
بسم الله الرحمن الرحيم

دانشکده مهندسی برق و رباتیک

رساله دکتری مهندسی برق- الکترونیک

استخراج ویژگی‌های مقاوم برای تشخیص گفتار در محیط‌های نویزی مبتنی بر تبدیل فوریه‌ی کسری

نگارنده: محسن صادقی

استاد راهنما

دکتر حسین مروی

اساتید مشاور

دکتر علیرضا احمدی‌فرد

Dr. Maaruf Ali

شهریور ۱۳۹۹

شماره : ۳/۱۷۵۶ ت.ب
تاریخ : ۹۹/۷/۳۰
ویرایش :

باسمه تعالی

مدیریت تحصیلات تکمیلی

فرم شماره ۱۲: صورت جلسه نهایی دفاع از رساله دکتری (Ph.D)

(ویژه دانشجویان ورودی های ۹۴ و ما قبل)

بدینوسیله گواهی می شود آقای/خانم **محسن صادقی** دانشجوی دکتری رشته **برق – الکترونیک** به شماره دانشجویی **۹۴۰۰۶۸۵** ورودی مهر ماه سال **۱۳۹۴** درتاریخ **۱۳۹۹/۰۶/۳۰** از رساله **نظری** ☒ / **عملی** ☐ خود با عنوان :

استخراج ویژگی های مقاوم برای تشخیص گفتار در محیط های نویزی مبتنی بر تبدیل فوریه کسری

دفاع و با اخذ نمره ۱۹/۲۵ به درجه : عالی نائل گردید.

الف) درجه عالی: نمره ۲۰-۱۹ ☑	ب) درجه بسیار خوب: نمره ۱۸/۹۹ – ۱۷ ☐
ج) درجه خوب: نمره ۱۶/۹۹– ۱۵ ☐	د) غیر قابل قبول و نیاز به دفاع مجدد دارد ☐
ه) رساله نیاز به اصلاحات دارد ☐	

ردیف	هیئت داوران	نام و نام خانوادگی	مرتبه علمی	امضاء
	دکتر محسن نوری	استاد/ اساتید راهنما	دانشیار	
	دکتر علیرضا اخوان فر	مشاور/ مشاورین	دانشیار	
	دکتر محمد رضا کریمی	استاد مدعو داخلی / خارجی	دانشیار	
	دکتر هادی گرایلو	استاد مدعو داخلی / خارجی	استادیار	
	دکتر سید علی سید ایوری	استاد مدعو داخلی / خارجی	دانشیار	
	دکتر عمار ابراهیمی	سرپرست (نماینده) تحصیلات تکمیلی دانشکده	استادیار	

مدیر محترم تحصیلات تکمیلی دانشگاه:

ضمن تأیید مراتب فوق مقرر فرمائید اقدامات لازم در خصوص انجام مراحل دانش آموختگی آقای/ خانم **محسن صادقی** بعمل آید.

نام و نام خانوادگی رئیس دانشکده:

تاریخ و امضاء و مهر دانشکده:

این پایان نامه را تقدیم می کنم به مهربانترین همراهان زندگیم، پدر، مادر، همسر عزیزم که حضورشان

همیشه گرما بخش روح من بوده است.

اکنون که به یاری پروردگار و یاری و راهنمایی اساتید بزرگ موفق به پایان این رساله شده‌ام وظیفه خود دانسته

که نهایت سپاسگزاری را از تمامی عزیزانی که در این راه به من کمک کرده‌اند را به عمل آورم:

در آغاز لازم است از پدر، مادر و همسر عزیزم به خاطر کمک های بی دریغشان تشکر کنم.

از استاد بزرگ جناب آقای دکتر حسین مروی که راهنمایی این رساله را به عهده داشته‌اند کمال تشکر را

دارم.

از جناب آقای دکتر علی رضا احمدی فرد که استاد مشاور این رساله بوده‌اند نیز قدردانی می نمایم.

تعهدنامه

اینجانب **محسن صادقی**دانشجوی دوره دکتری رشته **مهندسی برق** دانشکده مهندسی برق و رباتیک دانشگاه صنعتی شاهرود، نویسنده رساله **استخراج ویژگی‌های مقاوم برای تشخیص گفتار در محیط‌های نویزی مبتنی بر تبدیل فوریه‌ی کسری**تحت راهنمائی **دکتر حسین مروی** متعهد می‌شوم.

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان‌نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ

امضای دانشجو

چکیده

گفتار منبع اصلی برای ارتباط بین انسان‌ها به منظور نشان دادن ایده‌ها، احساس و تفکرشان به یکدیگر است. تکنولوژی بازشناسی گفتار این امکان را برای کامپیوتر فراهم کرده است که بتواند فرمان‌های گفتاری انسان را دریافت و آن‌ها را تفسیر کرده و واکنش مناسب را نشان دهد. به دلیل وجود نویز در محیط‌های واقعی با چالش عدم انطباق و شرایط نابرابر در دو حالت آزمون و آموزش شبکه برای کاربردهای واقعی مواجه هستیم. مقاومت در برابر نویز یک موضوع وسیع در تحقیقات سیستم‌های ASR[۱] می‌باشد که چندین دهه قدمت تحقیق و پژوهش دارد و پژوهشگران زیادی در این زمینه فعالیت نموده اند.

در این رساله، ابتدا به منظور بررسی مقاومت فورمنت‌های فریم‌های واکدار گفتار در محیط‌های نویزی با منابع مختلف، میزان جابه‌جایی فورمنت‌های فریم‌های واکدار نویزی نسبت به فریم‌های واکدار تمیز، اندازه گیری شده است. نشان داده شد که نویز سفید در تمامی سطوح سیگنال به نویز بررسی شده دارای بیشترین تاثیر روی فورمنت‌های واکدار سیگنال گفتار است. در ادامه الگوریتمی به منظور استخراج ویژگی مقاوم برای بازشناسی گفتار ارائه گردید. این ساختار پیشنهادی مبتنی بر تبدیل فوریه‌ی کسری و تابع ریشه است و FrRC[۲] نام گذاری شد. برای ارزیابی تئوری این روش پیشنهادی، یک رابطه‌ی ریاضی بین ویژگی‌های FrRC گفتار بدون نویز، نویز و گفتار نویزی بدست آورده شد و این رابطه با رابطه‌ی ریاضی مربوط به روش استخراج ویژگی MFCC[۳] در حالات مختلف مقایسه گردید. نتایج پیاده سازی سیستم بازشناسی گفتار مبتنی بر روش استخراج ویژگی FrRC حاکی از افزایش صحت بازشناسی نسبت به سایر روش های متداول استخراج ویژگی است. افزایش ۲۴/۶ و ۲۵/۳ درصدی صحت بازشناسی به ترتیب در مقایسه با روش های LPC و MFCC در محیط نویزی با نویز Babble با سطح سیگنال به نویز ۱۰-دسیبل گواه این ادعاست.

به منظور افزایش صحت بازشناسی گفتار در محیط های نویزی، الگوریتم دیگری که مبتنی بر تبدیل فوریه‌ی کسری و روش PNCC[۴] است، با نام AFPNCC[۵] معرفی، تحلیل و سپس پیاده سازی گردید. در الگوریتم پیشنهادی AFPNCC، براساس نوع و شدت نویز، ضریب آلفا تبدیل فوریه‌ی کسری موجود در الگوریتم، توسط بهینه‌ساز تکامل

[۱] Automatic Speech Recognition

[۲] Fractional Root Coefficients

[۳] Mel-Frequency Cepstral Coefficient

[۴] Power Normalized Cepstral Coefficient

[۵] Adaptive Fractional Power Normalized Cepstral Coefficient

تفاضلی که در بدنه ساختار الگوریتم پیشنهادی قرار دارد، استخراج می‌شود. نتایج پیاده سازی این الگوریتم بهبود صحت بازشناسی گفتار هم در محیط نویزی و هم بدون نویز را نشان می‌دهد. نتایج عددی حاصل از شبیه سازی سیستم بازشناس گفتار مبتنی بر الگوریتم استخراج ویژگی AFPNCC نیز نشان دهنده‌ی افزایش ۱۶ درصدی صحت بازشناسی نسبت به الگوریتم PNCC در محیط نویزی با نویز Pink و سطح سیگنال به نویز ۵- دسیبل است.

کلمات کلیدی: بازشناسی مقاوم گفتار، استخراج ویژگی مقاوم، ویژگی‌های کپسترال، تبدیل فوریه کسری، بهینه ساز تکامل تفاضلی

لیست مقالات مستخرج از رساله

مقالات ژورنالی استخراجی از رساله:

۱- ارائه یک روش نوین و کارآمد استخراج ویژگی برای بازشناسی گفتار مقاوم مبتنی بر تبدیل فوریه کسری و بهینه ساز تکامل تفاضلی، محسن صادقی، حسین مروی و علیرضا احمدی‌فرد، نشریه مدل‌سازی در مهندسی. (پذیرفته شده و در حال چاپ).

2- M. Sadeghi, H. Marvi, and M. Ali, "The effect of different acoustic noise on speech signal formant frequency location," International Journal of Speech Technology vol. 21, no. 3, pp. 741-752, Aug. 2018.

3- M. Sadeghi et al., "A New Feature Extraction Algorithm for Robust Speech Recognition Based on Mel Scale Root Fractional Fourier Transform," Circuits, Systems, and Signal Processing Journal. (Revised).

مقالات کنفرانسی :

4- M. Sadeghi and H. Marvi, "Optimal MFCC features extraction by differential evolution algorithm for speaker recognition," In 3rd Iranian Conference on Intelligent Systems and Signal Processing (ICSPIS), pp. 169-173, Dec. 2017.

فهرست مطالب

فهرست جدول ها

فهرست شکل ها

علائم اختصاری

ASR	Automatic Speech Recognition
CNN	Convolutional Neural Network
SNR	Signal to Noise Ratio
GMM	Gussian Mixture Model
HMM	Hidden Markov Model
MFCC	Mel-Frequency Cepstral Coefficient
LPC	Linear prediction coefficient
CMS	Cepstral mean subtraction
PCA	Principal component analysis
VTLN	Vocal Tract Length Normalization
AMFCC	Autocorrelation Mel-Frequency Cepstral Coefficient
MMSE	Minimum-Mean-Square-Error
PNN	Probabilistic Neural Network
PSO	Particle Swarm Optimization
DE	Differential Evolution
FRFT	Fractional Fourier Transform
MSM	Mean Square Movement
SVD	Singular Value Decomposition
SVM	Support Vector Machine
DTW	Dynamic Time Wrapping
PNCC	Power Normalized Cepstral Coefficient
FFT	Fast Fourier Transform
DWT	Discrete Wavelet Transform
WPT	Wavelet Packet Transform
DWPD	Discrete Wavelet Packet Decomposition
MMSE	Minimum Mean Square Error
LSF	Line Spectral Frequency
FrRC	Fractional Root Coefficient
APNCC	Adaptive Power Normalized Cepstral Coefficient
ARCC	Adaptive Root Cepstral Coefficient
MSM	Mean Square Movement
DHF	Dynamic Harmonic Feature
CMN	Cepstral Mean Normalization
PSD	Power Spectral Density
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
PE	Pre-Emphasis

فصل اول: مقدمه

۱-۱- مقدمه

بازشناسی گفتار فرآیندی به منظور تبدیل یک سیگنال ورودی دریافت شده به دنباله‌ای از حروف است که این فرآیند یک الگوریتم پیاده سازی شده بصورت یک برنامه کامپیوتری می‌باشد. به عبارت دیگر سیستم بازشناسی گفتار به کامپیوتر اجازه می‌دهد تا حروفی که یک شخص در یک میکروفن یا تلفن صحبت می‌کند را شناسایی و آنها را به متن قابل خواندن تبدیل کند. سیستم بازشناسی گفتار کاربردهای با ارزش زیادی که در آن نیاز به ارتباط بین ماشین و انسان می‌باشد را پشتیبانی می کند [۲-۱]. امروزه تکنولوژی‌های گفتار در گستره‌ی وسیعی از کاربردهای تجاری و صنعتی مورد استفاده قرار می‌گیرد. تاریخچه این تکنولوژی گویای این مطلب است که اولین سیستم بازشناسی گفتار در آزمایشگاه Bell در سال ۱۹۵۰ طراحی و پیاده‌سازی گردید. از آنجایی که سیگنال گفتار ذاتاً ناایستا است، بازشناسی گفتار ناشی از اختلاف جنسیت، حالت احساسی، لهجه، تلفظ، تفسیر، تو دماغی صحبت کردن ، فرکانس گام، سطح صدا و تنوع سرعت در افراد بعنوان یک کار سخت و پیچیده محسوب می‌شود [۳].

سیستم‌های بازشناسی گفتار خودکار در شرایط خاصی چون محیط‌های بدون نویز، بدون بازتاب صدا و یا تنوع کانال‌ها بسیار خوب عمل می‌کنند. این شرایط ایده‌ال به سختی می‌تواند در عمل برآورده شود. محیط‌های آکوستیک واقعی چالش‌های مختلفی به سیستم‌های بازشناسی گفتار تحمیل می-کنند. لذا مقاوم بودن سیستم‌های بازشناسی گفتار باید برای کاربردهای عملی مورد توجه قرار گیرد [۳-۱].

در کاربردهای واقعی اساسی‌ترین چالش در سیستم‌های پردازش گفتار، وجود نویزهای متفاوت با سطوح سیگنال به نویز مختلف که در بسیاری از موارد هیچ دانش قبلی در مورد این نویزها وجود ندارد، می‌باشد. چالش نویز در سیستم‌های ASR[۱] در محیط‌های واقعی به دلیل حضور نویز وجود

[۱] Automatic Speech Recognition

دارد و این مسئله باعث ایجاد عدم انطباق و شرایط نابرابر در دو محیط آزمون و آموزش شبکه برای کاربردهای جهان واقعی می‌گردد. مقاومت در برابر نویز یک موضوع وسیع در تحقیقات سیستم‌های ASR می‌باشد که چندین دهه قدمت تحقیق و پژوهش دارد [۴]. اخیراً تحقیقات زیادی بر روی توسعه سیستم‌های ASR به منظور طراحی سیستم با بیشترین مقاومت در برابر فاکتورهایی که باعث ایجاد شرایط نابرابر در دو فاز آموزش و آزمون می‌شود، صورت پذیرفته است که در بین این فاکتورها مهم‌ترین و چالشی‌ترین فاکتور ناخواسته همان بحث نویز می‌باشد[۴]. اگر از چندین سطح از زنجیره پردازش گفتار در سیستم ASR استفاده گردد، میزان عدم انطباق شرایط در دو فاز آموزش و آزمون کاهش می‌یابد[۴].

رویکردهایی که برای مقابله با این چالش وجود دارد را می توان در سه دسته کلی دسته‌بندی نمود که این سه دسته عبارتند از [۳و۴]:

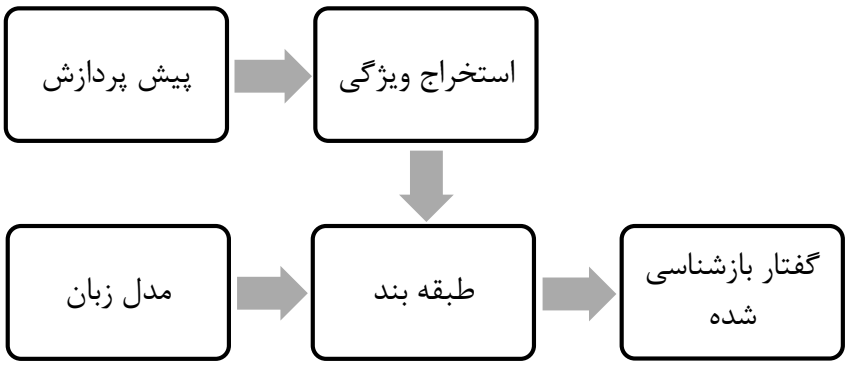
- بهبود گفتار
- تطبیق مدل گفتار
- استخراج ویژگی مقاوم

در ارتباطات اگر سیگنال گفتار تحت تاثیر نویز قرار گرفته باشد، بازسازی سیگنال گفتار بسیار سخت می‌باشد. مهم‌ترین خاستگاه حذف نویز از سیگنال گفتار این است که پیام اصلی بتواند بطور صحیح و قابل فهم دریافت گردد [۵].

۱-۲- بررسی اجمالی سیستم بازشناسی گفتار

توسعه در تکنولوژی گفتار ناشی از این است که انسان تمایل دارد مدل‌های مکانیکی به منظور ایجاد ارتباط کلامی قابلیت شبیه سازی پیدا کنند. پردازش گفتار به کامپیوتر این اجازه را می‌دهد که از فرمآنهای صوتی و زبآنهای مختلف انسان پیروی و تبعیت نماید. شکل ۱-۱ یک مدل پایه از سیستم

بازشناسی گفتار را نشان می‌دهد. این مدل از طبقات مختلفی از قبیل پیش پردازش، استخراج ویژگی گفتار، طبقه بند و همچنین مدل زبان تشکیل شده است [۲و۶].



شکل ۱-۱: طبقات بازشناسی گفتار [۲]

بعد از پیش پردازش، طبقه استخراج ویژگی بردارهای لازم را که در طبقه مدل زبان مورد استفاده قرار می‌گیرند، استخراج می‌کند. این بردارها برای اینکه صحت بالاتری داشته باشند باید در برابر نویز مقاوم باشند. بلوک طبقه بندی نیز گفتار را از طریق ویژگی‌های استخراج شده و مدل زبان تشخیص می‌دهد [۲و۳].

۱-۳- استخراج ویژگی برای بازشناسی گفتار

پردازش گفتار شامل تکنولوژی‌های مختلفی از قبیل کدگذاری گفتار، سنتز گفتار، بازشناسی گفتار و بازشناسی گوینده است. پردازش گفتار از دو قسمت بنیادی استخراج و طبقه بندی تشکیل می‌شود. مهم‌ترین معیار برای یک سیستم پردازش گفتار، روش استخراج ویژگی استفاده شده در آن است. چرا که روش استخراج ویژگی نقش بسیار مهم و اساسی در صحت سیستم پردازش گفتار دارد. استخراج ویژگی از سیگنال گفتار بدین معناست که از سیگنال گفتار یکسری ضرایب بردارهای ویژگی استخراج شده که این ویژگی‌ها فقط شامل اطلاعات مورد نیاز برای شناسایی گفتارهای دریافت شده است. ویژگی‌های استخراج شده از سیگنال گفتار باید معیارهای خاصی را برآورده نماید که مهم‌ترین این معیارها شامل موارد زیر می‌باشند[۲و۷]:

- ویژگی‌های استخراج شده باید به راحتی قابل اندازه گیری باشند.
- ویژگی‌های استخراج شده باید منطبق با زمان باشد.
- ویژگی‌های استخراجی باید در برابر نویز و سایر شرایط محیطی مقاوم باشند.

بردارهای ویژگی سیگنال‌های گفتار معمولاً از تکنیک‌های تحلیل طیفی استفاده می‌کنند. تکنیک‌های مختلفی برای استخراج ویژگی در پردازش گفتار استفاده می‌شود که برخی از کاربردی‌ترین این روش‌های استخراج ویژگی عبارتند از: FFT[۱]، LPC[۲]، MFCC[۳]، DWT[۴]، WPT[۵]، DWPD[۶]، PNCC[۷]، PLP[۸] و LSF[۹].

از آنجایی که در این رساله از روش LPC برای یافتن فورمنت‌های فریم‌های واکدار سیگنال گفتار استفاده شده است در ادامه به تشریح این روش استخراج ویژگی و همچنین روش استخراج ویژگی متداول در پردازش گفتار که روش MFCC می‌باشد، پرداخته شده است.

۱-۳-۱- روش استخراج ویژگی MFCC

ویژگی‌های صدای انسان می‌تواند توسط روش MFCC استخراج گردد. ویژگی‌های استخراج شده توسط این روش طیف توان زمان کوتاه صدای انسان را نشان می‌دهد. باندهای فرکانسی در این روش، در مقیاس Mel بصورت یکسان و برابر است چراکه تقریب دقیقی از صدای انسان می‌باشد. برای تبدیل فرکانس نرمال f به مقیاس Mel از رابطه ۱-۱ استفاده می‌شود. ویژگی MFCC در بازشناسی گفتار

---

[۱] Fast Fourier Transform
[۲] Linear Predictive Coding
[۳] Mel Frequency Cepstral Coefficient
[۴] Discrete Wavelet Transform
[۵] Wavelet Packate Transform
[۶] Discrete Wavelet Packet Decomposition
[۷] Power Normalized Cepstral Coefficient
[۸] Perceptual Linear Prediction
[۹] Line Spectral Frequency

بسیار مورد استفاده قرار می‌گیرد اما مشکل اساسی این روش استخراج ویژگی حساسیت زیاد آن به نویز است[۳ و ۴].

$$m = 2595 \log_{10}(1 + \frac{f}{100})$$

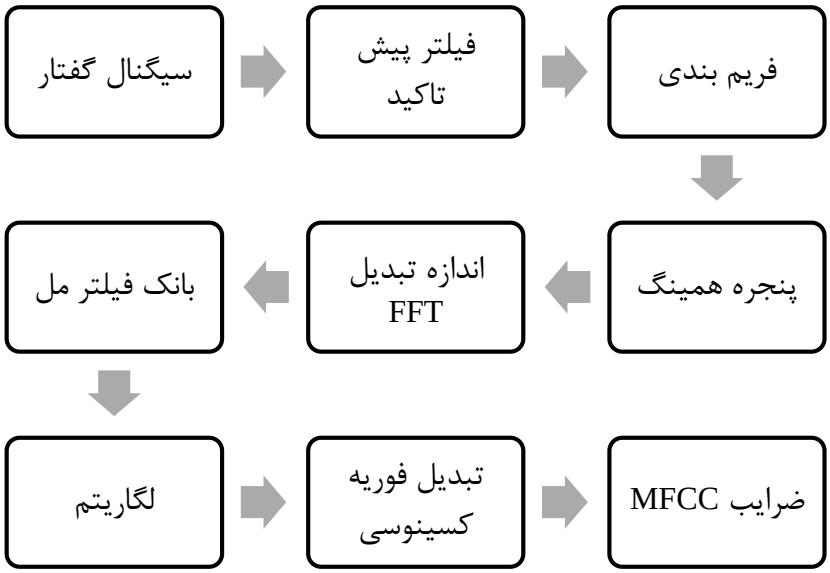
در این روش برای دستیابی به تغییرات بین فریم‌های مختلف از دلتا کپستروم استفاده می‌شود. مزیت دلتا در داشتن تغییرات زمانی حاصل‌جمع انرژی و MFCC بعنوان ویژگی‌های جدید بوده که نشان دهنده‌ی سرعت و شتاب حاصل‌جمع انرژی و MFCC است.

قدم بعدی اعمال DCT بر روی لگاریتم انرژی (E_k) است که از طریق فیلترهای میان‌گذر مثلثی برای داشتن L ضریب کپسترال مقیاس Mel بدست می‌آیند[۴ و ۶].

فرمول DCT بصورت رابطه ۱-۲ می‌باشد:

$$C_m = \sum_{k=1} N \cos\left[m \times (k - 0 \cdot 5) \times \frac{\pi}{N}\right] \times E_k \qquad m = 1.2 \ldots . \mathrm{L} \qquad (۲-۱)$$

در رابطه فوق، N تعداد فیلترهای میان‌گذر مثلثی، L تعداد ضرایب کپسترال مقیاس Mel هستند. مقدار m معمولاً برابر با ۲ در نظر گرفته می‌شود. روند مراحل استخراج ویژگی MFCC در شکل ۱-۲ نشان داده شده است [۴ و ۶]. در الگوریتم MFCC، ابتدا سیگنال گفتار فریم بندی و توسط پنجره همینگ پنجره گذاری می‌شود، سپس اندازه طیف فریم‌های سیگنال گفتار، استخراج شده و این طیف از بانک فیلتر مل عبور می‌کند. در مرحله بعد لگاریتم این طیف ها بدست آورده می‌شود و در نهایت با اعمال تبدیل فوریه گسسته کسینوسی بر روی خروجی‌های بلوک قبل، ضرایب MFCC بدست می آیند [۳].



شکل ۱-۲: بلوک دیاگرام قسمت‌های مختلف الگوریتم استخراج ویژگی MFCC [۳]

بسیاری از توسعه دهندگان بازشناسی گفتار از تکنیک MFCC به منظور استخراج ویژگی بهره می‌برند. روش MFCC تا حد امکان سعی دارد در جایی که فرکانس‌ها به طور غیرخطی در سراسر طیف صوتی حل شده‌اند از گوش انسان تقلید نماید. به همین خاطر از فیلترهای Mel جهت تبعیت از رابطه فضایی توزیع سلول گوش انسان استفاده می‌شود. از این رو مقیاس فرکانسی Mel در فرکانس‌های زیر یک کیلوهرتز بصورت یک مقیاس خطی و برای فرکانس‌های بیش از یک کیلو هرتز بصورت مقیاس لگاریتمی در نظر گرفته می‌شود [۲و۶].

متدولوژی MFCC در جایی که بردار ویژگی به طور مجزا از طریق هر فریم محاسبه می‌شود، مبتنی بر تحلیل زمان کوتاه است.

الف- مزایای تکنیک استخراج ویژگی MFCC

- ایجاد تمایز خوب
- وابستگی کم بین ضرایب
- به دلیل اینکه براساس ویژگی‌های خطی نمی‌باشد همانند سیستم ادراک شنوایی انسان است.

- ویژگی‌های آوایی لازم می‌تواند جمع آوری شود[۴و۸].

ب- معایب تکنیک استخراج ویژگی MFCC

- مقاومت کم در برابر نویز

- از آنجایی که این روش تنها طیف توان را در نظر گرفته و طیف فاز را نادیده می‌گیرد، به همین خاطر یک نمایش محدود را ایجاد می‌کند[۴و۸].

۱-۳-۲- روش استخراج ویژگی LPC

یکی از تکنیک‌هایی که غالباً در طبقه استخراج ویژگی برای تحلیل سیگنال مورد استفاده قرار می‌گیرد، تکنیک LPC می‌باشد. این تکنیک یک تکنیک در حوزه زمان به شمار می‌آید که در آن به تحلیل ساختار چاکنای انسان به منظور ایجاد یک تخمین دقیق از پارامترهای گفتار پرداخته می‌شود. این روش یکی از موفق‌ترین روش‌ها درکاربردهای وسیع امروزی است. این روش، یک روش قوی برای تحلیل گفتارهای کد شده با کیفیت خوب و نمونه‌های با نرخ بیت کم است [۹و۱۰] . مزایای متعددی برای استفاده از روش LPC وجود دارد که از این بین می‌توان به موارد زیر اشاره نمود [۹و۱۰]:

- روش LPC، زمان محاسباتی کوتاه‌تر و کاراتری برای پارامترهای سیگنال ایجاد می‌کند.

- این روش قادر است که مشخصات مهم را از سیگنال ورودی به خوبی استخراج کند.

هدف اصلی این روش، تقریب یک نمونه گفتار دریافت شده بصورت یک ترکیب خطی از نمونه‌های قبلی گفتار می‌باشد. اگر S(i) دامنه در لحظه i ام باشد در اینصورت [۹و۱۰]:

$$S(i) = a_1 * S(i-1) + a_2 * S(i-2) + \cdots + a_p * S(i-p) \qquad (۳-۱)$$

معمولاً مقدار p مقداری بین ۱۰ تا ۱۲ در نظر گرفته می‌شود[۹و۱۰].

در واقع در یک سیستم LTI می‌توان رابطه بین ورودی و خروجی هر سیستم را بصورت یک رابطه تفاضلی به شکل رابطه ۱-۴ ارائه نمود.

$$\sum_{i=0}^{p} a(i)y(n-i) = \sum_{j=0}^{q} b(j)x(n-j) \qquad (۴-۱)$$

در حالت کلی روش پیشگویی خطی LP[۱] تخمین خروجی در یک لحظه خاص به ازای خروجی‌های قبلی و ورودی زمآنهای مختلف می‌باشد. این ارتباط در رابطه ۱-۵ آورده شده است. برای این رابطه می‌توان سه مدل مختلف، مدل تمام قطب (AR[۲])، مدل تمام صفر (MA[۳]) و مدل صفر و قطب (ARMA[۴]) تعریف نمود.

$$y(n) = \sum_{j=0}^{q} b(j)\, x(n-j) - \sum_{i=1}^{p} a(i)y(n-i) \qquad (۵-۱)$$

در رابطه ۱-۵، به a(i) و b(j) ضرایب پیشگویی خطی می‌گویند.

در این رساله ما برای سیگنال گفتار از مدل تمام قطب (AR) استفاده کرده‌ایم. پس با توجه به این نکته می‌توان مدل تولید گفتار یا همان کانال صوتی را بصورت رابطه ۱-۶ بیان نمود.

$$H(z) = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} \qquad (۶-۱)$$

در این روش با مینیمم کردن مجموع مربعات تفاضل (در یک بازه‌ی محدود) بین نمونه‌های گفتار واقعی و مقادیر محاسبه شده، یک مجموعه پارامتر منحصر به فرد یا ضرایب پیش بینی شده بدست می‌آید. مقادیر ضرایب پیش بینی شده، بلوک‌های سازنده LPC گفتار را تشکیل می‌دهند [۹و۱۰]. شکل ۱-۳ مراحل انجام روش استخراج ویژگی LPC را نشان می‌دهد[۹و۱۰].

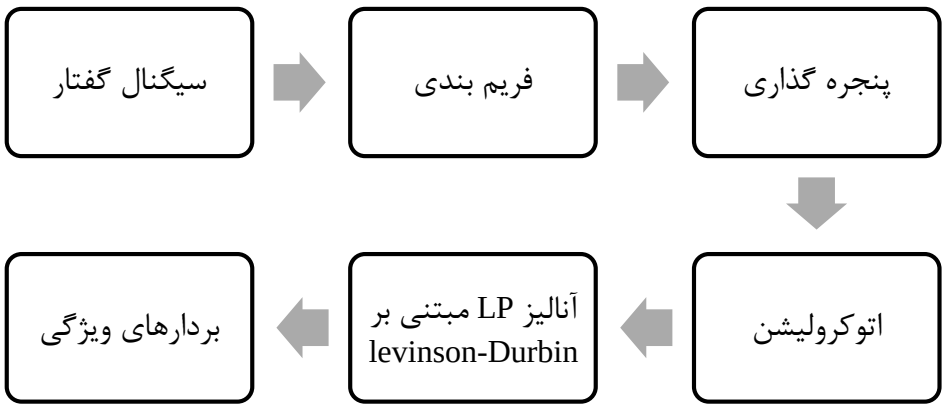
---

[۱] Linear Prediction

[۲] Autoregressive Model

[۳] Moving Average Model

[۴] Autoregressive Moving Average Model



شکل ۱-۳: بلوک دیاگرام قسمت‌های مختلف الگوریتم استخراج ویژگی LPC [۹و۱۰]

در این روش، ابتدا به دلیل خاصیت ناایستا بودن سیگنال گفتار، این سیگنال فریم‌بندی شده و به منظور کاهش اعوجاج ابتدا و انتهای فریم‌ها ناشی از مرحله فریم بندی، پنجره گذاری می‌شوند. سپس تابع خودهمبستگی بر روی این فریم‌ها اعمال می‌گردد و سپس توسط الگوریتم Levinson-Durbin، ضرایب پیشگویی خطی استخراج می‌شود و بردارهای ویژگی LPC را تشکیل می‌دهد.

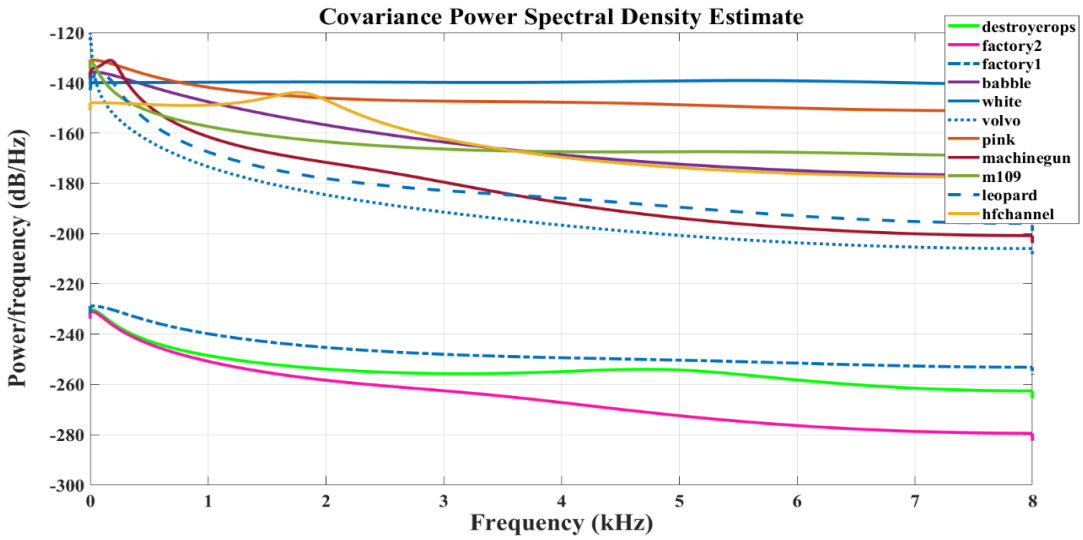
۱-۴- نویزهای مخرب متداول در سیگنال صوتی

همانطورکه در بخش قبل توضیح داده شد، یکی از مهم‌ترین چالش‌هایی که در پردازش گفتار وجود دارد، بحث وجود نویز در سیگنال صوتی می‌باشد. به همین خاطر بررسی میزان تاثیر نویزهای مختلف در پارامترهای یک سیگنال صوتی حیاتی و مهم است. در ادامه این بخش چند نمونه از انواع نویزهای صوتی متداول به صورت مختصر اشاره و تشریح شده است.

- **نویز سفید (White Noise):** مشخصه نویز سفید در داشتن انرژی یکسان در هر هرتز از پهنای باند است. این نویز توسط Wandel و Goltermann با نمونه‌گیری کیفیت بالا از مولد نویز آنالوگ استخراج شده است [۱۱و۱۲].

- **نویز Pink**: این نویز بطور معمول چگالی طیف توان که با نسبت 1/f متناسب است را نشان می‌دهد. این نویز نیز توسط Wandel و Goltermann با نمونه‌گیری کیفیت بالا از مولد نویز آنالوگ استخراج شده است [۱۱و۱۲].
- **نویز Leopard**: این نویز توسط تانک کانادایی Leopard که با سرعت ۱۱۲ کیلومتر بر ساعت در حال حرکت است، ایجاد شده است. سطح صدا در طی فرآیند ضبط برابر با ۱۱۴ دسیبل بوده است [۱۱و۱۲].
- **نویز M109**: نویز M109 نیز مانند نویز Leopard، نویزی می‌باشد که توسط یک تانک M109 که با سرعت ۳۰ کیلومتر بر ساعت در حرکت است، ایجاد شده است. سطح صدا در طی فرآیند ضبط، ۱۰۰ دسیبل بوده است [۱۱].
- **نویز Machine Gun**: این نویز حاصل از یک تیربار کالیبر ۰/۵ می‌باشد که بصورت پی‌درپی در حال تیراندازی است [۱۱].
- **نویز Volvo**: نویز Volvo، نویز حاصل از صدای یک خودروی ۳۴۰ Volvo که با سرعت ۱۲۰ کیلومتر بر ساعت با دنده ۴ در یک جاده آسفالتی در شرایط بارانی حرکت می‌کند، می‌باشد [۱۱].
- **نویز HF Channel**: نویز HF Channel نویز استخراج شده از یک کانال رادیویی HF بعد از عمل دمدولاسیون است [۱۱].
- **نویز همهمه (Babble)**: منبع نویز همهمه، صحبت ۱۰۰ انسان در یک سالن غذاخوری است. از آنجایی که شعاع سالن بیش از ۲ متر می‌باشد، صدای افراد اندکی قابل شنیدن است[۱۱].
- **نویز Factory 1**: منبع ایجاد این نویز، صدای تجهیزات برش صفحه و تجهیزات جوشکاری الکتریکی است[۱۱].
- **نویز Factory 2**: منبع ایجاد این نویز، صداهای موجود در یک سالن تولید خودرو است[۱۱].
- **نویز Destroyer**: منبع ایجاد این نویز، صداهای موجود در موتورخانه یک مجتمع است[۱۱].

در شکل ۱-۴، تابع چگالی طیف توان (PSD) مربوط به هرکدام از نویزهای فوق نمایش داده شده است.



شکل ۱-۴: تابع چگالی طیف توان برای نویزهای مطرح شده [۱۲]

همانطورکه در شکل ۱-۴ به وضوح مشخص است، PSD نویز سفید در تمامی فرکانس‌ها اثر تقریبا برابری دارد. به همین خاطر انتظار داریم که اثر این نویز نسبت به سایر نویزهای مطرح شده شدیدتر باشد. اما نویزهای Volvo و Machine Gun تنها در یک گستره‌ی فرکانسی محدود اثر گذار هستند. به همین خاطر این انتظار نیز وجود دارد که اثر این دو نویز از بقیه کمتر باشد. تایید این فرضیات در قسمت نتایج شبیه سازی بصورت دقیق ارائه خواهد شد.

در طی دو دهه اخیر محققان حوزه پردازش گفتار تلاش زیادی برای بهبود عملکرد سیستم‌های خودکار بازشناسی گفتار (ASR) در شرایط تمیز انجام داده‌اند. مقاوم‌سازی سیستم بازشناسی گفتار نسبت به تنوعات مختلف گفتاری از قبیل تنوع گوینده، لهجه، نویز محیط، کانال انتقال نیز از دیگر حوزه‌های فعال در بحث بازشناسی گفتار است[۱۳و۱۴]. بیشتر تحقیقات انجام شده در زمینه مقاوم‌سازی بازشناسی گفتار نسبت به تنوعات، روی سه تکنیک عمده‌ی بهسازی گفتار، استخراج ویژگی مقاوم و جبرانسازی پارامترهای مدل صوتی متمرکز شده است [۱۴].

سیستم‌های بازشناسی گفتار در شرایط آزمایشگاهی برای داده‌های تمیز به نرخ قابل قبولی از کارایی می‌رسند اما هنگام استفاده از آنها در محیط‌های مختلف و در کاربردهای واقعی، به خاطر وجود انواع مختلف اثرهای محیطی صحت سیستم افت شدیدی می‌کند. شرایط نامناسبی همچون تفاوت میان گوینده‌های مختلف، نویز محیط، میکروفن و اثرات کانال انتقال باعث این عدم کارایی می‌شوند. این مساله در حالت کلی به خاطر عدم تطابقی است که بین حالت آموزش سیستم و حالت استفاده یا آزمون بوجود می‌آید. برای جبران این عدم تطابق و بهبود صحت این سیستم‌ها در شرایط واقعی و همچنین بر حذر داشتن سیستم از نیاز به مجموعه‌ی عظیمی از داده‌های آموزشی، تکنیک‌های مختلفی ارائه شده است. هر کدام از این روش‌ها دارای محدودیت‌هایی هستند و فقط در شرایط محدودی کارا می‌باشند به طوری که برای سیستمی در محیطی خاص که با صحت قابل قبولی عمل می‌کند، ممکن است در محیطی دیگر قادر به رسیدن به آن صحت نباشد. روش‌های بهبود گفتار سعی در حذف نویزهای موجود در سیگنال گفتار قبل از وارد شدن این سیگنال به سیستم بازشناسی دارند. این روش‌ها، اگرچه در اغلب موارد بار محاسباتی سیستم را چندان بالا نمی‌برند اما معمولاً دارای کارایی و عملکرد محدودی هستند. در روش‌های استخراج ویژگی‌های مقاوم سعی می‌شود با توجه به تأثیر نویز بر خود سیگنال و یا ویژگی‌های آن، ویژگی‌هایی از سیگنال گفتار استخراج شود که نویز بر آنها بی‌اثر یا کم‌اثر باشد. با این کار بردارهای ویژگی مشابهی برای نمونه‌های گفتار آموزشی و آزمایشی به دست می‌آید و در نتیجه سیگنال گفتار به پارامترهایی ذاتاً مقاوم در مقابل نویز تبدیل می‌شود. بدین منظور تأثیر محیط بر مشخصه‌های سیگنال گفتار را از طریق تأثیر آنها بر خود سیگنال و یا ویژگی‌ها مشخص می‌نمایند. رویکردهای مبتنی بر مدل سعی در مقاوم سازی مدل‌های گفتار با نزدیک کردن مدل‌های موجود از گفتار تمیز به گفتار نویزی دارند. این روش‌ها نیاز به داده‌های تطبیق از محیط جدید و همچنین بار محاسباتی بالاتری نسبت به سایر روش‌ها دارند که معمولاً به بهبود بهتری از صحت در مقایسه با سایر روش‌ها می‌انجامند [۱۵و۱۶].

۱-۵- اهداف رساله

همانطورکه در بخش‌های قبل تشریح شد، یکی از چالش‌های اساسی در بازشناسی گفتار وجود نویز در محیط است که این نویز به شدت باعث کاهش صحت بازشناسی در محیط‌های واقعی می‌گردد. در این رساله به منظور بهبود عملکرد سیستم بازشناسی گفتار در محیط‌های نویزی که با نام سیستم‌های بازشناسی گفتار مقاوم معرفی می‌شوند، روش‌های استخراج ویژگی مقاومی پیشنهاد شده است. در این رساله سه روش استخراج ویژگی مقاوم پیشنهاد داده شده است. این روش‌های استخراج ویژگی پیشنهادی مبتنی بر تبدیل فوریه کسری زمان کوتاه و همچنین تابع ریشه است و این ساختارها در فصل چهارم و پنجم بطور کامل تشریح و مورد ارزیابی قرار داده شده‌اند.

۱-۶- ساختار رساله

ادامه این رساله به فصل‌های زیر تقسیم شده است. در فصل دوم مروری بر کارهای گذشته که در حوزه مقاوم سازی بازشناسی گفتار طی سالیان اخیر توسط سایر محققین در این زمینه صورت پذیرفته است، خواهیم داشت. فصل سوم این رساله به بخشی از تئوری‌های مورد نیاز به منظور طراحی سیستم‌های بازشناسی گفتار مقاوم پرداخته است. در فصل چهارم، در ابتدا به بررسی میزان تاثیر نویزهای مختلف بر روی جابه‌جایی فورمنت‌های فریم‌های واکدار سیگنال گفتار پرداخته شده است و سپس الگوریتم پیشنهادی(Fractional Root Coefficients (FrRC تشریح و مورد ارزیابی نتایج شبیه سازی قرار گرفته و در نهایت الگوریتم استخراج ویژگی Adaptive Fractional Power Normalized Cepstral Coefficients (AFPNCC) مورد بررسی و تحلیل قرار گرفته است و در نهایت در فصل پنجم رساله، نتیجه‌گیری کلی از کارهای انجام شده و همچنین کارهایی که می‌تواند در ادامه این رساله انجام گیرد، صورت پذیرفته است.

فصل دوم: مروری بر کارهای گذشته

۲-۱- مقدمه

برای کاربردهای عملی، مقاومت در برابر نویز به یک چالش بسیار بزرگ بدل شده است، چرا که ASR ملزم به کارکردن در محیط‌های بسیار دشوار آکوستیک است[۴]. تکنیک‌های زیادی به منظور کاهش حساسیت ویژگی‌ها به نویزهای خارجی و موجود در محیط پیشنهاد شده است. در برخی از دیدگاه‌ها، به منظور جبران سازی تاثیرات نویز یک تبدیل مستقیم بر روی بردارهای ویژگی اعمال می‌گردد [۱۷]. برخی مواقع، تبدیل بر روی حوزه کپستروم از قبیل نرمال‌سازی میانگین کپسترال (CMN[۱]) صورت می‌پذیرد [۱۸]. در انواع دیگر این چنین تکنیک‌هایی، تبدیل بر روی لگاریتم طیف یا لگاریتم انرژی‌های بانک فیلتر(LFBE[۲]) از قبیل بردار سری‌های تیلور [۱۸] و تحلیل‌های بانک فیلتر مل وزن دار اعمال می‌شود [۱۹و۲۰].

برخی دیگر از روش‌ها در سطح طیف کار می‌کنند. این روش‌ها سعی در کاهش تاثیر نویز جمع شونده بر روی طیف گفتار و سپس ویژگی‌های استخراجی دارند. تفریق طیفی [۱۸و۲۱] و تکنیک‌های متفاوت فیلترینگ طیف از روش‌های شناخته شده این تکنیک‌ها می‌باشند. تفریق طیفی، یک تخمین از طیف نویز را از طیف توان گفتار به منظور حذف تاثیر نویز، کم می‌کند. خودهمبستگی فاز (PAC[۳]) یک نمونه‌ی دیگر از این تکنیک‌ها است که اخیراً معرفی شده است [۲۲]. در این روش هدف ایجاد ضرایب خودهمبستگی با حساسیت کم نسبت به نویز جمع شونده می‌باشد [۲۲و۲۳].

نمونه‌هایی از کارهای انجام شده در حوزه بازشناسی مقاوم گفتار در محیط‌های نویزی در ادامه تشریح شده است.

---

[۱] Cepstral Mean Normalization

[۲] Log Filterbank Energy

[۳] Phase Autocorrelation

۲-۲- تفاضل میانگین ضرایب کپسترال (CMS[۱])

این روش به عنوان یک شبه استاندارد جهت مقاوم‌سازی ویژگی‌های کپسترال بکار می‌رود. نویز پیچشی در حوزه‌ی زمان معادل نویز جمع‌شونده در حوزه‌ی طیف و کپسترال است. مقاوم‌سازی این روش با کم کردن مقدار میانگین ضرایب کپسترال از مقادیر این ضرایب صورت می‌گیرد. علت این کار تغییر مقدار میانگین ویژگی‌ها در حضور نویز جمع‌شونده در حوزه کپسترال است [۲۴]. بردار ویژگی تفاضل میانگین ضرایب کپسترال از روی N عنصر موجود در بردار ضرایب کپسترال و گفتار عبور کرده از کانال با کم‌کردن میانگین $\mu_y = \frac{1}{N}\sum_{i=1}^{N} C_y(i)$ از هر یک از مولفه‌های بردار اصلی کپسترال به صورت رابطه ۲-۱ بدست می‌آید تا بردار ویژگی‌ها به میانگین نرمال سازی شوند[۲۴].

$$\widetilde{C_y} = C_y - \mu_y \quad . \quad y = 1. \ldots . T \tag{۱-۲}$$

در رابطه فوق، y، T، C_y و $\widetilde{C_y}$ به‌ترتیب بیانگر شماره فریم، تعداد فریم ها، ضرایب کپسترال فریم yام و ضرایب CMS فریم yام هستند.

۲-۳- تحلیل اجزای اصلی (PCA[۲])

تحلیل اجزای اصلی یا PCA، یک روش شناخته شده آماری در کاهش تعداد ویژگی‌ها و خوشه بندی می‌باشد. هدف این روش این است که با تبدیلی خطی و معکوس‌پذیر روی ویژگی‌ها، آنها را به فضای جدیدی ببرد که در آن فضا، ویژگی‌های جدید دارای بیشترین تغییرات نسبت به همدیگر باشند. به عبارتی، این روش در جستجوی بردارهایی است که خطای نگاشت به فضای جدید را مینیمم و واریانس داده‌های بعد از نگاشت را ماکزیمم نماید و به نوعی بردارهای فضای جدید را از همدیگر مستقل کند. این تبدیل به صورت رابطه ۲-۲ روی بردار N بعدی X اعمال می‌شود [۲۵].

---

۱ Cepstral Mean Subtraction

۲ Principal Component Analysis

$$Y = T(X - \mu_x) \qquad (۲-۲)$$

در رابطه ۲-۲، T ماتریس تبدیل مورد نظر می‌باشد که از N بردار ویژه مربوط به ماتریس همبستگی داده‌های آموزشی با مقادیر ویژه به صورت زیر بدست آمده است:

$$T = [\emptyset_1 \emptyset_2 \dots \emptyset_N]^T \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \qquad (۳-۲)$$

خاصیت اصلی این تبدیل، متعامد کردن بردارهای نگاشت‌شده توسط تبدیل است. با استفاده از تئوری ماتریس‌ها می‌توان نشان داد که میانگین بردار Y برابر صفر و ماتریس کواریانس آن نیز قطری و مقادیر روی قطر اصلی این ماتریس، برابر مقادیر ویژه بدست آمده از ماتریس کواریانس X است [۲۵]. یکی دیگر از توانایی‌های عمده این تبدیل کاهش بعد و فشرده سازی است. از آنجا که سطرهای ماتریس تبدیل T به ترتیب میزان واریانس داده‌ها قرار گرفته‌اند و به ترتیب، از بالا به پایین، از اهمیت آنها کاسته می‌شود، می‌توان به جای استفاده از کل ماتریس N*N از نمونه کاهش یافته‌ی آن استفاده کرد و سطرهای پایینی آن را که دارای اهمیت کمتری اند حذف کرد تا ماتریسی R*N (R≤N) از آن باقی بماند که فقط از R نمونه اول و با اهمیت‌تر آن که متناسب با مقادیر ویژه بزرگترند، استفاده شود. با این کار در واقع از اجزای کم اهمیت‌تر داده‌ها صرف نظر شده است و داده‌ها کاهش بعد می‌یابند[۲۵].

استفاده از این تبدیل در استخراج ویژگی‌های مبتنی بر MFCC می‌تواند به سه صورتِ زیر باشد: به عنوان آخرین بلوک مرحله استخراج ویژگی‌ها، بعد از تبدیل گسسته کسینوسی و به جای تبدیل گسسته کسینوسی باشد [۲۶و۲۷] که نتایج بکارگیری آن در این قسمت‌ها، نشان دهنده‌ی بهتر بودن استفاده از این روش به عنوان آخرین بلوک مراحل استخراج ویژگی است [۲۶]. در [۲۶و۲۷] نشان داده شده است که تجمیع روش PCA با روش استخراج ویژگی MFCC باعث افزایش مقاومت ویژگی‌های استخراجی می‌شود.

۲-۴- نرمال‌سازی طول مسیر صوتی گوینده (VTLN[۱])

یکی از منابعی که باعث کاهش کارایی و ناسازگاری در سیستم‌های بازشناسی می‌شود، اختلاف موجود بین گوینده‌های مختلف است. این مساله نه تنها ناشی از اختلاف موجود در دستگاه‌های فیزیکی و فیزیولوژیکی متفاوت آنهاست، بلکه ناشی از مسایل دیگری همچون تغییرات زبانی بین گوینده‌های مختلف مانند نحوهٔ تلفظ، لهجه، استرس و . . . می‌باشد. یکی از منابع اصلی اختلاف بین گوینده‌های متفاوت، طول مسیر صوتی (VTL) می‌باشد. این اختلاف باعث می‌شود که صدای افراد مختلف از نظر زیر و بمی در ظاهر با هم متفاوت باشد، مساله‌ای که در مقایسهٔ گفتار یک بچه با یک فرد بالغ یا یک زن با یک مرد کاملاً مشهود است. روش نرمال‌سازی طول مسیر صوتی یا VTLN از روش‌های رایج برای از بین بردن و یا حداقل کم اثر کردن اختلاف ناشی از طولهای مختلف مسیر صوتی در افراد مختلف است [۲۸و۲۹].

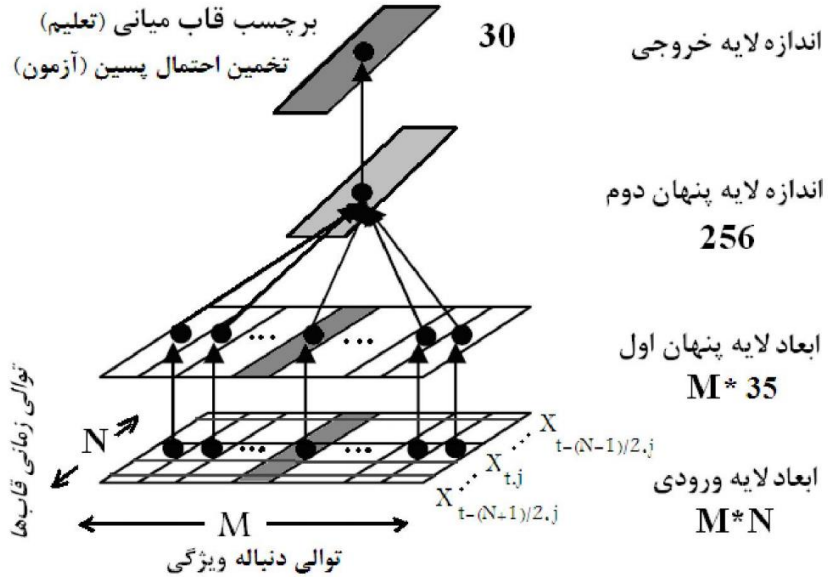
۲-۵- استفاده از ویژگی الگوی زمانی به دست آمده از ساختار شبکه عصبی بهینه شده MTMLP

در [۳۰]، ویژگی‌های الگوهای زمانی با استفاده از خروجی مقدار احتمال پسین واجی ساختار بهینه شده شبکه عصبی MTMLP، از مجموعه بردارهای بازنمایی مبتنی بر طیف (مانند ویژگی گفتاری LFBE) و همچنین مبتنی بر کسپتروم (مانند ویژگی گفتاریMFCC) استخراج شده است. با ترکیب اطلاعات الگوهای زمانی (دینامیک زمان بلند) به دست آمده از حوزه‌های لگاریتم طیف و کپستروم به بردار ویژگی‌های پایه بازشناسی، شامل ویژگی‌های گفتاری متداول MFCC و مشتقات زمانی اول و دوم آن (دینامیک زمان کوتاه)، نشان داده شده است که صحت بازشناسی واج در شرایط دادگان آزمون تمیز، حدود ۱ درصد نسبت به نتایج بهترین سیستم پایه بازشناسی بهبود یافته است. این در حالی است

[۱] Vocal Tract Length Normalization

که ویژگی‌های به دست آمده از روش پیشنهادی این مقاله، صحت بازشناسی بالاتری را در شرایط نویزی مختلف (تا حدود ۱۳ درصد) نتیجه داده است و این بیانگر مقاومت بالای این الگوریتم پیشنهادی در برابر نویز می‌باشد.

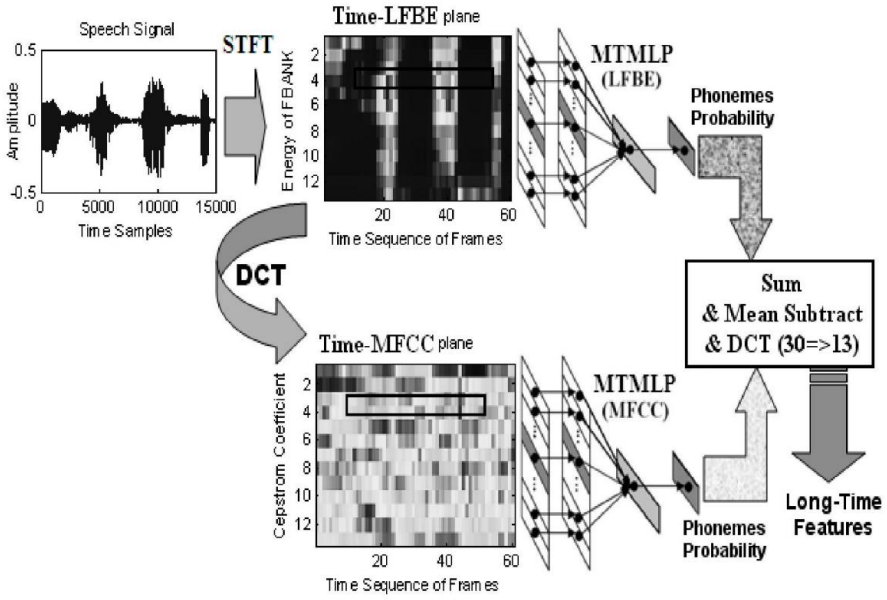
یکی از روش‌های مناسب برای استخراج ویژگی الگوهای زمانی، استفاده از مدل شبکه عصبی می‌باشد. در مقاله [۳۰] از مدل TMLP [۳۱] که ساختار آن مبتنی بر بخش تونوتوپیک سیستم شنوایی انسان است، در جهت استخراج ویژگی الگوهای زمانی استفاده شده است. این مدل در شکل ۲-۱ نشان داده شده است. در واقع این مدل تنها از یک طبقه MLP چهار لایه استفاده کرده است. بنابراین در این مدل، آموزش تنها با یک مرحله تعلیم صورت می‌پذیرد.



شکل ۲-۱: ساختارمدلTMLP با توالیN قاب زمانی وبرداروِیژگی ورودیM بعدی برای هرقاب‌زمانی [۳۰]

با توجه به دیدگاه الگوهای زمانی، از آن جا که اطلاعات موجود در ورودی مدل شبکه عصبی TMLP، به تعداد زیادی از بردارهای بازنمایی مربوط هستند، بنابراین تعلیم مناسب‌تر نگاشت شبکه عصبی پیشنهاد گردیده است. در این الگوریتم علاوه بر استفاده از برچسب واج قاب میانی ورودی، از اطلاعات واجی مربوط به قاب‌های قبل و بعد قاب میانی نیز در لایه خروجی شبکه استفاده گردیده است. در این روش، تعداد نرون‌های لایه خروجی به نسبت ساختار TMLP افزایش یافته است. بنابراین در هنگام

محاسبه مقادیر احتمالات پسین، می‌توان میانگین وزن داری از احتمالات مربوط به قاب‌های قبل و بعد را به قاب میانی افزود. این مدل بهبود یافته TMLP که با MTMLP معرفی شده است، می‌تواند به هموار سازی نتایج احتمالاتی خروجی کمک کند. نتایج میزان صحت بازشناسی قاب با استفاده از دو مدل TMLP و MTMLP توسط بردارهای ویژگی MFCC در دادگان آزمون تمیز، گویای این مطلب است که ساختار MTMLP توانسته است صحت بازشناسی بالاتری را سبب گردد. از آنجایی که استفاده همزمان بردارهای ویژگی متمایز که حاوی اطلاعات متفاوتی از یک سیگنال هستند، می‌تواند به افزایش کارایی سیستم‌های بازشناسی منجر شود [۳۲و۳۳]، در [۳۰] یک الگوریتم پیشنهادی از ترکیب ویژگی-های حوزه کپستروم(MFCC) و حوزه طیف (LFBE) ارائه شده است. این الگوریتم پیشنهادی که برای استخراج ویژگی الگوهای زمانی ارائه شده در شکل (۲-۲) نشان داده شده است [۳۰].



شکل ۲-۲: روش پیشنهادی برای استخراج ویژگی الگوهای زمانی [۳۰]

در [۳۰]، ابتدا با پیشنهاد مدل بهبود یافته‌ی شبکه عصبی MTMLP، نتایج تشخیص احتمالات واجی مدل شبکه عصبی TMLP بهبود داده شد. سپس با استفاده از تعریف مدل ترکیبی، اطلاعات الگوهای زمانی به دست آمده از مجموعه بردارهای ویژگی MFCC و LFBE ترکیب شده و نشان داده شده است که با استفاده از این الگوریتم پیشنهادی، تشخیص احتمالات واجی بهبود یافت.

۲-۶- استفاده از ویژگی‌های هارمونیک

مقاله [۳۵] با عنوان بازشناسی مقاوم گفتار با استفاده از ویژگی‌های هارمونیک در سال ۲۰۱۳ ارائه گردید. نویسندگان این مقاله یک سیستم بازشناسی گفتار با استفاده از اطلاعات مربوط به ساختار هارمونیک برای شناسایی ویژگی‌های هارمونیک در محیط‌های نویزی پیشنهاد دادند. الگوریتم پیشنهادی ابتدا عناصر هارمونیکی موجود در داخل سیگنال‌های گفتار را با استفاده از کانولوشن تابع سینوسی استخراج می‌کند. با تنظیم فرکانس تابع سینوسی برابر با فرکانس اصلی سیگنال گفتار، اجزای هارمونیک می‌توانند استخراج گردند. سیگنال بازسازی شده به دست آمده با جمع آوری اجزای هارمونیک استخراج شده دارای درجه بالایی از همبستگی با سیگنال اصلی بدست آمده است. اندازه انرژی فریم استخراج شده از اجزای هارمونیک، تحت پردازش بیشتری قرار گرفته تا ویژگی‌های هارمونیک دینامیکی را تغییر داده و سپس به همراه سایر روش‌های استخراج ویژگی از قبیل MFCC و PLP بعنوان یک بردار ویژگی در نظر گرفته شود. نتایج پیاده سازی این الگوریتم نشان داد، که الگوریتم پیشنهادی صحت بازشناسی مقاوم‌تر و بیشتری را در محیط‌های نویزی حاصل می‌کند [۳۵].

اطلاعات فرکانس گام هر فریم توسط الگوریتم تخمین فرکانس گام، ETSI[۱] استخراج می‌گردد. رابطه ۲-۴ به منظور استخراج اجزاء هارمونیک از فریم پنجره‌گذاری شده همینگ مورد استفاده قرار گرفته است [۳۵].

$$VH(t_0.i) = \sum_{i=0}^{B}[f(i)V(t_0.i)] \qquad (۴-۲)$$

تمامی فریم‌های گفتار توسط این رابطه با پریود فرکانس گام(T) متفاوت مورد پردازش قرار می‌گیرد. اگر یک فریم گفتاری خاص بعنوان فریم واکدار طبقه بندی شود، اطلاعات فرکانس گام شناسایی شده بعنوان پریود فرکانس گام(T) استفاده می‌شود. اگر این فریم‌ها بعنوان فریم‌های بی‌واک طبقه بندی گردد، پریود فرکانس گام(T) بعنوان میانگین تمامی دوره‌های فرکانس گام شناسایی شده در داخل آن سیگنال گفتار

[۱] European Telecommunication Standards Institute

خاص تنظیم می‌شود [۳۵]. اجزای هارمونیک در هر فریم استخراج می‌شود. انرژی فریم لگاریتمی پس از استخراج اجزای هارمونیک در هر فریم محاسبه می‌گردد. اندازه گیری انرژی فریم لگاریتمی هر یک از فریم هایی که اجزا هارمونیکی آنها استخراج شده اند توسط رابطه ۲-۵ بدست می‌آید [۳۵].

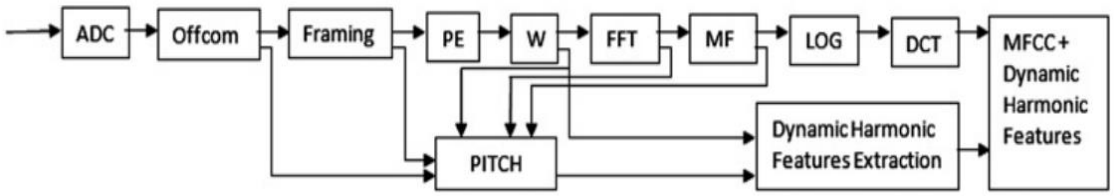
$$EHC(x.m) = Ln(\sum_{t=1}^{M'} HC(m.t^2)) \quad (۵-۲)$$

در رابطه فوق، EHC بیانگر میزان انرژی لگاریتمی فریم x ام و m امین جز هارمونیک نمونه های M′ می‌باشد. یک مقدار کف در محاسبه انرژی استفاده شده که اطمینان حاصل می‌کند که نتیجه برای EHC کمتر از ۱/۶ نیست. ویژگی‌های هارمونیک دینامیک (DHF[۱]) توسط رابطه ۲-۶ بدست می‌آید [۳۵].

$$DHF(m) = \frac{EHC(x+1.m) - EHC(x-1.m)}{2} \quad (۶-۲)$$

این ویژگی با ویژگی‌های MFCC، PLP یا RASTA-PLP به منظور ایجاد ویژگی‌های ترکیبی، تجمیع می‌شوند.

در شکل ۲-۳ بلوک دیاگرام الگوریتم استخراج ویژگی‌های MFCC و DHF نشان داده شده است [۳۵].



شکل ۲-۳: بلوک دیاگرام الگوریتم استخراج ویژگی‌های MFCC و DHF [۳۵]

نتایج پیاده‌سازی این الگوریتم پیشنهادی گویای این است که روش استخراج ویژگی MFCC+DHF بهترین صحت بازشناسی سیگنال‌های گفتار را در محیط‌های بدون نویز و روش استخراج ویژگی RASTA-PLP+DHF بیشترین مقاومت را در سیستم بازشناسی گفتار از خود نشان می‌دهد. اگرچه

[۱] Dynamic Harmonic Frequency

استخراج هارمونیک با استفاده از مدل هارمونیک ثابت کرده است که در بهبود نرخ بازشناسی گفتار الگوریتم پیشنهادی موفق است، اما مدل هارمونیک فقط می‌تواند سیگنال‌های واکدار گفتار بیان شده را نشان دهد، در حالی که دیگر مدل‌های گفتار مانند مدل HNM[1]، می‌تواند هر دو سیگنال واکدار و بی‌واک گفتار را نمایش دهد [۳۵].

۲-۷- استفاده از یک سرکوبگر نویز کپسترال MMSE

نویسندگان [۳۶] یک سیستم بازشناسی مقاوم گفتار با استفاده از یک سرکوبگر نویز کپسترال MMSE[2] در سال ۲۰۰۸ ارائه دادند. در این مقاله یک الگوریتم غیرخطی کارآمد و موثر سرکوب نویز که با استفاده از معیار بهینه‌سازی حداقل مربع خطا (MMSE)، برای بازشناسی گفتار مقاوم در برابر نویز در نظر گرفته شده است. الگوریتم پیشنهاد شده در این مقاله با هدف به حداقل رساندن خطا برای کپسترال فرکانس مل به جای طیف تبدیل فوریه گسسته مورد استفاده قرار گرفته است که در خروجی بانک فیلتر مل عمل می‌کند. به عنوان یک نتیجه، آمار استفاده شده برای تخمین عامل سرکوب، نسبت به سرکوبگر E&M log-MMSE[3] بسیار متفاوت می‌باشد. الگوریتم پیشنهادی در این مقاله به طور قابل توجهی کارآمدتر از سرکوبگر E&M log-MMSE است. چراکه تعداد کانال‌ها در بانک فیلتر فرکانس مل بسیار کوچکتر (۲۳ مورد در این مقاله) از تعداد ضرایب DFT (۲۵۶ ضریب) است. در این مقاله از دادگان Aurora-3 استفاده شده است[۳۶]. نتایج تجربی حاکی از کاهش ۴۸ درصدی خطای بازشناسی WER[4] نسبت به سطح استاندارد ICSLP02، ۲۶ درصد نسبت به CMN و ۱۳ درصد نسبت به روش سرکوبگر نویز E&M's log-MMSE می‌باشد. ساختار این الگوریتم‌ها در شکل‌های ۲-۴ تا ۲-۹ نشان داده شده است [۳۶].

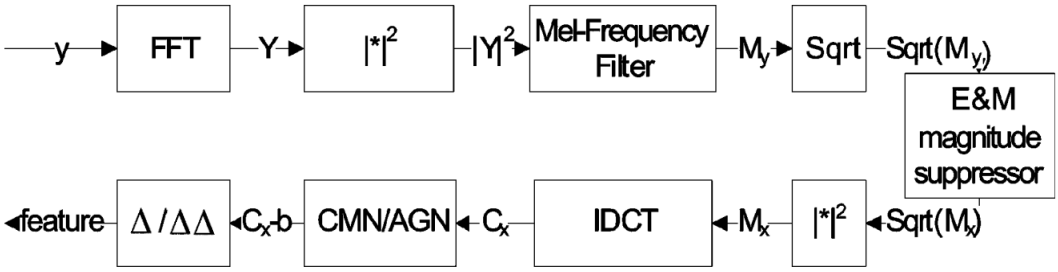
---

[1] Harmonic plus Noise Model

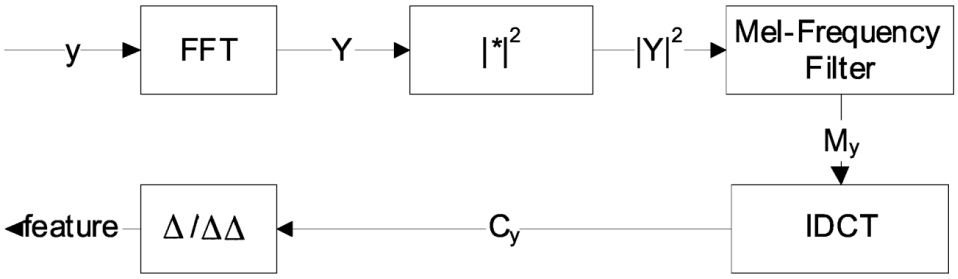
[2] Minimum Mean Square Error

[3] Ephraim and Malah

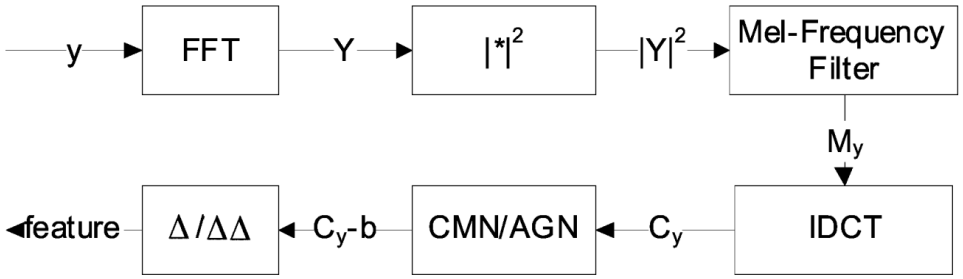
[4] Word Error Rate



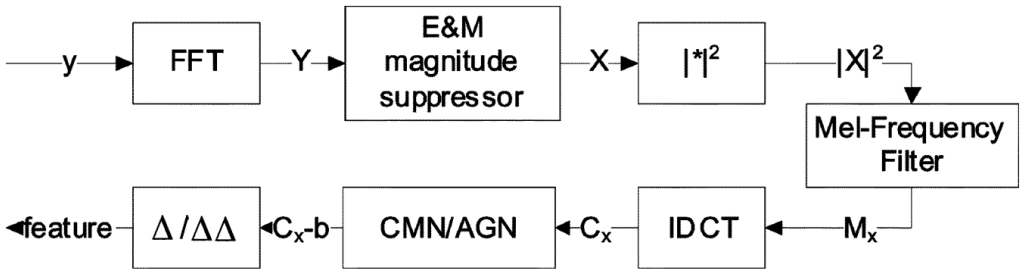
شکل ۲-۴: استخراج ویژگی موازی که دامنه سرکوبگر E&M's log-MMSE بصورت مستقیم به طیف دامنه خروجی بانک فیلتر اعمال شده است [۳۶]



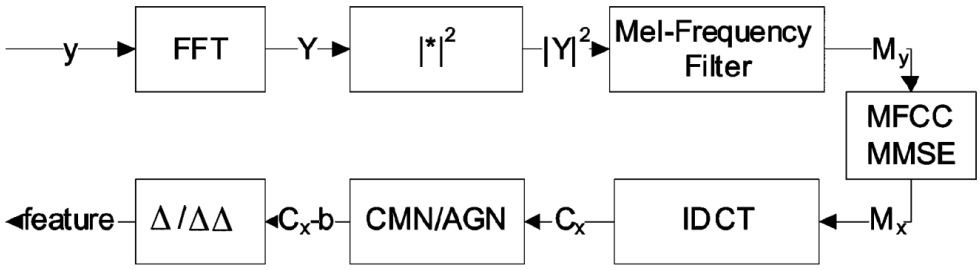
شکل ۲-۵: استخراج ویژگی موازی برای سیستم پایه ICSLP02 [۳۶]



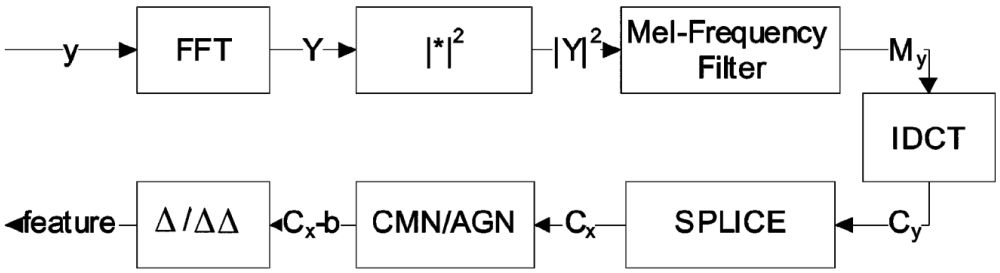
شکل ۲-۶: استخراج ویژگی موازی برای سیستم پایه CMN [۳۶]



شکل ۲-۷: استخراج ویژگی موازی برای سیستم E&M's log-MMSE [8] که سرکوب کننده به ضرایب DFT اعمال شده است [۳۶]



شکل ۲-۸: استخراج ویژگی موازی برای سیستم MFCC-MMSE [۳۶]



شکل ۲-۹: استخراج ویژگی موازی برای سیستم‌های SPLICE [۳۶]

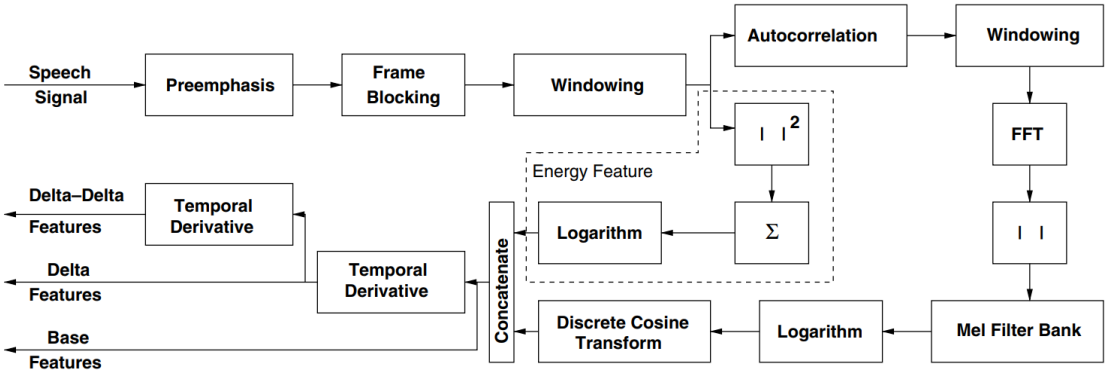
همانطورکه بیان شد در این مقاله یک الگوریتم غیرخطی کاهش نویز توسط معیار MMSE در حوزه MFCC برای بازشناسی گفتار اتوماتیک مقاوم پیشنهاد گردید [۳۶].

۲-۸- روش استخراج ویژگی AMFCC[۱]

در [۳۷] یک روش استخراج ویژگی مقاوم در برابر نویز جمع شونده پس زمینه برای کاربرد بازشناسی اتوماتیک گفتار پیشنهاد شده است. نویز پس زمینه باعث تخریب ضرایب خودهمبستگی سیگنال گفتار می‌شود و اغلب برای تخمین طیف، بیشترین تاثیر را روی ضرایب تاخیرهای پایین و کمترین تاثیر را روی ضرایب اتوکرولویشن با تاخیرهای بیشتر دارد. دامنه طیف دنباله خودهمبستگی با تاخیر بیشتر پنجره‌گذاری شده در اینجا بعنوان یک تخمین از طیف توان سیگنال گفتار استفاده شده است. تخمین طیف توان از طریق فیلتر بانک مل، عملگر لگاریتم و DCT برای دستیابی به ضرایب کسپترال، پردازش بیشتری شده است. این ضرایب کپسترال بعنوان ضرایب کپسترال فرکانسی مل خودهمبستگی معرفی

[۱] Autocorrelation Mel Frequency Cepstral Coefficient

گردیده‌اند. در این مقاله عملکرد بازشناسی گفتار ویژگی‌های AMFCC بر روی دادگان Aurora نشان داد که این روش استخراج ویژگی در محیط‌های بدون نویز شبیه روش استخراج ویژگی MFCC عمل می‌کند اما عملکرد بازشناسی روش استخراج ویژگی AMFCC در محیط‌های نویزی بهتر از روش استخراج ویژگی MFCC می‌باشد. بلوک دیاگرام روش استخراج ویژگی AMFCC بصورت شکل ۲-۱۰ می‌باشد [۳۷].



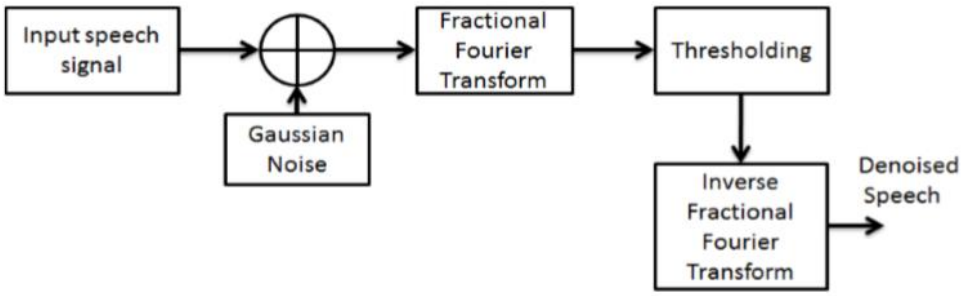
شکل ۲-۱۰: بلوک دیاگرام روش استخراج ویژگی AMFCC [۳۷]

همانطورکه در این شکل مشاهده می‌شود، ابتدا سیگنال گفتار ورودی وارد فیلتر پیش تاکید با تابع تبدیل $H(z) = 1 - 0/97z^{-1}$ می‌شود. به منظور تحلیل زمان کوتاه سیگنال گفتار پیش پردازش شده، از فریم‌بندی با طول ۳۲ میلی ثانیه و انتقال زمانی ۱۰ میلی ثانیه استفاده شده است. بعد از اعمال پنجره همینگ، فریم‌های پنجره‌گذاری شده به بلوک خودهمبستگی به منظور ایجاد دنباله خودهمبستگی با طول ۳۲ میلی ثانیه اعمال می‌گردد. بعد از این کار تاخیرهای زمانی کمتر از ۲ میلی ثانیه حذف شده و پنجره DDR به سایر ضرایب باقی‌مانده که از ۲ میلی ثانیه تا ۳۲ میلی ثانیه می‌باشند (دنباله خودهمبستگی تاخیر بالاتر) اعمال می‌گردد. در این مقاله طیف دامنه دنباله خودهمبستگی پنجره‌گذاری شده محاسبه شده و بعنوان تخمین طیف توان سیگنال مورد استفاده قرار می‌گیرد. این طیف سپس توسط بانک فیلتر مل، لگاریتم و عملگر DCT برای دستیابی به ۱۲ ضریب AMFCC مورد پردازش قرار می‌گیرد. علاوه بر این لگاریتم انرژی سیگنال گفتار پنجره گذاری شده محاسبه شده و به ۱۲ ضریب AMFCC به منظور دستیابی به ۱۳ ضریب افزوده می‌گردد [۳۷].

در این روش ضرایب MFCC از دنباله خودهمبستگی سیگنال نویزی بعد از حذف برخی از ضرایب تاخیر کمتر (زمان کمتر) استخراج گردیده است [۳۸]. این ضرایب تاخیر کمتر نشان داد که بیشترین تاثیر نویز بر روی این ضرایب پایین می‌باشد. مقدار آستانه تاخیر ۳ میلی ثانیه بود. همانطور که در این مقاله گزارش شده است، این روش برای نویزهای Car و Subway با استفاده از پایگاه داده Aurora2 کار می‌کند اما برای نویزهای Babble و exhibition استفاده نشده است. دلیل اصلی موفقیت محدود AMFCC در نویزهایی مانند Babble و exhibition این بود که مشخصات خودهمبستگی نویز خیلی شبیه به سیگنال گفتار است که باعث می‌شود که جداسازی آنها بسیار مشکل باشد [۳۸].

۲-۹- حذف نویز با استفاده از تبدیل فوریه کسری

در مقاله [۳۹] یک روش به منظور حذف و یا کاهش نویز موجود در گفتار پیشنهاد شده است. بلوک دیاگرام فرآیند پیشنهادی در این مقاله در شکل ۲-۱۱ نشان داده شده است. در این روش، آستانه سخت بر روی طیف دریافتی بعد از تبدیل فوریه کسری اعمال گردیده است [۳۹]. بعد از آستانه گذاری، معکوس تبدیل فوریه کسری به منظور دستیابی به سیگنال گفتار بدون نویز اعمال شده است.



شکل ۲-۱۱: بلوک دیاگرام فرآیند حذف نویز [۳۹]

در این مقاله نویز سفید گاوسی مورد بررسی قرار گرفته است که بر روی سیگنال گفتار اثر می‌گذارد. نتایج شبیه سازی با مقادیر مختلف SNR[1] و آستانه‌گذاری سخت حاکی از بهبود ۳/۶ دسیبل مقدار PSNR[2] با اعمال تبدیل فوریه کسری و آستانه‌گذاری است.

۲-۱۰-استفاده از یک چارچوب تعمیم یافته به منظور بازشناسی مقاوم گفتار

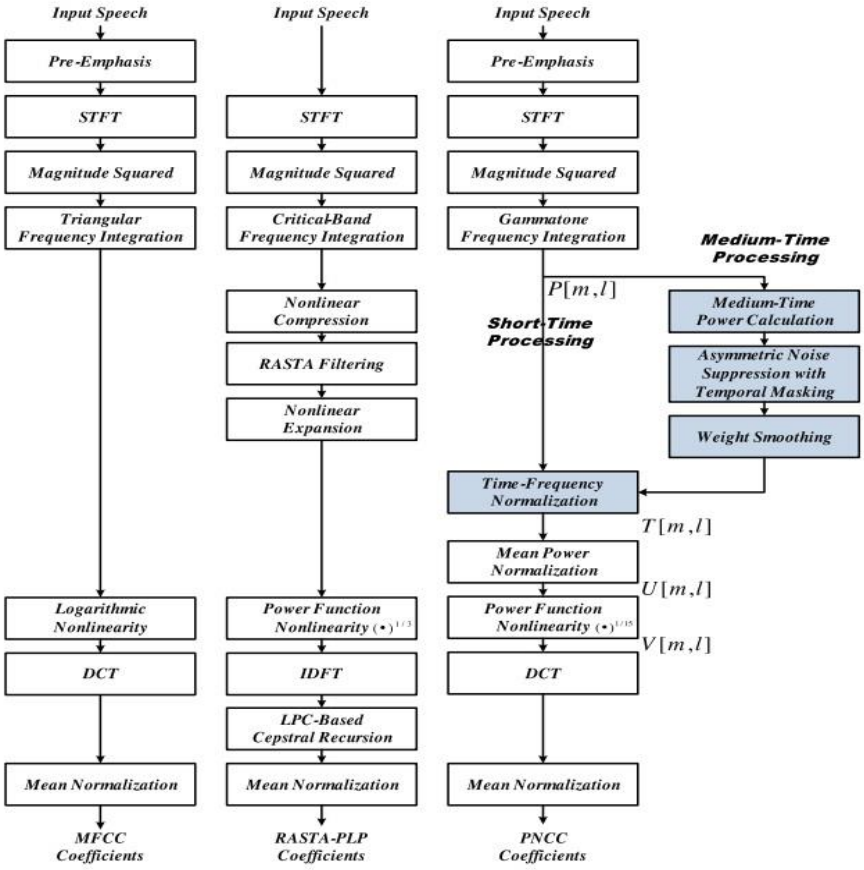
در مقاله [۴۰] یک سیستم مقاوم در برابر نویز برای بازشناسی گفتار با استفاده از یک چارچوپ تعمیم یافته ارائه شده است. در این سیستم فرض بر این است که در طول آموزش، علاوه بر داده‌های آموزشی، نوعی اطلاعات "ممتاز" موجود است و می‌تواند برای هدایت روند آموزش استفاده شود. این امر باعث بهبود عملکرد سیستم در زمان آزمون می‌گردد. در مورد کارکرد بازشناسی گفتار نویزی، اطلاعات ممتاز از یک مدل به نام "معلم" به دست می‌آید که تنها در مورد گفتار تمیز آموزش دیده است. نتایج پیاده‌سازی نشان داده است که ساختار پیشنهادی میزان خطا را به طرز قابل توجهی کاهش داده است و در نتیجه یک ساختار موثر در کاهش اثر نویز می‌باشد [۴۰].

۲-۱۱- روش استخراج ویژگی PNCC

شکل ۲-۱۲ ساختار پردازش MFCC مرسوم، پردازش PLP و پردازش PNCC جدید که ارائه شده است را مقایسه می‌کند. نوآوری اصلی پردازش PNCC شامل طراحی مجدد تابع غیر خطی شدت نرخ است، به نحوی که فعالیت‌های آکوستیکی پس زمینه مبتنی بر تحلیل زمان متوسط را حذف کند [۴۲].

[1] Signal to Noise Ratio

[2] Peak Signal to Noise Ratio



شکل ۲-۱۲: بلوک دیاگرام روش های استخراج ویژگی MFCC، RASTA-PLP و PNCC [۴۲]

همانطور که در شکل ۲-۱۲ می‌توان دید، مراحل اولیه‌ی پردازش PNCC مشابه مراحل متناظر در تحلیل‌های MFCC و PLP است، با این تفاوت که تحلیل‌های فرکانسی با استفاده از فیلتر گاماتن انجام می‌شود. این کار با مجموعه‌ای از عملگرهای غیر خطی متغیر با زمان انجام می‌پذیرد که از تحلیل زمانی با طول بیشتر استفاده کرده و کار تفریق نویز را به خوبی انجام می‌دهد. مراحل نهایی پردازش هم مانند MFCC و PLP است با این تفاوت که از توانی با ضریب ۱/۱۵ استفاده می‌شود[۴۲].

در ساختار PNCC پیشنهادی، رویکرد جدیدی برای جبرانسازی نویز ارائه شده است که سرکوب نویز نامتقارن (ANS[۱]) نامیده می‌شود. این پروسه براساس این واقعیت بدست می‌آید که توان گفتار در هر کانال با سرعت بیشتری نسبت به توان نویز پس زمینه در کانال مشابه تغییر می‌کند. لذا می‌توان

[۱] Asymmetric Noise Suppression

گفت طیف فرکانسی مدولاسیون گفتار بالاتر از نویز است. با توجه به این مشاهدات الگوریتم‌های بسیاری مانند پردازش RASTA-PLP گسترش یافته است که از شیوه‌های فیلتر کردن بالاگذر یا میان گذر به طور مستقیم یا ضمنی درحوزه‌ی مدولاسیون طیف استفاده می‌کند. ساده ترین راه برای رسیدن به این هدف اعمال فیلتر بالاگذر به هر کانال است. اثر این کار حذف اجزای با تغییرات کند است که نمودار همان نویز می‌باشد. یک مشکل اساسی در استفاده از فیلتر بالاگذر خطی برای توان این است که خروجی فیلتر می‌تواند منفی شود. مقادیر منفی برای ضرایب توان درفرم ریاضیاتی مساله، مشکل‌ساز هستند چرا که توان ذاتاً مقداری مثبت است. علاوه بر این، این مساله باعث ایجاد مشکلاتی در کاربرد غیرخطی و سنتز مجدد گفتار خواهد شد. مگر اینکه مقدار کف مناسبی برای ضرایب توان در نظر گرفته شود. به جای فیلتر کردن در حوزه‌ی توان، می‌توان پس از اعمال لگاریتم، فیلتر کردن را انجام داد، همان کاری که در نرمال سازی کردن میانگین کپسترال معمولی در پردازش MFCC انجام می‌شود. اما این رویکرد برای محیط‌های با نویز جمع شونده مناسب نمی‌باشد. تفریق طیفی روش دیگری است که برای کاهش اثرات نویزی که توان آن به آرامی تغییر می‌کند استفاده می‌شود. در تکنیک‌های تفریق طیفی، سطح نویز از روی توان بخش‌های غیر گفتار یا با استفاده از رویکرد به روز رسانی پیوسته تخمین زده می‌شود. در رویکرد معرفی شده با استفاده از فیلتر غیرخطی نامتقارن و تفریق آن از توان آنی، تخمینی از کف نویز متغیر با زمان به دست آورده می‌شود. نتایج شبیه سازی ارائه شده، نشان می‌دهد که صحت بازشناسی این روش پیشنهاد شده، در حضور تمامی نویزهای بررسی شده، صحت بازشناسی بالاتری را به نسبت MFCC نتیجه داده است [۴۲].

۲-۱۲- نتیجه‌گیری

در این فصل از رساله ساختارهای استخراج ویژگی مقاوم در برابر نویز برای کاربرد بازشناسی گفتار مورد بررسی قرار گرفت. همانطور که در این الگوریتم‌ها مشاهده می‌شود، اکثر این الگوریتم‌ها در حوزه کپستروم بوده و در واقع ساختار تعمیم یافته و بهبود یافته الگوریتم استخراج ویژگی MFCC می‌باشند.

دلیل این مقدار مقبولیت روش استخراج ویژگی MFCC را می توان در عملکرد خوب این الگوریتم در محیط‌های بدون نویز و کم نویز برای کاربرد بازشناسی گفتار دانست. به همین خاطر در ساختارهای بررسی شده در این فصل، سعی در بهبود عملکرد الگوریتم MFCC در محیط‌های نویزی توسط اضافه نمودن و تغییر بخش هایی از این الگوریتم شده است. در این رساله الگوریتم‌هایی به منظور بازشناسی مقاوم گفتار در محیط‌های نویزی، پیشنهاد خواهد شد که نسخه بهبود یافته الگوریتم‌های حوزه کپستروم است و این الگوریتم‌ها مبتنی بر تبدیل فوریه کسری و تابع ریشه است.

فصل سوم: مبانی نظری

۳-۱- مقدمه

برای پیاده‌سازی و ارزیابی یک سیستم بازشناسی گفتار نیاز به داشتن دانش کافی در مورد اجزای سازنده سیستم بازشناسی و همچنین پارامترهای ارزیابی مختلف می‌باشد. در این فصل به تشریح اصلی‌ترین بلوک‌های سازنده و برخی از پارامترهای ارزیابی سیستم بازشناسی گفتار پرداخته شده است.

۳-۲- شبکه عصبی

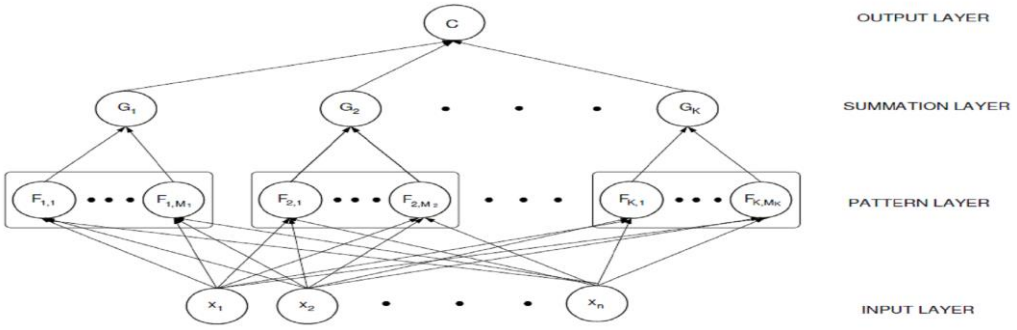
شبکه‌های عصبی زیستی مجموعه‌های بسیار عظیم از پردازشگرهای موازی به نام نرون هستند که به صورت هماهنگ برای حل مسئله عمل می‌کنند و توسط سیناپس‌ها (ارتباط‌های الکترومغناطیسی) اطلاعات را منتقل می‌کنند. یک شبکه عصبی مصنوعی شامل اجزای سازنده‌ی لایه‌ها و وزن‌ها می‌باشد. رفتار شبکه نیز وابسته به ارتباط بین اعضا است. در حالت کلی در شبکه‌های عصبی سه نوع لایه نرونی وجود دارد: لایه ورودی، لایه‌های پنهان و لایه خروجی. وزن‌های بین واحدهای ورودی و پنهان تعیین می‌کند که چه وقت یک واحد پنهان باید فعال شود و عملکرد واحد خروجی بسته به فعالیت واحد پنهان و وزن ارتباط بین واحد پنهان و خروجی می‌باشد. ساختارهای مختلفی برای شبکه عصبی وجود دارد که در ادامه به تشریح دو شبکه PNN[۱] و SVM[۲] که در الگوریتم‌های پیشنهادی برای استخراج ویژگی مقاوم برای بازشناسی گفتار استفاده شده پرداخته شده است. دلیل انتخاب طبقه‌بند PNN را می‌توان عملکرد یادگیری خوب و سریع با داده‌های آموزش کم دانست و دلیل انتخاب طبقه‌بند SVM، عملکرد خوب این طبقه‌بند برای کاربرد بازشناسی گفتار است.

[۱] Probabilistic Neural Network

[۲] Support Vector Machine

۳-۲-۱- شبکه عصبی احتمالاتی PNN

شبکه‌های عصبی احتمالاتی برای کاربردهای متنوع بازشناسی الگو به طور موفقیت آمیزی مورد استفاده قرار گرفته اند که از این کاربردها می‌توان به بازشناسی تصاویر، بازشناسی بافت، پردازش سیگنال، مالی و پزشکی اشاره نمود [۴۳و۴۴]. شبکه عصبی PNN یک طبقه‌بند الگو می‌باشد که در آن از استراتژی تصمیم‌گیری بسیار کاربردی بیز با تخمین‌گر غیرپارامتری پارزن [۴۵] برای تخمین توابع چگالی احتمال کلاس‌های متفاوت استفاده می‌گردد [۴۶]. مهم‌ترین مزایای PNN عبارتند از: فرآیند آموزش آن سریع بوده و ذاتاً یک ساختار موازی است که نمونه‌های آموزش می‌توانند بدون آنکه شبکه نیاز به آموزش مجدد وسیع داشته باشد، اضافه یا کم شود. براساس این واقعیات و مزایا می‌توان با قاطعیت بیان نمود که شبکه عصبی PNN می‌تواند بعنوان یک شبکه عصبی با نظارت در کاربردهای بازشناسی گوینده و گفتار به خوبی عمل کند. شبکه عصبی احتمالاتی در سال ۱۹۹۰ توسط Specht معرفی گردید. PNN بعنوان یک پیاده‌سازی تحلیل جداسازی kernel تعریف شده است که در آن عملیات بصورت یک شبکه مستقیم چند لایه با چهار لایه سازماندهی شده است. این لایه‌ها به شرح لایه ورودی، لایه الگو، لایه جمع‌کردن و لایه خروجی می‌باشند. در شکل ۳-۱ یک معماری پایه از شبکه عصبی PNN ارائه گردیده است [۴۷].



شکل ۳-۱: معماری پایه شبکه عصبی PNN [۴۷]

لایه اول محل قرارگیری ویژگی‌های استخراجی بعنوان ورودی شبکه عصبی است. لایه دوم که همان لایه الگو می‌باشد از توابع گاوسی که در آن مجموعه نقاط داده بعنوان مراکز می‌باشند، تشکیل شده است. با فرض اینکه X و W_i به یک طول واحد نرمال‌سازی شده باشند، خروجی واحد الگو بصورت رابطه زیر قابل نمایش می‌باشد[۴۸].

$$f(X.W_i) = \exp\left[-\frac{(X-W_i)^T(X-W_i)}{2\sigma^2}\right] \quad (۱-۳)$$

در رابطه فوق، i، σ، W و X به ترتیب بیانگر تعداد الگوها، پارامتر اسموسینگ، بردار وزن و بردار ورودی می‌باشند.

در این ساختار، هر دسته یک واحد جمع مرتبط به خودش را دارد و هر واحد جمع تنها به واحدهای الگو دسته‌ی خودش متصل است. واحدهای جمع، خروجی توابع کرنل واحدهای الگو متصل شده به خودش را جمع می‌کند. لایه خروجی در واقع یک لایه تصمیم‌گیر است که با توجه به مقادیر خروجی لایه قبل که بصورت باینری می‌باشد، بزرگترین مقدار را انتخاب کرده و سپس برچسب کلاس مربوطه را مشخص می‌کند [۴۸].

۳-۲-۲- طبقه‌بند ماشین بردار پشتیبان SVM

نسخه اولیه شبکه عصبی SVM در سال ۱۹۶۳ توسط vapnik ابداع شد و در سال ۱۹۹۵ توسط vapnik و cortes برای حالت غیرخطی تعمیم داده شد. شبکه عصبی بردار پشتیبان امروزه به ابزار بسیار قدرتمند برای پیاده سازی الگوریتم‌های یادگیری ماشین تبدیل شده است. طبقه‌بندهای SVM باتوجه به ساختار ساده و عدم نیاز به عملیات ریاضی پیچیده و سبک بودن بارمحاسباتی امروزه کاربردهای زیادی پیدا کرده است. این روش از جمله روش‌های جدیدی است که در سالیان اخیر کارایی خوبی نسبت به روش‌های قدیمی‌تر از جمله شبکه‌های عصبی پرسپترون نشان داده است. مبنای کاری طبقه‌بند SVM دسته‌بندی کردن خطی داده‌ها است و در جداسازی خطی داده‌ها سعی می‌شود خطی انتخاب گردد که

حاشیه اطمینان ماکزیمم گردد. معادله مرز بین دو کلاس را می‌توان بصورت رابطه ۳-۲ نوشت. برای تحقق هدف بیشینه شدن حاشیه اطمینان بین دو کلاس، رابطه لاگرانژ ۳-۳ باید کمینه گردد[۴۹].

$$y = w \cdot X + b \qquad (۲-۳)$$

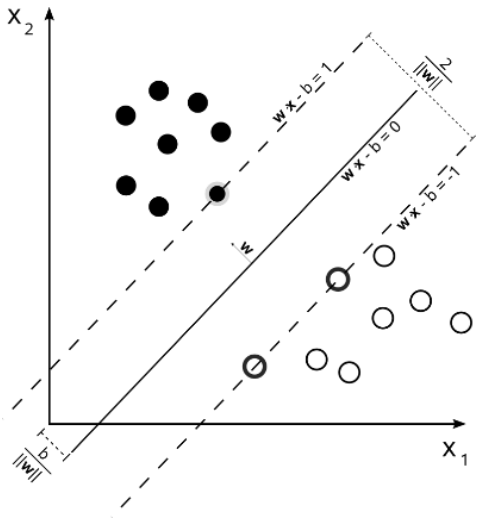
$$\underset{w.b.\alpha}{min}\{\frac{1}{2}\left|\left|w\right|\right|^2 - \sum_{i=1}^{n} \alpha_i[y_i(w \cdot X_i - b) - 1]\} \qquad (۳-۳)$$

برای کمینه کردن رابطه ۳-۳، باید قسمت دوم که همان سیکما می‌باشد، بیشینه شود. بدین منظور به دنبال α_iهایی هستیم که این قسمت را بیشینه نماید. به ازای هرکدام از داده‌ها یک α_i محاسبه می‌گردد که α_i های غیر صفر بیان کننده داده‌های مرزی و یا همان بردارهای پشتیبان هستند. پس از محاسبه این α_iها می توان w و b را با استفاده از روابط ۳-۴ و ۳-۵ محاسبه کرد.

$$w = \sum_i \alpha_i y_i X_i \qquad (۴-۳)$$

$$b = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} (w \cdot X_i - y_i) \qquad (۵-۳)$$

در رابطه ۳-۴، i در محدوده تعداد بردارهای پشتیبان تعریف می شود[۴۹].



شکل ۳-۲: تابع جداکننده، بردارهای پشتیبان و حاشیه اطمینان ماشین بردار پشتیبان[۴۹]

این شبکه‌ها به طورکلی به دو دسته hard SVM و soft SVM تقسیم می‌شوند. در شبکه‌های hard SVM حاشیه اطمینان باید به نحوی تعیین گردد که هیچکدام از داده‌ها اشتباه طبقه بندی نشوند. این قضیه در صورتی صحیح است که داده‌ها بصورت خطی قابل تفکیک باشند اما برای حالتی که کلاس‌ها بصورت خطی تفکیک پذیر نباشند از soft SVM استفاده می‌شود.

بطورکلی برای داده‌هایی که بصورت خطی جداپذیر نیستند، داده‌ها را با استفاده از یک تابع انتقال از فضای غیرخطی به فضای خطی نگاشت می‌دهیم. لذا رابطه مرز به رابطه تعمیم یافته ۳-۶ تبدیل می‌شود[۴۹].

$$g(x) = w^t\varphi(x) + b = \sum_{i\in SV} \alpha\varphi(x_i)^t\varphi(x) + b \ = \sum_{i\in SV} \alpha k(x_i.x) + b \qquad (۶\text{-}۳)$$

در رابطه ۳-۶، $k(x_i.x)$ همان کرنل می‌باشد.

۳-۳- الگوریتم‌های بهینه‌ساز

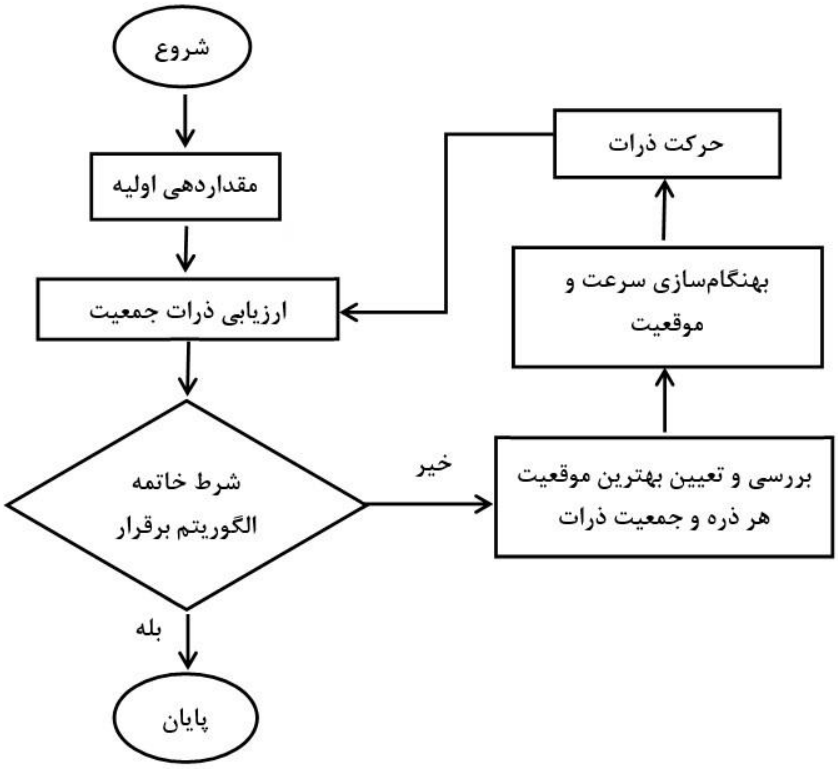
هدف از بهینه‌سازی یافتن بهترین جواب قابل قبول، با توجه به محدودیت‌ها و نیازهای مسأله است. برای یک مسأله، ممکن است جواب‌های مختلفی موجود باشد که برای مقایسه آنها و انتخاب جواب بهینه، تابعی به نام تابع هدف تعریف می‌شود. انتخاب این تابع به طبیعت مسأله وابسته است. به عنوان مثال، زمان سفر یا هزینه از جمله اهداف رایج بهینه‌سازی شبکه‌های حمل و نقل می‌باشد. به هرحال، انتخاب تابع هدف مناسب یکی از مهم‌ترین گام‌های بهینه‌سازی است. گاهی در بهینه‌سازی چند هدف به طور همزمان مد نظر قرار می‌گیرد؛ این‌گونه مسائل بهینه‌سازی را که دربرگیرنده چند تابع هدف هستند، مسائل چند هدفی می‌نامند. ساده‌ترین راه در برخورد با این گونه مسائل، تشکیل یک تابع هدف جدید به صورت ترکیب خطی توابع هدف اصلی است که در این ترکیب میزان اثرگذاری هر تابع با وزن اختصاص یافته به آن مشخص می‌شود. هر مسأله بهینه‌سازی دارای تعدادی متغیر مستقل است که آنها را متغیرهای طراحی می‌نامند که با بردار n بعدی X نشان داده می‌شوند. هدف از بهینه‌سازی تعیین

متغیرهای طراحی است، به گونه‌ای که تابع هدف کمینه یا بیشینه شود. در ادامه دو روش بهینه سازی پرکاربرد PSO[۱] و DE[۲] تشریح شده است.

۳-۳-۱- الگوریتم بهینه‌سازی ازدحام ذرات PSO

PSO از دسته الگوریتم‌های بهینه‌سازی است که بر مبنای تولید تصادفی جمعیت اولیه عمل می-کنند[۵۰]. در این الگوریتم با الگوگیری و شبیه سازی رفتار پرواز دسته جمعی (گروهی) پرندگان یا حرکت دسته جمعی (گروهی) ماهی‌ها بنا نهاده شده است. هر عضو در این گروه توسط بردار سرعت و بردار موقعیت در فضای جستجو تعریف می‌گردد. در هر تکرار زمانی، موقعیت جدید ذرات با توجه به بردار سرعت و بردار موقعیت در فضای جستجو تعریف می‌گردد. در هر تکرار زمانی، موقعیت جدید ذرات با توجه به بردار سرعت فعلی، بهترین موقعیت یافت شده توسط آن ذره و بهترین موقعیت یافت شده توسط بهترین ذره موجود در گروه، به روزرسانی می‌گردد. این الگوریتم ابتدا برای پارامترهای پیوسته تعریف شده بود اما با توجه به اینکه در برخی از کاربردها با پارامترهای گسسته سروکار داریم، این الگوریتم به حالت گسسته نیز بسط داده شده است. الگوریتم بهینه‌سازی ازدحام ذرات در حالت گسسته با (BPSO[۳]) معرفی می‌گردد[۵۰]. در این الگوریتم موقعیت هر ذره با مقدار باینری صفر و یا یک نشان داده می‌شود. در BPSO مقدار هر ذره می‌تواند از صفر به یک و یا از یک به صفر تغییر کند. سرعت هر ذره نیز به عنوان احتمال تغییر هر ذره به مقدار یک تعریف می‌گردد. روندنمای این الگوریتم بهینه‌ساز در شکل ۳-۳ نشان داده شده است.

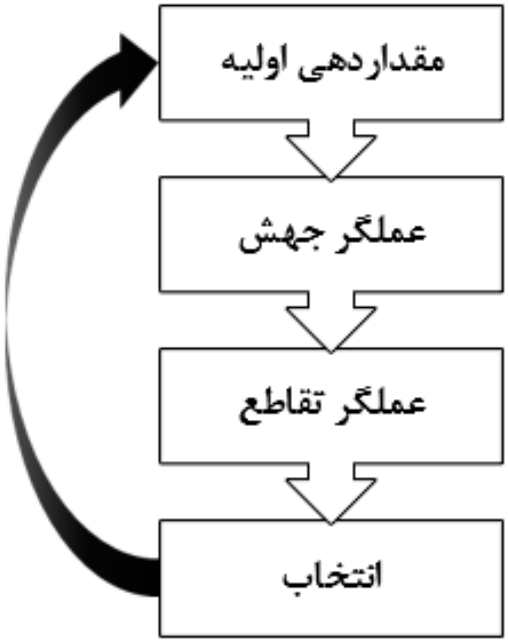
[۱] Particle Swarm Optimization
[۲] Differential Evolution
[۳] Binary Particle Swarm Optimization



شکل ۳-۳: روندنمای الگوریتم بهینه‌ساز PSO [۵۰]

۳-۳-۲- الگوریتم بهینه‌سازی تکامل تفاضلی (DE)

الگوریتم‌های تکاملی عموماً الگوریتم‌هایی هستند که در تمامی مسائل بهینه‌سازی می‌توانند نقش ایفا کنند. این الگوریتم‌ها قادر هستند که در مسائل واقعی و ریاضی به جواب‌های نزدیک بهینه دست یابند [۵۱]. دلیل این مقدار مقبولیت روش‌های بهینه‌سازی تکاملی به نسبت روش‌های کلاسیک و تحلیلی در این است که روش‌های کلاسیک و تحلیلی در بسیاری از مسائل برای بهینه‌سازی به زمان محاسباتی زیادی احتیاج دارند. الگوریتم‌های تکاملی انواع مختلفی دارند که یکی از آنها که اخیراً ارائه شده است، الگوریتم تکامل تفاضلی است. دلیل ارائه‌ی این الگوریتم را غلبه بر عیب اصلی الگوریتم ژنتیک که فقدان جستجوی محلی می‌باشد می‌توان دانست. ترتیب اعمال عملگرهای جهش و بازترکیب و همچنین نحوه عملکرد عملگر انتخاب را می‌توان به عنوان تفاوت‌های اساسی بین الگوریتم ژنتیک با الگوریتم تکامل تفاضلی دانست. در شکل ۳-۴ روال اجرای این الگوریتم ارائه گردیده است [۵۱].



شکل ۳-۴: روند اجرای الگوریتم DE [۲و۵۱]

الگوریتم DE از یک عملگر تفاضلی جهت تولید جمعیت جدید بهره می‌برد و این عملگر باعث ایجاد تبادل اطلاعات بین اعضای جمعیت می‌گردد. یکی از مزایای این الگوریتم، داشتن حافظه است که اطلاعات جواب‌های مناسب را در جمعیت فعلی حفظ می‌کند. در این الگوریتم، همه‌ی اعضای جمعیت شانس برابر برای انتخاب شدن به عنوان یکی از والدین را دارا می‌باشند. بدین صورت که نسل نوزاد با نسل والد از نظر میزان شایستگی که توسط تابع هدف سنجیده می‌شود، مقایسه می‌گردد. سپس بهترین اعضا به عنوان نسل بعدی وارد مرحله بعد می‌گردند. مهم‌ترین ویژگی الگوریتم DE، سرعت بالا، سادگی و قدرت بالای آن می‌باشد. این روش تنها با تنظیم سه پارامتر، شروع به کار می‌کند. این سه پارامتر عبارتند از NP، F، Cr که به ترتیب بیانگر اندازه جمعیت، وزن جهش و احتمال انجام تقاطع (بازترکیب) هستند و این احتمال در تفاضل دو بردار ضرب می‌گردد. معمولاً پارامتر F مقداری بین ۰ تا ۲ و پارامتر Cr مقداری بین ۰ تا ۱ در نظر گرفته می‌شود [۲و۵۱].

۳-۴- تبدیل فوریه کسری (FRFT[۱])

تبدیل فوریه کسری که با مخفف FRFT شناخته می‌شود، خانواده‌ای از تبدیل‌های خطی است که از تبدیل فوریه مشتق شده و در شاخه‌ی تحلیل هارمونیک در ریاضیات مورد استفاده قرار می‌گیرند. تبدیل فوریه کسری را می‌توان به صورت توان nام تبدیل فوریه (که در آن n عدد صحیح است) تعریف کرد. این تبدیل در مسائل مختلفی از جمله طراحی فیلتر، پردازش سیگنال، حل مسئله‌ی فاز و بازشناخت الگو کاربرد دارد.

تبدیل FRFT می‌تواند برای به‌دست آوردن نسخه کسری کانولوشن، ضریب همبستگی و دیگر عملگرها مورد استفاده قرار گیرد. این تبدیل برای اولین بار توسط ادوارد کندن [۵۲] با حل مسئله تابع گرین برای دوران فاز-زمان و همچنین توسط نامیاس [۵۳] و نوربرت وینر [۵۴] که بر روی مسئله چندجمله‌ای هرمایت کار می‌کردند، مطرح شد. با این حال این تبدیل تا سال ۱۹۹۳ که توسط چندین گروه مجدداً معرفی شد [۵۵]، استفاده‌ای در پردازش سیگنال نداشت. از آن زمان به بعد علاقه فراوانی به گسترش قضیه نمونه برداری شانون توسط این تبدیل برای سیگنال‌هایی که در دامنه کسری فوریه محدودند، نشان داده شده است.

برای هر عدد حقیقی، تبدیل فوریه کسری با زاویه α برای تابع f که با $F_\alpha(u)$ نشان داده می‌شود، به صورت زیر تعریف می‌گردد:

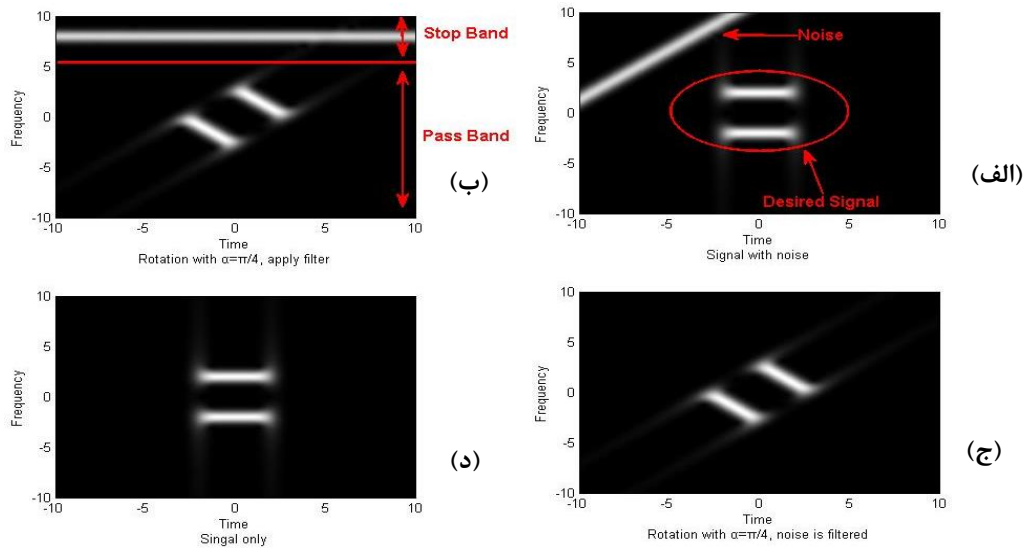
$$F_\alpha(u) = \int_{-\infty}^{+\infty} f(t) K_\alpha(t.u) dt \qquad (۳-۷)$$

$$K_\alpha(u.t) = \begin{cases} \delta(t-u) & if\ \alpha\ is\ a\ multiple\ of\ 2\pi \\ \delta(t+u) & if\ \alpha+\pi\ is\ a\ multiple\ of\ 2\pi \\ \sqrt{\frac{1-jcot\alpha}{2\pi}}\, e^{j\frac{t^2+u^2}{2}cot\alpha - jut\ cosec(\alpha)} & if\ \alpha\ is\ not\ a\ multiple\ of\ \pi \end{cases} \qquad (۳-۸)$$

[۱] Fractional Fourier Transform

در صورتی که زاویه α مضرب صحیحی از عدد π باشد، تابع‌های کتانژانت و کسکانت در فرمول بالا واگرا خواهند شد. با این حال این مشکل با قرار دادن حد تابع که تابع زیر انتگرال را به یک تابع دلتای دیراک تبدیل می‌کند، حل خواهد شد.

به دلیل ماهیت چرخشی این تبدیل، از این روش می‌توان برای فیلتر کردن سیگنال و حذف نویز استفاده نمود. همانطور که در شکل ۳-۵ نشان داده شده است، سیگنال نویزی در منحنی زمان-فرکانس وجود دارد، و برای فیلتر کردن نویز نمی‌توان بصورت مستقیم و بدون هیچ واسطه‌ای اقدام نمود، چرا که با این کار قسمتی از سیگنال اصلی نیز حذف خواهد شد. اما اگر بر روی سیگنال نویزی تبدیل فوریه کسری با زاویه برابر با ۴۵ درجه اعمال گردد، تصویر (ب) حاصل خواهد شد، در این حالت تمامی منحنی به اندازه ۴۵ درجه در جهت عقربه‌های ساعت می‌چرخد. با این کار می‌توان با اعمال یک فیلتر ساده پایین‌گذر که باندگذر و باندتوقف آن براساس دامنه فرکانسی مورد نیاز برای حذف نویز باشد، نویز را حذف نمود. در این حالت تصویر(ج) حاصل می‌گردد و در نهایت با اعمال مجدد تبدیل فوریه کسری، سیگنال اصلی بدون نویز با همان وضعیت اولیه بدست می‌آید [۵۶].



شکل ۳-۵: استفاده از تبدیل فوریه کسری برای حذف نویز الف) سیگنال به همراه نویز ب) اعمال تبدیل فوریه کسری با زاویه۴۵+ درجه ج)سیگنال فیلتر شده به منظور حذف نویز د)اعمال تبدیل فوریه کسری با زاویه ۴۵- درجه [۵۶]

در جدول ۳-۱ تبدیل فوریه کسری چند تابع مهم بدست آورده شده است. با توجه به این جدول می می‌توان مشاهده نمود که به ازای برخی از αها، مقدار تبدیل فوریه کسری این توابع با تبدیل فوریه معمولی یکسان خواهد شد.

جدول ۳-۱- تبدیل فوریه کسری برخی از توابع پرکاربرد

سیگنال	تبدیل فوریه کسری با ضریب α
$\delta(t-\tau)$	$\sqrt{\frac{1-jcot\alpha}{2}}.e^{j\frac{\tau^2+u^2}{2}\cot(\alpha)}e^{-jcsc(\alpha\tau u)}$
$e^{-j(at^2+bt+c)}$	$\sqrt{\frac{1-jcot\alpha}{j2a-jcot\alpha}}.e^{j\frac{2a\cot(\alpha)-1}{2(2\cot(\alpha)-2a)}u^2}e^{-j\frac{bcsc(\alpha)}{\cot(\alpha)-2a}}e^{-j\frac{b^2}{2(2\cot(\alpha)-2a)}-jc}$
1	$\sqrt{1+jtan(\alpha)}.e^{-\frac{j}{2}u^2\tan(\alpha)}$
Cos(vt)	$\sqrt{1+jtan(\alpha)}.e^{-\frac{j}{2}(u^2+v^2)\tan(\alpha)}\cos(uv\sec(\alpha))$
Sin(vt)	$\sqrt{1+jtan(\alpha)}.e^{-\frac{j}{2}(u^2+v^2)\tan(\alpha)}\sin(uv\sec(\alpha))$

در جدول ۳-۲- خاصیت های مهم تبدیل فوریه کسری ارائه شده است. از آنجایی که تبدیل فوریه کسری حالت بسط یافته تتبدیل فوریه معمولی می باشد، به همین خاطر این روابط دقیقا شبیه روابط تبدیل فوریه معمولی می باشد.

جدول ۳-۲- خاصیت های تبدیل فوریه کسری

خاصیت	حوزه زمان	تبدیل فوریه کسری
خطینگی	$ax(t)+by(t)$	$aX_\alpha(u)+bY_\alpha(u)$
جابه جایی زمانی	$x(t-T)$	$e^{j\frac{T^2}{2}\cot(\alpha)}e^{-\mathrm{juTcsc}(\alpha)}X_\alpha(u-Tcos(\alpha))$
مدولاسیون	$e^{j2\pi Ft}x(t)$	$e^{-j\frac{v^2\sin(\alpha)\cos(\alpha)}{2}+iuvcos(\alpha)}X_\alpha(u-vsin(\alpha))$
مشتق	$\frac{d}{dt}x(t)$	$\cos(\alpha)\frac{d}{du}X_\alpha(u)+jusin(\alpha)X_\alpha(u)$
انتگرال	$\int_{-\infty}^{t}x(t)dt$	$\sec(\alpha).e^{-j\frac{u^2}{2}\tan(\alpha)}\int_{-\infty}^{u}e^{j\frac{v^2}{2}\tan(\alpha)}X_\alpha(v)dv$
ضرب زمانی	$tx(t)$	$ucos(\alpha)X_\alpha(u)+jsin(\alpha)\frac{d}{du}X_\alpha(u)$
رابطه پارسوال	$\int_{-\infty}^{+\infty}x(t)y^*(t)dt=\int_{-\infty}^{+\infty}X_\alpha(u)Y_\alpha^*(u)du$	

۳-۵- پارامتر Mean Square Movement (MSM)

ویژگی‌هایی که توسط روش‌های مختلف استخراج ویژگی بدست آمده‌اند، توسط پارامتر MSM[۱] می‌توانند مورد ارزیابی قرار گیرند تا میزان مقاومت روش‌های مختلف استخراج ویژگی در برابر نویزهای مختلف و یا هر عامل ایجاد کننده اعوجاج دیگری بدست آید[۱۲] و [۵۷].

پارامتر میانگین مجذور جابه‌جایی (MSM) که در این رساله پیشنهاد شده است، میزان تغییرات دو سیگنال هم طول را می‌سنجد. این پارامتر از طریق رابطه ۳-۱۱ قابل محاسبه است. هر چقدر میزان MSM کمتر باشد، بدین معناست که آن روش نسبت به نویزهای مورد بررسی مقاوم‌تر بوده و حساسیت کمتری از خود نشان می‌دهند [۱۲].

$$AverageFormant_{Original\ Signal}(k) = \frac{1}{N}\sum_{i=1}^{N} Formant_{Original\ Signal}(i.k) \qquad (۹-۳)$$

$$AverageFormant_{Noisy\ Signal}(k) = \frac{1}{N}\sum_{i=1}^{N} Formant_{Noisy\ Signal}(i.k) \qquad (۱۰-۳)$$

$$MSM = \frac{\sum_{m=1}^{M}\{[AverageFormant_{Original\ Signal}(m)]-[AverageFormant_{Noisy\ Signal}(m)]\}^2}{M} \qquad (۱۱-۳)$$

در روابط فوق N، k و M به ترتیب بیانگر تعداد فریم‌های واکدار، شماره فورمنت و تعداد فورمنت‌ها به ازای هر فریم می‌باشند [۱۲].

[۱] Mean Square Movement

فصل چهارم : الگوریتم های پیشنهادی

۴-۱-مقدمه

وجود نویز در محیط سبب می‌شود که عملکرد سیستم بازشناسی گفتار به نسبت محیط تمیز به طرز چشم‌گیری کاهش یابد. برای اینکه سیستم بازشناسی گفتار بتواند در محیط نویزی عملکرد قابل قبولی را نتیجه دهد باید دانش کافی در مورد نوع، مشخصات و نحوه تاثیر نویز موجود در محیط بر روی سیگنال صوتی چه در حوزه زمان و چه در حوزه فرکانس کسب نمود. به همین خاطر در این فصل از رساله، ابتدا تاثیر نویزهای مختلف با سطوح سیگنال به نویز مختلف را بر روی یکی از مهم‌ترین پارامترهای سیگنال گفتار در حوزه فرکانس که فورمنت‌های سیگنال گفتار است به منظور سنجش میزان مقاومت گفتار در برابر نویزهای مختلف، مورد بررسی قرار داده‌ایم. سپس در بخش دوم این فصل یک روش استخراج ویژگی مقاوم برای بازشناسی گفتار در محیط نویزی که مبتنی بر تبدیل فوریه کسری می باشد، ارائه شده است. روش استخراج ویژگی پیشنهادی FrRC، در حضور نویزهای مختلف با سطوح سیگنال به نویز مختلف با سایر روش‌های استخراج ویژگی متداول، مقایسه شده و در نهایت تحلیلی ریاضی از ارتباط بین ویژگی‌های بدون نویز و نویزی FrRC ارائه شده است. در بخش آخر این فصل از رساله، ساختار پیشنهادی دوم با نام AFPNCC به منظور استخراج ویژگی مقاوم ارائه شده است و نتایج پیاده سازی این الگوریتم با سایر الگوریتم های استحراج ویژگی دیگر ارائه و گزارش شده است.

۴-۲- بررسی میزان مقاومت فورمنت‌های یک سیگنال گفتار در برابر تغییر مکان نسبت به نویزهای مختلف متداول

فورمنت‌ها از اهمیت ویژه‌ای در پردازش گفتار برخوردار هستند. برای بازشناسی گوینده، گفتار و بسیاری از کاربردهای دیگر، مشخص نمودن محل دقیق فورمنت‌ها بسیار حیاتی و مهم می‌باشد. در محیط واقعی مجموعه‌ای از نویزها بر روی سیگنال گفتار تاثیر می‌گذارند، لذا مشخص نمودن میزان تاثیر نویزهای

مختلف بر روی تغییر مکان فورمنت‌ها از اهمیت ویژه ای برخوردار است. بسیاری از کاربردها نظیر استخراج فورمنت به شدت در دهه گذشته مورد مطالعه قرار گرفته است. دیدگاه‌های مختلفی برای بدست آوردن فورمنت‌ها پیشنهاد شده است. اکثر این دیدگاه‌ها از اعمال برخی از روش‌های یافتن قله بر روی طیف روش پیشگویی خطی (LP[1]) بهره می‌برند. برخی دیگر از الگوریتم‌ها براساس مدل‌های مخفی مارکوف (HMM) می‌باشند که برای یافتن فورمنت‌ها، از معیار تصادفی استفاده می‌کنند[۵۸و۵۹]. در این بخش از رساله، برای استخراج فورمنت‌های سیگنال گفتار از روش LPC استفاده شده است.

۴-۲-۱- روش استخراج فورمنت‌های یک سیگنال گفتار با استفاده از روش LPC

روش LPC یکی از روش‌هایی است که برای بدست آوردن طیف فرکانسی استفاده می‌شود. این روش امروزه موفق‌ترین روش با کاربردهای وسیع می‌باشد[۱۲]. مزایای زیادی برای استفاده از روش LPC وجود دارد که از این قبیل می‌توان به موارد زیر اشاره نمود[۱۲]:

- روش LPC بهبود بیشتری در تقریب ضرایب طیف ایجاد می‌کند.
- روش LPC زمان کمتری برای محاسبه پارامترهای سیگنال صرف می‌کند.
- روش LPC قادر به دریافت و استخراج ویژگی‌ها و مشخصات مهم سیگنال‌های ورودی می‌باشد[۱۲و۳۲].

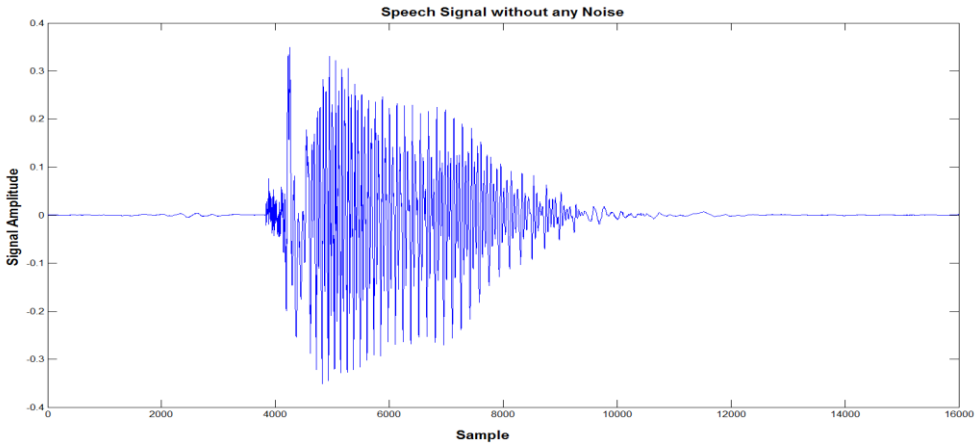
در علم گفتار و فونوتیک، فرکانس‌های فورمنت در واقع یک تشدید آکوستیک از مدل چاکنای انسان می‌باشد و این فرکانس‌ها بصورت قله دامنه طیف فرکانسی صدا اندازه‌گیری می‌شود. در این پیاده‌سازی استخراج فورمنت‌ها از طریق روش LPC تخمین زده شده است. مدل چاکنای بصورت یک فیلتر خطی

[1] Linear Prediction

با رزونانس‌ها مدل شده است، که فرکانس این رزونانس‌های چاکنای ، فرکانس‌های فورمنت نامیده می-شوند[۶۰]. در واقع برای یافتن فورمنت‌ها با استفاده از روش LPC از پاسخ فرکانسی مدل تولید گفتار بهره برده می‌شود. در این روش از روی پاسخ فرکانسی چاکنای، قله‌های دامنه را مشخص کرده و فرکانس نظیر این قله‌ها را بعنوان فورمنت‌های سیگنال گفتار در نظر می‌گیرند.

۴-۲-۲- روش استفاده شده برای بررسی میزان تاثیر نویز بر روی مکان فرکانس فورمنت‌های یک سیگنال گفتار

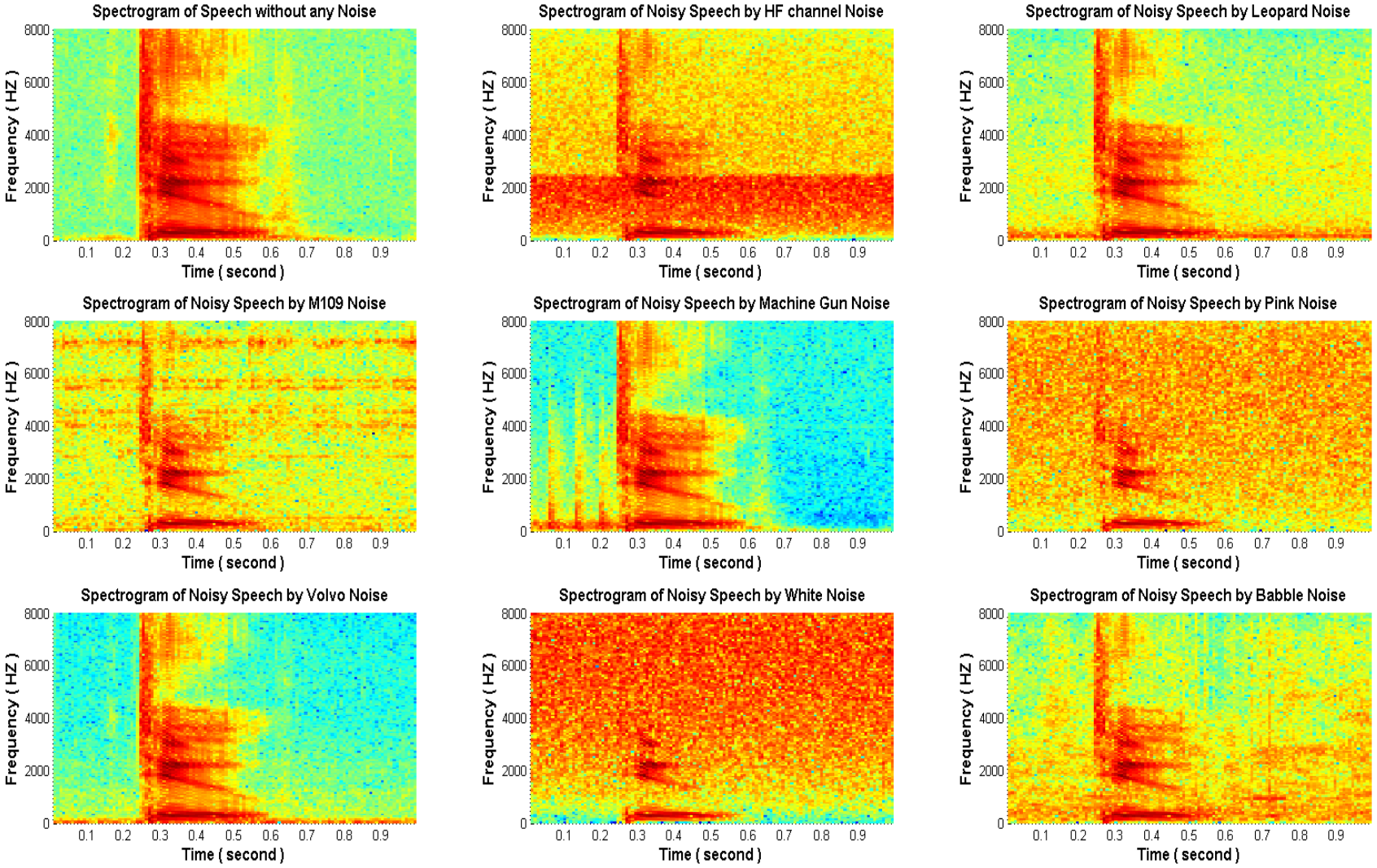
در این پیاده‌سازی، ابتدا سیگنال گفتار را با طول ۲۵۶ نمونه فریم‌بندی کرده و از بین فریم‌های سیگنال گفتار، ۱۶ فریم واکدار را جدا کرده‌ایم[۱۲]. شکل موج سیگنال گفتار ورودی بدون نویز در شکل ۴-۱ آورده شده است. فرکانس نمونه‌برداری این سیگنال ۱۶کیلوهرتز می‌باشد.



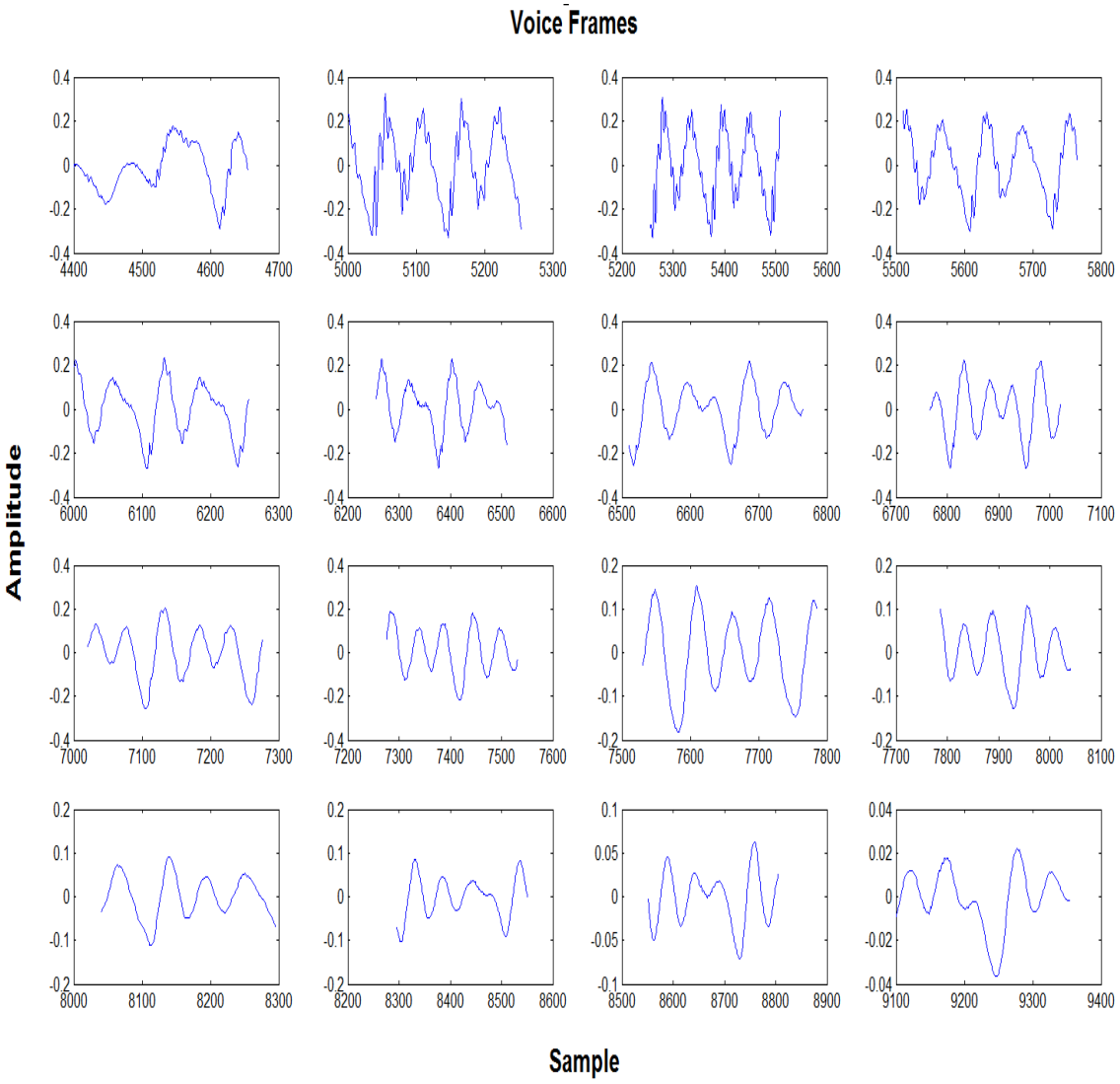
شکل ۴-۱: شکل موج سیگنال گفتار ورودی فریم‌بندی نشده بدون نویز [۱۲]

در شکل ۴-۲، طیف‌نگاره سیگنال گفتار بدون نویز نشان داده شده در شکل (۴-۱) به همراه طیف نگاره سیگنال‌های گفتار نویزی شده توسط ۸ نوع مختلف نشان داده شده است.

فریم‌های واکدار استخراج شده از این سیگنال گفتار در شکل ۴-۳ نشان داده شده است. همانطور که در این شکل‌ها مشخص است، می‌توان یک رفتار شبه پریودیک در شکل آنها مشاهده نمود و این رفتار شبه پریودیک بیانگر این است که فریم‌های در نظر گرفته شده فریم‌های واکدار می‌باشند[۱۲].



شکل ۴-۲: طیف نگاره سیگنال گفتار بدون نویز و هشت نوع سیگنال گفتار نویزی شده [۱۲]



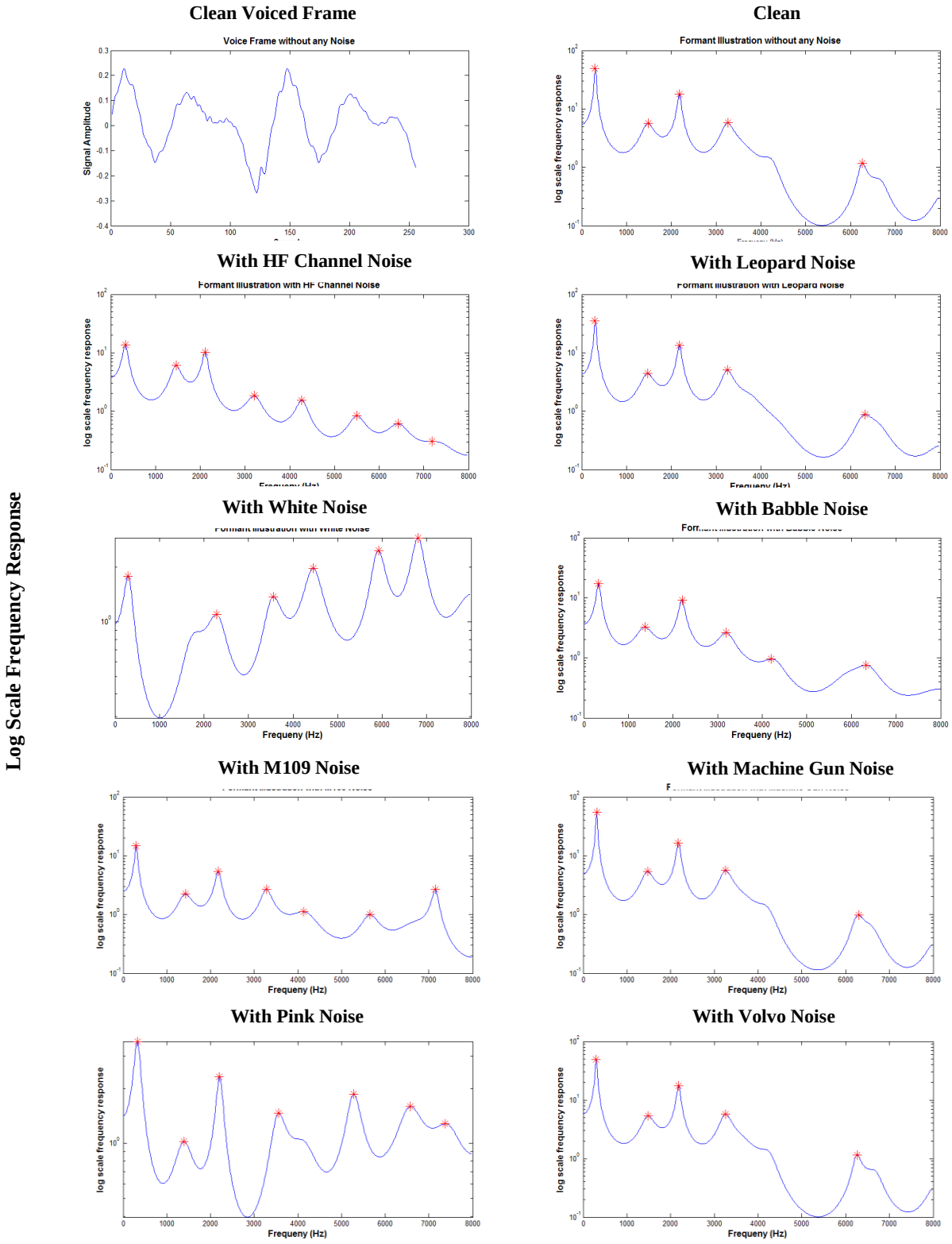
شکل ۴-۳: فریم‌های واکدار استخراج شده بدون اعمال هیچ منبع نویزی [۱۲]

بعد از اینکه فریم‌های واکدار استخراج گردید، تمامی فریم‌ها در یک پنجره همینگ با طول ۲۵۶ نمونه ضرب می‌شود. بعد از این کار برای هرکدام از فریم‌های استخراج شده که در پنجره همینگ ضرب شده است، پاسخ فرکانسی مدل تولید گفتار با استفاده از روش تمام قطب LPC بدست آورده شده است[۱۲]. سپس فرکانس متناظر با ۳ قله اول دامنه پاسخ فرکانسی هر فریم که بیانگر سه فورمنت هر فریم می‌باشد استخراج شده است. لازم به ذکر است تا این لحظه هیچ یک از نویزها با سیگنال گفتار اصلی تجمیع نشده است. تا این مرحله برای هر فریم ۳ فورمنت استخراج شده است. در مرحله بعد چون ۱۶ فورمنت اول، ۱۶ فورمنت دوم و ۱۶ فورمنت سوم بدست آمده است، برای هر فورمنت یک میانگین گیری انجام می‌گیرد تا فورمنت‌ها به سه فورمنت، میانگین فورمنت اول، دوم و سوم تبدیل

گردد. تا این مرحله برای کل فریم‌های واکدار استخراج شده بدون نویز جمعاً سه فورمنت استخراج گردید[۱۲].

در مرحله بعد تمامی ۱۶ فریم واکدار بدون نویز را هر بار با هرکدام از نویزهای HF Channel، Leopard، M109، Machine Gun، Pink، Volvo، Whiteو Babble نمونه به نمونه جمع می‌کنیم. در این حالت ۱۶ فریم واکدار نویزی شده برای هرکدام از نویزهای مذکور داریم. برای هرکدام از دسته نویزها همان کاری را که برای فریم‌های بدون نویز که در بالا ذکر شد، انجام می‌دهیم. با این کار جمعاً ۹ دسته میانگین فورمنت اول تا سوم بصورت بدون نویز، نویزی HF Channel، Leopard، M109، Machine Gun، Pink، Volvo، Babbleو White خواهیم داشت. مقادیر این میانگین فورمنت‌ها با SNRهای 10dB، 5dBو 15dB به ترتیب در جداول ۴-۱ تا ۴-۳ آورده شده‌اند[۱۲].

در شکل ۴-۴، یک فریم واکدار بدون نویز به همراه پاسخ فرکانسی و محل فورمنت‌های آنها، در حالت بدون نویز و همچنین با افزودن ۸ نوع نویز مطرح شده نشان داده شده است.

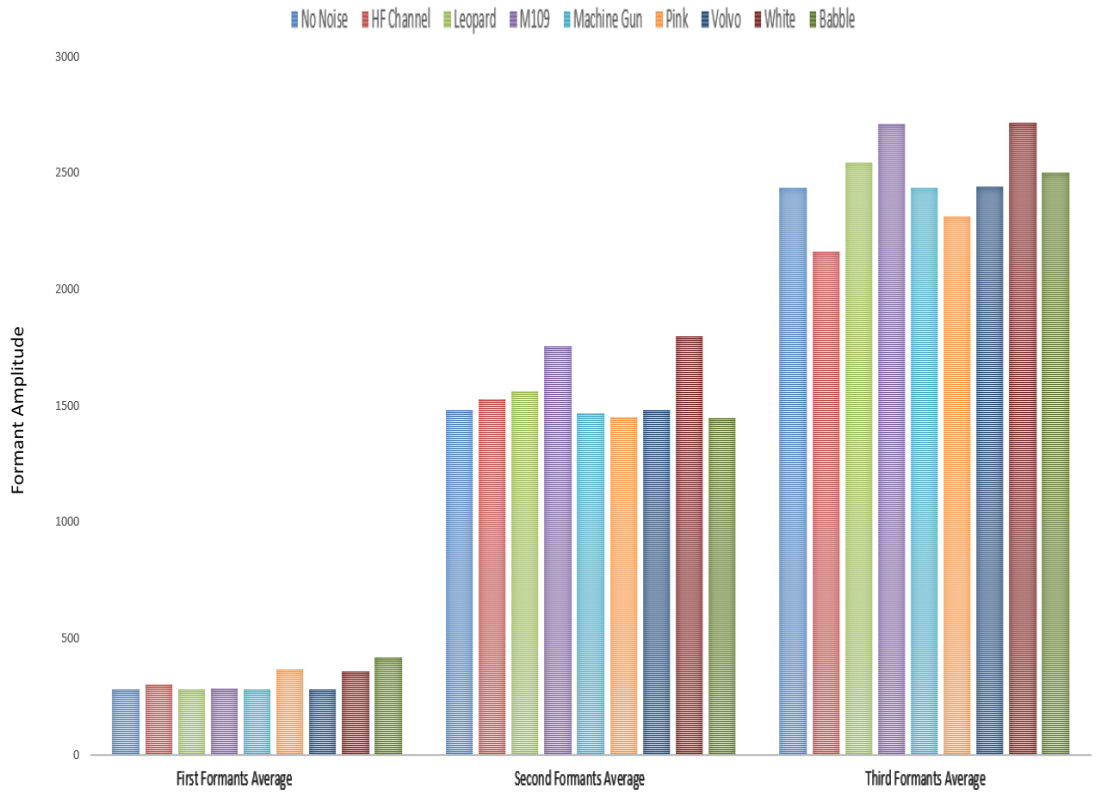


شکل ۴-۴: شکل موج یک فریم واکدار بدون نویز به همراه پاسخ فرکانسی و محل فورمنت‌های آنها، در حالت بدون نویز و همچنین با افزودن ۸ نوع نویز مطرح شده [۱۲]

- لازم به ذکر است که در شکل فوق، فورمنت‌ها با ضربدر قرمز رنگ مشخص شده‌اند.

جدول ۴-۱: میزان میانگین فورمنت‌های سیگنال‌های مختلف بدون نویز و با اعمال نویزهای مختلف (برحسب هرتز) با SNR = 10dB [۱۲]

	Type of Signal								
Formant Average Frequency (Hz)	Clean Signal	Clean Signal + HF Channel Noise	Clean Signal + Leopard Noise	Clean Signal + M109 Noise	Clean Signal + Machine Gun Noise	Clean Signal + Pink Noise	Clean Signal + Volvo Noise	Clean Signal + White Noise	Clean Signal + Babble Noise
First Formant	281.2	302.7	281.2	285.1	282.2	367.1	280.2	357.4	420.9
Second Formant	1482.4	1527.3	1564.4	1756.8	1470.7	1452.1	1481.4	1797.8	1447.2
Third Formant	2439.4	2165	2543.9	2711.9	2436.5	2315.4	2441.4	2717.7	2500.9



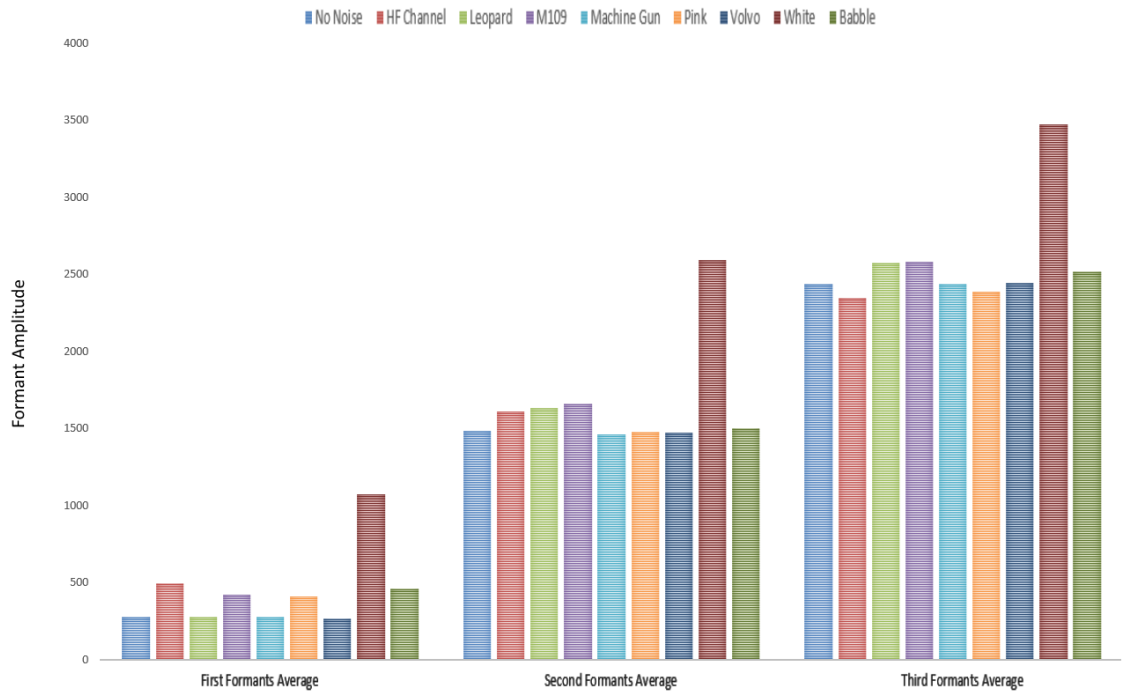
شکل ۴-۵: نمودار مستطیلی میانگین فورمنت‌های اول تا سوم به ازای نویزهای مختلف در SNR = 10dB [۱۲]

باتوجه به شکل ۴-۵ می‌توان مشاهده نمود که که وضعیت میانگین فورمنت‌های اول تا سوم در SNR = 10dB به ازای نویزهای مختلف چگونه می‌باشد. بعنوان نمونه می‌توان گفت که میانگین فورمنت‌های سوم به ازای نویزهای Volvo و Machine Gun نسبت به حالت بدون نویز کمترین تغییرات را دارد از

طرف دیگر می‌توان مشاهده نمود که نویز سفید باعث بیشترین جابه‌جایی و تغییر در هر سه میانگین فورمنت‌ها می‌شود[۱۲].

جدول ۴-۲: میزان میانگین فورمنت‌های سیگنال‌های مختلف بدون نویز و با اعمال نویزهای مختلف (برحسب هرتز) با SNR = 5 dB [۱۲]

	Type of Signal								
Formant Average Frequency (Hz)	Clean Signal	Clean Signal + HF Channel Noise	Clean Signal + Leopard Noise	Clean Signal + M109 Noise	Clean Signal + Machine Gun Noise	Clean Signal + Pink Noise	Clean Signal + Volvo Noise	Clean Signal + White Noise	Clean Signal + Babble Noise
First Formant	281.2	497.1	276.3	419.9	277.3	408.2	269.5	1071.3	458.9
Second Formant	1482.4	1610.3	1634.7	1661.1	1463.8	1476.5	1475.5	2595.70	1502.9
Third Formant	2439.4	2346.6	2576.1	2583.9	2436.5	2388.6	2443.3	3472.6	2520.5

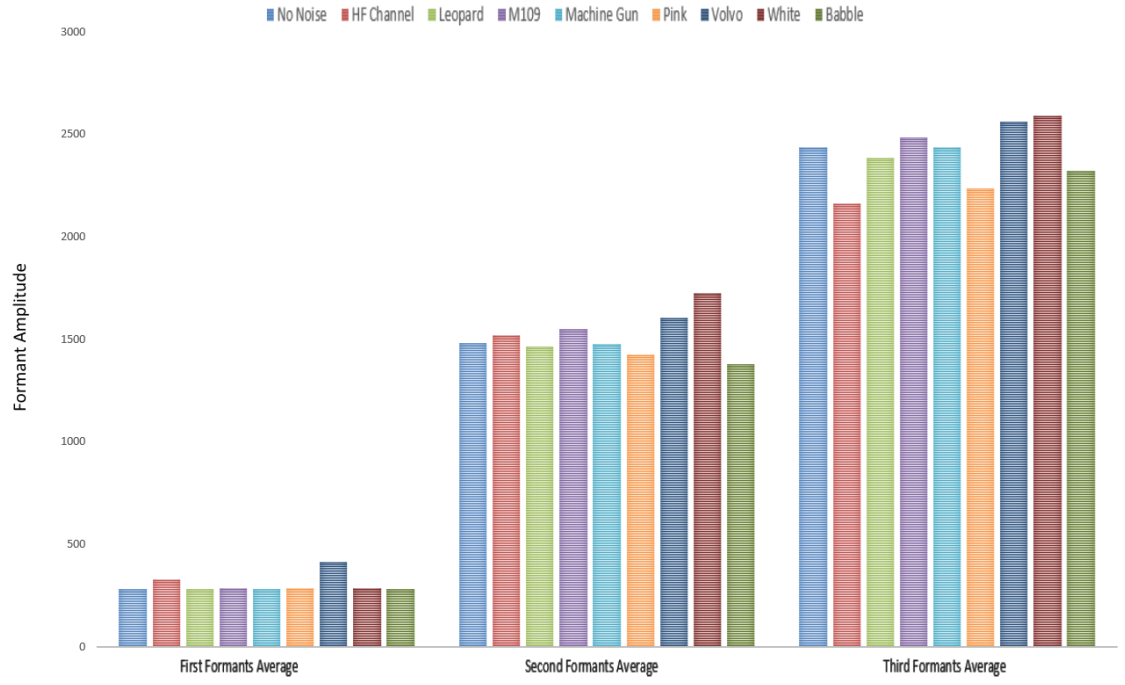


شکل ۴-۶: نمودارمستطیلی میانگین فورمنت‌های اول تا سوم به ازای نویزهای مختلف در SNR = 5dB [۱۲]

در شکل ۴-۶ نیز می‌توان مشاهده کرد به ازای SNR = 5dB نویز White بیشترین تغییر را در میانگین فورمنت‌ها ایجاد می‌کند. بعبارت دیگر فورمنت‌ها کمترین مقاومت را به ازای نویز White از خودشان نشان می‌دهند.

جدول ۴-۳: میزان میانگین فورمنت‌های سیگنال‌های مختلف بدون نویز و با اعمال نویزهای مختلف (برحسب هرتز) با SNR = 15 dB [۱۲]

	Type of Signal								
Formant Average Frequency (Hz)	Clean Signal	Clean Signal + HF Channel Noise	Clean Signal + Leopard Noise	Clean Signal + M109 Noise	Clean Signal + Machine Gun Noise	Clean Signal + Pink Noise	Clean Signal + Volvo Noise	Clean Signal + White Noise	Clean Signal + Babble Noise
First Formant	281.2	327.1	280.2	285.1	282.2	288.1	415	286.1	282.2
Second Formant	1482.4	1522.4	1463.8	1549.8	1477.5	1427.7	1605.4	1728.5	1377.9
Third Formant	2439.4	2162.1	2385.7	2484.3	2437.5	2235.3	2561.5	2590.8	2322.2



شکل ۴-۷: نمودار مستطیلی میانگین فورمنت‌های اول تا سوم به ازای نویزهای مختلف در SNR = 15dB [۱۲]

همانطورکه در شکل ۴-۷ مشخص است، میانگین فورمنت‌های اول تا سوم از سیگنال صوتی با SNR برابر با ۱۵ دسیبل در حضور نویز Machine Gun کمترین تغییرات را نسبت به بدون نویز دارا می‌باشد.

به عبارت دیگر، فورمنت‌ها نسبت به این نویز بسیار مقاوم بوده و کمترین حساسیت را نشان می‌دهند. اما می‌توان دید که نویز White بیشترین تاثیر را بر روی فورمنت‌ها به نسبت حالت بدون نویز دارد و باعث بیشترین جابه‌جایی فورمنت‌ها در مقایسه با حالت بدون نویز می‌شود[۱۲].

در این پیاده‌سازی برای اینکه دید صحیح‌تر و دقیق‌تری از میزان مقاومت فورمنت‌ها در برابر نویزهای مختلف داشته باشیم از پارامتر MSM یا همان میانگین مربعات جابه‌جایی استفاده شده است. با استفاده از رابطه زیر MSM هر کدام از منابع نویز نسبت به فریم‌های بدون نویز سنجیده می‌شود. هر چقدر میزان MSM برای هرکدام از منابع نویز کمتر باشد، بدین معناست که فورمنت‌های سیگنال نسبت به آن نویز مقاوم‌تر بوده و حساسیت کمتری از خود نشان می‌دهند. رابطه ۳-۱۱ این ارتباط را نشان می‌دهد. در جداول ۴-۴ تا ۴-۶ مقادیر MSMهای منابع نویز مختلف به ترتیب به ازای SNRهای 10dB، 5dB و 15dB ارائه شده است[۱۲].

جدول ۴-۴: مقایسه بین MSMها برای منابع نویز مختلف با SNR = 10dB [۱۲]

Noise Source	HF Channel	Leopard	M109	Machine Gun	Pink	Volvo	White	Babble
MSM	25927.5	5882.5	49851.1	48.9	7894.5	1.9	60920.1	8174.2

باتوجه به جدول ۴-۴ که براساس SNR = 10dB بدست آمده است، می‌توان اینطور بیان نمود که نویز Volvo و بعد از آن نویز Machine Gun دارای کمترین اثر برروی MSM یا همان جابه‌جایی فورمنت‌ها به نسبت حالت بدون نویز هستند. اما با قاطعیت می‌توان اظهار نمود که نویز سفید بیشترین تاثیر را بر روی MSM دارد و این نتیجه نیز منطقی می‌باشد چراکه نویز سفید در تمامی فرکانس‌های PSD تاثیرگذار است[۱۲].

جدول ۴-۵: مقایسه بین MSMها برای منابع نویز مختلف با SNR = 5dB [۱۲]

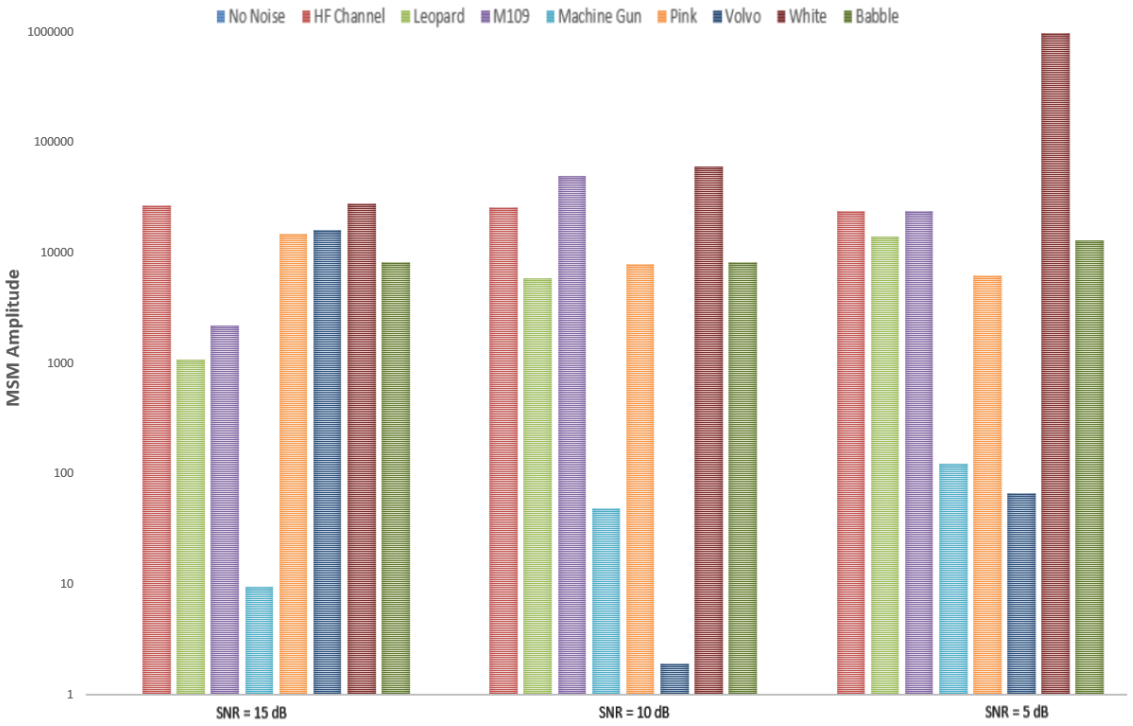
Noise Source	HF Channel	Leopard	M109	Machine Gun	Pink	Volvo	White	Babble
MSM	23850.4	13974.8	24018.9	122.7	6243.4	66.4	977021.8	12860

در جدول ۴-۵ که در SNR = 5dB بدست آمده است، می توان نکته فوق را دوباره تکرار کرده و تایید نمود که نویزهای Volvo و Machine Gun کمترین MSM اما نویز White بیشترین MSM را دارا می‌باشند[۱۲].

جدول ۴-۶: مقایسه بین MSMها برای منابع نویز مختلف با SNR = 15dB [۱۲]

Noise Source	HF Channel	Leopard	M109	Machine Gun	Pink	Volvo	White	Babble
MSM	26876.4	1076.7	2191.2	9.5	14898.3	15980.4	27832.6	8217.5

نتایج مربوط به MSMها در جدول ۴-۶ که در SNR = 15dB بدست آمده است را می‌توان اینطور تحلیل کرد که در این SNR، نویز Machine Gun همچنان کمترین مقدار MSM را دارد، در حالی که در این سیگنال به نویز، میزان MSM در نویز Volvo به شدت افزایش یافته است. اما همچنان نویز سفید دارای بیشترین میزان جابه‌جایی فورمنت‌ها است. چراکه MSM آن نسبت به بقیه نویزها بزرگ‌تر می‌باشد[۱۲].برای اینکه بتوان سه جدول فوق را باهم بطور کامل مقایسه کرد و مشاهده‌ی دقیق‌تری از میزان تاثیر SNRهای گوناگون در MSMها داشت، نمودار مستطیلی مربوط به تمامی MSMها در هر سه SNR = 10dB, 5dB , 15dB در شکل ۴-۸ نمایش داده شده است[۱۲]. لازم به ذکر است از آنجایی که دامنه MSMها باهم مخصوصاً به نسبت نویز سفید خیلی تفاوت دارد، برای مشاهده‌ی بهتر دامنه‌های MSMها، محور عمودی شکل (۴-۸) بصورت لگاریتمی مدرج شده است. همانطورکه در شکل ۴-۸ مشاهده می‌شود، MSM به ازای نویز Machine Gun و Volvo در دو SNR = 10dB, 5dB دارای کمترین مقدار می‌باشد. اما در SNR = 15dB کمترین مقدار برای MSM، مربوط به نویز Machine Gun است. در این SNR میزان MSM به ازای نویز Volvo به نسبت حالات قبل بشدت افزایش پیدا کرده است. اما می‌توان گفت نویز سفید به ازای هر سه SNR مطرح شده، دارای بیشترین MSM می باشد. بعبارت دیگر می توان این گونه بیان کرد که نویز سفید مستقل از میزان SNR، بیشترین اثر را بر روی جابه جایی فورمنت ها دارد. اما نکته دیگر این است که با افزایش میزان SNR، مقدار MSM نویز سفید به شدت کاهش می یابد[۱۲].



شکل ۴-۸: نمودار لگاریتمی مستطیلی میزان MSMها در برابر نویزهای مختلف با سه SNR = 15dB, 10dB, 5dB [۱۲]

۴-۳- الگوریتم استخراج ویژگی پیشنهادی اول: Fractional Root Coefficient (FrRC)

در این پیاده‌سازی یک روش استخراج ویژگی جدید با نام FrRC برای کاربرد بازشناسی گفتار در محیط‌های نویزی معرفی شده است. این روش استخراج ویژگی پیشنهادی مبتنی بر تبدیل فوریه کسری زمان کوتاه و تابع ریشه می‌باشد. از آنجایی که انتخاب ضریب تبدیل کسری و تابع ریشه برای تحلیل‌های مناسب سیگنال‌های چند جزئی از قبیل گفتار همچنان مورد بحث است، به همین خاطر در روش FrRC با استفاده از الگوریتم‌های فرا ابتکاری پارامترهای بهینه آلفا و گاما به ترتیب برای تبدیل فوریه کسری و تابع ریشه بدست آورده شده است. همچنین توسط روش تحلیلی و ریاضی کارآیی روش پیشنهادی FrRC بررسی و اثبات شده است.

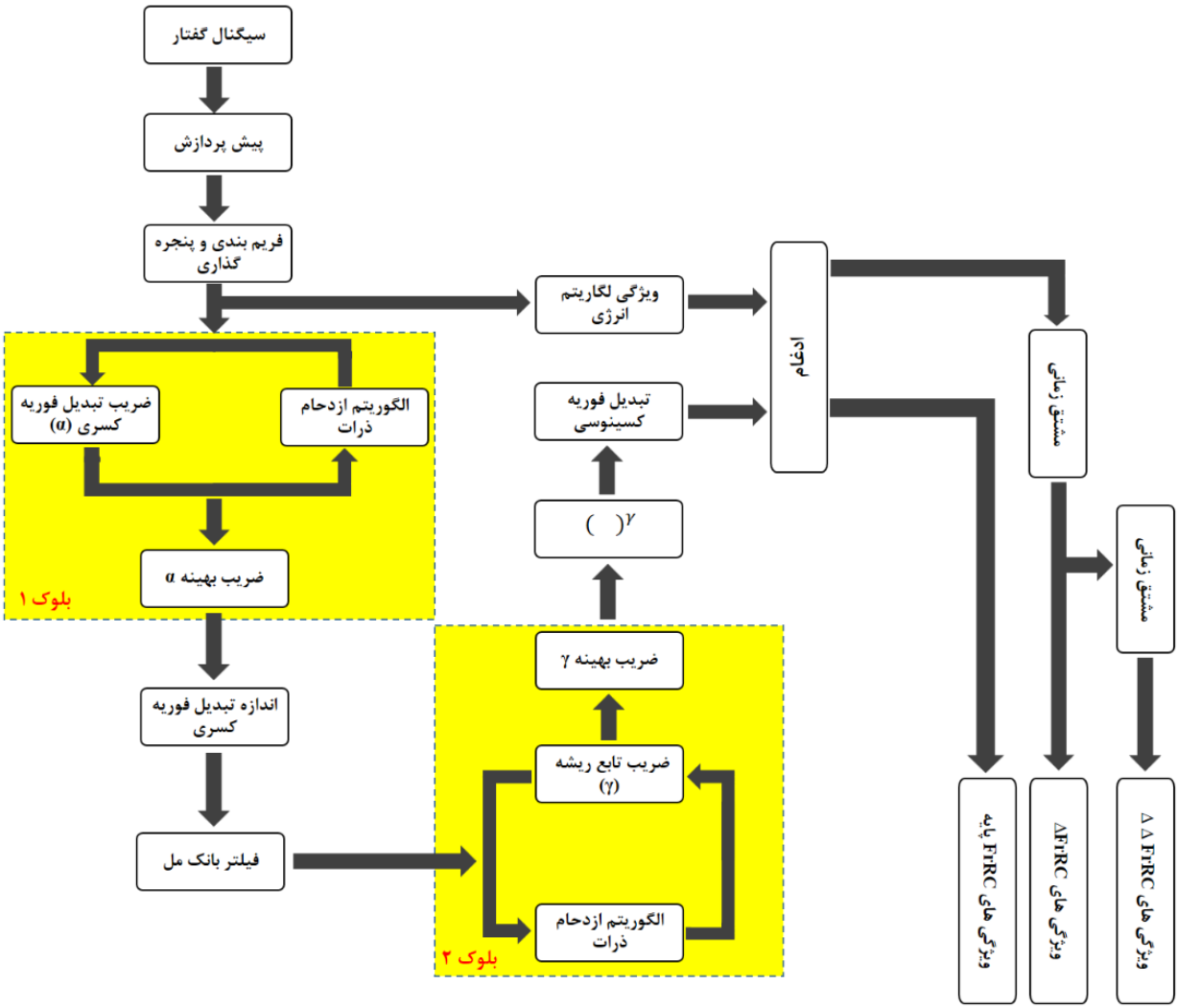
۴-۳-۱- تحلیل تئوری روش استخراج ویژگی پیشنهادی اول

در این بخش به تشریح روش استخراج ویژگی پیشنهادی پرداخته می‌شود. در ابتدا بلوک دیاگرام کلی و طبقات استفاده شده در روش استخراج ویژگی پیشنهادی FrRC معرفی شده، سپس نحوه استفاده از الگوریتم فراابتکاری برای دست‌یابی به ضرایب مناسب تبدیل فوریه کسری و تابع ریشه تشریح می‌گردد. در بخش بعدی نیز تحلیل ریاضی مربوط به روش استخراج ویژگی FrRC بیان می‌شود.

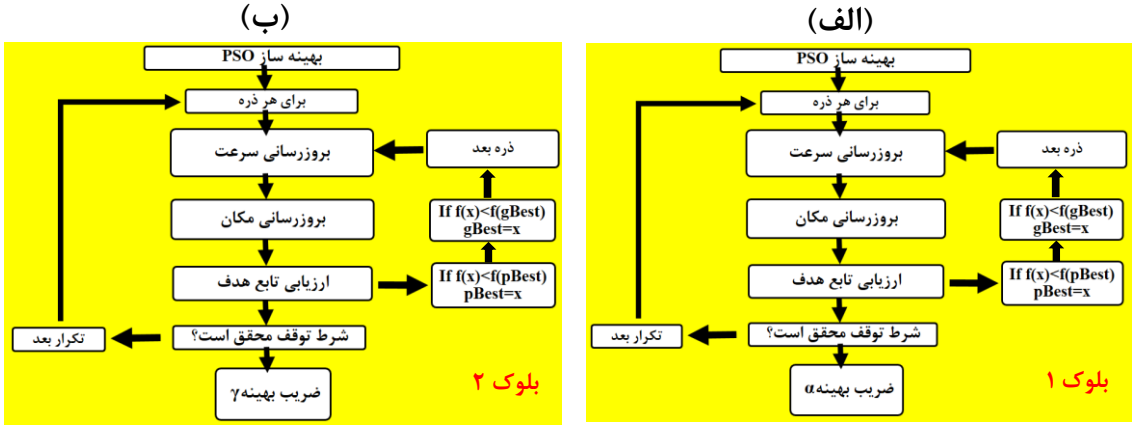
۴-۳-۱-۱- روش استخراج ویژگی پیشنهادی مبتنی بر تبدیل فوریه کسری و تابع ریشه(FrRC)

در شکل ۴-۹ روش استخراج ویژگی پیشنهادی مبتنی بر تبدیل فوریه کسری نمایش داده شده است. الگوریتم FrRC دارای دو مرحله است. در مرحله اول بصورت برون خط، مقادیر بهینه ضرایب آلفا و گاما به‌ترتیب برای تبدیل فوریه کسری و تابع ریشه توسط دو بلوک ۱ و ۲ بدست می‌آیند. در مرحله دوم که بصورت برخط است، دو بلوک ۱ و ۲ از ساختار الگوریتم FrRC حذف می‌گردند و تنها مقادیر بهینه بدست آمده توسط آن دو بلوک (در مرحله اول) برای استخراج ویژگی FrRC مورد استفاده قرار می‌گیرند. همانطور که در این شکل مشاهده می‌شود، در این روش ابتدا برروی سیگنال‌های گفتار ورودی پیش پردازش‌هایی از قبیل اعمال فیلتر پیش تاکید، هم طول کردن و ... اعمال می‌گردد، سپس سیگنال‌های صوتی به فریم‌های زمانی شامل تعداد اختیاری نمونه تقسیم می‌شود. در بیشتر سیستم‌ها، هم پوشانی بین فریم‌ها برای هموارسازی گذرای فریم به فریم استفاده می‌شود. سپس فریم‌ها توسط پنجره همینگ، پنجره‌گذاری می‌شوند تا ناپیوستگی در لبه‌ها حذف شود. در مرحله بعد سیگنال‌های گفتار که به فریم‌های هم طول تقسیم شده‌اند وارد بلوک ۱ که جزئیات این بلوک در شکل ۴-۱۰-الف نشان داده شده است می‌شوند. در این بلوک با استفاده از روش بهینه‌سازی ازدحام ذرات، ضریب بهینه برای تبدیل فوریه کسری با توجه به تابع هدف که در اینجا کمینه شدن خطای بازشناسی است، بدست می‌آید.

سپس با توجه به ضریب بدست آمده از بلوک ۱، برای هر فریم، تبدیل فوریه کسری برای استخراج مولفه‌های فرکانسی از سیگنال حوزه زمان محاسبه اعمال می‌گردد. به دلیل اینکه تبدیل فوریه کسری این سیگنال‌های صوتی بصورت مختلط است، بعد از اعمال تبدیل فوریه کسری بر روی هر فریم، اندازه آنها محاسبه می‌شوند. در مرحله بعد، فیلتر مقیاس مل فرکانسی بر فریم‌های اندازه تبدیل فوریه کسری اعمال می‌شود. این مقیاس تقریباً برای فرکانس‌های زیر یک کیلوهرتز بصورت خطی است و برای فرکانس‌های بیش از یک کیلو هرتز بصورت لگاریتمی می‌باشد



شکل ۴-۹: بلوک دیاگرام روش استخراج ویژگی پیشنهادی FrRC

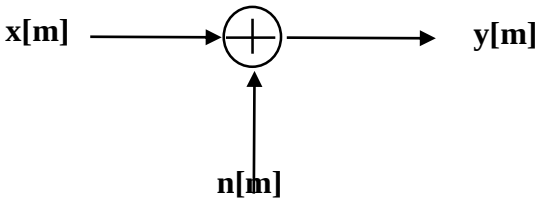


شکل ۴-۱۰: ساختار داخلی بلوک ۱ و بلوک ۲ استفاده شده در روش استخراج ویژگی پیشنهادی FrRC به منظور یافتن ضرایب بهینه آلفا و گاما

سپس سیگنال‌های عبور کرده از فیلتر مقیاس مل، وارد بلوک ۲ که جزئیات این بلوک در شکل ۴-۹-ب نشان داده شده است، می‌شوند. در این بلوک از طریق روش بهینه‌سازی PSO، ضریب بهینه گاما برای تابع ریشه با توجه به تابع هدف تعریف شده برای الگوریتم که همان تابع هدف بلوک ۱ می‌باشد، بدست می‌آید. لازم به ذکر است که دو بلوک ۱ و ۲ بصورت همزمان و وابسته به هم اجرا می‌گردند و در واقع الگوریتم بهینه‌سازی استفاده شده در این ساختار پیشنهادی، یک الگوریتم بهینه سازی با دو متغیر خروجی است. در این الگوریتم بهینه ساز، مقدار جمعیت اولیه ۱۰ و تعداد تکرار برابر با ۱۰۰ در نظر گرفته شده است. سپس با توجه به ضریب بهینه گاما استخراج شده از بلوک ۲، بر روی سیگنال‌های عبور کرده از بانک فیلتر مل، تابع ریشه اعمال می‌گردد. سپس بر روی سیگنال‌های خروجی طبقه قبل، تبدیل فوریه کسینوسی اعمال می‌شود. تا این مرحله ویژگی‌های پایه FrRC بدست می‌آیند. سپس از طریق بلوک تجمیع، ویژگی‌های FrRC پایه با لگاریتم انرژی تجمیع می‌شود که با این کار ویژگی انرژی در کنار ویژگی‌های FrRC پایه، ویژگی‌های FrRC را تشکیل می‌دهند. در کنار ویژگی‌های FrRC می‌توان از مشتق‌های اول و دوم این ویژگی نیز در کنار FrRC استفاده نمود.

۴-۳-۱-۲- تحلیل ریاضی الگوریتم FrRC

در این قسمت، ما از ویژگی‌های FrRC بعنوان ویژگی آکوستیک برای بسط رابطه بین گفتار تمیز و نویزی استفاده کرده‌ایم. شکل ۴-۱۱ یک مدل زمانی برای گفتار تخریب شده توسط نویز جمع شونده را نشان می‌دهد. سیگنال گفتار نویزی مشاهده شده y[m] که اندیس زمانی آن بصورت گسسته می‌باشد از طریق سیگنال گفتار تمیز x[m] با نویز جمع‌شونده n[m] توسط رابطه زیر ایجاد می شود.



شکل ۴-۱۱: مدل زمانی برای گفتار تخریب شده توسط نویز جمع‌شونده

$$y[m] = x[m] + n[m] \quad (۱-۴)$$

بعد از اعمال تبدیل فوریه کسری گسسته زمان کوتاه، رابطه هم ارزی ۴-۲ در حوزه کسری ایجاد می‌گردد.

$$Y_\alpha[k] = X_\alpha[k] + N_\alpha[k] \quad (۲-۴)$$

در رابطه فوق $X_\alpha[k]$ ، $N_\alpha[k]$ و $Y_\alpha[k]$ از طریق روابط ٤-۳ تا ٤-۵ بدست می‌آیند.

$$X_\alpha[k] = \sum_{m=-\infty}^{+\infty} A_\alpha e^{j\pi(m^2 cot\alpha - 2Km\, csc\alpha + K^2 cot\alpha)} x[m] \quad (۳-۴)$$

$$N_\alpha[k] = \sum_{m=-\infty}^{+\infty} A_\alpha e^{j\pi(m^2 cot\alpha - 2Km\, csc\alpha + K^2 cot\alpha)} n[m] \quad (۴-۴)$$

$$Y_\alpha[k] = \sum_{m=-\infty}^{+\infty} A_\alpha e^{j\pi(m^2 cot\alpha - 2Km\, csc\alpha + K^2 cot\alpha)} y[m] \quad (۵-۴)$$

در روابط فوق $A_\alpha = \frac{1}{\sqrt{|sin\alpha|}} e^{-j(\frac{\pi}{4} sgn(sin\alpha) - \frac{\alpha}{2})}$ و $p \in R$. $\alpha = \frac{p\pi}{2}$ می‌باشند. در اینجا، k اندیس فرکانسی است. طیف توان گفتار نویزی توسط رابطه ۴-۶ بدست می‌آید.

$$|Y_\alpha[k]|^2 = |X_\alpha[k] + N_\alpha[k]|^2 = (|X_\alpha[k]||N_\alpha[k]|)(|X_\alpha[k]||N_\alpha[k]|)^* = \quad (٦-٤)$$
$$|X_\alpha[k]|^2 + |N_\alpha[k]|^2 + 2|X_\alpha[k]||N_\alpha[k]|cos\beta_k$$

در رابطه فوق β_k زاویه بین دو متغیر مختلط $X_\alpha[k]$ و $N_\alpha[k]$ است. با فرض اینکه این دو متغیر از هم مستقل هستند و هیچ گونه وابستگی به هم ندارند، بنابراین این رابطه با توجه به [۴] بصورت رابطه ۴-۷ قابل بازنویسی است.

$$|Y_\alpha[k]|^2 = |X_\alpha[k]|^2 + |N_\alpha[k]|^2 \quad (٧-٤)$$

حذف کردن بخش فاز یک کار رایج فرمول‌بندی گفتار نویزی در حوزه طیف توان است [۲۵],[۴]. در الگوریتم‌های بهبود گفتار به سادگی قابل مشاهده است که حذف این بخش یک علت جزئی برای کاهش عملکرد بهبود در SNRهای پایین (در حد صفر دسیبل) می‌باشد [۴] [۲۶].

با اعمال یک مجموعه فیلترهای مقیاس مل (L تعداد کل فیلترها) به طیف توان رابطه ٤-۷، ما l امین انرژی‌های فیلتر مل را برای گفتار نویزی، گفتار تمیزو نویز بصورت روابط ۴-۸ تا ۴-۱۰ داریم.

$$\left|\overline{Y_\alpha[l]}\right|^2 = \sum_k W_k[l]|Y_\alpha[k]|^2 \quad (٨-٤)$$

$$\left|\overline{X_\alpha[l]}\right|^2 = \sum_k W_k[l]|X_\alpha[k]|^2 \quad (٩-٤)$$

$$\left|\overline{N_\alpha[l]}\right|^2 = \sum_k W_k[l]|N_\alpha[k]|^2 \quad (١٠-٤)$$

پس، رابطه زیر در حوزه بانک فیلتر مل برای l امین خروجی بانک فیلتر مل برقرار است.

$$\left|\overline{Y_\alpha[l]}\right|^2 = \left|\overline{X_\alpha[l]}\right|^2 + \left|\overline{N_\alpha[l]}\right|^2 \quad (١١-٤)$$

با در نظر گرفتن تابع ریشه به طرفین رابطه ۴-۱۱، در حوزه بانک فیلتر مل و با اعمال تابع ریشه با استفاده از علامت‌گذاری برداری رابطه ۴-۱۲ بدست آمد.

$$\left(\left|\overline{Y_\alpha[l]}\right|^2\right)^\gamma = \left(\left|\overline{X_\alpha[l]}\right|^2 + \left|\overline{N_\alpha[l]}\right|^2\right)^\gamma \quad (١٢-٤)$$

در رابطه فوق γ ضریب تابع ریشه می‌باشد. از آنجایی که این ضریب یک عدد حقیقی بین صفر و یک است، می‌توان این رابطه را براساس بسط دو جمله‌ای توسط تابع گاما بازنویسی کرد و به رابطه ۴-۱۳ دست یافت.

$$\left(\left|\overline{Y_\alpha[l]}\right|^2\right)^\gamma = \left(\left|\overline{N_\alpha[l]}\right|^2 + \left|\overline{X_\alpha[l]}\right|^2\right)^\gamma = \sum_{n=0}^{\infty} \frac{\mathcal{T}(\gamma+1)}{\mathcal{T}(\gamma-n+1)} \frac{\left|\overline{N_\alpha[l]}\right|^{2n} \left|\overline{X_\alpha[l]}\right|^{2(\gamma-n)}}{n!} \qquad (۱۳-۴)$$

با ساده‌سازی رابطه فوق براساس تابع گاما و روابط حاکم بر این عملگر می‌توان رابطه ۴-۱۳ را بصورت رابطه ۴-۱۴ بازنویسی نمود.

$$\left(\left|\overline{Y_\alpha[l]}\right|^2\right)^\gamma = \left|\overline{X_\alpha[l]}\right|^{2\gamma} + \frac{\mathcal{T}(\gamma+1)}{\mathcal{T}(\gamma)} \frac{\left|\overline{N_\alpha[l]}\right|^2 \left|\overline{X_\alpha[l]}\right|^{2(\gamma-1)}}{1!} + \frac{\mathcal{T}(\gamma+1)}{\mathcal{T}(\gamma-1)} \frac{\left|\overline{N_\alpha[l]}\right|^4 \left|\overline{X_\alpha[l]}\right|^{2(\gamma-2)}}{2!} + \cdots = \qquad (۱۴-۴)$$

$$\left(\left|\overline{Y_\alpha[l]}\right|^2\right)^\gamma = \left(\left|\overline{X_\alpha[l]}\right|^2\right)^\gamma + \sum_{n=1}^{\infty} \frac{\left|\overline{N_\alpha[l]}\right|^{2n} \left|\overline{X_\alpha[l]}\right|^{2(\gamma-n)}}{n!} \prod_{m=0}^{n-1} (\gamma - m)$$

رابطه ۴-۱۴ را می‌توان بصورت رابطه ۴-۱۵ بازنویسی نمود که این عبارت بعد از اعمال تابع ریشه می‌باشد.

$$\widetilde{Y_\alpha} = \widetilde{X_\alpha} + \sum_{n=1}^{\infty} \frac{\widetilde{X_\alpha}\left(\sqrt[\gamma]{\left(\frac{\widetilde{N_\alpha}}{\widetilde{X_\alpha}}\right)^n}\right)}{n!} \prod_{m=0}^{n-1} (\gamma - m) \qquad (۱۵-۴)$$

با اعمال تبدیل DCT بر روی رابطه ۴-۱۵، رابطه نویزی در حوزه کپستروم بصورت رابطه ۴-۱۶) بدست می‌آید.

$$\dot{Y_\alpha} = \dot{X_\alpha} + C\left(\sum_{n=1}^{\infty} \frac{C^{-1}(\dot{X_\alpha})\left(\sqrt[\gamma]{\left(C^{-1}\left(\frac{N_\alpha}{X_\alpha}\right)\right)^n}\right)}{n!} \prod_{m=0}^{n-1} (\gamma - m)\right) \qquad (۱۶-۴)$$

در رابطه فوق C، $\dot{X_\alpha}$، $\dot{N_\alpha}$ و $\dot{Y_\alpha}$ به ترتیب ماتریس DCT، ویژگی‌های FrRC گفتار تمیز، ویژگی‌های FrRC نویز و ویژگی‌های FrRC گفتار نویزی می‌باشد. رابطه ۴-۱۶، رابطه بین ویژگی‌های FrRC گفتار نویزی را با گفتار تمیز و نویز نشان می‌دهد.

در [۴]، تحلیل ریاضی بین ویژگی‌های MFCC گفتار تمیز، نویز و گفتار نویزی بصورت رابطه (۴-۱۷) گزارش شده است.

$$\ddot{Y} = \ddot{X} + C\left(\log\left(1 + e^{C^{-1}(\ddot{N}-\ddot{X})}\right)\right) \quad (۴-۱۷)$$

در رابطه فوق C، $\ddot{X}$، $\ddot{N}$ و $\ddot{Y}$ به ترتیب ماتریس DCT، ویژگی‌های MFCC گفتار تمیز، ویژگی‌های MFCC نویز و ویژگی‌های MFCC گفتار نویزی می‌باشد.

برای اینکه میزان مقاوم بودن روش استخراج ویژگی FrRC به نسبت روش MFCC سنجیده شود، در ادامه به مقایسه حدی حالات مختلف روابط (۴-۱۶) و (۴-۱۷) در سطوح سیگنال به نویز مختلف پرداخته شده است.

$$\text{If SNR} = A \Rightarrow \left(\frac{\dot{X}_\alpha}{\dot{N}_\alpha}\right) = K \Rightarrow \dot{Y}_\alpha = \dot{X}_\alpha + C\left(\sum_{n=1}^{\infty} \frac{C^{-1}(\dot{N}_\alpha)K^{\left(1-\frac{n}{\gamma}\right)}}{n!} \prod_{m=0}^{n-1}(\gamma - m)\right) \quad (۴-۱۸)$$

$$\text{If } A \gg 0\text{dB} \Rightarrow \left(\frac{\dot{N}_\alpha}{\dot{X}_\alpha}\right) \longrightarrow 0 \Rightarrow \sqrt[\gamma]{\left(C^{-1}\left(\frac{\dot{N}_\alpha}{\dot{X}_\alpha}\right)\right)^n} \rightarrow 0 \Rightarrow \ddot{Y}_\alpha \rightarrow \ddot{X}_\alpha \quad (۴-۱۹)$$

$$\text{If } A \gg 0\text{dB} \Rightarrow \ddot{N} \ll \ddot{X} \Rightarrow e^{-C^{-1}(\ddot{X})} \rightarrow 0 \Rightarrow C\left(\log\left(1 + e^{-C^{-1}(\ddot{X})}\right)\right) \rightarrow 0 \Rightarrow \ddot{Y} \rightarrow \ddot{X} \quad (۴-۲۰)$$

روابط (۴-۱۹) و (۴-۲۰) که برای حالت حدی SNR>>0dB بدست آمده است، نشان دهنده این حقیقت می‌باشد که در محیط‌هایی که شدت نویز به نسبت سیگنال خیلی کمتر است، ویژگی‌های استخراج شده از گفتار نویزی شبیه به ویژگی‌های استخراج شده از گفتار تمیز بدست می‌آیند.

$$\text{If } A = 0\text{dB}$$

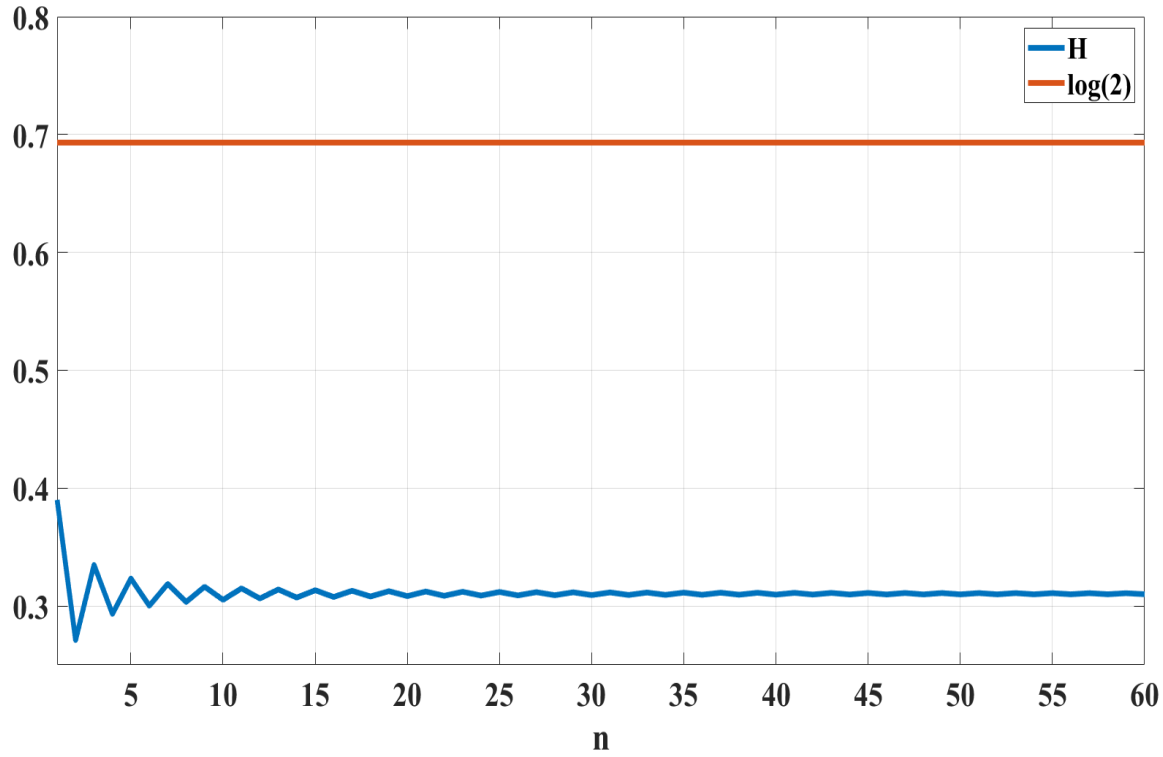
$$\Rightarrow K = 1 \Rightarrow \dot{Y}_\alpha = \dot{X}_\alpha + C\left(\sum_{n=1}^{\infty} \frac{C^{-1}(\dot{N}_\alpha)}{n!} \prod_{m=0}^{n-1}(\gamma - m)\right)$$

$$\Rightarrow \dot{Y}_\alpha = \dot{X}_\alpha + \dot{N}_\alpha\left(\sum_{n=1}^{\infty} \frac{\prod_{m=0}^{n-1}(\gamma - m)}{n!}\right). \textit{Assume } H = \sum_{n=1}^{\infty} \frac{\prod_{m=0}^{n-1}(\gamma - m)}{n!} \quad (۴-۲۱)$$

$$\Rightarrow \dot{Y}_{\alpha} = \dot{X}_{\alpha} + H \times \dot{N}_{\alpha}$$

$$\text{If } A = 0\text{dB} \Rightarrow \ddot{N} = \ddot{X} \Rightarrow e^{C^{-1}(\ddot{N}-\ddot{X})} = 1 \Rightarrow \ddot{Y} = \ddot{X} + \log(2) \qquad (۲۲-۴)$$

با توجه به رابطه (۴-۲۱) مشاهده می‌شود که در سطح سیگنال به نویز برابر با صفر دسیبل، برای روش استخراج ویژگی پیشنهادی FrRC، ویژگی‌های گفتار نویزی از حاصل‌جمع ویژگی‌های گفتار تمیز با H برابر ویژگی های FrRC نویز بدست می‌آید. از طرف دیگر با توجه به رابطه (۴-۲۲) در سطح سیگنال به نویز صفر دسیبل، ویژگی‌های گفتار نویزی MFCC برابر با حاصل‌جمع ویژگی‌های گفتار تمیز با یک عدد ثابت (log(2)) حاصل می‌گردد. حال در بین این دو روش استخراج ویژگی، روشی مقاوم‌تر است که کمترین تاثیر را روی ویژگی‌های گفتار بدون نویز ایجاد کند. بدین منظور، نمودار این دو عدد ثابت در شکل (۴-۱۲) رسم شده است.



شکل ۴-۱۲: مقایسه مقدار ثابت روش استخراج ویژگی FrRC (H) با مقدار ثابت روش استخراج ویژگی MFCC (log2) به ازای nهای مختلف

لازم به ذکر است که شکل (۴-۱۲) به ازای γ برابر با ۰/۳۹ ترسیم شده است. با توجه به شکل (۴-۱۲) می توان مشاهده نمود که مقدار H با افزایش n که در روابط مربوط به ویژگی های FrRC موجود است به یک عدد ثابت حدود ۰/۳۱ همگرا می‌شود که نسبت به عدد ثابت موجود در رابطه MFCC (log(2)) که مقداری در حدود ۰/۶۹ می باشد، مقدار کوچکتری است. از آنجایی که در سطح سیگنال به نویز صفر دسیبل، مقدار دامنه نویز و گفتار تمیز مقادیری بسیار کوچکتر از یک می‌باشند، به همین خاطر در روش FrRC، با ضرب شدن عدد ثابت H در ویژگی‌های FrRC سیگنال نویز، تاثیر نویز به شدت کاهش می‌یابد. در حالی که در روش MFCC، تمامی ویژگی‌های گفتار نویزی به نسبت گفتار بدون نویز حدوداً به اندازه ۰/۶۹ بزرگتر می‌شود. بنابراین می‌توان اینطور نتیجه گرفت که ویژگی‌های استخراج شده توسط روش استخراج ویژگی FrRC به نسبت روش MFCC در سطح سیگنال به نویز صفر دسیبل، کمتر تحت تاثیر نویز قرار می‌گیرند و یا به عبارت دیگر ویژگی‌هایی مقاوم‌تر از MFCC نتیجه می‌دهند.

$$If\ A \ll 0dB \ \Rightarrow\ K^{\left(1-\frac{n}{\gamma}\right)} = \frac{K}{K^{\left(\frac{n}{\gamma}\right)}}\ .0 < \gamma < 1\ .n \gg 1 \Rightarrow \left(\frac{n}{\gamma}\right) \gg 1\ .\ K \ll 1 \Rightarrow\ K^{\left(\frac{n}{\gamma}\right)} \ll 1$$

$$\Rightarrow K^{\left(\frac{n}{\gamma}\right)} \ll K\ \Rightarrow \frac{K}{K^{\left(\frac{n}{\gamma}\right)}} \gg 1\ .assume\ \frac{K}{K^{\left(\frac{n}{\gamma}\right)}} = T(n).\ T(n) \gg 1$$

$$\Rightarrow \dot{\mathrm{Y}}_{\alpha} = \dot{\mathrm{X}}_{\alpha} + \ \dot{N}_{\alpha}\left(\sum_{n=1}^{\infty}\frac{T(n)}{n!}\prod_{m=0}^{n-1}(\gamma - m)\right).H = \ \sum_{\mathrm{n}=1}^{\infty}\frac{\prod_{\mathrm{m}=0}^{\mathrm{n}-1}(\gamma - \mathrm{m})}{\mathrm{n}!}$$

$$\Rightarrow \dot{\mathrm{Y}}_{\alpha} = \dot{\mathrm{X}}_{\alpha} + \ \mathrm{H} \times \mathrm{T(n)} \times \dot{\mathrm{N}}_{\alpha} \qquad (۲۳-۴)$$

$$\mathrm{If\ A} \ll 0\mathrm{dB} \Rightarrow \ddot{\mathrm{N}} \gg\ \ddot{\mathrm{X}} \Rightarrow \mathrm{C}\left(\log\left(1 + \mathrm{e}^{\mathrm{C}^{-1}(\ddot{\mathrm{N}} - \ddot{\mathrm{X}})}\right)\right) \rightarrow \mathrm{C}\left(\log\left(1 + \mathrm{e}^{\mathrm{C}^{-1}(\ddot{\mathrm{N}})}\right)\right)$$

$$\Rightarrow \ddot{\mathrm{Y}} = \ddot{\mathrm{X}} + \ \mathrm{C}\left(\log(1 + \mathrm{e}^{\mathrm{C}^{-1}(\ddot{\mathrm{N}})})\right) \qquad (۲۴-۴)$$

در محیط‌های نویزی شدید که سطح سیگنال به نویز خیلی کمتر از صفر دسیبل می‌باشد، با توجه به رابطه (۴-۲۴) مشاهده می‌گردد، که ویژگی‌های MFCC گفتار نویزی به شدت از طریق ویژگی‌های MFCC نویز تخریب می‌شوند. از آنجایی که در این سطح سیگنال به نویز پایین، سطح دامنه نویز از

سطح دامنه گفتار بسیار بالاتر می‌باشد به همین خاطر با توجه به رابطه نمایی و لگاریتمی روش استخراج ویژگی MFCC که در شکل (۴-۲۴) نشان داده شده است، ویژگی‌های MFCC نویز تقویت می‌شود که سبب تخریب زیاد ویژگی‌های MFCC گفتار تمیز می‌گردد.

در رابطه (۴-۲۳) که مربوط به ویژگی‌های FrRC است، دو عدد ثابت که هر دو وابسته به مقدار n می‌باشند در ویژگی‌های FrRC نویز ضرب می‌شوند. با توجه به نمودار H که در شکل (۴-۱۲) نمایش داده شده است، مقدار این عدد ثابت در حدود ۰/۳۱ است و با توجه به رابطه (۴-۲۳) مقدار عدد ثابت T(n) با افزایش مقدار n کاهش پیدا می‌کند. بنابراین این دو عدد ثابت، نقش دو تضعیف کننده ویژگی-های نویز را دارد که باعث کاهش اثر ویژگی‌های FrRC نویز بر روی ویژگی‌های گفتار تمیز می‌گردند.

با توجه به نکات گفته شده، می‌توان به این نتیجه رسید که ویژگی‌های استخراج شده توسط روش استخراج ویژگی پیشنهادی FrRC در محیط‌های با سطح سیگنال به نویز بسیار پایین می‌تواند بسیار مقاوم‌تر به نسبت روش استخراج ویژگی MFCC نقش آفرینی کند.

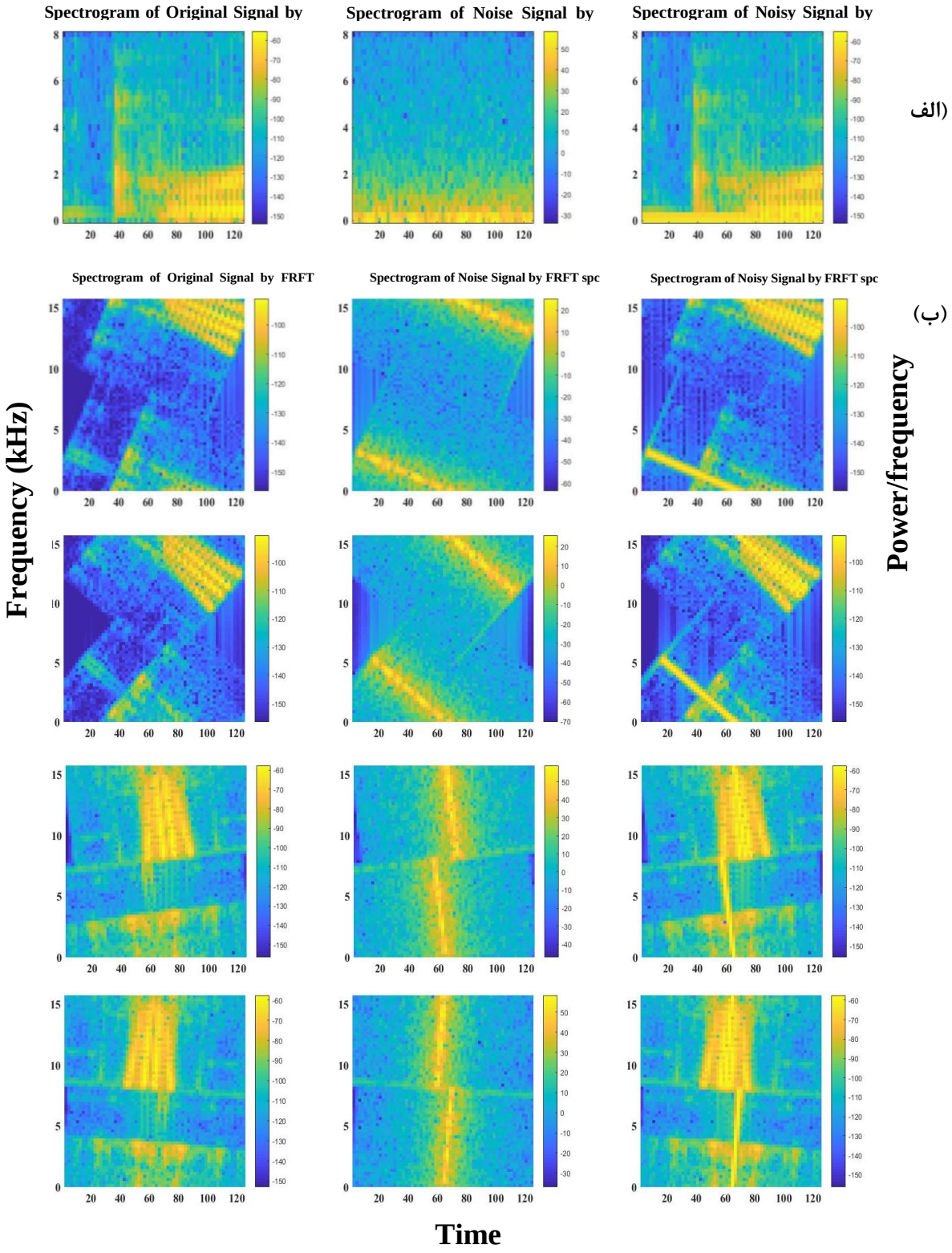
۴-۳-۲- شبیه سازی

در این بخش ابتدا به نمایش طیف نگاره مبتنی بر تبدیل فوریه معمولی و تبدیل فوریه کسری پرداخته شده و نشان داده می‌شود که تبدیل فوریه کسری به چه نحوی می‌تواند طیف سیگنال را مورد چرخش قرار دهد تا بتوان از این مزیت به منظور استفاده بهینه برای کاربرد مقاوم‌سازی استفاده نمود. سپس در بخش دوم، نتایج حاصل از شبیه سازی سیستم بازشناسی گفتار مبتنی بر روش‌های مختلف استخراج ویژگی در قیاس با روش استخراج ویژگی پیشنهادی FrRC مورد بررسی قرار خواهد گرفت.

۴-۳-۲-۱- طیف نگاره

طیف نگاره یک نمایش طیفی متغیر با زمان است که نحوه تغییرات چگالی طیف سیگنال را با زمان نشان می‌دهد. محور عمودی و محور افقی به ترتیب زمان و فرکانس می‌باشند. یک طیف نگاره معمولی

از طریق اعمال تبدیل فوریه زمان کوتاه (STFT) بر روی سیگنال زمانی ایجاد می‌گردد که در شکل ۴-۱۲-الف نشان داده شده است. در شکل ۴-۱۲-الف طیف نگاره معمولی برای سیگنال تمیز، نویز و سیگنال نویزی نمایش داده شده است. در این پیاده‌سازی علاوه بر طیف نگاره معمولی از طیف نگاره مبتنی بر تبدیل فوریه کسری نیز استفاده شده است. در واقع در این طیف نگاره از تبدیل فوریه کسری زمان کوتاه به جای تبدیل فوریه معمولی زمان کوتاه استفاده می‌شود. همانطورکه در بخش ۴-۳ توضیح داده شد، این تبدیل فوریه حالت بسط یافته‌ی تبدیل فوریه معمولی می‌باشد و از طریق پارامتر α قابلیت چرخش دارد. همانطورکه در شکل ۴-۱۲-ب تا ۴-۱۲-ه مشاهده می‌شود، طیف نگاره گفتار تمیز، نویز و گفتار نویزی به ازای α=0.25-0.45- 0.92- 1.05 ترسیم شده است.



شکل۴-۱۲: طیف نگاره گفتار تمیز، نویز Factory و گفتار نویزی به ترتیب از چپ به راست – شکل (الف) طیف نگاره با استفاده از FFT، شکل (ب) تا (ه) طیف نگاره با استفاده از FrFT به ترتیب با $\alpha = 0.25 - 0.45$

در واقع هدف از تکنیک استخراج ویژگی پیشنهادی بدست آوردن مجموعه‌ای از ویژگی‌های مقاوم در برابر نویزهای مختلف با سطوح SNR مختلف برای دست‌یابی به صحت بازشناسی خوب در محیط نویزی است. ایده اصلی روش استخراج ویژگی پیشنهادی FrRC استفاده هدفمند از خاصیت چرخش تبدیل فوریه کسری می‌باشد. بدین صورت که با چرخاندن زمان-فرکانس به مقدار بهینه‌ای که توسط الگوریتم ارائه شده در شکل (۴-۱۰) بدست می‌آید و سپس با استفاده از یک فیلتر ثابت، اثر نویز تا حد قابل قبولی در محیط‌های نویزی کاهش می‌یابد. نتایج پیاده‌سازی این الگوریتم در ادامه فصل ارائه شده است.

۴-۳-۲-۲- دادگان

به منظور ارزیابی میزان صحت بازشناسی از دادگان TIMIT، ۹ فونوم که از هرکدام از این فونوم‌ها ۵۳ صوت موجود می‌باشد استفاده شده است. لازم به ذکر است که فرکانس نمونه‌برداری این فونوم‌ها ۱۶ کیلوهرتز می‌باشد. از بین مجموع داده‌ها، ۹۰ درصد داده‌ها به منظور آموزش سیستم و ۱۰ درصد مابقی به منظور آزمون سیستم استفاده شده است. لیست فونوم‌های استفاده شده در این قسمت در جدول ۷-۴ نشان داده شده است. طبقه‌بند استفاده شده در این سیستم بازشناسی یک ماشین بردار پشتیبان با کرنل غیرخطی بصورت یک چند جمله‌ای درجه ۴ است.

جدول ۴-۷: فونوم‌های استفاده شده در این سیستم بازشناسی گفتار

Phoneme List								
Ao	Eh	en	Hv	Ix	Iy	Jh	Sh	Ux

۴-۳-۲-۳- شبیه سازی الگوریتم بهینه‌ساز به منظور یافتن ضرایب بهینه آلفا و گاما

در این بخش روش‌های استخراج ویژگی و همچنین روش استخراج ویژگی پیشنهادی FrRC با پنج نویز مختلف M109، Babble، Factory1، Factory2 و Destroyer در پنج سطح سیگنال به نویز از dB -10تا 10dB به منظور ارزیابی سیستم بازشناسی استفاده شده‌اند. لازم به ذکر است نویزهای استفاده

شده از پایگاه Noisex-92 تهیه گردیده است. با اجرای الگوریتم بهینه‌سازی تشریح شده در بخش ۴-۳-۱-۱ به منظور دستیابی به ماکزیمم صحت بازشناسی در محیط نویزی برای آلفا تبدیل فوریه کسری و ضریب گاما تابع ریشه به مقادیری دست یافته‌ایم که این مقادیر در جدول ۴-۸ آورده شده است.

جدول ۴-۸: مقادیر بهینه بدست آمده برای ضریب آلفا تبدیل فوریه کسری و ضریب گاما تابع ریشه

Noise Type	α	γ
M109	1.0518	0.45473
Babble	0.92519	0.39955
Destroyer ops	1.4064	0.3205
Factory1	3	0.43728
Factory2	3	0.39306

باتوجه به جدول ۴-۸ مشاهده می‌شود که به ازای نویزهای مختلف، ضرایب بهینه آلفا و گاما می‌تواند متفاوت باشد. برای ادامه کار از بین ضرایب آلفا و گاما استخراجی، مقادیری را در نظر گرفته‌ایم که برای همه نویزهای بررسی شده صحت بازشناسی بالایی را نتیجه دهد. به همین خاطر، برای آلفا مقدار ۰/۹۲۵۱۹ و برای گاما مقدار ۰/۳۹۹۵۵ در نظر گرفته شده است.

۴-۳-۲-۴- نتایج شبیه سازی سیستم بازشناسی گفتار با استفاده از روش‌های استخراج ویژگی مختلف و همچنین پیشنهادی (FrRC)

نتایج بدست آمده از شبیه سازی سیستم بازشناسی به ازای روش‌های استخراج ویژگی مختلف به همراه روش استخراج ویژگی پیشنهادی در جداول ۴-۹ تا ۴-۱۳ ارائه و گزارش شده است.

همانطورکه در جدول ۴-۹ مشاهده می‌شود، صحت بازشناسی سیستم در حضور نویز M109 با استفاده از روش استخراج ویژگی پیشنهادی FrRC در تمامی سطوح SNR به جز 10dB، بالاترین مقدار را در قیاس با سایر روش‌های استخراج ویژگی مرسوم نتیجه داده است.

جدول ۴-۹: نتایج شبیه سازی سیستم بازشناسی (صحت بازشناسی (درصد)) با روش‌های مختلف استخراج ویژگی به همراه روش پیشنهادی FrRC درحضور نویز M109 با پنج سطح سیگنال به نویز از -10dB تا 10dB

Noise Type		M109				
SNR (dB)		-10dB	-5dB	0dB	5dB	10dB
Feature Extraction Methods	LPC[۲۱ و ۲۲]	27.46	36.68	40.46	43.60	46.75
	MFCC [۱۲ و ۲۰ ،۱۸]	28.93	31.44	38.15	48.63	**62.26**
	FMFCC[۶۱]	23.69	33.12	42.97	51.36	60.58
	RMFCC[۶۲]	21.17	32.49	45.07	55.13	57.23
	LSF[۲۴]	27.04	31.03	38.78	52.41	58.07
	PLP[۱۹ و ۲۰]	24.53	37.31	43.81	46.5	52.20
	Proposed (FrRC)	**36.27**	**50.31**	**55.13**	**55.34**	55.13

جدول ۴-۱۰: نتایج شبیه سازی سیستم بازشناسی (صحت بازشناسی (درصد)) با روش‌های مختلف استخراج ویژگی به همراه روش پیشنهادی FrRC در حضور نویز Babble با پنج سطح سیگنال به نویز از -10dB تا 10dB

Noise Type		Babble				
SNR (dB)		-10dB	-5dB	0dB	5dB	10dB
Feature Extraction Methods	LPC[۲۱ و ۲۲]	28.09	39.20	47.17	51.15	51.99
	MFCC [۱۲ و ۲۰ ،۱۸]	28.93	33.75	43.39	55.97	65.19
	FMFCC[۶۱]	24.11	37.31	50.73	59.75	**65.82**
	RMFCC[۶۲]	21.17	30.82	45.49	54.50	57.02
	LSF[۲۴]	33.54	45.28	52.41	**61.21**	62.68
	PLP[۱۹ و ۲۰]	12.788	24.53	35.01	39.20	43.39
	Proposed (FrRC)	**35.01**	**46.54**	**54.71**	56.81	57.86

باتوجه به نتایج بدست آمده در جدول ۴-۱۰ که در حضور نویز Babble می‌باشد، مشاهده می‌شود که در محیط نویزی شدید با SNRهای صفر و کمتر از صفر دسیبل، باز هم روش استخراج ویژگی پیشنهادی بالاترین صحت بازشناسی را رقم زده است.

نتایج جدول ۴-۱۱ حاکی از این است که سیستم بازشناسی گفتار در حضور نویز Factory1 به ازای روش استخراج ویژگی FrRC توانسته است به نسبت تمامی روش‌های استخراج ویژگی دیگر در تمامی سطوح سیگنال به نویز بالاترین مقاومت در برابر نویز و یا بعبارت دیگر بالاترین صحت بازشناسی را ایجاد کند.

جدول ۴-۱۱: نتایج شبیه سازی سیستم بازشناسی (صحت بازشناسی (درصد)) با روش‌های مختلف استخراج ویژگی به همراه روش پیشنهادی FrRC در حضور نویز Factory1 با پنج سطح سیگنال به نویز از -10dB تا 10dB

Noise Type		Factory1				
SNR (dB)		-10dB	-5dB	0dB	5dB	10dB
Feature Extraction Methods	LPC[۲۱ و ۲۲]	19.49	20.75	28.09	35.43	42.97
	MFCC [۱۲ و ۲۰ ،۱۸]	18.45	25.99	32.07	39.20	55.13
	FMFCC[۶۱]	12.58	17.19	24.53	35.01	41.72
	RMFCC[۶۲]	16.35	23.27	37.31	49.05	55.55
	LSF[۲۴]	17.40	25.78	35.01	46.75	55.13
	PLP[۱۹ و ۲۰]	19.49	29.77	38.78	45.49	48.01
	Proposed (FrRC)	**29.35**	**34.17**	**46.96**	**55.34**	**56.18**

جدول ۴-۱۲: نتایج شبیه سازی سیستم بازشناسی (صحت بازشناسی (درصد)) با روش‌های مختلف استخراج ویژگی به همراه روش پیشنهادی FrRC در حضور نویز Factory2 با پنج سطح سیگنال به نویز از -10dB تا 10dB

Noise Type		Factory2				
SNR (dB)		-10dB	-5dB	0dB	5dB	10dB
Feature Extraction Methods	LPC[۲۱ و ۲۲]	39.41	47.38	49.05	51.57	54.92
	MFCC [۱۲ و ۲۰ ،۱۸]	20.75	28.09	37.52	49.89	64.36
	FMFCC[۶۱]	40.88	53.46	**62.05**	**65.20**	**66.87**
	RMFCC[۶۲]	23.48	41.09	50.31	58.49	59.12
	LSF[۲۴]	32.28	46.33	54.09	56.18	59.96
	PLP[۱۹ و ۲۰]	38.15	44.86	48.43	52.41	55.76
	Proposed (FrRC)	**48.43**	**54.51**	55.55	56.81	57.02

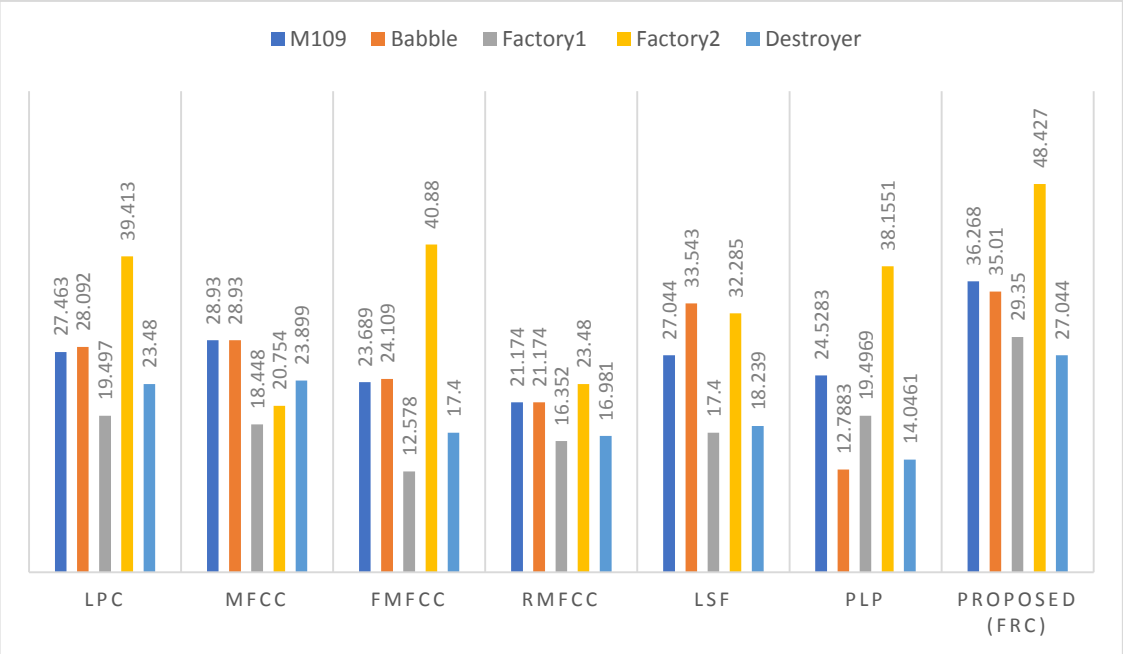
همانطورکه در جدول ۴-۱۲ مشاهده می‌گردد، سیستم بازشناسی گفتار در محیط نویزی با نویز Factory2 به ازای روش استخراج ویژگی FrRC در سطوح SNR کمتر از صفر در قیاس با سایر روش‌های استخراج ویژگی مرسوم، بیشترین صحت بازشناسی را نتیجه داده است. در حالی که در SNR های زیاد، سایر روش‌های استخراج ویژگی از قبیل MFCC بالاترین صحت بازشناسی را دارد.

سیستم بازشناسی گفتار در حضور نویز Destroyer به ازای SNRهای کمتر و مساوی صفر دسیبل بالاترین مقاومت و بعبارت دیگر بالاترین صحت بازشناسی را نتیجه داده است.

جدول ۴-۱۳: نتایج شبیه سازی سیستم بازشناسی (صحت بازشناسی (درصد)) با روش‌های مختلف استخراج ویژگی به همراه روش پیشنهادی FrRC در حضور نویز Destroyer با پنج سطح سیگنال به نویز از -10dB تا 10dB

Noise Type		Destroyer				
SNR (dB)		-10dB	-5dB	0dB	5dB	10dB
Feature Extraction Methods	LPC[۲۱ و ۲۲]	23.48	30.19	36.27	42.14	46.96
	MFCC [۱۲ و ۲۰ ،۱۸]	23.90	28.72	35.43	44.44	**63.31**
	FMFCC[۶۱]	17.40	24.53	32.70	42.35	47.38
	RMFCC[۶۲]	16.98	24.53	34.80	**49.26**	56.18
	LSF[۲۴]	18.24	28.30	36.06	45.07	55.76
	PLP[۱۹ و ۲۰]	14.04	18.86	25.78	35.85	41.09
	Proposed (FrRC)	**27.04**	**31.03**	**36.27**	41.51	49.89

برای اینکه بتوان بطور واضح‌تر درک صحیحی از میزان سودمندی روش استخراج ویژگی پیشنهادی FrRC در محیط‌های نویزی با نویز شدید بدست آورد، مقایسه‌ای بین صحت‌های بازشناسی روش‌های استخراج ویژگی مرسوم با FrRC در حضور نویزهای مختلف در سطح SNR = -10dB انجام شده است که این مقایسه در شکل ۴-۱۳ نشان داده شده است.



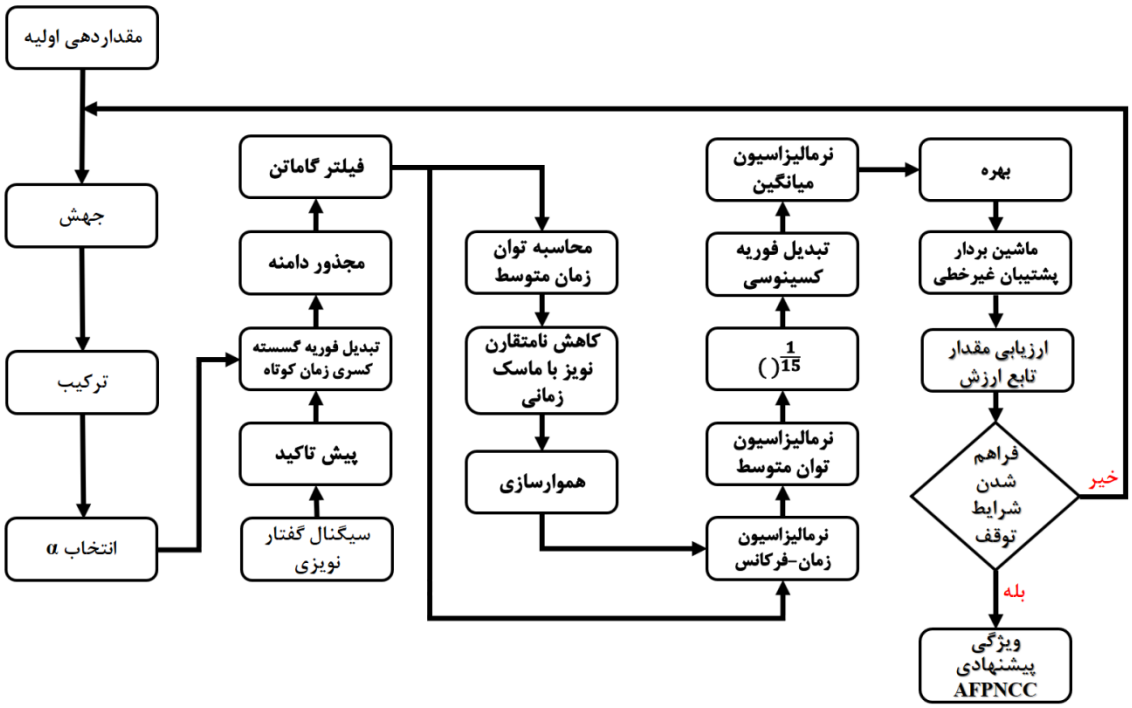
شکل ۴-۱۳: صحت‌های بازشناسی (درصد) روش‌های استخراج ویژگی مرسوم به همراه روش FrRC در حضور نویزهای مختلف در سطح SNR = -10dB

همانطورکه در شکل۴-۱۳ مشاهده می‌شود، روش استخراج ویژگی پیشنهادی مستقل از نوع نویز، بیشترین مقاومت در برابر نویز در محیط‌های نویزی شدید را از خود نشان داده و در نتیجه باعث بیشترین صحت بازشناسی در محیط نویزی می‌شود.

۴-۴- الگوریتم پیشنهادی سوم استخراج ویژگی: Adaptive Fractional Power Normolized Cepstral Coefficient (AFPNCC)

به منظور بهبود عملکرد روش الگوریتم پیشنهادی FrRC، الگوریتم دیگری با نام AFPNCC ارائه گردید. در این الگوریتم علاوه بر تکیه بر خاصیت چرخش تبدیل فوریه کسری و تابع ریشه، یک مرحله کاهش اثر نویز که توسط بدست آوردن کف نویز و تفریق آن از سیگنال اصلی است، لحاظ گردیده است. در این بخش به بررسی و تشریح اجزای الگوریتم استخراج ویژگی پیشنهادی که با نام ضرایب کپسترال توان نرمال سازی شده کسری وفقی (AFPNCC) معرفی شده است، می‌پردازیم. ساختار این الگوریتم پیشنهادی در شکل ۴-۱۴ نشان داده شده است. از آنجایی که در این روش استخراج ویژگی، از الگوریتم بهینه‌سازی به منظور یافتن ضریب تبدیل فوریه کسری با توجه به شدت و نوع نویز ورودی استفاده شده

است، این روش پیشنهادی یک روش تطبیقی براساس شدت و نوع نویز است. علاوه بر این، روش استخراج ویژگی پیشنهادی بهبود یافته‌ی روش استخراج ویژگی PNCC می‌باشد.



شکل ۴-۱۴: بلوک دیاگرام روش استخراج ویژگی پیشنهادی AFPNCC

نوآوری انجام شده در این ساختار نسبت به ساختار PNCC کلاسیک را می‌توان استفاده از تبدیل فوریه‌ی کسری با ضریب بهینه‌ی بدست آمده توسط بهینه ساز تکامل تفاضلی به جای تبدیل فوریه‌ی کلاسیک، ایجاد یک حلقه تطبیقی به منظور استخراج ویژگی‌های مقاوم نسبت به نوع و شدت نویز و نرمال‌سازی نهایی بعد از استخراج ویژگی توسط بلوک بهره به منظور تقویت ویژگی‌ها برای افزایش صحت سیستم ASR دانست.

در این الگوریتم، ابتدا پارامترهای الگوریتم بهینه سازی تکامل تفاضلی [۳۵-۳۶] از قبیل جمعیت اولیه، تعداد تکرار و ... مقداردهی اولیه می‌شوند. سپس عملیات جهش و تقاطع بر روی جمعیت اولیه اعمال می‌گردد. در مرحله‌ی بعد یک جواب اولیه که در واقع مقدار ضریب تبدیل فوریه‌ی کسری می‌باشد

بدست آمده و در بلوک STDFrFT[1] وارد می‌شود. بعد از این قسمت، یک فیلتر پیش تاکید بصورت $H(z)=1-0.97z^{-1}$ به سیگنال گفتار نویزی ورودی اعمال می‌شود. سپس از سیگنال فیلتر شده، تبدیل فوریه‌ی کسری گسسته زمان کوتاه گرفته شده و پس از آن تبدیل فوریه‌ی گسسته کسری زمان کوتاه سیگنال با استفاده از پنجره‌های همینگ با طول ۲۵ میلی ثانیه با ۱۰ میلی ثانیه همپوشانی بین فریم‌ها و ابعاد تبدیل فوریه کسری هم بعد با طول هر فریم محاسبه می‌گردد. خروجی مجذور دامنه STDFrFT محاسبه شده توسط فیلتر بانک گاماتن ۴۰ کانال که فرکانس مرکزی آنها به صورت خطی در پهنای باند مستطیلی معادل [۳۷] بین ۲۰۰ هرتز تا ۸۰۰۰ هرتز می‌باشند، وزن‌دار می‌شوند. علت استفاده از فیلتربانک گاماتن صحت نسبتاً بالای سیستم بازشناس گفتار در حضور نویز سفید می‌باشد [۲۰].

مرحله بعدی پردازش مربوط به پردازش زمان متوسط می‌باشد. در این مرحله، یک تحلیل بلند مدت زمانی (به عنوان مثال ۵ فریم) برای تخمین سطح پایین نویز و برای تفریق کردن آن از توان آنی سیگنال ورودی انجام شده است. سیستم شنیداری انسان روی قسمت ابتدایی پوش توان ورودی تمرکز می‌کند. لذا الگوریتم‌هایی برای بهبود این پوش ارائه شده است. برای تحقق این امر در AFPNCC یک قله متحرک برای هر کانال فرکانسی (l کانال) بدست آورده می‌شود. اگر توان آنی کمتر از پوش مدنظر باشد آن را حذف می‌کنیم. در بلوک ماسک گذاری زمانی ابتدا توان قله برخط برای هر کانال بدست آورده می‌شود ($\tilde{Q}_p[m.l]$). ماسک‌گذاری زمانی برای سگمنت‌های گفتار از طریق رابطه ۴-۲۵ بدست می‌آید [۲۳].

$$\tilde{Q}_{tm}[m.l]=\begin{cases}\tilde{Q}_0[m.l]. & \tilde{Q}_0[m.l]\geq\lambda_t\tilde{Q}_p[m-1.l]\\ \mu_t\tilde{Q}_p[m-1.l]. & \tilde{Q}_0[m.l]<\lambda_t\tilde{Q}_p[m-1.l]\end{cases} \qquad (۴-۲۵)$$

لازم به ذکر است که در رابطه ۴-۲۵، $\tilde{Q}_0[m.l]$، m، l و λ_t به ترتیب خروجی سیگنال عبوری از یکسوکننده نیم موج خطی، اندیس فریم، اندیس کانال و ضریب فراموشی هستند.

[1] Short Time Discrete Fractional Fourier Transform

اگر ضریب فراموشی کمتر یا مساوی ۰/۸۵ و μ_t کوچکتر یا مساوی ۰/۲ باشد، در اینصورت صحت بازشناسی برای حالت بدون نویز و نویزی ثابت است. در غیراینصورت صحت بازشناسی کاهش می‌یابد. حالتی که μ_t برابر با ۰/۲ و ضریب فراموشی برابر با ۰/۸۵ باشد، حالت ایده‌آل محسوب می‌شود و برای پیاده‌سازی PNCC استاندارد مورد استفاده قرار می‌گیرد. تاثیر ماسک گذاری زمانی در محیط‌های پرانعکاس ملموس است [۲۳].

خروجی پردازش زمان متوسط یک تابع تبدیل است که سیگنال اصلی در طبقه نرمال سازی زمان-فرکانس را مدوله می‌کند. سپس طبقه نرمال‌سازی میانگین توان، توان سیگنال را با تقسیم کردن به توان کل نرمال سازی می‌کند.

در طبقه بعد، یک تابع توان غیرخطی با ضریب ۱/۱۵ به ورودی اعمال می‌گردد. آزمایشات انجام شده در [۳۸] نشان می‌دهد که این ضریب برای فشار صدا، زمانی که صحت بازشناسی در حال بهینه شدن است، یک تناسب منطقی برای داده‌های فیزیولوژیکی فراهم می‌کند. آزمایشات نشان داده که استفاده از ترکیب غیرخطی تابع توان باعث بهبود صحت بازشناسی شده است [۴۰-۳۹]. سپس عملیات تبدیل فوریه گسسته کسینوسی و همچنین نرمال‌سازی میانگین بر روی سیگنال خروجی بلوک قبل انجام می‌پذیرد. تا این مرحله ویژگی‌های اولیه‌ی الگوریتم پیشنهادی AFPNCC ایجاد شده است. در مرحله بعد ویژگی‌های استخراج شده با عبور از طبقه بهره به منظور افزایش دامنه، در یک عدد ثابت ضرب می‌شوند. سپس این ویژگی‌ها به طبقه‌بند ماشین بردار پیشتیبان با کرنل چند جمله‌ای درجه ۴ با ضریب فولدینگ ۱۰ اعمال می‌شود. کرنل استفاده شده، توسط رابطه ۴-۲۶ قابل محاسبه می‌باشد.

$$\text{کرنل} = \left(\frac{1}{\text{تعداد ویژگی ها}} \times \left(\text{ماتریس ویژگی ها}\right)^t \times \text{بردار برچسب ها}\right)^4 \qquad (۴-۲۶)$$

در رابطه ۴-۲۶ منظور از بردار برچسب‌ها، برداری است که در آن، کلاس هر ویژگی قرار می‌گیرد.

خروجی طبقه‌بند، صحت بازشناسی و همچنین مقدار تابع ارزش یا همان Fitness که برابر با (صحت بازشناسی - ۱۰۰) است را نتیجه می‌دهد. در مرحله بعد اگر شرایط توقف الگوریتم که همان کمینه شدن تابع ارزش و ماکزیمم تکرار (۲۰۰ تکرار) است، فراهم باشد، الگوریتم پایان می‌یابد و ویژگی‌های AFPNCC که منجر به بیشینه شدن صحت بازشناسی می‌شود ایجاد می‌گردند. در غیر اینصورت الگوریتم ادامه پیدا کرده و تمامی فرآیند تشریح شده مجدداً تا فراهم شدن شرایط توقف ادامه می‌یابد.

۴-۴-۱- شبیه سازی و ارزیابی الگوریتم پیشنهادی

۴-۴-۱-۱- تنظیمات شبیه سازی

به منظور بررسی مقاومت و اثربخشی الگوریتم استخراج ویژگی پیشنهادی AFPNCC، عملکرد الگوریتم‌های استخراج ویژگی مختلف در محیط های نویزی با نویزها و سطوح SNR مختلف مورد ارزیابی قرار گرفته است. برای این هدف، هر مدل طبقه‌بندی با ۹۰ درصد از مجموع کل داده‌ها آموزش یافته و همچنین با ۱۰ درصد مابقی داده‌ها مورد آزمون قرار گرفته است.

نتایج آزمایشات و تحلیل‌ها با ۶ نوع نویز مختلف که شامل نویزهای Babble، HF-Channel، White، Pink، Factory و Destroyer (Noisex-92 [39]) می‌باشد، با ۶ سطح SNR متفاوت از ۵- دسی‌بل تا ۵۰ دسی‌بل ارائه شده است.

در این سیستم، طول فریم‌ها و همچنین میزان همپوشانی بین هر فریم به ترتیب ۲۰ میلی ثانیه و ۱۰ میلی ثانیه در نظر گرفته شده است. همچنین تعداد ویژگی های استخراج شده از هر فریم ۱۳ عدد می‌باشد.

۴-۴-۱-۲- دادگان

برای ارزیابی نتایج، از دادگان TI DIGITS [۴۲] استفاده شده است. این دادگان شامل ارقام انگلیسی با گوینده‌هایی در سنین مختلف از کودک تا بزرگسال و جنسیت مذکر و مونث می‌باشد. تعداد داده‌های موجود در این دادگان ۴۵۵۴ صوت در ۱۱ کلاس، ارقام ۰ تا ۹ و همچنین کلمه 'Oh' با فرکانس نمونه‌برداری ۸ کیلو هرتز می‌باشد.

۴-۴-۱-۳- شبیه سازی و نتایج

در واقع هدف از تکنیک استخراج ویژگی پیشنهادی بدست آوردن مجموعه‌ای از ویژگی‌های مقاوم در برابر نویزهای مختلف با سطوح SNR مختلف برای دست‌یابی به صحت بازشناسی مناسب در محیط نویزی است.

از آنجایی که آموزش و آزمون سیستم بازشناسی گفتار بر روی یک پایگاه داده انجام شده است، به همین خاطر سیستم ارزیابی شده یک سیستم بازشناسی گفتار وابسته به متن می‌باشد.

در این بخش از روش‌های استخراج ویژگی و همچنین روش استخراج ویژگی پیشنهادی AFPNCC با شش نویز مختلف HF-Channel، Babble، Factory1، White، Pink و Destroyer در شش سطح سیگنال به نویز از 5dB-تا 50dB به منظور ارزیابی سیستم بازشناسی در نرخ صحت بازشناسی استفاده شده است. لازم به ذکر است نویزهای مورد استفاده از پایگاه Noisex-92 تهیه شده است. با اجرای الگوریتم پیشنهادی تشریح شده در بخش ۴-۴ به منظور دستیابی به ماکزیمم صحت بازشناسی در محیط نویزی به مقدار ۰/۹۷ برای آلفا در تبدیل فوریه کسری دست یافته‌ایم.

نتایج بدست آمده از شبیه سازی سیستم بازشناسی به ازای روش‌های استخراج ویژگی مختلف به همراه روش استخراج ویژگی پیشنهادی در جداول ۴-۱۴ تا ۴-۱۶ ارائه و گزارش شده است.

باتوجه به نتایج بدست آمده در جدول ۴-۱۴ که در حضور نویزهای Babble و Factory می‌باشد، مشاهده می‌شود که در محیط نویزی با نویز Babble، در SNRهای بزرگتر از پنج دسی‌بل، صحت بازشناسی الگوریتم استخراج ویژگی پیشنهادی AFPNCC از تمامی روش‌های استخراج ویژگی دیگر بالاتر بوده و برای SNRهای کمتر از پنج دسی‌بل الگوریتم پیشنهادی بعد از الگوریتم PNCC در جایگاه دوم از لحاظ صحت بازشناسی قرار دارد. سیستم بازشناس گفتار مبتنی بر الگوریتم استخراج ویژگی AFPNCC در حضور نویز Factory در سطوح سیگنال به نویز بیشتر از صفر دسی‌بل، بالاترین صحت بازشناسی را نتیجه داده است و برای سطوح سیگنال به نویز کمتر از صفر دسی‌بل بعد از الگوریتمPNCC از لحاظ مقاومت در برابر نویز در جایگاه دوم قرار دارد.

جدول ۴-۱۴: نتایج شبیه سازی سیستم بازشناسی (صحت بازشناسی (درصد)) با روش های مختلف استخراج ویژگی به همراه روش پیشنهادی AFPNCC در حضور نویزهای Babble و Factory با شش سطح سیگنال به نویز از 5dB- تا 50dB

روش های استخراج ویژگی	نویز	Babble					
	SNR (dB)	-۵dB	۰dB	۵dB	۱۰dB	۱۵dB	تمیز(۵۰dB)
	LPC [۲۱ و ۲۲]	۴۱/۴۸	۵۰/۸۵	۵۶/۳۴	۶۲/۷۱	۶۸/۸۱	۷۹/۶۶
	MFCC [۱۸،۱۲و۲۰]	۶۶/۰۵	۷۶/۷۰	۸۴/۱۶	۸۹/۷۰	۹۳/۴۷	**۹۸/۴۶**
	PNCC [۶۱]	**۸۲/۵۸**	**۸۹/۵۰**	**۹۲/۹۷**	۹۴/۵۷	۹۵/۴۳	۹۷/۸۵
	LPCC [۲۱ و ۲۲]	۴۷/۱۶	۵۴/۰۸	۶۱/۱۱	۶۷/۶۷	۷۴/۲۶	۸۴/۶۷
	LSF [۲۴]	۴۹/۹۱	۵۶/۹۶	۶۳/۶۳	۷۰/۷۳	۷۷/۸۸	۸۹/۹۸
	Rasta-PLP [۱۹ و ۲۰]	۲۴/۹۶	۳۰/۴۵	۳۵/۵۷	۴۰/۲۵	۴۵/۲۵	۹۲/۶۲
	Proposed (AFPNCC)	۷۸/۱۳	۸۸/۰۳	۹۲/۶۸	**۹۵/۰۸**	**۹۶/۳۷**	۹۸/۲۲
	نویز	Factory					
	SNR (dB)	-۵dB	۰dB	۵dB	۱۰dB	۱۵dB	تمیز(۵۰dB)
	LPC [۲۱ و ۲۲]	۲۲/۴۲	۳۷/۳۹	۵۱,۶۰	۶۲/۹۷	۷۱/۰۸	۷۸/۶۷
	MFCC [۱۸،۱۲و۲۰]	۳۸/۲۳	۵۸/۷۳	۸۰/۱۷	۹۱/۳۰	۹۵/۷۸	**۹۸/۲۸**
	PNCC [۶۱]	**۷۶/۲۸**	**۸۷/۹۲**	۹۲/۸۴	۹۴/۷۷	۹۶/۰۴	۹۸/۱۷
	LPCC [۲۱ و ۲۲]	۴۵/۸۷	۶۳/۱۱	۷۴/۱۷	۷۹/۷۱	۸۲/۹۴	۸۶/۰۷
	LSF [۲۴]	۵۰/۳۹	۶۲/۲۵	۷۲/۲۲	۷۹/۹۹	۸۴/۹۳	۹۰/۵۳
	Rasta-PLP [۱۹ و ۲۰]	۲۷/۵۱	۳۲/۳۹	۳۶/۹۵	۴۱/۳۰	۴۷/۱۹	۹۲/۶۴
	Proposed (AFPNCC)	۷۴/۸۳	۸۷/۸۵	**۹۳/۳۹**	**۹۵/۴۹**	**۹۶/۵۳**	۹۸/۲۴

نتایج گزارش شده در جدول ۴-۱۵ گویای این مطلب است که روش استخراج ویژگی پیشنهادی توانسته است صحت بازشناسی سیستم بازشناس گفتار را در حضور نویزهای Destroyer و HF-Channel در تمامی سطوح سیگنال به نویز در بالاترین سطح صحت و بالاترین مقاومت در برابر نویز قرار دهد. این روش استخراج ویژگی حتی در محیط بدون نویز نیز بالاترین صحت بازشناسی را نتیجه داده است.

علاوه بر نویزهای مطرح شده در جداول ۴-۱۴ و ۴-۱۵، نتایج پیاده‌سازی الگوریتم بازشناسی گفتار با استفاده از روش‌های استخراج ویژگی مختلف و پیشنهادی در حضور نویزهای Pink و White در جدول ۴-۱۶ گزارش شده است.

جدول ۴-۱۵: نتایج شبیه سازی سیستم بازشناسی (صحت بازشناسی (درصد)) با روش های مختلف استخراج ویژگی به همراه روش پیشنهادی AFPNCC در حضور نویزهای Destroyer و HF-Channel با شش سطح سیگنال به نویز از 5dB- تا 50dB

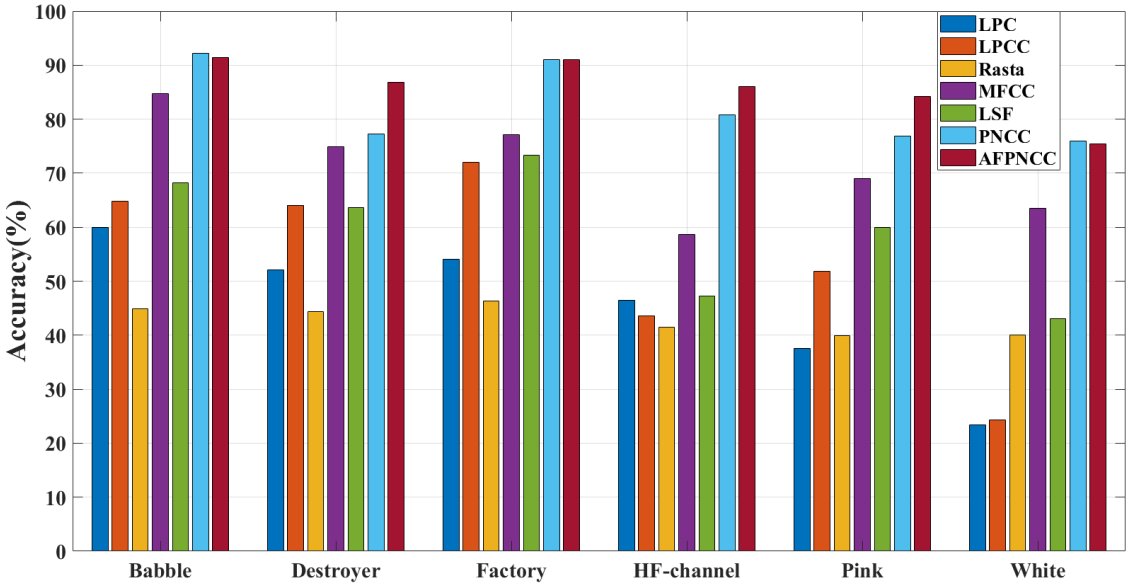
	نویز	Destroyer					
	SNR (dB)	-۵dB	۰dB	۵dB	۱۰dB	۱۵dB	تمیز(۵۰dB)
روش های استخراج ویژگی	LPC [۲۱ و ۲۲]	۲۲/۹۴	۳۴/۶۰	۴۸/۲۸	۶۰/۴۵	۶۸/۳۱	۷۸/۳۰
	MFCC [۲۰ و۱۸،۱۲]	۳۷/۵۹	۵۷/۵۷	۷۴/۸۸	۸۶/۸۹	۹۴/۰۷	۹۸/۳۱
	PNCC [۶۱]	۵۳/۴۰	۶۸/۶۷	۷۷/۵۸	۸۳/۱۸	۸۶/۷۳	۹۳/۸۷
	LPCC [۲۱ و ۲۲]	۳۲/۸۵	۴۷/۹۳	۶۳/۳۰	۷۳/۸۰	۷۹/۹۵	۸۶/۳۶
	LSF [۲۴]	۳۱/۲۹	۴۵/۹۶	۶۰/۸۴	۷۲/۵۹	۸۰/۱۰	۹۰/۶۴
	Rasta-PLP [۱۹ و ۲۰]	۲۵/۴۳	۳۰/۶۹	۳۴/۸۷	۳۹/۰۸	۴۴/۰۹	۹۲/۳۱
	Proposed (AFPNCC)	**۶۲/۳۸**	**۸۰/۳۵**	**۸۹/۳۷**	**۹۴/۴۰**	**۹۶/۴۶**	**۹۸/۳۵**
	نویز	HF Channel					
	SNR (dB)	-۵dB	۰dB	۵dB	۱۰dB	۱۵dB	تمیز(۵۰dB)
	LPC [۲۱ و ۲۲]	۱۶/۰۳	۲۸/۵۹	۴۱/۹۴	۵۲/۱۰	۵۹/۹۲	۷۹/۸۰
	MFCC [۲۰ و۱۸،۱۲]	۱۶/۶۸	۳۰/۲۶	۵۳/۱۴	۷۰/۶۸	۸۲/۴۷	**۹۸/۴۶**
	PNCC [۶۱]	۶۰/۱۰	۷۵/۱۶	۸۱/۵۷	۸۵/۶۴	۸۷/۹۹	۹۳/۸۳
	LPCC [۲۱ و ۲۲]	۱۲/۰۵	۲۰/۸۶	۳۴/۲۱	۴۸/۳۳	۶۱/۳۳	۸۴/۹۳
	LSF [۲۴]	۱۴/۰۷	۲۲/۴۶	۳۷/۲۴	۵۳/۲۹	۶۶/۵۵	۹۰/۰۷
	Rasta-PLP [۱۹ و ۲۰]	۱۷/۹۶	۲۴/۵۹	۳۱/۷۵	۳۸/۱۰	۴۴/۸۴	۹۱/۹۲
	Proposed (AFPNCC)	**۶۷/۴۸**	**۷۹/۵۷**	**۸۶/۵۸**	**۹۱/۱۰**	**۹۳/۷۲**	۹۸/۱۵

با توجه به جدول ۴-۱۶، مشاهده می شود که صحت بازشناسی روش استخراج ویژگی پیشنهادی به نسبت سایر روش‌های استخراج ویژگی در سطوح SNR مختلف مقدار بیشتری بدست آمده است.

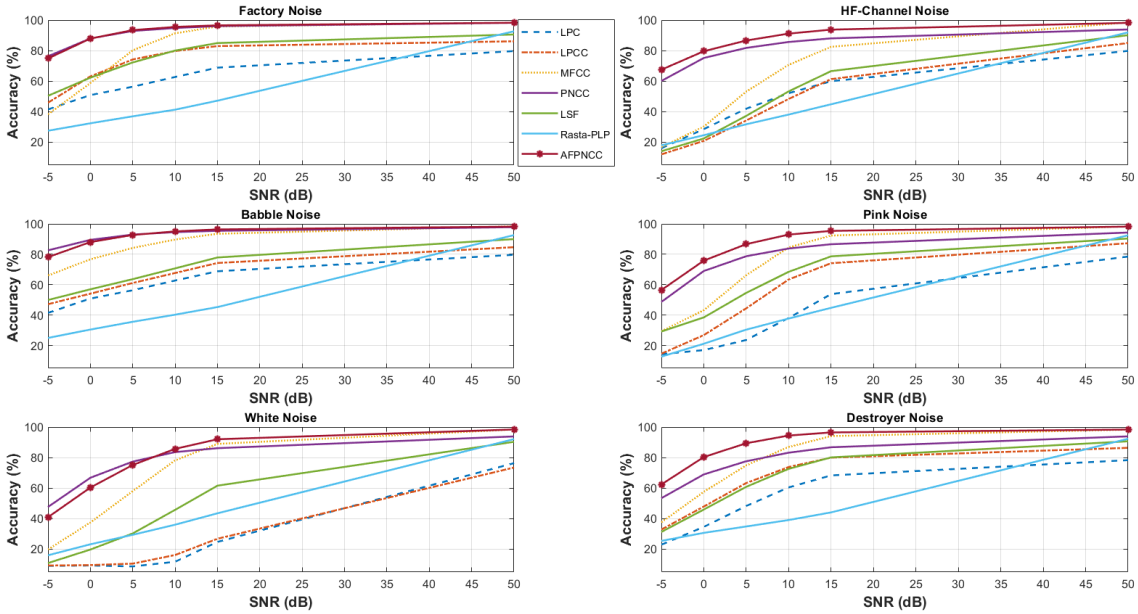
برای اینکه بتوان به درک صحیحی از میزان سودمندی روش استخراج ویژگی پیشنهادی AFPNCC در محیط‌های نویزی با سطوح سیگنال به نویز مختلف رسید، مقایسه‌ای بین صحت‌های بازشناسی روش‌های استخراج ویژگی مرسوم با AFPNCC در حضور نویزهای مختلف در سطوح SNR از ۵- دسی‌بل تا ۵۰ دسی‌بل انجام شده که این مقایسه در شکل ۵-۵ نشان داده شده است. همانطور که در شکل ۴-۱۵ مشاهده می‌شود، روش استخراج ویژگی پیشنهادی در حضور نویزهای مختلف دارای صحت بازشناسی بالایی بوده و می‌تواند بعنوان یک روش بسیار کارامد در محیط‌های نویزی و بدون نویز مورد استفاده قرار گیرد.

جدول ۴-۱۶: نتایج شبیه سازی سیستم بازشناسی (صحت بازشناسی (درصد)) با روش های مختلف استخراج ویژگی به همراه روش پیشنهادی AFPNCC در حضور نویزهای Pink و White با شش سطح سیگنال به نویز از 5dB- تا 50dB

	نویز	Pink					
	SNR (dB)	-۵dB	۰dB	۵dB	۱۰dB	۱۵dB	تمیز (۵۰dB)
روش های استخراج ویژگی	LPC [۲۱ و ۲۲]	۱۴/۲۵	۱۶/۹۵	۲۳/۶۰	۳۸/۰۷	۵۳/۸۰	۷۸/۵۰
	MFCC [۱۸،۱۲ و۲۰]	۲۹/۴۰	۴۳/۲۱	۶۶/۱۸	۸۴/۵۲	۹۲/۳۱	**۹۸/۳۳**
	PNCC [۶۱]	۴۸/۶۴	۶۸/۹۹	۷۸/۷۲	۸۳/۷۹	۸۶/۶۰	۹۴/۳۱
	LPCC [۲۱ و ۲۲]	۱۴/۵۸	۲۶/۸۱	۴۴/۴۴	۶۳/۴۱	۷۴/۱۱	۸۷/۲۶
	LSF [۲۴]	۲۹/۱۴	۳۸/۴۵	۵۴/۶۱	۶۸/۴۴	۷۸/۶۳	۹۰/۳۶
	Rasta-PLP [۱۹ و ۲۰]	۱۲/۶۴	۲۱/۲۴	۳۰/۳۹	۳۷/۷۲	۴۴/۷۵	۹۲/۵۱
	AFPNCC	**۵۶/۴۵**	**۷۵/۸۲**	**۸۶/۶۷**	**۹۲/۹۵**	**۹۵/۳۹**	۹۸/۲۲
	نویز	White					
	SNR (dB)	-۵dB	۰dB	۵dB	۱۰dB	۱۵dB	تمیز (۵۰dB)
	LPC [۲۱ و ۲۲]	۹/۰۹	۹/۳۳	۱۰/۷۱	۱۱/۷۷	۲۴/۷۷	۷۶/۳۳
	MFCC [۱۸،۱۲ و۲۰]	۱۹/۷۸	۳۷/۴۸	۵۸/۰۱	۷۸/۱۷	۸۸/۹۵	۹۸/۳۷
	PNCC [۶۱]	**۴۷/۶۵**	**۶۶/۷۷**	**۷۷/۳۶**	۸۳/۵۷	۸۶/۱۴	۹۳/۸۳
	LPCC [۲۱ و ۲۲]	۹/۳۱	۹/۵۹	۱۰/۵۴	۱۶/۱۸	۲۶/۸۳	۷۳/۴۷
	LSF [۲۴]	۱۰/۹۱	۱۹/۷۶	۳۰/۳۲	۴۵/۷۶	۶۱/۵۹	۹۰/۲۹
	Rasta-PLP [۱۹ و ۲۰]	۱۶/۰۱	۲۳/۲۱	۲۹/۴۲	۳۶/۰۳	۴۳/۵۰	۹۲/۰۷
	AFPNCC	۴۰/۹۳	۶۰/۳۶	۷۵/۱۴	**۸۵/۶۸**	**۹۱/۹۸**	**۹۸/۴۰**



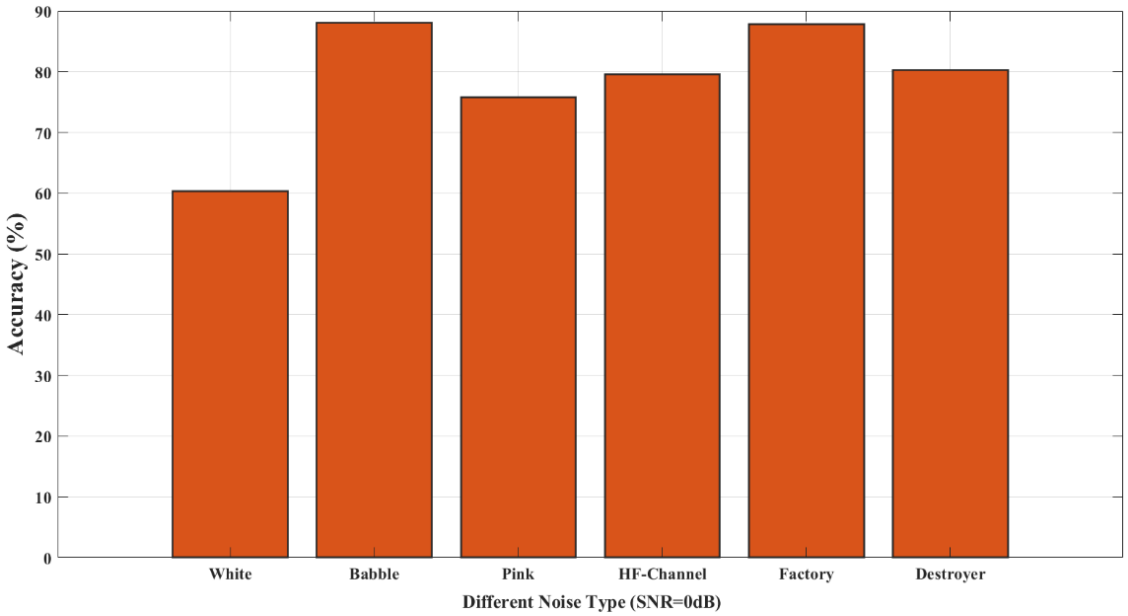
شکل ۴-۱۵: صحت بازشناسی روش‌های استخراج ویژگی مرسوم به همراه روش پیشنهادی AFPNCC در حضور نویزهای مختلف در سطوح SNR از ۵- دسی‌بل تا ۵۰ دسی‌بل



شکل ۴-۱۶: میانگین صحت بازشناسی سیستم بازشناسی گفتار برای روش‌های استخراج ویژگی مختلف و الگوریتم پیشنهادی در حضور نویزهای مختلف

در شکل ۴-۱۶ میانگین صحت بازشناسی به ازای سطوح SNR مختلف در حضور نویزهای Babble، Destroyer، Factory، HF-Channel، Pink و White برای روش‌های استخراج ویژگی مختلف به همراه

روش استخراج ویژگی پیشنهادی AFPNCC نمایش داده شده است. در این شکل مشاهده می‌شود که میانگین صحت بازشناسی روش استخراج ویژگی پیشنهادی در حضور نویز Babble و White بعد از PNCC در رتبه دوم قرار دارد و در سایر نویزها با اختلاف چشم گیری در رتبه اول صحت بازشناسی قرار می‌گیرد.



شکل ۴-۱۷: صحت بازشناسی روش استخراج ویژگی پیشنهادی AFPNCC در حضور نویزهای مختلف با سطح سیگنال به نویز صفر دسی‌بل

در شکل ۴-۱۷ که صحت بازشناسی سیستم بازشناسی گفتار به ازای نویزهای مختلف با سطح سیگنال به نویز صفر دسی‌بل ترسیم شده است، می‌توان مشاهده نمود که در حضور تمامی این نویزها صحت بازشناسی مقدار بالایی دارد. علاوه بر این می توان بیان کرد که نویز White بیشترین تاثیر در کاهش صحت بازشناسی و نویزهای Babble و Factory کمترین تاثیر را دارند.

علت برتری روش استخراج ویژگی پیشنهادی در استفاده‌ی هدفمند از تبدیل فوریه‌ی کسری با ضریب بهینه منطبق بر نوع و شدت نویز، استفاده از حلقه تطبیقی وابسته به نوع و شدت نویز و همچنین در نرمال‌سازی دامنه ویژگی‌های بهینه‌ی استخراج شده توسط طبقه‌ی بهره است. تجمیع این راهکارها در کنار هم توانسته است از روش پیشنهادی یک روش مقاوم در برابر نویز با صحت بازشناسی بالا رقم بزند.

در روش استخراج ویژگی پیشنهادی، نویز تنها در چند فریم محدود تاثیر خواهد داشت و همین امر در کنار سایر راه‌کارهای اعمال شده در این روش توانسته است منجر به این عملکرد خوب شود.

۴-۵- نتیجه‌گیری

در بخش اول این فصل از رساله، به بررسی میزان مقاومت فورمنت‌های فریم‌های واکدار سیگنال گفتار نسبت به نویز پرداخته شد. برای بررسی میزان مقاومت فورمنت‌ها در برابر نویزهای مختلف، نویزهای مختلف را به ۱۶ فریم واکدار از سیگنال گفتار اولیه اعمال کرده و به ازای اختلاط هر نویز با فریم‌های واکدار، فورمنت‌ها استخراج گردید. به ازای هر فریم واکدار در حالت بدون نویز ۳ فورمنت ابتدایی استخراج شد و در نهایت سه فورمنت ابتدایی تمامی فریم‌های واکدار بدون نویز میانگین‌گیری شده و با این کار در کل فریم‌های واکدار بدون نویز ۳ عدد بعنوان میانگین فورمنت اول، دوم و سوم تعیین شد. این کار برای تمامی فریم‌ها با اختلاط با منابع نویز دیگر نیز انجام شد. بنابراین برای ۹ حالت که عبارتند از: حالت بدون نویز و ۸ منبع نویز دیگر، از هرکدام سه عدد بعنوان میانگین فورمنت‌های اول، دوم و سوم تعیین شد. مسئله‌ی اختلاط فریم‌های بدون نویز با نویزها در سه SNR = 10dB, 5dB, 15dB انجام شده است تا بتوان میزان تاثیر SNR را نیز در میزان جابه‌جایی فورمنت‌ها مشاهده نمود. در نهایت با تعریف پارامتری با نام پارامتر MSM میزان مقاومت در برابر جابه‌جایی فورمنت‌ها ناشی از اعمال منابع نویز مختلف سنجیده شد و مشاهده گردید که فورمنت‌ها در برابر نویزهای Volvo و machinegun کمترین حساسیت را در دو SNR = 10dB ,5dB دارند اما در SNR = 15dB میزان MSM برای نویز Volvo به شدت افزایش یافت در حالیکه در این SNR، MSM نویز Machine Gun نسبت به تمامی حالات دیگر SNR از همه‌ی نویزهای دیگر کمتر می‌باشد. اما نکته جالب توجه این است که نویز سفید در تمامی SNRهای مطرح شده دارای بیشترین مقدار MSM است و این به این معناست که نویز سفید بیشترین تاثیر را بر روی جابه‌جایی فورمنت‌ها دارد.

در بخش دوم، یک روش جدید مقاوم در برابر نویز برای کاربرد بازشناسی گفتار در محیط‌های نویزی ارائه و پیشنهاد گردید. روش استخراج ویژگی پیشنهادی که بعنوان یک روش مقاوم در برابر نویز جمع-شونده با نام FrRC به منظور بازشناسی گفتار معرفی شده است، مبتنی بر تبدیل فوریه کسری زمان کوتاه و تابع ریشه می‌باشد. در این روش با استفاده از خاصیت چرخش تبدیل فوریه کسری، فیلتر موجود در این ساختار به نحوی عمل کرد که باعث بهبود عملکرد سیستم بازشناسی گفتار به نسبت سایر روش‌های استخراج ویژگی متداول در محیط های با سطوح سیگنال به نویز کم شد. به عنوان نمونه صحت بازشناسی گفتار برای الگوریتم FrRC نسبت به روش استخراج ویژگی MFCC در سطح سیگنال به نویز ۱۰- دسیبل، در محیط‌های نویزی با نویزهای M109، Babble، Factory1، Factory2 و Destroyer به ترتیب ۲۵٪، ۲۱٪، ۵۹٪، ۱۳۳٪ و ۱۳٪ افزایش یافته است. همچنین رابطه ریاضی بین ویژگی‌های FrRC گفتار تمیز، نویز و گفتار نویزی بدست آورده شد و این رابطه بصورت تئوری در سطوح مختلف سیگنال به نویز با رابطه‌ی بین ویژگی‌های MFCC گفتار تمیز، نویز و گفتار نویزی مقایسه شد. دلیل برتری روش FrRC نسبت به روش MFCC در محیط‌های نویزی نیز با تحلیل‌های ریاضی نشان داده شد.

در بخش سوم، روشی مقاوم برای استخراج ویژگی مبتنی بر تبدیل فوریه کسری برای کاربرد بازشناسی گفتار ارائه شد. ساختار پیشنهادی که با نام AFPNCC معرفی شد، در واقع حالت بهبود داده شده‌ی الگوریتم PNCC می‌باشد. در روش پیشنهادی از الگوریتم تکامل تفاضلی به منظور یافتن ضریب آلفا تبدیل فوریه کسری استفاده شده است. ضریب چرخش بدست آمده از این الگوریتم بهینه‌ساز برای تبدیل فوریه کسری در کنار سایر بلوک‌های الگوریتم استخراج ویژگی پیشنهادی AFPNCC باعث بهبود عملکرد سیستم بازشناس گفتار در محیط نویزی شده است. ساختار پیشنهادی نه تنها در سطوح سیگنال به نویز کم باعث بهبود عملکرد بازشناسی شده بلکه در محیط‌های با شدت نویز کم هم عملکرد سیستم بازشناسی گفتار را بهبود داده است. به عنوان نمونه در محیط‌های نویزی آلوده شده توسط نویزهای Destroyer، HF Channel، Pink با سطح سیگنال به نویز ۵- دسیبل، صحت بازشناسی

الگوریتم پیشنهادی نسبت به الگوریتم PNCC، به ترتیب ۱۶/۸٪، ۱۲/۳٪ و ۱۶٪ افزایش یافته است. همچنین به عنوان نمونه‌ای از محیط‌هایی با شدت نویز کم (SNR=15dB) به ترتیب برای محیط‌های نویزی Babble، Destroyer، HF-Channel، Pink و White، صحت بازشناسی سیستم بازشناس گفتار با استفاده از روش پیشنهادی AFPNCC به ترتیب ۱٪، ۱۱/۲٪، ۶/۵٪، ۱۰/۲٪ و ۶/۷٪ نسبت به روش PNCC افزایش یافته است.

فصل پنجم: نتیجه‌گیری و پیشنهادهایی برای ادامه کار

۵-۱- مقدمه

بازشناسی خودکار گفتار می‌تواند در وضعیتی که شرایط آموزش و آزمون سیستم یکسان باشد، عملکرد بالایی را نتیجه دهد. به همین خاطر هر چه شرایط این دو مرحله به هم نزدیک‌تر باشند، صحت بازشناسی سیستم گفتار بیشتر خواهد شد. با این حال در محیط‌های واقعی به دلیل وجود نویز، عملکرد کلی سیستم بازشناسی گفتار به طور قابل توجهی کاهش خواهد یافت. تحقیقات اخیر نشان داد که اکثر روش‌های استخراج ویژگی از قبیل روش Mel-Frequency Cepstral Coefficients (MFCC)، LPC، LPCC و ... تنها در محیط‌های بدون نویز می‌توانند صحت بازشناسی بالایی را نتیجه دهند ولی در محیط‌های پر نویز عملکرد سیستم بازشناسی گفتار به شدت کاهش می‌یابد. لذا در انجام این رساله با هدف کاهش این مشکلات، الگوریتم‌هایی معرفی گردید که قادر هستند عملکرد سیستم بازشناسی گفتار را در محیط‌های نویزی به حد چشم‌گیری بهبود ببخشند. بعبارت دیگر الگوریتم‌های پیشنهادی که برای استخراج ویژگی بازشناسی گفتار پیشنهاد شده‌اند در برابر نویز مقاوم هستند. الگوریتم‌های پیشنهادی مبتنی بر تبدیل فوریه کسری می‌باشند و در آنها از الگوریتم های فرا ابتکاری برای دست‌یابی به ضرایب بهینه و مقاوم بهره برده شده است. در ادامه این فصل ابتدا عملکرد الگوریتم‌های پیشنهادی به طور خلاصه جمع‌بندی شده و نهایتاً پیشنهادهایی برای ادامه کار مطرح خواهد شد.

۵-۲- جمع‌بندی الگوریتم‌های پیشنهادی در رساله

در این رساله الگوریتم‌هایی به منظور استخراج ویژگی مقاوم در محیط‌های نویزی مبتنی بر تبدیل فوریه کسری پیشنهاد شده است. استفاده از روش های بهینه سازی فرا ابتکاری نظیر الگوریتم های ازدحام ذرات و تکامل تفاضلی در بدنه الگوریتم های استخراج ویژگی پیشنهادی باعث ایجاد ویژگی هایی بهینه با حداکثر مقاومت در برابر نویز که صحت بازشناسی بالایی را در سیستم بازشناسی گفتار نتیجه داده است، شده است.

در بخش اول فصل چهارم، شبیه سازی‌های اولیه‌ای به منظور بررسی میزان مقاومت فورمنت‌های یک سیگنال گفتار در برابر تغییر مکان نسبت به نویزهای مختلف مورد بررسی و تجزیه و تحلیل قرار گرفت. بدین معنا که در این شبیه سازی‌ها، با اعمال نویزهای مختلف به سیگنال اصلی، میزان جابه‌جایی مکان فورمنت‌ها بررسی شده است. در این پیاده‌سازی پارامتری با نام MSM تعریف شده است که بیانگر میزان انحراف و جابه‌جایی فورمنت‌ها ناشی از اعمال نویزهای مختلف می‌باشد. تمامی اعمال انجام شده در این تحقیق، در سه SNR مختلف با مقادیر 10dB، 5dB و 15dB صورت پذیرفته است تا بتوان تاثیر SNR را در پارامتر MSM و میزان جابه‌جایی فورمنت‌ها مشاهده نمود. نتایجی که در این پیاده‌سازی بدست آمده، گویای این مطلب می‌باشد که فرکانس فورمنت‌ها در هر سه مقدار SNR ، نسبت به نویز machine gun بسیار مقاوم بوده اما از طرف دیگر بیشترین تاثیر و جابه‌جایی روی فرکانس فورمنت‌ها را نویز white سبب می‌شود.

در بخش دوم فصل چهارم، یک روش استخراج ویژگی جدید با نام Fractional Root Coefficient (FrRC) برای کاربرد بازشناسی گفتار در محیط‌های نویزی معرفی شد. این روش استخراج ویژگی پیشنهادی مبتنی بر تبدیل فوریه کسری زمان کوتاه و تابع ریشه می‌باشد. از آنجایی که انتخاب ضریب تبدیل کسری و ریشه برای تحلیل‌های مناسب سیگنال‌های چند جزئی از قبیل گفتار همچنان مورد بحث می‌باشد به همین خاطر در روش FrRC با استفاده از الگوریتم‌های فرا ابتکاری پارامترهای بهینه آلفا و گاما به ترتیب برای تبدیل فوریه کسری و تابع ریشه بدست آورده شده است. از دادگان TIMIT و Noisex-92 به منظور ارزیابی میزان مقاومت و صحت بازشناسی سیستم بازشناسی گفتار استفاده شده است. نتایج شبیه سازی حاکی از مقاومت بیشتر و صحت بازشناسی بالاتر روش استخراج ویژگی FrRC در قیاس با سایر روش‌های استخراج ویژگی در محیط‌های نویزی شدید می‌باشد. در این بخش علاوه بر تحلیل نتایج شبیه سازی، از طریق روابط ریاضی رابطه‌ای بین ویژگی‌های بدون نویز و نویزی الگوریتم استخراج ویژگی پیشنهادی FrRC اثبات گردید.

در بخش سوم فصل چهارم، یک الگوریتم استخراج ویژگی جدید که الگوریتم استخراج ضرایب کپسترال توان نرمال سازی شده کسری وفقی (AFPNCC) نامیده می‌شود، بعنوان یک روش مقاوم در برابر نویز برای کاربرد بازشناسی گفتار ارائه شده است. این روش استخراج ویژگی پیشنهادی مبتنی بر تبدیل فوریه گسسته کسری زمان کوتاه می‌باشد. در این روش پیشنهادی با استفاده از الگوریتم فرا ابتکاری تکامل تفاضلی، پارامتر بهینه α برای تبدیل فوریه کسری با توجه به کلاس نویز موجود در محیط بصورت وفقی بدست می‌آید. همچنین از دادگان TI Digit و Noisex-92 به منظور ارزیابی میزان مقاومت و صحت بازشناسی سیستم بازشناسی گفتار خودکار استفاده شده است. نتایج شبیه سازی بیانگر مقاومت بیشتر و صحت بازشناسی بالاتر روش استخراج ویژگی پیشنهادی در قیاس با سایر روش‌های استخراج ویژگی در محیط‌های نویزی و بدون نویز می‌باشد. در سیستم ASR پیشنهادی از طبقه‌بند ماشین بردار پشتیبان با کرنل غیرخطی استفاده شده است.

۵-۳- پیشنهادهایی برای بهبود عملکرد سیستم بازشناسی گفتار در آینده

برای ادامه کار رساله در زمینه بازشناسی مقاوم گفتار برای بهبود عمکلرد سیستم بازشناسی می‌توان موارد زیر را در نظر گرفت:

۱- در نظر گرفتن ویژگی‌های فاز در کنار اندازه ویژگی‌ها: در اکثر روش‌های استخراج ویژگی مبتنی بر کپسترال، از ویژگی‌های فاز صرفنظر می‌شود و تنها به استفاده از ویژگی‌های اندازه بسنده می‌شود. از آنجایی که در نظر گرفتن تنها ویژگی‌های اندازه می‌تواند بخشی از اطلاعات را حذف کند و ویژگی‌های فاز نیز حاوی اطلاعات مفید است، در صورت استفاده از آنها در کنار ویژگی‌های اندازه، عملکرد سیستم بازشناسی می‌تواند هم در محیط نویزی و هم در محیط بدون نویز بهبود یابد.

۲- استفاده از تبدیلات زمان-فرکانس دیگر: در این رساله ما از تبدیل زمان-فرکانس، تبدیل فوریه کسری استفاده کردیم و توانستیم به ویژگی‌هایی دست یابیم که نسبت به روش های استخراج ویژگی دیگر مقاومت بیشتری را در برابر نویز از خود نشان دهد و باعث بهبود صحت بازشناسی گفتار گرددد. از ویژگی‌های زمان-فرکانس که می‌تواند برای کاربرد پردازش گفتار مفید باشد می‌توان به تبدیل HHT[۱] اشاره نمود.

[۱] Hilbert Huang Transform

مراجع

[1] Y. F. Al-Irhaim and E. G. Saeed, "Arabic word recognition using wavelet neural network," AL-Rafidain Journal of Computer Sciences and Mathematics, vol. 7, no. 3, pp. 81-90, 2010.

[2] M. Sadeghi and H. Marvi, "Optimal MFCC features extraction by differential evolution algorithm for speaker recognition," In 3rd Iranian Conference on Intelligent Systems and Signal Processing (ICSPIS), pp. 169-173, Dec. 2017.

[3] K. Mannepalli, P.N. Sastry, and M. Suman, "MFCC-GMM based accent recognition system for Telugu speech signals," International Journal of Speech Technology, vol. 19, no. 1, pp. 87-93, Nov. 2015.

[4] J. Li, L. Deng, Y. Gong, and R. Haeb-Umbach, "An overview of noise-robust automatic speech recognition," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 4, pp. 745-777, Apr. 2014.

[5] P. Kumar and S. Kansal, "Noise removal in speech signal using fractional Fourier transform," In IEEE International Conference on Information, Communication, Instrumentation and Control (ICICIC), pp. 1-4, Aug. 2017.

[6] R. Hibare and A. Vibhute, "Feature extraction techniques in speech processing: a survey," International Journal of Computer Applications, vol. 107, no. 5, pp. 1-8, Dec. 2014.

[7] V. Radha and C. Vimala, "A review on speech recognition challenges and approaches," World of Computer Science and Information Technology Journal, vol. 2, no. 1, pp. 1-7, 2012.

[8] A. Madan and D. Gupta, "Speech feature extraction and classification: A comparative review," International Journal of computer applications, vol. 90, no. 9, pp. 20-25, Mar. 2014.

[9] N. Desai, K. Dhameliya, and V. Desai, "Feature extraction and classification techniques for speech recognition: A review," International Journal of Emerging Technology and Advanced Engineering, vol. 3, no. 12, pp. 367-371, 2013.

[10] O. P. Prabhakar and N. K. Sahu, "A survey on: Voice command recognition technique," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 5, pp. 576-585, 2013.

[11] Signal Processing Information Base(http://spib.linse.ufsc.br/noise.html).

[12] M. Sadeghi, H. Marvi, and M. Ali, "The effect of different acoustic noise on speech signal formant frequency location," International Journal of Speech Technology, vol. 21, no. 3, pp. 741-752, Aug. 2018.

[13] M. H. Davis and O. Scharenborg, "Speech perception by humans and machines," In Speech perception and spoken word recognition, pp. 191-214, Oct. 2016.

[14] Ch. Lee et al., "Optimizing feature extraction for speech recognition," IEEE Transactions on Speech and audio processing, vol. 11, no. 1, pp. 80-87, Jan. 2003.

ویسی هادی، روش‌های مبتنی بر مدل برای سیستم‌های بازشناسی‌گفتارمقاوم به نویز، پایان نامه کارشناسی ارشد، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف، تهران-ایران،۱۳۸۴۰

[16] M. J. F. Gales, "Model-based techniques for noise robust speech recognition" (Doctoral dissertation, University of Cambridge), 1995.

[17] B. Nasersharif and A. Akbari, "A framework for robust MFCC feature extraction using SNR-dependent compression of enhanced Mel filter bank energies," In Ninth International Conference on Spoken Language Processing, Sep. 2006.
[18] A. Narayanan and D. Wang, "Investigation of speech separation as a front-end for noise robust speech recognition," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 4, pp. 826-835, Apr. 2014.
[19] H. Y. Cho and Y. H. Oh, "On the use of channel-attentive MFCC for robust recognition of partially corrupted speech," IEEE signal processing letters, vol. 11, no. 6, pp. 581-584, Jun. 2004.
[20] J. Chen, Y. Wang, and D. Wang, "A feature study for classification-based speech separation at low signal-to-noise ratios," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 12, pp. 1993-2002, Dec. 2014.
[21] J. Chen, K. K. Paliwal, and S. Nakamura, "Sub-band based additive noise removal for robust speech recognition," In Seventh European Conference on Speech Communication and Technology, Sep. 2001.
[22] D. Gupta, P. Bansal, and K. Choudhary, "The state of the art of feature extraction techniques in speech recognition," In Speech and language processing for human-machine communications, vol. 664, pp. 195-207, 2018.
[23] X. Pan, H. Zhao, and Y. Zhou, "The application of fractional Mel cepstral coefficient in deceptive speech detection," PeerJ, 3: e1194, Aug. 2015.
[24] N. R. Shabtai, B. Rafaely, and Y. Zigel, "The effect of reverberation on the performance of cepstral mean subtraction in speaker verification," Applied acoustics, vol. 72, no. 2, pp. 124-126, Feb. 2011.
[25] H. Veisi and H. Sameti, "The integration of principal component analysis and cepstral mean subtraction in parallel model combination for robust speech recognition," Digital Signal Processing, vol. 21, no. 1, pp. 36-53, Jan. 2011.
[۲۶] ویسی هادی, ثامتی حسین, ابوطالبی حمیدرضا, استفاده ازتحلیل اجزای اصلی در بازشناسی‌گفتارپیوسته زبان فارسی برای مقاوم سازی ویژگی‌ها و کاهش تعداد آنها, دهمین کنفرانس سالانه انجمن کامپیوتر ایران, تهران-ایران, زمستان ۱۳۸۳،
[27] C. A. Ynoguti and F. Violaro, "On the use of principal component analysis over mel cepstral coefficients," Telecomunicaçoes, Revista do Instituto Nacional de Telecomunicaçoes, vol. 5, no. 2, pp. 13-17, 2002.
[28] P. Zhan and A. Waibel, "Vocal tract length normalization for large vocabulary continuous speech recognition," Carnegie-Mellon univ Pittsburgh PA school of computer science, 1997.
[29] L. Welling and et al., "A study on speaker normalization using vocal tract normalization and speaker adaptive training," In IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2, pp. 797-800, May. 1998.
[۳۰] فرشاد الماس گنج، یاسر شکفته، "بازشناسی مقاوم گفتار با استفاده از ویژگی الگوهای زمانی به دست آمده از ساختار شبکه عصبی بهینه شدهMTMLP"، هوش محاسباتی در مهندسی برق، ۵، ۳۵-۲۳، دانشگاه اصفهان، ۱۳۹۳.
[31] B. Y. Chenl, Q. Zhu, and N. Morgan, "Tonotopic multi-layered perceptron: A neural network for learning long-term temporal features for speech recognition," In IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, pp. I-945, Mar. 2005.
[32] K. Rakesh, S. Dutta, S., and K. Shama, "Gender Recognition using speech processing techniques in LABVIEW," International Journal of Advances in Engineering & Technology, vol. 1, no. 2, pp. 51-63, May. 2011.

[33] D. Gargouri, M. A. Kammoun, and A. B. Hamida, "A comparative study of formant frequencies estimation techniques," In Proceedings of the 5th WSEAS international conference on Signal processing, Istanbul, Turkey, pp. 15-19, May. 2006.
[34] D. Palaz, M. M. Doss, and R. Collobert, "Convolutional neural networks-based continuous speech recognition using raw speech signal," In IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4295-4299, Apr. 2015.
[35] Y. H. Goh, P. Raveendran, and S. S. Jamuar, "Robust speech recognition using harmonic features," IET Signal Processing, vol. 8, no. 2, pp. 167-175, Apr. 2014.
[36] D. Yu et al., "Robust speech recognition using a cepstral minimum-mean-square-error-motivated noise suppressor," IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 5, pp. 1061-1070, Jul. 2008.
[37] B. J. Shannon and K. K. Paliwal, "Feature extraction from higher-lag autocorrelation coefficients for robust speech recognition," Speech Communication, vol. 48, no. 11, pp. 1458-1485, Nov. 2006.
[38] G. Farahani, "Robust feature extraction using autocorrelation domain for noisy speech recognition," Signal & Image Processing: An International Journal, vol. 8, no. 1, pp. 23-44, Feb. 2017.
[39] P. Kumar and S. Kansal, "Noise removal in speech signal using fractional Fourier transform," In IEEE International Conference on Information, Communication, Instrumentation and Control, pp. 1-4, Aug. 2017.
[40] K. Markov and T. Matsui, "Robust speech recognition using generalized distillation framework," In International Speech Communication Association (Interspeech), pp. 2364-2368, Sep. 2016.
[41] Y. Qian et al., "Very deep convolutional neural networks for noise robust speech recognition," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 12, pp. 2263-2276, Dec. 2016.
[42] C. Kim and R. M. Stern, "Power-normalized cepstral coefficients (PNCC) for robust speech recognition," IEEE/ACM Transactions on audio, speech, and language processing, vol. 24, no. 7, pp. 1315-1329, July 2016.
[43] Y. Sun et al., "Indoor sound source localization with probabilistic neural network," IEEE Transactions on Industrial Electronics, vol. 65, no. 8, pp. 6403-6413, Aug. 2018.
[44] Y. Wang et al., "Quantification and segmentation of brain tissues from MR images: a probabilistic neural network approach," IEEE transactions on image processing, vol. 7, no. 8, pp. 1165-1181, Aug. 1998.
[45] E. Parzen, "On estimation of a probability density function and mode. The annals of mathematical statistics," vol. 33, no. 3, pp. 1065-1076, Sep. 1962.
[46] G. Dede and M. H. Sazlı, "Speech recognition with artificial neural networks. Digital Signal Processing," vol. 20, no. 3, pp. 763-768, May 2010.
[47] X. G. Li et al., "The Application of Probabilistic Neural Network in Speech Recognition Based on Partition Clustering," In Applied Mechanics and Materials, vol. 263, pp. 2173-2178, Dec. 2012.
[48] M. J. F. Gales, "Model-based techniques for noise robust speech recognition," (Doctoral dissertation, University of Cambridge), Sep. 1995.
[49] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, no. 3, pp. 273-297, Sep. 1995.
[50] D. Wang, D. Tan, and L. Liu, "Particle swarm optimization algorithm: an overview. Soft Computing," vol. 22, no. 2, pp. 387-408, Jan. 2017.

[51] A. A. Abou El Ela, M. A. Abido, and S. R. Spea, "Optimal power flow using differential evolution algorithm," Electric Power Systems Research, vol. 80, no. 7, pp. 878-885, July 2010.
[52] E. U. Condon, "Immersion of the Fourier transform in a continuous group of functional transformations," Proceedings of the National academy of Sciences of the United States of America, vol. 23, no. 3, pp. 158-164, Mar. 1937.
[53] V. Namias, "The fractional order Fourier transform and its application to quantum mechanics," IMA Journal of Applied Mathematics, vol. 25, no. 3, pp. 241-265, Mar. 1980.
[54] N. Wiener, "Hermitian polynomials and Fourier analysis," Journal of Mathematics and Physics, vol. 8, no. 1, pp. 70-73, Apr. 1929.
[55] L. B. Almeida, "The fractional Fourier transform and time-frequency representations," IEEE Transactions on signal processing, vol. 42, no. 11, pp. 3084-3091, Nov. 1994.
[56] E. Sejdić, I. Djurović, and L. Stanković, "Fractional Fourier transform as a signal processing tool: An overview of recent developments," Signal Processing, vol. 91, no. 6, pp. 1351-1369, Jun. 2011.
[57] M. Müller, "Information retrieval for music and motion," Heidelberg: Springer, vol. 2, 2007.
[58] K. Weber, S. Bengio, and H. Bourlard, "HMM2-extraction of formant structures and their use for robust ASR," In Seventh European Conference on Speech Communication and Technology, Sep. 2001.
[59] A. Acero, "Formant analysis and synthesis using hidden Markov models," In Sixth European Conference on Speech Communication and Technology, Sep. 1999.
[60] D. Gargouri, M. A. Kammoun, and A. B. Hamida, "A comparative study of formant frequencies estimation techniques," In Proceedings of the 5th WSEAS international conference on Signal processing, Istanbul, Turkey, pp. 15-19, May 2006.
[61] M. Tamazin, A. Gouda, and M. Khedr, "Enhanced automatic speech recognition system based on enhancing power-normalized cepstral coefficients," Applied Sciences, vol. 9, no. 10, May 2019.
[62] D. Darabian, H. Marvi, and M. Sharif Noughabi, "Improving the performance of MFCC for Persian robust speech recognition," Journal of AI and Data Mining, vol. 3, no. 2, pp. 149-156, 2015.

Abstract

Speech is the main source of communication between human beings in order to show their ideas, feelings and thoughts to each other. Speech Recognition Technology allows the computer to receive, interpret, and respond appropriately to human speech commands. Due to the noise existence in real environments, we face the challenge of noncompliance and unequal conditions in both test and train modes for real-world applications. Noise robustness is a broad subject in ASR systems research that has been around for decades and has been researched by many. researchers. In this thesis, first in order to investigate the robustness of formants of voiced frames of speech in noisy environments with different sources, the amount of displacement of the formants of noisy voiced frames compared to clean voiced frames has been measured. it was shown that white noise at all signal-to-noise levels has the greatest impact on the voiced formats of the speech signal. After that, an algorithm was proposed to extract the robust feature for speech recognition. This proposed structure is based on fractional Fourier transform and root function which was named FrRC. For theoretical justification of the proposed method, a mathematical relationship was obtained between the FrRC features of clean speech, noise and noisy speech, and this relationship was compared with the mathematical relationship of the MFCC feature extraction method in different cases. The results of implementation of the speech recognition system based on the FrRC feature extraction method indicate an increase in recognition accuracy compared to other feature extraction methods. An increase of 24.6% and 25.3% of the recognition accuracy compared to LPC and MFCC methods in noisy environment with Babble noise and with signal to noise level of -10dB, respectively, is proof of this claim.

In order to increase the accuracy of speech recognition in noisy environments, another algorithm based on Fourier transform and the Power Normalized Cepstral Coefficient method, called Adaptive Fractional Power Normalized Cepstral Coefficient (AFPNCC), was introduced, analyzed and then implemented. In the proposed AFPNCC algorithm, based on the type and intensity of noise, the alpha coefficient of fractional Fourier transform in the algorithm is extracted by the differential evolution optimizer located in the body of the proposed algorithm structure. The results of the implementation of this

algorithm show the improvement of speech recognition accuracy in both noisy and clean environments. Numerical results obtained from the simulation of speech recognition system based on AFPNCC feature extraction algorithm also show a 16 and 92% increase in recognition accuracy compared to PNCC and MFCC algorithms in noisy environment with Pink noise and signal to noise level of 5dB, respectively.

Keywords: Robust speech recognition, Robust feature extraction, Cepstral features, Fractional Fourier Transform, Differential Evolution optimizer.

Shahrood University of Technology

Faculty of Electrical and Robotic Engineering

PhD Thesis

Robust Features Extraction for Speech Recognition in Noisy Environments Based on Fractional Fourier Transform

By:

Mohsen Sadeghi

Supervisor:

Dr. Hossein Marvi

Advisors:

Dr. Alireza Ahmadi Fard

Dr. Maaruf Ali

August 2020