علوم و مهندسی کامپیوتر
نشریه علمی پژوهشی انجمن کامپیوتر ایران
مجلد ۳، شماره ۳ (الف)، پائیز ۱۳۸۴
صفحات ۳۱-۴۱
**مقاله عادی**

# بهبود کارائی سیستم های تعیین هویت گوینده در شرایط نویزی

محمد مهدی همایون پور　　ابراهیم شریف نبوی

آزمایشگاه سیستم‌های هوشمند صوتی و گفتاری
دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیر کبیر، تهران، ایران

## چکیده

در این مقاله به بررسی بعضی از روشهایی پرداخته میشود که بتوانند کارایی سیستم های تعیین هویت گوینده را در شرایط نویزی که در آنها تاثیر نویز، تاثیرات کانال انتقال و نوع میکروفون وجود دارند را تا حد زیادی حفظ نمایند تا بدین وسیله این سیستم ها بتوانند در محیطهای نویزی، تلفنی و مانند آن استفاده شوند. جهت بررسی روشهای مقاوم‌سازی سیستم‌های تعیین هویت گوینده، گفتار ۵۰ گوینده برای آموزش مدلهای گویندگان و آزمایش سیستم، از دادگان گفتاری فارس‌دات تلفنی استفاده گردید. ابتدا سکوت از گفتار گویندگان حذف و سپس جهت اعمال شرائط متفاوت آموزش و آزمایش، نسبت سیگنال به نویز فایلهای صوتی به مقادیر ۵، ۱۰، ۱۵ و ۲۰ دسی‌بل تغییر داده شد. ضرائب LPCC، LFCC، MFCC، از گفتار گویندگان استخراج و مدلهای GMM گویندگان ساخته شدند. سپس آزمایشهای تعیین هویت انجام گردیدند که از بین ویژگیهای فوق ضرائب MFCC از همه بهتر تشخیص داده شدند. همچنین با توجه به اینکه ضریب اول کپستروم که با انرژی فریم آنالیز مرتبط است، در تعیین هویت گوینده مفید نمی‌باشد، حذف آن از بردار ویژگی به میزان ۱۰،۴٪ بر افزایش راندمان سیستم تاثیر مثبت داشت. همچنین روشهای مختلفی مانند توالی اتوکورولیشن نسبی جهت بهسازی گفتار قبل از استخراج ویژگی و اعمال روشهای مختلفی جهت مقاوم‌سازی ضرائب کپسترال مانند تکنیک وزندهی خطی، تکنیک لیفتر کردن میانگذر، روش تفاضل میانگین در حوزه کپسترال، تکنیک PFL و ضرائب دینامیکی کپسترال مورد بررسی قرار گرفتند. اکثر این روشها باعث بهبود عملکرد سیستم تعیین هویت شدند. روش وزندهی خطی از بین این روشها نسبت به روش استفاده از ضرائب MFCC بتنهائی و بدون حذف ضریب اول کپستروم، بهترین روش مقاوم‌سازی در برابر نویز و تاثیرات کانال انتقال تشخیص داده شد. در این مقاله همچنین ترکیبهای جدیدی از روشهای فوق ارائه گردیده که با توجه به نتایج بدست آمده، این روشهای ترکیبی در افزایش مقاومت سیستم در برابر فاکتورهای محیط بسیار موثر می‌باشند. بهترین نتیجه بدست آمده ترکیب ضرائب MFCC و اعمال روش وزندهی خطی در یک بردار ویژگی همراه با ضرائب دلتا MFCC می باشد.

**کلمات کلیدی:** تعیین هویت گوینده، شرایط غیر متعارف، نویز، ضرایب کپسترال، دادگان فارس دات، وزن دهی ضرایب

## ۱- مقدمه

در صورتیکه بخواهیم سیستم های تعیین هویت گوینده در محیط های واقعی که از این محیط ها به محیط های غیر متعارف یا محیط های غیر آزمایشگاهی تعبیر می کنیم مورد استفاده قرار گیرند لازم است که آنها را در مقابل فاکتورهای مزاحم محیط مقاوم کنیم. این فاکتورها شامل منابع نویز جمع‌شونده، نویز کانولوشنال، انعکاس اتاق، خصوصیات ابزار کسب و انتقال گفتار نظیر میکروفن، تلفن و کانال‌های ارتباطی و اعوجاجات ناشی از آنها بوده و عموما خصوصیات طیفی صدا را تغییر داده و بر کارائی سیستمهای تعیین هویت تاثیر می‌گذارند. همچنین باید توجه داشت که شرائط تهیه داده‌های آموزشی یک سیستم تعیین هویت شناخته شده هستند، درحالیکه شرائط آزمایش معمولا ناشناخته و متنوع می‌باشند. عدم تطابق صوتی با توجه به تفاوت محیط‌های آموزش و آزمایش نیز می‌تواند کارائی شناسایی گوینده را به میزان زیادی کاهش دهد. سیستم‌هایی که از طریق خط تلفن کار می‌کنند، با مشکل بیشتری مواجه هستند و باید با انواع نویزهایی که توسط خط تلفن تولید می‌شوند، مقابله نمایند. تلفن محدودیت‌های خاص خود را بر سیگنال و روشهای تشخیص آن اعمال می‌کند. البته مشکلات دیگری نظیر نویز سوئیچینگ و انعکاس خط نیز تا حدودی وجود دارند. اما اکثر این نویزها به دلیل محدودیت پهنای باند، اثر زیادی بر سیگنال نمی‌گذارند. محدودیت پهنای باند

باعث می‌شود، فرکانس‌های بالاتر از KHz ۴، که برای تشخیص گوینده مفید هستند، حذف شوند. شکل پاسخ فرکانسی دهنی، گوشی و خط تلفن، یکنواخت نیست. عملا ممکن است در دفعات مختلف، مکالمه از مسیرهای مختلف با پاسخهای فرکانسی متفاوت انجام گیرد. البته راه‌حل‌های زیادی برای غلبه بر این مشکلات بطور جداگانه ارائه شده‌اند. در این مقاله سعی شده است روشهایی برای کاهش تاثیر منفی شرایط نویزی بر کارائی سیستم های تعیین هویت بتنهائی یا در ترکیب با یکدیگر مورد بررسی قرار گیرند. بخش ۲ این مقاله به تشریح تکنیکهای مقابله با نویز و شرایط غیر متعارف اختصاص دارد. در بخش ۳ به نحوه انجام آزمایشات و نتایج بدست آمده می‌پردازیم. در نهایت، بخش ۴ به ارائه پیشنهادات و نتیجه‌گیری می پردارد.

## ۲- تکنیکهای مقابله با نویز و شرایط غیر متعارف

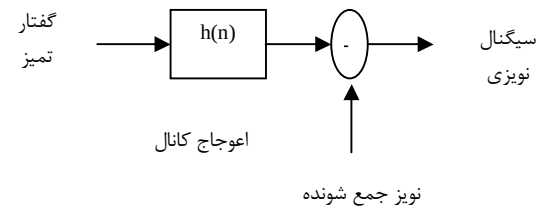
روشهای فراوانی در زمینه مقاومت در برابر نویز و شرایط غیر متعارف ارائه شده است. این روشها می‌توانند به چهار مجموعه عمومی تقسیم‌بندی شوند. مجموعه اول، تخمین سیگنال تمیز از سیگنال گفتار نویزی می‌باشد که به تکنیکهای بهسازی گفتار معروفند. این روشها قبل از مرحله استخراج ویژگی انجام می‌گیرند. مجموعه دوم، استفاده از پارامترها و ویژگیهای مقاوم در برابر نویز را توصیه می‌کنند که مزیت آنها در این است که نیازی به پردازش اضافی وجود ندارد. مجموعه سوم، شامل تکنیکهای تقویت پارامترها و حذف تاثیر کانال و نوع میکروفن از ویژگی‌ها می‌باشد و مجموعه آخر تکنیکهایی هستند که در مرحله مدلسازی استفاده می‌شوند و به تکنیکهای مبتنی بر مدل معروف هستند. در بخشهای بعد به بیان بعضی از این روشها که در این مقاله مورد ارزیابی قرار گرفته اند می پردازیم.

### ۲-۱ استفاده از توالی اتوکورولیشن نسبی

در استفاده از توالی اتوکورولیشن نسبی (RAS)[۱] که جزء تکنیکهای بهسازی گفتار می‌باشد، فرض بر اینست که نویز جمع شونده و کانال ثابت هستند[۱]. نحوه نویزی شدن سیگنال صحبت در شکل ۱ نمایش داده شده است و بصورت زیر مدل می‌شود:

$$y(m,n) = x(m,n) \otimes h(n) + w(m,n) \quad 0 \le m \le M-1, O \le n \le N-1 \quad (۱)$$

که m اندیس فریم، n اندیس زمانی گسسته، x(m,n) گفتار تمیز، y(m,n) گفتار نویزی شده، h(n) پاسخ ضربه کانال، w(m,n) نویز جمع شونده و $\otimes$ عملگر کانولوشن می‌باشند.




شکل ۱- بلوک دیاگرام نویزی شدن سیگنال گفتار

اگر x(m,n) وh(n) و w(m,n) مستقل باشند، اتوکورولیشن سیگنال نویزی شده می‌تواند بصورت زیر بدست آید:

$$r_{yy}(m,k) = r_{xx}(m,k) \otimes h(k) \otimes h(-k) + r_{ww}(m,k) \quad 0 \le m \le M-1, o \le K \le N-1 \quad (۲)$$

که $r_{yy}(m,k)$ ، $r_{xx}(m,k)$ و $r_{ww}(k)$ سریهای اتوکورولیشن زمان کوتاه گفتار نویزی شده، گفتار تمیز و نویز جمع شونده می‌باشند و k اندیس سری اتوکورولیشن هر فریم است. چون نویز جمع شونده ثابت فرض می‌شود پس $r_{ww}(m,k)$ برای همه فریمهای با اندیس m بدون تغییر فرض می شود و بنابراین اندیس فریم می تواند حذف شده و فرمول بصورت زیر در آید:

$$r_{yy}(m,k) = r_{xx}(m,k) \otimes h(k) \otimes h(-k) + r_{ww}(k) \quad 0 \le m \le M-1 \quad , \quad 0 \le k \le N-1 \quad (۳)$$

در اینجا تنها سری اتوکورولیشن یکطرفه برای هر فریم محاسبه می‌شود:

$$r_{yy}(m,k) = \frac{1}{N-K} \sum_{j=0}^{N-1-K} y(m,j) y(m,j+k) \quad 0 \le m \le M-1 \quad , \quad 0 \le k \le N-1 \quad (۴)$$

با مشتق‌گیری از دو طرف فرمول (۳) بر اساس m خواهیم داشت:

$$\frac{\partial}{\partial m} r_{yy}(m,k) = \frac{\partial}{\partial m} \{ r_{xx}(m,k) \} \otimes h(k) \otimes h(-k) \quad 0 \le m \le M-1 \quad 0 \le k \le N-1 \quad (۵)$$

توالی $\{\partial r_{yy}(m,k)/\partial m\}_{k=0}^{N-1}$ سری اتوکورولیشن نسبی (RAS) گفتار نویزی در فریم mام نامیده می‌شود. فرمول (۵) نشان می‌دهد که اثر نویز جمع شونده از RAS گفتار نویزی شده، حذف می‌گردد ولی نویز کانولوشنال همچنان باقی است. البته می‌توان نویز کانولوشنال را با تبدیل RAS در حوزه ضرائب کپسترال حذف کرد.

### ۲-۲ ضرایب مرکز ثقل زیرباند طیفی

افزوده شدن نویز به سیگنال گفتار، طیف توان گفتار را در همه فرکانسها تحت تاثیر قرار می‌دهد ولی این تاثیر در قسمتهای با دامنه بزرگترِ طیف (فرمانتها) کمتر است (زیرا SNR در نواحی فرمانت بیش از سایر نواحی است). وقتی ضرائب کپسترال از فرمانتها مانند نواحی دیگر استفاده می‌کنند، نسبت به نویز حساس خواهند بود ولی اگر فرمانتها به عنوان جزئی از بردار ویژگی بکار گرفته شوند با توجه به اینکه نقاط نظیر فرمانتها در طیف کمتر از سایر قسمتها تحت تاثیر نویز قرار می‌گیرند، این مشکل حل خواهد شد. ضرائب مرکز ثقل زیر باند طیفی SSC[۲] شباهت زیادی به فرکانسهای فرمانت دارند و می‌توانند به سادگی و با اطمینان (بدون هیچ خطای تخمین) از طیف توان سیگنال گفتار استخراج شوند[۲]. این ضرائب جزء پارامترهای ذاتا مقاوم در برابر نویز می‌باشند.

جهت محاسبه این ویژگیها باند فرکانسی را (۰ تا $F_s/2$ که $F_s$ فرکانس نمونه‌برداری بر حسب هرتز می‌باشد) به تعداد ثابتی زیرباند تقسیم کرده و مرکزثقل هر زیرباند را با استفاده از طیف توان سیگنال گفتار محاسبه می‌نمائیم. اگر باند فرکانسی را به M زیرباند تقسیم کرده و لبه‌های پایین و بالای هر زیرباند mام را به ترتیب $l_m$ و $h_m$ بنامیم، mامین SSC بصورت زیر محاسبه می‌شود:

$$SSC_m = \frac{\int_{l_m}^{h_m} f w_m(f) p^g(f) df}{\int_{l_m}^{h_m} w_m(f) p^g(f) df} \quad (۶)$$

که p(f) طیف توان و $g$ ثابت کنترل بازه پویای طیف توان می‌باشد. اگر $g < 1$ باشد، بازه پویای طیف توان کاهش می‌یابد. همچنین $w_m(f)$ شکل فیلتر اعمالی است که می‌تواند مستطیلی یا مثلثی باشد.

### ۲-۳ روش وزندهی خطی

ایده مبنا در استفاده از روشهای وزندهی کپسترال که همگی جزء تکنیکهای تقویت پارامترها و حذف تاثیر کانال و نوع میکروفن از ویژگی‌ها می‌باشند، توجه به حساسیت ضرائب کپسترال مرتبه پایین نسبت به شیب طیفی و حساسیت ضرائب

کپسترال مرتبه بالا نسبت به نویز می‌باشد[۳]. تکنیکهای وزندهی با ضرب ضرائب کپسترال $C(n)$ در پنجره w(n) و استفاده از کپسترومهای وزندار به عنوان بردار ویژگی تعریف می‌شوند. این عملیات وزندهی با عنوان لیفتر کردن[۳] نیز شناخته می‌شود. اولین نتیجه لیفتر کردن، استخراج یک بردار ویژگی با بعد متناهی از ضرائب کپسترال می‌باشد. همچنین انتخابهای دقیق w(n) مقاومت در برابر نویز را افزایش می‌دهد. چندین نوع روش وزندهی وجود دارد که روش وزنـــدهی خطی یا LW[۴] یکی از آنهاست. پنجره w(n) در تکنیک LW بصورت زیر تعریف می‌شود:

$$w(n)=\begin{cases} n & n=1,2,...,L \\ 0 & otherwise \end{cases} \qquad (۷)$$

که در آن L طول پنجره می‌باشد. روش LW هر ضریب کپسترال را با استفاده از اندیس n آن جزء وزندهی می‌نماید لذا نقش اجزاء با مرتبه کمتر کمرنگ می‌شود.

## ۲-۴ روش لیفتر کردن میانگذر

روش لیفتر میانگذر یا BPL[۵] یکی دیگر از روشهای وزندهی ضرائب کپسترال می‌باشد و پنجره w(n) در آن بصورت زیر تعریف می‌شود[۴]:

$$w(n)=\begin{cases} 1+\frac{L}{2}\sin(\frac{np}{L}) & n=1,2,...,L \\ 0 & otherwise \end{cases} \qquad (۸)$$

در این تکنیک اعضاء بردار ضرائب کپسترال توسط یک تابع سینوسی، وزندهی می‌شوند. بنابراین ضرائب کپسترال مرتبه پایینتر و همچنین ضرائب مرتبه بالاتر کم اهمیت‌تر خواهند شد. در تکنیکهای وزندهی LW و BPL وزنها تنها تابعی از اندیس کپسترال هستند و هیچ برداشت صریحی از تغییرات آنی که حاصل از شرایط محیطی متفاوت می‌باشند، در ضرائب کپسترال وجود ندارند.

## ۲-۵ تفاضلِ میانگین در حوزهٔ کپسترال

یکی از روشهایی که برای مقابله با اثرِ کانال انتقال و یا نویز کانولوشنال پیشنهاد شده است، روش تفاضلِ میانگین در حوزهٔ کپسترال و یا CMS[۶] است[۵]. در واقع مشخصهٔ فرکانسی کانال انتقال در طیف سیگنال گفتار ضرب می‌شود و این ضرب، معادلِ کانولوشنِ پاسخ ضربهٔ کانال با سیگنال گفتار می‌باشد (نویز کانولوشنال). اگر فرض کنیم که عبارت گفته شده به قدر کافی طولانی و نیز از لحاظ فونتیکی متعادل باشد[۷]، آنگاه بطور تقریبی می‌توان در نظر گرفت که میانگین بردارهای کپسترال در طول زمان با شرط فوق صفر است. همچنین می‌توان در نظر گرفت که بردارهای کپسترال بصورت جمع شونده تحت تأثیرِ بردارِ کپسترالِ مربوط به مشخصهٔ کانال واقع می‌شوند. یعنی میانگین بردارهای کپسترال در یک عبارت طولانی و از لحاظِ فونتیکی متعادل، برابر بردار کپسترالِ مربوط به مشخصهٔ کانال می‌باشد. با کم‌کردن این میانگین از بردارهای کپسترال عبور داده شده از کانال از خود این بردارهای کپسترال، اثرِ کانالِ انتقال بطور تقریبی از بین رفته و بردار کپسترالِ اصلیِ بدون اعوجاج بدست خواهد آمد. این عمل را تفاضلِ میانگین در حوزهٔ کپسترال می‌نامند که برای جبرانِ اثرِ کانال انتقال و یا برای حذف نویز کانولوشنال بکار می‌رود. روش CMS بر این فرض استوار است که عبارت گفتاری، طولانی و از لحاظ فونتیکی متعادل باشد، مثلا در بازشناسی گوینده مستقل از متن که گوینده در هر تعیین هویت، گفتاری کافی بیان می‌نماید، می‌توان در نظر گرفت که میانگین بردارهای کپسترال در یک عبارت ثابت، بطور تقریبی یک بردارِ کپسترال ثابت خواهد شد. در این صورت میانگینِ بردارهای کپسترال در این عبارت، حاوی اطلاعات با تغییرات کم مانند مشخصه کانال و میکروفن خواهد بود که تفریق این بردار میانگین (بردار ثابت) از بردارهای کپسترال اولیه منجر به حذف مشخصه کانال خواهد شد.

## ۲-۶ روش لیفترکردن پس‌فیلتر

ایده تکنیک PFL[۸] که در دسته روشهای تقویت پارامترها و حذف تاثیر کانال و نوع میکروفن از ویژگی‌ها قرار دارد، مبتنی بر این واقعیت است که بطور مفهومی، مقاومت در برابر نویز در نواحی فرمانت (پیکهای طیف) بیشتر از نواحی دیگر است. پس‌فیلتر از A(x) معکوس تابع تبدیل H(z) مجرای گفتار بدست می‌آید و تابع انتقال آن بصورت زیر می‌باشد[۶]:

$$H_{pfl}(z)=\frac{A(z/b)}{A(z/a)} \quad 0<b<a\le 1 \qquad (۹)$$

اگر A(z) حداقل فاز باشد، $H_{pfl}(z)$ نیز حداقلْ فاز خواهد بود. بنابراین، PFL علّی است و ضرائب کپسترال جدید بصورت زیر بدست می‌آیند:

$$C_{pfl}(0)=0 \qquad (۱۰)$$

$$C_{pfl}(n)=C(n)\left[a^{n}-b^{n}\right] \qquad 0<n<p \qquad (۱۱)$$

تکنیک PFL یک تکنیک وزندهی یا لیفتر کردن ضرائب کپسترال است و در برابر اثرات کانال و نویز بسیار مقاوم می‌باشد.

## ۲-۷ ضرایب دینامیکی کپسترال

مشتقاتِ ضرایب کپسترال برای نشان‌دادن دینامیک مجرای گفتار و نحوهٔ تغییرات زمانی آن مناسب و مفید هستند. اثر کانال انتقال در واقع یک عنصر جمع شونده در حوزهٔ لگاریتم طیف است که تغییرات کندی در طول زمان دارد و فیلتر دلتا (مشتق اول) آن را از بین می‌برد. در واقع می‌توان گفت که ضرایب دلتا کپسترال فاقد اثرات مربوط به مشخصات کانال انتقال بوده و نسبت به مشخصهٔ کانالِ انتقال مقاوم هستند. این پدیده در مورد ضرایب دِلتادِلتا کپسترال یا ضرایب شتاب یا مشتقات دوم نیز صادق است. در واقع ضرایب دلتا کپسترال و ضرایب دلتادلتا کپسترال سرعت و شتاب تغییرات مجرای گفتار را مدل می‌کنند و بکارگیری آنها در سیستم‌های بازشناسی راندمان این سیستم‌ها را بالا می‌برد. نکتهٔ بسیار مهمی که در مورد این ضرایب وجود دارد این است که بازای هر طولی از پنجرهٔ رگرسیون، فیلتر دلتا بعضی از مؤلفه‌های فرکانسی مفید را حذف می‌نماید و بردارهای کپسترال را دچار اعوجاج می‌کند، بدین معنی که بسیاری از اطلاعات مفید مربوط به مجرای گفتار در ضرایب دلتا کپسترال از بین می‌روند. به همین خاطر تنها بکار بردن ضرایب دلتا کپسترال در یک سیستم بازشناسی موجب فقدان بسیاری از اطلاعات مفید می گردد، از این رو در سیستم‌های بازشناسی، ضرایب دلتا کپسترال به تنهایی استفاده نمی‌شوند، بلکه توأم با ضرایب کپسترالِ استاتیک استفاده می‌گردند. بعنوان مثال اگر بردار کپسترالِ استاتیک یک بردار ۱۲ بعدی باشد، بردار ویژگی نهایی با اضافه کردن ضرایب دلتا کپسترال و دلتادلتا کپسترال به بردار کپسترالِ استاتیک، یک بردار ۳۶ بعدی خواهد بود. فیلتر دِلتا در واقع برای از بین‌بردن اثرات نویز کانولوشنال و مشخصهٔ کانال و مؤلفه‌های با تغییرات آهسته در حوزهٔ لگاریتم طیف است و نمی‌تواند برای مقابله با نویزهای جمعی بکار رود.

## ۲-۸ روشهای ترکیبی

ترکیب روشهای مختلف با توجه به امتیازات و کارائی هر کدام سبب بهبود عملکرد مجموعه تعیین هویت گوینده خواهد شد. با توجه به سیگنال گفتاری و نوع نویز یا اعوجاجات موثر در تخریب سیگنال گفتار می‌توان ترکیبی از روشهای مختلف را آزمود. معمولا ترکیب روشها، چون باعث تجلی نقاط قوت و پوشش نقاط ضعف روشهای مورد استفاده می‌شود، مفید بوده و توصیه می گردند.

## ۳- آزمایشات و نتایج

### ۳-۱ دادگان مورد استفاده

در این مقاله، از دادگان فارس‌دات تلفنی که مشخصات کامل آن در ادامه ارائه خواهد شد، استفاده شده است. این دادگان حاوی صدای ۶۴ گویشور فارسی زبان در دو حالت فی‌البداهه و رسمی به میزان ۷ ساعت و۵۶ دقیقه، با در نظر گرفتن توزیع سن، جنس و لهجه‌های مختلف کشور، می‌باشد. در هنگام ضبط این دادگان، گویشوران از شهر یا محل اقامت خود با مرکز تهیه دادگان در تهران تماس تلفنی گرفته‌اند. برای هر گویشور به طور متوسط ۷/۵ دقیقه داده‌ی گفتاری تلفنی موجود می‌باشد. داده‌ها در حالت فی‌البداهه شامل اطلاعات عمومی درباره‌ی گویشور، محل اقامت وی و گفتن یک خاطره است. همچنین در حالت رسمی، داده‌ها شامل اعداد اصلی، ایام هفته، اسامی ماه، حروف الفبا، ۵۰ کلمه‌ی پربسامد، شش جمله‌ی فارس‌دات غیر تلفنی و کلیه هجاهای باز زبان فارسی است. برای ضبط صدا از کارت صوتی ساند بلاستر ۱۶ بیتی با فرکانس نمونه برداری ۱۱۰۲۵ هرتز استفاده شده است.

### ۳-۲ استخراج ویژگی، مدل کردن گویندگان و چگونگی محاسبه کارایی

به منظور بررسی روشهای مقاوم‌سازی سیستم‌های تعیین هویت گوینده لازم است که بتوان نسبت سیگنال به نویز فایل‌های صوتی را تعیین و در صورت نیاز تغییر داد. پس از محاسبه نسبت سیگنال به نویز فایل صوتی هر گوینده، با افزودن ضریبی از نمونه‌های نویز سفید، نسبت سیگنال به نویز را کاهش دادیم. در این مقاله نسبت سیگنال به نویز فایل‌ها به مقادیر ۲۰، ۱۵، ۱۰ و ۵ دسی‌بل تغییر داده شده‌اند. ۵۰ گوینده از دادگان فارس‌دات تلفنی انتخاب و بمنظور مدل کردن گویندگان از روش مدل مخلوط گوسی GMM با ۶۴ مخلوط گوسی و ۶۰ ثانیه داده آموزشی استفاده شد. باقیمانده فایل گفتاری هر گوینده به قسمتهای ۲ ثانیه‌ای بعنوان گویشهای آزمایش شکسته شد. در این مقاله انواع ضرائب کپسترال به عنوان ویژگی مورد بررسی قرار گرفتند. در هر آزمایش تعیین هویت، ویژگیهای هر گویش استخراج و با مدل GMM کلیه گویندگان مقایسه و با توجه به مقدار درستنمائی[۹] حاصل، گوینده گویش تست تعیین گردید. در این مقاله ارزیابی روشهای مقاوم سازی با محاسبه کارایی سیستم تعیین هویت صورت گرفته است. کارایی تعیین هویت بر حسب درصد، بسادگی از طریق محاسبه تعداد دفعات تشخیص صحیح هویت هر گوینده تقسیم بر تعداد تستهای تعیین هویت، ضربدر ۱۰۰ بدست آمده است.

### ۳-۳ انتخاب نوع ویژگی مقاوم

تعیین هویت گویندگان توسط روشهای مختلف استخراج ویژگی LPCC[۱۰]، LFCC[۱۱]، MFCC وSSC انجام گردید. در تمامی آزمایشات، طول پنجره فریم ۳۰ میلی‌ثانیه و طول شیفت فریم ۱۵میلی‌ثانیه در نظر گرفته شده است. بجز روش SSC که ۳ ویژگی با استفاده از آن استخراج گردید طول بردار ویژگی در سایر روشها ۱۲ بوده است. همچنین از پنجره همینگ استفاده شده و ضریب پیش‌تاکید ۰/۹۷۵ در نظر گرفته شده است. نتایج مربوط به راندمان تعیین هویت گوینده برا بازاء شرائط متفاوت بر اساس SNR محیط آموزش و آزمایش در جداولی جمع‌آوری گردیده است. در این جداول که جدول ۱ و ۲ نمونه‌ای از آنها است، در سطر اول مقدار SNR محیط آزمایش مشخص شده است. به همین ترتیب ستون اول SNR محیط آموزش را نشان می‌دهد. همچنین سطر آخر میانگین هر ستون را که همان میانگین راندمان تعیین هویت گوینده در محیط آزمایش با SNR ثابت و محیط‌های آموزش با SNR متفاوت است، نشان می‌دهد و ستون آخر نیز میانگین هر سطر را که همان میانگین راندمان تعیین هویت گوینده در محیط آموزش با SNR ثابت با توجه به محیط‌های آزمایش متفاوت است، نشان می‌دهد. در جدول ۱ راندمانهای تعیین هویت در زمانی که همان داده‌های آموزش برای آزمایش استفاده می‌شوند مشاهده می‌گردند ولی جدول ۲ راندمانهای تعیین هویت را در زمانی که داده‌های آزمایشی متفاوت با داده‌های آموزشی هستند، نمایش می‌دهد.

جدول ۱- راندمان تعیین هویت گوینده با استفاده از ویژگی MFCC برای داده‌های آموزشی بر حسب درصد

| MFCC | 20dB | 15dB | 10dB | 5dB | میانگین |
|---|---|---|---|---|---|
| 20dB | ۱۰۰/۰ | ۹۴/۰ | ۵۶/۰ | ۱۸/۰ | ۶۷/۰ |
| 15dB | ۹۴/۰ | ۱۰۰/۰ | ۱۰۰/۰ | ۴۲/۰ | ۸۴/۰ |
| 10dB | ۴۴/۰ | ۱۰۰/۰ | ۱۰۰/۰ | ۹۶/۰ | ۸۵/۰ |
| 5dB | ۱۸/۰ | ۳۸/۰ | ۹۰/۰ | ۱۰۰/۰ | ۶۱/۵ |
| میانگین | ۶۴/۰ | ۸۳/۰ | ۸۶/۵ | ۶۴/۵ | - |

جدول ۲- راندمان تعیین هویت گوینـــده با استفاده از ویـــژگی MFCC برای داده های آزمایش بر حسب درصد

| MFCC | 20dB | 15dB | 10dB | 5dB | میانگین |
|---|---|---|---|---|---|
| 20dB | ۹۲/۰ | ۶۳/۵ | ۳۳/۱ | ۱۲/۱ | ۵۰/۲ |
| 15dB | ۸۱/۱ | ۸۴/۹ | ۶۶/۶ | ۲۴/۵ | ۶۴/۳ |
| 10dB | ۳۵/۳ | ۷۳/۵ | ۸۶/۰ | ۴۶/۴ | ۶۰/۳ |
| 5dB | ۱۱/۷ | ۲۴/۸ | ۵۶/۰ | ۷۷/۲ | ۴۲/۴ |
| میانگین | ۵۵/۰ | ۶۱/۷ | ۶۰/۴ | ۴۰/۱ | - |

به منظور مقایسه بهتر نتایج حاصل از تعیین هویت بازاء ویژگیهای مختلف LPCC، LFCC، MFCC، SSC، میانگین راندمانهای تعیین هویت را بازاء SNRهای مختلف برای داده‌های آموزش و آزمایش بدست آوردیم. این نتایج در جدول ۳ ارائه شده اند. هر راندمان ارائه شده در این جدول حاصل میانگین راندمانهای تعیین هویت در شرائطی که SNR داده‌های آموزش برابر یا بیشتر از SNR داده‌های آزمایش باشند بدست آمده است (بعنوان نمونه در جداول ۱و ۲ این راندمانها بصورت پررنگ مشخص شده‌اند و در بالای قطر اصلی جدول قرار دارند).

جدول ۳- میانگین راندمان تعیین هویت گوینده بازاء ویژگی‌های مختلف برحسب درصد

| نوع ویژگی | LPCC | LFCC | MFCC | SSC+MFCC |
|---|---|---|---|---|
| داده‌های آموزش | ۶۷/۲ | ۷۱/۰ | ۸۰/۶ | ۷۶/۴ |
| داده‌های آزمایش | ۴۴/۶ | ۴۶/۹ | ۵۸/۶ | ۵۶/۱ |

نتایج تعیین هویت بازاء ضرائب LFCC با ۲۰ فیلتر در بانک فیلتر مربوطه گویای کارائی بهتر این ضرائب نسبت به ضرائب LPCC می‌باشد. نتایج ارزیابی ویژگی MFCC برای داده‌های آموزشی و آزمایشی درحالیکه تعداد فیلترها در بانک فیلتر برای این ضرائب نیز ۲۰ در نظر گرفته شده است و نمونه‌های طیف فرکانسی خارج از محدوده ۲۰۰ و ۵۰۰۰ هرتز صفر می‌شوند، حاکی از بهتر بودن راندمان ضرائب MFCC نسبت به سایر ویژگیهایی است که راندمان استفاده از آنها قبل از این ارائه شد. بخاطر برتری این ویژگی و نیز مبنا قرار گرفتن این ویژگی در آزمایشات بعدی در این مقاله، جزئیات دقیقتری از کارایی این ویژگی در جداول ۱ و ۲ آورده شده است.

یکی دیگر از ویژگی‌هایی که می‌تواند برای مقاوم‌سازی سیستم‌های تعیین هویت بکار رود ضرائب مرکز ثقل زیرباند طیفی می‌باشد. همانطور که شرح داده شد، ضرائب مرکز ثقل زیرباند طیفی یا همان ضرائب SSC شباهت زیادی به فرکانسهای فرمانت دارند. با توجه به اینکه تاثیر نویز بر بخشی از طیف که حاوی فرکانسهای فرمانت است، کمتر از سایر قسمتها می‌باشد، لذا از ضرائب SSC همراه با ضرائب MFCC در یک بردار شامل ۱۵ ویژگی (۱۲ ضریب MFCC و ۳ ضریب SSC) استفاده گردید. با توجه به نتایج بدست آمده و مقایسه آنها با راندمان سیستم تعیین هویتی که از ۱۵ ضریب MFCC به عنوان بردار ویژگی استفاده می‌کند، ملاحظه می‌شود که استفاده از این ضرائب تنها زمانی که محیط‌های آموزش و آزمایش از نسبت سیگنال به نویز بالائی برخوردارند، باعث افزایش کارائی سیستم می‌گردد. ولی هنگامیکه میانگین راندمان آنها مورد مقایسه قرار می‌گیرد (جدول ۴ را ببینید) استفاده از ۱۵ ضریب MFCC در یک بردار ویژگی راندمان بهتری را نسبت به استفاده از بردار ویژگی شامل ۱۲ ضریب MFCC و ۳ ضریب SSC، با داده‌های آزمایش نشان می‌دهد.

جدول ۴- میانگین راندمان تعیین هویت گوینده با بردار ویژگی شامل ۱۵ ضریب MFCC و بردار ویژگی شامل ۱۲ ضریب MFCC و ۳ ضریب SSC بر حسب درصد

| بردار ویژگی | ۱۵ ضریب MFCC | ۱۲ ضریب MFCC و ۳ ضریب SSC |
|---|---|---|
| داده‌های آموزش | ۷۳/۸ | ۷۶/۴ |
| داده‌های آزمایش | ۵۶/۴ | ۵۶/۱ |

ضرائب MFCC راندمان بهتری را برای تعیین هویت گوینده در محیط‌های نویزی و گفتار منتقل شده برروی خط تلفن نسبت به ویژگی‌های LPCC، LFCC و SSC ارائه می‌نمایند. ضرائب کپسترال مبتنی بر معیار مل بر اساس آنالیز بانک فیلتر و ادراک گوش انسان از فرکانس کار می‌کنند. لازم بذکر است در تحقیق صورت گرفته در [۷] ضرائب LPCC نسبت به ضرائب MFCC کارائی بهتری را در تعیین هویت گوینده در محیط‌های تمیز و گفتار غیرتلفنی ارائه نموده‌اند. در زمانی که SNR سیگنال گفتاری بالا است و از طرفی گفتار، تلفنی نیست و باند بالای فرکانسی توسط تلفن محدود نگشته است، ضرائب LPCC کارائی بهتری نسبت به ضرائب MFCC برای تعیین هویت گوینده دارند چراکه ضرائب LPCC از اطلاعات وابسته به گوینده، موجود در باند بالای فرکانسی به همان میزانی استفاده می‌نمایند که از اطلاعات موجود در باندهای پائین فرکانسی؛ درحالیکه رزولوشن و دقت فرکانسی در هنگام استخراج ضرائب MFCC در باندهای بالای فرکانسی کمتر از باندهای پائین می‌باشد. دیگر اینکه در محیط‌های نویزی از میزان پایداری محاسبات ضرائب LPCC کاسته می‌شود و لذا کارائی آن کم می‌شود، درحالیکه ضرائب MFCC بدلیل خاصیت میانگین‌گیری از نمونه‌های فرکانسی در هر فیلتر از بانک فیلتر، از خاصیت کاهش تاثیر نویز برخوردار می‌باشند.

همچنین نتایج نشان می‌دهند که در مواردی که SNR داده‌های آموزش با داده‌های آزمایش یکسان است، راندمان تعیین هویت حداکثر مقدار خود را داراست. همچنین به نظر می‌رسد که در حالتیکه SNR داده‌های آموزشی بالاتر از SNR داده‌های آزمایشی است، راندمان بهتری نسبت به مواردی که SNR داده‌های آموزشی کمتر از SNR داده‌های آزمایشی باشد، بدست می‌آید.

در آزمایش دیگری تاثیر حذف ضریب اول کپستروم بردار ویژگی مورد بررسی قرار گرفت. ضریب اول کپستروم بیانگر انرژی فریم می‌باشد. از آنجا که انرژی فریم کمتر حاوی اطلاعات وابسته به گوینده می‌باشد لذا وجود یا حذف آن از بردارهای ویژگی می‌تواند حائز اهمیت باشد. بدین نحو که بردارهایی از ضرائب MFCC با ۱۳ ویژگی استخراج گردید و سپس ضریب اول کپستروم را از این مجموعه حذف کردیم. همچنین محدودیت حذف نمونه‌های طیف فرکانسی خارج از محدوده ۲۰۰ و ۵۰۰۰ هرتز را نیز برداشتیم. نتیجه حاصل حاکی از بهبود کارائی ضرائب کپسترال می‌باشد. این بهبود در جدول ۵ مشاهده می‌شود.

جدول ۵- میانگین راندمان تعیین هویت گوینده با توجه به حذف یا عدم حذف اولین ضریب کپستروم بر حسب درصد

| ضرائب MFCC | بدون حذف اولین ضریب کپستروم | با حذف اولین ضریب کپستروم |
|---|---|---|
| داده‌های آموزش | ۸۰/۶ | ۸۷/۸ |
| داده‌های آزمایش | ۵۸/۶ | ۶۸/۹ |

همچنین ضرائب MFCC را علاوه بر ۲۰ فیلتر در بانک فیلتر با ۲۴ و ۲۸ فیلتر در بانک فیلتر استخراج کردیم که با توجه به نتایج بدست آمده که در جدول ۶ قابل مشاهده است ضرائب MFCC که با استفاده از ۲۴ فیلتر استخراج گردیده‌اند، بازاء داده‌های آزمایشی کارائی بیشتری را دارا می‌باشند.

جدول ۶- میانگین راندمان تعیین هویت گوینده بازاء تعداد فیلترها در بانک فیلتر برحسب درصد

| تعداد فیلترها در بانک فیلتر | ۲۰ | ۲۴ | ۲۸ |
|---|---|---|---|
| داده‌های آموزش | ۸۷/۸ | ۸۷/۴ | ۸۷/۰ |
| داده‌های آزمایش | ۶۸/۹ | ۶۸/۹ | ۶۸/۶ |

بهمین نحو ضرائب SSC را همراه با ضرائب MFCC که با استفاده از بانک فیلتر شامل ۲۴ فیلتر استخراج گردیده‌اند و ضریب اول کپستروم حذف گردیده است در یک بردار شامل ۱۵ ویژگی (۱۲ ضریب MFCC و ۳ ضریب SSC) آزمایش کردیم. با توجه به نتایج بدست آمده و مقایسه آنها با راندمان سیستم تعیین هویتی که از ۱۵ ضریب MFCC به عنوان بردار ویژگی استفاده می‌کند (در حالیکه بانک فیلتر شامل ۲۴ فیلتر می‌باشد و ضریب اول کپستروم حذف گردیده است) این نتیجه حاصل شد که تاثیر اضافه کردن این سه ضریب SSC کمتر از تاثیر اضافه کردن سه ضریب MFCC به بردارهای ویژگی ۱۲ عضوی می‌باشد (جدول ۷ را ببینید).

جدول ۷- میانگین راندمان تعیین هویت گوینده گوینده با بردار ویژگی شامل ۱۵ ضریب MFCC و بردار ویژگی شامل ۱۲ ضریب MFCC و ۳ ضریب SSC با توجه به حذف اولین ضریب کپستروم بر حسب درصد

| بردار ویژگی | ۱۵ ضریب MFCC | ۱۲ ضریب MFCC و ۳ ضریب SSC |
|---|---|---|
| داده‌های آموزش | ۹۰/۶ | ۷۸/۰ |
| داده‌های آزمایش | ۷۴/۱ | ۶۰/۵ |

استفاده از بانک فیلتری شامل ۲۴ فیلتر و حذف ضریب اول کپستروم در استخراج ضرائب LPCC و LFCC نیز مورد آزمایش قرار گرفت. نتایج حاکی از بهبود راندمان این دو نوع ویژگی می‌باشد. ولی با توجه به نتایج بدست آمده که در جدول ۸ نیز مشاهده می‌شود، تصمیم گرفته شد که ۱۳ ضریب MFCC در حالتیکه بانک فیلتر شامل ۲۴ فیلتر می‌باشد استخراج گردیده و سپس ضریب اول کپستروم از این بردار حذف شود و برداری با ۱۲ ویژگی برای ادامه آزمایشات مورد استفاده قرار گیرد.

جدول ۸- میانگین راندمان تعیین هویت گوینده بازاء ضرائب LPCC، LFCC و MFCC با توجه به حذف اولین ضریب کپستروم بر حسب درصد

| نوع ویژگی | LPCC | LFCC | MFCC |
|---|---|---|---|
| داده‌های آموزش | ۷۶/۸ | ۸۲/۴ | ۸۷/۴ |
| داده‌های آزمایش | ۵۲/۵۰ | ۵۵/۶ | ۶۸/۹ |

## ۳-۴ روشهای مقاوم‌سازی ویژگی‌ها در برابر نویز

### ۳-۴-۱ توالی اتوکورولیشن نسبی

همانطور که اشاره شد، در روشهای بهسازی گفتار سعی می‌گردد گفتار تمیز از گفتار نویزی شده تخمین زده شود. استفاده از توالی اتوکورولیشن نسبی باعث می‌شود که با فرض ثابت بودن نویز جمع‌شونده، اثر این نویز از گفتار نویزی‌شده حذف گردد. در این مقاله روش RAS پیاده‌سازی گردیده و ضرائب MFCC بجای گفتار اصلی از توالی اتوکورولیشن استخراج گردیده‌اند. اما نتایج حاصل، نشان می‌دهد که راندمان استفاده از ضرائب MFCC که از گفتار اصلی بدست آمده باشند بالاتر از هنگامی است که از این روش استفاده می‌شود (جدول ۹ را ببینید).

جدول ۹- میانگین راندمان تعیین هویت گوینده بااستفاده از روش توالی اتوکورولیشن نسبی بر حسب درصد

| ضرائب MFCC با استفاده از توالی اتوکورولیشن نسبی | ضرائب MFCC با حذف ضریب اول کپسترال | ضرائب MFCC بدون حذف ضریب اول کپسترال | بردار ویژگی |
|---|---|---|---|
| ۸۱/۶ | ۸۷/۴ | ۸۰/۶ | داده‌های آموزش |
| ۵۳/۶ | ۶۸/۹ | ۵۸/۶ | داده‌های آزمایش |

### ۳-۴-۲ تکنیک وزندهی خطی

از دیگر تکنیکهای مقابله با نویز و عوامل مزاحم روی خط تلفن می‌توان به روشهای تقویت پارامترها و حذف تاثیر کانال انتقال و نوع میکروفن اشاره کرد. در این روشها، ویژگیهای استخراج شده، وزن داده می‌شوند تا در برابر نویز جمع‌شونده، نویز کانولوشنال، اعوجاجات کانال انتقال و خصوصیات میکروفن تقویت گردند. لازم به ذکر است که در دادگان فارس‌دات تلفنی که SNR آنرا با مقادیر مختلف SNR تغییر داده‌ایم، اکثر عوامل مزاحم فوق وجود دارند.

از جمله روشهای مقابله با اثرات فوق می‌توان به تکنیک وزندهی خطی یا LW اشاره کرد. ایده مبنا در استفاده از روشهای وزندهی کپسترال، توجه به حساسیت ضرائب کپسترال مرتبه پائین نسبت به شیب طیفی و حساسیت ضرائب مرتبه بالا نسبت به نویز می‌باشد. در جدول ۱۰ می‌توان نتایج حاصل از پیاده‌سازی روش LW را برروی ضرائب MFCC و مقایسه آن با نتایج حاصل از ضرائب MFCC با حذف ضریب اول کپسترال مشاهده نمود.

جدول ۱۰- میانگین راندمان تعیین هویت گوینده با استفاده از روش وزندهی خطی با حذف ضریب اول کپسترال از بردار ویژگی بر حسب درصد

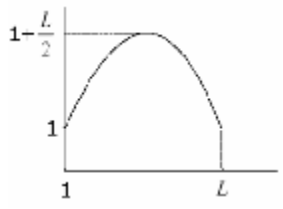
| ضرائب MFCC با اعمال روش وزندهی خطی | ضرائب MFCC با حذف ضریب اول کپسترال | بردار ویژگی |
|---|---|---|
| ۹۲/۴ | ۸۷/۴ | داده‌های آموزش |
| ۷۰/۷ | ۶۸/۹ | داده‌های آزمایش |

همانطور که در جداول ۱۰ مشاهده می‌شود، راندمان تعیین هویت با استفاده از این روش افزایش یافته و منجر به بهبود کارائی سیستم گردیده است. این بهبود در شرائطی که محیط آموزش و آزمایش یکسان نیست، کاملا مشخص می‌باشد. البته در شرائط یکسانِ محیط آموزش و آزمایش، شاهد کاهش راندمان تعیین هویت هستیم که این کاهش بسیار جزئی و قابل اغماض می‌باشد. بطورکلی اعمال روش وزندهی خطی برروی ضرائب MFCC به بهبود مقاومت سیستم در برابر نویز و مقابله با اثرات شرائط غیرمتعارف کمک بسزائی می‌نماید.

### ۳-۴-۳ تکنیک لیفتر کردن میانگذر

یکی دیگر از روشهای وزندهی کپسترال، تکنیک لیفتر کردن میانگذر یا BPL می‌باشد. پنجره $W_1(n)$ در این روش، بصورت شکل ۲ و رابطه زیر تعریف می‌گردد:

$$w_1(n)=\begin{cases}1+\frac{L}{2}\sin(\frac{np}{L}) & n=1,2,...,L\\ 0 & otherwise\end{cases} \qquad (۱۲)$$

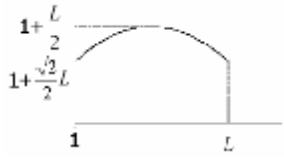



شکل۲- پنجره $W_1(n)$

جدول ۱۱ نتایج حاصل از راندمان تعیین هویت بازاء این روش و سایر روشهای لیفتر کردن میانگذر را نشان میدهد. در نتایج حاصل از ضرب این پنجره در ضرائب MFCC شاهد افزایش کارائی سیستم تعیین هویت گوینده هستیم. هرچند بهبود حاصل نسبت به اعمال روش LW کمتر می‌باشد.

در این مقاله، شکل پنجره $W_1(n)$ را بصورت پنجره $W_2(n)$ تغییر دادیم. نتیجه این تغییر در راندمان تعیین هویت در جدول ۱۱ دیده میشود. نتایج بدست آمده حاکی از بهبود راندمان تعیین هویت گویندگان نسبت به ضرائب MFCC است. این بهبود حتی از روش BPL اصلی نیز قابل توجه‌تر می‌باشد.

$$w_2(n)=\begin{cases}1+\frac{L}{2}\sin(\frac{(n+6)p}{24L}) & n=1,2,...,L \quad , \quad L=12\\ 0 & otherwise\end{cases} \qquad (۱۳)$$




شکل ۳- پنجره $W_2(n)$

در اصلاح دیگری شکل پنجره را بصورت شکل پنجره $W_3(n)$ تغییر دادیم:

$$w_3(n)=\begin{cases}1+\frac{L}{2}\sin(\frac{np}{2L}) & n=1,2,...,L\\ 0 & otherwise\end{cases} \qquad (۱۴)$$

شکل ۴- پنجره $W_3(n)$

نتایج بدست آمده بازهم حاکی از بهبود راندمان تعیین هویت گویندگان نسبت به ضرائب MFCC می باشد. از دیگر پنجره‌هایی که در این مقاله برای روش BPL پیشنهاد نمودیم، پنجره $W_4(n)$ با شکل ۵ و رابطه زیر می‌باشد. در جدول ۱۱ نتیجه استفاده از این پنجره نیز دیده میشود.

$$w_4(n)=\begin{cases}1+\frac{L}{2}\sin(\frac{np}{2L}+\frac{p}{2}) & n=1,2,...,L\\ 0 & otherwise\end{cases} \quad (۱۵)$$

شکل ۵- پنجره $W_4(n)$

نتایج حاصل از اعمال این پنجره، اگرچه از بهبود کارائی سیستم نسبت به ضرائب MFCC خبر می‌دهد ولی بهتر از نتایج حاصل از پنجره $W_2(n)$ نیست.

جدول ۱۱- میانگین راندمان تعیین هویت گوینده با استفاده از روش لیفتر کردن میانگذر بدون حذف ضریب اول کپسترال از بردار ویژگی بر حسب درصد

| نوع پنجره | $W_1(n)$ | $W_2(n)$ | $W_3(n)$ | $W_4(n)$ | ضرائب MFCC با حذف ضریب اول کپسترال |
|---|---|---|---|---|---|
| داده‌های آموزش | ۹۳/۰ | ۹۲/۸ | ۹۳/۰ | ۹۲/۲ | ۸۷/۴ |
| داده‌های آزمایش | ۷۰/۵ | ۷۰/۶ | ۷۰/۳ | ۷۰/۵ | ۶۸/۹ |

مقایسه نتایج موجود در جدول ۳، راندمان بالاتر پنجره $W_2(n)$ را بازاء داده‌های آزمایشی نشان می‌دهد. از آنجا که تاثیر تغییر در خصوصیات کانال انتقال و نوع میکروفن در ضرائب کپسترال مرتبه پائین بیشتر ظاهر می‌شود و تاثیر نویزهای جمع‌شونده در ضرائب مرتبه بالا ظاهر می‌گردد، لذا با توجه به اینکه لیفتر $W_2(n)$ منجر به کارائی بسیار خوبی نسبت به سایر لیفترها شده، می‌توان نتیجه گرفت که دادن وزن کمتر به ضرائب کپسترال مرتبه پائینتر و ضرائب کپسترال مرتبه بالاتر موجب حذف تاثیرات مزاحم فوق شده و در بهبود کارائی تعیین هویت بسیار موثر می‌باشد.

### ۳-۴-۴ روش تفاضل میانگین در حوزه کپسترال

روش CMS برای مقابله با اثر کانال انتقال و یا نویز کانولوشنال پیشنهاد شده است. با توجه به تلفنی بودن دادگان مورد استفاده در این مقاله، استفاده از این روش مورد بررسی قرار گرفت. نتایج حاصل از استفاده از این روش در جدول ۱۲ مشاهده می‌شود. این روش در محیط‌های آزمایش با نسبت سیگنال به نویز پائین، افزایش کارائی را منجر می‌شود؛ ولی در محیطهای با SNR بالا باعث کاهش کارائی سیستم تعیین هویت می‌گردد. ولی در مجموع با توجه به میانگینهای موجود در جدول ۱۲ این روش بهتر از ضرایب MFCC بدون اعمال روش تفاضل میانگین در حوزه کپسترال عمل نکرده است.

جدول ۱۲- میانگین راندمان تعیین هویت گوینده با استفاده از روش تفاضل میانگین در حوزه کپسترال با حذف ضریب اول کپسترال از بردار ویژگی بر حسب درصد

| ضرائب MFCC با اعمال روش CMS | ضرائب MFCC | بردار ویژگی |
|---|---|---|
| ۸۱/۸ | ۸۷/۴ | داده‌های آموزش |
| ۵۱/۵ | ۶۸/۹ | داده‌های آزمایش |

با توجه به اینکه در روش CMS میانگین ضرائب کپسترال از هر ضریب کم می‌شود، لذا تعداد ضرائب کپسترال که در محاسبه میانگین استفاده می‌شوند، مهم می‌باشد. لذا در این قسمت آزمایشات را بازاء گویشهای تست ۲ ثانیه‌ای و ۵ ثانیه‌ای انجام دادیم تا تاثیر طول داده‌های تست در میانگین‌گیری در روش CMS مشخص گردد. نتایج حاصل در جدول ۱۳ گویای آن است که افزایش حجم داده‌های تست منجر به تخمین بهتر میانگین ضرائب کپسترال و در نتیجه بهبود نتایج می‌شود.

جدول ۱۳- میانگین راندمان تعیین هویت گوینده با توجه به زمان گویشهای تست بدون وجود ضریب اول کپستروم بر حسب درصد

| زمان گویشهای تست | ۲ ثانیه | ۵ ثانیه |
|---|---|---|
| داده‌های آموزش | ۸۱/۸ | ۸۱/۸ |
| داده‌های آزمایش | ۵۱/۵ | ۶۱/۳ |

### ۳-۴-۵ تکنیک PFL

همانطورکه بیان گردید، PFL یک روش وزندهی یا لیفتر کردن ضرائب کپسترال می‌باشد که مبتنی بر این واقعیت عمل می‌کند که مقاومت در برابر نویز در نواحی فرمانت (پیکهای طیف) بیش‌از سایر نواحی است. همانطور که گفته شد، در این تکنیک دو پارامتر $a$ و $b$ وجود دارند که مقادیر متنوع این دو پارامتر در نتایج تعیین هویت گوینده تاثیرگذار می‌باشند. در این مقاله راندمان تعیین هویت بازاء مقادیر $a$، ۱ و ۰/۸ و مقدار $b$، ۰، ۰/۲۵، ۰/۵ و ۰/۷۵ بدست آمد که نتایج حاصل در جدول ۱۴ ارائه شده‌اند.

جدول ۱۴- میانگین راندمان تعیین هویت گوینده با استفاده از ویژگی MFCC و اعمال روش PFL بازاء مقادیر مختلف $a$ و $b$ در حالیکه ضریب اول کپستروم حذف شده و بر حسب درصد

| $a$ | ۱ | | | | ۰.۸ | | | | ضرائب MFCC |
|---|---|---|---|---|---|---|---|---|---|
| $b$ | ۰ | ۰/۲۵ | ۰/۵ | ۰/۷۵ | ۰ | ۰.۲۵ | ۰.۵ | ۰.۷۵ | |
| داده‌های آموزش | ۹۲/۶ | ۹۳/۰ | ۹۲/۶ | ۹۲/۲ | ۸۸/۲ | ۸۷/۸ | ۸۷ | ۶۶/۳ | ۸۷/۴ |
| داده‌های آزمایش | ۷۰/۶ | ۷۰/۵ | ۷۰/۶ | ۷۰/۶ | ۶۱/۳ | ۶۱/۸ | ۶۰/۵ | ۲۳/۱ | ۶۸/۹ |

همانطورکه مشاهده می‌شود، در شرائطی که مقدار $a$ را ۱ درنظر بگیریم، نتایج بر بهبود عملکرد سیستم تعیین هویت گوینده با اعمال روش PFL تاکید دارند. ولی با مقدار $a$ =۰/۸ با کاهش کارائی روبرو هستیم. نتایج این جدول گویای آن است که بازاء $a$ =۱ و $b$ = ۰/۷۵، بهترین کارائی بدست می‌آید.

### ۳-۴-۶ ضرائب دینامیکی کپسترال

استفاده از ضرائب دینامیکی کپسترال همراه با ویژگیهای اصلی در یک بردار ویژگی بر کارائی سیستم تعیین هویت گوینده خواهد افزود. جدول ۱۵ نتایج سیستم تعیین هویتی را نشان می‌دهند که بردار ویژگی آن شامل ۱۲ ضریب MFCC و ۱۲ ضریب دلتا MFCC می‌باشد و بهبود نتایج کاملا مشخص می‌باشد. همچنین استفاده از ضرائب دلتادلتا MFCC همراه با ضرائب MFCC و ضرائب دلتا MFCC در یک بردار ویژگی ۳۶تائی، یکی دیگر از راه‌هائی است که در تقویت پارامترهای MFCC در مقابل نویز موثر می‌باشد. جدول ۱۵ نتایج حاصل از این روش را نیز نشان می‌دهد. طبق جدول اضافه کردن مشتق دوم نسبت به حالتی که ضرائب MFCC همراه مشتق اول آنها استفاده می‌شوند، بهبودی را نشان نمی‌دهد. جمع‌بندی کلی از جدول فوق تاثیر مثبت اضافه کردن مشتق اول به ویژگیهای کپسترال در بهبود کارائی تعیین هویت می‌باشد. دلیل این موضوع اینست که ضرایب دینامیکی یا مشتقاتِ ضرایب کپسترال برای نشان‌دادن دینامیک مجـــرای

جدول ۱۵- میانگین راندمان تعیین هویت گوینده با استفاده از ویژگی MFCC و ضرائب دینامیکی کپسترال با حذف ضریب اول کپستروم بر حسب درصد

| بردار ویژگی | MFCC | Δ MFCC MFCC + | MFCC + Δ MFCC + ΔΔ MFCC |
|---|---|---|---|
| داده‌های آموزش | ۸۷/۴ | ۸۹/۲ | ۹۰/۲ |
| داده‌های آزمایش | ۶۸/۹ | ۷۰/۰ | ۶۹/۴ |

گفتار و نحوهٔ تغییرات زمانی آن مناسب و مفید هستند. در واقع ضرایب دلتا کپسترال و ضرایب دلتادلتا کپسترال سرعت و شتاب تغییرات مجرای گفتار را مدل می‌کنند و بکارگیری آنها در سیستم‌های بازشناسی راندمان این سیستم‌ها را بالا می‌برد.

### ۳-۴-۸ مقایسه روشهای پیاده‌سازی شده

جهت مقایسه بهتر روشهای پیاده‌سازی شده تا کنون، نتایج حاصل از استفاده از آنها در استخراج ضرائب MFCC و یا اعمال آنها بر ضرائب MFCC جهت تقویت این ویژگی‌ها در برابر نویز جمع‌شونده، نویز کانولوشنال، اعوجاجات کانال انتقال و خصوصیات میکروفن، در جدول ۱۶ جمع‌آوری شده است.

جدول ۱۶- میانگین راندمان تعیین هویت گوینده بازاء روش‌های مختلف مقاوم‌سازی با حذف ضریب اول کپستروم بر حسب درصد

| تکنیک مورد استفاده | MFCC | RAS | LW | BPL | CMS | PFL | Δ MFCC MFCC + |
|---|---|---|---|---|---|---|---|
| داده‌های آموزش | ۸۷.۴ | ۸۱.۶ | ۹۲.۴ | ۹۲.۸ | ۸۱.۸ | ۹۲.۲ | ۸۹.۲ |
| داده‌های آزمایش | ۶۸.۹ | ۵۳.۶ | ۷۰.۸ | ۷۰.۶ | ۵۱.۵ | ۷۰.۶ | ۷۰.۰ |

با توجه به جدول روشهای LW و PFL بهترین راندمان را بازاء داده های آزمایشی بین روشهای مورد بررسی دارا می‌باشند.

### ۳-۴-۹ روشهای ترکیبی

ترکیب روشهای مقابله با نویز و عوامل مزاحم روی خط تلفن می‌تواند در حذف اثرات نویز جمع‌شونده، کانولوشنال، اثرات کانال ارتباطی و نیز نوع میکروفن موثر باشد. همانطور که قبلا توضیح داده شد، استفاده از تکنیکهای تقویت پارامتر و حذف تاثیر کانال انتقال و نوع میکروفن می‌تواند کارائی سیستم تعیین هویت را افزایش دهد. معمولا ترکیب روشها، چون باعث تجلی نقاط قوت و پوشش نقاط ضعف روشهای مورد استفاده می‌شود، مفید می‌باشد. لذا در این مقاله تکنیکهای تقویت پارامتر و حذف تاثیر کانال انتقال و نوع میکروفن را جهت افزایش کارائی سیستم باهم ترکیب کردیم.

#### ۳-۴-۹-۱ ترکیب روش وزندهی خطی و ضرائب دینامیکی کپسترال

نتیجه ترکیب دو روش وزندهی خطی و ضرائب دینامیکی کپسترال از گروهِ تکنیکهای تقویت پارامتر و حذف تاثیر کانال انتقال و نوع میکروفن در یک سیستم تعیین هویت گوینده در جدول ۱۷ آمده است. بنابر راندمان‌های موجود در جدول، این ترکیب جدید نتایج قابل توجه‌ای را در بر دارد. روش ترکیبی در اکثر موارد بهتر از استفاده از روشهای وزندهی خطی و ضرائب دینامیکی کپسترال بتنهائی می‌باشد. ولی همانطور که در جدول ۱۷ ملاحظه می‌شود میانگین راندمان این روش ترکیبی از روش LW بتنهائی در مواردی که گویشهای آزمایش متفاوت با گویشهای آموزش در نظر گرفته می‌شوند، پائینتر می‌باشد اما هنگامی که همان گویشهای آموزش مورد آزمایش قرار می‌گیرند، روش ترکیبی راندمان بالاتری دارد. در این روش ترکیبی، یک بردار ویژگی ۲۴تائی شامل ۱۲ ضریب MFCC که روش وزندهی خطی بر آنها اعمال شده و ۱۲ ضریب دلتا MFCC، مورد استفاده قرار گرفته است.

جدول ۱۷- میانگین راندمان ترکیب دو روش وزندهی خطی و ضرائب دینامیکی کپسترال با حذف ضریب اول کپستروم بر حسب درصد

| روش ترکیبی | Δ MFCC MFCC + | LW | ضرائب MFCC | ضرائب MFCC |
|---|---|---|---|---|
| ۹۳/۱ | ۸۹/۲ | ۹۲/۴ | ۸۷/۴ | داده‌های آموزش |
| ۷۱/۰ | ۷۰/۰ | ۷۰/۷۶ | ۶۸/۹ | داده‌های آزمایش |

#### ۳-۴-۹-۲ ترکیب روش لیفتر کردن میانگذر و ضرائب دینامیکی کپسترال

روش ترکیبی دیگری که در این مقاله برای اولین بار آزمایش شد، ترکیب دو روش لیفتر کردن میانگذر و ضرائب دینامیکی کپسترال می‌باشد. نتایج بدست آمده که در جدول ۱۸ قابل ملاحظه است، حاکی از بهبود نتایج نسبت به اعمال روش لیفتر کردن میانگذر و یا استفاده از ضرائب دینامیکی کپسترال همرا با ضرائب MFCC به تنهائی است.

جدول ۱۸- میانگین راندمان ترکیب دو روش لیفتر کردن میانگذر و ضرائب دینامیکی کپسترال با حذف ضریب اول کپستروم بر حسب درصد

| روش ترکیبی | MFCC + Δ MFCC | BPL | ضرائب MFCC | |
|---|---|---|---|---|
| ۹۲/۶ | ۸۹/۲ | ۹۲/۸ | ۸۷/۴ | داده‌های آموزش |
| ۷۲/۴ | ۷۰/۰ | ۷۰/۶ | ۶۸/۹ | داده‌های آزمایش |

البته با استفاده از این ترکیب در محیط‌های آموزش و آزمایش با نسبت سیگنال به نویز بالا بهبود نتایج برجسته تر از بهبود نتایج در محیط‌های آموزش و آزمایش با نسبت سیگنال به نویز پائین می‌باشد.

#### ۳-۴-۹-۳ ترکیب روش PFL و ضرائب دینامیکی کپسترال

استفاده از روش PFL در بهبود کارائی سیستم تعیین هویتی که ضرائب MFCC را به عنوان ویژگی از گفتار ورودی استخراج و استفاده می‌نماید، بسیار موثر می‌باشد. در این مقاله از اعمال این روش بر ضرائب MFCC و ضرائب دلتا MFCC در یک بردار ویژگی استفاده کردیم. همانطور که در جدول ۱۹ مشاهده می‌شود، این ترکیب، بیشتر از هر یک از دو روش PFL و ضرائب دینامیکی کپسترال باعث بهبود نتایج تعیین هویت گوینده در شرائط نویزی می‌شود.

جدول ۱۹- میانگین راندمان ترکیب دو روش PFL و ضرائب دینامیکی کپسترال با حذف ضریب اول کپستروم بر حسب درصد

| روش ترکیبی | MFCC + Δ MFCC | PFL | ضرائب MFCC | ضرائب MFCC |
|---|---|---|---|---|
| ۹۲/۸ | ۸۹/۲ | ۹۲/۲ | ۸۷/۴ | داده‌های آموزش |
| ۷۱/۰ | ۷۰/۰ | ۷۰/۶ | ۶۸/۹ | داده‌های آزمایش |

در نهایت جهت مقایسه بهتر راندمان روشهای ترکیبی با یکدیگر در جدول ۲۰ میانگین راندمان تعیین هویت گوینده با استفاده از روشهای ترکیبی که تا کنون بیان شد بازاء داده‌های آموزش و آزمایش بر حسب درصد ارائه شده است.

جدول ۲۰- میانگین راندمان تعیین هویت گوینده با استفاده از روشهای ترکیبی با حذف ضریب اول کپستروم بر حسب درصد

| نوع ترکیب | راندمان برحسب درصد | |
|---|---|---|
| | داده‌های آموزش | داده‌های آزمایش |
| ترکیب روش وزندهی خطی و ضرائب دینامیکی کپسترال | ۹۳/۱ | ۷۱/۰ |
| ترکیب روش لیفتر کردن میانگذر و ضرائب دینامیکی کپسترال | ۹۲/۶ | ۷۰/۴ |
| ترکیب روش PFL با مقادیر ۱= $a$ و ۰/۷۵= $b$ و ضرائب دینامیکی کپسترال | ۹۲/۸ | ۷۱/۰ |

همانطور که در جدول فوق مشاهده می‌شود ترکیب روش وزندهی خطی و ضرائب دینامیکی کپسترال بهترین راندمان را ارائه می‌نماید و بعداز آن از ترکیب روش PFL با مقادیر ۱= $a$ و ۰،۷۵= $b$ و ضرائب دینامیکی کپسترال راندمان بالاتری نسبت به سایر روشها برخوردار است. در جدول ۲۱ نتایج حاصل از تعیین هویت بازاء بهترین ویژگی‌های مورد استفاده در این بخش اعم از ترکیبی و غیرترکیبی جهت مقایسه آورده شده است.

جدول ۲۱- میانگین راندمان تعیین هویت گوینده بهترین ویژگی و روشها با حذف ضریب اول کپستروم بر حسب درصد

| ویژگی یا روش | راندمان برحسب درصد | |
|---|---|---|
| | داده‌های آموزش | داده‌های آزمایش |
| ضرائب MFCC | ۸۷/۴ | ۶۸/۹ |
| روش وزندهی خطی | ۹۲/۴ | ۷۰/۸ |
| ترکیب روش وزندهی خطی و ضرائب دینامیکی کپسترال | ۹۳/۱ | ۷۱/۰ |
| ترکیب روش PFL با مقادیر ۱= $a$ و ۰/۷۵= $b$ و ضرائب دینامیکی کپسترال | ۹۲/۸ | ۷۱/۰ |

با توجه به جدول ۲۱ می‌توان نتیجه گرفت که اگر ضریب اول کپستروم را از بردار ویژگی حذف کرده سپس از روش ترکیبی حاصل از روش وزندهی خطی و ضرائب دینامیکی کپسترال استفاده نمائیم راندمان سیستم تعیین هویت گوینده در برابر نویز به میزان قابل توجهی افزایش خواهد داشت که نشان از مقاومت بهتر سیستم در برابر شرائط نویزی دارد. به همین شکل استفاه از ترکیب روش PFL با مقادیر مناسب برای دو پارامتر $a$ و $b$ و ضرائب دینامیکی کپسترال توصیه می‌گردد. نکته حائز اهمیت دیگری که می‌توان نتیجه گرفت آن است که معمولا راندمان تعیین هویت در زمانیکه SNR داده‌های آموزش و آزمایش بهم نزدیک هستند بالاتر از زمانی است که SNR داده‌های آموزش و آزمایش باهم اختلاف دارند، حتی اگر SNR داده‌های آموزشی بسیار بالا باشد.

## ۴- نتیجه‌گیری

بر اساس نتایج مربوط به استفاده از انواع ضرائب کپسترال مانند ضرائب LPCC، LFCC، MFCC و SSC در تعیین هویت گوینده برروی داده‌های تلفنی و نویزی، ضرائب MFCC از بالاترین راندمان در تعیین هویت گوینده برخوردار می‌باشند. در زمانی که SNR سیگنال گفتاری بالا است و از طرفی گفتار، تلفنی نیست و باند بالای فرکانسی توسط تلفن محدود نگشته است، ضرائب LPCC کارائی بهتری نسبت به ضرائب MFCC برای تعیین هویت گوینده دارند چراکه ضرائب LPCC از اطلاعات وابسته به گوینده، موجود در باند بالای فرکانسی به همان میزانی استفاده می‌نمایند که از اطلاعات موجود در باندهای پائین فرکانسی؛ درحالیکه رزولوشن و دقت فرکانسی در هنگام استخراج ضرائب MFCC در باندهای بالای فرکانسی کمتر از باندهای پائین می‌باشد. در محیط‌های نویزی از میزان پایداری محاسبات ضرائب LPCC کاسته می‌شود و لذا کارائی آن کم می‌شود، درحالیکه ضرائب MFCC بدلیل خاصیت میانگین‌گیری از نمونه‌های فرکانسی در هر فیلتر از بانک فیلتر، از خاصیت کاهش تاثیر نویز برخوردار می‌باشند.

ضریب اول کپستروم بیانگر انرژی فریم می‌باشد. از آنجا که انرژی فریم کمتر حاوی اطلاعات وابسته به گوینده می‌باشد لذا حذف آن از بردارهای ویژگی باعث بهبود کارائی ضرائب کپسترال می‌گردد. حذف نمونه‌های طیف فرکانسی خارج از محدوده بایستی متناسب با فرکانس نمونه‌برداری داده‌های صوتی باشد. توجه به این مسئله از حذف اطلاعات مفید در تعیین هویت گوینده جلوگیری می‌نماید. همچنین در حالتیکه فرکانس نمونه‌برداری از گفتار ۱۱۰۲۵ هرتز باشد، با توجه به نتایج بدست آمده، ضرائب MFCC که با استفاده از بانک فیلتر حاوی ۲۴ فیلتر استخراج گردیده‌اند، بهترین کارائی را از خود نشان می‌دهند. همانطور که بیان شد، در روشهای بهسازی گفتار سعی می‌گردد گفتار تمیز از گفتار نویزی شده تخمین زده شود. استفاده از توالی اتوکورولیشن نسبی باعث می‌شود که با فرض ثابت بودن نویز جمع‌شونده، اثر این نویز از گفتار نویزی‌شده حذف گردد. اما نتایج حاصل از پیاده‌سازی این روش، نشان می‌دهد که راندمان استفاده از ضرائب MFCC که از گفتار اصلی بدست آمده باشند بالاتر از هنگامی است که از این روش استفاده می‌شود.

در روشهای تقویت پارامترها و حذف تاثیر کانال انتقال و نوع میکروفن، ویژگیهای استخراج شده، دستکاری می‌شوند تا در برابر نویز جمع‌شونده، نویز کانولوشنال، اعوجاجات کانال انتقال و خصوصیات میکروفن تقویت گردند. از جمله روشهای مقابله با اثرات فوق می‌توان از تکنیک وزندهی خطی یا LW نام برد. ایده مبنا در استفاده از روشهای وزندهی کپسترال، توجه به حساسیت ضرائب کپسترال مرتبه پائین نسبت به تاثیر کانال و نوع میکروفن و حساسیت ضرائب مرتبه بالا نسبت به نویز می‌باشد. با توجه به نتایج اعمال روش وزندهی خطی برروی ضرائب MFCC به بهبود مقاومت سیستم در برابر نویز و مقابله با اثرات شرائط نویزی کمک بسزائی می‌نماید. اعمال این روش بر ضرائب MFCC به عنوان بهترین روش مقاوم‌سازی بین روشهای پیاده‌سازی شده، بطور میانگین باعث افزایش راندمان سیستم به میزان ۱،۸۵٪ بازاء داده های آموزشی در محیط‌های با سیگنال به نویز مختلف گردید.

یکی دیگر از روشهای وزندهی کپسترال، تکنیک لیفتر کردن میانگذر می‌باشد. همانطورکه گفته شد، در این روش از یک پنجره جهت وزندهی ضرائب کپسترال استفاده می‌شود. در نتایج حاصل از این روش لیفتر کردن شاهد افزایش کارائی سیستم تعیین هویت گوینده هستیم، هرچند بهبود حاصل نسبت به اعمال روش وزندهی خطی کمتر می‌باشد. از آنجا که تاثیر تغییر در خصوصیات کانال انتقال و نوع میکروفن در ضرائب کپسترال مرتبه پائین بیشتر ظاهر می‌شود و تاثیر نویزهای جمع‌شونده در ضرائب مرتبه بالا ظاهر می‌شوند، لذا می‌توان نتیجه گرفت دادن وزن کمتر به ضرائب کپسترال مرتبه پائینتر و ضرائب کپسترال مرتبه بالاتر موجب حذف تاثیرات مزاحم فوق شده و در بهبود کارائی تعیین هویت بسیار موثر می‌باشد.

روش PFL نیز یک روش وزندهی یا لیفتر کردن ضرائب کپسترال می‌باشد که مبتنی بر این واقعیت عمل می‌کند که مقاومت در برابر نویز در نواحی فرمانت (پیکهای طیف) بیش‌از سایر نواحی است. همانطور که گفته شد، در این تکنیک دو پارامتر $a$ و $b$ وجود دارند که مقادیر متنوع این دو پارامتر در نتایج تعیین

هویت گوینده تاثیرگذار می‌باشند. راندمان بدست آمده از ین روش وزندهی نیز حاکی از موثر بودن این روش در تعیین هویت گوینده در شرائط نویزی می‌باشد. با توجه به نتایج حاصل از روشهای وزندهی که در این مقاله مورد بررسی قرار گرفتند، مطالعه و ابداع روشهای وزندهی که بطور مناسب و با توجه به اثرات نویز به ضرائب کپسترال وزن می‌دهند، در افزایش راندمان و مقاومت سیستمهای تعیین هویت گوینده در برابر نویز بسیار موثر می‌باشد.

روش تفاضل میانگین در حوزه کپسترال برای مقابله با اثر کانال انتقال و یا نویز کانولوشنال پیشنهاد شده است. نتایج نشان می‌دهد که این روش در محیط‌های آزمایش با نسبت سیگنال به نویز پائین، افزایش کارائی را منجر می‌شود؛ ولی در محیطهای با SNR بالا باعث کاهش کارائی سیستم تعیین هویت می‌گردد. همچنین با توجه به اینکه در روش CMS میانگین ضرائب کپسترال از هر ضریب کم می‌شود، لذا تعداد ضرائب کپسترال که در محاسبه میانگین استفاده می‌شوند، مهم می‌باشد. نتایج حاصل گویای آن است که افزایش حجم داده‌های تست منجر به تخمین بهتر میانگین ضرائب کپسترال و در نتیجه بهبود نتایج می‌شود.

استفاده از ضرائب دینامیکی کپسترال همراه با ویژگیهای اصلی در یک بردار ویژگی بر کارائی سیستم تعیین هویت گوینده خواهد افزود. دلیل این موضوع اینست که ضرایب دینامیکی یا مشتقاتِ ضرایب کپسترال برای نشان‌دادن دینامیک مجرای گفتار و نحوهٔ تغییرات زمانیِ آن مناسب و مفید هستند. در واقع ضرایب دلتا کپسترال و ضرایب دلتادلتا کپسترال سرعت و شتاب تغییرات مجرای گفتار را مدل می‌کنند و بکارگیری آنها در سیستم‌های بازشناسی راندمان این سیستم‌ها را بالا می‌برد. نتایج حاکی از تاثیر مثبت اضافه کردن مشتق اول به ویژگیهای کپسترال در بهبود کارائی تعیین هویت می‌باشند.

ترکیب روشهای مقابله با نویز و عوامل مزاحم روی خط تلفن می‌تواند در حذف اثرات نویز جمع‌شونده، کانولوشنال، اثرات کانال ارتباطی و نیز نوع میکروفن موثر باشد. همانطور که قبلا توضیح داده شد، استفاده از تکنیکهای تقویت پارامتر و حذف تاثیر کانال انتقال و نوع میکروفن می‌تواند کارائی سیستم تعیین هویت را افزایش دهد. معمولا ترکیب روشها، چون باعث تجلی نقاط قوت و پوشش نقاط ضعف روشهای مورد استفاده می‌شود، مفید می‌باشد. ترکیب روشهای وزندهی خطی، لیفترکردن میانگذر و PFL با ضرائب دلتا کپسترال در یک بردار ویژگی در این مقاله انجام گرفت و نشان داد که روشهای ترکیبی عمدتا دارای راندمان بالاتری نسبت به روشهای سازنده آنها به تنهائی می‌باشند. بهترین نتیجه بدست آمده ترکیب ضرائب MFCC و اعمال روش وزندهی خطی یا روش PFL بر آنها در یک بردار ویژگی همراه با ضرائب دلتا MFCC می‌باشد که موجب افزایش ۵،۴٪ در میانگین راندمان سیستم بازاء داده های آموزشی و ۲/۴٪ بازاء داده های آزمایشی گردید.

اگر ضریب اول کپستروم را از بردار ویژگی حذف کرده سپس از روش ترکیبی حاصل از روش وزندهی خطی و ضرائب دینامیکی کپسترال استفاده نمائیم، سیستم تعیین هویت گوینده بهترین راندمان را در شرائط نویزی دارا می‌باشد. به همین شکل استفاده از ترکیب روش PFL با مقادیر مناسب برای دو پارامتر آن و ضرائب دینامیکی کپسترال توصیه می‌گردد. همچنین نتایج نشان می‌دهند که در مواردی که SNR داده‌های آموزش با داده‌های آزمایش یکسان است، راندمان تعیین هویت حداکثر مقدار خود را داراست. همچنین به نظر می‌رسد که در حالتیکه SNR داده‌های آموزشی بالاتر از SNR داده‌های آزمایشی است، راندمان بهتری نسبت به مواردی که SNR داده‌های آموزشی کمتر از SNR داده‌های آزمایشی باشد، بدست می‌آید. نکته حائز اهمیت دیگری که می‌توان نتیجه گرفت آن است که معمولا راندمان تعیین هـــویت در زمانیکه SNR داده‌های آمـــوزش و آزمایش بهم نزدیک هستند بالاتر از زمـــانی است که SNR داده‌های آمـــوزش و آزمایش باهم اختلاف دارند، حتی اگر SNR داده‌های آموزشی بسیار بالا باشد.

## قدردانی

این مقاله حاصل از پروژه تحقیقاتی طرح ماده ۱۰۲ به شماره ۱۵۱۷ می باشد که از طرف سازمان مدیریت و برنامه ریزی و وزارت فناوری اطلاعات و ارتباطات حمایت گردیده است.

## مراجع

[1] K. H. You, H. C. Wang, "Robust Features Derived from Temporal Trajectory Filtering for Speech Recognition under the Corruption of Additive and Convolutional Noises," *ICASSP*, pp. 577-580, 1998.

[2] K. K. Paliwal, "Spectral Subband Centroid Features for Speech Recognition", *ICASSP*, pp. 617-620, 1998.

[3] H. Hermansky and N. H. Morgan, "RASTA Processing of Speech," *IEEE Transaction on Speech and Audio Processing*, vol. 2, No. 4, 1994.

[4] B. H. Juang, L. R. Rabiner and J. Wolpon, "On the use of Band-pass Liftering in Speech Recognition," *IEEE Transaction on ASSP*, ASSP-35, pp. 947-954, 1987.

[5] B. S. Atal, "Effectiveness of Linear Prediction Analysis pf Speech Wave for Automatic Speaker Identification and Verification," *JASA*, vol. 55, pp. 1304-1312, 1974.

[6] V. Ramamoorthy, N. S. Jayant, R. V. Cox and M. M. Sondhi, "Enhancement of ADPCM Speech Coding with Backward-Adaptive Algorithms for Postfiltering and Noise Feedback," *Journal of Selected Areas in Communications*, vol. 6, no. 2, pp. 364-382, 1988.

[۷] ه. رزازان ، م. م. همایونپور، **تعیین هویت گوینده مجموعه باز: بررسی روشهای افزایش سرعت و دقت** ، پایان نامه شناسی ارشد، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر، ۱۳۸۲.

---

[1] Relative Autocorrolation Sequence
[2] Spectral Subband Centroid
[3] Liftrering
[4] Linear Weighting
[5] BandPass Liftering
[6] Cepstral Mean Subtraction
[7] Phonetically-Balanced
[8] Post Filter
[9] Likelihood
[10] Linear Prediction Cepstral Coefficients
[11] Linear Frequency Cepstral Coefficients

**محمد مهدی همایونپور** در سال ۱۳۳۹ در شهر شیراز متولد شد. تحصیلات تا مقطع دیپلم را در شهر شیراز سپری و دیپلم متوسطه خود را در سال ۱۳۵۸ دریافت کرد. وی تحصیلات خود در مقطع کارشناسی را در رشته مهندسی برق (الکترونیک) در دانشگاه صنعتی امیرکبیر (سال ۱۳۶۶)، کارشناسی ارشد را در رشته برق (مخابرات)، از دانشگاه خواجه نصیرالدین طوسی (سال ۱۳۶۹)، کارشناسی ارشد دوم خود را در زمینه فونتیک (۱۳۷۴) در دانشگاه سوربون جدید در فرانسه و همزمان دورهٔ دکترای خود را در دانشگاه پاریس ۱۱ در زمینه مهندسی برق (۱۳۷۴) بپایان رسانید. نامبرده از سال ۱۳۷۴ در سمت عضو هیأت علمی دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیر کبیر به تدریس و تحقیق مشغول می‌باشد. ایشان علاوه بر تدریس، راهنمایی پروژه های کارشناسی، کارشناسی ارشد و دکتری در زمینه های مهندسی کامپیوتر و فناوری اطلاعات و نیز هدایت تعداد زیادی پروژه های صنعتی و ملی را برعهده داشته است. نامبرده عضو انجمن های علمی کامپیوتر، ارتباطات و فناوری اطلاعات و رمز می باشد و مسئولیت های اجرایی متعدد از جمله ریاست و معاونت های آموزشی و پژوهشی دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیر کبیر و شرکت در برگزاری چندین کنفرانس و مسابقه علمی را بر عهده داشته و موفق به انتشار بیش از ۸۰ مقاله علمی- پژوهشی در مجلات و کنفرانس های علمی داخل و خارج از کشور گردیده است.

آدرس پست الکترونیکی نامبرده عبارتست از:

homayoun@ce.aut.ac.ir

آدرس سایت اینترنتی:

www.autice.com

**ابراهیم شریف نبوی** در سال ۱۳۵۷ در شهر مشهد متولد شد. تحصیلات تا مقطع دیپلم را در همان شهر سپری و دیپلم متوسطه خود را در سال ۱۳۷۵ دریافت کرد. وی تحصیلات خود در مقطع کارشناسی را در رشته مهندسی کامپیوتر-گرایش نرم‌افزار در دانشگاه آزاد اسلامی-واحد مشهد (سال ۱۳۷۶-۱۳۸۰) و مقطع کارشناسی ارشد را در رشته مهندسی کامپیوتر-گرایش معماری سیستمهای کامپیوتری در دانشگاه صنعتی امیرکبیر (سال ۱۳۸۰-۱۳۸۳) بپایان رسانید. نامبرده از سال ۱۳۸۳ در سمت کارشناس مرکز سیستمهای پایه در شرکت ایزایران به فعالیت مشغول است. زمینه های تخصصی مورد علاقه ایشان پردازش گفتار، متن‌بازها و سیستم‌عامل می‌باشد.