

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه یزد

دانشکده مهندسی برق و کامپیوتر

گروه مهندسی مخابرات

پایان نامه

برای دریافت درجه کارشناسی ارشد

مهندسی برق مخابرات

عنوان پایان نامه

**توسعه روشهای مبتنی بر فرض توزیع غیر گوسی برای بهسازی**

**گفتار**

استاد راهنما: دکتر حمیدرضا ابوطالبی

استاد مشاور: دکتر محمدرضا تابان

نگارنده: محبوبه کارگریان مروستی

مهر ۱۳۹۱

## تقدیم

به پدرم

که عالمانه به من می‌آموزد تا چگونه در عرصه زندگی، ایستادگی را تجربه نمایم

و به مادرم

درباری بی‌کران فداکاری و عشق که وجودم برایش همه رنج بود و وجودش برایم همه  
مهر

و به همسرم:

که سایه مهربانیش سایه‌سار زندگی‌م می‌باشد، او که اسوه صبر و تحمل بوده و مشکلات  
مسیر را برایم تسهیل نمود.

## تشکر و قدردانی

اولین سپاس به پیشگاه حضرت دوست که هر چه هست از اوست.

با تقدیر و تشکر شایسته از استاد فرهیخته و فرزانه جناب آقای دکتر حمیدرضا ابوطالبی که با نکته‌های دلاویز و گفته‌های بلند صحیفه‌های سخن را علم‌پرور نمود و همواره راهنما و راه‌گشای نگارنده در اتمام و اکمال پایان‌نامه بوده است. سپاس از استاد گرانقدر، جناب آقای دکتر محمدرضا تابان استاد مشاور اینجانب که با دقت و راهنمایی‌های سازنده‌شان بنده را در این پایان‌نامه همراهی کردند. و سپاس از زحمات جناب آقای دکتر مسعودرضا آقابزرگی و دکتر قانعی داور داخلی و خارجی این پایان‌نامه.

همچنین از همسرم که محیطی سرشار از سلامت و امنیت و آرامش و آسایش برای من فراهم آورده است تشکر می‌نمایم. و از فرزند کوچکم امیرعلی که به دلیل مشغله درسی من اذیت شد عذرخواهی می‌کنم. در آخر از خداوند منان برای تمامی این عزیزان موفقیت و بهروزی روزافزون تو ام با سلامتی مسئلت دارم.

## چکیده:

با توجه به کاربرد روزافزون سیستم‌های گفتاری در زندگی امروزی، کارهای بسیاری در زمینه بهسازی و بازشناسی گفتار صورت گرفته که در همه این کاربردها وجود نویز و پژواک باعث کاهش کارایی سیستم‌های گفتاری می‌شود. تاکنون تعداد زیادی از الگوریتم‌های بهسازی گفتار براساس تخمین دامنه طیف سیگنال گفتار و همچنین تخمین ضرایب حقیقی و موهومی طیف گفتار مبتنی بر معیار MMSE و MAP و با فرض گوسی بودن مؤلفه‌های دکارتی ضرایب DFT سیگنال گفتار ارائه شده‌اند. در سال‌های اخیر محققان به دنبال وارد کردن مدل‌های احتمالاتی دقیق‌تری برای مؤلفه‌های دکارتی ضرایب DFT سیگنال گفتار هستند.

در این پایان‌نامه، ابتدا به بررسی سیستم‌های تک‌کاناله تخمین دامنه سیگنال تمیز با فرض گوسی بودن مؤلفه‌های دکارتی ضرایب DFT سیگنال گفتار و همچنین سیستم تک‌کاناله تخمین مؤلفه‌های دکارتی ضرایب سیگنال تمیز با همان فرض گوسی بودن ضرایب می‌پردازیم. در ادامه نگاهی به سیستم‌های بهسازی که مبتنی بر فرض غیرگوسی برای مؤلفه‌های دکارتی ضرایب DFT سیگنال گفتار هستند، داریم. در این زمینه، نخست به کمک معیار Kullback به دنبال تابع توزیعی هستیم که بر هیستوگرام دامنه گفتار منطبق شود. در این راستا با استفاده از توزیع سوپرگوسین و با تغییر پارامترهایش از نظر معیار Kullback به توزیع نمایی رسیدیم که بر هیستوگرام دامنه گفتار منطبق‌تر است و تخمین‌گر دامنه طیف سیگنال گفتار را با فرض توزیع نمایی برای دامنه گفتار براساس معیار MMSE بدست آوردیم. در ادامه این پایان‌نامه، روش تخمین مرتبه متغیر MMSE با فرض سوپرگوسین بودن دامنه گفتار را مورد بازنگری قرار داده و تخمین‌گرهای جدید را استخراج می‌کنیم. شبیه‌سازی‌ها نشان می‌دهند که با در نظر گرفتن توزیع نمایی برای دامنه گفتار نسبت به حالتی که ضرایب توزیع لاپلاس یا گوسی دارند بهسازی بیشتری از نظر معیار segSNR بدست می‌آید.

کلمات کلیدی: تابع چگالی احتمال سوپرگوسین، بهسازی گفتار، تخمین دامنه گفتار، معیار

Kullback، تخمین ضرایب گفتار.

## فهرست مطالب:

### فصل اول: پیش‌گفتار ..... ۱

۱-۱- عوامل محیطی موثر در تخریب سیگنال گفتار ..... ۴

۱-۲- وارد کردن مدل آماری سیگنال گفتار ..... ۵

۱-۳- ساختار پالمی‌نامه ..... ۹

### فصل دوم: مروری بر روش تخمین سیگنال گفتار بر اساس توزیع‌های گوسی و غیرگوسی

مؤلفه‌های دکارتی ضرایب DFT سیگنال گفتار تمیز و استفاده از رویکردهای MMSE و

MAP ..... ۱۰

۱-۲- مدل‌های آماری مؤلفه‌های دکارتی برای ضرایب DFT سیگنال گفتار ..... ۱۲

۲-۲- مدل‌هایی برای دامنه ضرایب DFT سیگنال گفتار ..... ۱۷

۲-۳- تخمین سیگنال گفتار با رویکرد MMSE ..... ۲۰

۲-۳-۱- تخمین مؤلفه‌های دکارتی ضرایب DFT سیگنال گفتار با رویکرد MMSE ..... ۲۰

۲-۳-۲- تخمین دامنه طیفی گفتار با رویکرد MMSE ..... ۲۷

۲-۴- تخمین سیگنال گفتار با رویکرد MAP ..... ۳۶

۲-۴-۱- تخمین مؤلفه‌های دکارتی ضرایب DFT سیگنال گفتار با رویکرد MAP ..... ۳۷

۲-۴-۲- تخمین دامنه طیفی گفتار با رویکرد MAP ..... ۳۹

۲-۵- روش تخمین MMSE مرتبه متغیر دامنه طیف سیگنال ..... ۴۰

۲-۶- جمع‌بندی ..... ۴۲

### فصل سوم: تخمین دامنه طیف سیگنال گفتار با در نظر گرفتن توزیع‌نمایی برای دامنه

طیف سیگنال گفتار ..... ۴۴

۳-۱- معیار Kullback ..... ۴۵

۳-۲- تخمین MMSE با فرض توزیع سوپرگوسین برای دامنه ..... ۴۸

۳-۲-۱- تخمین دقیق MMSE بر اساس فرض سوپرگوسین ..... ۴۹

۳-۳- تخمین MMSE مرتبه متغیر دامنه طیف سیگنال گفتار با فرض مدل آماری

سوپرگوسین ..... ۵۲

.....۵۴.....	۴-۳- شبیه‌سازی و ارزیابی
.....۵۷.....	۳-۴-۱ ارزیابی عملکرد تخمین‌گر superMMSE
.....۶۲.....	۳-۴-۲ ارزیابی عملکرد تخمین‌گر $\beta$ -order superMMSE
.....۶۶.....	<b>فصل چهارم: جمع‌بندی و نتیجه‌گیری</b>
.....۶۶.....	۴-۱- کارهای انجام شده در این پلن‌نامه
.....۶۷.....	۴-۲- دستاوردهای این پلن‌نامه
.....۶۸.....	۴-۳- پیشنهادهایی برای ادامه کار

## فهرست شکل‌ها:

- شکل (۱-۲): هیستوگرام قسمت حقیقی ضرایب DFT سریگنال گفتار تم‌ی سه گوینده مرد و سه گوینده زن میانگین‌گیری شده با طول شش دقیقه، (... گوسی، (- . -) لاپلاسینو ( -- ) گاما ( $f=8000\text{Hz}$  و  $L=256$ ) شکل (ب) بزرگنمایی شده بخشی از شکل (الف) است [۱۸]..... ۷
- شکل (۱-۲): هیستوگرام دامنه طیف گفتار حدود یک ساعت گفتار و در محدوده طیفی بین ۵۰۰-۲۰۰۰ هرتز [6] ..... 8
- شکل (۱-۲): تابع چگالی احتمال گامای تعمیم‌یافته برای (a) با  $\gamma=1$  و  $\nu=\{0.25, 0.5, 0.75, 1\}$  و برای (b) با  $\gamma=2$  و  $\nu=\{0.1, 0.2, 0.3, 0.5\}$ ، همچنین تابع چگالی‌ها به واریانس یک نرمالیزه شده‌اند [۵]..... ۱۳
- شکل (۲-۲): نمودار تابع توزیع IG نمودار (----)  $\alpha=10$  و نمودار (.....) با  $\alpha=5$  و نمودار (---) با  $(\alpha=2.5)$  [۲۸] ..... ۱۵
- شکل (۳-۲): نمودار تابع توزیع NIG و مقایسه با تابع توزیع گوسی ..... ۱۶
- شکل (۴-۲): تابع چگالی احتمال گاما یکطرفه برای دامنه برای (a) با  $\gamma=1$  و  $\nu=\{0.8, 1, 1.5\}$  و برای (b) با  $\gamma=2$  و  $\nu=\{0.5, 1, 1.5\}$ ، همچنین تابع چگالی‌ها به واریانس یک نرمالیزه شده‌اند [۵]..... ۱۷
- شکل (۵-۲): مقایسه  $E\{S_R|Y_R\}$  در حالت مدل گفتار لاپلاس با فیلتر وینر برای سه مقدار SNR پیشین (15, 0, -10 db) و  $\sigma_N^2 + \sigma_S^2 = 2$  [۲۰]..... ۲۳
- شکل (۶-۲): مقایسه  $E\{S_R|Y_R\}$  با مدل گفتار گاما و فیلتر وینر برای سه SNR پیشین (15, 0, -) و (10 db)  $\sigma_N^2 + \sigma_S^2 = 2$  [۳۱]..... ۲۵
- شکل (۷-۲): مشخصات ورودی و خروجی تخمین‌گرهای DFT مختلط ( $\sigma_N^2 + \sigma_S^2 = 2$ ) (الف)  $\gamma=1$  و (ب)  $\gamma=2$  [۵]..... ۲۷
- شکل (۱-۳): هیستوگرام دامنه طیفی گفتار ۳۰ ثانیه ..... ۴۶
- شکل (۲-۳): مقایسه بین تابع چگالی احتمال رایلی و سوپرگوسین با تغییر پارامترها ..... ۴۶
- شکل (۳-۳): تحقق زمان گسسته از یک فرآیند نویز سفید تولید شده از یک کامپیوتر ..... ۵۵
- شکل (۴-۳): تحقق تبدیل فوریه زمان گسسته از یک فرآیند نویز صورتی ..... ۵۵



شکل (۳-۵): اسپکتروگرام نویز صورتی (چپ) و نویز سفید (راست) ..... ۵۶

شکل (۳-۶): مقایسه روش پیشنهادی superMMSE با روش های  $\beta$ -order MMSE-STSA، LSA، MMSE-STSA، LAPMMSE براساس توزیع گوسی، برای نویز سفید گوسی توسط معیار segSNR ..... ۵۹

شکل (۳-۷): مقایسه روش پیشنهادی superMMSE با روش های  $\beta$ -order MMSE-STSA، LSA، MMSE-STSA، LAPMMSE براساس توزیع گوسی، برای نویز سفید گوسی توسط معیار PESQ ..... ۵۹

شکل (۳-۸): مقایسه روش پیشنهادی superMMSE با روش های  $\beta$ -order MMSE-STSA، LSA، MMSE-STSA، LAPMMSE براساس توزیع گوسی، برای نویز سفید گوسی توسط معیار LLR ..... ۶۰

شکل (۳-۹): مقایسه روش پیشنهادی superMMSE با روش های  $\beta$ -order MMSE-STSA، LSA، MMSE-STSA، LAPMMSE براساس توزیع گوسی، برای نویز صورتی توسط معیار segSNR ..... ۶۰

شکل (۳-۱۰): مقایسه روش پیشنهادی superMMSE با روش های  $\beta$ -order MMSE-STSA، LSA، MMSE-STSA، LAPMMSE براساس توزیع گوسی، برای نویز صورتی توسط معیار PESQ ..... ۶۱

شکل (۳-۱۱): مقایسه روش پیشنهادی superMMSE با روش های  $\beta$ -order MMSE-STSA، LSA، MMSE-STSA، LAPMMSE براساس توزیع گوسی، برای نویز صورتی توسط معیار LLR ..... ۶۱

شکل (۳-۱۲): مقایسه روش پیشنهادی superMMSE با روش های  $\beta$ -order MMSE-STSA، LSA،  $\beta$ -order LAPMMSE و  $\beta$ -order MMSE-STSA براساس توزیع گوسی، برای نویز سفید گوسی توسط معیار segSNR ..... ۶۳

شکل (۳-۱۳): مقایسه روش پیشنهادی superMMSE با روش های  $\beta$ -order MMSE-STSA، LSA،  $\beta$ -order LAPMMSE و  $\beta$ -order MMSE-STSA براساس توزیع گوسی، برای نویز سفید گوسی توسط معیار PESQ ..... ۶۳

شکل (۳-۱۴): مقایسه روش پیشنهادی  $\beta$ -order superMMSE با روش های MMSE-STSA، LSA،  $\beta$ -order LAPMMSE و  $\beta$ -order MMSE-STSA براساس توزیع گوسی، برای نویز سفید گوسی توسط معیار LLR . ..... ۶۴

شکل (۳-۱۵): مقایسه روش پیشنهادی  $\beta$ -order superMMSE با روش های MMSE-STSA، LSA،  $\beta$ -order LAPMMSE و  $\beta$ -order MMSE-STSA براساس توزیع گوسی، برای نویز صورتی توسط معیار segSNR ..... ۶۴

شکل (۳-۱۶): مقایسه روش پیشنهادی  $\beta$ -order superMMSE با روش های MMSE-STSA، LSA،  $\beta$ -order LAPMMSE و  $\beta$ -order MMSE-STSA براساس توزیع گوسی، برای نویز صورتی توسط معیار PESQ ..... ۶۵

شکل (۳-۱۷): مقایسه روش پیشنهادی  $\beta$ -order superMMSE با روش های MMSE-STSA، LSA،  $\beta$ -order LAPMMSE و  $\beta$ -order MMSE-STSA براساس توزیع گوسی، برای نویز صورتی توسط معیار LLR ..... ۶۵

## فهرست جدول‌ها:

- جدول ( ۱-۳): استفاده از Kullback- Leibler Divergence بین تابع چگالی‌های متفاوت و هیستوگرام دامنه طیفی گفتار ..... ۴۹
- جدول (۲-۳): نام تخمین‌گر، مدل احتمالاتی DFT گفتار و شماره فرمول تخمین‌گر ..... ۵۶

## کلمات اختصاری

ApLapMMSE	.....	Approximate Laplacian MMSE
JMAP	.....	Joint Maximum a Posteriori
IG	.....	Inverse Gaussian
LLR	.....	Log-Likelihood Ratio
LPC	.....	Linear Prediction Coefficients
MAP	.....	Maximum a Posteriori
MMSE	.....	Minimum Mean Square Error
MMSE-LSA	.....	MMSE-Log Spectral Amplitude
MNIG	.....	Multivariate Normal Inverse Gaussian
NIG	.....	Normal Inverse Gaussian
OM-LSA	.....	Optimally-Modified Log Spectral Amplitude
PESQ	.....	Perceptual Evaluation of Speech Quality
RIG	.....	Rayleigh Inverse Gaussian
STSA	.....	Short time spectral amplitude
SNR	.....	Signal to Noise Ratio
SegSNR	.....	segmental SNR

# فصل اول

## پیش‌گفتار

تعداد زیادی از سیستم‌های پردازش گفتاری مثل کدینگ گفتار و بازشناسی گفتار براساس سیگنال‌های گفتار عاری از نویز ایجاد شده و عملکرد آنها در محیط‌های نویزی به سرعت کاهش می‌یابد؛ بنابراین پاکسازی سیگنال گفتار از نویز با استفاده از شیوه‌های بهسازی گفتار مورد توجه قرار گرفته است.

منظور از بهسازی سیگنال گفتار<sup>۱</sup>، عبارت است از تلاش برای بالا بردن کیفیت و قابلیت فهم این سیگنال و یا بهبود عملکرد سیستم‌های ارتباط گفتاری، در شرایطی که گفتار تحت تأثیر عوامل مخرب (نویز<sup>۲</sup> و پژواک<sup>۳</sup>) قرار گرفته باشد. بنابراین هدف بهسازی گفتار با توجه به

- 
- 1- Speech Enhancement
  - 2- Noise
  - 3- Reverberation

کاربردهای خاص متفاوت است [۱]. تاکنون روش‌های متنوعی برای بهسازی گفتار در قالب دو دسته الگوریتم‌های تک‌کاناله<sup>۱</sup> و الگوریتم‌های چندکاناله<sup>۲</sup> ارائه شده است. در الگوریتم‌های تک‌کاناله (همچون تبدیل موجک<sup>۳</sup>، تفریق طیفی<sup>۴</sup>، فیلتر وینر<sup>۵</sup> و...) تنها یک سیگنال برای پردازش موجود می‌باشد و معمولاً عملیات بهسازی با فرض ایستادن بودن نویز در فاصله بین دو ناحیه سکوت<sup>۶</sup> متوالی و براساس اختلاف طیف سیگنال گفتار و نویز صورت می‌پذیرد. در این الگوریتم‌ها، به دلیل در دسترس بودن تنها یک کانال، امکان بهسازی بیشتر از حد مشخصی وجود ندارد. این در حالی است که در الگوریتم‌های چندکاناله، با هدف ارتقای SNR نهایی و در نتیجه نیل به کیفیت بالاتر، از دو یا چند میکروفون برای ضبط صدا استفاده می‌شود. در این زمینه نیز تاکنون روش‌های مختلفی با بهره‌گیری از فیلترهای<sup>۷</sup> و فیلتر وینر چندکاناله و ... ارائه شده که هر یک، ضمن دارا بودن ویژگی‌های مطلوب، ناکارآمدی‌های خاص خود را نیز به دنبال دارند.

اکثر الگوریتم‌های بهسازی گفتار تک‌کاناله در حوزه فرکانس [۲] و معمولاً از نوع تخمین‌گر طیف توان نویز [۳]، تخمین‌گر دامنه طیف سیگنال تمیز [۴] و تخمین‌گر مؤلفه‌های دکارتی ضرایب طیف سیگنال هستند [۵]. به دلیل طبیعت غیر ایستادن سیگنال‌های گفتار نیاز داریم که سیگنال‌های نویزی در فریم‌های کوتاه مدت پردازش شوند. بعضی از تخمین‌گرهای سیگنال گفتار قاعده تخمین آماری را بر اساس مدل آماری مؤلفه‌های دکارتی ضرایب تبدیل فوریه گسسته<sup>۸</sup> (DFT) بکار می‌برند. مثال‌های معروف در این زمینه، الگوریتم‌های مبتنی بر معیار حداقل خطای میانگین مربعی<sup>۹</sup> (MMSE) [۴] و ماکزیمم‌سازی احتمال پسین<sup>۱۰</sup> (MAP) [۶] و فیلتر وینر [۱] است که در این الگوریتم‌ها دامنه و یا خود مؤلفه‌های دکارتی ضرایب DFT سیگنال گفتار را بر اساس معیار حداقل خطای میانگین مربعی تخمین می‌زنند.

- 
- 1- Single-Channel Algorithms
  - 2- Multi-Channel Algorithms
  - 3- Wavelet Transform
  - 4- Spectral Subtraction
  - 5- Wiener Filter
  - 6- Silence
  - 7- Adaptive Filter
  - 8- discrete Fourier transform
  - 9- Minimum Mean Square Error
  - 10- Maximum a posteriori

تاکنون تعداد زیادی از الگوریتم‌های بهسازی گفتار در حوزه فرکانس پیشنهاد شده‌اند [۸]. در این پایان‌نامه روی تخمین‌گرهای دامنه طیف کوتاه مدت<sup>۱</sup> (STSA) با استفاده از معیار MMSE و MAP متمرکز شده‌ایم که در دو دهه‌ی اخیر توجه بسیار زیادی از محققان را به خود جلب کرده است [۴و۶]. در کارهای قبلی، Ephraim و Malah تخمین‌گر دامنه طیف کوتاه مدت را با فرض مدل آماری گوسی ارائه داده‌اند [۴]. فرض مدل آماری گوسی تقریب خوبی برای ضرایب DFT نویز است [۹]. اما برای سیگنال گفتار که اندازه معمولی فریم‌های DFT استفاده شده برای آن کوتاه است (۴۰ms-۱۰ms) فرض مدل آماری گوسی مناسب نمی‌باشد. در [۱۰-۱۲] نشان داده شده است که قسمت حقیقی و موهومی ضرایب سیگنال گفتار می‌توانند توسط توزیع‌های بهتری مانند توزیع گاما یا لاپلاسین مدل شوند. همچنین با استفاده از معیار Kullback در [۱۳] نشان داده شده است که دامنه طیف گفتار بهتر است که توسط توزیع سوپرگوسین مدل شود. این مسأله، محققان را بر آن داشت تا بر اساس مدل‌های دقیق‌تری مثل لاپلاسین یا گاما یا سوپرگوسین یک تخمین‌گر بهینه MMSE-STSA و یا تخمین‌گرهای دامنه طیف کوتاه مدت (STSA) با استفاده از رویکرد MAP [۱۳] بدست آورند.

## ۱-۱ عوامل محیطی مؤثر در تخریب سیگنال گفتار

دو مورد از مهم‌ترین عوامل مخرب سیگنال گفتار نویز و انعکاس می‌باشند. در اینجا به بررسی اجمالی از نویز خواهیم پرداخت.

در یک تعریف جامع، به میدان‌های صوتی در غیاب سیگنال گفتار مطلوب، میدان نویزی اطلاق می‌شود [۱۵]. در حقیقت نویز، مجموعه‌ای از سیگنال‌های مزاحم است که با قرار گرفتن در زمینه‌ی گفتار تمیز، منجر به افت کیفیت آن می‌شود. در محیط‌های واقعی، این عامل مخرب، از منابع متنوعی سرچشمه می‌گیرد. از این رو میدان‌های نویزی از تنوع بسیار وسیعی برخوردار هستند و می‌توان آنها را از دیدگاه‌های مختلفی تقسیم‌بندی نمود.

---

1- Short Time Spectral Amplitude

در دیدگاه اول براساس میزان همبستگی<sup>۱</sup> سیگنال‌های نویز در نقاط مختلف از فضا، نویزهای محیط به دو دسته نویزهای همبسته و نویزهای ناهمبسته افراز می‌شوند؛ در یک میدان نویزی همبسته، امواج، بدون هیچ انعکاس و یا پراکندگی<sup>۲</sup>، به سمت آرایه ای از میکروفون‌ها<sup>۳</sup> منتشر می‌شوند به نحوی که این سیگنال‌ها در نقاط مختلف فضا از جمله روی میکروفون‌ها، کاملاً به یکدیگر همبسته می‌باشند [۱۶]. پس می‌توان چنین نتیجه گرفت که یک میدان نویزی همبسته، از منبعی در فضای آزاد و بدون مانع عمده در پخش صدا، حاصل می‌شود، مانند انواع نویز جهت‌دار. این در حالی است که در میدان‌های نویزی ناهمبسته، سیگنال‌های نویز در نقاط مختلف فضا، فاقد همبستگی هستند. از نمونه‌های این دسته از میدان‌های نویزی می‌توان به نویز الکتریکی میکروفون‌ها اشاره کرد.

از منظر دیگر و براساس میزان پویایی نویز<sup>۴</sup>، می‌توان آن را به دو دسته ایستاد<sup>۵</sup> و ناپایدار<sup>۶</sup> افراز نمود. ماهیت ناپایدار برخی از نویزها باعث شده تا از الگوریتم‌های وفقی برای تطابق زمانی با شرایط محیط استفاده کنیم.

نویزهای محیط در قالب دو دسته نویزهای جهت‌دار<sup>۷</sup> و همه‌جهته<sup>۸</sup> نیز تقسیم می‌شوند [۱۷]. همانطور که از نام آنها نیز مشخص است، در دسته اول، سیگنال‌های مزاحم از یک جهت مشخصی از فضا دریافت می‌شوند، در حالی که در دسته دوم، این سیگنال‌ها از جهت‌های مختلفی از محیط دریافت خواهند شد.

میدان نویزی پخشنده<sup>۹</sup>، برآیند سیگنال‌های ناخواسته‌ای است که با احتمال مساوی و انرژی یکنواخت در همه جهت‌های محیط منتشر شده و همگی دارای دامنه برابر و فاز تصادفی هستند؛ این مدل که تلفیقی از اجزای همبسته و ناهمبسته است، به دلیل کاربرد زیادی که در شرایط

- 
- 1- Correlation
  - 2- Scattering
  - 3- Microphone Array (MA)
  - 4- Dynamic Noise
  - 5- Stationary
  - 6- Non-Stationary
  - 7- Directional Noise
  - 8- Omni-Directional Noise
  - 9- Diffuse



واقعی (همچون محیط‌های کاری و یا محیط داخل خودرو) دارد، از اهمیت بسیار بالایی برخوردار می‌باشد [۷].

طیف وسیعی از الگوریتم‌های بهسازی گفتار، براساس تحلیل فرکانسی سیگنال‌های نویزی استوارند. در این راستا، نویز را می‌توان به دو دسته سفید و رنگی افراز نمود. همان‌طور که می‌دانیم نویز سفید از طیف فرکانسی یکنواختی برخوردار بوده و نمونه‌های آن دو به دو ناهمبسته می‌باشند. در محیط‌های واقعی بیشتر با نویزهایی با طیف فرکانسی غیریکنواخت (رنگی) مواجه هستیم. از نمونه نویزهای رنگی می‌توان به نویز همهمه یا نویز خودرو اشاره کرد.

## ۱-۲ وارد کردن مدل آماری سیگنال گفتار

در سال ۱۹۸۴، Ephraim و Malah تخمین‌گر دامنه طیف کوتاه مدت (STSA) با فرض مدل آماری گوسی ارائه دادند [۴]. فرض مدل آماری گوسی تقریب خوبی برای مؤلفه‌های دکارتی ضرایب DFT نویز است [۹]. از آنجایی که گفتار یک سیگنال غیر ایستادن است اگر آن را پنجره‌بندی کنیم می‌توان آن را در بازه‌های کوچک شبه ایستادن فرض کرد. برای سیگنال گفتار که اندازه معمولی فریم‌های DFT استفاده شده در پردازش‌های معمول گفتاری کوتاه هستند (۴۰ms-۱۰ms) فرض مدل آماری گوسی مناسب نمی‌باشد. Porter و Boll [۱۰]، Brehm و Stammer [۱۸] و Martin [۱۱] نشان داده‌اند که هنگامی که طول فریم‌های گفتار بین ۴۰ms-۱۰ms هستند، قسمت حقیقی و موهومی ضرایب سیگنال گفتار می‌توانند توسط توزیع‌های بهتری مانند توزیع گاما یا لاپلاسیان مدل شوند. همچنین Loannis و Paul [۱۴] و Lotter [۱۳] نشان داده‌اند که دامنه طیفی گفتار بهتر است که با توزیع سوپرگوسین، گاما یا CHI مدل شود. توزیع‌های گوسی، گاما، لاپلاسیان، سوپرگوسین، CHI و رایلی به ترتیب به صورت زیر تعریف می‌شوند:

الف) مدل گوسی (میانگین صفر و واریانس  $\lambda_x$ ):

$$P_X(x) = \frac{1}{\sqrt{\pi\lambda_x}} \exp\left(-\frac{x^2}{\lambda_x}\right) \quad (1-1)$$

ب) مدل گاما:

$$P_X(x) = \frac{\sqrt[4]{3}}{2\sqrt[4]{2\lambda_x}\sqrt{\pi}} |x|^{1/2} \exp\left(-\frac{\sqrt{3}|x|}{\sqrt{2}\sqrt{\lambda_x}}\right) \quad (2-1)$$

ج) مدل لاپلاسین:

$$P_X(x) = \frac{1}{2\sqrt{\lambda_x}} \exp\left(-\frac{|x|}{\sqrt{\lambda_x}}\right) \quad (3-1)$$

د) مدل سوپرگوسین:

$$P(X) = \frac{\mu^{\nu+1}}{\Gamma(\nu+1)\sigma_s^{\nu+1}} X^\nu \exp\left(-\frac{\mu X}{\sigma_s}\right) \quad (4-1)$$

ه) مدل CHI Distribution:

$$P(X) = \frac{2}{\theta^a \Gamma(a)} X^{2a-1} \exp\left(-\frac{X^2}{\theta}\right) \quad (5-1)$$

و) مدل گامای تعمیم یافته برای دامنه:

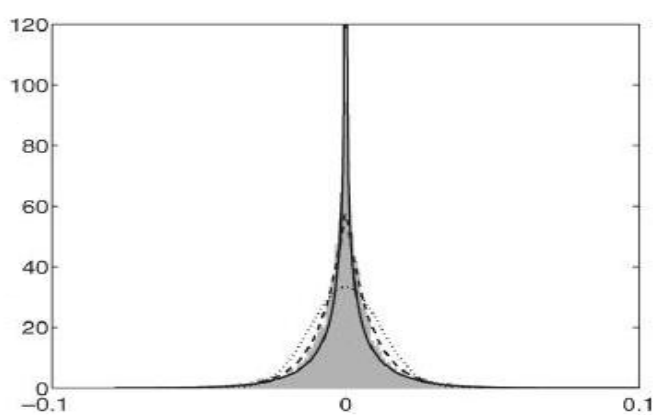
$$P(X) = \frac{1}{\theta^a \Gamma(a)} X^{a-1} \exp\left(-\frac{X}{\theta}\right) \quad (6-1)$$

ی) توزیع رایلی:

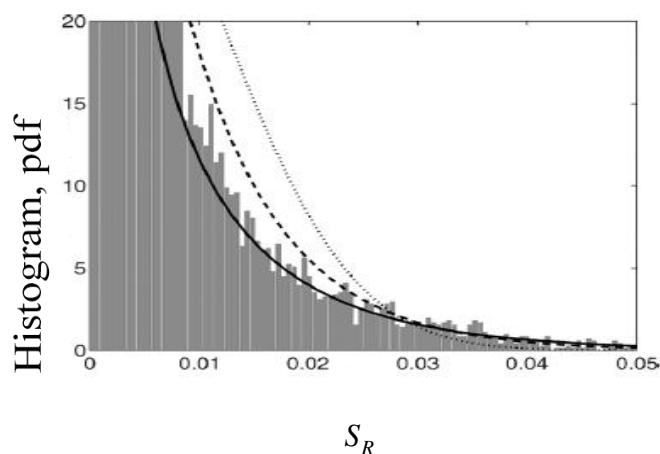
$$P(X) = \frac{X}{\sigma_s^2} \exp\left(-\frac{X^2}{2\sigma_s^2}\right) \quad (7-1)$$

سه رابطه اول مدلی برای مؤلفه‌های دکارتی ضرایب DFT گفتار هستند که  $x$  نشان دهنده بخش حقیقی یا موهومی است و روابط بعدی مدلی برای دامنه طیفی گفتار هستند و متغیر  $X$  معرف دامنه است.  $\sigma_s^2 = \lambda_x$  واریانس متغیر تصادفی  $x$  است. پارامترهای  $\nu$  و  $\mu$  در رابطه (4-1) شکل تابع چگالی احتمال را تعیین می کند بطوریکه با  $\nu=1$ ،  $\mu=2.5$  توزیع مؤلفه‌های دکارتی ضرایب DFT گفتار لاپلاس و با  $\nu=0.01$ ،  $\mu=1.5$  توزیع مؤلفه‌های دکارتی ضرایب DFT گفتار گاما خواهد شد. رابطه (5-1) توزیع CHI با  $2a$  درجه آزادی و پارامتر مقیاس  $\sqrt{\frac{\theta}{2}}$  می باشد [5]. در رابطه (6-1) پارامتر  $a$  شکل توزیع را مشخص می کند و رابطه ممان دوم و پارامتر  $\theta$  بصورت  $E[S^2] = \theta^2 a(a+1)$  می باشد که  $E[.]$  بیانگر امید ریاضی است.

شکل (۱-۱) نمودار هیستوگرام قسمت حقیقی ضرایب DFT سیگنال گفتار تمیز سه گوینده مرد و سه گوینده زن میانگین گیری شده با طول شش دقیقه را نشان می دهد (اندازه فریم  $L=256$  و فرکانس نمونه برداری  $f=8000Hz$ ). همان طوری که از این شکل ها مشخص است توزیع گوسی برای مؤلفه های دکارتی ضرایب DFT سیگنال گفتار مناسب نیست و توزیع های دیگری مثل توزیع گاما، توزیع لاپلاسین و غیره را می توان برای مدل کردن مؤلفه های دکارتی ضرایب DFT استفاده کرد.



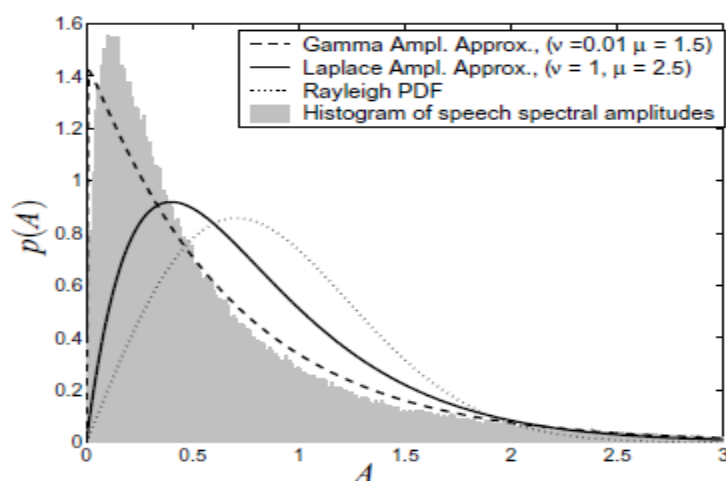
(الف)



(ب)

شکل (۱-۱): هیستوگرام قسمت حقیقی ضرایب DFT سیگنال گفتار تمیز سه گوینده مرد و سه گوینده زن میانگین گیری شده با طول شش دقیقه، (....) گوسی، (—) لاپلاسین و (----) گاما ( $L=256$ ) و شکل (ب) بزرگنمایی شده بخشی از شکل (الف) است [۱۸].

شکل (۲-۱) نمودار هیستوگرام دامنه طیفی سیگنال گفتار تمیز از مجموعه داده‌ای حدود یک ساعت گفتار و در محدوده فرکانسی بین ۵۰۰-۲۰۰۰ هرتز را نشان می‌دهد. همان‌طوری که از این شکل مشخص است توزیع رایلی برای دامنه طیفی سیگنال گفتار مناسب نیست و توزیع‌های دیگری مثل توزیع سوپرگوسین با تغییر پارامترهایش را می‌توان برای مدل کردن دامنه طیفی گفتار استفاده کرد.



شکل (۲-۱): هیستوگرام دامنه طیف گفتار حدود یک ساعت گفتار و در محدوده طیفی بین ۵۰۰-۲۰۰۰ هرتز [6]

در این پایان‌نامه توزیع گوسی را برای مؤلفه‌های دکارتی ضرایب DFT سیگنال نویز و توزیع سوپرگوسین را برای دامنه طیفی سیگنال گفتار تمیز در نظر گرفتیم. نتایج نشان می‌دهند که تخمین‌گرهایی که توزیع نمایی (  $\nu=0, \mu=2.5$  ) در (۴-۱) را برای دامنه طیفی گفتار در نظر گرفته‌اند، نسبت به تخمین‌گرهای مشابه که فرض توزیع لاپلاس (  $\nu=1, \mu=2.5$  ) در (۴-۱) را برای مؤلفه‌های دکارتی ضرایب DFT گفتار در نظر گرفته‌اند، و نسبت به تخمین‌گرهایی که در آن فرض گوسی بودن مؤلفه‌های دکارتی ضرایب DFT گفتار در نظر گرفته شده، عملکرد بهتری دارند.