



افزایش دقت بازشناسی اعداد در حضور نویز با استفاده از انواع ویژگی‌های مقاوم

هادی ویسی امین فاضل دهکردی حسین ثامتی

دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف

veisi@mehr.sharif.edu afazel@mehr.sharif.edu sameti@sharif.edu

۱. مقدمه

سیستم‌های بازشناسی گفتار در شرایط آزمایشگاهی برای داده‌های تمیز به نرخ قابل قبولی از کارایی می‌رسند اما هنگام استفاده از آنها در عمل، به خاطر وجود انواع مختلف اثرهای محیطی دقت سیستم افت شدیدی می‌کند. شرایط نامناسبی همچون 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 دقت این سیستم‌ها در شرایط واقعی و همچنین بر حذر داشتن سیستم از نیاز به مجموعه 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 دقت قابل قبولی عمل می‌کند، ممکن است در محیطی دیگر قادر به رسیدن به آن دقت نباشد. می‌توان این روش‌ها را در یک نگاه کلی در سه گروه استخراج ویژگی‌های مقاوم، روش‌های بهبود گفتار و روش‌های مبتنی بر مدل قابل دسته بندی هستند [۱]. در روشهای استخراج ویژگی‌های مقاوم که در این مقاله مورد بررسی قرار می‌گیرند، سعی می‌شود با توجه به تأثیر نویز بر خود سیگنال ویا ویژگی‌های آن، ویژگی‌هایی از سیگنال گفتار استخراج شود که نویز بر آن‌ها بی‌اثر یا کم‌اثر باشد. با این

چکیده: در این مقاله به مرور روشهای مختلف استخراج ویژگی‌های مقاوم به نویز برای بازشناسی گفتار پرداخته شده است و توانایی هر کدام از این روش‌ها در شرایط نویزی مختلف روی یک سیستم بازشناسی اعداد ارزیابی شده‌اند. برای این سیستم بازشناسی که مبتنی بر مدل مخفی مارکوف است، از روش MFCC به عنوان مبنا برای استخراج ویژگی‌های مورد نیاز سیستم استفاده می‌شود. روشهای استخراج ویژگی مقاوم به نویز مختلفی مانند تفاضل میانگین ضرایب کپسترا، ضرایب کپسترا حاصل از پیشگویی خطی، تحلیل طیف نسبی، ضرایب کپسترا ریشه‌ای و تحلیل اجزای اصلی در نویزهای مختلف با نسبت سیگنال به نویزهای متفاوت مورد ارزیابی و مقایسه قرار گرفته‌اند. برای انجام آزمایش‌ها از دادگان گفتاری Aurora 2 استفاده شده است که شامل دادگان تمیز و نویزی با اثر کانال و نویزهای جمع‌شونده مختلف در نسبت سیگنال به نویزهای مختلف می‌باشد. نتایج بکارگیری این روش‌ها منجر به بهبود دقت بازشناسی در شرایط مختلف نویزی می‌شود هرچند هر کدام از این روش‌ها برای برخی شرایط نویزی بهتر از برخی شرایط دیگر عمل می‌کنند.

کلمات کلیدی: بازشناسی مقاوم گفتار، استخراج ویژگی‌های مقاوم، بازشناسی اعداد تلفنی.

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. در این مقاله به بررسی عملکرد این روشها در حضور نویزهای جمع شونده^۱ مختلف و نویز کانال تلفنی شبیه سازی شده، بر روی سیستم بازشناسی اعداد انگلیسی پرداخته شده است. در ادامه، ابتدا در بخش دوم، به معرفی دادگان Aurora 2 و سیستم بازشناسی مورد استفاده پرداخته می‌شود. در بخش سوم روشهای مختلف استخراج ویژگی‌های مقاوم مورد بررسی قرار می‌گیرند و نتایج آنها در حضور نویزهای جمع شونده و نویز کانال بیان می‌شوند. مقایسه عملکرد روشهای مورد بررسی با تغییر کانال و در حضور نویز جمع شونده در بخش چهارم انجام می‌شود و سرانجام در بخش پنجم، مقایسه این روشها به همراه نتیجه‌گیری و جمع‌بندی مطالب بیان شده آورده شده است.

۲. معرفی دادگان و سیستم بازشناسی

دادگان Aurora 2 [۲] برای بررسی میزان دقت سیستم‌های بازشناسی گفتار در محیط‌های نویزی و بویژه برای بررسی کارایی الگوریتم‌های استخراج ویژگی مقاوم در شرایط نویزی مختلف مورد استفاده قرار می‌گیرد. این دادگان از دادگان گفتاری TIDigits استفاده می‌کند و از اعداد متصل که به وسیله گوینده‌های انگلیسی زبان آمریکایی بیان شده‌اند، تشکیل شده است. در این دادگان ابتدا TIDigits به ۸ کیلو هرتز زیرنمونه‌برداری^۲ شده است و پس از آن فیلتری جهت شبیه‌سازی خط تلفن بر روی دادگان اعمال گردیده است. دو نوع فیلتر به نامهای G.712 و MIRS [۲] برای انجام کار شبیه سازی استفاده شده است. گفتار حاصل از اعمال فیلترهای فوق در این دادگان گفتار تمیز دادگان را تشکیل می‌دهند. سپس به این دادگان تمیز نویزهایی با نسبت سیگنال به نویزهای مختلف

بطور جداگانه اضافه گردیده است تا مجموعه‌هایی با نسبت سیگنال به نویزهای -۵، ۰، ۵، ۱۰، ۱۵ و ۲۰ دسی‌بل را بسازند. نویزهای مورد استفاده عبارتند از نویز همهمه (babble)، رستوران (restaurant)، نمایشگاه (exhibition)، ماشین (car)، خیابان (street)، فرودگاه (airport)، مترو (subway) و نویز ایستگاه قطار (train-station).

در این دادگان مجموعه‌های آموزش و آزمایش بطور جداگانه مشخص شده‌اند. داده‌های آموزشی این دادگان خود بر دو نوع دادگان آموزشی تمیز و دادگان آموزشی ترکیبی (شامل تمیز و نویزی) می‌باشد. این دو مجموعه آموزشی هر دو به وسیله فیلتر G.712 فیلتر شده‌اند و شامل ۸۴۴۰ جمله می‌باشند. در دادگان آموزشی ترکیبی ۸۴۴۰ جمله به بیست زیر مجموعه ۴۲۲ جمله‌ای دسته‌بندی شده‌اند که این بیست زیر مجموعه در بر گیرنده چهار نوع نویز مترو، همهمه، ماشین و نمایشگاه در چهار نسبت سیگنال به نویز مختلف ۵، ۱۰، ۱۵ و ۲۰ دسی‌بل و چهار زیر مجموعه بدون نویز می‌باشند. دادگان آزمایش نیز خود شامل سه مجموعه A، B و C است. دادگان آزمایشی A شامل ۴۰۰۴ جمله فیلتر شده با فیلتر G.712 می‌باشد که به چهار زیر مجموعه ۱۰۰۱ جمله‌ای تقسیم‌بندی شده‌اند. علاوه بر زیرمجموعه‌های تمیز این مجموعه، به هر یک از این زیر مجموعه‌ها نیز نویزهای مترو، همهمه، ماشین و نمایشگاه اضافه شده‌اند که برای هر دسته نویزی هم مجموعه‌هایی در نسبت‌های مختلف سیگنال به نویز -۵، ۰، ۵، ۱۰، ۱۵ و ۲۰ دسی‌بل وجود دارند تا با هم هفت حالت مختلف را تشکیل دهند. دادگان آزمایش B نیز دقیقاً شبیه به روش فوق به دست آمده‌اند با این تفاوت که در اینجا از نویزهای رستوران، خیابان، فرودگاه و نویز ایستگاه قطار استفاده شده است. در این حالت عدم انطباق بین دادگان آموزش و آزمایش حتی برای حالت آموزش به وسیله دادگان آموزشی ترکیبی هم وجود دارد، چرا که دادگان آموزشی ترکیبی فقط شامل نویزهای مجموعه آزمایش A است. در دادگان آزمایش C دو زیر مجموعه از چهار زیر مجموعه ۱۰۰۱ جمله‌ای وجود دارند. در این مجموعه داده‌های گفتار و نویز به وسیله فیلتر MIRS، فیلتر شده‌اند و سپس نویزهای مترو و خیابان با نسبت سیگنال به

^۱ - Additive

^۲ - Down sampling

نویزهای ۵-، ۰، ۵، ۱۰، ۱۵ و ۲۰ به گفتار اضافه شده است. در اینجا نیز حالت تمیز، یعنی بدون حضور نویز جمع شونده وجود دارد. این دسته از دادگان آزمایش برای نشان دادن اثر نویز کانالهای متفاوت بر روی سیستم تشخیص گفتار، تهیه شده است.

سیستمی بازشناسی که در این مقاله در آزمایشات مورد استفاده قرار گرفته است، HTK می باشد [۳]. این سیستم مدل سازی واحدهای آوایی را با استفاده از مدل مخفی مارکوف انجام می دهد. HMM های به کار برده شده در این سیستم از نوع چگالی پیوسته و با تلفیق های گوسی می باشند که پرش بین حالت های آن فقط به صورت چپ به راست مجاز می باشد. هر عدد زبان انگلیسی با یک HMM دارای ۱۶ حالت و ۳ تلفیق گوسی مدل شده است.

۳. روشهای استخراج ویژگی های مقاوم

در این بخش روشهای مختلف استخراج ویژگی های مقاوم به همراه نتایج حاصل از پیاده سازی آنها روی سیستم بازشناسی اعداد انگلیسی بررسی می شوند. MFCC^۱ به عنوان روش مبنای استخراج ویژگی استفاده شده است که درصد دقت^۲ بازشناسی سیستم با آن برای مجموعه های آزمایش A و B در دادگان، برای نویزهای مختلف و در نسبت سیگنال به نویزهای متفاوت به صورت جدول ۱ آورده شده است.

جدول ۱: نتایج بازشناسی با MFCC (مبنای)

SNR ▴ Noise ▾	clean	20	15	10	5	0	-5	Avg
Subway	98.9	97.1	93.5	78.7	52.2	26.0	11.2	65.4
Babble	99.0	90.2	73.8	49.4	26.8	9.3	1.6	50.0
Car	99.0	97.4	90.0	67.0	34.1	14.5	9.4	58.8
Exhibition	99.2	96.4	92.0	75.7	44.8	18.1	9.6	62.3
Restaurant	98.9	90.0	76.2	54.8	31.0	11.0	3.5	52.2
Street	99.0	95.7	88.5	67.1	38.5	17.8	10.5	59.6
Airport	99.0	90.6	77.0	53.9	30.3	14.4	8.2	53.3
Train-station	99.2	94.7	83.7	60.3	27.9	11.6	8.5	55.1

۱-۳ تفاضل میانگین ضرایب کپسترال^۳

یک روش معمول برای مقاوم سازی ویژگی های کپسترال در مقابل نویز جمع شونده در این حوزه و بطور معادل نویز

پیشگی^۴ در حوزه زمان، کم کردن مقدار میانگین آنها از مقدارشان است [۴]. علت این کار تغییر مقدار میانگین ویژگی ها در حضور نویز جمع شونده است [۱]. بردار ویژگی تفاضل میانگین ضرایب کپسترال از روی N عنصر موجود در بردار ضرایب کپسترال $\vec{C}_{y,i}$ و گفتار عبور کرده از کانال $y(n)$ با کم کردن میانگین $\mu_y = \frac{1}{N} \sum_{i=1}^N \vec{C}_{y,i}$ از هر یک از مولفه های بردار اصلی کپسترال به صورت رابطه (۱) بدست می آید تا بردار ویژگی ها به میانگین نرمالیزه شوند^۵.

$$\vec{C}_{MN,i} = \vec{C}_{y,i} - \mu_y, i = 1, \dots, N \quad (1)$$

علاوه بر میانگین، می توان این نرمال سازی را برای واریانس هم انجام داد ولی معمولاً انجام این کار بهبود زیادی را به همراه ندارد. بکارگیری این روش روی دادگان آزمایش مورد استفاده نتایج جدول ۲ را به همراه داشته است. اعمال این روش، به غیر از کاهش اندک نتایج در نویزهای مترو و نمایشگاه، بهبود قابل توجهی را برای سایر نویزها به دنبال داشته است.

جدول ۲: نتایج بازشناسی با MFCC-MN

SNR ▴ Noise ▾	clean	20	15	10	5	0	-5	Avg
Subway	99.0	96.3	91.6	77.7	51.0	24.7	14.0	64.9
Babble	99.2	97.6	93.7	83.5	58.2	25.2	12.2	67.1
Car	99.0	97.1	93.0	78.1	46.7	21.1	12.9	64.0
Exhibition	99.2	95.3	89.5	74.5	46.2	20.3	10.5	62.2
Restaurant	99.0	97.5	94.4	85.2	63.1	30.7	12.6	68.9
Street	99.2	97.2	93.3	81.9	56.0	25.4	12.7	66.5
Airport	99.0	97.3	94.9	86.9	63.3	31.7	15.4	69.8
Train-station	99.2	97.8	94.0	82.9	56.2	25.4	12.7	66.9

۲-۳ ضرایب کپسترال حاصل از پیشگویی خطی

بدست آوردن ضرایب کپسترال حاصل از پیشگویی خطی مبتنی بر درک انسان^۶ [۶] یکی از روش های مبتنی بر مدل شنوایی می باشد. ایده اصلی روش PLP این است که قدرت تفکیک فرکانسی و حساسیت نسبت به تغییر فرکانس در گوش انسان در فرکانس های مختلف، یکسان نیست و به علاوه حساسیت گوش نسبت به شدت و انرژی صوت در

⁴ - Convolutional

⁵ - Mean Normalization (MN)

⁶ - Perceptual Linear Prediction (PLP)

¹ - Mel Frequency Cepstral Coefficient

² - Accuracy

³ - Cepstral Mean Subtraction (CMS)

فرکانسهای مختلف متفاوت است. بر این اساس نشان داده شده است که در گوش انسان میزان احساس بلندی با ریشه سوم انرژی آن متناسب است [۶]. از آنجا که این روش هم بر پایه طیف زمان کوتاه گفتار می باشد، مشابه سایر تکنیک های مبتنی بر طیف زمان کوتاه، در مقابل اثرات کانال های مخابراتی آسیب پذیر است که نتایج بکارگیری آنهم این مساله را نشان می دهد. نتایج حاصل از بکارگیری این روش در جدول ۳ آورده شده است. استفاده از این روش منجر به کاهش نتیجه نسبت به حالت مبنا در همه شرایط نویزی شده است. یک راه بهبود کارایی این روش در حضور نویز کانال، اعمال یک فیلتر بالاگذر یا میانگذر است که منجر به روش بعدی می شود.

جدول ۳: نتایج بازشناسی با PLP

SNR ▴ Noise ▾	clean	20	15	10	5	0	-5	Avg
Subway	98.8	97.3	91.7	75.7	48.1	25.6	12.6	64.3
Babble	99.1	89.3	70.7	46.0	20.7	4.7	-0.3	47.2
Car	98.9	95.4	81.2	54.3	25.8	10.3	7.3	53.3
Exhibition	99.3	95.9	87.5	65.1	30.8	10.6	6.7	56.5
Restaurant	98.8	88.5	71.3	47.4	22.3	3.2	-2.0	47.1
Street	99.1	95.4	85.8	62.4	34.0	14.1	7.0	56.8
Airport	98.9	90.0	73.6	48.1	23.3	8.5	2.9	49.3
Train-station	99.3	92.8	77.1	50.9	22.7	8.0	7.0	51.1

۳-۳ تحلیل طیف نسبی

یکی از روش های مقابله با نویز با سرعت کم در کانال های متعدد، RASTA^۱ می باشد. ایده این روش از این واقعیت سرچشمه می گیرد که نویز پیچشی تقریباً حالت ایستان دارد، بنابراین با اعمال یک فیلتر بالاگذر یا میانگذر در روش PLP، می توان تقریباً اثر نویز پیچشی را حذف کرد. نتایج استفاده از این روش، که RASTA-PLP نیز گفته می شود، در جدول ۴ آورده شده است. این روش برای نویزهای مترو، نمایشگاه و ماشین نتایج را نسبت به حالت مبنا پایین آورده است.

برای اینکه بتوان در این روش، نویز جمع شونده را نیز در کنار نویز پیچشی در نظر بگیریم از حالت تعمیم یافته RASTA با نام J-RASTA می توان استفاده کرد [۵]. در روش پایه RASTA، از یک لگاریتم و عکس آن در مراحل پردازش استفاده می شود در حالی که در J-RASTA این مورد به

صورت $y_i = \log(1 + Jx_i)$ اصلاح می شود که در آن x_i سیگنال ورودی است. عکس این تابع به صورت تقریبی به صورت $x_i = \frac{\exp(y_i)}{J}$ در نظر گرفته می شود. از آنجا که عملکرد این روش به شدت به مقدار J وابسته است بایستی برای هر سطح نویز، مقدار بهینه پارامتر J را بدست آورد. اگر چه روش هایی برای یافتن مقدار بهینه J وجود دارد ولی هنوز یک روش کلی برای این کار وجود ندارد و معلوم نیست که این روش ها عملکرد مناسبی را در همه محدوده وسیع تغییرات میزان و نوع نویز داشته باشند [۵]. نتایج اعمال این روش به صورت نشان داده شده در جدول ۵ می باشد که منجر به بهبود خوبی برای برخی نویزها و به طور مشخص برای نویز ماشین شده است، روی نویز همههمه تقریباً بدون تاثیر بوده و نتایج نویزهای نمایشگاه و رستوران را پایین آورده است. مقدار J بکار گرفته شده در اینجا 1.6×10^{-6} بوده است.

جدول ۴: نتایج بازشناسی با RASTA-PLP

SNR ▴ Noise ▾	clean	20	15	10	5	0	-5	Avg
Subway	98.7	95.6	89.8	70.4	39.2	20.0	12.2	60.8
Babble	98.9	96.4	91.9	75.5	41.8	19.5	11.2	62.2
Car	98.8	96.4	89.7	62.7	29.2	17.5	9.7	57.7
Exhibition	99.1	95.7	89.0	70.4	34.3	16.0	8.3	59.0
Restaurant	98.7	96.3	92.1	79.4	49.6	23.6	11.7	64.5
Street	98.9	96.3	89.8	71.1	40.9	19.3	10.6	61.0
Airport	98.8	96.6	92.5	79.1	47.7	24.5	13.4	64.6
Train-station	99.1	96.9	90.9	72.4	39.6	20.2	10.9	61.4

جدول ۵: نتایج بازشناسی با J-RASTA

SNR ▴ Noise ▾	clean	20	15	10	5	0	-5	Avg
Subway	98.6	96.1	90.4	77.5	60.2	37.6	16.2	68.1
Babble	98.8	86.4	71.1	52.3	31.1	9.4	-6.8	48.9
Car	98.8	97.3	94.7	86.9	70.0	41.6	15.0	72.0
Exhibition	99.1	90.0	79.9	63.4	40.4	15.7	0.6	55.6
Restaurant	98.6	82.2	66.2	49.7	28.5	10.0	-5.9	47.0
Street	98.8	94.1	88.0	74.6	55.2	30.4	12.7	64.8
Airport	98.8	89.4	80.0	63.1	43.2	22.6	4.2	57.3
Train-station	99.1	93.2	88.5	78.4	58.3	33.8	12.3	66.2

۳-۴ ضرایب کپسترال ریشه ای^۲

یکی دیگر از راه حل هایی که برای استفاده از ضرایب کپسترال در محیط های نویزی به کار گرفته شده، استفاده از برخی توابع دیگر بجای لگاریتم در محاسبه این ضرایب با روش MFCC

^۲ - Root Cepstral Coefficient (RCC)

^۱ - Relative SpecTrAl

۳-۵ تحلیل اجزای اصلی

تحلیل اجزای اصلی یا PCA^1 ، یک روش شناخته شده آماری در کاهش تعداد ویژگی‌ها و کلاستر کردن 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 نماید. این تبدیل به صورت رابطه (۳) روی بردار N بعدی X اعمال می‌شود:

$$Y = T.(X - \mu_X) \quad (3)$$

که در آن T ماتریس تبدیل مورد نظر است که از N بردار ویژه $\{\Phi_i, i=1,2,\dots,N\}$ مربوط به ماتریس کواریانس داده‌های آموزشی با مقادیر ویژه $\{\lambda_i, i=1,2,\dots,N\}$ به صورت زیر بدست آمده است:

$$T = [\Phi_1 \ \Phi_2 \ \dots \ \Phi_N]^T, \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N \quad (4)$$

خاصیت اصلی این تبدیل متعامد کردن بردارهای نگاشت شده Y توسط تبدیل است. از توانایی‌های دیگر این تبدیل کاهش بعد و فشرده سازی است. از آنجا که سطرهای ماتریس تبدیل T به ترتیب میزان واریانس داده‌ها قرار گرفته‌اند، می‌توان به جای استفاده از کل ماتریس $N \times N$ از نمونه کاهش یافته و ماتریسی $R \times N$ ($R \leq N$) که فقط از R بردار اول و با اهمیت‌تر تشکیل شده است، استفاده شود.

استفاده از این تبدیل در استخراج ویژگی‌های مبتنی بر MFCC می‌تواند به سه صورت: به عنوان آخرین بلوک مرحله استخراج ویژگی‌ها، بعد از تبدیل گسسته کسینوسی و به جای تبدیل گسسته کسینوسی باشد [۷، ۸] که نتایج بکارگیری آن در این قسمت‌ها، نشان دهنده بهتر بودن استفاده از این روش به عنوان آخرین بلوک مراحل استخراج ویژگی است. از اینرو در این آزمایش‌ها این تبدیل را بعد از محاسبه مشتق(های) ضرایب کپسترال اعمال می‌کنیم. در آزمایش‌های انجام شده که از

می‌باشد. یکی از این توابع که در اینجا استفاده می‌شود، تابع ریشه می‌باشد. نحوه استخراج این ضرایب به صورت نشان داده شده در رابطه (۲) می‌باشد که در آن به جای گرفتن لگاریتم از انرژی فیلتر بانک‌ها $E[j]$ ، ریشه $\sqrt{E[j]}$ آن محاسبه می‌شود. سپس مشابه MFCC تبدیل گسسته کسینوسی اعمال شده است. در این رابطه M تعداد فیلتر بانک‌ها و N تعداد ضرایب می‌باشد.

$$RCC[i] = \sum_{j=1}^M \left[(E[j])^r \cos\left(\frac{i(j-0.5)}{M} \pi\right) \right], i=1,2,\dots,N \quad (2)$$

انجام این کار در کنار بهبود کارایی سیستم، منجر بهبود سرعت استخراج ویژگی‌ها می‌شود. نتایج بازنمایی حاصل با استفاده از این روش در جدول ۶ آورده شده است که در صورت اعمال تفاضل میانگین روی این ضرایب، نتایج به صورت جدول ۷ در می‌آیند. اعمال RCC به تنهایی، مشابه حالت قبل، برای نویزهای مترو و نمایشگاه منجر به کاهش نتیجه شده است و برای سایر موارد هم بهبود نسبتاً اندکی را به همراه داشته است ولی وقتی با روش تفاضل میانگین ضرایب ترکیب شود منجر به بهبود قابل توجهی در همه شرایط نویزی شده است.

جدول ۶: نتایج بازنمایی با RCC

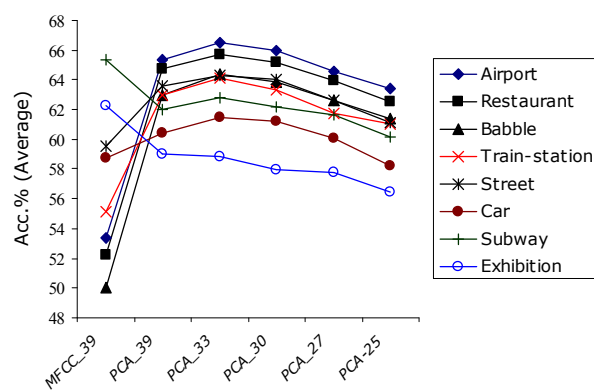
SNR ▸ Noise ▼	clean	20	15	10	5	0	-5	Avg
Subway	99.1	94.2	81.8	56.7	32.4	16.4	9.6	55.7
Babble	99.0	95.2	84.0	60.0	29.2	6.7	-0.9	53.3
Car	98.8	97.8	94.0	77.1	41.3	14.1	5.3	61.2
Exhibition	99.1	94.9	87.8	67.1	34.2	13.5	7.6	57.7
Restaurant	99.1	93.8	82.3	59.6	29.0	6.4	-1.1	52.7
Street	99.0	96.2	90.1	71.8	42.7	20.0	9.6	61.3
Airport	98.8	94.8	85.1	63.6	32.6	9.5	-1.8	54.6
Train-station	99.1	95.1	87.5	70.0	36.3	10.3	0.9	57.0

جدول ۷: نتایج بازنمایی با RCC_MN

SNR ▸ Noise ▼	clean	20	15	10	5	0	-5	Avg
Subway	99.1	97.0	93.8	83.1	61.3	30.2	12.8	68.2
Babble	99.0	97.8	92.8	81.4	56.1	26.0	10.9	66.3
Car	99.0	98.0	95.5	86.9	64.1	26.8	10.5	68.7
Exhibition	99.2	95.7	90.9	78.4	51.0	19.8	9.4	63.5
Restaurant	99.1	97.3	91.8	79.8	57.2	28.3	11.6	66.5
Street	99.0	97.1	93.8	84.0	61.6	30.1	11.9	68.2
Airport	99.0	98.0	95.1	86.6	64.9	33.9	13.5	70.1
Train-station	99.2	97.4	94.0	84.9	62.9	29.4	11.7	68.5

¹ - Principal Component Analysis

تفاضل میانگین ضرایب هم به همراه PCA استفاده شده است، با تعداد مختلف ویژگی‌ها کار بازشناسی انجام شده است. نتایج کار برای نویزهای مختلف به صورت میانگین دقت روی نسبت سیگنال به نویزهای مختلف در شکل ۱ به همراه نتیجه مبنای آنها آورده شده است. این نتایج نشان می‌دهند که تعداد بهینه در اینجا ۳۳ ویژگی است. این کاهش تعداد علاوه بر بهبود دقت، بعثت کاهش محاسبات مراحل بعد، باعث افزایش سرعت سیستم نیز می‌شود. نتایج کامل اعمال این تبدیل در حالت استفاده از ۳۳ ویژگی در جدول ۸ آورده شده است. مشابه اغلب روشهای مورد بررسی، در اینجا نیز، برای نویزهای مترو و نمایشگاه نتایج کاهش یافته است ولی در سایر موارد بهبودهای قابل توجهی بدست آمده است.



شکل ۱: میانگین دقت سیستم در روش MFCC-PCA-MN برای تعداد مختلف ویژگی‌ها و مقایسه آنها با روش مبنای MFCC

جدول ۸: نتایج بازشناسی با MFCC-PCA-MN با ۳۳ ویژگی

SNR ▴ Noise ▾	clean	20	15	10	5	0	-5	Avg
Subway	98.8	95.8	90.1	74.1	45.9	21.7	13.3	62.8
Babble	98.8	97.1	91.6	79.3	50.5	21.8	11.8	64.4
Car	98.6	96.4	91.6	73.6	41.2	18.6	10.5	61.5
Exhibition	98.8	94.0	85.3	68.4	38.8	17.0	9.3	58.8
Restaurant	98.8	96.4	92.4	81.2	54.9	24.8	11.3	65.7
Street	98.8	96.7	90.9	77.3	51.4	22.9	11.9	64.3
Airport	98.6	97.1	93.3	81.9	54.9	26.9	12.9	66.5
Train-station	98.8	96.8	92.2	78.7	48.9	22.1	11.5	64.2

۴. روشهای استخراج ویژگی و تغییر کانال

نتایج ارایه شده در فوق برای ترکیب نویز کانال G.712 و نویزهای جمع‌شونده در نسبت سیگنال به نویزهای مختلف بود. برای بررسی کارایی این روشها در شرایطی که کانال انتقال

صحت تغییر می‌کند از مجموعه آزمایش C که فقط دو نوع نویز مترو و خیابان را دارد و کانال شبیه سازی آن MIRS می‌باشد، استفاده شده است. نتایج بازشناسی روی این مجموعه برای دو نوع نویز مترو و خیابان به صورت نشان داده شده در جدول‌های ۹ و ۱۰ می‌باشد. روشهای J-RASTA و -RCC MN منجر به بهبود نتایج در هر دو حالت نویزی شده‌اند. به علاوه روش MFCC-MN در نویز خیابان منجر به بالا بردن دقت بازشناسی شده است. در سایر موارد اعمال روشها باعث کاهش نتیجه شده است. استفاده از روش تفاضل میانگین ضرایب چه به تنهایی و چه در کنار سایر روشها بهبود قابل ملاحظه‌ای را به دنبال داشته است که می‌توان این مساله را در تفاوت نتایج بکارگیری RCC و RCC-MN مشاهده کرد.

جدول ۹: نتایج بازشناسی با روشهای مختلف استخراج ویژگی برای نویز

مترو در کانال MIRS

SNR ▴ Method ▾	clean	20	15	10	5	0	-5	Avg
MFCC-39	99.1	93.5	86.8	73.9	51.3	25.4	11.8	63.1
RCC	99.1	91.8	79.8	54.1	26.1	11.0	8.1	52.9
Rasta-PLP	98.6	95.6	90.0	72.3	40.2	19.8	12.0	61.2
PCA-MN-33	98.8	95.6	88.9	71.0	42.4	20.7	11.7	61.3
PLP	98.8	93.8	87.6	74.4	50.3	20.5	9.1	62.1
MFCC-MN	99.0	95.9	89.6	72.9	44.5	22.3	12.2	62.3
RCC-MN	99.1	96.1	91.5	78.7	55.6	27.2	11.6	65.7
J-Rasta	98.7	97.1	93.2	84.3	65.8	39.2	18.3	70.9

جدول ۱۰: نتایج بازشناسی با روشهای مختلف استخراج ویژگی برای نویز

خیابان در کانال MIRS

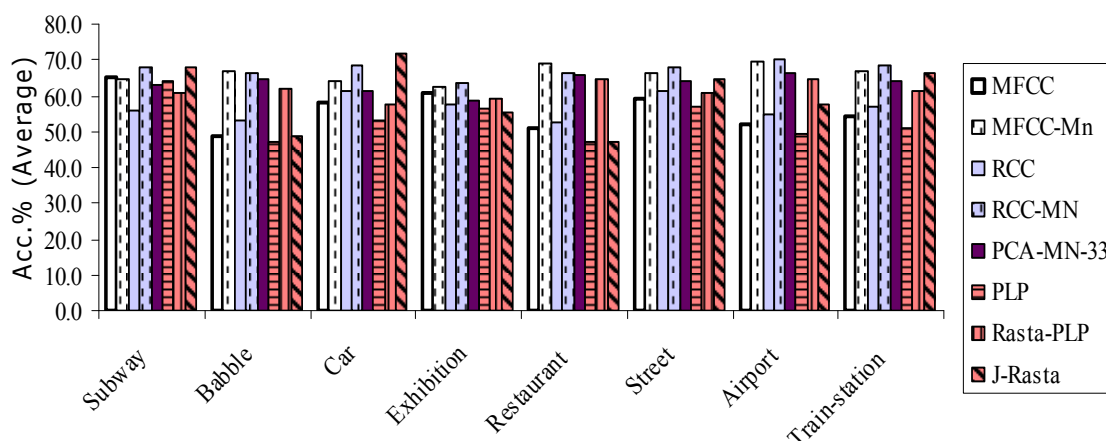
SNR ▴ Method ▾	clean	20	15	10	5	0	-5	Avg
MFCC-39	99.0	95.1	88.9	74.4	49.2	22.9	11.2	63.0
RCC	99.0	94.0	87.0	72.4	49.4	24.0	11.5	62.5
Rasta-PLP	98.9	95.7	90.6	71.7	39.3	20.1	10.6	61.0
PCA-MN-33	98.9	95.8	90.5	76.4	48.6	22.4	11.8	63.5
PLP	99.2	94.7	88.8	74.2	51.5	23.9	10.3	63.2
MFCC-MN	99.2	97.0	92.6	78.0	51.6	24.5	11.8	64.9
RCC-MN	99.0	96.6	92.3	80.2	57.4	28.2	13.4	66.7
J-Rasta	98.6	94.4	89.1	76.9	58.2	33.0	15.5	66.5

۵. مقایسه، نتیجه گیری و جمع بندی

مساله مقاوم سازی سیستم‌های بازشناسی گفتار در این مقاله با ارزیابی روشهای مختلف استخراج ویژگی در شرایط مختلف نویزی مورد بررسی قرار داده شد. استفاده از تفاضل میانگین ضرایب کلاً باعث بهبود کارایی سیستم بازشناسی در شرایط نویزی موجود شده است. سایر روشها هم در برخی شرایط

نویزی نسبت به برخی شرایط دیگر بهتر جواب داده اند. برای نویزهای مترو و نمایشگاه، استفاده از روشهای مورد بررسی بجز در یکی-دو مورد آنهم بصورت جزئی، نتایج بازشناسی را پایین می آورد که علت این مساله می تواند دارا بودن انرژی نسبتاً یکسان این دو نوع نویز در فرکانس های مختلف باشد در

حالی که انرژی سایر نویزها بیشتر در فرکانس های پایین می باشد. مقایسه میانگین نرخ دقت بازشناسی روی نسبت سیگنال به نویزهای مختلف در شرایط نویزی مختلف، برای روشهای مورد بررسی در این مقاله بر روی مجموعه های آزمایش A و B به صورت نشان داده در شکل ۲ می باشد.

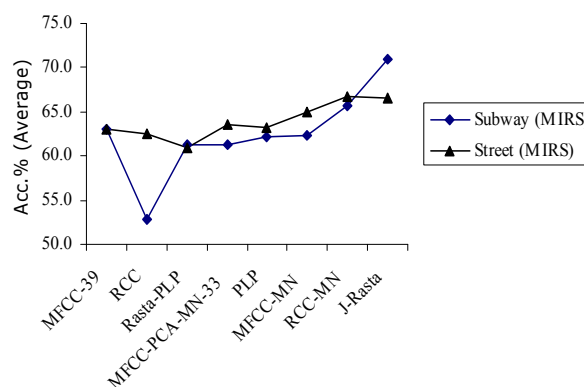


شکل ۲: مقایسه میانگین دقت بازشناسی (در نسبت سیگنال به نویزهای مختلف) روشهای مختلف استخراج ویژگی برای شرایط نویزی متفاوت

منابع

- [1]. M. J. F. Gales, *Model-based techniques for noise robust speech recognition*, Ph.D Thesis, Univ. of Cambridge, 1995.
- [2]. H. G. Hirsch, D. Pearce, "The AURORA Experimental Framework for the Performance Evaluations of Speech Recognition Systems under Noisy Conditions", ISCA ITRW ASR2000, Paris, Sep. 2000.
- [3]. S. Young, G. Everman, D. Kershaw, G Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, P. Woodland, *The HTK Book (for HTK Version 3.2)*, Cambridge University, 2002.
- [4]. A. Garcia and R. Mammone, "Channel-Robust Speaker Identification Using Modified-Mean Cepstral Mean Normalization with Frequency Warping", Proc. ICASSP, pp.325-328, 1999.
- [5]. J. Koehler and et al, "Integrated RASTA-PLP into speech recognition", ICASSP'94, Vol. 1, PP. 421-424, 1994
- [6]. H. Hermansky, "Perceptual Linear Predictive Analysis of Speech", Journal of Acoustical Society of America, pp. 1738-1752, April 1990.
- [7]. R. Hariharan, I. Kiss, O.Viikki, "Noise Robust Speech Parameterization Using Multiresolution Feature Extraction", IEEE Transaction ON Speech and Audio Processing, VOL. 9, NO. 8, Nov. 2001.
- [8]. C. A. Ynoguti, F. Violaro, "On The Use Of Principal Component Analysis Over Mel Cepstral Coefficients", Telecomunicações, Revista do Instituto Nacional de Telecomunicações, ISSN 1516-2338, Vol. 05, no 2, pp 13-17, Dec. 2002.

بصورت مشابه، مقایسه کارایی این روشها در شرایط تغییر کانال، با آزمایش روی مجموعه آزمایش C در شکل ۳ آورده شده است. نتایج این مقایسه بیانگر موفق بودن عملکرد روش J-RASTA به نسبت سایر روشهاست. این مساله بویژه برای تغییر کانال در حضور نویز مترو و در نسبت سیگنال به نویزهای پایین مشخص تر است. به علاوه روش تفاضل میانگین ضرایب و بکارگیری آن با روشهای دیگر مانند RCC و MFCC، همانطور که انتظار آن نیز وجود دارد، در حالت کلی منجر به بهبود دقت شده است.



شکل ۳: مقایسه روشهای مختلف استخراج ویژگی با تغییر کانال