

بهبود روش MFCC برای تشخیص گوینده در عملیات جاسوسی

ابوالفضل چمن مطلق^۱، عباس نجفی پور^۲

تاریخ دریافت: ۹۲/۰۹/۱۲

تاریخ پذیرش: ۹۲/۱۰/۱۶

چکیده

در این مقاله، روشی جدید برای استخراج ویژگی از سیگنال صوت و تشخیص گوینده صدا ارائه شده است که کاربرد وسیعی در علوم جاسوسی و پدافند غیرعامل دارد (بدین صورت که با تشخیص گوینده یک صدای ضبط شده می توان اولاً مانع تقلید صدای افراد خاص مانند فرماندهان شد و ثانیاً می توان در استراق سمع مکانی که چندین نفر در حال گفتگوی مهمی هستند، گوینده هر صدا را تشخیص داد). در واقع روش ارائه شده، بهبود یافته روش MFCC می باشد. از آن جایی که با توجه به آزمایشات، بیشتر اطلاعات صوت در فرکانس های پایین آن ذخیره می شود و فرکانس های بالای آن اطلاعات تفکیک کننده زیادی ندارد، محققان معمولاً ویژگی های صوت را از فرکانس های پایین سیگنال صوت استخراج می کنند. در روش ارائه شده در این مقاله، فیلتر فرکانسی Mel که در روش MFCC استفاده می شود، تغییر و بهبود داده شده و در انتها نتایج روش ارائه شده (MMFCC) با روش MFCC مقایسه شده و بهبود آن اثبات شده است. نتایج حاصله برای ۲۰ گوینده متفاوت مؤید این است که برای روش بهبود یافته خطی درصد تشخیص گوینده حدود ۴/۵٪ و برای روش بهبود یافته نهائی حدود ۹٪ ارتقاء یافته است.

کلیدواژه ها: جاسوسی، پدافند غیرعامل، تشخیص گوینده، فیلتر Mel، روش MFCC، استخراج ویژگی

۱- استادیار و عضو هیئت علمی دانشگاه جامع امام حسین (ع)

۲- دانش آموخته کارشناسی ارشد دانشکده برق دانشگاه علم و صنعت ایران - نویسنده مسئول

۱- مقدمه

امروزه تشخیص گوینده صدا کاربرد وسیعی در علوم جاسوسی، پدافند غیرعامل و همچنین سیستم‌های امنیتی دارد. برخی از این کاربردها به شرح ذیل می‌باشد.

■ کاربرد در علوم جاسوسی و استراق سمع: بدین صورت که در عملیات استراق سمع می‌توان شخصی را که در هر لحظه صحبت می‌کند شناسایی کرده و نتیجه‌گیری و برداشت درست‌تری از صحبت‌ها داشت.

■ کاربرد در پدافند غیرعامل: دفاع غیرعامل به مجموعه اقداماتی اطلاق می‌گردد که به استفاده از جنگ‌افزار نیاز ندارد و با اجرای آن می‌توان از وارد شدن خسارات به تجهیزات و تأسیسات حیاتی و حساس نظامی و غیرنظامی و تلفات انسانی جلوگیری نموده و یا میزان این خسارات و تلفات را به حداقل ممکن کاهش داد. از آنجایی که صدای هر فرد مانند اثر انگشت او دارای مجموعه مشخصات منحصر به فردی است، چنانچه بتوان این ویژگی‌های منحصر به فرد را استخراج کرد، بستری ایجاد می‌شود که دشمنان توانایی تقلید صدای افراد مهم از جمله فرماندهان و... را نداشته باشند و لذا از انتشار فرامین باطل جلوگیری شود.

■ کاربرد در سیستم‌های امنیتی: جهت کنترل تردد به مناطق حساس و امنیتی می‌توان از صدای افراد و تشخیص گوینده آن استفاده کرد تا تنها افرادی بتوانند به آن مناطق وارد شوند که صدایشان در سیستم کنترل تردد ضبط شده باشد.

در خصوص جایگاه تشخیص گوینده صوت در علوم نظامی می‌توان گفت که هم کاربرد دفاعی برای آن وجود دارد و هم کاربرد جاسوسی (که هم می‌تواند دفاعی باشد و هم غیر دفاعی). در کاربرد دفاعی، تشخیص گوینده در زمره علوم دفاع غیرعامل (پدافند غیرعامل) می‌باشد، زیرا بدون استفاده از جنگ‌افزار می‌توان با به‌کار بستن آن مانع برخی از صدمات به نیروهای خودی شد.

استخراج ویژگی‌های متمایز کننده از سیگنال صوت، کاری است که اغلب روش‌های تشخیص گوینده از آن استفاده می‌کنند. برای استخراج مشخصات تأثیرگذار از سیگنال صوتی معمولاً هر فریم از آن دریافت شده و به حوزه فرکانس برده می‌شود و بر طبق ایده‌ای خاص، اطلاعاتی مهم و تأثیرگذار از آن سیگنال صوتی استخراج می‌شود که در عین حال بتواند شامل اطلاعاتی باشد که گوینده آن سیگنال را به خوبی شرح داده و هرگونه اطلاعات اضافی را حذف کند.

منبع تحریک سیگنال صدا دار، یک فشار هوای پریودیک است که می‌توان آن را به صورت ایده‌آل یک قطار ضربه در نظر گرفت. فرکانس این قطار ضربه pitch نام دارد که یک مشخصه مربوط به

گوینده است. یعنی pitch در یک سیگنال صوتی برای یک گوینده در کلمات مختلف از گفتار او یکی است [۱].

فرکانس Formant نیز از دیگر خصوصیات است که از سیگنال صوت استخراج می‌شود و وابسته به گوینده است [۱].

طیف زمان کوتاه^۱ نیز یک مشخصه سه‌بعدی از صوت است که بعدها آن، فرکانس انرژی و زمان است. در عمل این طیف همان طیف سیگنال یا به عبارت دیگر، توان دوم اندازه تبدیل فوریه یک سیگنال در بازه‌های کوچک زمانی است که در هر یک از بازه‌ها برداری به دست می‌آید که انرژی را تابعی از فرکانس نشان می‌دهد. لازم به ذکر است که بعد زمان هم در بازه‌های مختلف آن پنهان است یعنی هر یک از بازه‌های کوچکی که از آن استخراج می‌شود نماینده زمانی خاص هستند. باید یادآوری شود که فرکانس Formant در این طیف خود را به صورت بیشینه‌های محلی نشان می‌دهد [۱].

همبستگی طیفی^۲ سیگنال صوت نیز میزان تشابه فرکانسی مربوط به صوت یک گوینده را با صوت خودش نشان می‌دهد که در گویندگان مختلف متفاوت است و در برخی از مقالات از آن استفاده شده است [۲].

روش‌های LPCC^۳ و یا LPC^۴ از روش‌های رایج دیگر در حوزه تشخیص گوینده هستند که در آنها آنالیز با ترکیب اطلاعات مختلف موجود در صوت مانند pitch و فرکانس Formant و ... صورت می‌گیرد. این آنالیز مانند یک فیلتر تمام قطب (فیلتری که تابع تبدیل آن فقط قطب دارد و صفر ندارد) دیجیتال است. در این آنالیز نمونه k ام سیگنال صوتی به وسیله p نمونه قبل آن پیشگویی می‌شود که p همان مرتبه آن فیلتر است [۱-۳].

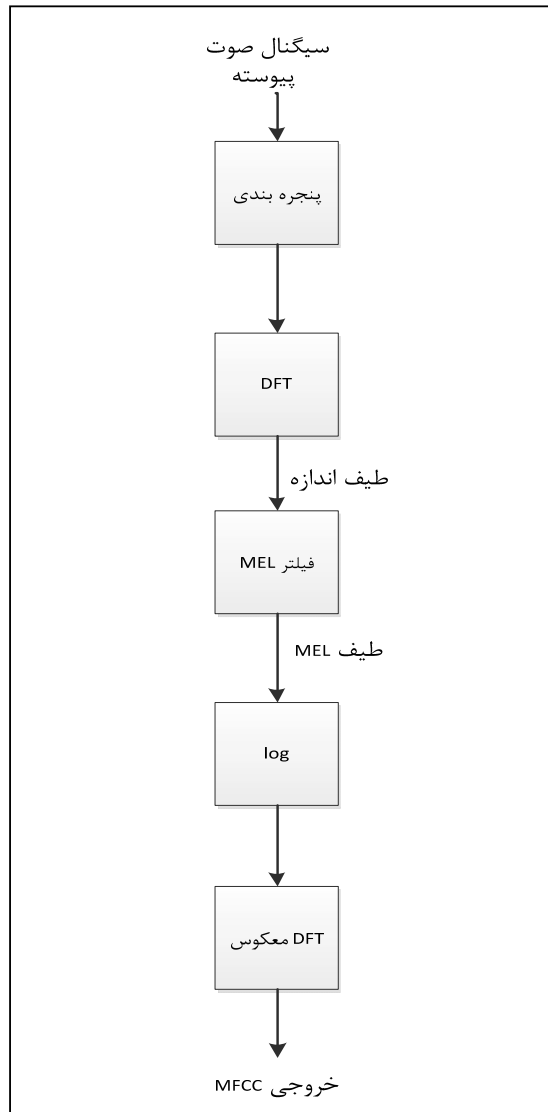
روش دیگری که بر پایه استخراج ویژگی به تشخیص گوینده می‌پردازد، روش MFCC^۵ است که در آن اندازه تبدیل فوریه سیگنال صوت در فیلتر Mel ضرب می‌شود [۱].

تشخیص گوینده معمولاً در چهار مرحله انجام می‌شود که عبارت‌اند از [۴ و ۵]:

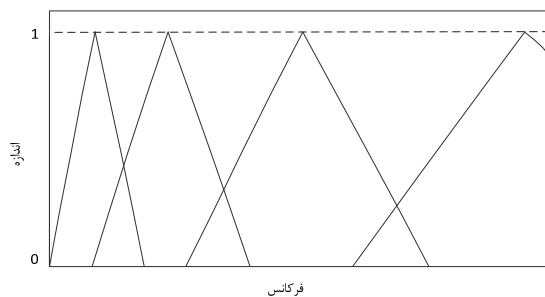
- پنجره‌بندی سیگنال صوت
- پیش‌پردازش
- تحلیل اصلی
- مدل‌سازی و تطبیق

در این مقاله، پس از پنجره‌بندی و پیش‌پردازش سیگنال صوت، از روش MFCC به‌عنوان تحلیل اصلی استفاده شده است و پیشنهادی نیز برای بهبود این روش داده می‌شود. ادامه مقاله به این صورت سازماندهی شده است:

1- Short Time Spectrum
2- Spectral Correlation
3- Loop Periodical Central Capacity
4- Loop Periodical Capacity
5- MEL- Frequency Cepstrum Coefficients



شکل ۱- مراحل روش MFCC



شکل ۲- بانک فیلتر MEL [۱]

در بخش دوم، روش MFCC تشریح شده است. بخش سوم روش پیشنهادی را معرفی می‌کند و در بخش چهارم نحوه شبیه‌سازی تشریح شده است. پس از آن در بخش پنجم نتایج شبیه‌سازی و سپس جمع‌بندی مقاله آورده شده است.

۲- روش MFCC

در تحقیق‌های انجام‌شده در زمینه سیگنال صوتی، محققان به این موضوع دست یافتند که در یک سیگنال صوتی، اطلاعات تأثیرگذارتر سیگنال، بیشتر در فرکانس‌های پایین وجود دارد و در نتیجه، برای دستیابی به اطلاعات مفیدتر از سیگنال، باید روی این قسمت از سیگنال تأکید بیشتری انجام گیرد. این ایده روشی را به نام MFCC به وجود آورده است که یکی از مهم‌ترین و پرکاربردترین روش‌ها در حوزه مطالعات تشخیص گوینده صدا می‌باشد [۱].

در این آنالیز ابتدا اندازه تبدیل فوریه پنجره داده مورد نظر از سیگنال صوت محاسبه می‌شود. سپس، اندازه تبدیل فوریه در یک فیلتر بانک به نام Mel ضرب می‌شود و تعدادی ضرب (که به تعداد فیلترهای ما بستگی دارد) استخراج می‌شود. سپس لگاریتم ضرایب به دست آمده و با استفاده از یکی از روش‌های کاهش ابعاد، تعداد کمتری ضریب از آن استخراج می‌شود.

فیلتر بانک Mel (که در ادامه توضیح داده خواهد شد) تأثیر فرکانس‌های پایین سیگنال صوت را در ویژگی‌های استخراجی از صوت بیشتر می‌کند و فرکانس‌های بالاتر را بیشتر تضعیف کرده و تأثیر آنها را در ویژگی‌های استخراجی کم می‌کند. لازم به ذکر است که دلیل استفاده از رابطه Mel این است که این رابطه همان رابطه شنوایی گوش انسان است.

شکل (۱) بلوک دیاگرام کلی روش MFCC را نشان می‌دهد.

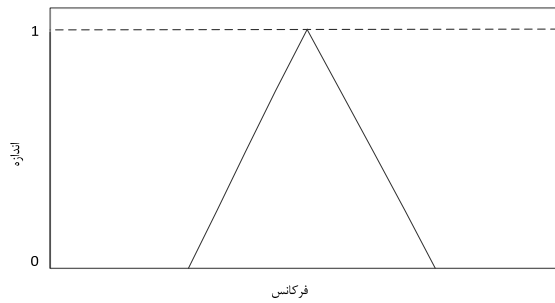
بنابراین طیف فرکانسی سیگنال صوت در بانک فیلتر Mel ضرب می‌شود. اگر سیگنال صوت به صورت $x(n)$ نمونه‌برداری شود، تبدیل فوریه گسسته (DFT) آن به صورت زیر خواهد بود:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi nk}{N}}, \quad k = 0, 1, \dots, N-1 \quad (1)$$

که در آن، N تعداد نمونه‌ها در هر پنجره سیگنال می‌باشد.

بانک فیلتر Mel نیز به صورت زیر می‌باشد:

که همان‌طور که دیده می‌شود، تمرکز بیشتری در فرکانس‌های پایین دارد. فیلتر بانک Mel به گونه‌ای است که تعداد فیلترهای موجود در آن در فرکانس‌های پایین، بسیار بیشتر از تعداد فیلترهای موجود در فرکانس بالا می‌باشد.

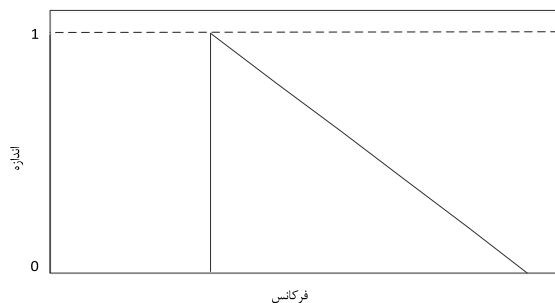


شکل ۳- فیلتر Mel [۱]

که همان طور که دیده می شود بیشترین بهره فیلتر در فرکانس میانی آن می باشد.

۳- روش پیشنهادی (MMFCC)

همان طور که دیده می شود، بیشترین بهره فیلتر Mel در فرکانس مرکزی فیلتر است. همان طور که گفته شد بیشتر اطلاعات متمایزکننده سیگنال صوت در فرکانس های پایین آن ذخیره شده است. بنابراین در اینجا پیشنهاد می شود که از فیلترهایی به صورت شکل (۴) برای بانک فیلتری استفاده شود تا در هر بازه فرکانسی، فرکانس های پایین تر بهره بیشتری داشته و در ضرایب استخراجی مؤثرتر باشند.



شکل ۴- فیلتر پیشنهادی با تابع خطی

بنابراین، بانک فیلتری پیشنهادی به صورت شکل (۵) خواهد بود (همان طور که در شکل دیده می شود، فیلترهای استفاده شده در فرکانس های بالاتر دارای پهنای باند کمتر هستند تا تأثیر فرکانس های بالاتر در ویژگی های استخراجی کمتر شود). همان طور که دیده می شود بیشترین بهره فیلترهای استفاده شده در بانک فیلتری، مربوط به پایین ترین فرکانس در بازه مربوطه می باشد. هر یک از فیلترها را می توان به صورت زیر طراحی نمود.

این فیلتر بانک به ازای هر فیلتر موجود در آن یک ضریب استخراج می کند، به این صورت که فیلتر مورد نظر در سیگنال ضرب شده و انرژی سیگنال به دست آمده محاسبه می شود و به عنوان یک ضریب نوشته می شود. پس در نتیجه به تعداد فیلترهای موجود در فیلتر بانک، ضریب استخراج می شود.

با توجه به اینکه تعداد فیلترهای این فیلتر بانک در فرکانس های کم، بسیار بیشتر از فرکانس های بالا است، بنابراین تعداد ضرایب استخراج شده از فرکانس های کم بیشتر شده و این موضوع باعث تأکید بیشتر روی فرکانس های کم می شود.

هر یک از فیلترهای بانک فیلتر Mel را می توان به صورت زیر ساخت [۵]:

$$H_i(k) = \begin{cases} 0 & \text{for } k < f_{b_{i-1}} \\ \frac{k - f_{b_{i-1}}}{f_{b_i} - f_{b_{i-1}}} & \text{for } f_{b_{i-1}} \leq k \leq f_{b_i} \\ \frac{f_{b_{i+1}} - k}{f_{b_{i+1}} - f_{b_i}} & \text{for } f_{b_i} \leq k \leq f_{b_{i+1}} \\ 0 & \text{for } k > f_{b_{i+1}} \end{cases} \quad (2)$$

که در آن، i شماره فیلتر است و:

$$f_b(i) = \left(\frac{N}{F_s} \right)^{\frac{1}{M}} \left(f_{mel}(f_{low}) + j \cdot \frac{\hat{f}_{mel}(f_{high}) - \hat{f}_{mel}(f_{low})}{M+1} \right) \quad (3)$$

$$f_{mel} = 1127 \cdot \ln \left(1 + \frac{f_{lin}}{700} \right) \quad (4)$$

$$\hat{f}_{mel}^{-1} = 700 e^{\frac{\hat{f}_{mel}}{1127}} - 1 \quad (5)$$

با داشتن این بانک فیلتری، ضرایب MFCC به صورت زیر محاسبه می شوند:

$$C_j = \sum_{i=1}^M X_i \cdot \cos \left(\left(j \cdot (i - 0.5) \cdot \frac{\pi}{M} \right) \right); j = 1, 2, \dots, J \quad (6)$$

که M تعداد فیلترها در بانک فیلتری است و J تعداد ضرایب استخراجی است. X_i نیز به صورت لگاریتمی زیر محاسبه می شود:

$$X_i = \log \left(\sum_{k=0}^{N-1} |X(k)| \cdot |H_i(k)| \right); i = 1, 2, \dots, M \quad (7)$$

هر یک از فیلترهای Mel استفاده شده در بانک فیلتری به صورت زیر می باشد:

$$H_i(k) = \begin{cases} 0 & \text{for } k < f_{b_{i-1}} \\ e^{\frac{-s}{f_{b_{i+1}} - f_{b_{i-1}}}(k - f_{b_{i-1}})} & \text{for } f_{b_{i-1}} \leq k \leq f_{b_{i+1}} \\ 0 & \text{for } k > f_{b_{i+1}} \end{cases} \quad (9)$$

۴- شبیه‌سازی

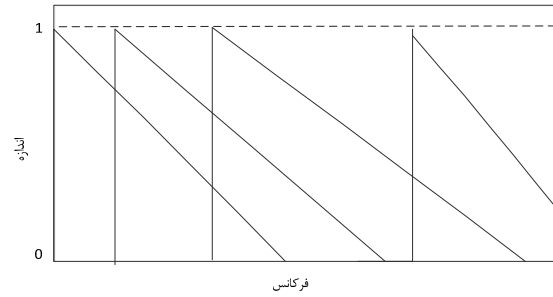
برای شبیه‌سازی روش MFCC و همچنین روش پیشنهادی MMFCC، از بانک داده و شبکه عصبی به شرح ذیل استفاده شده است (لازم به ذکر است که تمامی شبیه‌سازی‌ها در toolbox نرم‌افزار Matlab به نام Neural Networks انجام شده است):

۴-۱- بانک داده

در این مقاله، از بانک داده TIMIT جهت شبیه‌سازی‌ها و آزمایش روش‌های پیشنهادی استفاده شده است. این بانک داده از معتبرترین بانک‌های داده مورد استفاده در مقالات پردازش صوت و تشخیص گوینده است. در شبیه‌سازی‌ها، ۱۰ جمله مختلف که توسط ۲۰ گوینده متفاوت به زبان انگلیسی گفته و ضبط شده‌اند، مورد استفاده قرار گرفته است. مجموعاً ۲۰۰ صوت ضبط شده است که مدت زمان هر یک از صوت‌های ضبط‌شده کمتر از ۲۰ ثانیه می‌باشد و ۱۰ جمله گفته‌شده توسط هر گوینده با گوینده دیگر متفاوت است. گوینده‌ها همگی مذکر بوده و میانگین سنی ۳۳ سال دارند. از آنجایی که تشخیص بین صدای کودک و بزرگسال و همچنین بین مؤنث و مذکر ساده‌تر از تشخیص بین صدای دو مرد یا دو بزرگسال از هم است، پس می‌توان گفت که بانک داده استفاده‌شده، شرایط نسبتاً سختی را برای تشخیص گوینده ایجاد کرده است. در بانک داده استفاده‌شده، اطلاعاتی در مورد SNR صوت‌های ضبط‌شده داده نشده است ولی همه آنها در یک محیط و توسط یک دستگاه ضبط شده‌اند. از ۱۰ جمله‌ای که توسط هر فرد بیان شده است، ۷ جمله آن برای آموزش و ۳ جمله برای تست شبکه عصبی استفاده شده‌اند (یعنی مجموعاً ۱۴۰ جمله برای آموزش و ۶۰ جمله برای تست).

۴-۲- شبکه عصبی

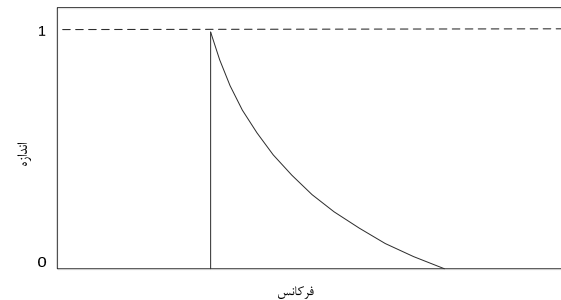
در این مقاله از شبکه عصبی سه‌لایه back propagation استفاده شده است که لایه اولی دارای ۷ نورون (نمایشگر تعداد ویژگی‌های استخراجی از هر پنجره داده)، لایه پنهان دارای ۱۷ نورون و لایه خروجی دارای ۲۰ نورون است، که برابر تعداد گویندگان می‌باشد. تابع استفاده شده برای لایه ورودی و لایه پنهان، تابع tansig بوده است که دارای مقداری بین -۱ تا ۱ می‌باشد. تابع مورد استفاده در لایه خروجی، logsig می‌باشد که مقدارش بین ۰ و ۱ متغیر است. مقدار خروجی هر نورون در لایه خروجی، نمایانگر احتمال این است که صوت مربوط به گوینده متناظر با آن نورون باشد. روش استفاده شده برای آموزش این شبکه عصبی، روش LM است.



شکل ۵- بانک فیلتر پیشنهادی با تابع خطی

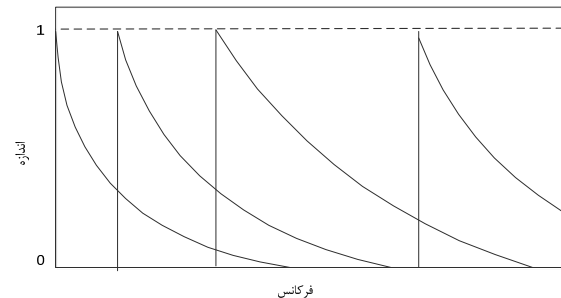
$$H_i(k) = \begin{cases} 0 & \text{for } k < f_{b_{i-1}} \\ \frac{f_{b_{i+1}} - k}{f_{b_{i+1}} - f_{b_{i-1}}} & \text{for } f_{b_{i-1}} \leq k \leq f_{b_{i+1}} \\ 0 & \text{for } k > f_{b_{i+1}} \end{cases} \quad (8)$$

که در آن، $i = 1, 2, \dots, Q$ و Q تعداد فیلترهاست. برای اینکه باز هم تأثیر فرکانس‌های پایین‌تر در هر بازه فرکانسی بیشتر دیده شود، می‌توان فیلتر را به صورت نمایی شکل (۶) تعریف کرد.



شکل ۶- فیلتر پیشنهادی با تابع نمایی

در این حالت، بانک فیلتری به صورت شکل (۷) خواهد بود:



شکل ۷- بانک فیلتر پیشنهادی با تابع نمایی

که برای طراحی آن می‌توان از فرمول زیر استفاده کرد:

می‌توان از انتشار فرامین غلط با تقلید صوت و همچنین ورود جاسوسان به مناطق امنیتی جلوگیری نموده و یا میزان اشتباه را به حداقل ممکن کاهش داد.

الگوریتم‌های تشخیص گوینده، علاوه بر کاربرد دفاعی در پدافند غیرعامل، کاربردهای غیر دفاعی در استراق سمع در عملیات جاسوسی و همچنین کاربردهای غیر نظامی در کنترل تردد دارند.

در این مقاله، روشی برای بهبود الگوریتم روش MFCC به منظور تشخیص صدا پیشنهاد شده است. نتایج حاصله برای ۲۰ گوینده متفاوت مؤید آن است که برای روش بهبودیافته خطی، درصد تشخیص گوینده حدود ۴/۵٪ و برای روش بهبود یافته نمایی، حدود ۹٪ بهبود یافته است.

مراجع

- Petersen, Richard, "Mel Frequency Cepstral Coefficients: An Evaluation of Robustness of MP3 Encoded Music", Journal of Informatics and Mathematical Modeling, Technical University of Denmark, Denmark, (2002).
- Politecnico di Bari, " Frame Length Selection in Speaker Verification Task", Human-Machine Interaction Systems Conf., Dept of Electrical and Electronic Engineering, (2008).
- Rajesh M. Hegde, Hema A. Murthy, R. Gadde., "Significance of Joint Features Derived from the Modified Group Delay Function in Speech Processing", IEEE Trans. Signal Process., vol. 14, no. 2, (2010).
- TomiKinnunen, "Spectral Features for Automatic Text-Independent Speaker Recognition", PhD Thesis, University of Joensuu, Department of Computer Science, Finland, (2003).
- MuzhirShabanAl-Ani&Thabit Sultan Mohammed & M. Aljebory, "Speaker Identification: A Hybrid Approach Using Neural Networks and Wavelet Transform". Journal of Computer Science, (2007).
- Réda&BoukelifAoued, "Artificial Neural Network & Mel-Frequency Cepstrum Coefficients-Based Speaker Recognition", IEEE Trans. Signal Process., vol. 7, no. 5, pp.554-568, (2009).
- Todor, Ganchev., Nikos Fakotakis and George Kokkinakis, "Comparative Evaluation of Various MFCC Implementations on the Speaker Verification Task", IEEE Trans. Signal Process., vol. 17, no. 13, (2011).

۵- مقایسه نتایج حاصل از شبیه‌سازی‌ها

جدول زیر نتایج (درصد موفقیت در تشخیص گوینده) حاصل از شبیه‌سازی سه روش را با هم مقایسه می‌کند:

جدول ۱- مقایسه شبیه‌سازی‌ها

تعداد گویندگان	روش استخراج ویژگی	درصد تشخیص
۲۰	MFCC	۸۳٪
۲۰	MMFCC با تابع خطی	۸۷/۵٪
۲۰	MMFCC با تابع نمایی	۹۱/۳٪

همان‌طور که مشاهده می‌شود درصد موفقیت در تشخیص گوینده، در روش MFCC، ۸۳٪ بوده است که این مقدار برای روش پیشنهادی MMFCC با تابع خطی به ۸۷/۵٪ (۴/۵ درصد بهبود) و برای تابع نمایی به ۹۱/۳٪ (تقریباً ۹ درصد بهبود) رسیده است که پیشرفت قابل ملاحظه‌ای مشاهده می‌شود.

البته طبق جدول فوق، در همه موارد درصد تشخیص گوینده، پایین به نظر می‌رسد که این امر به دلیل زیر است: در این مقاله از روش‌های مستقل از متن یا text independent استفاده شده است. این بدین معنی است که شبکه عصبی به کار رفته فقط با یک متن گفته شده توسط افراد مختلف آموزش داده نشده است، بلکه جملات مختلفی جهت آموزش به کار گرفته شده‌اند (یعنی ۱۰ جمله‌ای که توسط هر شخص گفته شده با ۱۰ جمله گفته شده با شخص دیگر متفاوت بوده است) و این امر سبب سخت‌تر شدن تشخیص گوینده می‌شود. درحالی که در بسیاری از مقالات، افراد مختلف تنها یک جمله ثابت را می‌گویند و بنابراین تشخیص گوینده بسیار ساده‌تر و درصد تشخیص هم بهتر می‌شود.

۶- جمع‌بندی و نتیجه‌گیری

همان‌طور که گفته شد، با استفاده از روش‌های تشخیص گوینده صوت، می‌توان بستر امنیتی مناسبی برای دفاع غیرعامل (پدافند غیرعامل) ایجاد کرد، بدین معنی که بدون استفاده از جنگ‌افزار یا تجهیزات نظامی و با اجرای الگوریتم‌های تشخیص گوینده صدا

Improving MFCC Method for Speaker Recognition in Spying Operations

A. CHaman Motlagh¹

A. Najafi Pour²

Abstract

In this article, a new method is introduced for extracting features from vocal signals and discriminating speakers, which has a vast application in spy operations and passive defense sciences (for example; not only by using this method, it is possible to prevent the imitation of particular people's voice such as commanders but also the detection of every speakers' voice when eavesdropping the place where a number of people are discussing an important matter can be made possible). The introduced method is the improved version of MFCC method. Experimental studies say that the most useful information of vocal signals is in their low frequencies and their high frequencies are not useful in speaker recognition procedures. The method introduced in this article, the Mel frequency filter which is used in MFCC, is changed and improved. The results of MFCC and the proposed method are compared for 20 speakers and the speaker recognition percentage has improved approximately 4.5% for the linear and 9% for the exponential proposed methods.

Key Words: *Spying, Passive Defense, Speaker Recognition, Mel Filter, MFCC Method, Feature Extraction*

1- Assistant Professor and Academic Member, Imam Hussein Comprehensive University

2- M.S in Electrical Engineering, Faculty of Electrical Engineering, Tehran Science and Technology - Writer in Charge