناسب نباشد ولی پارامترهای توزیع سیگنال تمیز به صورت درون خطی، تخمین زده می شوند.

در کل مقایسه و بررسی نتایج بیانگر عملکرد قابل قبول روش پیشنهادی است و این مطلب نشان می دهد که تقریب سیگنال تمیز با توزیع LMM مناسب تر از تقریب آن با توزیع های گوسی یا

لاپلاس است. البته این مناسب تر بودن از جهت بهبود نتایج می باشد. زیرا از لحاظ پیچیدگی محاسباتی، زمان اجرای برنامه ها و حجم حافظه اشغالی، سه روش دیگر مناسب تر هستند.

۶-۳- پیشنهادات برای کارهای آینده

- در روش پیشنهادی در این پایان نامه، مدل LMM برای سیگنال تمیز و مدل گوسی را برای نویز در نظر گرفته شد. پیشنهاد دیگر این است، که نویز را هم با توزیع لاپلاس، مخلوطی از لاپلاس یا مخلوطی از گوسی مدل کرد.
- در این پایان نامه، برای تخمین سیگنال تمیز از تخمین گر MMSE استفاده شد. استفاده از تخمین گر MAP شاید باعث کاهش محاسبات و بهبود نتایج شود.
- در این پایان نامه، مقدار دهی اولیه به EM الگوریتم به صورت تصادفی انجام شد. می توان مقدار دهی اولیه را با روش های دیگر مثل الگوریتم ژنتیک انجام داد که این کار احتمالا منجر به ایجاد نتایج بهتری خواهد شد.
- در روش پیشنهادی، از تخمین گر MMSE برای تخمین بخش حقیقی و موهومی طیف سیگنال تمیز استفاده شد. در حالی که می توان دامنه و فازسیگنال تمیز را تخمین زد. در این حالت باید توزیع دامنه و فاز ضرایب طیف را با فرض اینکه توزیع بخش حقیقی و موهومی LMM باشد، محاسبه کرد.
- در این پایان نامه، از روش ردیابی کمینه ها برای تخمین نویز استفاده شده است. می توان به جای این روش از روش های دیگر تخمین نویز نیز استفاده کرد.
- عملکرد روش را می توان علاوه بر تغییر N با تغییر پارامتر دیگری مثل تغییر طول قاب نیز سنجید.
- روش پیشنهادی را می توان با فرض عدم قطعیت گفتار هم به دست آورد.روش پیشنهادی در حوزه طیف عمل می کند. شاید بتوان این روش را به حوزه های دیگر مثل حوزه زمان، کپسترال و .... گسترش داد.

فهرست منابع

[۱] Accardi. A. J and Cox. R. V, (۱۹۹۹), A Modular Approach to Speech Enhancement with an Application to Speech Coding, In IEEE International Conference on Acoustics, Speech and Signal Processing, pp. ۲۰۱-۲۰٤, Arizona, USA

[۲] Akbari.A and Lebouquin.R,(۱۹۹٦),Speech enhancement using a Wiener Filtering under signal presence uncertainty, European signal processing Conference, pp ۹۷۱-۹۷٤, Italy

[۳] Andrianakis.I, (۲۰۰۷), Baysian Algorithm for Speech Enhancement, Ph. D Thesis, Univecity os Southampton

[٤] Berouti. M and Schwartz. R, Makhoul. J and Cambridge . M. A, (۱۹۷۹) , Enhancement of Speech Corrupted by Acoustic Noise, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. ۲۰۸-۲۱۱, Washington,DC, USA

[٥] Bilmes. J. A, (۱۹۹۸), A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, Technical report, Department of Electrical Engineering and Computer Science, California

[٦] Boll. S ,(۱۹۷۹), Suppression of Acoustic Noise in Speech Using Spectral Subtraction**, IEEE Transaction on Aocostic, Speech and Signal Processing**, Vol. ۲۷, No. ۲, pp. ۱۱۳-۱۲۰

[۷] Breithaupt. C and Martin. R, (۲۰۰۳), MMSE Estimation of Magnitude-Squared DFT Coefficients with Super Gausssian Priors, ICASSP, pp. ۸۹٦-۸۹۹, Hong Kong, China .

[۸] Cappe. O, (۱۹۹٤), Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor, **IEEE Transaction on Aocostic, Speech and Signal Processing**, Vol. ۲, No. ۲, pp. ۳٤٥-۳٤۹

[۹] Chen. B, (۲۰۰٥), Speech Enhancement Using A Laplacian-Based MMSE Estimator, Ph. D. Thesis, Department of Electrical Engineering, Texas, USA, University of Texas

[۱۰] Chen. B, Loizou. P. C, (۲۰۰۷), A Laplacian-Based MMSE Estimator for Speech Enhancement, **Speech Communication,** Vol. ٤۹, No. ۲, pp. ۱۳٤-۱٤۳

[۱۱] Cohen. I, (۲۰۰۲), Optimal Speech Enhancement Under Signal Presence Uncertainty

Under Spectral Amplitude Estimator, **IEEE Signal Processing Letters**, Vol. ۹, No. ٤, pp.۱۱۳-۱۱٦

[١٢] Cohen. I, (٢٠٠٤), On the Decision-Directed Estimation Approach of Ephraim and Malah, ICASSP, pp. ٢٩٣-٦, Montreal, Canada .

[١٣] Cohen. I, (٢٠٠٥b), Relaxed Statistical Model for Speech Enhancement and apriori SNR Estimation, **IEEE Transaction on Aocostic, Speech and Signal Processing**, Vol. ١٣, No. ٥ part ٢, pp. ٣٣٦-٣٥٠

[١٤] De Moor. B, (١٩٩٣), The Singular Value Decomposition and Long and Short Space of Noisy Matrices, **IEEE Transaction on Aocostic, Speech and Signal Processing**, Vol. ٤١, No. ٩, pp. ٢٨٢٦-٢٨٣٨

[١٥] Dempster. A. P, Laird. N. M and Rubin. D. B, (١٩٧٧). Maximum Likellihood from Incomplete Data via the EM Algorithm, **Journal of the Royal Statistical Society,** Vol. ٣٩, No. ١, pp. ٤٥-٦٧

[١٦] Dendrinos.M, Bakamidis.S, and Carayannis. G, (١٩٩١), Speech Enhancement from Noise: a regenerative approach, **speech communication**, Vol. ١٠, No. ١, pp. ٤٥-٧٦

[١٧] Drucker. H, (١٩٦٨), Speech Processing in a High Ambient Noise Environment, **IEEE Transaction on Audio and Electracoustics**, Vol. ١٦, No. ٢, pp. ١٦٥-٨

[١٨] Ephraim. Y, (١٩٩٢a), A Bayesian Estimation Approach for Speech Enhancement Using Hidden Markove Models, **IEEE Transaction on Aocostic, Speech and Signal Processing**, Vol. ٤٠, No. ٤, pp. ٧٢٥-٧٣٥

[١٩] Ephraim. Y, (١٩٩٢b), Gain Adapted Hidden Markove Models for Recognition of Clean and Noisy Speech , **IEEE Transaction on Aocostic, Speech and Signal Processing**, Vol. ٤٠, No. ٦, pp. ١٣٠٣-١٣١٦

[٢٠] Ephraim. Y, (١٩٩٢c), Statistical Model-Based Speech Enhancement System , **Proceeding of the IEEE** , Vol. ٨٠, No. ١٠, pp. ١٥٢٦-١٥٥٥

[٢١] Ephraim. Y, Cohen. I, (٢٠٠٥), Recent Advancements in Speech Enhancement, In the Electrical Engineering HandBook, CRC Press

[٢٢] Ephraim. Y and Malah. D, (١٩٨٤), Speech Enhancement Using a Minimum Mean Square Error Short Time Spectral Amplitude Estimator, IEEE International Conference on Acoustics, Speech and Signal Processing , pp. ١١٠٩-١١٢١, San Diego, California

[٢٣] Ephraim. Y and Malah. D, (١٩٨٥), Speech Enhancement Using a Minimum Mean

Square Error Log Spectral Amplitude Estimator, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. ٤٤٣-٥, Tampa, Florida, USA

[٢٤] Ephraim. Y and Malah. D, Juang. B. H, (١٩٨٩), On the Application of Hidden Markove Models for Enhancing Noisy Speech, **IEEE Transaction on Aocostic, Speech and Signal Processing**, Vol. ٣٧, No. ١٢, pp. ١٨٤٦-١٨٥٦

[٢٥] Ephraim. Yand Van Trees. H. L, (١٩٩٥), A Signal Subspace Approach for Speech Enhancement, **IEEE Transaction on Aocostic, Speech and Signal Processing**, Vol. ٣,

No. ١٢, pp. ٢٥٢-٢٦٦

[٢٦] Eekelens. J. S, Jensen. J and Heusdens. R, (٢٠٠٧b). Speech Enhancement Based on Rayleigh Mixture Modeling of Speech Spectral Amplitude Distributions, In European Signal Processing Conference, pp. ٦٥-٩

[٢٧] Haigh. J. A and Mason. J. S, (١٩٩٣), Robust Voice Activity Detection Using Cepstral Feature, In IEEE International Conference on Computer, Communication, Control and Power Engineering, pp. ٣٢١-٤

[٢٨] Hendriks. R. C, (٢٠٠٨), Advances in DFT-Based Single-Microphon Speech Enhancement,Ph. D Thesis, Technischm Univerity, Delft

[٢٩] Hirsch. H. G and Ehrlicher. C, (١٩٩٥), Noise Estimation Technique for Robust Speech Recognition, IEEE International Conference on Acoustics,

Speech and Signal Processing, pp. ١٥٣-١٥٦, Detroit, Michigan, USA

[٣٠] Hu. Y and Loizou. P, (٢٠٠٦), Subjective Comparison of Speech Enhancement Algorithms, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. ١٥٣-٦, Toulouse, France

[٣١] Hu. Y and Loizou. P, (٢٠٠٨), Evaluation of Objective Quality Measures for Speech Enhancement, , **IEEE Transaction on Aocostic, Speech and Signal Processing**, Vol. ١٦, No. ١, pp. ٢٥٢-٢٦٦

[٣٢] Junqua. J. C, Reaves. B and Mark. B (١٩٩١), A Study of Endpoint Detection Algorithm in Adverse Conditions: Incidence on a DTW and HMM Recognizer, In European on Speech Communication and Technology, pp. ١٣٧١-١٣٧٤, Genova, Italy

[٣٣] Kondoz. A. M, (٢٠٠٤), Digital Speech: Coding for low Bit Rate Communication System, Wiley .

[٣٤] Laird.D, Dempster.A.P, Rubin.D.B,(١٩٩٧), Maximum Likelihood from Incomplete D- ata via the EM Algorithm, **Journal of the Royal Statistical Society**,Vol. ٣٩, No.١,

pp.١-٣٨.

[٣٥] Lim. J and Oppenhime. A, (١٩٧٩), Enhancement and Bandwidth Compression of Noisy Speech, **Proceeding of the IEEE**, Vol. ٦٧, No. ١٢, pp. ١٥٨٦-١٦٠٤

[٣٦] Lin. L, Holmes. W. H and Ambikairajah. E, (٢٠٠٣a), Adaptive Noise Estimation Algorithm for Enhancement, **Electrical Letters,** , Vol. ٣٩, No. ٩, pp. ٧٥٤-٧٥٥

[٣٧] Loizou. P. C, (٢٠٠٧), **Speech Enhancement:Theory and Practice,** CRC Press

[٣٨] Lotter. T, Vary. P, (٢٠٠٥), Speech Enhancement by MAP Spectral Amplitude Estimation Using a Super-Gaussian Speech Model, **EURASIP Journal on Applied Signal Processing**, Vol. ٢٠٠٥, No. ٧, pp. ١١١٠-١١٢٦

[٣٩] Malah. D, Cox. R,V and Accardi. A. J, (١٩٩٩), Tracking Speech Presence Uncertainty to Improve Speech Enhancement in Non-Stationary Noise Environment, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. ٧٨٩-٧٩٢, Arizona, USA

[٤٠] Martin . R, (١٩٩٤), Spectral Subtraction Based on Minimum Statistic, European signal processing Conference, pp ١١٨٢-٨٥, Edinburgh, Scotland

[٤١] Martin . R, (٢٠٠١), Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistic, **IEEE Transaction on Aocostic, Speech and Signal Processing**, Vol. ٩, No. ٥, pp. ٥٠٤-٥١٢

[٤٢] Martin . R, (٢٠٠٣), Statistical Method for Enhancement of Noisy Speech, In International Workshop on Acoustic Echo and noise control, pp. ٤٣-٦٥, Kyoto, Japan .

[٤٣] Martin . R and Breithaupt. C, (٢٠٠٣), Speech Enhancement in the DFT domain using Laplacian Speech Priori, In International Workshop on Acoustic Echo and noise control, pp. ٨٧-٩٠, Kyoto, Japan .

[٤٤] Martin.R,(٢٠٠٥), Speech Enhancement Based on Minimum Mean Square Error and Supergaussian Priors, **IEEE Transaction on Speech and Audio Processsing**, Vol.١٣, No.٥

[٤٥] McAulay. R and Malpass. M, (١٩٨٠), Speech Enhancement Using a Soft-Decision Noise Suppression Filter, **IEEE Transaction on Aocostic, Speech and Signal Processing**, Vol. ٩, No. ٥, pp. ٥٠٤-٥١٢

[٤٦] Mitianoudis. N and Stathaki. T, (٢٠٠٤), Overcomplete source separation using Laplasian Mixture Model, IEEE Signal Processing Letters, Vol. ١, No. ٢ .

[٤٧] Paliwal. K. K and Basu. A, (١٩٨٧), A Speech Enhancement Method Based on Kalman Filtering, IEEE International Conference on Acoustics, Speech and Signal

Processing, pp. ۱۷۷-۱۸۰, Dallas, Texas, USA

[٤٨] Paliwal. K and Alsteris. L, (۲۰۰٥), On the Usefulness of STFT Phase Spectrum in Human Listening Tests, **Speech Communication,** Vol. ٤٥, No. ۲, pp. ۱٥۳-۱۷۰

[٤٩] Papoulis,A and Pillai. S, (۲۰۰۲), **Probability, Random Variables and Stochastic Processes**, NewYork, McGraw-Hill

[٥۰] Porter. J and Boll. S, (۱۹۸٤), Optimal Estimator for Spectral Restoration of Noisy Speech, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. ۱۸A. ۲. ۱ – ۱۸A. ۲. ٤, San Diego, California, USA

[٥۱] Rabiner. L and Sambur. M, (۱۹۷۷), Voiced-Unvoiced Detection Using the Itakura

LPC Distance Measure, ICASSP, pp. ۳۲۳-٦

[٥۲] Rangachari. S, (۲۰۰٤), Noise Estimation Algorithms For Highly NON-Stationary Environment, Ph. D Thesis, University of Texas at Dallas

[٥۳] Ris . C and Dupont. S, (۲۰۰۱), Assesssing Local Noise Level Estimation Methods: Application to Noise Robust ASR. **Speech Communication**, V۰l.۳٤, No.۱, pp.۱٤٤-۱٥۸

[٥٤] Sohn. J and Sung. W, (۱۹۹۸), A Voice Activity Detector Employing Soft Decision Based Noise Spectrum Adaptation, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. ۳٦٥-۸, Seattle, USA

[٥٥] Soon. I and Yeo.C.K, (۱۹۹۹), Improved Noise Suppression Filter Using Self-Adaptive Estimator of probability of Speech Absence, **Signal Processing.** Vol. ۷٥, No. ۲, pp. ۱٥۱-۱٥۹

[٥٦] Sudirga. R, (۲۰۰۹), A Speech Enhancement System Based on Statistical and Acoustic-Phonetic Knowledg, M. S Thesis, Queen's University, Ontario, Canada .

[٥۷] Tucker. R, (۱۹۹۲), Voice Activity Detection Using periodicity Measure, **IEEE Proceeding on Communications, Speech and Vision**, Vol. ۱۳۹, No. ٤, pp. ۳۷۷-۳۸۰

[٥۸] Vary. P, (۱۹۸٥), Noise Suppression by Spectral Magnitude Estimation Mechanism and Theoretical Limits, **Signal Processing.** Vol. ۸, No. ٤,

[٥۹] Veisi.H, (۲۰۰٥), Model-Based Method for Noise Robust Speech Recognition Systems, M.S. Thesis, Computer Engineering Department, Sharif Universitie of Technology, Tehran, Iran

[٦۰] Wiener. N, (۱۹٤۹), **Extrapolation, Interpolation and Smoothing of Stationary Time Series: Engineering Application,** MIT Press, Cambridge, Mass

[٦١] Wolf. P. J and Godsill. S. J, (٢٠٠١), Simple Alternative to the Ephraim and Malah Suppression Rule for Speech Enhancement, **IEEE Workshop on Statistical Signal Processing,** Vol. ٢, pp. ٤٩٦-٤٩٩

[٦٢] Wolf. P. J and Godsill. S. J, (٢٠٠٣a), Efficient Alternative to the Ephraim and Malah Suppression Rule for Audio Signal Enhancement, **EURASIP Journal on Applied Signal Processing,** Vol. ٢٠٠٣, No. ١٠, pp. ١٠٤٣-١٠٥٥

[٦٣] You. C. H and Rahardja. S, (٢٠٠٣), Adaptive B-Order MMSE Estimator for Speech Enhancement, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. ٨٢٥-٨٥٥, Hong Kong, China

[٦٤] http:// WWW. Wikipedia.org/ Laplace Distribution. Html

[٦٥] http:// WWW. Wikipedia.org/ TIMIT. Html

[۶۶] اسفندیان. ن، (۱۳۸۶)، استفاده از آنالیز LPC در روش تبدیل موجک، جهت غنی سازی سیگنال گفتار، پایان نامه کارشناسی ارشد، دانشکده فنی مهندسی، دانشگاه مازندران.

Summary

The name " speech enhancement " refers to larg group of methods that improve the quality and intelligibility of noisy speech by suppressing the background noise from noisy signal.

There are a lot of methods and papers about speech enhancement, but estimation of the clean signal from the noisy signal is still one of the insoluble problem in speech processing field .

One of the most important methods for speech enhancement, are statistical methods, because they have better performance.

In these methods, a distribution for clean speech and noise are assumed. Then an estimator are used for estimating the clean signal from noisy signal.

In this thesis, an estimator for speech enhancement in Discrete Fourier Transform (DFT) domain is proposed. It is shown that the complex DFT coefficients of clean speech can be modeled more accurate, by Laplacian Mixture Model than Gaussian, Laplacian and GMM distributions. Then, an analytical solution for estimating the complex DFT coefficients with the MMSE (Minimum Mean Square Error) estimator is derived, when the clean speech DFT coefficients are mixture of Laplacians distributed and the DFT coefficients of noise are Gaussian distributed. The derived MMSE estimator is non-linear and it is shown that, this estimator has better performance than the estimators which are based on Laplacian or Gaussian models.

For estimating LMM's parameter the Expectation Maximization (EM) Algorithm is used. Indeed, the parameters of LMM are estimated offline with clean speech from TIMIT data base and the parameters of noise are estimated online with minimal tracking based method.

Finally, the proposed algorithm is evaluated in term of Segmental SNR, LLR(Log Likellihood Ratio) and PESQ (Perceptual Evaluation Of Speech Quality) and then the proposed method is compared with Laplacian and Gaussian based methods.

The results of this comparison shows that the proposed method has an acceptable performance.

Important words:

Speech enhancement, Laplacian Mixture Model, Expectation Maximization, MMSE estimator, Voice Activity Detector, Evaluation methods.