بهبود کیفیت گفتار نویزی باند محدود با تلفیق الگوریتم‌های سری تیلور برداری و گسترش پهنای باند

سارا پورمحمدی، منصور ولی و محسن قدیانی

چکیده: در مقاله حاضر با تلفیق دو دیدگاه سری‌های تیلور برداری و گسترش پهنای باند مصنوعی، ایده جدیدی در زمینه بهبود کیفیت سیگنال گفتار باند محدود تخریب‌شده توسط نویز ارائه شده است. بدین ترتیب که ابتدا پارامترهای بازنمایی MFCC استخراج‌شده از گفتار نویزی باند محدود به روش سری‌های تیلور برداری اصلاح شده و سپس با استفاده از مدل گسترش پهنای باند مبتنی بر GMM، بردارهای بازنمایی گفتار باند گسترده برای این پارامترهای اصلاح‌شده تخمین زده می‌شوند. سپس به کمک دو معیار اندازه‌گیری PESQ و LSD، میزان شباهت پوش طیف و سیگنال گفتار تخمین زده شده باند گسترده با پوش طیف باند گسترده و گفتار تمیز مرجع سنجیده می‌شود. نتایج به دست آمده از پیاده‌سازی این الگوریتم به وضوح بیانگر کارایی مناسب ایده پیشنهادی در جهت بهبود کیفیت بردارهای بازنمایی گفتار باند محدود آلوده به نویز و نزدیک‌تر کردن آنها به بردارهای ویژگی سیگنال گفتار باند گسترده مرجع هستند.

کلید واژه: **سری‌های تیلور برداری، گسترش پهنای باند، گفتار نویزی باند محدود، مدل ترکیب گوسی.**

۱- مقدمه

نویز جمع‌شونده و کانال انتقال با پهنای باند محدود از جمله مهم‌ترین عوامل کاهش کیفیت سیگنال گفتار در کاربردهای عملی هستند. نویز جمعی ناشناخته از طریق شیفت‌دادن میانگین طیفی و افزایش واریانس کل توزیع احتمال سیگنال روی آن تأثیر می‌گذارد. از طرف دیگر محدودیت پهنای باند کانال انتقال نیز با حذف مؤلفه‌های فرکانس بالای سیگنال گفتار، کیفیت و ادراک‌پذیری آن را تا حدود زیادی کم می‌کند [۱].

برای حذف این عوامل مزاحم محیطی تاکنون روش‌های متعددی پیشنهاد شده‌اند. برخی از این تکنیک‌ها مانند سری‌های تیلور برداری (VTS)[1] و تئوری دادگان مفقود بیشتر بر روی حذف اثر نویز و اعوجاج کانال از روی سیگنال گفتار متمرکز شده‌اند و برخی دیگر مانند مجموعه تکنیک‌های گسترش پهنای باند (BWE)[2] تنها به جبران محدودیت پهنای باند فرکانسی ناشی از کانال انتقال توجه دارند. اما از آنجا که در اکثر کاربردهای عملی، کانال انتقال، تلفیقی از دو پدیده کاهش پهنای باند و افزودن نویز به سیگنال گفتار را به همراه دارد و هر دو عامل به طور

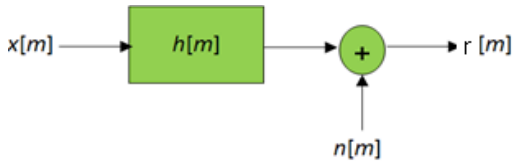
این مقاله در تاریخ ۵ تیر ماه ۱۳۹۱ دریافت و در تاریخ ۳ اسفند ماه ۱۳۹۱ بازنگری شد.

سارا پورمحمدی، دانشکده فنی و مهندسی، دانشگاه شاهد، تهران، (email: s_pourmohammadi@yahoo.com).

منصور ولی، دانشکده فنی و مهندسی، دانشگاه خواجه نصیر الدین طوسی، تهران، (email: mansour.vali@eetd.kntu.ac.ir).

محسن قدیانی، دانشکده فنی و مهندسی، دانشگاه شاهد، تهران، (email: mohsenghadyani@ymail.com).

1. Vector Taylor Series
2. Bandwidth Extension



شکل ۱: مدل تأثیر تنوعات محیطی بر روی سیگنال گفتار [۲].

هم‌زمان بر روی کاهش کیفیت گفتار تأثیرگذار هستند، نیاز است تا با تلفیق مناسب این دو تکنیک، گام‌های مؤثرتری برای جبران‌سازی تنوعات محیطی ذکرشده برداشته شود. مقاله حاضر به شرح چگونگی ترکیب دو ایده سری‌های تیلور برداری و گسترش پهنای باند با هدف نزدیک‌تر کردن پارامترهای بازنمایی گفتار تخریب‌شده باند محدود به پارامترهای گفتار تمیز باند گسترده، و در نتیجه بهبود کیفیت سیگنال دریافتی در خروجی کانال انتقال اختصاص یافته است. در این تحقیق از تکنیک سری‌های تیلور برداری برای حذف نویز بردارهای بازنمایی MFCC[3] استفاده شده و گسترش پهنای باند با استفاده از مدل ترکیب گوسی (GMM)[4] انجام شده است.

در ادامه، ابتدا الگوریتم سری‌های تیلور برداری به عنوان ابزاری توانمند برای جبران اثرات نویز جمع‌شونده از روی بردارهای بازنمایی به صورت خلاصه مورد بررسی قرار می‌گیرد. سپس در بخش ۴ شرح کوتاهی در زمینه مجموعه روش‌های گسترش پهنای باند سیگنال گفتار با تأکید بر تکنیک مدل ترکیب گوسی ارائه می‌گردد. بخش ۵ به شرح ایده ترکیبی معرفی‌شده در این تحقیق پرداخته و نتایج حاصل از پیاده‌سازی آنها بر روی دادگان گفتاری واقعی را مورد بررسی قرار می‌دهد. در انتها نیز بحث و تحلیل نتایج ارائه خواهد گردید.

۲- معرفی

وقتی سیگنال گفتار مرجع تحت اثر تنوعات محیطی قرار گیرد، نمایش برداری گفتار تخریب‌شده در حالت کلی به صورت (۱) خواهد بود

$$r = x + f(x, a_1, a_2, \ldots) \qquad (1)$$

که در آن x بردار بازنمایی گفتار مرجع و f تابع تنوعات محیطی است و پارامترهای $a_1, a_2, \ldots$ عوامل مزاحم تأثیرگذار بر روی سیگنال گفتار هستند. با محدودکردن این تنوعات به نویز جمع‌شونده و کانال انتقال خطی ناشناخته اما تغییرناپذیر با زمان که دارای پهنای باند محدود است، شکل ۱ برای نمایش نحوه تأثیرگذاری شرایط محیطی بر روی گفتار مرجع نتیجه خواهد شد.

برای چنین سیستمی ارتباط بین طیف توان دادگان گفتاری تمیز و تخریب‌شده به صورت (۲) نشان داده می‌شود

3. Mel Frequency Cepstral Coefficients
4. Gaussian Mixture Model

$$|R(\omega)| = |X(\omega)|.|H(\omega)|^{2} + |N(\omega)| \quad (۲)$$

که در آن $X(\omega)$، $N(\omega)$ و $R(\omega)$ به ترتیب چگالی طیف توان گفتار تمیز باند کامل، نویز و گفتار تخریب‌شده - سیگنال گفتاری که از کانال با پهنای باند محدود عبور کرده و نویز نیز به آن افزوده شده است - هستند و $|H(\omega)|^{2}$ طیف توان کانال خطی انتقال است. با لگاریتم‌گیری از دو طرف (۲) و جایگزینی $\log|X(\omega)|$، $\log|H(\omega)|^{2}$، $\log|N(\omega)|$ و $\log|R(\omega)|$ به ترتیب با x، h، n و r (۳) در حوزه لگاریتم طیف توان به دست می‌آید

$$r = x + h + \log(1 + \exp(n - x - h)) \quad (۳)$$

با اندکی عملیات محاسباتی می‌توان نگاشت گفتار تمیز و نویزی را از حوزه لگاریتم طیف به حوزه کپستروم که در کاربردهای پردازش گفتار متداول‌تر است منتقل نمود [۲]

$$r_c = x_c + h_c + C\log(1 + \exp(C^{-1}(n_c - x_c - h_c))) \quad (۴)$$

که اندیس c نشان‌دهنده حوزه کپستروم، C ماتریس تبدیل کسینوسی گسسته[1] (DCT) و C^{-1} معکوس آن است. در تحقیق حاضر روند بازیابی گفتار تمیز باند گسترده در حوزه کپستروم مورد بررسی قرار می‌گیرد.

تکنیک مورد استفاده برای حذف نویز از دادگان گفتاری باند محدود در این تحقیق، سری‌های تیلور برداری است که در بخش بعدی به اختصار معرفی خواهد شد.

۳- الگوریتم سری‌های تیلور برداری

تکنیک سری‌های تیلور برداری از جمله کاراترین روش‌های حذف اثر تنوعات مزاحم محیطی سیگنال گفتار به شمار می‌رود که به ویژه در بازشناخت گفتار، در هر دو حوزه اصلاح پارامترهای بازنمایی و جبران‌سازی مدل بازشناسی کاربرد دارد. این ایده الگوریتمی تحلیلی و مبتنی بر محاسبات دقیق ریاضی را برای تخمین و حذف پارامترهای مزاحم محیطی اثرگذار بر روی کیفیت گفتار و بازشناسی آن معرفی می‌کند. ایده مذکور را اولین بار مورنو برای نگاشت بردارهای بازنمایی نویزی به تمیز به کار گرفت [۳] و از آن پس همواره یکی از مهم‌ترین زمینه‌های مورد علاقه در مباحث مربوط به بهبود کیفیت گفتار و بازشناسی مقاوم بوده است.

به طور خلاصه می‌توان مهم‌ترین مزایای استفاده از روش VTS برای جبران‌سازی تنوعات محیطی تأثیرگذار بر روی سیگنال گفتار را چنین بر شمرد:

۱) نیاز به کمترین حجم اطلاعات گفتاری و دانش اولیه در مورد مدل تنوعات محیطی.
۲) تحلیلی و مبتنی بر فرمول‌بندی دقیق ریاضی‌بودن الگوریتم که امکان اجرای گام به گام و ساخت‌یافته آن را فراهم آورده است.

با فرض وجود داده‌های در دسترس زیر:

- مجموعه‌ای از بردارهای بازنمایی‌های آلوده به نویز و اثر کانال در حوزه کپستروم $R = \{r_0, r_1, \ldots, r_{S-1}\}$
- تابع چگالی احتمال[2] (PDF) گفتار تمیز
- مجموعه‌ای از مقادیر اولیه برای نویز و کانال شامل میانگین و کواریانس نویز و بردار کانال

و با در نظر گرفتن مدل فرض‌شده برای اثر تنوعات مزاحم، الگوریتم گام به گام جبران‌سازی پارامترهای محیطی (که شامل محدودیت پهنای باند و نویز جمع‌شونده هستند) به ترتیب زیر بیان می‌گردد:

۱) تعلیم مدل GMM برای گفتار تمیز و تخمین پارامترهای تابع توزیع احتمال آن
۲) انتخاب مقادیر اولیه دسته پارامترهای مجهول
۳) بسط تابع تنوعات محیطی برای هر یک از توزیع‌های گوسی گفتار تمیز حول $\{\mu_{n.}, \Sigma_{n.}, h_{n.}\}$
۴) تخمین پارامترهای توزیع احتمال بردار ویژگی گفتار نویزی $(\mu_{r,k}, \Sigma_{r,k})$
۵) انجام یک بار الگوریتم EM[3] برای تخمین مجدد دسته مجهولات $\{\mu_n, \Sigma_n, h\}$
۶) در صورت همگراشدن تابع شباهت مقادیر بهینه مجهولات حاصل شده‌اند. در غیر این صورت دسته مجهولات اولیه $\{\mu_{n.}, \Sigma_{n.}, h_{n.}\}$ با مقادیر جدید به دست آمده آنها شامل $\{\mu_n, \Sigma_n, h\}$ جایگذاری شده و الگوریتم برای تکرار به مرحله ۳ باز می‌گردد.
۷) تخمین بردارهای بازنمایی تمیز از روی معادل‌های نویزی با استفاده از مقادیر تقریب زده شده نویز و کانال به کمک تکنیک کمینه‌سازی میانگین مربعات خطا [۳].

شکل ۲ بلوک دیاگرام استخراج بردارهای بازنمایی اصلاح‌شده به ازای هر یک از بردارهای ویژگی تخریب‌شده در حوزه لگاریتم طیف را نمایش می‌دهد.

تابع توزیع احتمال بردارهای بازنمایی گفتار تمیز مطابق با (۵) به صورت مجموعی از توزیع‌های گوسی فرض می‌شود

$$p(x) = \sum_{k=1}^{K} p(k).N(x, \mu_{x,k}, \Sigma_{x,k}) \quad (۵)$$

در این رابطه K، $\mu_{x,k}$ و $\Sigma_{x,k}$ به ترتیب برابرند با تعداد کل توزیع‌های گوسی و بردار میانگین و ماتریس کواریانس kامین توزیع. اکنون فرض می‌شود که تأثیر نویز و کانال بر روی گفتار تمیز، مدل توزیع GMM آن را بر هم نمی‌زند و می‌توان توزیع احتمال گفتار تخریب‌شده را نیز به صورت مجموعی از توزیع‌های گوسی تکی فرض کرد [۴]

$$p(r) = \sum_{k=1}^{K} p(k).N(r, \mu_{r,k}, \Sigma_{r,k}) \quad (۶)$$

در (۶) r، $p(k)$، $\mu_{r,k}$ و $\Sigma_{r,k}$ به ترتیب نماینده بردار بازنمایی گفتار نویزی، احتمال اولیه، بردار میانگین و ماتریس کواریانس kامین توزیع گوسی هستند. بنابراین در صورتی که پارامترهای مدل توزیع احتمال گفتار تمیز در اختیار باشند، می‌توان با یک نگاشت مؤلفه‌های نظیر برای گفتار نویزی به ازای هر یک از توزیع‌های گوسی آن را به دست آورده و در طی این فرایند پارامترهای تابع تنوعات محیطی را نیز تخمین زد.

تخمین پارامترهای توزیع به روش ماکزیمم شباهت[4] امکان‌پذیر است. اساس عملکرد این روش، حداکثرسازی احتمال مشاهدات تولیدشده از یک توزیع با مجموعه‌ای از پارامترها است. این کار به وسیله تنظیم پارامترها به گونه‌ای که شباهت بین آنها به حداکثر مقدار خود برسد انجام می‌پذیرد. الگوریتم EM این تخمین را به صورت مداوم تکرار کرده و در هر بار تکرار میزان شباهت را افزایش می‌دهد تا مقادیر مطلوب ۳ پارامتر حاصل گردد.

1. Discrete Cosine Transform
2. Probability Density Function
3. Expectation Maximization
4. Maximum Likelihood

پس از تخمین پارامترهای گفتار تمیز می‌توان (۳) را به صورت بسط سری تیلور مرتبه اول تابع تنوعات حول نقطه اولیه انتخابی $(n_.,x_.,h_.)$ تخمین زد. رابطه (۷) نشان‌دهنده این تقریب است

$$\begin{aligned} r \cong x + f(n_.,x_.,h_.) + \nabla_x f(n_.,x_.,h_.)x \\ + \nabla_n f(n_.,x_.,h_.)n + \nabla_h f(n_.,x_.,h_.)h \end{aligned} \quad (۷)$$

با استفاده از (۷) می‌توان بردار میانگین و ماتریس کواریانس هر یک از توزیع‌های گوسی توزیع احتمال گفتار نویزی را به ترتیب با (۸) و (۹) بیان کرد [۳]

$$\begin{aligned} \mu_{r,k} = (I + \nabla_x f)'\mu_{r,k} + (\nabla_n f)'\mu_n \\ + (\nabla_h f)'h + g(n_.,x_.,h_.) \end{aligned} \quad (۸)$$

$$\Sigma_{r,k} = (I + \nabla_x f)'\Sigma_{x,k}(I + \nabla_x f) + (\nabla_n f)'\Sigma_n(\nabla_n f) \quad (۹)$$

در روابط فوق $\mu_{r,k}$، $\Sigma_{r,k}$، $\mu_{x,k}$ و $\Sigma_{x,k}$ به ترتیب میانگین و کواریانس kامین توزیع گفتار تخریب‌شده و گفتار تمیز، و μ_n، Σ_n و h پارامترهای نویز و اعوجاج خطی هستند. $(\nabla_x f)'$، $(\nabla_n f)'$ و $(\nabla_h f)'$ نیز ترانهاده گرادیان تابع تنوعات محیطی به ترتیب نسبت به بردار گفتار تمیز، نویز و کانال هستند و I بردار واحد هم‌مرتبه با بردارهای گرادیان است. اما علاوه بر این دو مؤلفه نیاز به تخمینی صحیح از احتمال پسین[1] هر یک از توزیع‌های گوسی وجود دارد. به این منظور تابع چگالی احتمال گفتار تخریب‌شده به صورت (۱۰) نمایش داده می‌شود

$$p(r|k) = \sum_{k=1}^{K} p(k)p(r|k,\lambda) = \sum_{k=1}^{K} p(k)N(r,\mu_{r,k},\Sigma_{r,k}) \quad (۱۰)$$

که در آن $p(k)$، λ و N به ترتیب عبارتند از احتمال اولیه kامین توزیع گوسی، دسته پارامترهای مجهول و تابع احتمال گفتار نویزی به ازای kامین توزیع. می‌توان احتمال پسین هر توزیع مدل GMM را طبق (۱۱) تخمین زد [۵]

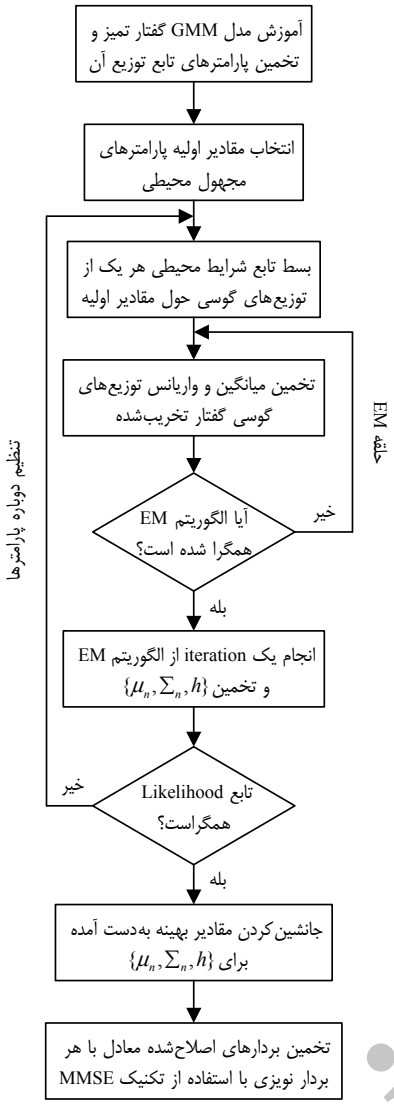
$$p(r|k) = \frac{p(k)p(r|k,\lambda)}{\sum_{k'=1}^{K} p(k')p(r|k',\lambda)} \quad (۱۱)$$

پس از تخمین مؤلفه‌های گفتار نویزی شامل احتمالات اولیه و پسین مدل توزیع و میانگین و کواریانس گفتار تخریب‌شده، فرایند تخمین پارامترهای شرایط محیطی با استفاده از الگوریتم EM تکراری و تلاش در جهت حداکثرسازی شباهت بین نمونه‌های مشاهده‌شده و نمونه‌های مرجع انجام می‌شود. این روند با به کارگیری یک الگوریتم ماکزیمم تخمین که به صورت تکراری پارامترهای نویز و اعوجاج کانال را به روز می‌کند امکان‌پذیر است.

در تحقیق حاضر الگوریتم VTS تنها به منظور حذف اثر نویز از روی بردارهای بازنمایی سیگنال گفتار استفاده شده و گسترش پهنای باند آنها با استفاده از تکنیک‌های BWE انجام می‌شود. از این رو الگوریتم EM تنها دو مؤلفه بردار میانگین و ماتریس کواریانس نویز را به صورت تکراری تخمین می‌زند. معادلات نهایی تخمین دو پارامتر مذکور در هر بار تکرار الگوریتم تخمین در (۱۲) و (۱۳) بیان شده‌اند [۶]

$$\hat{\mu}_n = \frac{\sum_{t=1}^{T}\sum_{k=1}^{K} p(k|r_t,\lambda)\mu_n(r_t,k,\lambda)}{\sum_{t=1}^{T}\sum_{k=1}^{K} p(k|r_t,k,\lambda)} \quad (۱۲)$$

1. A Posteriori Probability



شکل ۲: بلوک دیاگرام بازیابی بردارهای بازنمایی تمیز از دادگان نویزی به روش سری‌های تیلور برداری [۳].

$$\begin{aligned} \hat{\Sigma}_n = \\ \frac{\sum_{t=1}^{T}\sum_{k=1}^{K} p(k|r_t,\lambda)[\Sigma_n(r_t,k,\lambda) + \mu_n(r_t,k,\lambda)\mu_n'(r_t,k,\lambda)]}{\sum_{t=1}^{T}\sum_{k=1}^{K} p(k|r_t,k,\lambda)} \\ - \hat{\mu}_n\mu_n' \end{aligned} \quad (۱۳)$$

که در آن $\hat{\mu}_n$ و $\hat{\Sigma}_n$ به ترتیب بردار میانگین و ماتریس کواریانس نویز در هر بار تکرار الگوریتم EM هستند. فرایند تکراری تخمین تا همگرایی دو مؤلفه نویز و دستیابی به مقادیر بهینه آنها ادامه می‌یابد.

اکنون که تخمین مناسبی از نویز جمع‌شونده با سیگنال گفتار در حوزه لگاریتم طیف حاصل شد، راه حل مناسب برای تقریب بردار بازنمایی تمیز از معادل نویزی آن استفاده از روش حداقل‌سازی میانگین مربعات خطا (MMSE)[2] است که (۱۴) نشان‌دهنده عبارت نهایی مورد استفاده برای آن است [۳]

$$\begin{aligned} \hat{x}_{mmse} &= r - \sum_{k=1}^{K} p(k|r)g(n,x,h) \\ &= r - \sum_{k=1}^{K} p(k|\lambda)(\mu_{x,k} - \mu_{r,k}) \end{aligned} \quad (۱۴)$$

2. Minimum Mean Square Error

در (۱۴) r و $\hat{x}_{mmse}$ به ترتیب بردار بازنمایی نویزی و اصلاح‌شده هستند. پس از حذف نویز از روی بردارهای بازنمایی تخریب‌شده باند محدود، پارامترهای بازنمایی اصلاح‌شده با باند فرکانسی محدود حاصل شده‌اند و در مرحله بعد باید بردارهای ویژگی اصلاح‌شده باند گسترده را تخمین زد. این کار با استفاده از ایده گسترش پهنای باند گفتار امکان‌پذیر خواهد بود.

٤- گسترش پهنای باند سیگنال گفتار

سیگنال گفتار در شرایط معمول پهنای باند صفر تا ۱۰ کیلوهرتز دارد اما در شرایطی مانند انتقال از طریق خط تلفن، پهنای باند آن محدود می‌شود. این محدودیت موجب از دست رفتن بخشی از اطلاعات سیگنال گفتار شده و در نتیجه کیفیت و ادراک‌پذیری آن را کاهش می‌دهد. مجموعه تکنیک‌های گسترش پهنای باند با افزودن بخش‌های از دست رفته طیف سیگنال گفتار باند محدود، گفتار باند گسترده را بازسازی و در نتیجه کیفیت و ادراک‌پذیری آن را تا حد زیادی به گفتار مرجع نزدیک می‌کنند.

سیستم تولید گفتار انسان شامل دو بخش اصلی است. بخش اول تولید سیگنال تحریک منبع و بخش دوم فیلتر لوله صوتی است که در مسیر تحریک یادشده قرار گرفته است [۷]. بر این اساس مدل پیشنهادی برای بازسازی سیگنال گفتار باند گسترده از روی گفتار باند باریک نیز شامل دو مرحله مجزا خواهد بود: یکی گسترش سیگنال تحریک[1] و دیگری گسترش پوش طیف گفتار[2] [۸]. تخمین پوش طیف به طور معمول مرحله دشوارتری از تخمین سیگنال تحریک بوده و به کیفیت پارامترهای بازنمایی مورد استفاده برای تخمین وابسته است. در این تحقیق نیز بر بخش گسترده‌سازی پوش طیف تمرکز شده است و الگوریتم تلفیقی پیشنهادی در مرحله بازسازی پوش طیف سیگنال باند گسترده مورد ارزیابی قرار می‌گیرد.

پوش طیف به طور معمول با ضرایب فیلتر لوله صوتی یا AR[3] معرفی می‌شود اما استفاده از بردارهای بازنمایی مانند ضرایب کپسترال[4]، ضرایب LPC[5]، ضرایب LSF[6]، ضرایب MFCC و یا ضرایب خودهمبستگی نیز مرسوم هستند [۹]. در صورت استفاده از پارامترهای بازنمایی نام‌برده، پس از تخمین ضرایب سیگنال باند گسترده در مرحله بازتولید گفتار باید به ضرایب AR تبدیل شوند.

بردارهای بازنمایی LSF به طور گسترده در گسترش پهنای باند گفتار مورد استفاده قرار می‌گیرند. مهم‌ترین ویژگی این دسته بردارهای بازنمایی، قابلیت بازگشت‌پذیری ساده‌تر آنها به ضرایب LP نسبت به دیگر بردارهای ویژگی از جمله MFCC است. اما اخیراً به دلایل متعدد استفاده از بردارهای بازنمایی MFCC مورد توجه قرار گرفته است. به طور مثال از نتایج به دست آمده از [۹] که گسترش پهنای باند را با انواع مختلف ضرایب بازنمایی مورد مقایسه قرار داده است، این گونه استنباط می‌گردد که پارامترهای MFCC در نمایش اطلاعات متقابل[7] و تفکیک‌پذیری[8] نسبت به دیگر بردارهای بازنمایی عملکرد بهتری دارند. همچنین در [۱۰] نشان داده شده است که قطعیت[9] نتایج حاصل از ضرایب MFCC، به طور میانگین دو برابر نتایج حاصل از ضرایب LSF است. در [۱۱] نیز یک سیستم گسترش پهنای باند به روش GMM با استفاده از هر دو دسته بردارهای بازنمایی MFCC و LSF پیاده‌سازی شده و برتری ضرایب MFCC نسبت به LSF در نمایش همبستگی موجود بین باندهای مختلف فرکانسی ثابت شده است. لذا می‌توان نتیجه‌گیری کرد که سودبردن از بردارهای بازنمایی MFCC در گسترش پهنای باند، با وجود دشواری بازتولید سیگنال گفتار، بر استفاده از ضرایب مشابه مانند LSF یا LPC برتری دارد. به همین دلیل در این بخش روش معکوس تبدیل کسینوسی گسسته (IDCT)[10] برای بازتولید پوش طیف توسط ضرایب MFCC با بهترین کیفیت توصیف گردیده است.

روش‌های متنوعی برای تخمین پوش طیف باند گسترده معرفی شده است. روش کتاب کد[11] یکی از اولین روش‌های پیشنهادی است که شامل یک مجموعه از پیش تعیین شده از پوش طیف‌های باند محدود و باند بالای مربوطه می‌باشد. اطلاعات پوش برای هر فریم از سیگنال باند محدود، با همه داده‌های کتاب کد مقایسه شده و نمونه دارای بهترین تطابق انتخاب می‌شود [۷]. ایده مدل مخفی مارکوف (HMM)[12] نیز از جمله تکنیک‌های رایج در بازشناسی گفتار بوده و از آن در گسترش پهنای باند استفاده می‌شود. در HMM یک مدل آماری نحوه وابستگی موجود بین بردار ویژگی استخراج‌شده از سیگنال باند محدود و هر حالت از مدل HMM را بیان می‌کند [۱۲]. همچنین مدل ترکیب گوسی روش متداولی در تخمین پارامترهای پوش طیف باند بالا به شمار می‌رود. در روش GMM می‌توان تابع توزیع احتمال دادگان را مدل کرد. GMM به وسیله الگوریتم EM آموزش می‌بیند و سپس یک تخمین‌زننده خطای میانگین مربعات بین پوش طیف اصلی و تخمین زده شده را مینیمم می‌کند. اگرچه استفاده از مدل‌های آماری در تخمین پوش طیف مرسوم‌تر است اما استفاده از شبکه عصبی نیز نتایج قابل قبولی در پی داشته است. در روش شبکه عصبی یک نگاشت از بردارهای باند محدود به باند گسترده انجام می‌پذیرد [۱۳]. بخش بعد به توصیف نحوه گسترش پوش طیف سیگنال گفتار به روش مدل ترکیب گوسی اختصاص دارد.

٤-۱ گسترش پوش طیف به روش GMM

پس از اصلاح بردارهای بازنمایی به روش VTS، با ترکیب بردار بازنمایی باند محدود x و باند بالای y ماتریس دادگان تعلیم به صورت $z^T = [x^T y^T]$ ساخته می‌شود. ماتریس z به عنوان ورودی مدل تخمین به کار رفته و توزیع GMM مفروض برای آن مشابه با بخش ۳ مدل می‌شود، با این تفاوت که ماتریس‌های کواریانس جهت مدل‌کردن همبستگی موجود بین مؤلفه‌های توزیع و دقیق‌تر بودن تخمین به صورت کامل[13] (غیر قطری) در نظر گرفته می‌شوند. در صورتی که پارامترهای تابع چگالی احتمال در دسترس باشند، می‌توان y را بر اساس مجموعه مشاهدات x با استفاده از روش مینیمم‌سازی میانگین مربعات خطا تخمین زد. اما از آنجا که راه حلی تحلیلی برای یافتن پارامترهای مورد اشاره وجود ندارد، این کار مشابه با تکنیک VTS با استفاده از الگوریتم EM تکراری انجام می‌گیرد.

1. Expansion of Excitation
2. Expansion of Envelope
3. Auto Regressive
4. Cepstral Coefficients
5. Linear Predictive Coefficients
6. Line Spectral Frequencies
7. Mutual Information
8. Separability
9. Certainity
10. Inverse Discrete Cosine Transform
11. Codebook
12. Hidden Marcov Model
13. Full

پس از تخمین پارامترهای مدل GMM می‌توان پوش طیف توان گفتار باند بالا را به کمک روش MMSE بازسازی نمود. در این تکنیک هدف کمینه‌سازی میانگین مربعات فاصله موجود بین بردارهای بازنمایی باند گسترده تخمین زده شده $\hat{y}$ و بردارهای باند گسترده واقعی y است و از آن جهت که از ماتریس کواریانس کامل استفاده شده است، معادله آن به صورت (۱۵) خواهد بود

$$\hat{y}_{mmse} = \sum_{k=0}^{K} p(k|x,\lambda)[\mu_k^y + \Sigma_k^{yx}(\Sigma_k^{xx})^{-1}(x-\mu_k^x)] \qquad (۱۵)$$

که در آن $P(k|x,\lambda)$ احتمال kامین توزیع مدل GMM مفروض است که بردارهای بازنمایی x را تولید نموده و خود از (۱۶) حاصل می‌شود

$$P(k|x,\lambda) = \frac{p(k)p(x|\mu_k^x,\Sigma_k^{xx})}{\sum_{k'=1}^{K} p(k)p(x|\mu_{k'}^x,\Sigma_{k'}^{xx})} \qquad (۱۶)$$

μ_k^y و μ_k^x بخشی از بردار میانگین و Σ_k^{yx} و Σ_k^{xx} بخشی از ماتریس کواریانس kامین مؤلفه می‌باشند که مطابق (۱۷) و (۱۸) از تجزیه μ_m^z و Σ_m^z به دست می‌آیند [۷]

$$\mu_k^z = \begin{bmatrix} \mu_k^x \\ \mu_k^y \end{bmatrix} \qquad (۱۷)$$

$$\Sigma_k^z = \begin{bmatrix} \Sigma_k^{xx} & \Sigma_k^{xy} \\ \Sigma_k^{yx} & \Sigma_k^{yy} \end{bmatrix} \qquad (۱۸)$$

۴-۲ بازتولید گفتار باند گسترده

پس از تخمین ضرایب بازنمایی در حوزه کپستروم به کمک تکنیک GMM، لازم است تا پارامترهای بازنمایی به دست آمده به ضرایب فیلتر لوله صوتی تبدیل شوند و امکان بازتولید و ارزیابی گفتار با استفاده از مدل پیش‌بین خطی فراهم گردد.

تعدادی از مراحل ساخت ضرایب MFCC قابل بازگشت هستند اما بقیه مراحل برگشت‌پذیر نبوده و موجب از دست رفتن اطلاعات مفید موجود در سیگنال گفتار می‌شوند. با این وجود همچنان ممکن است که تخمینی منطقی و هر چند هموارشده از دامنه طیف توان گفتار به دست آورد. با استفاده از این دامنه طیف نیز می‌توان تخمینی از ضرایب فیلتر لوله صوتی حاصل نمود. بنابراین در حالت کلی محاسبه ضرایب فیلتر لوله صوتی از روی پارامترهای بردار بازنمایی MFCC فرایندی ۲مرحله‌ای است:

۱) بازیابی طیف سیگنال از روی بردارهای بازنمایی MFCC

۲) محاسبه ضرایب مدل پیش‌بین خطی از روی دامنه طیف مذکور

اولین گام در فرایند فوق، استفاده از معکوس تبدیل کسینوسی گسسته است که در (۱۹) بیان گردیده است

$$\log \hat{Y}_k = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} c_n \cos\left(\frac{(2k+1)n\pi}{2N}\right) \quad , \quad 0 < k < N-1 \qquad (۱۹)$$

که در این رابطه $\hat{Y}_k$، c_n و N به ترتیب طیف توان تخمین زده شده، بردارهای بازنمایی MFCC و تعداد فیلترها در فیلتر بانک می‌باشند. برای معکوس‌سازی عمل لگاریتم از عملگر نمایی استفاده می‌شود. با اعمال تابع نمایی می‌توان تخمینی از بردارهای خروجی فیلتر بانک مل در اختیار داشت. مرحله بعدی شامل تخمین دامنه طیف توان است. از آنجا که فرایند معکوس‌سازی با استفاده از تعداد محدودی از پارامترهای بردار بازنمایی انجام می‌گیرد، در این مرحله برای تخمین طیف تنها همین تعداد محدود فیلتر بانک در اختیار می‌باشد، به همین دلیل برای بازیابی طیفی با کیفیت بالا لازم است تا بین پارامترهای انرژی فیلتر بانک‌ها درون‌یابی صورت گیرد.

درون‌یابی با گرفتن تبدیل DCT معکوس رزولوشن بالا (رابطه (۲۰)) انجام می‌شود. این درون‌یابی Mel-scale هایی با رزولوشن بسیار دقیق نتیجه می‌دهد که از روی آنها باندهای فرکانسی مجزا با فاصله‌گذاری خطی (و نه مل) تخمین زده می‌شوند

$$\log \hat{Y}_{k'} = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} c_n \cos\left(\frac{(2k'+1)n\pi}{2iN}\right) \qquad (۲۰)$$
$$, \quad 0 < k' < iN-1$$

در (۲۰)، i فاکتور درون‌یابی و N تعداد فیلتر بانک‌ها است و بنابراین تعداد لگاریتم‌های انرژی iN برابر می‌شود. فاکتور درون‌یابی به وسیله رزولوشن مقیاس مل مطلوب تعیین خواهد شد. با استفاده از (۲۱)، فرکانس از مقیاس خطی به مقیاس مل تبدیل می‌گردد

$$f_{mel} = 2595 \log_{10}\left(1 + \frac{f_{HZ}}{700}\right) \qquad (۲۱)$$

سپس فاکتور رزولوشن از (۲۲) محاسبه می‌شود [۱۴]

$$i = \frac{f_{mel2} - f_{mel1}}{N+1} \qquad (۲۲)$$

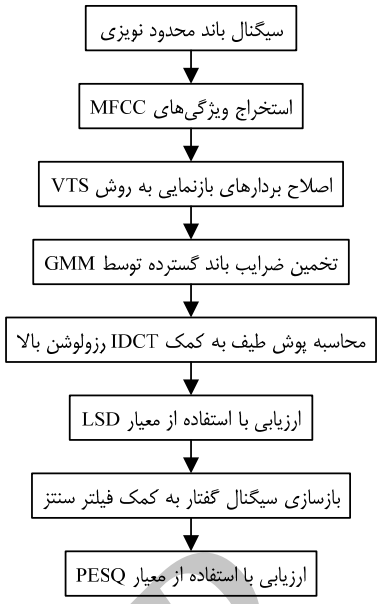
مرحله بعدی شامل تبدیل دامنه طیف به ضرایب AR است. در آنالیز LP[1] LP[1] معمولی ضرایب فیلتر لوله صوتی با حداقل‌نمودن میانگین مربعات خطای pامین مرتبه از مدل پیش‌بین خطی حاصل می‌شوند. روش خودهمبستگی برای یافتن ضرایب فیلتر نیاز به $(p+1)$ ضریب اول خودهمبستگی دارد که از روی فریم پنجره‌گذاری شده محاسبه می‌شود. با به توان ۲ رساندن طیف و سپس محاسبه تبدیل فوریه معکوس آن، مجموعه ضرایب خودهمبستگی به دست می‌آیند [۱۵].

پس از تخمین ضرایب فیلتر لوله صوتی، سیگنال تحریک باند گسترده نیز با عبور گفتار باند محدود موجود از فیلتر آنالیز به دست می‌آید. روش تاکردن طیف[2] یکی از تکنیک‌های رایج گسترش سیگنال تحریک است که به دلیل حجم محاسبات کم و نتایج مناسب کاربرد زیادی دارد. این روش تصویری آیینه‌ای از طیف باند محدود اصلی درون باند بالا تولید می‌کند. پیاده‌سازی تکنیک تاکردن طیفی در حوزه زمان، با درج صفر بین هر دو نمونه متوالی انجام خواهد گرفت [۷].

پس از آن با در اختیار داشتن ضرایب تخمینی فیلتر لوله صوتی باند گسترده، فیلتر سنتز طراحی شده و با عبور سیگنال تحریک گسترده شده از این فیلتر و با استفاده از روش جمع همپوشان[3] سیگنال گفتار باند گسترده تخمینی بازسازی می‌شود. پس از بازسازی، سیگنال باند گسترده از یک فیلتر بالاگذر عبور داده می‌شود تا سیگنال گفتار باند بالای از دست رفته به دست آید.

از آنجا که سیگنال باند باریک در دسترس است، می‌توان از خود آن در خروجی سیستم بهره برد. در نتیجه به کمک درون‌یابی نرخ نمونه‌برداری سیگنال باند باریک افزایش داده می‌شود تا بتوان آن را با سیگنال باند بالا جمع کرد. حاصل جمع سیگنال گفتار باند باریک اصلی و سیگنال باند بالای فیلترشده تخمینی، سیگنال بازسازی شده باند گسترده خواهد بود.

1. Linear Predictive
2. Spectral Folding
3. Overlap - Add



شکل ۳: الگوریتم پیاده‌سازی تکنیک تلفیقی VTS و BWE.

۵- پیاده‌سازی الگوریتم پیشنهادی

نوآوری اصلی معرفی‌شده در تحقیق حاضر، تلفیق دو ایده سری‌های تیلور برداری و گسترش پهنای باند جهت جبران‌سازی اثر نویز جمع‌شونده و محدودیت پهنای باند است. به این ترتیب که بردارهای ویژگی گفتار نویزی باند باریک در محدوده فرکانسی صفر تا ۴ کیلوهرتز، ابتدا به روش سری‌های تیلور برداری حذف نویز می‌شوند و سپس پوش طیف باند بالای آنها به روش مدل ترکیب گوسی بازسازی می‌شود. سرانجام با استفاده از دو معیار اندازه‌گیری LSD[1] (رابطه (۲۳)) و PESQ[2] نتایج حاصل از پیاده‌سازی این الگوریتم با حالت جبران‌سازی نشده مقایسه می‌گردد

$$d_{LSD}=\sqrt{\frac{1}{2\pi}\int_{0}^{\pi}(10\log_{10}(Y(w))-10\log_{10}(\hat{Y}(w)))^2\,dw} \qquad (۲۳)$$

در (۲۳)، $Y(w)$ و $\hat{Y}(w)$ به ترتیب پوش طیف باند گسترده اصلی و تخمین زده شده هستند. شکل ۳ بلوک دیاگرام الگوریتم معرفی‌شده در این تحقیق را نشان می‌دهد. معیار LSD پوش طیف سیگنال باند گسترده تخمینی و اصلی را مقایسه کرده و میزان شباهت آنها را با مقداری عددی مشخص می‌سازد. بدیهی است هرچه این عدد به صفر نزدیک‌تر باشد تخمین بهتری از پوش طیف سیگنال باند گسترده به دست آمده است.

در معیار PESQ، از چند شنونده مختلف خواسته می‌شود به سیگنال گفتار گوش داده و به کیفیت آن از نظر مقدار نویز و قابلیت فهم، امتیازی بین صفر تا ۴/۵ تخصیص دهند به گونه‌ای که هرچه کیفیت سیگنال بهتر باشد، امتیاز بالاتری به آن اختصاص یابد. سپس با میانگین‌گیری از امتیازات به دست آمده، عدد PESQ به عنوان معیار ارزیابی شنیداری کیفیت گفتار مشخص خواهد شد [۱۶]. در این مقاله با انتخاب ۱۰ جمله از مجموعه دادگان گفتاری TIMIT که هر یک توسط ۱۰ گوینده مرد و زن ادا شده‌اند، معیار PESQ برای سنجش میزان بهبود ناشی از اعمال الگوریتم VTS در بازسازی گفتار باند گسترده اندازه‌گیری شده است.

برای ارزیابی عملکرد ایده معرفی‌شده برای نزدیک‌تر نمودن مؤلفه‌های بازنمایی گفتار باند محدود تخریب‌شده به گفتار تمیز باند گسترده، آزمایشات متنوعی طراحی و پیاده‌سازی شده‌اند. مجموعه دادگان گفتاری انگلیسی TIMIT که در فرکانس ۱۶ کیلوهرتز نمونه‌برداری شده به عنوان دادگان گفتاری مرجع باند کامل انتخاب شده‌اند. این دادگان خود به دو مجموعه آموزش و آزمون تقسیم شده است. دادگان آموزشی شامل ۲۰۶۴ جمله از ۲۵۸ گوینده زن و مرد مختلف است، در حالی که مجموعه آزمون متشکل از ۷۶۰ جمله بیان‌شده توسط گویندگان مرد و زن متفاوت می‌باشد. در این حالت سهم دادگان آموزش و آزمون به ترتیب ۷۳ و ۲۷ درصد خواهد بود.

تمامی عبارات اداشده در مجموعه دادگان گفتاری فریم‌بندی شده و از آنها بردارهای بازنمایی لگاریتم فیلتر بانک مل (LFBE)[3] استخراج می‌شود. طول فریم‌ها ۲۰ میلی‌ثانیه با همپوشانی ۵۰٪ در نظر گرفته شده است.

جهت ارزیابی ایده مطرح‌شده در تحقیق، به مجموعه دادگان تست نویز سفید گوسی در SNRهای مختلف افزوده شده و ضرایب بازنمایی LFBE از آنها استخراج می‌گردد. سپس با حذف ۴ مؤلفه مربوط به باند بالای هر بردار LFBE پارامترهای گفتار نویزی باند محدود در حوزه لگاریتم طیف توان به دست می‌آیند. در ادامه با اعمال تبدیل کسینوسی گسسته، پارامترهای MFCC متناظر با این ضرایب استخراج می‌شوند.

تعداد فیلترهای مورد استفاده در فیلتر بانک مقیاس مل با تعداد ضرایب بازنمایی مورد نیاز برابر است. تعداد ضرایب مورد استفاده برای هر بردار بازنمایی LFBE برابر با ۱۲ ضریب باند باریک و ۴ ضریب باند بالا انتخاب می‌گردد و در نتیجه هر فریم از گفتار باند گسترده با یک بردار بازنمایی ۱۶مؤلفه‌ای مشخص خواهد شد.

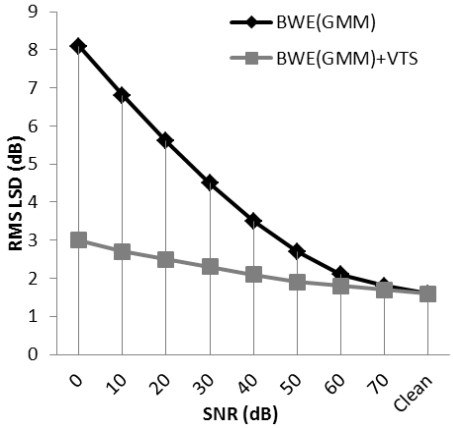
پس از استخراج بردارهای بازنمایی در حوزه لگاریتم فیلتر بانک، با اعمال تبدیل کسینوسی گسسته بر روی آنها، پارامترهای MFCC متناظر نیز به دست می‌آیند. از آنجا که در این تحقیق از مشتقات اول و دوم ضرایب کپستروم استفاده نمی‌شود، بردار بازنمایی MFCC گفتار باند گسترده نیز شامل ۱۶ مؤلفه خواهد بود. با جداسازی ۱۲ مؤلفه اول هر یک از این بردارها، پارامترهای ویژگی فریم‌های گفتاری باند باریک مورد استفاده برای اصلاح و گسترش پهنای باند استخراج می‌شوند.

۵-۱ جبران‌سازی نویز جمع‌شونده

جبران‌سازی نویز از روی پارامترهای بازنمایی LFBE مطابق الگوریتم گام به گام تشریح‌شده در بخش ۳ انجام گردید. مدل GMM نماینده تابع توزیع احتمال گفتار تمیز با مجموعه دادگان TIMIT تعلیم دیده و پارامترهای آن استخراج شدند. جهت اجتناب از پیچیدگی محاسبات، از همبستگی متقابل بین توزیع‌های گوسی صرف نظر شده و ماتریس کواریانس آنها قطری فرض شدند. انتخاب تعداد مؤلفه‌های گوسی مدل GMM گفتار مرجع نکته مهم دیگری است که بر روی دقت تخمین پارامترهای مجهول نویز تأثیر مستقیم داشته و در این تحقیق برابر ۱۲۸ انتخاب شد، چرا که افزایش بیشتر تعداد مؤلفه‌های گوسی منجر به بیش تعلیم مدل و افت کارایی الگوریتم پیشنهادی می‌گردد. در قدم بعدی مقادیر اولیه پارامترهای مزاحم محیطی به ترتیب زیر انتخاب شدند:

- بردار میانگین نویز: میانگین بردارهای بازنمایی ۱۰ فریم اول گفتار نویزی
- ماتریس کواریانس نویز: کواریانس اولین توزیع گوسی از مدل GMM گفتار تمیز (قطری)

1. Log Spectral Distance
2. Perceptual Evaluation of Speech Quality
3. Log of Filter Bank Energy



شکل ۴: بهبود کیفیت پوش طیف گفتار ناشی از اعمال الگوریتم VTS در بازسازی گفتار باند باریک به روش GMM با RMS LSD.

جدول ۱: مقایسه اندازه RMS LSD برای ارتقای کیفیت سیگنال‌های نویزی باند محدود با استفاده از الگوریتم VTS در حذف نویز و GMM در توسعه پهنای باند.

سیگنال به نویز (دسی‌بل)	BWE	BWE+VTS
تمیز	۱٫۶۱	۱٫۶۱
۷۰	۱٫۷۸	۱٫۶۸
۶۰	۲٫۱۰	۱٫۷۶
۵۰	۲٫۶۶	۱٫۹۳
۴۰	۳٫۴۸	۲٫۱۴
۳۰	۴٫۵۰	۲٫۳۱
۲۰	۵٫۶۴	۲٫۵۱
۱۰	۶٫۸۴	۲٫۶۹
۰	۸٫۱۰	۳٫۰۱

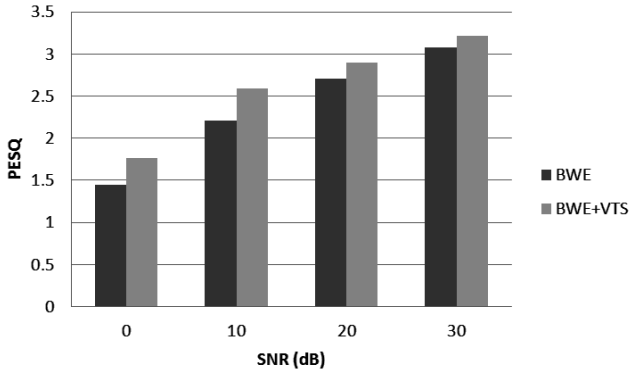
– بردار کانال باند محدود: با توجه به دانش اولیه به صورت برداری تمام صفر

در قدم بعد میانگین و کواریانس هر یک از توزیع‌های GMM گفتار تخریب‌شده مطابق (۸) و (۹) تخمین زده شدند و در ادامه با به کارگیری الگوریتم EM تکراری، تخمینی برای بردار میانگین و ماتریس کواریانس نویز مطابق با (۱۲) و (۱۳) به دست آمد. در نهایت نیز بردارهای بازنمایی اصلاح‌شده به کمک تکنیک MMSE تخمین زده شدند. سپس با اعمال تبدیل کسینوسی گسسته، ضرایب کپستروم متناظر نیز محاسبه گردیدند. حاصل این عمل بردارهای بازنمایی MFCC جبران‌سازی شده باند باریک هستند که جهت گسترش پهنای باند به مدل مبتنی بر GMM داده خواهند شد.

۵-۲ گسترش پهنای باند گفتار نویزی و اصلاح‌شده به روش GMM

جهت پیاده‌سازی تخمین پوش طیف سیگنال باند گسترده با استفاده از روش GMM، ابتدا باید ماتریس دادگان آموزش مدل GMM ساخته شود. برای ساخت این ماتریس از پارامترهای بازنمایی MFCC گفتار تمیز بهره گرفته شده است. بدین صورت که بردار بازنمایی باند محدود (x) و باند گسترده (y) در یک آرایه کنار هم قرار گرفته و ماتریس z متشکل از ضرایب MFCC برای تعلیم مدل تولید می‌شود.

مدل GMM معرفی‌شده در بخش قبلی با استفاده از ماتریس آموزش z و الگوریتم EM تعلیم می‌بینند تا پارامترهای مدل شامل بردار میانگین، ماتریس کواریانس و وزن‌های اولیه هر یک از توزیع‌های گوسی به دست



شکل ۵: بهبود کیفیت شنیداری گفتار ناشی از اعمال الگوریتم VTS در بازسازی سیگنال باند باریک به روش GMM با معیار PESQ برای SNR‌های پایین.

جدول ۲: مقایسه معیار PESQ برای ارتقای کیفیت سیگنال‌های نویزی باند محدود با استفاده از الگوریتم VTS در حذف نویز و GMM در توسعه پهنای باند.

سیگنال به نویز (دسی‌بل)	BWE	BWE+VTS
۳۰	۳٫۰۸	۳٫۲۲
۲۰	۲٫۷۱	۲٫۹۰
۱۰	۲٫۲۱	۲٫۵۹
۰	۱٫۴۵	۱٫۷۶

آیند. تعداد توزیع‌های گوسی مدل به مانند تکنیک VTS برابر ۱۲۸ انتخاب می‌شوند ولی بر خلاف آن به دلیل همبستگی قابل توجه موجود بین توزیع‌های گوسی، در این حالت ماتریس‌های کواریانس به صورت کامل (غیر قطری) فرض می‌شوند. همچنین جهت تست مدل پیشنهادی از بردارهای بازنمایی باند محدود نویزی و جبران‌سازی شده به روش VTS استفاده می‌شود.

پس از طی مراحل فوق، با به کارگیری تکنیک MMSE بردار بازنمایی باند گسترده MFCC تخمین زده می‌شود. جهت سنجش عملکرد الگوریتم پیشنهادی از دو معیار اندازه‌گیری LSD و PESQ استفاده شده است. معیار LSD اختلاف پوش طیف باند گسترده تخمین زده شده با پوش طیف باند گسترده اصلی را به صورت RMS و بر حسب دسی‌بل محاسبه می‌کند. از آنجا که برای استفاده از این رابطه نیاز به اطلاعات پوش طیف باند گسترده است، قبل از آن از بردارهای بازنمایی MFCC با استفاده از IDCT رزولوشن بالا، پوش طیف سیگنال بازسازی می‌شود.

جدول ۱ نتایج مربوط به اندازه‌گیری‌های RMS LSD در مقادیر SNR مختلف از صفر تا ۷۰ دسی‌بل برای بردارهای بازنمایی باند گسترده اصلاح‌نشده و اصلاح‌شده به روش VTS را مورد مقایسه قرار داده است. توجه به این نکته ضروری است که برای مجموعه دادگان گفتاری TIMIT، نسبت سیگنال به نویز ۸۰ دسی‌بل به عنوان گفتار تمیز مرجع در نظر گرفته شده است.

شکل ۴ نتایج حاصل از جدول ۱ را به صورت نموداری مورد مقایسه قرار داده است.

جدول ۲ و نمودار شکل ۵ به ترتیب به نمایش و مقایسه نتایج ارزیابی کیفیت گفتار باند گسترده نویزی و اصلاح‌شده به روش VTS به ازای SNR‌های پایین با استفاده از معیار PESQ اختصاص دارد.

نتایج حاصل از هر دو معیار عینی نشان‌دهنده این نکته هستند که هر چند گسترش پهنای باند سیگنال گفتار بر مبنای مدل ترکیب گوسی برای گفتار باند محدود عاری از نویز کارایی مناسبی دارد، اما با اضافه‌شدن نویز به سیگنال گفتار بازدهی آن به شدت کاهش می‌یابد. از همین رو تکنیک

سری‌های تیلور برداری جهت حذف نویز گفتار قبل از اعمال به سیستم گسترش پهنای باند مورد استفاده قرار گرفت. همان گونه که از نمودار شکل‌های ۴ و ۵ مشاهده می‌شود، الگوریتم پیشنهادی تأثیر بسیار مطلوبی در نزدیک‌تر ساختن پوش طیف گفتار باند گسترده بازسازی شده از سیگنال نویزی باند محدود به گفتار مرجع باند گسترده، و همچنین افزایش کیفیت سیگنال گفتار بازسازی شده دارد. از آنجا که تخمین پوش طیف مرحله اصلی از پروسه تخمین بردارهای بازنمایی باند گسترده به شمار می‌رود، استفاده از این الگوریتم به بهبود چشمگیر کیفیت سیگنال گفتار بازسازی شده منجر می‌گردد.

از طرفی الگوریتم معرفی‌شده در این تحقیق به ویژه در SNR های پایین بهبود به مراتب بهتری نسبت به SNR های متوسط و بالا نتیجه می‌دهد که این امر ناشی از توانایی تکنیک VTS در حذف نویز بهتر سیگنال گفتار در SNRهای پایین است.

٦- نتیجه‌گیری

گسترش پهنای باند سیگنال گفتار باند محدود باعث افزایش کیفیت و ادراک‌پذیری آن می‌گردد اما اگر این سیگنال باند محدود آغشته به نویز محیطی باشد، استفاده از روش‌های گسترش پهنای باند به تنهایی مؤثر نبوده و پوش طیف سیگنال بازسازی شده تفاوت زیادی با پوش طیف سیگنال باند گسترده مرجع خواهد داشت. بهره‌گیری از تکنیک VTS برای اصلاح بردارهای بازنمایی نویزی و سپس اعمال ضرایب بازنمایی اصلاح‌شده به مدل گسترش پهنای باند مبتنی بر GMM، منجر به بهبود چشمگیری در نتایج پیاده‌سازی می‌گردد. این بهبود به ویژه برای SNR های پایین که بازسازی آنها با مشکلات بیشتری روبه‌رو است، مطلوب‌تر می‌باشد که از جمله مزایای مهم الگوریتم پیشنهادی در تحقیق حاضر به شمار می‌رود. بررسی تلفیق الگوریتم سری‌های تیلور برداری برای اصلاح بردارهای بازنمایی و دیگر روش‌های گسترش پهنای باند در تحقیقات بعدی نویسندگان گزارش خواهد گردید.

مراجع

[1] M. Vali, S. A. Seyyed Salehi, and K. Karimi, "Robust speech recognition by modifying clean and telephone feature vectors using bidirectional neural network," *in Proc. Interspeech*, Pittsburgh, US, 17-21 Sep. 2006.

[2] R. M. Stern, B. Raj, and P. J. Moreno, "Compensation for environmental degradation in automatic speech recognition," in *Proc. of the Tutorial and Research Workshop*, pp. 33-42, 1997.

[3] P. J. Moreno, *Speech Recognition in Noisy Environment*, Ph.D. Thesis, pp. 79-96 and 121-126, 1996.

[4] P. J. Moreno, B. Raj, and R. M. Stern, "A vector taylor series approach for environment-independent speech recognition," in *Proc. ICASSP*, vol. 2, pp. 733-736, Atlanta, US, 7-10 May 1996.

[5] N. S. Kim, D. Y. Kim, B. G. Kong, and S. R. Kim, "Application of VTS to environment compensation with noise statistics," in *Proc. Interspeech*, 2001.

[6] D. Y. Kim, C. K. Un, and N. S. Kim, "Speech recognition in noisy environments using first-order vector taylor series," *Speech Communication*, vol.24, no.1, pp. 39-49, Apr. 1998.

[7] B. Iser and G. Schmidt, Bandwidth Extension of Telephony Speech, in Adaptive Signal Processing: Next Generation Solutions, eds. T Adali and S. Haykin, New York, Wiley, 2010.

[8] J. Peter and V. Peter, "On artificial bandwidth extension of telephone speech," *Signal Processing*, vol. 83, no. 8, pp. 1707-1719, 2003.

[9] P. Jax and P. Vary, "Feature selection for improved bandwidth extension of speech signal," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 697-700, Montreal, Canada, 2004.

[10] A. H. Nour-Eldin and P. Kabal, "Objective analysis of the effect of memory inclusion on bandwidth extension of narrowband speech," in *Proc. Interspeech*, pp. 2489-2492, Antwerp, Belgium, 2007.

[11] A. H. Nour-Eldin and P. Kabal, "Mel-frequency cepstral coefficient-based bandwidth extension of narrowband speech," in *Proc. Interspeech*, pp. 53-56, Brisbane, Australia, 22-26 Sep. 2008.

[12] H. Pulakka, U. Remes, K. Palomaki, M. Kurimo, and P. Alku, "Speech bandwidth extension using gaussian mixture model-based estimation of the highband mel spectrum," in *Proc. ICASSP*, pp. 5100-5103, 2011.

[13] A. Shahina and B. Yegnanarayana, "Mapping neural networks for bandwidth extension of narrowband speech," in *Proc. Interspeech*, pp 1435-1438, 2006.

[14] B. Milner and X. Shao, "Speech reconstruction from mel-frequency cepstral coefficients using a source-filter model," *InterSpeech*, pp. 2421-2424, Denver, US, 2002.

[15] L. Laaksonen, H. Pulakka, V. Myllyla, and P. Alku, "Development, evaluation, and implementation of an artificial bandwidth extension method of telephone speech in mobile terminal," *IEEE Trans. Consumer Electronics*, vol. 55, no. 2, pp. 780-787, May 2009.

[۱۶] ب. زمانی دهکردی، ا. اکبری و ب. ناصر شریف، "طرح دو فیلتر جدید برای بهبود کیفیت گفتار مبتنی بر توزیع احتمال پسین برای ضرایب موجک،" *نشریه علمی پژوهشی انجمن کامپیوتر ایران*، جلد ۶، شماره ۳- ب، صص. ۱۳-۱، پاییز ۱۳۸۷.

سارا پورمحمدی مدرک کارشناسی برق- الکترونیک خود را در سال ۱۳۸۶ و مدرک کارشناسی ارشد مهندسی پزشکی خود را در سال ۱۳۹۰ از دانشگاه شاهد اخذ نمود. پردازش سیگنال‌های دیجیتال (گفتار، تصویر و سیگنال‌های حیاتی)، شبکه‌های عصبی، الگوریتم‌های تکاملی و بازشناسی الگو از جمله زمینه‌های تحقیقاتی وی هستند.

منصور ولی در سال ۱۳۷۶ مدرک کارشناسی مهندسی برق- الکترونیک خود را از دانشگاه صنعتی اصفهان و در سال ۱۳۷۸ مدرک کارشناسی ارشد مهندسی برق- بیوالکتریک خود را از دانشگاه صنعتی شریف تهران دریافت نمود. وی در سال ۱۳۸۵ موفق به اخذ درجه دکتری مهندسی پزشکی- بیوالکتریک از دانشگاه صنعتی امیرکبیر گردید. دکتر ولی از سال ۱۳۸۵ به مدت یکسال در دانشکده فنی مهندسی دانشگاه اصفهان و از ۱۳۸۶ تا ۱۳۹۱ در دانشکده فنی مهندسی دانشگاه شاهد تهران مشغول به خدمت گردید. ایشان در حال حاضر عضو هیأت علمی دانشکده برق و کامپیوتر دانشگاه خواجه نصیر تهران می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: پردازش و بازشناخت گفتار، پردازش صوت در پزشکی و شبکه‌های عصبی مصنوعی.

محسن قدیانی مدرک کارشناسی برق- الکترونیک خود را در سال ۱۳۸۴ از دانشگاه صنعتی خواجه نصیرالدین طوسی و مدرک کارشناسی ارشد مهندسی پزشکی خود را در سال ۱۳۸۹ از دانشگاه شاهد دریافت نمود. وی هم‌اکنون دانشجوی دکتری مهندسی برق دانشگاه سمنان است. از جمله زمینه‌های تحقیقاتی مورد علاقه وی می‌توان شبکه‌های مخابراتی بی‌سیم، بهینه‌سازی و پردازش گفتار و تصویر را نام برد.