

ارائه یک روش نوین و کارآمد استخراج ویژگی برای بازشناسی گفتار مقاوم مبتنی بر تبدیل فوریه کسری و بهینه‌ساز تکامل تفاضلی

محسن صادقی^{۱*}، حسین مروی^۲ و علیرضا احمدی فرد^۳

اطلاعات مقاله	چکیده
دریافت مقاله: ۱۳۹۸/۰۹/۳۰ پذیرش مقاله: ۱۳۹۹/۰۲/۰۶	یکی از چالش‌های اساسی در تشخیص گفتار، استخراج ویژگی مقاوم نسبت به نویز می‌باشد. در این مقاله یک الگوریتم استخراج ویژگی جدید که الگوریتم استخراج ضرائب کپسترال توان نرمالیزه شده کسری وفقی 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. از آنجایی که انتخاب ضریب تبدیل کسری برای تحلیل‌های مناسب سیگنال‌های چند جزئی از قبیل گفتار همچنان مورد بحث است، در این روش پیشنهادی با استفاده از الگوریتم فراابتکاری 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 دقت بازشناسی سیستم بازشناس گفتار خودکار استفاده شده است. نتایج شبیه‌سازی بیانگر مقاومت بیشتر و دقت 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. لازم به ذکر است که تمامی شبیه‌سازی‌های انجام شده توسط نرم‌افزار MATLAB صورت گرفته است.
واژگان کلیدی: تبدیل فوریه کسری، الگوریتم تکامل تفاضلی، استخراج ویژگی مقاوم، بازشناسی مقاوم گفتار، طبقه بند، ASR، AFPNCC	

۱-مقدمه

در دهه‌های اخیر پس از معرفی مدل‌های پنهان مارکوف^۲ (به‌عنوان مثال [۱]) و مدل‌های زبان آماری^۳ (به‌عنوان مثال [۲])، عملکرد سیستم‌های تشخیص گفتار در محیط‌های آکوستیک به طرز چشمگیری بهبود یافته است. با این وجود، عملکرد اکثر سیستم‌های تشخیص گفتار در حضور نویزهای جمع‌شونده و اعوجاج کانال به شدت تحت تأثیر قرار می‌گیرد. یکی از چالش برانگیزترین مسایل که به‌طور قابل-توجهی دقت بازشناسی را کاهش می‌دهد، عدم انطباق شرایط محیط آموزش و آزمون می‌باشد. عوامل ایجاد این

اختلاف می‌تواند نویز جمع‌شونده، اعوجاج کانال گوینده‌های متفاوت، بازتاب و غیره باشد.

از آنجایی که نویزها بر روی انتقال و بازشناسی گفتار و صوت اثر می‌گذارند، باید بعنوان یک فاکتور مهم و تأثیرگذار مورد بررسی قرار گیرند. در طی سالیان اخیر، ده‌ها الگوریتم به-منظور مقابله با این مشکل ارائه و پیشنهاد شده است. بسیاری از الگوریتم‌های متداول جبران‌سازی نویز، بهبود قابل توجهی در دقت بازشناسی گفتار در حضور نویزهای شبه ایستا ایجاد کرده‌اند (به‌عنوان مثال [۱۰-۳]). در بسیاری از سیستم‌های توسعه داده شده برای بازشناسی

* پست الکترونیک نویسنده مسئول: Mohsen.Sadeghi@Shahroodut.ac.ir

۱. دانشجوی دکتری، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی شاهرود

۲. دانشیار، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی شاهرود

۳. دانشیار، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی شاهرود

^۲ Hidden Markov Model

^۳ Statistical Language Models

بازشناسی گفتار با استفاده از یک چارچوب تعمیم یافته ارائه داده است. در این سیستم فرض بر این است که در طول آموزش، علاوه بر داده‌های آموزشی، نوعی اطلاعات "ممتاز" وجود دارد و می‌تواند برای هدایت روند آموزش استفاده شود. این امر باعث بهبود عملکرد سیستم در زمان تست می‌شود. در مورد کارکرد تشخیص گفتار نویزی، اطلاعات ممتاز از یک مدل به نام "معلم" به دست می‌آید که تنها برای گفتار تمیز آموزش دیده است. در [۲۵] از شبکه‌های عصبی کانولوشنال خیلی عمیق به منظور بازشناسی مقاوم گفتار در برابر نویز استفاده شده است. در ساختار CNN^۵ پیشنهادی در این مقاله، تعداد لایه‌های کانولوشن به نسبت CNN های مرسوم بیشتر می‌باشد. علاوه بر این یک چارچوب انطباق با استفاده از بکارگیری همزمان شبکه CNN خیلی عمیق و ویژگی‌های کمکی i-vector^۶ و fMLLR^۶ گسترش می‌یابد. این تغییرات منجر به کاهش اساسی نرخ خطا نسبت به شبکه‌ی CNN استاندارد شده است. در [۱۲] میزان تأثیر نویزهای مختلف با سطوح سیگنال به نویز متفاوت بر روی فورمنت‌های گفتار مورد بررسی قرار گرفته است. در این مقاله از پارامترهای مختلفی به منظور نمایش میزان مقاومت و تأثیرپذیری فورمنت‌ها نسبت به نویزهای مختلف استفاده شده است. در نهایت از یک سیستم ASR به منظور اثبات عملی این نتایج بهره برده است. در مطالعه‌ی انجام شده در [۴۶] ساختار استخراج ویژگی PNCC به منظور افزایش مقاومت در برابر نویزهای محیطی مختلف از طریق ادغام فیلترینگ کانال گاماتن^۷ با مینی‌م‌سازی بایاس کانال مورد اصلاح قرار گرفته است. نتایج حاصل از شبیه‌سازی این ساختار گویای بهبود چشم‌گیر در دقت بازشناسی برای سطوح سیگنال به نویز پایین است. در این مقاله، یک مجموعه ویژگی جدید برای بازشناسی گفتار که ضرائب کپسترال توان نرمالیزه شده کسری وقفی نامیده‌ایم را تشریح خواهیم نمود. این الگوریتم جدید در واقع نسخه‌ی بهبود داده شده‌ی نسخه‌های مختلف الگوریتم PNCC متداول [۲۲] و [۲۶] می‌باشد. از آنجایی که PNCC متداول از روش‌های مبتنی بر فوریه‌ی متداول استفاده می‌کند، در

گفتار خودکار، شناسایی و بازشناسی گوینده بر اساس یکی از دو روش استخراج ویژگی ضرایب کپستروم فرکانس مل (MFCC) [۱۱-۱۴] و یا تخمین خطی ادراکی (PLP^۲) [۱۵] انجام می‌گیرد. در [۱۶] نتایج تحقیقات در حوزه‌ی تکنیک‌های دیجیتالی برای بازشناسی سیگنال نویزی تک میکروفونی در سه دسته شامل ویژگی‌های مقاوم در برابر نویز، بهبود گفتار و جبران-سازی مدل گفتار برای نویز بررسی شده است. در [۱۷] چندین روش استخراج ویژگی آکوستیک به منظور پیدا کردن مقاوم‌ترین پیکربندی front-end برای بازشناسی در محیط‌هایی با چندین نوع نویز و همچنین با نسبت سیگنال به نویز متفاوت مورد مقایسه و بررسی قرار گرفته است. در [۱۸] برخی از تکنیک‌های شناخته شده front-end و چندین تکنیک نرمالیزاسیون مقایسه شده است، اما تأثیر نویز در این بررسی در نظر گرفته نشده است. در [۱۹] شناسایی end-point و جداسازی گفتار/غیرگفتار با هم ترکیب شده و از آن در جهت بهبود گفتار در کاربردهای تجاری استفاده شده است. در [۲۰] یک روش استخراج ویژگی مقاوم در برابر نویز تحت عنوان PNCC^۳ مبتنی بر داده‌های شنیداری ارائه شده است. ویژگی‌های جدید PNCC شامل مواردی چون به کارگیری تابع غیرخطی به جای تابع لگاریتمی غیرخطی استفاده شده در ضرایب MFCC، یک الگوریتم جدید به منظور حذف تحریک پس-زمینه با استفاده از تخمین توان متوسط مبتنی بر نسبت میانگین حسابی به میانگین هندسی و تفریق توان زمان کوتاه پس‌زمینه می‌باشد. در [۲۱] یک بررسی کامل از تکنیک‌های مدرن مقاوم در برابر نویز برای سیستم‌های ASR^۴ توسعه یافته طی ۳۰ سال گذشته انجام گرفته است. این مقاله به تحلیل و طبقه‌بندی یک رنج وسیع از تکنیک‌های مقاوم در برابر نویز با استفاده از ۵ معیار متفاوت پرداخته است. در [۲۲] چندین تکنیک استخراج ویژگی front-end از قبیل MFCC، PLP و PNCC [۲۳] برای بازشناسی گفتار در محیط‌های تمیز و نویزی مورد مقایسه قرار گرفته است. [۲۴] یک سیستم مقاوم در برابر نویز برای

¹ Mel-Frequency Cepstral Coefficient

² Perceptual Linear Prediction

³ Power-Normalized Cepstral Coefficients

⁴ Automatic System Recognition

⁵ Convolutional Neural Network

⁶ feature space Maximum Likelihood Linear Regression

⁷ Gammatone Channel Filtering

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} t \\ \omega \end{bmatrix} \quad (1)$$

در رابطه فوق، u و v باید همیشه نسبت به یکدیگر متعامد باشند. اگر سیگنال در حوزه زمان یک پالس مربعی باشد، در اینصورت در حوزه فرکانس بصورت یک تابع سینک تحقق می‌یابد. تبدیل فوریه کسری هر تابعی به شکل x با زاویه α بصورت تابع $R^\alpha = X_\alpha$ تعریف می‌گردد. در اینصورت $X_\alpha(u)$ توسط رابطه (۲) به دست می‌آید [۳۰].

$$X_\alpha(u) = \int_{-\infty}^{+\infty} x(t) K_\alpha(t, u) dt \quad (2)$$

$$K_\alpha(t, u) = \sqrt{\frac{1 - j \cot\alpha}{2\pi}} e^{j \cot\alpha \frac{(t^2 + u^2)}{2}} - j u t c s c \alpha \quad (3)$$

در حالتی که α مضربی از π نباشد، رابطه (۳) صادق است. وقتی α مضربی از 2π باشد، $K_\alpha(t, u) = \delta(t-u)$ بوده و وقتی $\alpha + \pi$ مضربی از π باشد، $K_\alpha(t, u) = \delta(t+u)$ است [۳۰].

از آنجایی که گفتار واکدار یک ساختار هارمونیک دارد، تنظیم درجه‌ی تبدیل فوریه‌ی کسری به یک مقدار مرتبط با هارمونیک یا فورمنت‌ها می‌تواند باعث بهبود نمایش مشخصات متغیر با زمان گفتار گردد. در نتیجه می‌تواند مشخصات هارمونیک هارمونیک‌های گفتار را بهتر دنبال کند [۳۱]. مقدار بهینه درجه تبدیل فوریه کسری می‌تواند بسته به معیار بهینه‌سازی بدست‌آید که هر معیار دارای مزایا و معایب خاص خود است [۳۲-۳۱].

از آنجایی که سیگنال گفتار را می‌توان بصورت مجموعه‌ای از سیگنال‌های چیرپ ساده‌سازی نمود، لذا می‌توانیم بر روی نرخ چیرپ یا برخی دیگر از پارامترها توسط تبدیل فوریه کسری متمرکز شویم [۳۱]. انتخاب درجه تبدیل فوریه کسری برای تحلیل مناسب سیگنال‌های چندجزئی از قبیل گفتار هنوز مورد بحث می‌باشد. در [۳۳-۳۴] دو روش که-مبتنی بر فورمنت‌ها و فرکانس گام می‌باشد به منظور یافتن درجه FrFT پیشنهاد شده است.

۳- روش استخراج ویژگی پیشنهادی AFPNCC

در این بخش به بررسی و تشریح اجزای الگوریتم استخراج ویژگی پیشنهادی که با نام ضرائب کیسترال توان نرمالیزه

محیط‌های نویزی شدید نمی‌تواند کارایی خوبی از خود به نمایش بگذارد. اما در الگوریتم^۱ AFPNCC از تبدیل فوریه کسری که یک تبدیل زمان فرکانس بوده در کنار الگوریتم بهینه‌سازی تکامل تفاضلی^۲ به منظور دست‌یابی به ماکزیمم دقت بازشناسی گفتار در برابر نویزهای جمع-شونده شبه‌ایستا استفاده می‌شود.

این مقاله بصورت زیر سازماندهی شده است. در بخش دوم به بررسی تبدیل فوریه کسری پرداخته شده است. در بخش سوم روش استخراج ویژگی پیشنهادی AFPNCC تشریح می‌شود. سپس در بخش چهارم سیستم بازشناس گفتار با استفاده از روش‌های استخراج ویژگی متداول و همچنین روش استخراج ویژگی پیشنهادی شبیه‌سازی شده و در نهایت نتایج بدست‌آمده مورد بررسی قرار گرفته است. در بخش پنجم نتیجه‌گیری مختصری در مورد این الگوریتم پیشنهادی انجام شده است.

۲- تبدیل فوریه کسری

تبدیل فوریه کسری^۳ شکل تعمیم یافته‌ی تبدیل فوریه کلاسیک است که به دلیل انعطاف‌پذیری بیشتر نسبت به تبدیل فوریه کلاسیک در کاربردهای پردازش سیگنال بسیار مفید واقع می‌شود و مسایل عدیده‌ای که توسط تبدیل فوریه کلاسیک قابل حل نیست توسط این روش حل می‌شود [۲۷]. طی چند سال گذشته تبدیل فوریه کسری در کاربردهایی چون اپتیک و پردازش سیگنال سونار مورد استفاده قرار گرفته است. تبدیل فوریه کسری همچنین در دنبال کردن مشخصات دینامیک هارمونیک-های گفتار مفید می‌باشد [۲۸]. تبدیل FrFT کاربردهای زیادی در حل مسایل دیفرانسیل، انتشار پرتو اپتیک، روزنانس آینه کروی، تئوری پراش اپتیک، پردازش سیگنال اپتیک و شناسایی سیگنال نیز دارد [۲۹].

تبدیل فوریه کسری یک فرم بسط یافته‌ی تبدیل فوریه بوده که مبتنی بر زاویه چرخش α می‌باشد [۳۰]. زاویه چرخش α بصورت $\alpha\pi/2$ محاسبه می‌شود که $\alpha \in \mathbb{R}$ است. اگر محور زمان و فرکانس (t و ω) توسط زاویه α در جهت خلاف عقربه‌های ساعت چرخیده شود، در این صورت متغیرهای چرخش می‌تواند توسط u و v در فرم ماتریسی به صورت رابطه (۱) نشان داده شوند [۳۰]:

¹ Adaptive Fractional Power Normalized Cepstral Coefficients

² Differential Evolution Algorithm

³ Fractional Fourier Transform (FrFT)

مستطیلی معادل^۲ [۳۷] بین ۲۰۰ هرتز و ۸۰۰۰ هرتز می-باشند، وزن دار می‌شوند. علت استفاده از فیلتربانک گاماتن دقت نسبتاً بالای سیستم^۳ ASR در حضور نویز سفید می-باشد [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 پیک متحرک برای هر کانال فرکانسی (l) (کانال) بدست آورده می‌شود. اگر توان آنی کمتر از پوش مد-نظر باشد آن را حذف می‌کنیم. در بلوک ماسک‌گذاری زمانی ابتدا توان پیک برخط برای هر کانال بدست آورده می‌شود ($\bar{Q}_p[m, l]$). ماسک‌گذاری زمانی برای سگمنت‌های گفتار از رابطه (۴) بدست می‌آید [۲۳].

$$\bar{Q}_{tm}[m, l] = \begin{cases} \bar{Q}_0[m, l], & \bar{Q}_0[m, l] \geq \lambda_t \bar{Q}_p[m-1, l] \\ \mu_t \bar{Q}_p[m-1, l], & \bar{Q}_0[m, l] < \lambda_t \bar{Q}_p[m-1, l] \end{cases} \quad (4)$$

لازم به ذکر است که در رابطه (۴)، $\bar{Q}_0[m, l]$ ، λ_t و μ_t به ترتیب خروجی سیگنال عبوری از یکسوکننده نیم موج خطی، اندیس فریم، اندیس کانال و ضریب فراموشی هستند. اگر ضریب فراموشی کمتر یا مساوی ۰/۸۵ و μ_t کوچکتر یا مساوی ۰/۲ باشد، در اینصورت دقت بازشناسی برای حالت بدون نویز و نویزی ثابت است. در غیر اینصورت دقت 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. تأثیر ماسک‌گذاری زمانی در محیط-های پرنعکاس ملموس است [۲۳].

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

شده کسری وقتی (AFPNCC) معرفی شده است، می-پردازیم. ساختار این الگوریتم پیشنهادی در شکل (۱) نشان داده شده است. از آنجایی که در این روش استخراج ویژگی پیشنهادی، 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.

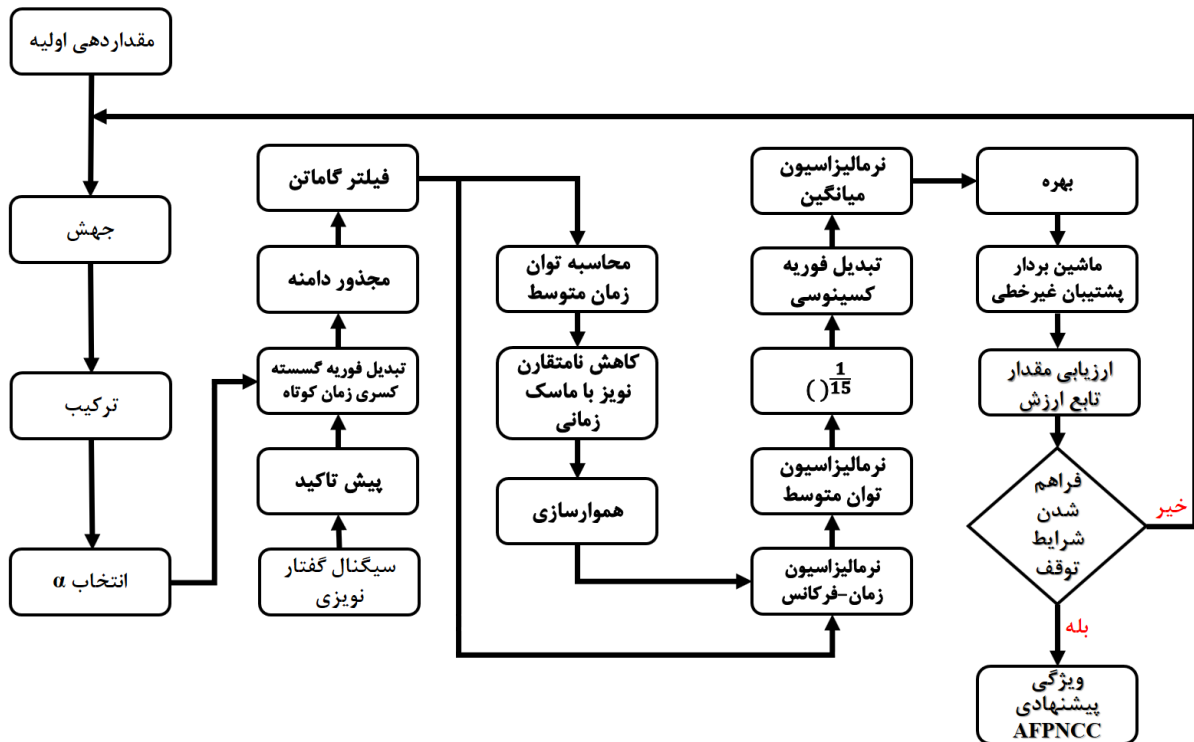
نوآوری انجام شده در این ساختار نسبت به ساختار PNCC کلاسیک را می‌توان استفاده از تبدیل فوریه کسری با ضریب بهینه‌ی بدست آمده توسط بهینه‌ساز تکامل تفاضلی به جای تبدیل فوریه کلاسیک، ایجاد یک حلقه تطبیقی به منظور استخراج ویژگی‌های مقاوم نسبت به نوع و شدت نویز و نرمالیزاسیون نهایی بعد از استخراج ویژگی توسط بلوک بهره به منظور تقویت ویژگی‌ها برای افزایش دقت سیستم ASR دانست. در قسمت شبیه‌سازی و نتایج، میزان بهبود عملکرد سیستم ASR توسط روش استخراج ویژگی پیشنهادی به نسبت سایر الگوریتم‌ها گزارش شده است.

در این الگوریتم، ابتدا پارامترهای الگوریتم بهینه‌سازی تکامل تفاضلی [۳۵-۳۶] از قبیل جمعیت اولیه، تعداد تکرار و ... مقادردهی 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 فوریه کسری می‌باشد بدست آمده و در بلوک STDFrFT^۱ وارد می‌شود. بعد از این قسمت، یک فیلتر پیش تأکید بصورت $H(z)=1-0.97Z^{-1}$ به سیگنال گفتار نویزی ورودی اعمال می‌شود. سپس از سیگنال فیلترشده، تبدیل فوریه کسری گسسته زمان کوتاه گرفته شده و پس از آن تبدیل فوریه کسری گسسته کسری زمان کوتاه سیگنال با استفاده از پنجره‌های همینگ با طول ۲۵ میلی‌ثانیه با ۱۰ میلی‌ثانیه 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

¹ Short Time Discrete Fractional Fourier Transform

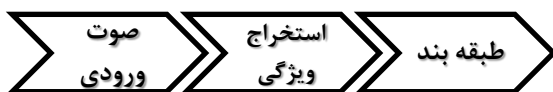
² Equivalent Rectangular Bandwidth

³ Automatic Speech Recognition



شکل ۱- بلوک دیاگرام روش استخراج ویژگی پیشنهادی AFPNCC

خروجی طبقه‌بند دقت بازشناسی و همچنین مقدار تابع ارزش یا همان Fitness که برابر با (دقت 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 دقت بازشناسی می‌شود ایجاد می‌گردند. در غیر اینصورت الگوریتم ادامه پیدا کرده و تمامی فرآیند تشریح شده مجدداً تا فراهم شدن شرایط توقف ادامه می‌یابد.



شکل ۲- بلوک دیاگرام سیستم بازشناس گفتار

۴- شبیه‌سازی و ارزیابی الگوریتم پیشنهادی

۴-۱- تنظیمات شبیه‌سازی

به منظور بررسی مقاومت و اثربخشی الگوریتم استخراج ویژگی پیشنهادی AFPNCC، عملکرد الگوریتم‌های استخراج ویژگی مختلف در محیط‌های نویزی با نویزها و سطوح SNR مختلف مورد ارزیابی قرار گرفته است. در اینجا از یک طبقه‌بند ماشین بردار پشتیبان با کرنل غیر-خطی چند جمله‌ای 10-Fold استفاده شده است. برای این

در طبقه بعد، یک تابع توان غیرخطی با ضریب $1/15$ به ورودی اعمال می‌گردد. آزمایشات انجام شده در [۳۸] نشان می‌دهد که این ضریب برای فشار صدا، زمانی که دقت باز-شناسی 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 دقت بازشناسی شده است [۳۹-۴۰]. سپس عملیات تبدیل فوری گسسته کسینوسی و همچنین نرمالیزاسیون 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 پشتیبان با کرنل چند جمله‌ای درجه ۴ با ضریب فولدینگ ۱۰ اعمال می‌شود. کرنل استفاده شده، توسط رابطه (۵) قابل محاسبه می‌باشد.

$$\text{کرنل} = \left(\frac{1}{\text{تعداد ویژگی‌ها}} \times \left(\text{ماتریس ویژگی‌ها} \right)^t \right)^4 \quad (5)$$

(بردار برچسب‌ها)

در رابطه (۵) منظور از بردار برچسب‌ها، برداری است که در آن، کلاس هر ویژگی قرار می‌گیرد.

سیستم بازشناس به‌ازای روش‌های استخراج ویژگی مختلف به‌همراه روش استخراج ویژگی پیشنهادی در جداول ۱ تا ۳ ارائه و گزارش شده است.

باتوجه به‌نتایج بدست آمده در جدول ۱ که در حضور نویز-های Babble و Factory می‌باشد، مشاهده می‌شود که در محیط نویزی با نویز Babble، در SNRهای بزرگتر از پنج دسی‌بل، دقت 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 دقت 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 دقت بازشناسی را نتیجه داده است و برای سطوح سیگنال به نویز کمتر از صفر دسی‌بل بعد از الگوریتم PNCC از لحاظ مقاومت در برابر نویز در جایگاه دوم قرار دارد. نتایج گزارش شده در جدول ۲ گویای این مطلب است که روش استخراج ویژگی پیشنهادی توانسته است دقت باز-شناسی سیستم بازشناس گفتار را در حضور نویزهای Destroyer و HF-Channel در تمامی سطوح سیگنال به نویز در بالاترین سطح دقت و بالاترین مقاومت در برابر نویز قرار دهد. این روش استخراج ویژگی حتی در محیط بدون نویز نیز بالاترین دقت بازشناسی را نتیجه داده است.

علاوه بر نویزهای مطرح شده در جداول ۱ و ۲، نتایج پیاده-سازی الگوریتم بازشناس گفتار با استفاده از روش‌های استخراج ویژگی مختلف و پیشنهادی در حضور نویزهای Pink و White در جدول ۳ گزارش شده است.

با توجه به‌جدول ۳، مشاهده می‌گردد که دقت بازشناسی روش استخراج ویژگی پیشنهادی به‌نسبت سایر روش‌های استخراج ویژگی در سطوح SNR مختلف مقدار بیشتری بدست آمده است. برای این‌که بتوان به‌درک صحیحی از میزان سودمندی روش استخراج ویژگی پیشنهادی AFPNCC در محیط‌های نویزی با سطوح سیگنال به‌نویز مختلف رسید، مقایسه‌ای بین دقت‌های بازشناسی روش-های استخراج ویژگی مرسوم با AFPNCC در حضور نویز-های مختلف در سطوح SNR از ۵- دسی‌بل تا ۵۰ دسی‌بل انجام شده که این مقایسه در شکل (۳) نشان داده شده است. همان‌طور که در شکل (۳) مشاهده می‌شود، روش استخراج ویژگی پیشنهادی در حضور نویزهای مختلف دارای دقت

هدف، هر مدل طبقه‌بندی با ۹۰ درصد از مجموع کل داده-ها 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 HF-Channel, Babble, White, Pink, Factory و Destroyer (Noise-92) بوده 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 به‌ترتیب ۲۰ میلی-ثانیه و ۱۰ میلی‌ثانیه در نظر گرفته شده است. تعداد ویژگی-های استخراج شده از هر فریم ۱۳ عدد می‌باشد.

۲-۴- 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 مؤنث 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.

۳-۴- شبیه‌سازی 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 دقت باز-شناسی مناسب در محیط نویزی است. طبقه‌بند استفاده شده در این سیستم بازشناسی یک ماشین بردار پشتیبان با کرنل غیرخطی بصورت یک چند جمله‌ای درجه ۴ می‌باشد. از آنجایی که آموزش و آزمون سیستم بازشناس گفتار بر روی یک پایگاه داده انجام شده است، به‌همین خاطر سیستم ارزیابی شده یک سیستم بازشناس 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, Pink, White و Destroyer در شش سطح سیگنال به نویز از ۵dB- تا ۵۰dB به‌منظور ارزیابی سیستم بازشناس در نرخ دقت بازشناسی استفاده شده است. لازم به‌ذکر است نویزهای مورد استفاده از پایگاه Noise-92 تهیه شده است. با اجرای الگوریتم پیشنهادی تشریح شده در بخش (۳) به‌منظور دستیابی به ماکزیمم دقت بازشناسی در محیط نویزی به مقدار ۰/۹۷ برای آلفا در تبدیل فوریه کسری دست یافته‌ایم. نتایج بدست آمده از شبیه‌سازی

بازشناسی بالایی بوده و می‌تواند بعنوان یک‌روش بسیار کار آمد در محیط‌های نویزی و بدون نویز مورد استفاده قرار گیرد.

جدول ۱- نتایج شبیه‌سازی سیستم بازشناس با روش‌های مختلف استخراج ویژگی به همراه روش پیشنهادی AFPNCC در حضور نویز- های Babble و Factory با شش سطح سیگنال به نویز از -5dB تا 5dB

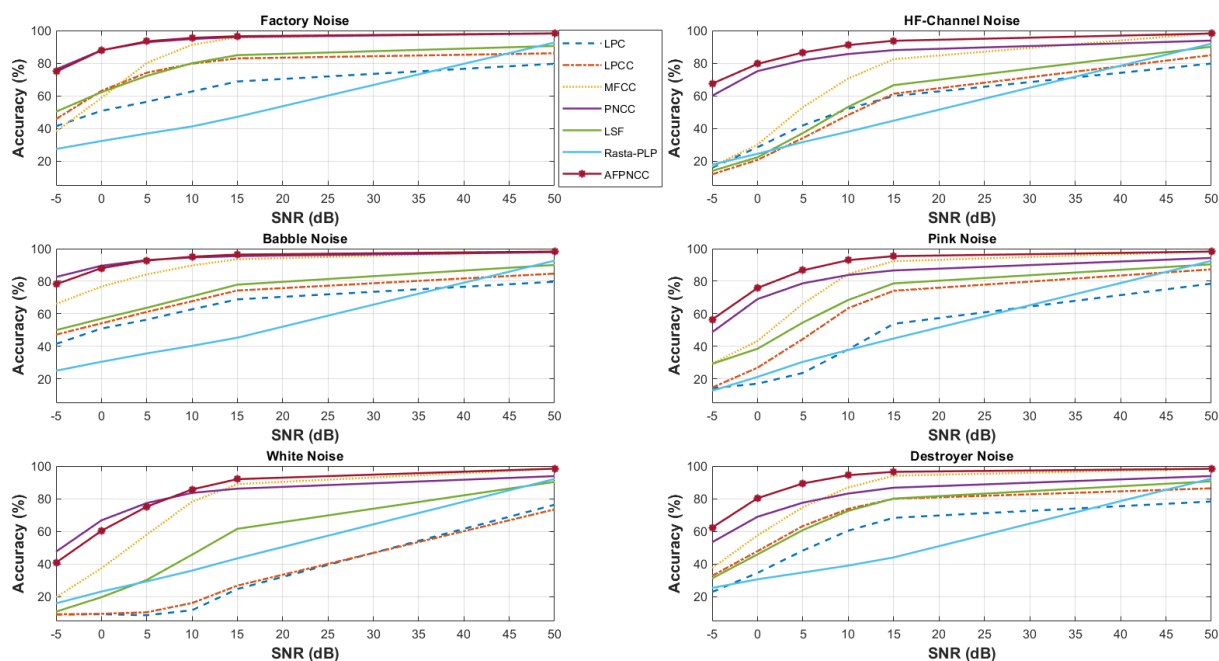
روش های استخراج ویژگی	Babble						
	نویز	-5dB	0dB	5dB	10dB	15dB	تمیز (50dB)
	SNR (dB)						
	LPC [۴۳]	۴۱/۴۸	۵۰/۸۵۶۴	۵۶/۳۴۶۱	۶۲/۷۱۴۱	۶۸/۸۱۸۶	۷۹/۶۶۶۲
	MFCC [۱۱-۱۲]	۶۶/۰۵۱۸	۷۶/۷۰۱۸	۸۴/۱۶۷۸	۸۹/۷۰۱۴	۹۳/۴۷۸۳	۹۸/۴۶۲۹
	PNCC [۲۲]	۸۲/۵۸۶۷	۸۹/۵۰۳۷	۹۲/۹۷۳۲	۹۴/۵۷۶۲	۹۵/۴۳۲۶	۹۷/۸۴۸
	LPCC [۴۳]	۴۷/۱۶۷۳	۵۴/۰۸۴۳	۶۱/۱۱۱۱	۶۷/۶۷۶۸	۷۴/۲۶۴۴	۸۴/۶۷۲۸
	LSF [۴۵]	۴۹/۹۱۲۲	۵۶/۹۶۰۹	۶۳/۶۳۶۴	۷۰/۷۲۹	۷۷/۸۸۷۶	۸۹/۹۸۶۸
	Rasta-PLP [۴۶]	۲۴/۹۶۷۱	۳۰/۴۵۶۷	۳۵/۵۷۳۱	۴۰/۲۵۰۳	۴۵/۲۵۶۹	۹۲/۶۲۱۹
	AFPNCC	۷۸/۱۲۹۱	۸۸/۰۳۲۵	۹۲/۶۸۷۷	۹۵/۰۸۱۲	۹۶/۳۷۶۸	۹۸/۲۲۱۳
نویز	Factory						
SNR (dB)	-5dB	0dB	5dB	10dB	15dB	تمیز (50dB)	
[۴۳]LPC	۲۲/۴۱۹۹	۳۷/۳۹۵۷	۵۱/۶۰۳	۶۲/۹۷۷۶	۷۱/۰۸۰۴	۷۸/۶۷۸۱	
[۱۱-۱۲]MFCC	۳۸/۲۳۰۱	۵۸/۷۳۹۶	۸۰/۱۷۱۳	۹۱/۳۰۴۳	۹۵/۷۸۳۹	۹۸/۲۸۷۲	
[۲۲]PNCC	۷۶/۲۸۴۶	۸۷/۹۲۲۷	۹۲/۸۴۱۵	۹۴/۷۷۳۸	۹۶/۰۴۷۴	۹۸/۱۷۷۴	
[۴۳]LPCC	۴۵/۸۷۱۸	۶۳/۱۰۹۴	۷۴/۱۷۶۵	۷۹/۷۱۰۱	۸۲/۹۳۸۱	۸۶/۰۷۸۲	
[۴۵]LSF	۵۰/۳۹۵۳	۶۲/۲۵۳	۷۲/۲۲۲۲	۷۹/۹۹۵۶	۸۴/۹۳۶۳	۹۰/۵۳۵۸	
[۴۶]Rasta-PLP	۲۷/۵۱۴۳	۳۲/۳۸۹۱	۳۶/۹۵۶۵	۴۱/۳۰۴۳	۴۷/۱۸۹۳	۹۲/۶۴۳۸	
AFPNCC	۷۴/۸۳۵۳	۸۷/۸۵۶۸	۹۳/۳۹۰۴	۹۵/۴۹۸۵	۹۶/۵۳۰۵	۹۸/۲۴۳۳	

جدول ۲- نتایج شبیه‌سازی سیستم بازشناس با روش‌های مختلف استخراج ویژگی به همراه روش پیشنهادی AFPNCC در حضور نویزهای Destroyer و HF-Channel با شش سطح سیگنال به نویز از -5dB تا 5dB

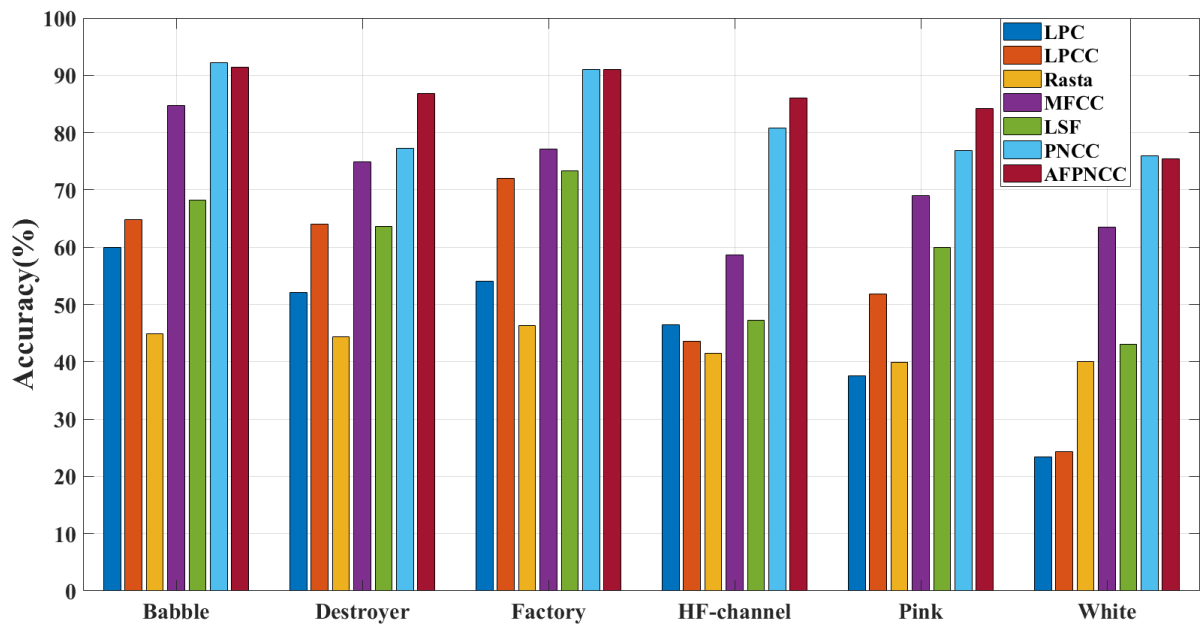
روش های استخراج ویژگی	Pink						
	نویز	-5dB	0dB	5dB	10dB	15dB	تمیز (50dB)
	SNR (dB)						
	LPC [۴۳]	۱۴/۲۵۱۲	۱۶/۹۵۲۱	۲۳/۶۰۵۶	۳۸/۰۷۶۴	۵۳/۷۹۸۹	۷۸/۵۰۲۴
	MFCC [۱۱-۱۲]	۲۹/۴۰۲۷	۴۳/۲۱۴۸	۶۶/۱۸۳۶	۸۴/۵۱۹۱	۹۲/۳۱۴۴	۹۸/۳۳۱۱
	PNCC [۲۲]	۴۸/۶۳۸۶	۶۸/۹۹۴۳	۷۸/۷۲۲	۸۳/۷۹۴۵	۸۶/۶۰۵۲	۹۴/۳۱۲۷
	LPCC [۴۳]	۱۴/۵۸۰۶	۲۶/۸۱۱۶	۴۴/۴۴۴۴	۶۳/۴۱۶۸	۷۴/۱۱۰۷	۸۷/۲۶۳۹
	LSF [۴۵]	۲۹/۱۳۹۲	۳۸/۴۴۹۷	۵۴/۶۱۱۳	۶۸/۴۴۵۳	۷۸/۶۳۴۲	۹۰/۳۶۰۱
	Rasta-PLP [۴۶]	۱۲/۶۴۸۲	۲۱/۲۴۳	۳۰/۳۹۰۹	۳۷/۷۲۵۱	۴۴/۷۵۱۹	۹۲/۵۱۲۱
	AFPNCC	۵۶/۴۵۵۹	۷۵/۸۲۳۵	۸۶/۶۷۱۱	۹۲/۹۵۱۳	۹۵/۳۸۸۷	۹۸/۲۲۱۳
نویز	White						
SNR (dB)	-5dB	0dB	5dB	10dB	15dB	تمیز (50dB)	
LPC [۴۳]	۹/۰۹۰۹	۹/۳۳۲۵	۱۰/۷۱۷۶	۱۱/۷۶۹۹	۲۴/۷۶۹۴	۷۶/۳۲۸۵	
MFCC [۱۱-۱۲]	۱۹/۷۸۴۸	۳۷/۴۸۳۵	۵۸/۰۱۴۹	۷۸/۱۷۳	۸۸/۹۵۴۸	۹۸/۳۷۵۱	
PNCC [۲۲]	۴۷/۶۵۰۴	۶۶/۷۷۶۵	۷۷/۳۶۰۶	۸۳/۵۷۴۹	۸۶/۱۴۴	۹۳/۸۲۹۶	
LPCC [۴۳]	۹/۳۱۰۵	۹/۵۹۶	۱۰/۵۴۰۲	۱۶/۱۸۳۶	۲۶/۸۳۳۶	۷۳/۴۷۲۹	
LSF [۴۵]	۱۰/۹۱۳۵	۱۹/۷۶۲۸	۳۰/۳۲۵	۴۵/۷۶۲	۶۱/۵۹۴۲	۹۰/۲۹۶۲	
Rasta-PLP [۴۶]	۱۶/۰۰۷۹	۲۳/۲۱۰۴	۲۹/۴۲۴۷	۳۶/۰۳۴۳	۴۳/۵۰۰۲	۹۲/۰۷۲۹	
AFPNCC	۴۰/۹۳۱	۶۰/۳۶۴۵	۷۵/۱۴۲۷	۸۵/۶۸۲۹	۹۱/۹۸۵۱	۹۸/۳۹۷	

جدول ۳- نتایج شبیه‌سازی سیستم بازشناس با روش‌های مختلف استخراج ویژگی به‌همراه روش پیشنهادی AFPNCC در حضور نویزهای Pink و White با شش سطح سیگنال به نویز از -۵dB تا ۵۰dB

روش‌های استخراج ویژگی	نویز	Destroyer ops					
	SNR (dB)	-۵dB	۰dB	۵dB	۱۰dB	۱۵dB	تمیز(۵۰dB)
	LPC [۴۳]	۲۲/۹۴۶۹	۳۴/۶۰۶۹	۴۸/۲۸۷۲	۶۰/۴۵۲۳	۶۸/۳۱۳۶	۷۸/۳۰۴۸
	MFCC [۱۱-۱۲]	۳۷/۵۹۳۳	۵۷/۵۷۵۸	۷۴/۸۷۹۲	۸۶/۸۹۰۶	۹۴/۰۷۱۱	۹۸/۳۰۹۲
	PNCC [۲۲]	۵۳/۴۰۳۶	۶۸/۶۷۲۳	۷۷/۵۸۰۱	۸۳/۱۷۹۶	۸۶/۷۳۶۹	۹۳/۸۷۳۵
	LPCC [۴۳]	۳۲/۸۵۰۲	۴۷/۹۳۵۹	۶۳/۳۰۷	۷۳/۸۰۳۲	۷۹/۹۵۱۷	۸۶/۳۶۳۶
	LSF [۴۵]	۳۱/۲۹۱۲	۴۵/۹۵۹۶	۶۰/۸۴۷۶	۷۲/۵۹۵۵	۸۰/۱۰۵۴	۹۰/۶۴۵۶
	Rasta-PLP [۴۶]	۲۵/۴۲۸۲	۳۰/۶۹۸۳	۳۴/۸۷۰۴	۳۹/۰۸۶۵	۴۴/۰۹۳۱	۹۲/۳۱۴۴
	AFPNCC	۶۲/۳۸۴۷	۸۰/۳۵۲	۸۹/۳۷۲	۹۴/۴۰۰۵	۹۶/۴۶۴۶	۹۸/۳۵۳۱
نویز		HF Channel					
	SNR (dB)	-۵dB	۰dB	۵dB	۱۰dB	۱۵dB	تمیز(۵۰dB)
	LPC [۴۳]	۱۶/۰۲۹۹	۲۸/۵۹۰۳	۴۱/۹۴۱۲	۵۲/۱۰۸	۵۹/۹۲۵۳	۷۹/۷۹۸
	MFCC [۱۱-۱۲]	۱۶/۶۸۸۶	۳۰/۲۵۹۱	۵۳/۱۴۰۱	۷۰/۶۸۵۱	۸۲/۴۷۶۹	۹۸/۴۶۲۹
	PNCC [۲۲]	۶۰/۱۰۱	۷۵/۱۶۴۷	۸۱/۵۷۲۳	۸۵/۶۳۹	۸۷/۹۸۸۶	۹۳/۸۲۹۶
	LPCC [۴۳]	۱۲/۰۵۵۳	۲۰/۸۶۰۸	۳۴/۲۱۱۷	۴۸/۳۳۱۱	۶۱/۳۳۰۷	۸۴/۹۳۶۳
	LSF [۴۵]	۱۴/۰۷۵۵	۲۲/۴۶۳۸	۳۷/۲۴۲	۵۳/۲۹۳۸	۶۶/۵۵۶۹	۹۰/۰۷۴۷
	Rasta-PLP [۴۶]	۱۷/۹۶۲۲	۲۴/۵۹۳۸	۳۱/۷۵۲۳	۳۸/۰۹۸۴	۴۴/۸۳۹۷	۹۱/۹۱۹۲
	AFPNCC	۶۷/۴۷۹۱	۷۹/۵۷۸۴	۸۶/۵۸۳۲	۹۱/۱۰۶۷	۹۳/۷۱۹۸	۹۸/۱۵۵۵



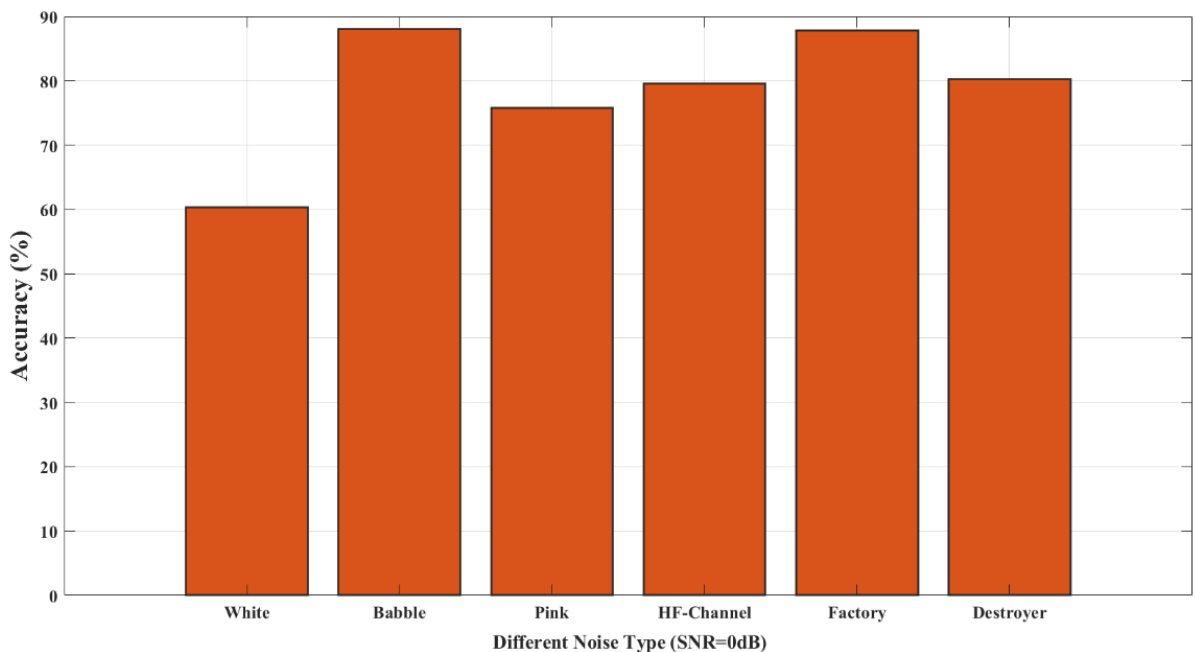
شکل ۳- دقت‌های بازشناسی روش‌های استخراج ویژگی مرسوم به‌همراه روش پیشنهادی AFPNCC در حضور نویزهای مختلف در سطوح SNR از -۵ دسی‌بل تا ۵۰ دسی‌بل



شکل ۴- میانگین دقت بازشناسی سیستم بازشناس گفتار برای روش‌های استخراج ویژگی مختلف و الگوریتم پیشنهادی در حضور نویز-های مختلف

مشاهده می‌شود که میانگین دقت 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 چشم‌گیری در رتبه اول دقت بازشناسی قرار می‌گیرد.

در شکل (۴) میانگین دقت بازشناسی به‌ازای سطوح SNR مختلف در حضور نویزهای Babble, Destroyer, HF-Channel, Factory, Pink و White برای روش‌های استخراج ویژگی مختلف به‌همراه روش استخراج ویژگی پیشنهادی AFPNCC نمایش داده شده است. در این شکل



شکل ۵- دقت‌های بازشناسی روش استخراج ویژگی پیشنهادی AFPNCC در حضور نویزهای مختلف با سطح سیگنال به نویز صفر دسی‌بل

می‌شود، همواره به دنبال روش استخراج ویژگی با بیشترین مقاومت در برابر نویز هستیم. بدین منظور در این مقاله روشی جدید و مقاوم در برابر نویز برای کاربرد بازشناسی گفتار در محیط‌های نویزی ارائه و پیشنهاد شده است. روش استخراج ویژگی پیشنهادی که با نام AFPNCC معرفی شده است، مبتنی بر تبدیل فوریه کسری زمان کوتاه بوده که در آن از روش‌های فرا ابتکاری تکامل تفاضلی به منظور یافتن ضرایب تبدیل فوریه کسری استفاده می‌شود. این روش یک روش استخراج ویژگی مقاوم و بهینه شده برای کاربرد بازشناسی گفتار می‌باشد. در واقع این روش پیشنهادی نسخه بهبود داده شده‌ی روش PNCC می‌باشد. الگوریتم پیشنهادی AFPNCC در این مقاله توسط راهکارهایی چون استفاده از خاصیت چرخش تبدیل فوریه کسری، استفاده از روش بهینه‌سازی تکامل تفاضلی به منظور انتخاب مقدار بهینه‌ی ضریب آلفای تبدیل فوریه کسری و استفاده از تبدیل فوریه کسری به نحوی که نویز بر روی تعداد محدودی از فریم‌های سیگنال گفتار (در حوزه فرکانس) تأثیرگذار، توانست دقت بازشناسی در محیط نویزی حتی با سطوح سیگنال به نویز کم را بهبود بخشد و نسبت به سایر روش‌های استخراج ویژگی عملکرد بهتری را ارائه دهد.

در شکل (۵) که دقت بازشناسی سیستم بازشناس گفتار به ازای نویزهای مختلف با سطح سیگنال به نویز صفر دسی-بل ترسیم شده است، می‌توان مشاهده نمود که در حضور تمامی این نویزها دقت بازشناسی مقدار بالایی دارد. علاوه بر این می‌توان بیان کرد که نویز White بیشترین تأثیر در کاهش دقت بازشناسی و نویزهای Babble و Factory کمترین تأثیر را دارند.

علت برتری روش استخراج ویژگی پیشنهادی در استفاده‌ی هدفمند از تبدیل فوریه کسری با ضریب بهینه منطبق بر نوع و شدت نویز، استفاده از حلقه تطبیقی وابسته به نوع و شدت نویز و همچنین در نرمالیزاسیون دامنه ویژگی‌های بهینه‌ی استخراج شده توسط طبقه‌ی بهره است. ادغام این راهکارها در کنار هم توانسته است از روش پیشنهادی یک روش مقاوم در برابر نویز با دقت بازشناسی بالا رقم بزند. در روش استخراج ویژگی پیشنهادی، نویز تنها در چند فریم محدود تأثیر خواهد داشت و همین امر در کنار سایر راهکارهای اعمال شده در این روش توانسته است منجر به این عملکرد خوب شود.

۵- نتیجه گیری

از آنجایی که اختلاف شرایط موجود در حالت آموزش و آزمون منجر به کاهش دقت بازشناسی در محیط‌های نویزی

مراجع

- [1] L. R. Rabiner and B.-H. Juang, "Fundamentals of Speech Recognition", Englewood Cliffs, NJ: Prentice Hall, 1993.
- [2] F. Jelinek, "Statistical methods for speech recognition", MIT press, 1997.
- [3] A. Acero and R. M. Stern, "Environmental robustness in automatic speech recognition", IEEE International Conference on Acoustics, Speech, and Signal Processing, Albuquerque, New Mexico, Vol. 2, April 1990, pp. 849-852.
- [4] P. J. Moreno, B. Raj, and R. M. Stern, "A vector Taylor series approach for environment-independent speech recognition", IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, May 1996, pp. 733-736.
- [5] P. Pujol, D. Macho, and C. Nadeu, "On real-time mean-and-variance normalization of speech recognition features", IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, May 2006, pp. 773-776.
- [6] R. M. Stern, B. Raj, and P. J. Moreno, "Compensation for environmental degradation in automatic speech recognition" ESCA Workshop on Robust Speech Recognition for Unknown Communication, April 1997, pp. 33-42.
- [7] R. Singh, R. M. Stern, B. Raj and G. Davis, "Signal and feature compensation methods for robust speech recognition", Noise Reduction in Speech Applications, April 2002, pp. 221-246.
- [8] J. Droppo, and A. Acero, "Noise robust speech recognition with a switching linear dynamic model", IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, Que., Canada, Vol. 1, May 2004, pp. 953-956.

- [9] S. Molau, M. Pitz, and H. Ney, "Histogram based normalization in the acoustic feature space", IEEE Workshop Automatic Speech Recognition Understanding, November 2001, pp. 21–24.
- [10] H. Misra, S. Ikbal, H. Bourlard, and H. Hermansky, "Spectral entropy based feature for robust ASR", IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, May 2004, pp. 193–196.
- [11] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 28, No. 4, August 1980, pp. 357–366.
- [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, September 2018, pp. 741–752.
- [۱۳] علی حریمی و خشایار یغمائی، "بهبود نرخ تشخیص احساس از روی گفتار با استفاده از محدودیت تفکیک جنسیتی گویندگان"، نشریه مدل سازی در مهندسی، دوره ۱۵، شماره ۴۸، بهار ۱۳۹۶، صفحه ۱۸۳–۲۰۰.
- [۱۴] سیدعلی سیلمانی ایوری، محمدرضا فدوی امیری و حسین مروی، "تولید سیگنال مصنوعی زلزله به کمک مدلی جدید در فشرده سازی و آموزش شبکه های عصبی مصنوعی"، نشریه مدل سازی در مهندسی، دوره ۱۴، شماره ۴۶، پاییز ۱۳۹۵، صفحه ۷۵–۸۵.
- [15] H. Hermansky, "Perceptual linear prediction analysis of speech", Journal of the Acoustical Society of America, Vol. 87, No. 4, April 1990, pp. 1738–1752.
- [16] Y. Gong, "Speech recognition in noisy environments: A survey", Speech communication, Vol. 16, No. 3, April 1995, pp. 261-291.
- [17] D. S. Kim, S. Y. Lee, and R. M. Kil, "Auditory processing of speech signals for robust speech recognition in real-world noisy environments", IEEE Transactions on speech and audio processing, Vol. 7, No. 1, January 1999, pp. 55-69.
- [18] B. Milner, "A comparison of front-end configurations for robust speech recognition", IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, May 2002, pp. I-797.
- [19] S. J. Lee, H. Chung, J. G. Park, H. Y. Jung, and Y. Lee, "A commercial car navigation system using Korean large vocabulary automatic speech recognizer", Asia-Pacific Signal and Information Processing Association, Sapporo, Japan, October 2009, pp. 286-289.
- [20] C. Kim, and R. M. Stern, "Feature extraction for robust speech recognition using a power-law nonlinearity and power-bias subtraction", Tenth Annual Conference of the International Speech Communication Association, Brighton, UK, September 2009, pp. 28-31.
- [21] 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, February 2014, pp. 745-777.
- [22] G. Sárosi, M. Mozsáry, P. Mihajlik, and T. Fegyó, "Comparison of feature extraction methods for speech recognition in noise-free and in traffic noise environment", 6th Conference on Speech Technology and Human-Computer Dialogue, Brasov, Romania, May 2011, pp. 18-21.
- [23] 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, July 2016, pp. 1315-1329.
- [24] K. Markov, and T. Matsui, "Robust Speech Recognition Using Generalized Distillation Framework", International Speech Communication Association, San Francisco, USA, September 2016, pp. 2364-2368.
- [25] Y. Qian, M. Bi, T. Tan, and K. Yu, "Very deep convolutional neural networks for noise robust speech recognition", IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 24, No. 12, August 2016, pp. 2263-2276.
- [26] C. Kim and R. M. Stern, "Feature extraction for robust speech recognition based on maximizing the sharpness of the power distribution and on power flooring", IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, USA, March 2010, pp. 4574–4577.
- [27] K. K. Sinha, "The Fractional Fourier Transform in Signal Processing", International Journal of Scientific and Research Publications, Vol. 3, No. 2, February 2013, pp. 1-3.
- [28] U. K. Agrawal, M. Chandra, and C. Badgaiyan "Fractional Fourier transform combination with MFCC based speaker identification in clean environment", International Journal of Advanced Science, Engineering and Technology, Vol. 1, No. 1, October 2012, pp. 26-28.

- [29] V. A. Narayanan, and K. M. M. Prabhu, "The fractional Fourier transform: theory, implementation and error analysis", *Microprocessors and Microsystems*, Vol. 27, No. 10, November 2003, pp. 511-521.
- [30] P. Kumar, and S. Kansal, "Noise removal in speech signal using fractional Fourier transform", *International Conference on Information, Communication, Instrumentation and Control*, Indore, India, August 2017, pp. 1-4.
- [31] D. J. Ma, X. Xie, and J. M. Kuang, "A novel algorithm of seeking FrFT order for speech processing", *IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, May 2011, pp. 3832-3835.
- [32] R. Saxena, and K. Singh, "Fractional Fourier transform: A novel tool for signal processing", *Journal of the Indian Institute of Science*, Vol. 85, No. 1, February 2005, pp. 11-26.
- [33] H. Yin, C. Nadeu, and V. Hohmann, "Pitch-and formant-based order adaptation of the fractional Fourier transform and its application to speech recognition", *EURASIP Journal on Audio, Speech, and Music Processing*, Vol. 2009, January 2009, pp. 1-14.
- [34] H. Yin, C. Nadeu, and V. Hohmann, "Using pitch and formants for order adaptation of fractional Fourier transform in speech signal processing", *V Jornadas en Tecnología del Habla*, 2008, pp. 71-74.
- [35] M. Sadeghi and H. Marvi, "Optimal MFCC features extraction by differential evolution algorithm for speaker recognition", *3rd Iranian Conference on Intelligent Systems and Signal Processing*, Shahrood, Semnan, Iran, December 2017, pp. 169-173.
- [۳۶] امید خطیبی بردسیری، سیدمحسن هاشمی و محمدرضا رزازی، "ارائه یک مدل جدید جهت تخمین تالش الزم برای توسعه سرویس-های نرم‌افزاری"، *نشریه مدل‌سازی در مهندسی*، دوره ۱۵، شماره ۴۹، تابستان ۱۳۹۶، صفحه ۲۴۵-۲۶۱.
- [37] B. C. Moore and B. R. Glasberg, "A revision of Zwicker's loudness model", *Acta Acustica united with Acustica*, Vol. 82, No. 2, March 1996, pp. 335-345.
- [38] C. Kim, "Signal processing for robust speech recognition motivated by auditory processing", *Carnegie Mellon Univ.*, Pittsburgh, PA USA, Dec. 2010 [Online]. Available: <http://www.cs.cmu.edu/~robust/Thesis/ChanwooKimPhDThesis.pdf>.
- [39] E. Principi, S. Squartini, and F. Piazza, "Power Normalized Cepstral Coefficients based supervectors and i-vectors for small vocabulary speech recognition", *International Joint Conference on Neural Networks*, Beijing, China, July 2014, pp. 3562-3568.
- [40] M. Bashirpour, and M. Geravanchizadeh, "Speech emotion recognition based on power normalized cepstral coefficients in noisy conditions", *Iranian Journal of Electrical and Electronic Engineering*, Vol. 12, No. 3, September 2016, pp. 197-205.
- [41] A. Varga and H. J. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems", *Speech communication*, Vol. 12, No. 3, July 1993, pp. 247-251.
- [42] <http://www ldc.upenn.edu/readme.files/tidigits.readme.html>.
- [43] I. Missaoui, and Z. Lachiri, "Gabor filterbank features for robust speech recognition", *International Conference on Image and Signal Processing*, Cherbourg, France, Vol. 8509, June 2014, pp. 665-671.
- [44] V.Z. Kėpuska, and H. A. Elharati, "Robust speech recognition system using conventional and hybrid features of MFCC, LPCC, PLP, RASTA-PLP and hidden markov model classifier in noisy conditions", *Journal of Computer and Communications*, Vol. 3, No. 6, May 2015, pp. 1-9.
- [45] V. F. S. de Alencar, and A. Alcaim "Transformations of LPC and LSF parameters to speech recognition features", *International Conference on Pattern Recognition and Image Analysis*, Berlin, Heidelberg, Vol. 3686, August 2005.
- [46] 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, January 2019, pp. 1-13.