مجله علمی پژوهشی «علوم و فناوری‌های پدافند نوین»
سال نهم، شماره ۴، زمستان ۱۳۹۷؛ ص ۴۴۸-۴۴۱

ارائه الگوریتم جدید مبتنی بر مدل مخلوط گوسی با استفاده از ویژگی‌های ضرایب کپسترال نرمالیزه شده توانِ بر مبنای فیلتر کاکلی در سیستم تصدیق هویت گوینده

جعفرخلیل‌پور[1*]، اسماعیل زارع‌زاده[2]

۱- استادیار دانشگاه خاتم الانبیاء(ص) آجا، ۲- دانشجو دکتری دانشگاه صنعتی امیرکبیر

(دریافت: ۹۶/۰۶/۲۶، پذیرش: ۹۶/۱۱/۰۳)

چکیده

در این مقاله، یک الگوریتم استخراج ویژگیِ مبتنی بر سیستم شنوایی، بر اساس یک تبدیل زمانی- فرکانسی به نام تبدیل شنوایی (AT) و ضرایب کپسترال نرمالیزه شده توان(PNCC)، که یک ویژگی موفق در زمینه تشخیص گفتار و گوینده بوده است، پیشنهاد می‌گردد. به طور معمول عملکرد مدل‌های صوتی که توسط داده‌های بدون نویز(تمیز) آموزش داده می‌شوند، وقتی در شرایط نویزی مورد آزمایش قرار می‌گیرند به طور فزاینده‌ای کاهش می‌یابد. ویژگی پیشنهادی که ضرایب کپسترال نرمالیزه شده توانِ مبتنی بر فیلتر کاکلی (CFPNCC) نامیده می‌شود تحت چنین شرایطی مقاومت بالایی را از خود بروز می‌دهد. ویژگی بارز الگوریتم پیشنهادی ترکیب مزیت‌های فیلتر بانک کاکلی با مزایای ویژگی PNCC است که مقاومت توأم در مقابل نویزهای ایستان و غیر ایستان را به همراه دارد. به‌گونه‌ای که آزمایش‌های انجام‌شده بر روی پایگاه دادگان استاندارد SSC نشان می‌دهد، در سیستم تصدیق گوینده مبتنی بر مدل مخلوط گوسی، این ویژگی بهتر از ویژگی PNCC عمل می‌کند و به‌طور کلی نسبت به سایر ویژگی‌های متداول در زمینه تشخیص گوینده مانند MFCC و RASTA-PLP در شرایط نویزی نرخ خطای پایین‌تری را داراست.

کلیدواژه‌ها: تصدیق گوینده، استخراج ویژگی، مقاوم به نویز

Presenting A New Algorithm Based on GMM-UBM With Cochlear Filter- PNCC Feature for Speaker Verification

J. Khalilpour*, E. Zarezadeh

Khatam al-Anbia University

(Received: 17/09/2017; Accepted: 23/01/2018)

Abstract

In this paper, an auditory-inspired feature extraction algorithm based on a recently published time-frequency transform, i.e., auditory transform (AT) and the power normalized cepstral coefficients (PNCC) is proposed. Usually, the performance of acoustic models trained in clean speech drops significantly when tested on noisy speech.The proposed feature, called Cochlear Filter PNCC (CFPNCC), has shown strong robustness in the acoustic mismatch situations. An important feature of the proposed algorithm is the combination of advantages of the cochlear filter with the advantages of the PNCC feature, which has the resistance to both stationary noise and non-stationary noise. As shown in our experiments, in a GMM-UBM speaker verification system, CFPNCC outperforms the original PNCC and achieves the best overall results on the SSC database compared to the conventional features such as MFCC and RASTA-PLP under noisy conditions.

Keywords: Speaker Verification, Feature Extraction, Noise Robust

*Corresponding Author E-mail: j_khalilpour@yahoo.com:.

Archive of SID
www.SID.ir

۱. مقدمه

تصدیق گوینده خودکار (ASV) به فرآیند شناخت (تشخیص) هویت یک شخص بر اساس صدای او اطلاق می‌گردد. تمامی سیستم‌های ASV از سه بخشِ: استخراج ویژگی، مدل‌سازی گوینده و تصمیم‌گیری بر اساس طبقه‌بندی الگویی[1] تشکیل می‌شود [۱]. با توجه به این‌که استخراج ویژگی گام اول این زنجیره را تشکیل می‌دهد، کیفیت سایر بخش‌ها (مدل‌سازی و طبقه‌بندی) وابستگی شدیدی به آن خواهد داشت.

الگوریتم‌های استخراج ویژگی (یا به اختصار ویژگی‌های) موفق بایستی اطلاعات تفکیک‌کننده کافی برای طبقه‌بندی را دارا باشند. همچنین به آسانی قابل مدل‌سازی بوده و در مقابل تغییرات صوتی مقاوم باشند. ضرایب کپسترال مبتنی بر فرکانسِ مل (MFCC) و ضرایب پیشگویی خطی (PLP) از جمله ویژگی‌های متداول در زمینه تشخیص گوینده هستند که در شرایطِ انطباق محیط‌های آموزش و آزمایش، عملکرد تشخیصِ هویت خوبی را از خود نشان می‌دهند. اما نرخ خطای آن‌ها در شرایط عدم انطباق به طور فزاینده‌ای بالا می‌رود. بسیاری از تحقیقات، تمرکز خود را در زمینه روش‌های افزایش مقاومت سیستم و مقابله با این مسئله قرار داده‌اند. روش‌های بهبود گفتار مانند کاهش (تفاضل) طیفی و استفاده از فیلتر وینر از جمله روش‌هایی هستند که از طریق افزایش کیفیت سیگنال به حل این مشکل پرداخته‌اند [۲].

راه دیگر افزایش مقاومت سیستم، بهبود دامنه طیفیِ ویژگی با وزن‌دهی نواحی معتبر [۳] یا حذف ناحیه‌های زمانی- فرکانسی غیر مهم(بدون اطلاعات مهم) با روش‌های حذف ویژه[2] [۶-۴] است. علاوه بر این تکنیک‌ها، روش نرمالیزه کردن ویژگی نیز می‌تواند به صورت ترکیبی با سایر روش‌ها, جهت پردازش بهتر سیگنال‌های صوتی مورد استفاده قرار گیرد. در نهایت روش‌هایی وجود دارند که از بازبینی در الگوریتم استخراج, جهت مقاوم کردن ویژگی استفاده می‌کنند [۸-۷].

در این مقاله تمرکز بر روی یک تبدیل زمانی-فرکانسیِ مبتنی بر سیستم شنوایی [۱۰-۹] و الگوریتم استخراج ویژگی PNCC است [۱۲-۱۱]. تبدیل شنوایی (AT) که موج عبوری از داخل حلزونی گوش را مدل‌سازی می‌کند، شامل یک جفت تبدیل مستقیم و معکوس است. در حالت تبدیل مستقیم, سیگنال ورودی به یک سری از سیگنال‌های زیرباند شکسته می‌شود. از مزایای AT می‌توان به مقاوم بودن در مقابل نویز, مستقل بودن از خرابی ناشی از هارمونیک‌های Pitch و نویز محاسباتی اشاره کرد. ویژگی PNCC به عنوان یک الگوریتم بهبودیافته مستخرج از MFCC معرفی شده ولی عملکرد بالایی را در سیستم‌های تشخیص گفتار [۱۳] و گوینده [۱۴] نسبت به آن ویژگی نشان داده است.

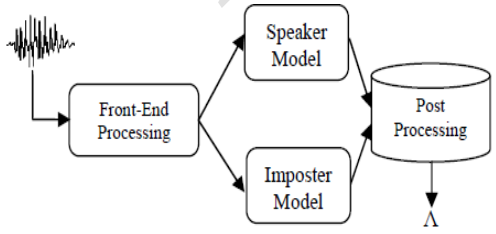
در این مقاله, یک ویژگی جدید به نام ضرایب کپسترال نرمالیزه شده مبتنی بر فیلتر کاکلی یا به اختصار CFPNCC پیشنهاد می‌گردد که از به‌کارگیری فیلتر تبدیل شنوایی در بخش مدل‌سازی حلزونی گوش در ساختار PNCC ایجاد و در سیستم تصدیق هویت گوینده مبتنی بر مدل مخلوط گوسی ارزیابی شده است. در بخش بعدی این مقاله, ابتدا مؤلفه‌های اساسی یک سیستم تصدیق هویت گوینده به اختصار توضیح داده می‌شود. سپس به شرح الگوریتم استخراج ویژگی پیشنهادی پرداخته می‌شود. ارزیابی عملکردِ سیستم پیشنهادی و نتیجه‌گیری، بخش‌های پایانی این مقاله خواهند بود.

۲. بررسی سیستم پایه تصدیق هویت گوینده

ساختار یک سیستم تصدیق هویت گوینده مبتنی بر مدل مخلوط گوسی در شکل (۱) نمایش داده شده است. این ساختار از سه بخش تشکیل شده است که در ادامه به بررسی هر کدام خواهیم پرداخت.

۲-۱. پردازش اولیه[3]

وظیفه این بخش استخراج ویژگی از سیگنال گفتار و به دست آوردن مشخصات گوینده از آن است. علاوه بر آن، تکنیک‌هایی برای کاهش اثر پیچیدگی[4] از ویژگی، مانند فیلتر کردن خطی نیز می‌تواند مورد استفاده قرار گیرد. خروجی مرحله پردازش اولیه یک رشته از بردارهای ویژگی به صورت $\mathbf{X}=\{\mathbf{x}_1,...,\mathbf{x}_t\}$ است که t بیانگر اندیس بردار در زمان‌های گسسته $t\in\{1,2,...,T\}$ است. از جمله پرکاربردترین ویژگی‌ها در سیستم‌های نوین تشخیص گوینده، می‌توان به MFCC، RASTA-PLP و PNCC اشاره کرد که از آن‌ها برای ارزیابی ویژگی پیشنهادی استفاده خواهد شد.



شکل ۱. مؤلفه‌های یک سیستم تصدیق هویت گوینده [۲۳].

[1] Pattern Classification
[2] Missing Feature Techniques
[3] Front End
[4] Confounding

Archive of SID

www.SID.ir

۲-۲. مدل‌سازی گوینده هدف و وانمودکنندگان[1]

یکی از گام‌های مهم در طراحی سیستم‌های تصدیق هویت گوینده، انتخاب یک مدل مناسب برای طبقه‌بندی (دسته‌بندی) گوینده هدف و وانمودکننده‌ها است. انتخاب این مدل بستگی شدیدی به نوع ویژگی انتخابی و همچنین نوع کاربرد سیستم دارد. برای سیستم‌های تصدیق هویت گوینده مستقل از متن که هیچ اطلاعات اولیه‌ای از جمله (صحبت) شخص گوینده نداریم، یکی از موفق‌ترین روش‌ها، طبقه‌بندی بر اساس مدل مخلوط گوسی (GMM) است. از مزایای استفاده از GMM به عنوان یک تابع چگالی احتمال، می‌توان به هزینه محاسباتی کم، داشتن پایه نظری شناخته‌شده، غیرحساس بودن به مشخصات زمانی سیگنال گفتار (در حالت مستقل از متن) و مدل‌سازی تنها بر اساس توزیع اطلاعات صوتی گوینده اشاره کرد [۱۵]. برای یک بردار ویژگی D بعدی $\mathbf{X}$، چگالی مخلوط مورد استفاده برای طبقه‌بندی به صورت زیر تعریف می‌شود:

$$p(\mathbf{x}|\lambda)=\sum_{i=1}^{M} w_i b_i(\mathbf{x}). \qquad (۱)$$

تابع چگالی یک ترکیب خطی وزن دار از M چگالی گوسی نمایی $b_i(\mathbf{x})$ با پارامترهای میانگین D×1 بعدی $\boldsymbol{\mu}_i$، ماتریس کواریانس D×D بعدی $\boldsymbol{\Sigma}_i$ به شکل زیر است:

$$b_i(\mathbf{x})=\frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}_i|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T(\boldsymbol{\Sigma}_i)^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)} \qquad (۲)$$

علاوه بر این، وزن‌های مخلوط w_i باید به گونه‌ای باشند که شرط $\Sigma w_i = 1$ برقرار باشد. در حالت کلی پارامترهای مدل به صورت $\lambda=\{w_i,\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\}$ بیان می‌گردند و با داشتن مجموعه داده‌های آموزش، با استفاده از الگوریتم حداکثر احتمال[2](EM) تخمین زده می‌شوند. به‌طور معمول بردارهای ویژگی $\mathbf{X}=\{\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_T\}$ مستقل از هم در نظر گرفته می‌شوند. بنابراین، تابع احتمال مدل λ به صورت زیر تبدیل می‌شود:

$$p(\mathbf{X}|\lambda)=\prod_{t=1}^{T} p(\mathbf{x}_t|\lambda), \qquad (۳)$$

نرخ احتمال $p(\mathbf{x}_t|\lambda)$ از طریق رابطه (۱) قابل محاسبه خواهد بود. به طور معمول مقدار میانگین احتمال، که از تقسیم $p(\mathbf{X}|\lambda)$ بر T به‌دست می‌آید، مورد استفاده قرار می‌گیرد. اگرچه مدل گوینده هدف λ_{hyp}، به وضوح از طریق داده‌های آموزشِ آن گوینده تخمین زده می‌شود، اما مدل‌سازی مربوط به وانمودکنندگان، به این دلیل که باید بیانگر مجموعه فضای همه گوینده‌های غیرهدف باشد، به‌طور مشخص تعریف نشده است. راه‌کار غالب برای این مسئله، جمع‌آوری داده‌های تعداد زیادی از گوینده‌های غیرهدف و ایجاد یک مدل واحد برای آن‌هاست. مدل پیش‌زمینه کلی[3](UBM) نامی است که برای این مدل انتخاب می‌شود [۱۵].

۲-۳. مرحله تصمیم‌گیری

با داشتن $\mathbf{X}$ به عنوان سیگنال صحبت ورودی و یک مدل برای گوینده هدف، وظیفه بخش تصمیم‌گیری به صورت تعیین احتمال تعلق $\mathbf{X}$ به گوینده S تعریف می‌شود. یکی از قدرتمندترین روش‌های تصمیم‌گیری، آشکارسازی نرخ احتمال[4](LR) است. در این روش، برای تعیین این‌که آیا گوینده هدف جمله $\mathbf{X}$ را گفته است، نرخ احتمال $p(\mathbf{X}|\lambda_s)$، با احتمال این‌که سایر گویندگان (غیرهدف) آن جمله را گفته باشند $p(\mathbf{X}|\lambda_{bkg})$، مقایسه می‌شود. مناسب‌ترین روش برای انجام این مقایسه، آزمایش LR است که به صورت زیر تعریف می‌شود:

$$\frac{p(\mathbf{X}|\lambda_s)}{p(\mathbf{X}|\lambda_{bkg})}=\begin{cases}\geq\theta & \text{accept}\\ <\theta & \text{reject,}\end{cases} \qquad (۴)$$

در این رابطه θ بیانگر آستانه تصمیم‌گیری است. با گرفتن لگاریتم از طرفین رابطه (۴) نرخ احتمال لگاریتمی زیر حاصل خواهد شد:

$$\log p(\mathbf{X}|\lambda_s)-\log p(\mathbf{X}|\lambda_{bkg})\underset{<}{\overset{\geq}{}}\log\theta=\Lambda. \qquad (۵)$$

برای یافتن Λ قاعده تصمیم‌گیری بیز[5] جهت یافتن بهینه‌ترین پاسخ مورد استفاده قرار گرفته است. زمانی که هیچ اطلاعات اولیه‌ای درباره میزان خطاها نداریم، قاعده بیز به صورت انتخاب مقادیری از Λ که نرخ خطای کلی[6] را به کمترین مقدار می‌رساند، اعمال می‌گردد:

$$\text{HTER}=\frac{1}{2}(\%\text{FA}+\%\text{FR}), \qquad (۶)$$

در این رابطه %FA نرخ پذیرش اشتباه و %FR نرخ رد اشتباه است.

۳. ضرایب کپسترال توان نرمالیزه شده مبتنی بر فیلتر کاکلی(CFPNCC)

ویژگی پیشنهادیِ ضرایب کپسترال توان نرمالیزه شده مبتنی بر فیلتر کاکلی(CFPNCC) با ترکیب الگوریتم ویژگی متداول PNCC و فیلتربانک تبدیل شنوایی(AT) به‌دست می‌آید. این کار از طریق تغییر فیلتربانک گاماتون که در بسیاری از ویژگی‌های مبتنی بر سیستم شنوایی برای مدل‌سازی حلزونی گوش استفاده شده، با فیلتر بانک تبدیل شنوایی صورت می‌گیرد. جزئیات روش استخراج الگوریتم پیشنهادی در ادامه شرح داده می‌شود.

[1] Imposter
[2] Expectation–Maximization
[3] Universal-Background-Model
[4] likelihood ratios (LR) detection

Archive of SID

www.SID.ir

۳-۱. الگوریتم استخراج ویژگی PNCC

شکل (۲) بلوک دیاگرام الگوریتم استخراج ویژگی PNCC را نمایش می‌دهد. این ساختار شامل سه بخش اصلی با نام‌های پردازش اولیه، جبران‌سازی اثر محیط[1] و پردازش نهایی[2] است. پردازش اولیه مراحل زیر را شامل می‌شود:

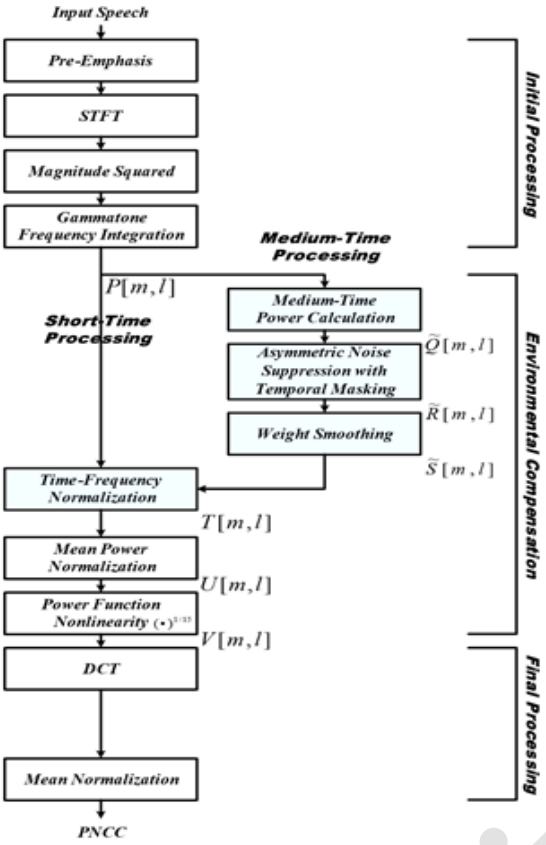
- عبور از یک فیلتر پیش تاکید به شکل $H(z)=1-0.97z^{-1}$.
- اعمال تبدیل فوریه زمان کوتاه[3](STFT) با استفاده از پنجره‌های همینگ با طول ۲۵٫۶ میلی‌ثانیه و درهم روی ۱۰ میلی‌ثانیه.
- وزن‌دهی مربع دامنه STFT با فیلتربانک ۴۰ تایی گاماتون که فرکانس‌های مرکزی آن به صورت خطی در معیار پهنای باند مستطیلی معادل (ERB) در بازه ۲۰۰Hz تا ۸۰۰۰ Hz فاصله‌گذاری شده‌اند.
- محاسبه توان طیفیِ زمان کوتاه $P[m,l]$، که m و l به ترتیب اندیس‌های فریم و کانال هستند.

در مرحله جبران‌سازی محیطی کمیتی به نام توانِ زمان-متوسط $\tilde{Q}[m,l]$، با محاسبه میانگین متحرک $P[m,l]$ به دست می‌آید که در جبران‌سازی اثر نویز به کار می‌رود. برای انجام این کار، فیلتر پایین گذر نامتقارن زمانی بر روی توان زمان-متوسط، اعمال و پوش پایینی آن به عنوان مدل نویز، تخمین زده می‌شود. نهایتاً نتیجه به دست آمده از کمیت $\tilde{Q}$ حذف می‌گردد. یکسوسازی نیم‌موج و تخمین آستانه ماسک زمانی در ادامه انجام می‌گیرد و از ترکیب آن با حداقل نویز تخمینی در مرحله قبل، کمیتی به نام $\tilde{R}[m,l]$ به دست می‌آید. جزئیات پیاده‌سازی‌های مربوط به این قسمت، در شکل (۳) نمایش داده شده است. برای انجام آسان‌تر محاسباتِ این بخش، تابع تبدیلی به شکل $\tilde{R}[m,l]/\tilde{Q}[m,l]$ از ترکیب اثر توأم الگوریتم حذف نویز نامتقارن و بلوک ماسک زمانی به دست می‌آید و پس از میانگین‌گیری زمانی و فرکانسی($\tilde{S}[m,l]$)، برای محاسبه خروجی بلوک در طیف زمان کوتاه $P[m,l]$ ضرب شود.

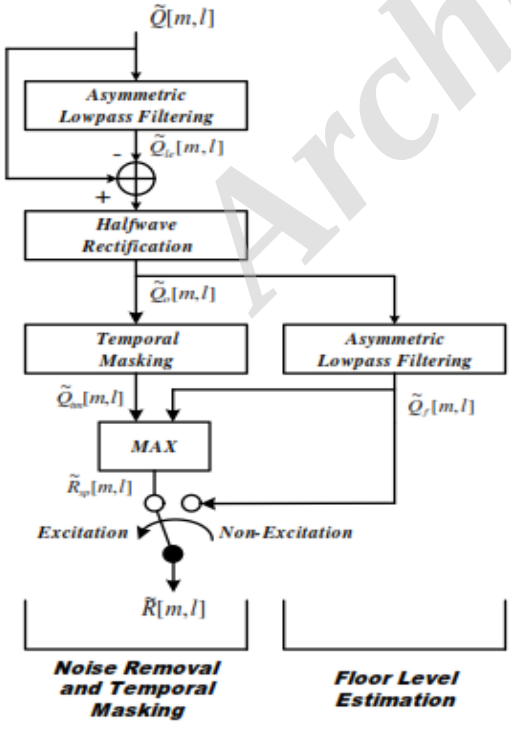
$$T[m,l]=P[m,l]\tilde{S}[m,l] \qquad (۷)$$

در گام بعدی برای انجام نرمالیزه‌سازی توانِ متوسط، میانگین متحرک در بین کانال‌های فرکانسی بر روی $T[m,l]$ محاسبه می‌شود(μ) و با تقسیم T بر آن (μ) کمیت U به دست می‌آید.

در بخش پردازش نهایی، با اعمال عملگر غیرخطیِ قانون توان به صورت ریشه ۱۵ ام($V[m,l]=U[m,l]^{1/15}$) و محاسبه تبدیل DCT و نرمالیزه کردن میانگین طیفی ضرایب PNCC محاسبه می‌شوند [۱۳].



شکل ۲. ساختار الگوریتم استخراج ویژگی PNCC [۱۳].



شکل ۳. بلوک دیاگرام مربوط به حذف نویز نامتقارن و ماسک زمانی در ساختار الگوریتم استخراج ویژگی PNCC[۱۳] .

[1] Bayes
[2] Half Total Error Rate
[3] Environmental Compensation
[4] Final Processing
[5] Short Time Fourier Transform(STFT)

Archive of SID

www.SID.ir

۳-۲. تبدیل شنوایی[1]

فیلتربانک کاکلی مورد استفاده در ویژگی پیشنهادی در واقع تبدیلِ مستقیم AT است که با فرض $f(t)$ به عنوان یک سیگنال ورودی جمع‌پذیر و $\psi_{a,b}(t)$ به عنوان پاسخ ضربه غشاء بازیلار گوش به صورت زیر تعریف می‌شود:

$$T(a,b)=\int f(t)\psi_{a,b}(t)dt. \qquad (۸)$$

در این رابطه

$$\psi_{a,b}(t)=\frac{1}{\sqrt{|a|}}\psi\left(\frac{b-t}{a}\right) \qquad (۹)$$

است. $T(a,b)$ موج متحرک در داخل BM را بیان می‌کند. فاکتور a یک متغیر مقیاس یا تأخیر است که با تغییر آن می‌توان فرکانس مرکزی پاسخ ضربه را انتقال داد. فاکتورهای b و $1/\sqrt{|a|}$ شیفت زمانی و انرژی نرمال‌سازی هستند. فیلتر کاکلی به صورت زیر تعریف می‌شود:

$$\psi_{a,b}(t)=\frac{1}{\sqrt{|a|}}\left(\frac{b-t}{a}\right)^{\alpha}\exp\left[-2\pi f_L\beta\left(\frac{b-t}{a}\right)\right]\cos\left[2\pi f_L\left(\frac{b-t}{a}\right)+\theta\right]u(-t). \qquad (۱۰)$$

α و β پارامترهایی مثبت و $u(t)$ تابع پله واحد است. از دیدگاه نظری انتقال، مقدار θ باید به گونه‌ای انتخاب گردند که شرط زیر را برقرار سازد:

$$\int_{-\infty}^{\infty}\psi(t)dt=0. \qquad (۱۱)$$

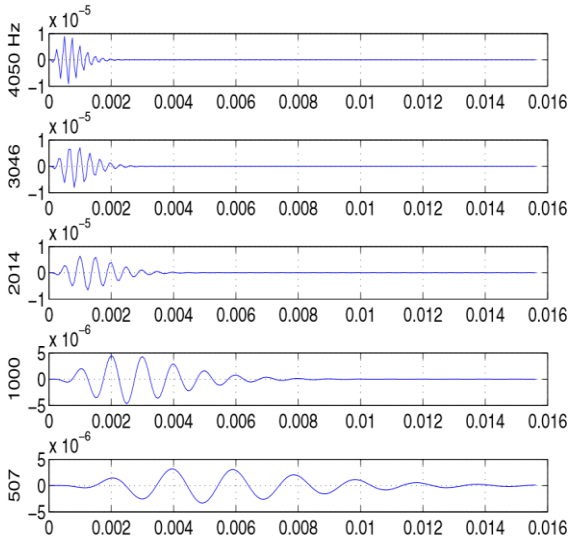
برای هر فیلتر با فرکانس مرکزی f_c مقدار متغیر a به صورت زیر تعریف می‌شود:

$$a=\frac{f_l}{f_c}, \qquad (۱۲)$$

f_l حداقل فرکانس مرکزی موجود در فیلتربانک است. توجه به این نکته ضروری است که توزیع فرکانسی فیلتر کاکلی می‌تواند خطی یا غیرخطی مانند ERB [۱۴]، Bark [۱۵]، Mel [۱۶]، لگاریتمی و غیره باشد. شکل (۴) پاسخ ضربه مربوط به پنج فیلتر کاکلی را که توسط رابطه (۱۰) محاسبه شده است، نمایش می‌دهد.

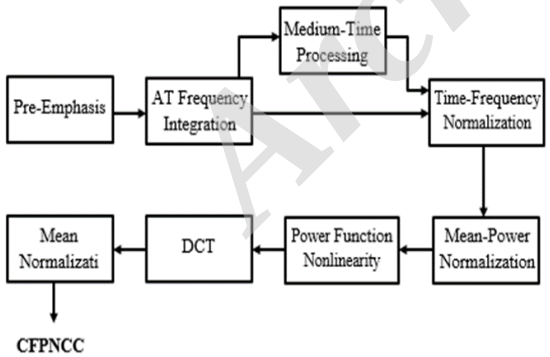
۳-۳. ضرایب پیشنهادی CFPNCC

بلوک دیاگرام شکل (۵) الگوریتم استخراج ویژگی پیشنهادی را نمایش می‌دهد. در این ساختار، در بخش پردازش اولیه، سیگنال عبوری از فیلتر پیش تاکید از طریق فیلتر کاکلی ۴۰ کاناله که فرکانس‌های مرکزی آن در محدوده ۲۰۰Hz تا ۸۰۰۰ در معیار ERB به طور خطی فاصله‌گذاری شده‌اند، تجزیه می‌گردد. توان طیفی زمان کوتاه، در ادامه از خروجی‌های فیلتربانک محاسبه می‌شود. مابقی مراحل مشابه بلوک‌های استفاده شده در ساختار PNCC است.



شکل ۴. پاسخ ضربه مربوط به BM در تبدیل AT با پارامترهای $\alpha=3$ و $\beta=0$. این فیلترها بیش‌ترین شباهت را به اندازه‌گیری‌های روان-صوتی[1] دارند [۱۹].

در ویژگی پیشنهادی، فیلتر کاکلی به جای فیلتر گاماتون برگزیده شده است. در تابع گاماتون پهنای باند هر فیلتر به فرکانس مرکزی آن فیلتر وابسته است و قابل‌تغییر نیست. درحالی‌که پهنای باند در AT، (معادله ۱۰) توسط پارامتری به نام β قابل‌دسترس و کنترل است. تنظیم پارامتر β نیز رابطه مستقیمی با عملکرد سیستم تشخیص دارد که با انتخاب مناسب آن می‌توان به تشخیص بهتری رسید. تبدیل AT نسبت به نویز بسیار مقاوم‌تر بوده و تحت تأثیر خرابی ناشی از هارمونیک‌های پیچ[2] و نویز محاسباتی قرار نمی‌گیرد.



شکل ۵. ساختار الگوریتم استخراج ویژگی پیشنهادی CFPNCC [۲۰].

بر اساس مطالعات صورت گرفته در منبع [۲۰]، پارامتر پهنای باند فیلتر β، ۰/۳۵ انتخاب شد. علاوه بر آن، خصوصیات دینامیکی CFPNCC نیز با افزودن ضرایب Δ به آن، در ساختار ویژگی مورد استفاده قرار گرفت.

[1] Psychological Measurements
[2] Pitch

Archive of SID
www.SID.ir

۴. نتایج آزمایش‌ها و ارزیابی

در این بخش ابتدا، تنظیمات آزمایشگاهی شامل شرح منابعِ مورد استفاده و معیارهای ارزیابی، مورد بررسی قرار گرفته است. سپس برتری ویژگی پیشنهادی در سیستم تصدیق هویت گوینده مبتنی بر مدل GMM-UBM نسبت به خود PNCC و دو ویژگی متداول دیگر نمایش داده شده است.

۴-۱. تنظیمات آزمایشگاهی

برای انجام آزمایش‌های تصدیق هویت گوینده از پایگاه دادگان استاندارد SSC [۲۱] استفاده شده است. SSC شامل ۱۷۰۰۰ جمله صوتی مربوط به ۳۴ گوینده (۱۸ مرد و ۱۶ زن و ۲۵۰ جمله برای هر شخص) است. در این تحقیق از ۱۰ جمله برای آموزش مدل و ۱۰ جمله مجزا برای آزمایش هر گوینده استفاده شده است. برای آموزش مدل UBM نیز تمامی داده‌های آموزش مربوط به ۳۴ نفر جمع شده و یک مدل واحد ایجاد گردیده است.

داده‌های مربوط به مرحله آزمایش با ترکیب سیگنال‌های تمیز با نویزهای سفید(ایستان) و کارخانه(غیرایستان) که از پایگاه دادگان استاندارد NOISX-92 [۲۲] اتخاذ شده است، در ۵ سطح SNRمختلف شامل حالت بدون نویز، ۵dB-، ۰dB، ۵ dB و ۱۰ dB ایجاد شده‌اند. ویژگی پیشنهادی این مقاله CFPNCC، از طریق مقایسه عملکرد تصدیقِ هویت، با ویژگی‌های MFCC، RASTA-PLP و خود PNCC مورد ارزیابی قرار گرفته است. برای پیاده‌سازی MFCC و RASTA-PLP از کدهای استاندارد و برای PNCC از انچه قبلا گزارش شده [۱۶]، استفاده شده و ضرایب مشتق اول به تمامی آن‌ها اضافه گردیده است. معیارهای ارزیابی سیستم، EERو MinDCF بوده‌اند. EER خطای مربوط به حالتی است که نرخ پذیرش اشتباه (P_{miss}) و نرخ رد اشتباه (P_{fa}) با هم برابرند و MinDCF که یکی از معیارهای ارزیابی استاندارد NIST در تشخیص گوینده است، به صورت ترکیب وزن‌داری از P_{miss} و P_{fa} به شکل $0.1\times P_{miss}+0.99\times P_{fa}$ تعریف می‌شود. علاوه بر این دو معیار نمودارهای DET مربوط به چندین حالت از شرایط آزمایشگاهی نیز جهت مقایسه بهتر آورده شده است.

۴-۱. نتایج آزمایشگاهی و بررسی آن

این بخش به تحلیل عملکرد CFPNCC در سیستم تصدیق هویت گوینده مبتنی بر GMM-UBM در قالب معیارهای اشاره‌شده، می‌پردازد.

همان‌گونه که در جدول (۱) مشاهده می‌شود، در حالت بدون نویز، CFPNCC همانند سایر ویژگی‌ها، عملکرد تقریباً کاملی را در معیار EER از خود نشان می‌دهد. جدول‌های (۲) تا (۵) مقایسه‌ای را بین نتایج تشخیص در دو معیار EER و MinDCF نشان می‌دهد. همان طور که مشاهده می‌شود. برای SNRهای بالا(۰، ۵، و ۱۰ دسی‌بل) تفاوت قابل‌ملاحظه‌ای بین نتایج ویژگی پیشنهادی و سه ویژگی دیگر وجود دارد که این اختلاف در $SNR=-5dB$ کاهش می‌یابد. بر طبق نتایج ارائه‌شده در جدول‌های (۲) و (۳) که مربوط به نویز سفید (ایستان) است، CFPNCC در هر دو معیار بهتر از سایر ویژگی‌ها عمل کرده است. جدول‌های (۴) و (۵) عملکرد سیستم تصدیق را در نویز غیرایستان کارخانه نشان می‌دهد. در این حالت ویژگی پیشنهادی در تمامی حالت‌های مربوط به معیار EER بهتر از سایرین عمل می‌کند. با این وجود در معیار MinDCF، این برتری در SNRهای بالا(۵ و ۱۰ دسی‌بل) قابل مشاهده است.

جدول ۱. نتایج تصدیق هویت گوینده در معیار EER بدون نویز.

نوع ویژگی	EER (%)
MFCC	۰/۱۸
RASTAPLP	۰/۳۰
PNCC	۰/۲۹
CFPNCC	۰/۲۹

جدول ۲. نتایج تصدیق هویت گوینده در معیار EER در حضور نویز سفید در چهار SNR مختلف.

نوع ویژگی	EER (%)			
	SNR (dB)			
	-۵	۰	۵	۱۰
MFCC	۵۰/۲۹	۴۶/۶۹	۴۶/۱۱	۳۵/۷۵
RASTAPLP	۴۸/۸۲	۴۷/۲۲	۴۳/۲۳	۳۵
PNCC	۳۹/۱۱	۲۶/۴۷	۱۳/۷۹	۶/۱۷
CFPNCC	۳۷/۲۹	۲۳/۳۳	۱۰/۷۴	۴/۷

جدول ۳. نتایج تصدیق هویت گوینده در معیار MinDCF در حضور نویز سفید در چهار SNR مختلف.

نوع ویژگی	MinDCF (%)			
	SNR (dB)			
	-۵	۰	۵	۱۰
MFCC	۹/۸۹	۹/۵۷	۹/۳۶	۷/۲۲
RASTAPLP	۱۰	۹/۸	۹/۶۲	۸/۲
PNCC	۹/۹۵	۹/۷۸	۷/۸۷	۳/۱۳
CFPNCC	۹/۹۰	۹/۴۶	۶/۴۴	۲/۵۴

جدول ۴. نتایج تصدیق هویت گوینده در معیار EER در حضور نویز کارخانه در چهار SNR مختلف.

نوع ویژگی	EER (%)			
	SNR (dB)			
	-۰/۵	۰	۵	۱۰
MFCC	۵۰/۳۰	۴۴/۱۱	۳۶/۶۸	۲۳/۱۷
RASTAPLP	۴۹/۸۳	۴۷/۰۵	۴۰/۷۶	۲۹/۷۲
PNCC	۴۳/۲۹	۲۹/۱۱	۱۲/۵۸	۵/۲۹
CFPNCC	۴۳/۲۳	۲۸/۵۲	۱۰/۸۸	۴/۴۱

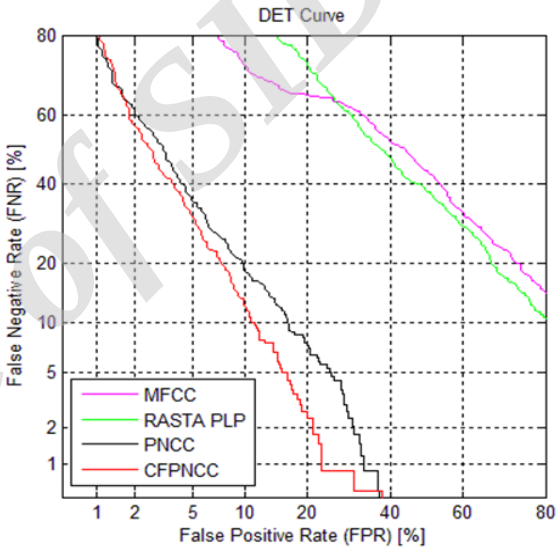
Archive of SID

www.SID.ir

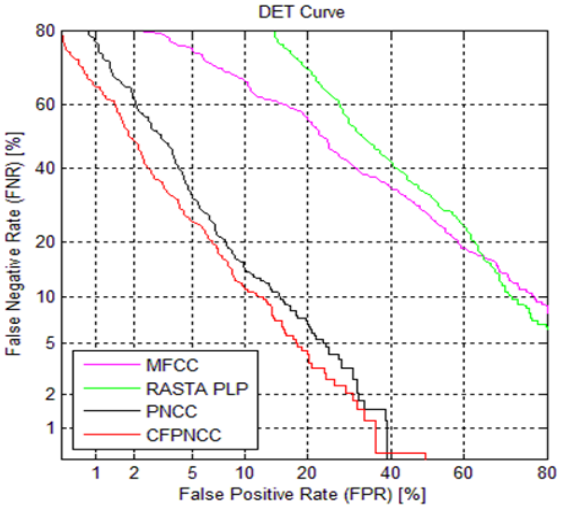
جدول ۵. نتایج تصدیق هویت گوینده در معیار MinDCF در حضور نویز کارخانه در چهار SNR مختلف.

نوع ویژگی	MinDCF (%) SNR (dB)			
	-۵	۰	۵	۱۰
MFCC	۹/۹۵	۹/۹۲	۹/۸۶	۹/۴۰
RASTAPLP	۱۰	۹/۹۷	۹/۹۰	۹/۷۰
PNCC	۹/۸۷	۹/۷۷	۷/۹۸	۳/۶۹
CFPNCC	۱۰	۹/۸۲	۷/۳۴	۳/۳۰

شکل‌های (۶) و (۷) منحنی‌های DET مربوط به نویزهای سفید و کارخانه را در $SNR = 5dB$ نمایش می‌دهد. نمودارها برتری ویژگی پیشنهادی را نسبت به ویژگی‌های MFCC، RASTA-PLP و PNCC به وضوح نشان می‌دهد.



شکل ۶. نمودارهای DET مربوط چهار ویژگی مختلف در سیستم تصدیق هویت گوینده در حضور نویز سفید ($SNR = 5dB$).



شکل ۷. نمودارهای DET مربوط چهار ویژگی مختلف در سیستم تصدیق هویت گوینده در حضور نویز کارخانه ($SNR = 5dB$).

۵. نتیجه‌گیری

در این مقاله یک ویژگی جدید به نام ضرایب کپسترال نرمالیزه شده توان مبتنی بر فیلتر کاکلی(CFPNCC) پیشنهاد گردیده و در بخش پردازش اولیه یک سیستم ASV مورد استفاده قرار گرفته است. روش پیشنهادی، با استفاده از تبدیل زمان- فرکانسی برگشت‌پذیر AT و ترکیب آن با الگوریتم استخراج ضرایب PNCC محاسبه می‌شود. نتایج آزمایش‌ها نشان می‌دهد که در شرایط عدم انطباق محیط‌های آموزش و آزمایش، این ویژگی بهتر از خود PNCC و دو ویژگی پرکاربرد MFCC و RASTA-PLP عمل می‌کند. به عنوان قدم بعدی عملکرد CFPNCC در سیستم تصدیق هویت جدید مبتنی بر IVector در حضور نویزهای مختلف و پایگاه دادگان دیگر مورد بررسی قرار می‌گیرد.

۶. مرجع‌ها

[1] Shao, Y.; Deliang, W. “Robust Speaker Identification Using Auditory Features and Computational Auditory Scene Analysis”; Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, 2008, 1589-1592.

[2] Hermansky, H. “Perceptual Linear Prediction (PLP) Analysis of Speech”; J. Acoust. Soc. Am. 1990, 87, 1738-1752.

[3] Alam, M. J.; Kenny, P.; O’Shaughnessy, D. “Robust Feature Extraction for Speech Recognition by Enhancing Auditory Spectrum”; Proc. Interspeech, Portland, Oregon, USA, 2012.

[4] Drygajlo, A.; El-Muliki, M. “Speaker Verification in Noisy Environments with Combined Spectral Subtraction and Missing Feature Theory”; Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, 1998, 121-124.

[5] Pullella, D.; Kuhne, M.; Togneri, R. “Robust Speaker Identification Using Combined Feature Selection and Missing Data Recognition”; Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, 2008, 4833-4836.

[6] Ribas, D. ; Villalba, J.; Lleida, E.; Calvo, J. “Missing Feature Techniques Combination for Speaker Recognition in Noisy Environment”; FALA 2010: VI Jornadas en Tecnolog´ıa del Habla and II Iberian SLTech Workshop, Vigo, Spain, 2010, 115-118.

[7] Liu, Y.; Liang, H.; Liu, J. “Improved Multitaper PNCC Feature for Robust Speaker Verification”; IEEE 9th International Symposium on Chinese Spoken Language Processing, 2014.

[8] Li, Q. “Solution for Pervasive Speaker Recognition”; SBIR Phase I Proposal, Submitted to NSF IT.F4, Li Creative Technologies, Inc. NJ, 2003

[9] Li, Q. “An Auditory-Based Transform for Audio Signal Processing”; Proc. IEEE Workshop Applications of Signal Processing to Audio and Acoustics, (New Platz, NY), 2009.

[10] Kinnunen, T.; Li, H. “An Overview of Text-Independent Speaker Recognition: from Features to Supervectors”; Speech Communication 2010, 52, 12-40.

[11] Jing, X.; Xiang, B.; Yang, H.; Zhou, P “Robust Speaker Verification Using Improved PNCC Based on GMM-UBM”; Int. J. Automation and Power Eng. 2015, 4, 14-19.

Archive of SID

www.SID.ir

[12] Jayanth, M.; Roja Reddy, B. "Speaker Identification based on GFCC using GMM-UBM"; Int. J. Eng. Sci. Invention 2016, 5, 62-65.

[13] Sadjadi, S.; Hansen, J. "Assessment of Single-Channel Speech Enhancement Techniques for Speaker Identification under Mismatched Conditions"; Proc. Interspeech, Makuhari, Japan, 2010, 2138-2141.

[14] Nasersharif, B.; Akbari, A. "SNR-Dependent Compression of Enhanced Mel Sub-band Energies for Compensation of Noise Effects on MFCC Features"; Pattern Recognition Letters 2007, 28, 1320-1326

[15] Chanwoo, K.; Stern, R. M. "Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition"; IEEE Trans. Audio, Speech and Language Processing 2016, 24, 1315-1329.

[16] Ambikairajah; E.; Jia Min K. K.; Vidhyasaharan S.; Haizhou, L. "PNCC-[vector-SRC Based Speaker Verification"; IEEE Signal & Information Processing Association Annual Summit and Conference, Asia-Pacific, 2012, 1-7.

[17] Reynolds, D. A.; Rose, R. C. "Robust Text-independent Speaker Identification Using Gaussian Mixture Speaker Models"; IEEE Trans. Speech Audio Process, 1995, 3, 72-83.

[18] Moore, B. C. J.; Glasberg, B. R. "Suggested Formula for Calculating Auditory Filter Bandwidth and Axcitation Patterns"; J. Acoust. Soc. Am. 1983, 74, 750-753.

[19] Zwicker, E.; Terhardt, E. "Analytical Expressions for Critical Band Rate and Critical Bandwidth as a Function of Frequency"; J. Acoust. Soc. Am. 1980, 68, 1523-1525

[20] Davis, S. B.; Mermelstein, P. "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences"; IEEE Trans. Acoustic, Speech and Signal Processing 1980, 28, 357-366.

[21] Li, Q.; Yan, H. "Robust speaker identification using an auditory-based feature"; Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing 2010, 4514-4517.

[22] Cooke, M.; Lee, T. W. "Speech Separation and Recognition Competition";http://www.eng.cam.ac.uk/comp.speech/Section l/Data/noisex.html (Available on July 05, 2015).

Archive of SID

www.SID.ir