بازشناسی مقاوم گفتار فارسی با استفاده از روش بردار سری تیلور VTS

محسن قدیانی[1] ، منصور ولی[2]
[1]دانشجوی کارشناسی ارشد دانشگاه شاهد، mohsenghadyani@ymail.com
[2]استادیار دانشگاه شاهد، vali@shahed.ac.ir

چکیده - تکنیک بردار سری تیلور *(VTS)* از جمله کاراترین روش‌های بازشناسی مقاوم گفتار به شمار می رود که در هر دو حوزه اصلاح بردارهای بازنمایی و اصلاح مدل صوتی بازشناسی کاربرد دارد. مقاله حاضر به شرح این تکنیک برای اصلاح بردارهای بازنمایی تخریب شده توسط نویز جمع شونده و کانال انتقال خطی از روی سیگنال گفتار فارسی و در نتیجه افزایش نرخ بازشناسی آن پرداخته است. به این ترتیب که پس از استخراج بردارهای بازنمایی تمیز و نویزی به روش متداول *MFCC*، به کمک تکنیک *VTS* بردارهای بازنمایی نویزی جبران سازی شده‌اند. یک مدل بازشناسی گفتار مبتنی بر شبکه عصبی *MLP* با دینامیک زمانی، توسط دادگان تمیز تعلیم داده شده است. نتایج حاصل از تست مدل بر روی بردارهای بازنمایی نویزی و اصلاح شده نشان داده است که استفاده از الگوریتم *VTS* در جبران‌سازی بردارهای بازنمایی، نه تنها منجر به بهبود چشمگیری در بازشناسی گفتار با *SNR* های پایین خواهد شد(20% بهبود برای *SNR=0*)، بلکه جبران‌سازی برای *SNR*های در حد *25dB* نیز منجر به افزایش بازشناسی نسبت به گفتار تمیز شده و کارآیی این روش برای غلبه بر تنوعات میکروفون و گوینده را نیز نمایان کرده است.

کلید واژه- بازشناسی مقاوم گفتار ، بردار بازنمایی ، بردار سری تیلور ، تابع شرایط محیطی ، نویز

1- مقدمه

هنگامی که از سیستم بازشناسی گفتار آموزش یافته در شرایط آزمایشگاهی در محیط واقعی استفاده شود، راندمان سیستم و نرخ بازشناسی به دلیل عدم انطباق دادگان آموزشی و داده جمع آوری شده در محیط واقعی به مقدار زیادی کاهش می‌یابد. از این رو مبحث مقاوم‌سازی در برابر نویز به عنوان یکی از ضرورت های کاربردی از زمینه‌های فعال تحقیقاتی در سالهای اخیر بوده است[1]. بر اساس تقسیم بندی یک سیستم بازشناسیِ الگو به دو بخش استخراج ویژگی و مدل بازشناسی، تکنیک های مقاوم سازی بازشناسی گفتار را در دو دسته کلی جای می دهیم:

الف- تکنیک‌های مبتنی بر اصلاح بردارهای بازنمایی صوتی (Feature-Based Techniques) که هدف از به کار گیری آن‌ها اصلاح و مقاوم سازی بردارهای بازنمایی است [2].

ب- تکنیک های مبتنی بر اصلاح مدل صوتی بازشناسی (Model-Based Techniques) که هدف از به کار گیری آن ها جبران سازی مدل بازشناسی در برابر نویز محیطی است[2].

به علاوه در بسیاری از موارد (مانند SNR های بسیار پایین یا سیگنال غرق در نویز) تمرکز بر روی یکی از این مجموعه تکنیک ها کافی نیست و باید به هر دو قسمت توجه شود که حاصل آن تکنیک های ترکیبی (Joint Techniques) خواهد بود.

در سال های اخیر استفاده از تکنیک بردار سری تیلور (Vector Taylor Series) جهت مقاوم‌سازی سیستم‌های بازشناسی گفتار در برابر تغییرات محیطی به صورت گسترده‌ای رواج یافته است. ایده بردار سری تیلور روشی تحلیلی و مبتنی بر محاسبات دقیق ریاضی برای تخمین و حذف پارامترهای مزاحم محیطی، که نرخ بازشناسی را به شدت کاهش می دهند، می-باشد[3]. این ایده را اولین بار Moreno در سال 1996 برای نگاشت بردارهای بازنمایی نویزی به تمیز به کار گرفت [4] و از آن پس همواره یکی از مهمترین زمینه‌های مورد علاقه در مباحث مربوط به بهبود کیفیت سیستم‌های بازشناسی گفتار بوده است. تکنیک بردار سری تیلور را به دو طریق می‌توان استفاده نمود [2]:

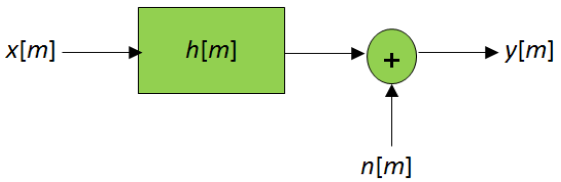
- اصلاح بردارهای بازنمایی و اعمال به مدل بازشناسی (مقاوم سازی بردارهای بازنمایی)
- اصلاح پارامترهای مدل بازشناسی (مقاوم سازی مدل)

در روش اول ابتدا بردارهای بازنمایی را جبران سازی کرده و سپس آن‌ها را بازشناسی می‌کنند، در حالی که در روش دوم بردارهای اصلاح نشده و آلوده به نویز به مدل مقاوم شده اعمال می شوند. الگوریتم مورد استفاده برای هر دو روش کاملا تحلیلی بوده و نتایج حاصل از پیاده سازی آن ها برای جبران سازی نویز جمع شونده و کانال انتقال خطی بسیار قابل توجه می باشند[3].

در مقاله حاضر، اصول و جزییات الگوریتم بردار سری تیلور برای اصلاح بردارهای بازنمایی صوتی مورد بحث قرار می گیرد. در ابتدا نحوه مدل کردن تنوعات محیطی موجود بر روی سیگنال گفتار و معادلات مرتبط برای بیان ریاضی اثر این تنوعات نشان داده خواهد شد. سپس الگوریتم تشریحی و گام به گام جبران سازی اثر کانال خطی انتقال و نویز جمع شونده روی بردارهای بازنمایی استخراج شده به روش MFCC بیان گردیده و نحوه تخمین پارامترهای نویز و کانال، بدون وجود هیچ گونه اطلاعات قبلی و تنها از روی دادگان نویزی مشاهده شده به طور کامل شرح داده خواهد شد. در ادامه نیز نتایج آزمایشات عملی پیاده سازی الگوریتم بر روی یک سیستم بازشناسی گفتار نمونه استاندارد، برای مجموعه دادگان فارس دات تخریب شده توسط نویز و به ازای SNR های مختلف ارائه شده و با نتایج حاصل از حالت جبران نشده مقایسه خواهد گردید.

2- توصیف مدل و مفروضات

در ابتدا مدل کلاسیک شکل (1) را در نظر می گیریم:



شکل 1: نحوه تاثیر تنوعات محیطی بر روی سیگنال گفتار

$x[m]$, $y[m]$ به ترتیب نماینده سیگنال گفتار تمیز و تخریب شده در حوزه زمان و $n[m]$, $h[m]$ نیز بیانگر اثر نویز جمع شونده و کانال انتقال هستند.

نویز جمعی ناشناخته از طریق شیفت دادن میانگین طیفی و افزایش واریانس کل توزیع روی سیگنال تأثیر می‌گذارد. تغییر در کانال تلفن، تغییر در میکروفون یا اضافه شدن یک گوینده باعث اغتشاش‌های کانوالوی در سیگنال گفتار می‌شود. این اغتشاش‌ها یک تغییر جمعی در حوزه لگاریتم سیگنال ایجاد کرده و منجر به یک آفست متغیر با زمان می‌شوند[1].

معادله (1) رابطه بین طیف توان گفتار نویزی و گفتار تمیز را نمایش می دهد:

$$Y(\omega) = X(\omega)\,|H(\omega)|^2 + N(\omega) \qquad (1)$$

که در آن $Y(\omega)$, $X(\omega)$ به ترتیب معرف طیف توان گفتار نویزی و تمیز و $H(\omega)$, $N(\omega)$ نیز به ترتیب نشان دهنده تابع انتقال کانال و طیف توان نویز جمع شونده هستند. معادله فوق در حوزه کپستروم به صورت زیر بیان می شود [5]:

$$y = x + g(x,h,n) \qquad (2)$$

که در آن x, y به ترتیب بردار کپستروم گفتار تمیز و نویزی و $g(x,n,h)$ تابع شرایط محیطی است که به شکل زیر بسط داده می شود:

$$g(x,h,n) = h + ln(1 + e^{n-x-h}) \qquad (3)$$

در معادله بالا i,h,n به ترتیب بردارهای کپستروم نویز، کانال و شماره توزیع گوسی اند. در این حالت فرض می کنیم که می‌توان تابع توزیع احتمال کپستروم گفتار تمیز را به صورت مجموعه‌ای از K توزیع گوسی تکی با میانگین و واریانس $\mu_{x,k}$ ، $\sum_{x,k}$ در نظر گرفت:

$$p(x_t) = \sum_{k=0}^{K-1} p_k\, N_{x_t}(\mu_{x,k}, \sum_{x,k}) \qquad (4)$$

در رابطه فوق p_k احتمال هر یک از k توزیع گوسی است.

همچنین، تابع احتمال نویز جمع شونده را با یک توزیع گوسی تکی مدل کرده و کانال انتقال را ناشناخته، اما خطی و تغییر ناپذیر با زمان فرض می کنیم [6].

هدف نهایی الگوریتم VTS ، تخمین پارامترهای تابع توزیع احتمال گفتار نویزی از روی گفتار تمیز است. پس از محاسبه تابع توزیع احتمال بردار بازنمایی گفتار تخریب شده (PDF(y)) می‌توان تابع توزیع احتمال بردار بازنمایی اصلاح شده برای کپستروم های دیده نشده سیگنال گفتار را به کمک روش می نیمم میانگین مربعات خطا (MMSE) به سادگی حساب کرد[6].

در حالتی که دسته پارامترهای محیطی (n, h) معلوم باشند، می-توان تابع توزیع احتمال گفتار نویزی را به طور مستقیم حساب کرد. اما در کاربردهای عملی (n, h) ناشناخته‌اند و در حقیقت جز درمواقعی که تابع شرایط محیطی شکل بسیار ساده‌ای داشته باشد، امکان محاسبه مستقیم PDF(y) وجود ندارد [7].

برای به دست آوردن PDF گفتار نویزی، با یک ساده سازی شکل کلی آن را به صورت GMM در نظر گرفته و تابع g را با بسط تیلور آن جانشین می کنیم. خواهیم داشت:

$$y = x + g(x_0,h,n) + g'(x_0,h,n)\,(x - x_0) + \ldots \qquad (5)$$

بدیهی است تنها در صورتی بسط تیلور تابع به طور دقیق با خود آن برابر خواهد بود که تعداد جملات بسط بی نهایت باشد. اما در صورتی که از تقریب های مرتبه پایین تر بسط نیز استفاده شود، تخمینی مناسب از مقدار تابع شرایط محیطی در هر نقطه به دست خواهد آمد [8]. در این مقاله از بسط سری تیلور مرتبه اول تابع شرایط محیطی (VTS-1) سود خواهیم برد.

برای محاسبه توزیع احتمال گفتار نویزی در حوزه کپستروم باید تعداد متناهی از جملات بسط (در این جا یک جمله) را در نظر گرفته و آن را به مرتبه خاصی محدود کنیم. در این حالت به روابط زیر برای محاسبه بردار میانگین و ماتریس کوواریانس کپستروم گفتار نویزی می رسیم [9]:

$$\mu_y = E(x) + g(x_0,h,n) + g'(x_0,h,n)\,E(x-x_0) + \ldots \quad (6)$$

$$\Sigma_y = E(x\,x^T) + E(g(x,h,n)\,g(x,h,n)^T) + 2E(x\,g(x,h,n)^T) - \mu_y\,\mu_y^T$$

که در آن $\boldsymbol{g}$ تابع شرایط محیطی و x_0, h, n به ترتیب بردار نویز، کانال انتقال و میانگین بردارهای بازنمایی تمیز هستند که هر سه معلوم اند. لذا توزیع احتمال y به سادگی قابل محاسبه است.

3 - شرح الگوریتم بردار سری تیلور

در بخش قبل روش تخمین پارامترها و توزیع احتمال گفتار نویزی و محاسبه میانگین و کوواریانس کپستروم را شرح دادیم، اما با این فرض که پارامترهای محیطی معلوم‌اند یا دست کم اطلاعاتی در مورد آنها داریم. درحالی که در هیچ یک از کاربردهای عملی چنین نیست. لذا باید از روشی جایگزین برای تخمین پارامترهای شرایط محیطی استفاده کرد که با استفاده از دادگان مشاهده شده، دسته پارامترهای مجهول را نتیجه دهد.

3-1 - توصیف الگوریتم

با فرض وجود داده های در دسترس زیر:

1- مجموعه ای ازکپستروم های آلوده به نویز و اثر انتقال کانال $Y = \{y_0, y_1, \ldots, y_{S-1}\}$.

2- تابع PDF گفتار تمیز.

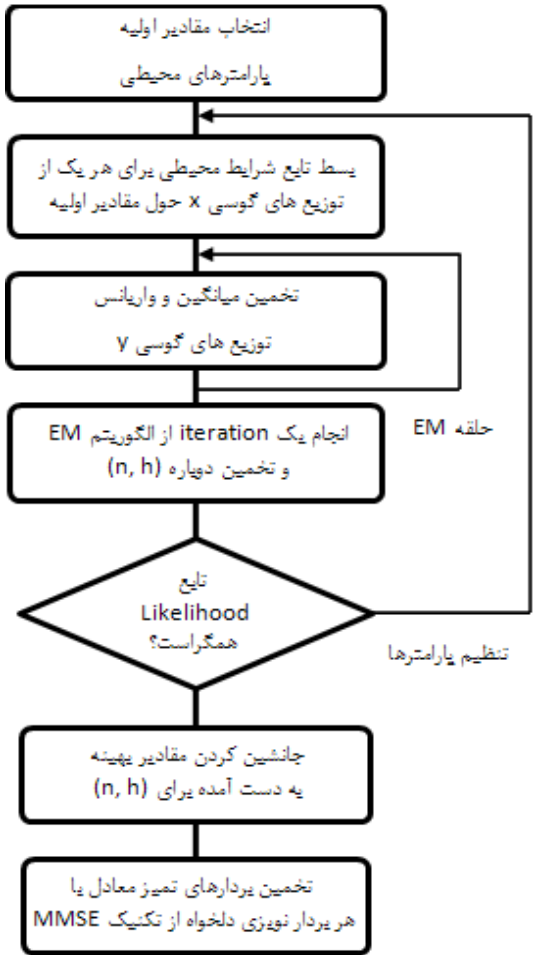
3- مجموعه ای از مقادیر اولیه برای نویز و کانال:

- میانگین نویز: می‌نیمم بردارهای نویزی دیده شده.
- میانگین کانال: اختلاف بین میانگین بردارهای بازنمایی نویزی و تمیز.
- میانگین بردارهای بازنمایی تمیز.

و با در نظر گرفتن مدل فرض شده برای اثر تنوعات مزاحم محیطی، الگوریتم گام به گام جبران سازی پارامترهای محیطی از روی کپستروم گفتار نویزی به ترتیب زیر تحقق می‌پذیرد[6]:

- انتخاب مقادیر اولیه برای $\{\mu_x, n_0, h_0\}$.
- بسط تابع g (با مرتبه انتخابی) برای هر یک از توزیع‌های گوسی x حول $\{\mu_x, n_0, h_0\}$.
- تخمین پارامترهای توزیع احتمال کپستروم گفتار نویزی.
- انجام یک iteration از الگوریتم EM برای تخمین دوباره دسته پارامترهای مجهول (n, h).
- درصورت همگرا شدن تابع حداکثر شباهت (Likelihood)، مقادیر بهینه برای (n, h) به دست آمده‌اند. در غیر این صورت (n_0, h_0) را با (n, h) جانشین کرده به مرحله 2 باز می گردیم.
- با معلوم شدن مقادیر پارامترهای محیطی، از روش حداقل میانگین مربعات خطا (MMSE) بردارهای گفتار تمیز معادل هر بردار نویزی دلخواه را محاسبه می کنیم.

شکل (2) بلوک دیاگرام این الگوریتم را نشان می‌دهد. در بخش بعد جزییات الگوریتم فوق تشریح خواهد شد.



شکل2 بلوک دیاگرام الگوریتم اصلاح بردارهای بازنمایی به روش VTS

3-2- تخمین پارامترهای مزاحم محیطی

همان طور که مشاهده شد با بسط دادن تابع تنوعات محیطی، y تابعی از دو پارامتر n, h خواهد بود. بنابراین با داشتن سیگنال‌هی گفتار نویزی، تابع Likelihood این گونه تعریف می‌شود:

(7)

$$L\left(Y=\{y_0,y_1,\dots,y_{S-1}\}\right)=\sum_{t=0}^{S-1}\log(p(y_t|h,n))$$

دسته پارامترهای محیطی مجهول (n, h) به کمک تکنیک Iterative Expectation Maximization و به صورت بیان شده در ادامه این بخش تخمین زده می‌شوند.

از معادله (6) چنین بر می‌آید که میانگین کپستروم نویزی، تابعی خطی از h , n است:

$$\mu_{k,y}=a_k+b_k\,n+c_k\,h \qquad (8)$$

که در آن داریم:

(9)

$$a_k=\mu_{k,x}+g\left(\mu_{k,x},h_0,n_0\right)-\nabla_h g\left(\mu_{k,x},h_0,n_0\right)h_0-\nabla_n g\left(\mu_{k,x},h_0,n_0\right)n_0$$

$$b_k=\boldsymbol{I}+\nabla_h g\left(\mu_{k,x},h_0,n_0\right)$$

$$c_k=\nabla_n g\left(\mu_{k,x},h_0,n_0\right)$$

که ∇_n و ∇_h به ترتیب مشتق تابع محیطی نسبت به دو پارامتر h,n هستند.

در حالی که ماتریس کوواریانس تنها به سه مقدار از قبل معلوم μ_x, n_0, h_0 بستگی دارد:

$$\Sigma_{k,y}=\left(\boldsymbol{I}+\nabla_h g\left(\mu_{k,x},n_0,h_0\right)\right)\Sigma_{k,x}\left(\boldsymbol{I}+\nabla_h g\left(\mu_{k,x},n_0,h_0\right)\right)^T \qquad (10)$$

تابع کمکی Q را به صورت زیر تعریف می کنیم:

$$Q(\varphi,\bar{\varphi})=E[L(Y,S|\bar{\varphi})|Y,\varphi] \qquad (11)$$

که در آن زوج (Y,S) نماینده دادگان کامل است: Y بردار مشاهدات نویزی است و S نشان می دهد کدام یک از توزیع های گوسی آن بردار مشاهدات را تولید کرده است. φ هم همان دسته دادگان نامعلوم است.

معادله بالا را می توان به این صورت بسط داد:

$$Q(\varphi,\bar{\varphi})=\sum_{t=0}^{S-1}\sum_{k=0}^{K-1}\frac{p(y_t,k|\varphi)}{p(y_t|\varphi)}\log(p(\,y_t,k|\varphi) \qquad (12)$$

که در آن L تعداد ابعاد بردار در حوزه کپستروم و a_k همان عبارت رابطه (9) است.

با ساده سازی بیشتر معادله (12) خواهیم داشت:

(13)

$$Q(\varphi,\bar{\varphi})=constant-\frac{1}{2}\sum_{t=0}^{S-1}\sum_{k=0}^{K-1}P[k|y_t,\varphi]\left(y_t-\left(a_k+b_k\,\overline{h}+c_k\,\overline{n}\right)\right)^T\Sigma_{k,y}^{-1}\left(y_t-\left(a_k+b_k\,\overline{h}+c_k\,\overline{n}\right)\right)$$

برای پیدا کردن(n, h) باید Q ماکزیمم شود. بنابراین از آن مشتق گرفته برابر صفر قرار می‌دهیم:

$$\nabla_{\bar{h}}\,Q(\varphi,\bar{\varphi})=\sum_{t=0}^{S-1}\sum_{k=0}^{K-1}P[k|y_t,\varphi]\,\Sigma_{k,y}^{-1}\left(y_t-\left(a_k+\bar{h}\right)\right)=0 \qquad (14)$$

$$\nabla_{\bar{h}}\,Q(\varphi,\bar{\varphi})=\sum_{t=0}^{S-1}\sum_{k=0}^{K-1}P[k|y_t,\varphi]\,\Sigma_{k,y}^{-1}\left(y_t-\left(a_k+\overline{h}\right)\right)=0$$

دو معادله بالا را می‌توان به فرم ماتریسی در یک دستگاه معادلات بیان کرد:

$$\begin{bmatrix}d\\g\end{bmatrix}=\begin{bmatrix}E&F\\H&J\end{bmatrix}\begin{bmatrix}\overline{h}\\\overline{n}\end{bmatrix} \qquad (15)$$

که در آن خواهیم داشت:

(16)

$$d=\sum_{t=0}^{S-1}\sum_{k=0}^{K-1}P[k|y_t,\varphi]\;c_k^T\;\Sigma_{k,y}^{-1}\left(y_t-a_k\right)$$

$$E=\sum_{t=0}^{S-1}\sum_{k=0}^{K-1}P[k|y_t,\varphi]\;c_k^T\;\Sigma_{k,y}^{-1}\,b_k$$

$$F=\sum_{t=0}^{S-1}\sum_{k=0}^{K-1}P[k|y_t,\varphi]\;c_k^T\;\Sigma_{k,y}^{-1}\,c_k$$

$$g=\sum_{t=0}^{S-1}\sum_{k=0}^{K-1}P[k|y_t,\varphi]\;b_k^T\;\Sigma_{k,y}^{-1}\left(y_t-a_k\right)$$

$$H=\sum_{t=0}^{S-1}\sum_{k=0}^{K-1}P[k|y_t,\varphi]\;b_k^T\;\Sigma_{k,y}^{-1}\,b_k$$

$$J=\sum_{t=0}^{S-1}\sum_{k=0}^{K-1}P[k|y_t,\varphi]\;b_k^T\;\Sigma_{k,y}^{-1}\,c_k$$

با حل دسته معادلات (16) ، به روابط زیر برای مقدار بهینه بردار کانال و نویز در هر iteration می رسیم:

$$\begin{cases}\bar{h}=\left(H-J\,F^{-1}\,E\right)^{-1}\left(g-J\,F^{-1}d\right.\\\bar{n}=\left(J-H\,F^{-1}\,E\right)^{-1}\left(g-H\,F^{-1}d\right.\end{cases} \qquad (17)$$

ماتریس $\begin{bmatrix} E & F \\ H & J \end{bmatrix}$ باید معکوس پذیر باشد، در غیر این صورت راه حلی برای مساله بهینه سازی طرح شده وجود نخواهد داشت [4]. ممکن است با حالاتی روبرو شویم که یا هیچ جوابی وجود نداشته یا آن که بی نهایت جواب به دست آید. این اتفاق وقتی می‌افتد که مقادیر محاسبه شده برای $n\ ,\ h$ به ∞ میل کنند. برای اجتناب از این قضیه حد بالا و پایینی برای این سه پارامتر در نظر می‌گیریم تا مقادیر به دست آمده از آن تجاوز نکنند. حدود یاد شده از آزمایشات تجربی حاصل می‌شوند [10].

پس از آن که دو پارامتر مجهول $(n,\ h)$ به دست آمدند، آنها را به جای $(n_0,\ h_0)$ جانشین کرده و الگوریتم ذکر شده را تا جایی ادامه می‌دهیم که مقادیر تخمینی برای $(n,\ h)$ تفاوت چشمگیری با مقادیر قبلی نداشته باشند. در این حالت مقادیر بهینه دسته پارامترهای شرایط محیطی حاصل شده اند [11].

3-3- جبران سازی بردارهای نویزی

اکنون که تمامی مجهولات به دست آمده اند، با استفاده از تکنیک حداقل میانگین مربعات خطا، توانایی آن را داریم که بردار بازنمایی اصلاح شده هم‌ارز با هر بردار بازنمایی تخریب شده را به دست آوریم [6، 12]. در حالت کلی داریم:

$$\hat{x}_{MMSE} = E(x|y) = \int_X x\, p(x|y)\, dx \qquad (18)$$

با در نظر گرفتن x به عنوان تابعی از $y,\ g$ و تقریب g با بسط سری تیلور مرتبه اول آن خواهیم داشت:

$$\hat{x}_{MMSE} = y - \sum_{k=0}^{K-1} P[k|y]\ g(\mu_{k,x}, h, n) \qquad (19)$$

با استفاده از معادله (19)، می‌توان بردار جبران شده معادل با هر کپستروم نویزی دلخواه را به راحتی تخمین زد.

4- نتایج آزمایشات.

جهت ارزیابی کارایی الگوریتم ارایه شده در این مقاله، از مجموعه دادگان فارس‌دات میکروفونی که در آن دو جمله مشترک توسط 100 گوینده مختلف (در مجموع 200 جمله) ادا شده اند استفاده شده است. دو جمله مشترک انتخابی عبارتند از:

- " با روشن شدن هوا تظاهر کنندگان به سوی مجلس شورای ملی شروع به راهپیمایی کردند ".
- " مگر مژده اول چراغ قوه را خاموش نکرد".

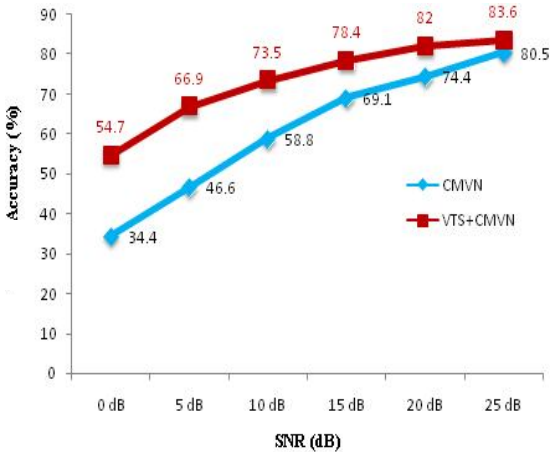
از این مجموعه دادگان 75% برای تعلیم مدل و 25% باقیمانده جهت آزمون آن اختصاص یافته اند. نرخ نمونه برداری سیگنال‌های گفتار 16 کیلو هرتز است.

روش کار بدین ترتیب است که از 200 جمله گفتار تمیز موجود، بردارهای بازنمایی 39 بعدی (شامل 12 ضریب MFCC به علاوه لگاریتم انرژی و مشتقات اول و دوم آن‌ها) استخراج شده و پس از نرمالیزه کردن بردارها به میانگین و واریانس (CMVN)، 75% از آنها به یک مدل مرجع بازشناسی مبتنی بر شبکه عصبیMLP تعلیم داده می شوند.

25% باقیمانده از دادگان گفتار تمیز را به صورت دستی با نویز سفید گوسی در SNR های صفر ، 5 ، 10 ، 15 ، 20 و 25 دسی‌بل مخلوط کرده و بردارهای بازنمایی MFCC نویزی را استخراج می‌نماییم. سپس بردارهای بازنمایی نویزی را پس از نرمالیزه کردن به میانگین و انحراف معیار، با استفاده از الگوریتم VTS اصلاح می‌کنیم. درنهایت این دو دسته دادگان اصلاح نشده و اصلاح شده به روش VTS، برای تست به مدل بازشناسی اعمال می‌شوند.

مدل بازشناسی، یک شبکه عصبی MLP تک لایه پنهان با ساختار نورونی < 34-100-39×7 > می‌باشد. در ورودی شبکه بردار بازنمایی فریم جاری به همراه بردارهای بازنمایی 3 فریم مجاور چپ و راست فریم جاری (در مجموع 7 فریم) قرار می گیرند. در لایه پنهان شبکه 100 نورون و در خروجی شبکه نیز به تعداد آواهای موجود در دادگان تعلیم یعنی 34 نورون در نظر گرفته شده است. نورون های هر دو لایه دارای توابع غیرخطی از نوع تانژانت هایپربولیک هستند. مقادیر وزن های اولیه شبکه به صورت تصادفی و در محدوده 1- تا 1 در نظر گرفته شده‌اند و ضریب یادگیری در شروع تعلیم برابر 0/001 و در حین تعلیم در هر تکرار با ضریب 0/95 کاهش می‌یابد که باعث تعلیم بهینه شبکه می‌گردد. ضریب مومنتم نیز برابر 0/2 در نظر گرفته شده است.

شبکه مذکور چندین بار با مقادیر وزن های تصادفی متفاوت اولیه تعلیم داده می‌شود و در نهایت نرخ صحت بازشناسی گفتار برای بردارهای بازنمایی تمیز برابر 82/7 درصد بدست آمده است. در نمودار شکل(3) نرخ بازشناسی بر اساس درصد صحت بازشناسی فریم‌های غیرسکوت برای دو دسته بردارهای بازنمایی نویزی و اصلاح شده و به ازای SNR های مختلف به صورت مقایسه ای بیان شده است.



شکل3: نتایج حاصل از بازشناسی بردارهای بازنمایی نویزی و جبران سازی شده به روش VTS به ازای SNR های مختلف

نتایج فوق نشان می دهد که بکارگیری روش VTS برای اصلاح بردارهای بازنمایی نه تنها برای بهبود بازنمایی‌های گفتار نویزی و نزدیکتر شدن نتایج به بازشناسی گفتار تمیز کمک می-کند بلکه بکارگیری این روش برای اصلاح دادگان گفتار نویزی 25 دسی بل ، نتایج بازشناسی بالاتری نسبت به نتیجه بازشناسی مدل بر روی گفتار تمیز (82/7%) دارد و این توانایی روش VTS را برای جبرانسازی اثر کانال را نیز نشان می‌دهد. زیرا بر روی دادگان گفتار میکروفونی فارس‌دات اثر تخریبی کانال ناشی از تنوعات میکروفون و گوینده نیز وجود دارد.

5- نتیجه گیری

با توجه به اهمیت مقاوم سازی سیستم‌های بازشناسی گفتار در برابر تنوعات محیطی موجود در کاربردهای عملی و نظر به عدم کارایی روش‌های متداول استخراج ویژگی مانند MFCC در برابر این تنوعات، در این مقاله تکنیک تحلیلی و مبتنی بر فرمول بندی دقیق ریاضی بردار سری تیلور VTS برای اصلاح و جبران سازی بردارهای بازنمایی صوتی ارائه گردید. نتایج حاصل از پیاده سازی این تکنیک بر روی مجموعه دادگان گفتاری فارسی تخریب شده با نویز سفید که اثر کانال انتقال را نیز در قالب تنوعات میکروفون و گوینده دارا می باشد، به خوبی نشان دهنده تاثیر آن در حذف اثر نویز جمع شونده و کانال انتقال خطی از روی سیگنال گفتار هستند. بهبود درصد بازشناسی به ویژه در SNR های پایین بسیار قابل توجه است. می‌توان با استفاده از مرتبه‌های بالاتر بسط سری تیلور تابع شرایط محیطی و نیز روش های موجود برای مقدار دهی اولیه صحیح‌تر پارامترهای تابع تنوعات محیطی، به نتایجی بهتر به ویژه در SNR های بسیار پایین دست یافت[12].

سپاسگزاری

بر خود لازم می‌دانیم از زحمات و حمایت‌های بی دریغ سرکار خانم مهندس پورمحمدی که بی یاری ایشان انجام پروژه فوق میسر نبود تشکر کنیم.

6- مراجع

[1] منصور ولی، *" بازشناسی مقاوم گفتار به منظور جبرانسازی تنوعات گفتار میکروفنی و تلفنی توسط شبکه‌های عصبی "*، پایان نامه دکتری، دانشگاه صنعتی امیرکبیر، بهار 1385.

[2] Pedro J. Moreno, Bhiksha Raj and Richard M. Stern, “A Vector Taylor Series Approach for Environment-Independent Speech Recognition”, Microsoft Research Center, pp, 1-4, 2008

[3] R. C. van Dalen and M. J. F. Gales, “Extended VTS for Noise-Robust Speech Recognition”, InterSpeech, pp, 1-4, 2009

[4] Pedro J.Moreno, “Speech Recognition in Noisy Environments”, Thesis, pp, 79-104, 1996

[5] Do Yeong Kim, Chang Kwan Un and Nam Soo Kim, “Speech Recognition in Noisy Environments Using First-Order Vector Taylor Series”, Speech Communication, pp, 1-4, 1998

[6] O. Kalinli, M. L. Seltzer and A. Acero, “Noise Adaptive Trainig Using Vector Taylor Series Approach for Noise Robust Automatic Speech Recognition”, Microsoft Research Center, pp, 1-3, 2009

[7] Nam Soo Kim, Do Yeong Kim, Byung Goo Kong and Sang Ryong Kim, “Application of VTS to environment Compansation with Noise Statistics”, ICASP, pp, 1-4, 1999

[8] Alex Acero, Li Deng andTrausti Kristjansson and Jerry Zhang, “HMM Adaptation Using Vector Taylor Series for Noisy Speech Recognition”, Microsoft Research, pp, 1-4, 2000

[9] Jasha Droppo, “Noise Robust Automatic Speech Recognition”, Microsoft Research Center, pp, 12-43, 2008

[10] R. C. van Dalen and M. J. F. Gales, “Covariance Modeling for Noise Robust Speech Recognition”, Interspeech, pp, 1-4, 2008

[11] Chandra Kaut Raut, Takuya Nishimoto, Shigeki Sagayama, “Model Composition by Lagrange Polynomial Approximation for Robust Speech Recognition in noisy Environment”, Interspeech, pp, 1-4, 2008

[12] Zhao Xianyu, Ou Zhijian and Wang Zuoying, “Using Vector Taylor Series with Noise Clustering for Speech Recognition in non-stationary Noisy Environments”, High Technology Letters, pp, 1-5, 2006