

استخراج ویژگی ترکیبی مقاوم در مقابل نویز برای سیستم‌های بازشناسی گوینده

محمد احدی
عضو هیأت علمی دانشکده مهندسی برق،
دانشگاه صنعتی امیرکبیر
تهران، ایران
sma@aut.ac.ir

امیرحسین حاج‌احمدی
دانشجوی کارشناسی ارشد، دانشکده مهندسی
کامپیوتر و فن‌آوری اطلاعات، دانشگاه صنعتی
امیرکبیر
تهران، ایران
hadjahmadi@aut.ac.ir

محمد مهدی همایون پور
عضو هیأت علمی دانشکده مهندسی کامپیوتر
و فن‌آوری اطلاعات، دانشگاه صنعتی امیرکبیر
تهران، ایران
homayoun@aut.ac.ir

سیستم‌های خودکار بازشناسی گفتار و گوینده در مقابل نویزها را می‌توان واسطه بین بازشناسی خودکار و کاربرد واقعی آن دانست [1]. علی‌رغم بهبود بسیار ایجاد شده در کارایی سیستم‌های خودکار بازشناسی این موضوع همچنان یکی از زمینه‌های فعال تحقیقاتی محسوب می‌شود که همچنان تلاش‌های بسیاری در این زمینه انجام می‌گیرد. می‌توان انگیزه اصلی از این تحقیق را چنین بیان کرد که "بازشناسی مقاوم واسطه بین بازشناسی و کاربرد واقعی آن است". تا زمانی که بازشناسی مقاوم به میزان کافی توسعه پیدا نکرده باشد، سیستم‌های بازشناسی در کاربردهای واقعی به کارایی مطلوب نخواهند رسید [1].

مجموعه تلاش‌های انجام شده برای مقاوم‌سازی سیستم‌های بازشناسی گفتار و گوینده را می‌توان به ۳ دسته تقسیم کرد. در دسته اول که استخراج ویژگی‌های مقاوم^۲ نامیده می‌شوند، سعی شده است تا مستقیماً از سیگنال گفتار ویژگی‌هایی مقاوم در مقابل نویزها استخراج شوند. در دسته دوم که نرمالسازی ویژگی‌ها^۳ نامیده می‌شوند، سعی می‌شود، تأثیر نویز تا حد امکان از ویژگی‌های استخراج شده حذف گردد. در دسته‌ای دیگر که تطبیق مدل^۴ نامیده می‌شوند، مقاوم‌سازی در مرحله ساخت مدلها صورت می‌گیرد [2].

تکنیک‌های مختلف استخراج ویژگی‌ها همواره مورد مطالعه بوده‌اند و بهبودهای بسیاری نیز پیدا کرده‌اند. استخراج ویژگی‌های مقاوم در مقابل نویز بنیادی‌ترین مسئله حل نشده در بازشناسی مقاوم گفتار و گوینده محسوب می‌شود [3]. سه معیار خوب بودن برای ویژگی‌های مقاوم عبارتند از [4]:

- میانگین^۵ ویژگی‌ها تحت تأثیر نویز تغییر نکند.
- واریانس مربوط به فضای ویژگی تحت تأثیر نویز افزایش نیابد.
- ویژگی‌ها از خاصیت جدایی‌پذیری برای الگوهای مختلف برخوردار باشند.

با کمک گرفتن از شیوه ادراک گوش ضرائب مل-کپسترال^۶ (MFCC) معرفی شدند و امروزه به عنوان روشی استاندارد برای استخراج

چکیده: با توجه به اهمیت مقاوم‌سازی سیستم‌های بازشناسی گوینده در مقابل انواع نویزهای محیطی در امکان استفاده عملی از آنها، در این مقاله روش استخراج پیک‌های طیف دنباله خودهمبستگی و نیز روش پس‌پردازشی MVA که هر دو از روش‌های مقاوم‌سازی سیستم‌های بازشناسی گفتار در مقابل نویز هستند، برای استفاده در سیستم‌های تعیین هویت گوینده با ویژگی‌های رایج MFCC مقایسه شده‌اند. نتایج نشان‌دهنده مقاومت بیشتر پیک‌های طیف دنباله خودهمبستگی در مقابل نویز همهمه و مقاومت بیشتر ضرائب MFCC نرمال شده با پس‌پردازش MVA در مقابل نویزهای کانولوشنی است. همچنین نتایج نشان می‌دهند که هر دو این روش‌ها علی‌رغم افزایش دقت بازشناسی در مقابل نویزها دقت بازشناسی در محیط تمیز را کاهش می‌دهند. به همین دلیل در این مقاله یک نوع روش استخراج ویژگی که از الحاق ویژگی‌های مذکور استفاده می‌کند، نیز پیشنهاد شده است که به میزان قابل توجهی می‌تواند بر مقاومت سیستم‌های تعیین هویت گوینده در مقابل انواع نویزهای محیطی بیفزاید، بدون آنکه کارایی آن در محیط‌های تمیز با افت قابل توجهی مواجه شود.

واژه‌های کلیدی: تعیین هویت گوینده، استخراج ویژگی، مقاوم‌سازی در مقابل نویز، ترکیب ویژگی‌ها و پیک‌های طیف خودهمبستگی.

۱- مقدمه

اولین سیستم بازشناسی گوینده در دهه ۱۹۶۰ در حدود ۱۰ سال پس از ایجاد اولین سیستم بازشناسی گفتار تولید شد. تحقیقات در زمینه مقاوم نمودن این نوع سیستمها نیز در دهه ۱۹۹۰ به اوج خود رسید و همچنان ادامه پیدا کرده است [1].

دقت تمامی سیستم‌های خودکار بازشناسی گوینده و گفتار^۱ در هنگام استفاده در محیط نویزی به شدت کاهش پیدا می‌کند. دلیل این افت دقت تفاوت بین داده‌های آموزش و داده‌های تست این سیستمها است. با توجه به وجود نویز در اکثر محیط‌های واقعی، مقاوم‌سازی^۲

گفتار از کارایی بالایی برخوردار هستند، به ترتیب در بخش‌های ۲، ۳ و ۴ بیان شده است. در بخش ۵ روش ترکیبی پیشنهادی تشریح شده است. سپس در بخش ۶ آزمایش‌های انجام شده جهت مقایسه مقاومت سیستم‌های تعیین هویت گوینده با استفاده از ویژگی‌های گفتاری مذکور تشریح می‌شوند. در انتها در بخش ۷ نتیجه‌گیری نهایی بیان شده است.

۲- ضرایب کپسترال مبتنی بر معیار مل (MFCC)

ایده اصلی در ضرایب کپستروم مبتنی بر معیار مل (MFCC) برگرفته از خواص گوش انسان در دریافت و فهم گفتار می‌باشد. عملکرد گوش انسان به گونه‌ای است که فرکانس ادراکی آن با فرکانس واقعی صدا متفاوت است [6].

ضرایب کپسترال مبتنی بر معیار مل در یک فریم زمانی داده شده بصورت بردار $C = (C[1] \dots C[D])^T$ نمایش داده می‌شوند که D در آن تعداد ضرایب کپسترال است و به صورت زیر محاسبه می‌گردند:

$$C = G \log_e Q \quad (1)$$

که بردار $Q = (Q[1] \dots Q[J])^T$ طیف قدرت فرکانس-مل است و برای بدست آوردن آن بایستی طیف فوریه فریم را بدست آورده از دامنه آن لگاریتم گرفت، سپس یک بانک فیلتر که توزیع فیلترهای آن بر پایه معیار مل است روی طیف اعمال و خروجی تک-تک فیلترها محاسبه شود. آنگاه این خروجی‌ها بردار Q را تشکیل می‌دهند. G ماتریسی با ابعاد $i \times j$ است که نمایش دهنده تبدیل کسینوسی گسسته است.

$$G_{ij} = \sqrt{\frac{2}{J}} \cos\left(\frac{pi}{J}(j - 0.5)\right), i = 1 \dots I, j = 1 \dots J \quad (2)$$

مراحل محاسبه ضرایب MFCC در شکل ۱ نشان داده شده است. در [5] بیان شده است که چون برای استخراج ویژگی‌های MFCC از یک بانک فیلتر مبتنی بر معیار مل استفاده می‌شود و خروجی‌های فیلترهای این بانک فیلتر به صورت غیریکنواخت تحت تأثیر نویزها قرار می‌گیرند، از مقاومت بالایی در مقابل نویزها برخوردار نیستند.

۳- استخراج پیک‌های طیف خود همبستگی

نحوه کامل استخراج پیک‌های طیف سیگنال گفتار در حوزه خودهمبستگی که به اختصار DAS^{۱۲} نامیده می‌شوند، در شکل ۲ نشان داده شده است. ساختار این روش شباهت بسیاری با روش استخراج ویژگی‌های MFCC دارد. با این تغییر که از دنباله خودهمبستگی بجای خود سیگنال استفاده می‌شود.

در این روش ابتدا دنباله خودهمبستگی سیگنال گفتار (\cdot) با استفاده از رابطه (۳) محاسبه می‌شود.

ویژگی‌های گفتاری شناخته می‌شوند. ضرایب MFCC در کاربردهای بسیاری نتایج بسیار خوبی را تولید کرده‌اند. ولی این ضرایب علی‌رغم کارایی خوب در سیستم‌های بازشناسی گوینده، به عنوان ضرایب مقاوم شناخته نمی‌شوند. در ضمن این ضرائب استاندارد سیگنال گفتار، تنها طیف سیگنال زمانی زمان کوتاه را مورد بررسی و تحلیل قرار می‌دهد و بسیاری از اطلاعات پویای^۸ سیگنال گفتار را در نظر نمی‌گیرند که این کمبود با اضافه کردن مشتقات زمانی ویژگی‌ها مانند ویژگی‌های دلتا و دلتا-دلتا رفع می‌شود. امروزه در بسیاری از سیستم‌های بازشناسی گفتار و گوینده از ویژگی‌های MFCC استفاده می‌شود.

در [5] بیان شده است که چون برای استخراج ویژگی‌های MFCC از یک بانک فیلتر مبتنی بر معیار مل استفاده می‌شود و خروجی‌های فیلترهای این بانک فیلتر به صورت غیریکنواخت تحت تأثیر نویزها قرار می‌گیرند، از مقاومت بالایی در مقابل نویزها برخوردار نیستند.

در اوائل دهه ۱۹۹۰ روشی برای مقاوم‌سازی ویژگی‌های MFCC معرفی شد، که به نرمالسازی میانگین کپسترال^۹ (CMN) شهرت یافت. در این روش برای یک مجموعه از بردارهای ویژگی کپسترال استخراج شده بردار میانگین محاسبه و از تک تک آنها کسر می‌گردد [5]. با حذف این بردار میانگین از تمام بردارهای ویژگی به ویژگی‌های مقاومتری دست یافته می‌شود. نوع تکامل یافته روش نرمالسازی میانگین کپسترال روش پس پردازش MVA^{۱۰} است که در بخش ۴ تشریح شده است.

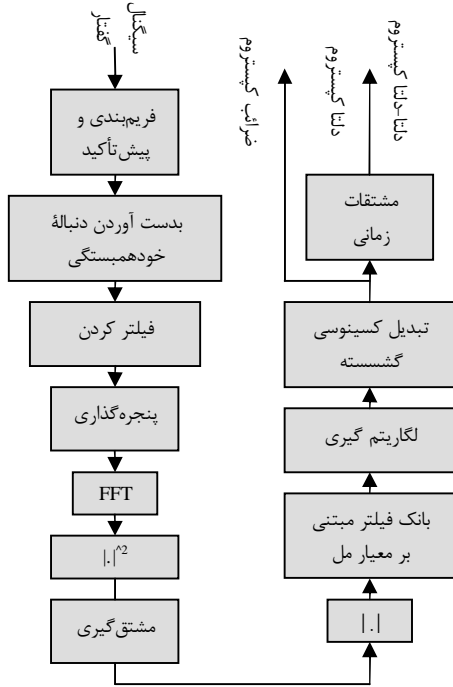
یکی از ویژگی‌های جدیدی که توانسته است، بر مقاومت سیستم‌های بازشناسی گفتار به مقدار قابل توجهی بیفزاید، پیک‌های سیگنال گفتار در حوزه خود همبستگی است که به صورت مختصر DAS^{۱۱} نامیده می‌شود. دلیل مقاوم بودن این ویژگی‌ها را می‌توان در تغییرات ناچیز خودهمبستگی سیگنال نویزی جستجو کرد [2]. در بخش ۲ نحوه استخراج این ویژگی‌ها از سیگنال گفتار بیان شده است.

نتایج آزمایش‌های انجام شده در [۱] و همین مقاله نشان می‌دهند که هرچند این دو روش در مقاوم‌سازی سیستم‌های تعیین هویت گوینده در مقابل نویزها از قدرت بالایی برخوردارند ولی در صورت استفاده در محیط‌های تمیز موجب کاهش دقت سیستم‌ها می‌شوند که این هم نمی‌تواند مطلوب باشد.

به همین دلیل در این مقاله روشی جدید معرفی شده است که با استفاده از ترکیب سه نوع ویژگی یعنی ضرائب کپسترال مبتنی بر معیار مل، ضرائب کپسترال مبتنی بر معیار مل نرمال شده با استفاده از روش پس پردازشی MVA و پیک‌های طیف دنباله خودهمبستگی سیگنال گفتار قادر است به مقدار قابل توجهی بر مقاومت سیستم‌های تعیین هویت گوینده در مقابل نویزهای مختلف بیفزاید.

در این مقاله ابتدا نحوه استخراج ویژگی‌های MFCC، روش جدید استخراج پیک‌های طیف دنباله خودهمبستگی سیگنال گفتار^{۱۲} و روش پس پردازشی MVA که در مقاوم‌سازی سیستم‌های بازشناسی خود

نتایج بیان شده در [2] گویای این مطلب است که این روش دقت بازشناسی گفتار را برای انواع نویز به طور میانگین حدود ۱۰٪ بهبود داده است. همچنین نتایج آزمایش‌هایی که در شکل ۴ نشان داده شده‌اند، بیانگر این مطلب هستند که با استفاده از پیک‌های طیف دنباله خودهمبستگی می‌توان بر مقاومت سیستم‌های تعیین هویت گوینده در مقابل نویزها افزود. ولی اگر به تنهایی استفاده شوند از دقت سیستم‌ها در محیط‌های بدون نویز می‌کاهند.



شکل (۲): دیاگرام بلوکی روش استخراج پیکهای گفتار در حوزه خودهمبستگی [2].

۴- تکنیک پس پردازشی MVA برای نرمال سازی

ویژگی‌های کپسترال

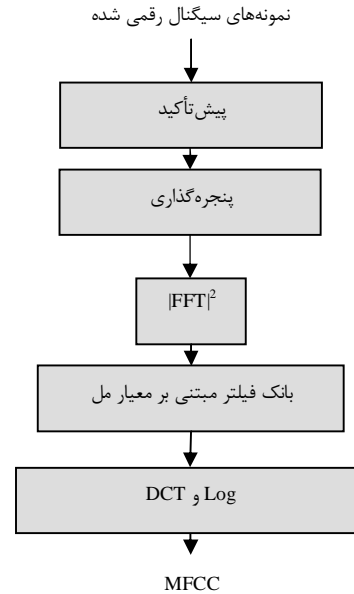
این روش پس پردازشی ترکیبی است از تفاضل میانگین^{۱۴}، نرمالیزه کردن واریانس^{۱۵} و فیلتر کردن ARMA^{۱۶} در دامنه کپسترال [7]. اگر $V^{(t)}$ بردار ویژگی فریم t باشد، آنگاه تفاضل میانگین به صورت زیر تعریف می‌شود:

$$\bar{V}^{(t)} = V^{(t)} - m \quad (6)$$

که بردار ویژگی تفاضل میانگین گرفته شده و μ در رابطه بالا بردار میانگین است. نرمالیزه کردن واریانس (VN) نیز به صورت زیر تعریف می‌شود:

$$r_{yy}(m, k) = \frac{1}{N-K} \sum_{i=0}^{N-1-K} y(m, i)y(m, i+k) \quad (3)$$

که در آن $y(m, i)$ نشان دهنده نمونه m ام از فریم m ام سیگنال گفتار ورودی، $r_{yy}(m, k)$ نمونه k ام از فریم m ام دنباله خودهمبستگی و N طول فریم است.



شکل (۱): دیاگرام بلوکی روش استخراج ضرائب کپسترال مبتنی بر معیار مل (MFCC) [6].

با اعمال یک فیلتر بالاگذر به مشتق دنباله خودهمبستگی بصورت زیر می‌توان به داده‌ای رسید که به مقدار زیادی تأثیر نویزها از آن حذف شده است.

$$\frac{\partial r_{yy}(m, k)}{\partial m} = \frac{\sum_{t=-L}^L tr_{yy}(m+t, k)}{\sum_{t=-L}^L t^2} \quad (4)$$

$$0 \leq m \leq M-1 \quad 0 \leq k \leq N-1$$

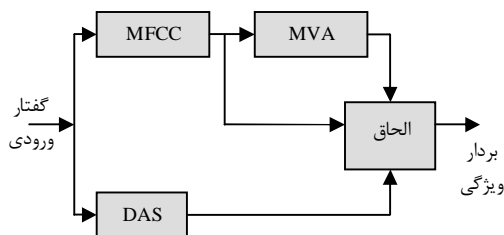
پس از محاسبه FFT دنباله حاصل و بدست آوردن اندازه آن، از دنباله حاصل با رابطه زیر مشتق گیری می‌شود.

$$Diff_Y(k) \approx \sum_{l=-Q}^P a_l Y(k+l), \quad 0 \leq k \leq K-1 \quad (5)$$

که در آن P و Q مرتبه معادله تفاضلی، a_l ها مقادیری حقیقی به عنوان وزن ضرائب و K طول تبدیل فوریه هستند. دنباله حاصل در انتها به بانک مبتنی بر معیار مل داده می‌شود و از خروجی‌های آن لگاریتم و تبدیل کسینوس گرفته می‌شود و ویژگی‌های نهایی بدست می‌آیند [2].

ساخت که هم در محیط‌های تمیز از قدرت بازشناسی بالایی برخوردار باشد، هم در مقابل نویزهای کانال و نویزهای جمع‌شوند، از مقاومت بالایی برخوردار باشد.

در شکل ۳ کلیت روش ترکیبی جدید استخراج ویژگی مقاوم برای مقاوم‌سازی سیستم‌های تعیین هویت گوینده نشان داده شده است. در این روش در ابتدا از سیگنال گفتار ورودی دو نوع ویژگی MFCC و DAS استخراج می‌شوند. سپس روی ویژگی‌های MFCC پس‌پردازش MVA اعمال می‌شود. در نهایت بردار ویژگی جدیدی شامل تعدادی از ویژگی‌های MFCC، DAS و MFCC نرمال شده با MVA تولید می‌گردد.



شکل (۳): نحوه استخراج ویژگی‌های ترکیبی پیشنهادی

۶- تست و آزمایش

چندین آزمایش برای مقایسه دقت و مقاومت ویژگی‌های MFCC، MFCC نرمال شده با تکنیک MVA، پیک‌های طیف دنباله خودهمبستگی و ویژگی‌های ترکیبی پیشنهادی با یکدیگر انجام شده است. در تمامی آزمایش‌ها از GMMهایی با ۱۲۸ مخلوط گوسی به عنوان مدل‌های بازشناسی گوینده، از ۳۵ سیگنال‌های گفتار ۴۰ ثانیه‌ای با ۳۵ گوینده برای آموزش مدل‌ها و از ۲۸ سیگنال گفتار ۳ ثانیه‌ای با ۳۵ گوینده که جمعاً ۹۸۰ سیگنال می‌شود، برای تست استفاده شده است. تمامی دادگان گفتاری مورد استفاده در این آزمایش‌ها از دادگان گفتار فارسی‌تلفنی استخراج شده‌اند. در آزمایشات انجام شده، جهت سنجش مقاومت سیستم‌ها نویزهای مهمه^{۱۷} و نویز ماشین^{۱۸} از دادگان نویز NoiseX و نویز سفید گوسی^{۱۹} که نوعی نویز کانال (کانولوشنی) است، به سیگنال‌های گفتار تست در SNRهای مختلف اضافه شده است. نویز مهمه در دادگان NoiseX از صحبت کردن همزمان ۱۰۰ نفر در یک فروشگاه، و نویز ماشین از صدای یک خودروی ولو ۳۴۰ در حال حرکت با سرعت ۱۲۰ کیلومتر بر ساعت بدست آمده‌اند [9].

در تمامی آزمایش‌های انجام شده مرتبه فیلتر ARMA در رابطه (۸) از مرتبه ۳ و مرتبه فیلتر رابطه (۴) نیز ۲ در نظر گرفته شده است. همچنین برای تشکیل بردار ویژگی ترکیبی پیشنهادی از ۱۳ ویژگی MFCC، ۱۳ ویژگی MFCC نرمال شده و ۱۳ ویژگی DAS استفاده شده است.

$$\hat{V}^{(t)}[d] = (s^2[2])^{-\binom{t}{2}} \bar{V}^{(t)}[d] \quad (7)$$

که در آن بردار ویژگی تفاضل گرفته شده میانگین و نرمالیزه شده واریانس است و $s^2[d]$ واریانس تخمین‌زده شده برای d امین عنصر از ویژگیها است.

فیلتر ARMA نیز به صورت زیر تعریف می‌شود:

$$\hat{V}^{(t)} = \frac{\hat{V}^{(t-m)} + \dots + \hat{V}^{(t-1)} + \hat{V}^{(t)} + \dots + \hat{V}^{(t+m)}}{2m+1} \quad (8)$$

که بردار ویژگی نرمال شده پس از عمل پس‌پردازش MVA است و m مرتبه فیلتر ARMA است.

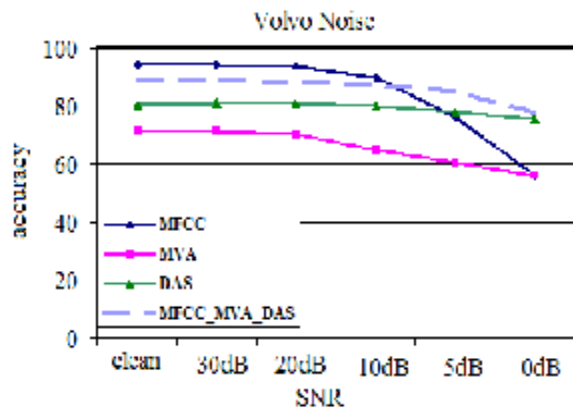
با استفاده از تفریق میانگین می‌توان تأثیر نویزهای کانولوشنی را تا حد قابل قبولی کاهش داد. [8] اگر نویز جمعی دارای سطح نویز کمی باشد، آنگاه با استفاده از لگاریتم انرژی می‌توان به پایداری بیشتری دست یافت ولی اگر سطح نویز جمعی زیاد باشد این تکنیک از قدرت چندانی برخوردار نخواهد بود. بعد از انجام این کار با نرمالیزه کردن واریانس ضرائب می‌توان به ویژگی‌های قابل قبولی دست یافت. نقش فیلتر ARMA بیشتر در هموار کردن تغییرات و نوسانات سریع سیگنال تولید شده است [7].

نتایج آزمایش‌های نشان داده شده در شکل ۴ بیان‌گر توان بالایی پس‌پردازش MVA در حذف تأثیرات ناشی از نویزهای کانولوشنی (نویز کانال) مانند نویز سفید گوسی هستند. روش پس‌پردازش MVA نیز مانند استخراج پیک‌های طیف دنباله خودهمبستگی از دقت سیستم‌های بازشناسی گوینده در محیط‌های بدون نویز می‌کاهد.

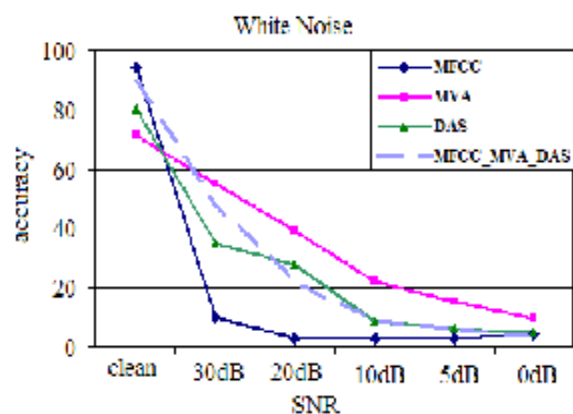
۵- روش پیشنهادی برای ترکیبی استخراج ویژگی‌های ترکیبی

با توجه به مطالب بیان شده در بخش‌های قبل و نتایج آزمایشاتی که در نمودارهای شکل ۴ مشاهده می‌شوند می‌توان دریافت که پیک‌های طیف خود همبستگی در مقاوم‌سازی سیستم‌های تعیین هویت گوینده در مقابل تأثیرات ناشی از نویز مهمه بهتر از ضرائب MFCC نرمال شده با عمل پس‌پردازش MVA عمل می‌کنند. ولی در مقابل نویزهای کانال وضعیت متفاوت است و روش MVA مقاوم‌تر از پیک‌های طیف خود همبستگی عمل کرده است. در ضمن نتایج آزمایش‌ها نشان می‌دهند که هر دوی این روش‌ها یعنی ویژگی‌های MFCC نرمال شده با MVA و پیک‌های طیف خود همبستگی علی‌رغم بهبود دقت سیستم‌های تعیین هویت گوینده در محیط‌های نویزی در محیط‌های تمیز و بدون نویز دقت بازشناسی را به مقدار قابل توجهی کاهش داده‌اند.

می‌توان با الحاق سه نوع ویژگی MFCC، MFCC نرمال شده با MVA و پیک‌های طیف دنباله خودهمبستگی یک بردار ویژگی جدید



(ب)



(ج)

شکل (۱): مقایسه دقت‌های بازشناسی ویژگی‌های MFCC، MFCC، DAS و ویژگی‌های ترکیبی پیشنهادی در مقابل نویزهای (الف) همهمه (ب) ماشین (ج) نویز سفید گاسی

۷- نتیجه‌گیری

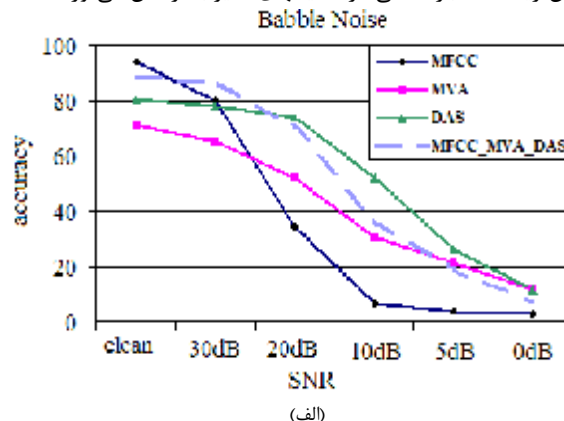
با توجه به اهمیت مقاوم‌سازی سیستم‌های تعیین هویت گوینده در مقابل نویزها برای استفاده‌های عملی، در این مجموعه سعی شد تأثیر دو روش استخراج پیک‌های طیف دنباله خودهمبستگی و روش پس‌پردازش MVA که در مقاوم‌سازی سیستم‌های خودکار بازشناسی گفتار بسیار موثر هستند، بر روی سیستم‌های تعیین هویت گوینده مورد بررسی قرار گیرد.

نتایج آزمایشات نشان دادند که روش پس‌پردازش MVA در مقاوم‌سازی سیستم‌های تعیین هویت گوینده در مقابل نویز سفید (کانولوشنی) و استخراج پیک‌های طیف خودهمبستگی نیز در مقاوم‌سازی سیستم‌های تعیین هویت گوینده در مقابل نویزهای همهمه و ماشین بسیار موثر هستند. ولی هر دوی این روش‌ها در محیط‌های تمیز در مقایسه با ویژگی‌ها رایج MFCC باعث کاهش دقت بازشناسی می‌شود.

نتایج آزمایش‌ها در نمودارهای شکل ۴ نشان داده شده‌اند. از نمودار قسمت (الف) از این شکل می‌توان دریافت که ویژگی‌های MFCC در مقابل نویز همهمه نسبت به ویژگی‌های MFCC، MFCC، DAS نرمال شده با روش MVA و ویژگی‌های ترکیبی پیشنهادی مقاوم‌تر هستند. همچنین به خوبی مشخص است که هرچند ویژگی‌های MFCC و MFCC نرمال شده نسبت به ویژگی‌های MFCC مقاوم‌تر هستند ولی در میزان آغستگی به نویز کم دقت بازشناسی آنها به میزان قابل توجهی کاهش یافته است. نکته دیگری که از این نمودار قابل استخراج است، کارایی ویژگی ترکیبی پیشنهادی است. این ویژگی علی‌رغم داشتن مقاومت خوب نسبت به نویز، در محیط‌های تمیز و بدون نویز نیز از راندمان بالایی برخوردار است.

در قسمت (ب) شکل ۴ نیز می‌توان مشاهده کرد که روش استخراج ویژگی ترکیبی پیشنهادی علی‌رغم اینکه در مقابل نویز ماشین نیز از مقاومت بالایی برخوردار است، در میزان آغستگی به نویز کم مانند دو روش DAS و پس‌پردازش MVA باعث کاهش دقت سیستم نمی‌شود. البته در این شکل کاملاً واضح است که پس‌پردازش MVA در مقابل نویز ماشین که نوعی نویز جمع‌شونده محسوب می‌شود، نتایج خوبی را تولید نمی‌کند.

در قسمت (ج) از شکل ۴ نیز می‌توان مشاهده کرد که پس‌پردازش MVA در مقابل نویز سفید گاسی که نوعی نویز کانولوشنی محسوب می‌شود، از مقاومت بسیار بالایی برخوردار است. این در حالی است که نویز سفید گاسی در سیستم‌های تعیین هویت گوینده عامل بسیار مخربی محسوب می‌شود و قادر است دقت بازشناسی ویژگی‌های MFCC را به میزان بسیار زیادی کاهش دهد. روش ترکیبی پیشنهادی در مقابل نویز سفید گاسی نیز مقاومت نسبی قابل قبولی را بدون کاهش زیاد دقت بازشناسی در محیط‌های تمیز به ارمغان می‌آورد.



(الف)

زیرنویس‌ها

- ¹ Automatic Speech and Speaker Recognition
- ² Robustness
- ³ Robust Feature Extraction
- ⁴ Feature Normalization
- ⁵ Model Adaptation
- ⁶ Mean
- ⁷ Mel-Cepstral
- ⁸ Dynamic
- ⁹ Cepstral Mean Normalization
- ¹⁰ Mean Normalization-Variance Normalization-ARMA Filtering
- ¹¹ Differentiation of Autocorrelation Sequence
- ¹² Spectral Peaks of the Autocorrelation Sequence
- ¹³ Differentiation of Autocorrelation Sequence
- ¹⁴ Mean Subtraction
- ¹⁵ Variance Normalization
- ¹⁶ Auto-Regression and Moving-Average
- ¹⁷ Babble
- ¹⁸ Volvo
- ¹⁹ White Gaussian Noise

همچنین روش جدیدی پیشنهاد شد که با الحاق ویژگی‌های MFCC، MFCC نرمال شده با پس‌پردازش MVA و پیک‌های طیف دنباله خودهمبستگی توانسته است بدون کاهش چشم‌گیر دقت بازشناسی در محیط‌های تمیز، تا حد زیادی بر مقاومت سیستم‌های تعیین هویت گوینده بیفزاید.

از آنجا که طول بردار ویژگی ترکیبی پیشنهادی بزرگ است و همه ویژگی‌های موجود درون آن از مقاومت یکسانی برخوردار نیستند، انتظار می‌رود با استفاده یک روش انتخاب ویژگی مقاوم بتوان به بردار ویژگی کوچکتری نیز دستیافت که از نظر مقاومت و دقت بازشناسی تفاوت چندانی بردار ویژگی فعلی نداشته باشد.

قدردانی

این مقاله مورد حمایت مالی مرکز صنایع نوین از وزارت صنایع جمهوری اسلامی ایران قرار گرفته است.

مراجع

- [۱] حاج‌احمدی، امیرحسین، همایون‌پور، محمد مهدی، فراهانی، غلامرضا، احدی، سید محمد، تعیین هویت گوینده مقاوم در مقابل نویز با استفاده از پیک‌های طیف خودکار سیگنال در حوزه خود همبستگی، دوازدهمین کنفرانس بین‌المللی سالانه انجمن کامپیوتر ایران، ۱۳۸۵.
- [1] Furui, S., "50 years of progress in speech and speaker recognition", SPECOM, Greece, pp.1-9, 2005.
 - [2] G. Farahani, S. M. Ahadi, M.M. Homayounpour, "Features based on filtering and spectral peaks in autocorrelation domain for robust speech recognition", Computer Speech and Language, Volume 21, Issue 1, pp. 187-205, 2006.
 - [3] Huang, K. C., Juang, Y. T., Chang, W. Ch., "Robust integration for speech features", Signal Processing 2006.
 - [4] Hermansky, H., Morgan, N., "RASTA Processing of Speech", IEEE Transactions On Speech And Audio Processing. Vol. 2. NO. 4. October 1994.
 - [5] Zhu, Q., Alwan, A., "Non-linear feature extraction for robust speech recognition in stationary and non-stationary Noise", Computer Speech and Language 17, pp. 381-402, 2003.
 - [6] Huang, X., Acero, A., Hon, H. W., "Spoken Language Processing", Prentice Hall 2001.
 - [7] Chen, C. P., Bilmes, J., "MVA Processing of Speech Features", UWEE Technical Report Number UWEETR-2003-2004, November 2003.
 - [8] Ding, G. H., Li, Ch., Xu, B., "Comparison of MLLR and CDCN for speech recognition in additive noise by experiments", ISCSLP 2002.
 - [9] http://spib.rice.edu/spib/select_noise.html