

یک روش دو مرحله‌ای جدید برای بهسازی گفتار مبتنی بر بهسازی ادراکی، بسته موجک و تخمین نویز ناایستاد و بهینه‌سازی ابرپارامترهای آن با الگوریتم دسته ذرات کوانتومی (QPSO)

الهه صاحبی همراه^۱، دانشجوی کارشناسی ارشد؛ سید جهان‌شاه کبودیان^۲، استادیار

۱- دانشکده مهندسی کامپیوتر و فناوری اطلاعات - دانشگاه رازی - کرمانشاه - ایران - e1372.sahebi@yahoo.com

۲- دانشکده مهندسی کامپیوتر و فناوری اطلاعات - دانشگاه رازی - کرمانشاه - ایران - kabudian@razi.ac.ir

چکیده: موضوع بهبود کیفیت صدا امروزه به یکی از موضوعات مهم و اساسی روز تبدیل شده است. تضعیف نویز به‌نحوی که اختلالی در سیگنال اصلی به‌وجود نیابد، یک چالش مهم برای بهبود صدا محسوب می‌شود. موضوع تحقیق این مقاله بررسی نویز موجود در سیگنال گفتار، حذف و یا کاهش آن از سیگنال گفتار نویزی است. در این مقاله یک روش دو مرحله‌ای جدید برای کاهش نویز موجود در سیگنال گفتار نویزی ارائه شده است. در این روش، در مرحله اول، سیگنال گفتار نویزی توسط روش ادراکی بی‌زین بهبود داده می‌شود و سپس بر روی سیگنال بهبود داده‌شده بسته موجک اعمال می‌شود و سیگنال را به زیر باندهای مجزا از هم تبدیل می‌کند. در مرحله دوم بر روی هر زیر باند روش بهسازی NNESE اعمال می‌شود. همچنین ابرپارامترهای روش NNESE در تمامی زیرباندها توسط الگوریتم فراابتکاری دسته ذرات کوانتومی (QPSO) بهینه می‌شود. بدین ترتیب هر زیر باند به‌طور جداگانه بهبود داده می‌شود. در نهایت سیگنال بهسازی‌شده با ترکیب زیر باندها حاصل می‌شود. روش پیشنهادی نسبت به روش‌های دیگران از لحاظ معیارهای PESQ و Segmental SNR بر روی بخشی از پایگاه داده TIMIT آغشته‌شده به نویزهای NoiseX-92 بهتر عمل می‌کند و منجر به کاهش بیشتر نویز از سیگنال می‌گردد.

واژه‌های کلیدی: پردازش گفتار، حذف نویز، بهبود کیفیت گفتار، بسته موجک، الگوریتم بهینه‌سازی دسته ذرات کوانتومی، QPSO.

A new two-stage method for speech enhancement based on perceptual enhancement, wavelet packet and nonstationary noise detection and QPSO-based optimization of its hyperparameters

E. Sahebi Hamrah¹, MSc student; S J. Kabudian², Assistant Professor

1- Faculty of Computer Engineering and Information Technology, Razi University, Kermanshah, Iran, e1372.sahebi@yahoo.com

2- Faculty of Computer Engineering and Information Technology, Razi University, Kermanshah, Iran, kabudian@razi.ac.ir

Abstract: Nowadays, speech enhancement has become one of the most important issues in signal processing. Noise reduction such that it does not disturb the original signal is an important challenge in speech enhancement. In this paper, we have proposed a new hybrid two-stage method for speech enhancement. In the proposed method, noisy speech signal is enhanced using perceptually-motivated Bayesian approach in the first stage. Then, the signal is decomposed into sub-bands using wavelet packet decomposition. In the second stage, each sub-band signal is enhanced using NNESE method. Hyperparameters of NNESE is optimized using QPSO metaheuristic. Finally, all enhanced sub-band signals are combined together using wavelet packet reconstruction. Proposed method is better than others in terms of PESQ and Segmental SNR criteria over a subset of TIMIT speech database polluted by NoiseX-92 noise database.

Keywords: Speech processing, noise reduction, speech enhancement, wavelet transform, QPSO optimization algorithm.

تاریخ ارسال مقاله: ۱۳۹۷/۰۱/۲۴

تاریخ اصلاح مقاله: ۱۳۹۷/۰۶/۱۰ و ۱۳۹۷/۰۹/۰۴

تاریخ پذیرش مقاله: ۱۳۹۷/۰۹/۲۶

نام نویسنده مسئول: سید جهان‌شاه کبودیان

نشانی نویسنده مسئول: ایران - کرمانشاه - طاقستان - دانشگاه رازی - دانشکده فنی و مهندسی رازی.

۱- مقدمه

گفتار یکی از ابزارهای ارتباطی بین انسانها است که در طول روز در دفعات متعددی از آن استفاده می‌شود. برقراری ارتباط گفتاری در محیط‌های آرام معمولاً بدون هیچ‌گونه مشکلی انجام می‌گیرد. زمانی که برقراری ارتباط در فاصله‌های دور و همچنین محیطی ناآرام همراه با عوامل محیطی مزاحم باشد، گفتار موردنظر آغشته به نویز می‌شود. نویز کیفیت سیگنال گفتار را تا حد زیادی کاهش می‌دهد در نتیجه توانایی شنونده برای درک و فهم سیگنال گفتار تضعیف می‌گردد. وجود نویز یکی از مباحث مهم در حوزه پردازش سیگنال‌های صوتی است که همواره مطرح بوده و راه‌حل‌های متعددی جهت مقابله با آن پیشنهاد شده است. وجود نویز باعث پایین آمدن کارایی سیستم‌های پردازش سیگنال شده و همچنین باعث ایجاد اختلال در گفتار می‌گردد [۱]. در سال‌های اخیر پیشرفت قابل توجهی در زمینه حذف نویز از سیگنال‌های صوتی صورت گرفته است. اهمیت حذف نویز با گسترش و توسعه تلفن‌های بی‌سیم یعنی موبایل نیز افزایش یافته است. فرض کنید سیگنال گفتار موردنظر حاوی اطلاعات مهمی باشد اما آغشته به نویز شود. همین امر باعث ایجاد مشکل در سیستم‌های بازشناسی گفتار و رمزگشایی سیگنال می‌شود. یا تصور کنید از یک سیستم شناسایی گوینده استفاده می‌شود اگر سیگنال نویزی باشد دقت شناسایی گوینده کاهش خواهد یافت. مطالعه و تحقیق در زمینه بهبود گفتار به ۴۰ سال پیش بازمی‌گردد [۲]. نحوه کار اکثر روش‌هایی که برای بهسازی گفتار استفاده می‌شوند به این صورت است که ابتدا با استفاده از تخمین‌گرها طیف نویز را تخمین می‌زنند و سپس با استفاده از فیلترها و سایر روش‌ها نویز را کاهش داده و سیگنال بهبود یافته را ایجاد می‌کنند [۳].

تفریق طیفی یکی از روش‌های کلاسیک بهسازی گفتار است. تفریق طیفی به‌عنوان روش پایه و متداول برای حذف نویزهای محیطی از سیگنال گفتار مطرح می‌شود. این روش دارای مزایای مهمی از جمله پیچیدگی کم، پیاده‌سازی راحت و ساده هست [۴، ۵]. روش بسته موجک روشی است که برای تحلیل زمانی-فرکانسی سیگنال و به‌منظور کاهش نویز از سیگنال نویزی به کار می‌رود. برای اینکه کاهش نویز بهتری انجام گیرد انتخاب تابع مناسب موجک، تعداد سطح تجزیه و تعیین نوع محاسبه مقدار آستانه و نحوه تخمین آن از جمله مسائل مهم در این روش است [۶، ۷]. روش فیلتر وینر یک روش جایگزین برای تفریق طیفی برای افزایش کیفیت سیگنال گفتار است [۸]. فیلتر وینر فیلتری است که با معیار $MMSE^1$ برای بازیابی سیگنال گفتار از گفتار نویزی استفاده می‌شود [۹]. روش فیلتر وینر تطبیقی برگرفته‌شده از روش فیلتر وینر است [۱۰]. روش بهسازی گفتار بر اساس آستانه گذاری (Thresholding) نیازمند برآورد دقیق مقدار آستانه (Threshold) و همچنین انتخاب تابع آستانه‌گذاری (Thresholding Function) مناسب است که روش‌های مختلفی بر این پایه نیز به وجود آمده است [۱۱]. از جمله روش‌های بهسازی گفتار که اخیراً مورد توجه قرار گرفته است روش‌های مبتنی بر روش EMD [۱۲] است. از جمله روش‌های مبتنی

بر این روش می‌توان به روش‌های EMD^۲ [۱۲] و EMDH^۳ [۱۳] اشاره کرد. استفاده از ویژگی‌های ادراکی نیز از کارهایی است که در راستای بهسازی سیگنال گفتار صورت می‌گیرد. استفاده از ویژگی ادراکی، اعوجاج در سیگنال گفتار را کاهش می‌دهد [۱۴، ۱۵]. روش‌های بسیاری در رابطه با استفاده از ساختار هارمونیک برای ایجاد بهبود گفتار نیز ارائه شده است [۱۶].

فیلترهای وفقی به منظور یادگیری در کاربردهای مختلفی چون تخمین کانال، تخمین موقعیت هدف، شناسایی سیستم و حذف نویز از سیگنال گفتار مورد استفاده قرار می‌گیرند. یکی از فیلترهای مشهور در این زمینه الگوریتم وفقی کمینه میانگین مربعات (LMS) می‌باشد که به علت عملکرد مناسب و پیچیدگی اندک محاسباتی به صورت گسترده در حذف نویز صوتی مورد استفاده قرار می‌گیرد [۱۷]. موضوع حذف نویز از سیگنال گفتار در سال‌های اخیر پیشرفت قابل توجهی داشته است و روش‌های بسیاری در این حوزه معرفی شده‌اند.

روش پیشنهادی در این مقاله یک روش تلفیقی مبتنی بر الگوریتم بهسازی گفتار NNESE^۵ [۱۸] و بسته موجک است که بر روی سیگنال‌های بهسازی شده توسط الگوریتم Perceptual Bayesian Estimator اعمال می‌شود. در روش پیشنهادی از الگوریتم فرابانتکاری QPSO^۶ [۱۹، ۲۰] برای بهینه‌سازی ابرپارامترهای روش ترکیبی استفاده می‌شود. هدف از ارائه این روش حذف نویزهای باقی‌مانده (Residual Noise) در سیگنال بهسازی شده است به نحوی که منجر به بهبود کیفیت سیگنال گفتار با معیارهای معمول گردد. در ادامه، ساختار مقاله به این صورت است: در بخش دوم روش NNESE که پایه‌ای برای روش پیشنهادی است توضیح داده می‌شود. در قسمت سوم روش ترکیبی پیشنهادی Wavelet-NNESE برای کاهش نویز باقیمانده معرفی می‌شود. در قسمت چهارم الگوریتم پیشنهادی نهایی و دومرحله‌ای در این مقاله معرفی می‌شود که از ترکیب روش بهسازی Bayesian Estimator و روش بهسازی پیشنهادی Wavelet-NNESE برای کاهش نویز به دست می‌آید. در قسمت پنجم و ششم الگوریتم بهینه‌سازی QPSO و نحوه استفاده از آن را که برای بهینه‌سازی ابرپارامترهای روش پیشنهادی استفاده شده است، توضیح داده خواهد شد. در قسمت هفتم روش پیشنهادی را مورد ارزیابی قرار داده و نتایج آزمایش‌ها شرح داده خواهد شد و در نهایت در قسمت آخر به نتیجه‌گیری مطرح می‌شود.

۲- الگوریتم NNESE

بیشتر روش‌هایی که برای حذف نویز و بهسازی سیگنال گفتار مورد استفاده قرار می‌گیرد شامل ۴ مرحله اصلی به شرح زیر می‌باشند:

۱. تجزیه سیگنال با استفاده از روش‌های موجود مانند STFT^۷، Wavelet، EMD و غیره

۲. تشخیص و آشکارسازی نویز موجود در سیگنال

۳. تضعیف مؤلفه‌های نویزی تشخیص داده‌شده در سیگنال

۴. بازسازی سیگنال گفتار بهسازی شده

$$\varepsilon(\rho) = \frac{1}{2}\rho + \frac{1}{\rho} \log\left(1 + \sqrt{1 - \exp(-\rho^2)}\right) \quad (2)$$

که با استناد به روش تخمین گر d-Dimensional Trimmed Estimator (DATE) مقدار ρ برابر با ۴ در نظر گرفته شده است. پس خواهیم داشت [۱۸]:

$$\varepsilon(\rho) = 3.4742$$

پس از اینکه سیگنال گفتار نویزی قاب بندی شد، مجموعه $\{y_q(k)\}$ که شامل دنباله‌ای از نمونه‌های هر قاب سیگنال گفتار نویزی می‌باشد را برحسب مقادیر دامنه (قدرمطلق سیگنال) به صورت صعودی مرتب کرده و Y_k نامیده می‌شود. برای هر قاب خواهیم داشت:

$$Y_1 \leq Y_2 \leq Y_3 \leq \dots \leq Y_N \quad (3)$$

مرحله ۲: تخمین انحراف معیار نویز

به منظور تخمین انحراف معیار باید سه گام اصلی اعمال شود که خواهیم داشت [۱۸]:

• محاسبه k_{min} :

برای محاسبه k_{min} فرض می‌شود که دنباله نمونه‌های مرتب شده تا نمونه k_{min} ام تشکیل دهنده مؤلفه‌های نویزی خواهد بود.

$$Y_1 \leq Y_2 \leq Y_3 \leq \dots \leq Y_{k_{min}} \quad (4)$$

طبق نامساوی Bienayme-Chebyshev-Markov برای محاسبه k_{min} خواهیم داشت [۱۸]:

$$k_{min} = \frac{N}{2} - hN \quad (5)$$

$$h = \frac{1}{\sqrt{4N(1-Q)}} \quad (6)$$

قضیه نامساوی یا نابرابری Bienayme-Chebyshev-Markov تضمین می‌کند که در هر نمونه تصادفی تقریباً تمامی مقادیر در نزدیکی میانگین خواهند بود. اگر بخواهیم به طور دقیق تر بیان نماییم می‌توان گفت که این نابرابری بیان می‌کند که: حداکثر مقادیری که در هر توزیع می‌تواند بیش از K برابر انحراف معیار با میانگین فاصله داشته باشد، $\frac{1}{K^2}$ است. مقدار Q درجه اطمینان (Confidence Degree) به طور پیش فرض برابر با ۰/۹۵ در نظر گرفته شده است [۱۸].

در روش NNESE [۱۸] بهسازی سیگنال بدون انجام مرحله اول صورت می‌گیرد؛ یعنی هیچ مرحله تجزیه‌ای بر روی سیگنال صورت نمی‌گیرد. مؤلفه‌های نویزی به طور مستقیم تخمین زده می‌شوند. سپس نویز تخمین زده شده از نمونه‌های سیگنال نویزی کم شده و منجر به کاهش مؤلفه‌های نویزی از سیگنال نویزی می‌شوند. این روش کیفیت سیگنال را ارتقاء می‌دهد. پس این روش شامل ۳ مرحله اصلی است:

۱. تشخیص نویز موجود در سیگنال نویزی با استفاده از تخمین انحراف معیار

۲. تضعیف نویز در قسمت‌های نویزی

۳. بازیابی سیگنال گفتار بهسازی شده

۲-۱- تشخیص نویز با استفاده از تخمین انحراف معیار

برای تشخیص نویز در سیگنال گفتار نویزی، مؤلفه‌های نویزی با استفاده از تخمین انحراف معیار به طور مستقیم تشخیص داده می‌شوند. به منظور تخمین انحراف معیار، ابتدا سیگنال نویزی به قاب‌های جدا از هم تقسیم بندی می‌شود. برای تقسیم بندی سیگنال نویزی به قاب‌های جدا از هم خواهیم داشت [۱۸]:

$$y_q(k) = y(k + qN) \quad (1)$$

y سیگنال گفتار آغشته شده به نویز و q به عنوان اندیس هر قاب و N تعداد نمونه‌های هر قاب است. در این روش مقدار N برابر ۵۱۲ در نظر گرفته شده است [۱۸].

پس از قاب بندی سیگنال با استفاده از تخمین گر d-Dimensional Trimmed Estimator (DATE) [۲۱] می‌توان انحراف معیار را به ازای هر قاب محاسبه نمود. تخمین گر DATE تنها قادر به تخمین انحراف معیار سیگنال گفتاری است که به نویز سفید گوسی آغشته شده است. در روش NNESE این محدودیت از بین رفته است و انحراف معیار هر قاب برای سیگنالی که آغشته به هر نوع نویز غیرایستاد باشد، قابل محاسبه است؛ بنابراین، این روش برای حذف نویزهای غیرایستاد گوسی/غیرگوسی و به خصوص نویزهای باقیمانده می‌تواند مناسب باشد. تخمین انحراف معیار در روش NNESE دو مرحله اصلی دارد [۱۸]:
مرحله ۱: آماده سازی دنباله‌ای از نمونه‌های سیگنال نویزی که شامل گام‌های زیر است [۱۸]:

• تعیین مقدار آستانه با استفاده از فرمول ۲ [۱۸]:

کم می‌شود. همین امر باعث حذف شدن مؤلفه‌های نویزی از سیگنال نویزی می‌شود [۱۸]:

$$\tilde{y}_q(k) = \begin{cases} \text{sign}(y_q(k)) \cdot (|y_q(k)| - \alpha\sigma_q), & \text{if } |y_q(k)| \geq \alpha\sigma_q \\ \text{sign}(y_q(k)) \cdot (\beta|y_q(k)|) & \text{otherwise} \end{cases} \quad (10)$$

رابطه فوق، یک تابع آستانه‌گذاری نرم (Soft Thresholding) است که یک مقدار کف (Floor) نیز در نظر می‌گیرد.

مقدار α به صورت پیش فرض ۰/۳۵ و مقدار β به صورت پیش فرض ۰/۶۵ قرار داده شده است.

در نهایت می‌توان سیگنال بهسازی شده به ازای همه قاب‌ها را طبق فرمول ۱۱ به صورت زیر بازسازی کرد:

$$\tilde{y}(k) = \sum_{q=0}^{M-1} \tilde{y}_q(k - qN) \quad (11)$$

بنابراین الگوریتم بهسازی گفتار NNESE دارای ۵ ابرپارامتر $\{\alpha, \beta, \rho, w, Q\}$ است. در قسمت بعد خواهیم گفت که ابتدا سیگنال گفتار توسط بسته موجک به زیر باندهای مختلف تجزیه می‌گردد و با توجه به خصوصیات طیفی متفاوت این زیر باندها در سیگنال گفتار و در نویزهای غیرایستاد، مقادیر این ۵ ابرپارامتر برای زیر باندهای مختلف، متفاوت خواهد بود. مجموعه تمامی این ابرپارامترها در زیر باندهای مختلف در کنار هم قرار گرفته و به عنوان یک ابرپارامتر بزرگ و کلی (Total Hyperparameter) در نظر گرفته می‌شود. این ابرپارامتر بزرگ در روش ترکیبی نهایی Bayesian-Wavelet-NNESE توسط الگوریتم بهینه‌سازی QPSO بهینه خواهند شد.

۳- پودمان^۸ ترکیبی پیشنهادی Wavelet-NNESE برای کاهش نویز باقیمانده و نایستاد

پودمان پیشنهادی این مقاله که روش Wavelet-NNESE نام دارد به این صورت است که ابتدا بر روی سیگنال گفتار نویزی بسته موجک اعمال می‌شود. از تبدیل موجک برای تجزیه سیگنال به زیر باندهای مجزا از هم استفاده می‌شود تا زیر باندهای مختلف موجک به صورت مستقل از هم پردازش شوند. سپس الگوریتم NNESE بر روی هر یک از زیر باندها اعمال می‌شود تا هر زیر باند از سیگنال به طور مستقل بهبود پیدا کند. پس از آن سیگنال بهبودیافته نهایی از تلفیق زیر باندهای بهبود داده شده به دست می‌آید. در روش NNESE پنج ابرپارامتر (Hyperparameter) وجود دارد که به این صورت هستند: ۱. اندازه پنجره

• بررسی شرط اینکه آیا یک مقدار k ($k \in \{k_{min} \dots, N\}$) وجود دارد که در شرط زیر صدق کند [۱۸]:

$$\|Y_{k-1}\| \leq \frac{\varepsilon(\rho)}{\lambda} \frac{\sum_{i=1}^k \|Y_i\|}{k} \leq \|Y_{k+1}\| \quad (7)$$

|| موجود در فرمول ۷ نشان‌دهنده نرم اقلیدسی است که در اینجا برای سیگنال حقیقی (Real) همان قدرمطلق می‌باشد و مقدار λ طبق فرمول زیر و توسط تابع گاما قابل محاسبه است [۱۸]:

$$\lambda = \sqrt{2} \Gamma(1) / \Gamma(0.5) = 0.7979 \quad (8)$$

در صورت برقراری شرط بالا (فرمول ۷) به صورت زیر خواهد بود:

$$b_q = k$$

و در صورت برقرار نبودن شرط مذکور b_q به صورت زیر خواهد بود:

$$b_q = k_{min}$$

• محاسبه انحراف معیار

پس از قاب‌بندی سیگنال و تعیین مقدار b_q ، می‌توان مقدار انحراف معیار نویز به ازای هر قاب را طبق فرمول زیر محاسبه نمود [۱۸]:

$$\sigma_q = \frac{\varepsilon(\rho)}{\lambda} \frac{\sum_{i=1}^{b_q} \|Y_i\|}{b_q} \quad (9)$$

۲-۲- انتخاب مؤلفه‌های نویزی

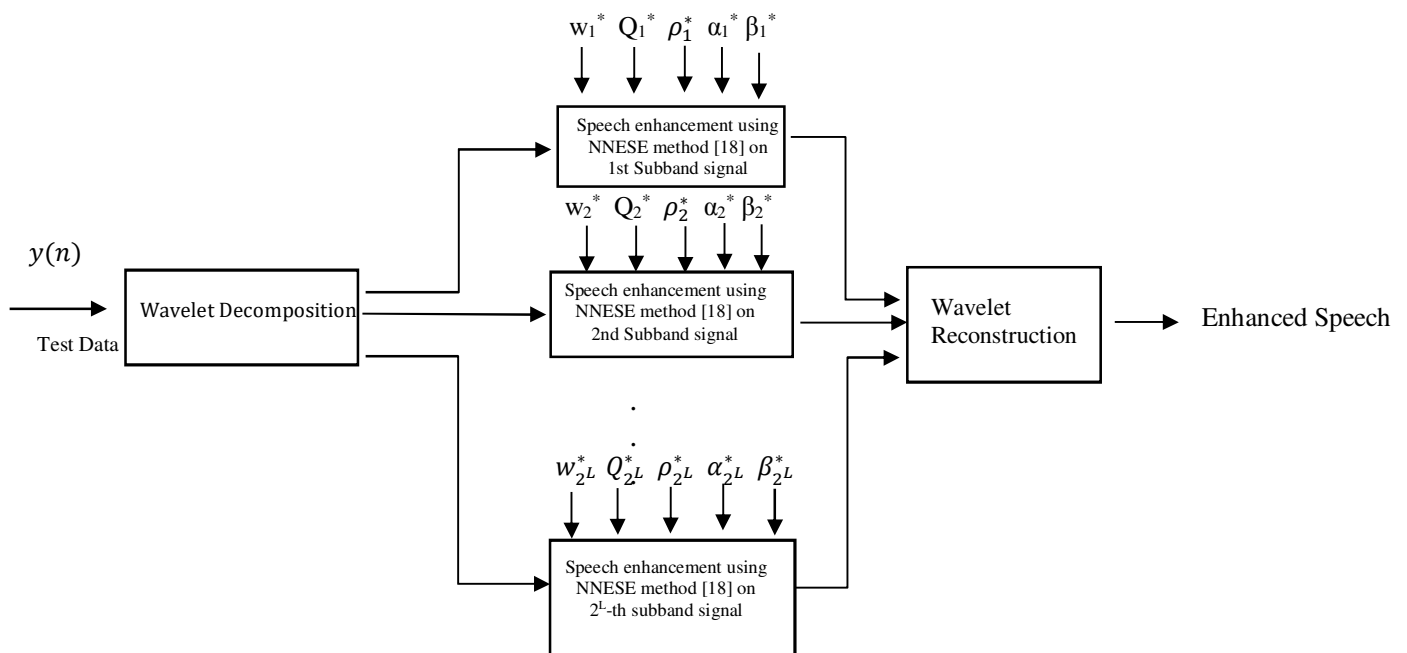
مقدار انحراف معیار به دست آمده از فرمول ۹ برای تشخیص بخش نویزی سیگنال استفاده می‌شود. این مقدار در یک ضریب ضرب می‌گردد و به عنوان سطح آستانه انتخاب می‌شود، طوری که مقدار پایین تر از این مقدار، نویز تشخیص داده می‌شود و مقدار بالاتر از آن به عنوان سیگنال گفتار در نظر گرفته می‌شود. نحوه اعمال مقدار آستانه و تشخیص مؤلفه‌های نویزی تحت یک تابع آستانه گذاری نرم در فرمول ۱۰ نشان داده شده است. با بررسی سیگنال‌های آغشته شده به انواع نویزهای غیرایستاد مشاهده شده است که انحراف معیار تخمین زده شده به درستی مؤلفه‌های نویزی را به طور مستقیم تشخیص می‌دهد [۱۸].

۳-۲- بازبایی سیگنال گفتار

برای بازبایی سیگنال بهسازی شده طبق فرمول ۱۰ ابتدا به ازای هر قاب، ضریبی از مقدار انحراف معیار، محاسبه شده و از نمونه‌های سیگنال اصلی

به‌ازای دادگان توسعه (که متفاوت از دادگان تست است) به بیشترین حد ممکن خود رسیده و در نتیجه بیشترین کاهش نویز را از سیگنال داشته باشد. در مرحله تست پس از تجزیه سیگنال‌های تست به زیر باندهای مجزا توسط تبدیل موجک، برای هر زیر باند، الگوریتم NNESE با مقادیر بهینه ابر پارامترهای به‌دست‌آمده از الگوریتم QPSO، بر روی سیگنال‌های تست اعمال می‌شود. پس‌از آن با استفاده از معکوس بسته موجک، سیگنال بهسازی شده را به‌دست می‌آید. بلوک دیاگرام روش Wavelet-NNESE در شکل ۱ مشاهده می‌شود. در مرحله تست تعداد تقسیم‌بندی سیگنال توسط بسته موجک به تعداد سطح تجزیه بستگی دارد که در این روش L یعنی تعداد سطح تجزیه برابر ۳ در نظر گرفته شده است که به میزان 2^L یعنی ۸ زیر باند ایجاد خواهد شد. از میان موجک‌های موجود از موجک خانواده Daubechies یعنی db10 استفاده شده است. در روش پیشنهادی، به‌ازای هر زیر باند، پنج ابرپارامتر وجود دارد. مجموعه ابرپارامترهای روش NNESE متناظر با زیرباندهای مختلف موجک را در کنار هم قرار داده تا یک بردار ۴۰ بعدی ابرپارامترها به نام بردار ابرپارامتر کلی به‌وجود آید. این بردار ۴۰ بعدی توسط الگوریتم بهینه‌سازی QPSO به‌گونه‌ای بهینه می‌گردد که بهترین میانگین PESQ و بهترین میانگین SegSNR به‌ازای دادگان توسعه به‌دست آید. پس از بهینه‌کردن این بردار ابر پارامتر کلی، مقدار آن ثابت (fixed) می‌شود و از آن‌پس برای بهسازی دادگان تست استفاده می‌گردد.

(w) که به‌طور پیش‌فرض برابر با ۳۲ میلی‌ثانیه در نظر گرفته شده است. ۲. مقدار درجه اطمینان (Q) که به‌طور پیش‌فرض برابر با ۰/۹۵ در نظر گرفته شده است. ۳. مقدار آلفا (α) که به‌طور پیش‌فرض برابر با ۰/۳۵ در نظر گرفته شده است. ۴. مقدار بتا (β) که به‌طور پیش‌فرض برابر ۰/۶۵ در نظر گرفته شده است. ۵. مقدار ρ با استناد به روش DATE [۲۱] که به‌طور پیش‌فرض برابر با ۴ در نظر گرفته شده است. مزیت روش Wavelet-NNESE آن است که می‌توان در آن مقادیر ابرپارامترها را برای هر زیرباند متفاوت در نظر گرفت؛ زیرا خصوصیات آماری و طیفی سیگنال‌های هر باند متفاوت است. در قسمت‌های بعدی گفته خواهد شد که برای بهینه‌سازی مجموعه تمامی ابرپارامترهای روش NNESE در تمامی زیر باندها به‌صورت توأم از الگوریتم QPSO [۲۰، ۱۹] استفاده شده است. با توجه به خصوصیات طیفی متفاوت زیر باندهای موجک در سیگنال گفتار و در نویزهای غیرایستاد و غیر سفید، مقادیر بهینه این ابرپارامتر برای زیر باندهای مختلف، متفاوت خواهد بود. مجموعه تمامی این ابرپارامترها در کنار هم قرار داده و آن بردار ابرپارامتر بزرگ و کلی (Total Hyperparameters Vector) نامیده می‌شود. این بردار یا ابرپارامتر بزرگ در روش ترکیبی Bayesian-Wavelet-NNESE توسط الگوریتم بهینه‌سازی QPSO بهینه خواهند شد. به‌نحوی که منجر به بهینه‌ترین حالت بر روی دادگان توسعه (Development) شود و معیارهای ارزیابی PESQ و Segmental SNR

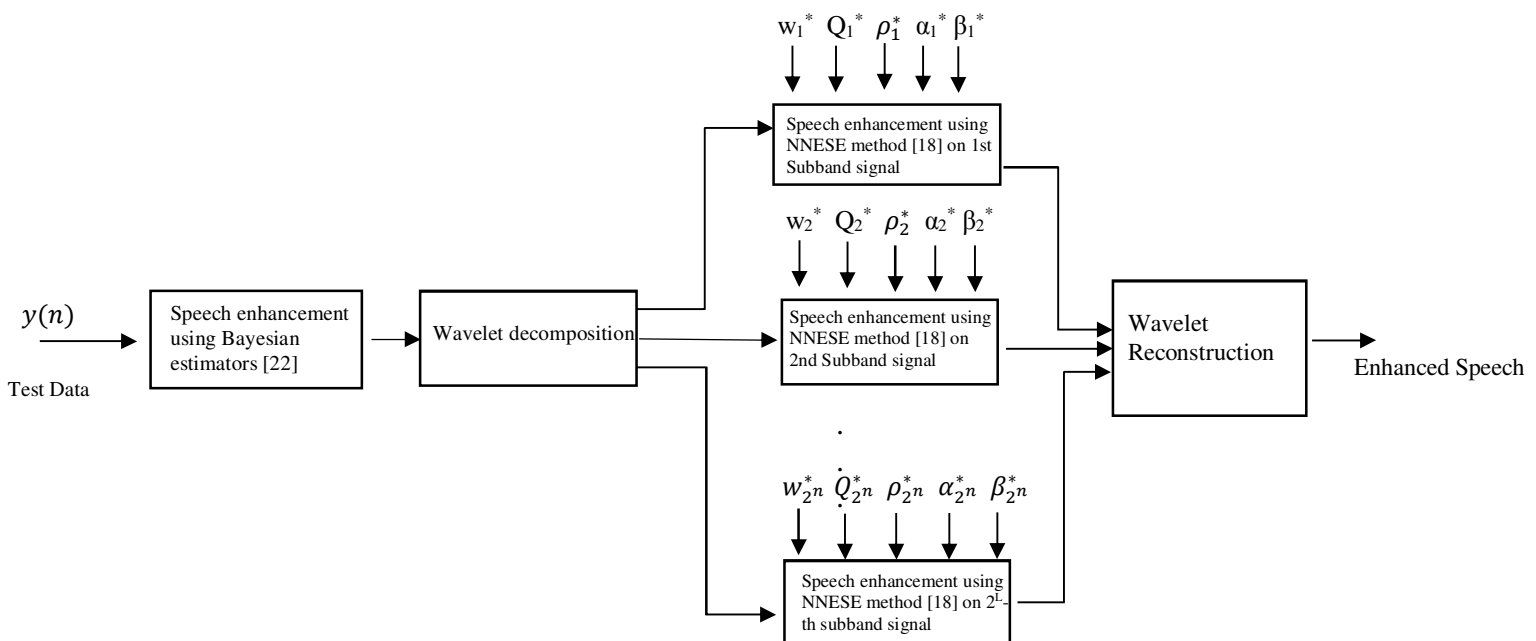


شکل ۱: بلوک دیاگرام روش پیشنهادی Wavelet-NNESE برای بهسازی سیگنال گفتار تست، ابر پارامترهای w^* ، Q^* ، ρ^* ، α^* ، β^* با توجه به الگوریتم QPSO بر روی داده‌های Development محاسبه می‌شوند.

در شکل ۲ قابل مشاهده است. برای درک بهتر و ساده‌تر سیستم ترکیبی پیشنهادی، الگوریتم این روش را می‌توان به صورت شبه کد در شکل ۳ مشاهده نمود. همان‌طور که در شکل ۳ مشاهده می‌کنید روش نهایی پیشنهادی این مقاله به این صورت است که سیگنال گفتار ابتدا با روش شناخته‌شده Bayesian Estimator بهبود داده می‌شود. سیگنال بهسازی شده توسط روش Bayesian سپس توسط بسته موجک به زیرباندهای مختلف تجزیه می‌شود. سیگنال گفتار در هر یک از این زیر باندها توسط روش NNESE مجدداً بهبود داده می‌شود تا نویزهای باقیمانده در آن زیرباند کاهش یابد. در مرحله آخر، سیگنال‌های بهبود داده‌شده در تمام زیر باندها باهم ترکیب شده و سیگنال بهسازی شده نهایی را می‌سازند. یکی از مسائل مهم در روش پیشنهادی تعیین مقادیر ابرپارامترها در هر زیرباند است. اولین و ساده‌ترین راه‌حل، آن است که مقادیر پارامترها در زیرباندهای مختلف یکسان فرض شود و برابر با مقادیر پیشنهادی در مقاله NNESE قرار داده شود. طبیعتاً این راه‌حل بهینه نخواهد بود زیرا خصوصیات آماری و طیفی سیگنال‌های زیر باندهای مختلف باهم متفاوت است. در بخش بعد الگوریتم بهینه‌سازی QPSO توضیح داده خواهد شد. از این الگوریتم برای بهینه‌سازی مقدار بردار ابرپارامترهای زیر باندها (\mathcal{P}) استفاده می‌شود تا بهترین مقادیر PESQ و SegSNR حاصل شود.

۴- سیستم نهایی پیشنهادی برای بهسازی گفتار: Bayesian-Wavelet-NNESE

روش بهسازی Perceptually-Motivated Bayesian Estimator یکی از روش‌های شناخته‌شده و قدرتمند در حوزه کاهش نویز و بهبود گفتار است. اطلاعات کامل‌تر و دقیق‌تر در رابطه با این روش در مقاله [۲۲] توضیح داده شده است. راه‌حل ترکیبی و پیشنهادی موردنظر ما بدین صورت است که، ابتدا روش Bayesian به‌عنوان یک روش پایه برای بهسازی گفتار مورد استفاده قرار می‌گیرد. سیگنال بهسازی شده حاصل از مرحله اول پس از اعمال بهسازی به دلایل مختلف بازمه دارای نویز باقیمانده و غیرایستنا است. در روش پیشنهادی این مقاله، پودمان پیشنهادی Wavelet-NNESE که در بخش ۳ به‌طور مفصل توضیح داده شد، به‌عنوان مرحله پس‌پردازش برای حذف نویزهای باقی‌مانده در مرحله اول به‌کار گرفته می‌شود. استفاده از این پودمان به خاطر آن است که روش NNESE روشی مناسب برای کاهش نویزهای غیرایستنا و غیرسفید است. اگر نویز جمع‌شده با سیگنال گفتار از نوع غیرسفید و غیرایستنا باشد، نویز باقیمانده نیز غیرسفید و غیرایستنا خواهد بود. این بدان دلیل است که روش‌های تخمین طیف نویز ایده‌آل نیستند. بلوک دیاگرام سیستم ترکیبی پیشنهادی



شکل ۲: بلوک دیاگرام روش نهایی پیشنهادی Bayesian-Wavelet-NNESE برای بهسازی سیگنال گفتار تست، ابر پارامترهای $w^*, Q^*, \rho^*, \alpha^*, \beta^*$ با توجه به الگوریتم QPSO بر روی داده‌های Development محاسبه می‌شوند.

% Bayesian-Wavelet_NNESE

% x is input Noisy Speech Signal

% z is output enhanced Speech Signal with Proposed Method

Input: x, L (Wavelet level), θ (total hyperparameter vector = $\{w_k, Q_k, \rho_k, \alpha_k, \beta_k\}, k=1,2,\dots,2^L$)

Output: z

$y = \text{Bayesian Speech Enhancement}(x)$; % Loizou's Perceptual Bayesian Estimator

SubBand Signals = Wavelet Packet Decomposition (y, L);

for $k = 1$ to 2^L

Enhanced SubBand Signals (k) = NNESE (SubBand signals (k), $w_k, Q_k, \rho_k, \alpha_k, \beta_k$);

end

$z = \text{Wavelet Packet Reconstruction}(\text{Enhanced SubBand Signals});$

شکل ۳: شبکه کد روش نهایی پیشنهادی Bayesian-Wavelet-NNESE

پتانسیلی با مرکزیت $p_{i,j}(t)$ تعریف می‌شود. در الگوریتم QPSO یک تابع چگالی احتمال $Q(\theta_{i,j}(t+1))$ محاسبه می‌شود که احتمال حضور یک ذره در موقعیت $(t+1)$ در زمان $t+1$ را نشان می‌دهد (با فرض یک مسئله d بعدی هر ذره با بردار موقعیت $\theta_i = (\theta_{i,1}, \theta_{i,2}, \dots, \theta_{i,d})$ نشان داده می‌شود).

$$Q(\theta_{i,j}(t+1)) = \frac{1}{2\beta |m_{best,j}(t) - \theta_{i,j}(t)|} e^{-\frac{|p_{i,j}(t) - \theta_{i,j}(t+1)|}{\beta |m_{best,j}(t) - \theta_{i,j}(t)|}} \quad (12)$$

β ضریب انبساط-انقباض است که مشخصه همگرایی الگوریتم را تعیین می‌کند. M_{best} میانگین بهترین موقعیت ذرات است که طبق فرمول ۱۳ محاسبه می‌شود و $p_{i,j}(t)$ مختصات جاذب محلی (Local Attractor) است.

$$m_{best}(t) = (m_{best,1}, m_{best,2}, \dots, m_{best,d}) \quad (13)$$

$$= \left(\frac{1}{M} \sum_{i=1}^M P_{i,1}(t), \frac{1}{M} \sum_{i=1}^M P_{i,2}(t), \dots, \frac{1}{M} \sum_{i=1}^M P_{i,d}(t) \right)$$

M در معادله بالا تعداد اعضای جمعیت و d بعد فضا را تعیین می‌کند.

$$p_{i,j}(t) = \frac{c_1 P_{i,j}(t) + c_2 P_{g,j}(t)}{c_1 + c_2} \quad (14)$$

P_i بهترین موقعیت برای ذره i ام (موقعیتی که بهترین مقدار تابع هدف را به دست می‌دهد) و بردار P_g بهترین موقعیت ذره در میان

۵- الگوریتم بهینه‌سازی QPSO

امروزه مفهوم بهینه‌سازی نقش اساسی را در علوم و مهندسی ایفا می‌کند. بهینه‌سازی کاربردهای زیادی از جمله در تخصیص منابع، زمان‌بندی‌ها، تصمیم‌گیری‌ها و غیره دارد. روش‌های مختلفی برای بهینه‌سازی وجود دارد. از جمله این الگوریتم‌ها می‌توان به بهینه‌سازی دسته ذرات با رفتار کوانتومی (QPSO) [۱۹] اشاره نمود. اخیراً مفاهیم مکانیک کوانتومی در مسائل بهینه‌سازی مورد استفاده قرار می‌گیرد. این الگوریتم می‌تواند در یافتن حل بهینه در فضای جستجو خیلی مؤثر باشد. هدف این الگوریتم یافتن مقدار بهینه بردار θ^* است به گونه‌ای که مقدار $f(\theta)$ یا تابع هدف (Objective function) بهینه شود. در این مقاله، بردار θ مجموعه ابرپارامترهای روش Wavelet-NNESE است و تابع هدف $f(\theta)$ معیارهای PESQ و SegSNR است که باید توسط روش QPSO بهینه شود. روش QPSO بر مبنای ترکیبی از روش PSO^1 با قوانین مکانیک کوانتومی کار می‌کند و با مدل‌سازی ذرات PSO در دنیای کوانتوم، دقت و قدرت جستجوی ذرات را نسبت به روش PSO افزایش می‌دهد [۲۰]. روش PSO یکی از قدیمی‌ترین الگوریتم‌ها در دسته تکنیک‌های مبتنی بر بهینه‌سازی آماری است. این روش اولین بار در سال ۱۹۹۵ بر پایه توجه به حرکت گروه‌هایی چون پرندگان، حشرات و غیره بنا نهاده شد. این الگوریتم سعی می‌کند که از فرآیند طبیعی کار و زندگی گروهی الهام بگیرد تا به یک مکان بهینه دست پیدا کند [۲۳]. در کنار قدرت بالا، سادگی روش QPSO نسبت به سایر روش‌های بهینه‌سازی را می‌توان از مزیت‌های دیگر این روش جدید عنوان کرد. در دنیای کوانتوم با توجه به اصل عدم قطعیت هایزنبرگ، امکان تعیین هم‌زمان سرعت و موقعیت ذره امکان‌پذیر نیست. آنچه در اینجا مطرح است احتمال حضور یک ذره در یک نقطه از فضای جواب است که با توجه به میدان

به‌عنوان مثال، اگر تعداد سطوح موجک ۳ باشد، ۸ زیرباند خواهیم داشت و بنابراین بردار کلی ابرپارامترها ۴۰ بعدی خواهد بود. این بردار ۴۰ بعدی توسط الگوریتم QPSO بهینه می‌شود تا بیشترین مقدار PESQ و SegSNR به دست آید. روش پیشنهادی Bayesian-Wavelet-NNESE دارای دو مرحله آموزش/توسعه (Train/Development) و تست (Test) می‌باشد. به‌همین دلیل دادگان گفتاری استفاده‌شده نیز به دو دسته دادگان آموزش/توسعه و دادگان تست تقسیم‌بندی می‌گردد. فرض می‌گردد که در دادگان آموزش/توسعه، به‌ازای هر فایل نویزی، فایل تمیز آن نیز وجود دارد تا بتوان معیارهای PESQ و SegSNR را محاسبه نمود، ولی در دادگان تست، فقط به فایل نویزی دسترسی وجود دارد. در مرحله آموزش سیستم، بردار کلی ابرپارامترها (θ) توسط الگوریتم QPSO به‌گونه‌ای بهینه می‌شود که روش Bayesian-Wavelet-NNESE بهترین کیفیت یا بیشترین مقدار PESQ و SegSNR را به‌ازای دادگان آموزشی به دست آورد. پس از اجرای الگوریتم QPSO به تعداد تکرار کافی و همگرایی آن و رسیدن به مقدار بهینه ابرپارامترها (θ^*)، مقدار θ^* ثابت نگاه داشته می‌شود و از آن پس برای بهسازی دادگان تست استفاده خواهد شد. بلوک دیاگرام مرحله آموزش و تست به‌ترتیب در شکل‌های ۴ و ۵ نشان داده شده‌اند. مرحله آموزش سیستم مطابق با شکل ۴ به‌صورت زیر است: ابتدا بردار ابرپارامتر برای یکی از اعضای جمعیت QPSO با مقدار θ_0 مقاردهی اولیه می‌شود و بقیه اعضای جمعیت مقادیر رندوم دارند. این مقدار اولیه θ_0 در بخش ۳ ذکر شده است. به‌ازای این مقدار، روش بهسازی پیشنهادی Bayesian-Wavelet-NNESE بر روی تمام فایل‌های گفتاری نویزی از دادگان آموزشی اعمال می‌شود و کیفیت آن‌ها بهبود می‌یابد. دادگان آموزشی شامل فایل‌های گفتاری متنوع از گویندگان مختلف (مرد و زن)، نویزهای مختلف و SNR های مختلف است.

جمعیت کل ذرات است. c_1 و c_2 بردارهای تصادفی هستند که در بازه [۰،۱] تولید می‌شوند. در نهایت موقعیت بعدی هر ذره طبق معادله ۱۵ محاسبه می‌شود [۲۰، ۱۹].

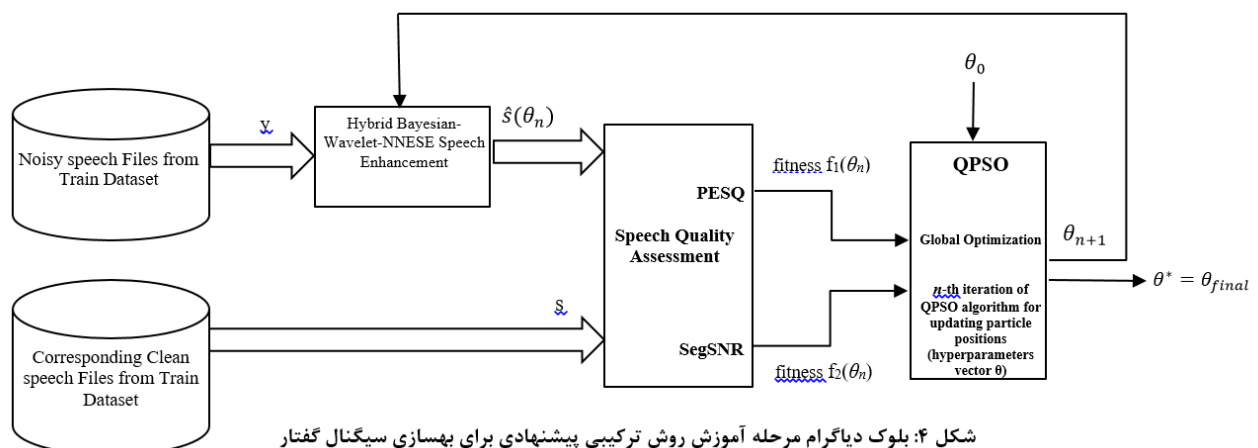
$$\theta_{i,j}(t+1) = p_{i,j}(t) \pm \beta \left| m_{best,i,j}(t) - \theta_{i,j}(t) \right| \ln \left(\frac{1}{u} \right), \quad u \sim U(0,1) \quad (15)$$

الگوریتم QPSO یک الگوریتم تکراری است و آن قدر تکرار خواهد شد تا به همگرایی برسد (با یک معیار از قبل تعریف‌شده).

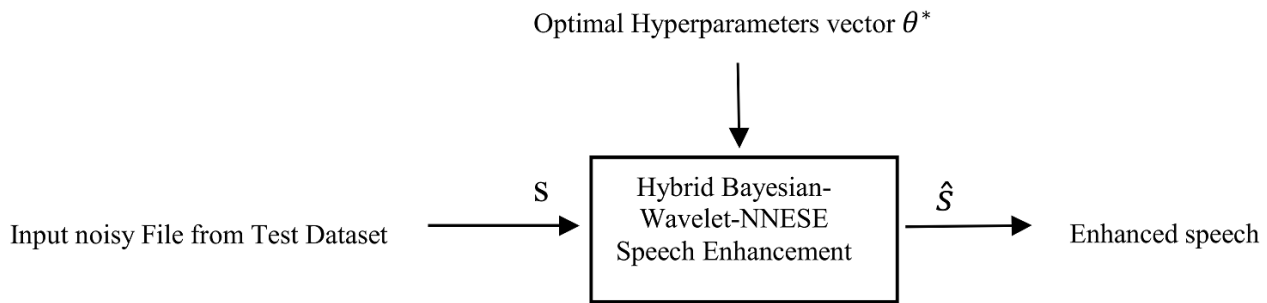
۶- بهینه‌سازی ابرپارامترهای روش پیشنهادی Bayesian-Wavelet-NNESE با الگوریتم QPSO

همان‌طور که در قسمت‌های قبل گفته شد، در روش هیبرید Bayesian-Wavelet-NNESE، پودمان بهسازی NNESE بر روی هر یک از زیرباندهای موجک اعمال می‌شود. پودمان NNESE برای هر یک از زیرباندهای موجک دارای ۵ ابرپارامتر $\{\alpha, \beta, \rho, w, Q\}$ است. واضح است که خصوصیات طیفی و آماری نویز و سیگنال گفتار در هر یک از این زیرباندها متفاوت است؛ بنابراین در این مقاله ابرپارامترهای هر یک از زیرباندها متفاوت فرض می‌شود. اگر تعداد سطوح موجک L باشد، آنگاه تعداد زیرباندها 2^L خواهد بود. ابرپارامترهای هر یک از زیرباندها در کنار هم قرار داده می‌شود تا یک بردار کلی از ابرپارامترها به نام بردار ابرپارامتر کلی (θ) (Total Hyperparameters Vector) با ابعاد 5×2^L تشکیل شود. بردار کلی ابرپارامترها (θ) به‌صورت زیر تشکیل می‌گردد:

$$\theta = [w_1 \ Q_1 \ \rho_1 \ \alpha_1 \ \beta_1 \ w_2 \ Q_2 \ \rho_2 \ \alpha_2 \ \beta_2 \ \dots \ w_{2^L} \ Q_{2^L} \ \rho_{2^L} \ \alpha_{2^L} \ \beta_{2^L}] \quad (16)$$



شکل ۴: بلوک دیاگرام مرحله آموزش روش ترکیبی پیشنهادی برای بهسازی سیگنال گفتار



شکل ۵: بلوک دیاگرام مرحله تست روش ترکیبی پیشنهادی برای بهسازی سیگنال گفتار

همچون [۱۸] NNESE، [۱۳] EMDH، [۱۳] Perceptually Motivated Spectral Subtraction، [۲۶] MSS-SMPR، [۲۲] Bayesian Estimator، [۲۷] Subtraction، [۲۸] MMSE-LSA، [۲۹] MMSE-SA، [۳۰] Geometric Approach to Spectral Subtraction Teager Energy، [۳۱] Modulation Spectral Subtraction Audio De-Noising، [۳۲] Operated Perceptual Wavelet Packet Using MATLAB Wavelets Toolbox [۳۳] مورد مقایسه قرار گرفته است. هریک از روش‌هایی که مورد مقایسه قرار گرفته است به این صورت بوده که ابتدا کد پیاده‌سازی شده هریک از این روش‌ها از مقاله مرجع آن گرفته شده است و سپس هر روش توسط نویسندگان مقاله با پایگاه داده معرفی شده اجرا و آزمایش شده است و سپس نتایج آن ثبت شده است.

۷-۲- معیارهای ارزیابی

برای ارزیابی و مقایسه روش‌ها از معیار غیر ادراکی نسبت سیگنال به نویز Segmental SNR و معیار ادراکی کیفیت گفتار PESQ [۳۴] استفاده می‌شود. معیار PESQ معیاری است که کیفیت گفتار را مورد بررسی قرار می‌دهد هرچه میزان PESQ بیشتر باشد نشان‌دهنده این است که گفتار از لحاظ کیفیت شنیداری بهتر شده است. معیار Segmental SNR جز دسته روش‌های ارزیابی کمی به شمار می‌رود. این معیار مقادیر نسبت سیگنال به نویز را در سیگنال بهسازی شده به‌ازای هر قاب محاسبه کرده و میانگین این مقادیر را به‌ازای قاب‌های مختلف به‌دست می‌آورد هرچه این عدد بزرگتر شود نشان‌دهنده این است که میزان نویز موجود در سیگنال نویزی کاهش پیدا کرده و در نتیجه کیفیت شنوایی بالاتر می‌رود. نحوه محاسبه این معیار در فرمول ۱۷ نشان داده شده است [۳].

$$SNR_{seg} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \frac{\sum_{n=Nm}^{Nm+N-1} x^2[n]}{\sum_{n=Nm}^{Nm+N-1} (x[n] - \hat{x}[n])^2} \quad (17)$$

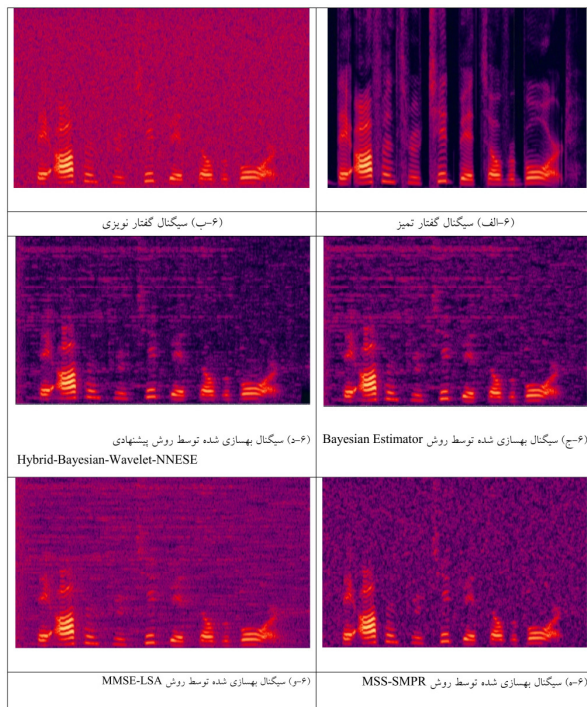
پس از بهبود فایل‌ها با روش پیشنهادی و با مقدار ابرپارامتر فعلی، کیفیت متوسط تمامی فایل‌های آموزشی بهسازی شده با معیار PESQ (تابع هدف f_1) و معیار SegSNR (تابع هدف f_2) به‌دست می‌آید. با توجه به توابع هدف f_1 و f_2 که برای هر یک از اعضای جمعیت QPSO به‌دست آمده است، مقادیر Mbest، Pbest و Gbest در الگوریتم QPSO به‌دست می‌آید و بردار ابرپارامتر برای هر یک از اعضای جمعیت با فرمول (۱۵) تغییر پیدا می‌کند و مقادیر جدیدی برای θ به‌دست می‌آید. این کار در یک حلقه (Loop) آن قدر تکرار می‌گردد که الگوریتم همگرا شود و یا به تعداد تکرار حداکثر برسد. تعداد اعضای جمعیت QPSO برابر ۲۰ و حداکثر تعداد تکرار برابر ۲۰۰ در نظر گرفته شده است. قابل توجه آن است که در الگوریتم QPSO استفاده شده، بهینه‌سازی با دو هدف انجام شده است (Bi-Objective Optimization). بدین معنی که از نظر الگوریتم QPSO ابرپارامتری بهتر است که به‌طور هم‌زمان هم مقدار PESQ و هم مقدار SegSNR را بهبود دهد. پس از پایان الگوریتم QPSO، آخرین مقدار Gbest (Global Best) به‌عنوان مقدار بهینه θ^* برای بردار ابرپارامترها در نظر گرفته می‌شود و از آن پس مقدار آن ثابت فرض شده و برای بهسازی فایل‌های نویزی در مرحله تست طبق شکل ۵ استفاده می‌شود.

۷-۲- آزمایش‌ها و نتایج

۷-۱- پایگاه داده

برای ارزیابی روش پیشنهادی از دو مجموعه آموزش/توسعه و تست استفاده شده است. مجموعه دادگان آموزشی/توسعه شامل ۱۰ سیگنال گفتار (شامل ۵ سیگنال زن و ۵ سیگنال مرد) از گویندگان پایگاه داده TIMIT [۲۴] و مجموعه دادگان تست شامل ۱۰ نمونه سیگنال متفاوت دیگر (شامل ۵ سیگنال زن و ۵ سیگنال مرد) از پایگاه TIMIT است. گویندگان دادگان آموزش و تست کاملاً متفاوت هستند. سیگنال‌های مذکور را با ۷ نوع نویز، f16, white, factory2, pink, volvo, babble, buccaneer2 [۲۵] NoiseX-92 از پایگاه داده [۲۵] NoiseX-92، آورده کرده‌ایم. نسبت سیگنال به نویز (SNR)، برابر با مقادیر ۱۰، ۰، ۱۰+، ۲۰+ دسی‌بل بوده است. روش پیشنهادی با روش‌های دیگر

برای هر الگوریتم محاسبه شده است. RTF مدت زمان اجرای هر الگوریتم به‌ازای یک ثانیه از گفتار می‌باشد که در این جدول قرار داده شده است. آزمایشات توسط یک سیستم ۶۴ بیتی با امکانات Core i7-5500U 2.4GHz و در اختیار داشتن حافظه 8GB و همچنین نرم افزار متلب نسخه 2012a انجام شده است.



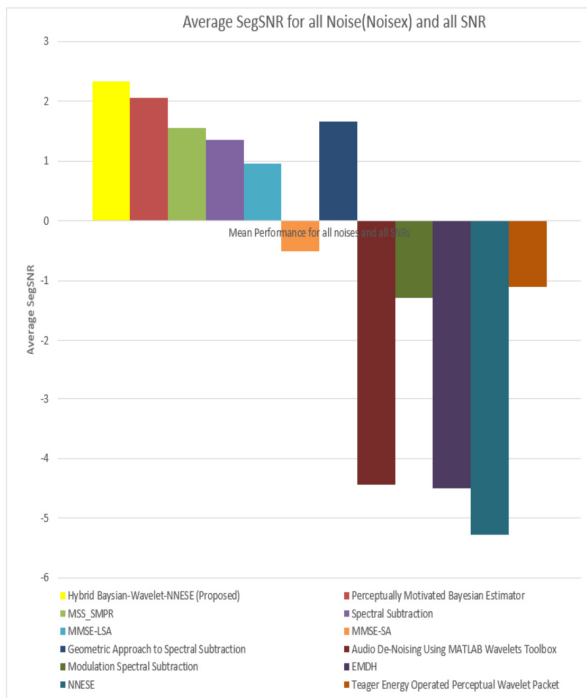
شکل ۶: ارزیابی و مقایسه روش‌ها با یکدیگر، (الف-۶) سیگنال گفتار تمیز، (ب-۶) سیگنال گفتار نویز، (ج-۶) سیگنال بهسازی شده توسط روش Bayesian Estimator، (د-۶) سیگنال بهسازی شده توسط روش Bayesian-Wavelet-NNESE، (ه-۶) سیگنال بهسازی شده توسط روش MMSE-LSA، (و-۶) سیگنال بهسازی شده توسط روش MSS-SMPR

همان‌طور که در جدول ۱ مشاهده می‌شود روش پیشنهادی ما با معیارهای موردنظر و در با مقایسه با سایر روش‌ها عملکرد بهتری داشته است و منجر به حذف نویز بهتر و افزایش کیفیت بیشتر در سیگنال گفتار نویزی شده است. برای درک بهتر جدول ۱ مقادیر به‌صورت نمودار رسم شده است. نمودار ۱ در شکل ۷ میانگین معیار PESQ را به‌ازای تمام SNR و تمام نویزها نشان می‌دهد. همان‌طور که از نمودار ۱ مشخص است روش پیشنهادی ما نسبت به سایر روش‌های مورد ارزیابی بهتر عمل کرده است و منجر به بهسازی بهتر شده است. نمودار ۲ در شکل ۸ میانگین معیار Segmental SNR را به‌ازای تمام SNR و تمام نویزها نشان می‌دهد.

که در فرمول ۱۷ $x[n]$ سیگنال تمیز، $\hat{x}[n]$ سیگنال بهسازی‌شده، N طول هر قاب، M تعداد کل قاب‌های سیگنال می‌باشد. در محاسبه SegSNR قسمت‌های سکوت سیگنال استفاده نشده است و فقط از قاب‌هایی که در ناحیه گفتاری و غیرسکوت قرار دارند استفاده شده است (اندازه‌گیری SegSNR با جعبه ابزار VoiceBox انجام شده که برای تشخیص سکوت از الگوریتم Sohn استفاده می‌کند). این حالت برای مقایسه روش‌ها منصفانه‌تر است. در محاسبه معیارهای فوق هم سیگنال نویزی و هم سیگنال تمیز را باید در اختیار داشت. هرچقدر مقدار معیارهای فوق‌الذکر بیشتر باشد، نشان می‌دهد که سیگنال بهسازی‌شده به سیگنال اصلی (تمیز) نزدیک‌تر است.

۷-۳- ارزیابی و مقایسه روش‌ها

در اولین آزمایش، بهبود حاصل‌شده در طیف‌گفتار توسط روش پیشنهادی به‌صورت بصری نشان داده شده است. در روش پیشنهادی ترکیبی در مرحله اول روش بیزین مقداری از نویز جمعی سیگنال را حذف می‌نماید. نویز باقی‌مانده در سیگنال با احتیاط بیشتری توسط روش NNESE که پارامترهای آن در هر زیرباند از بسته موجک بهینه می‌شود تضعیف می‌گردد. در واقع همان‌طور که در شکل ۶ نشان داده شده است روش NNESE نویز باقی‌مانده در سیگنال را بیشتر تضعیف می‌کند در حالی که طبق شکل به ساختار هارمونیک گفتار آسیبی وارد نمی‌کند. در شکل ۶ یک سیگنال گفتار تمیز (الف-۶) و سیگنال گفتار نویزی متناظر با آن (ب-۶) (با نسبت سیگنال به نویز 10DB نویز white) نشان داده شده است که توسط روش Bayesian Estimator (ج-۶) و روش پیشنهادی Bayesian-Wavelet-NNESE (د-۶) بهسازی شده است. همان‌طور که دیده می‌شود روش پیشنهادی توانسته است که تا حدی نویزهای باقیمانده در روش Bayesian را حذف نماید و علاوه بر این باعث وضوح بیشتر طیف سیگنال شده و برخی از هارمونیک‌های سیگنال گفتار را واضح‌تر کند (Clarify) و بازایی نماید (Restore). شکل‌های (الف-۶) و (ب-۶) نیز به‌ترتیب طیف سیگنال بهسازی‌شده با روش‌های MSS-SMPR و MMSE-LSA را نشان می‌دهد. همان‌طور که ملاحظه می‌گردد طیف روش پیشنهادی نسبت به روش‌های دیگر بهتر است. همان‌طور که در توضیح روش پیشنهادی گفته شد، ابتدا با استفاده از الگوریتم بهینه‌سازی QPSO، ابرپارامترهای روش Bayesian-Wavelet-NNESE بر روی مجموعه دادگان آموزش (توسعه) تنظیم گردید و سپس کارایی سیستم با دو معیار PESQ و SegSNR بر روی مجموعه دادگان تست محاسبه گردید. جدول ۱ میانگین معیارهای PESQ و SegSNR را به‌ازای تمام فایل‌های صوتی تست و تمام انواع نویزها از پایگاه داده Noisex و تمام SNR نشان می‌دهد. در جدول ۱ علاوه بر محاسبه معیارهای مطرح‌شده مقدار RTF (Real Time Factor) نیز



شکل ۸: نمودار میانگین SegmentalSNR به ازای همه نویزها و همه دسی بل‌ها

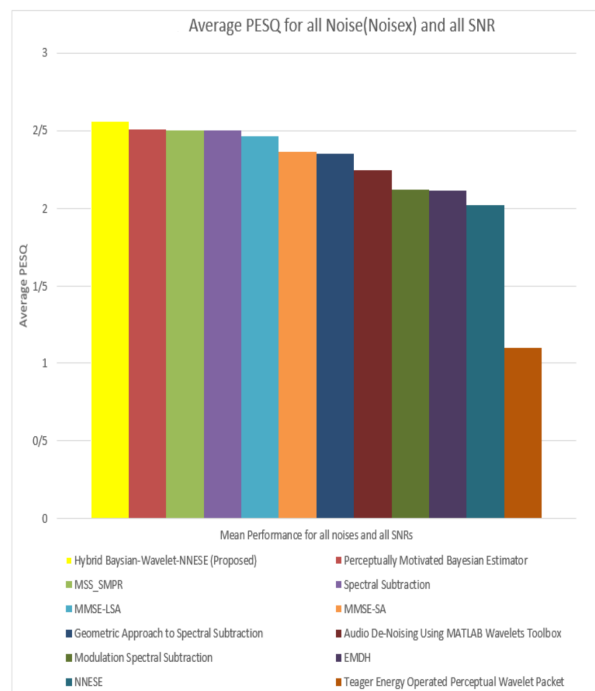
همان‌طور که از نمودار ۲ مشخص است روش پیشنهادی ما نسبت به سایر روش‌های مورد ارزیابی بهتر عمل کرده است و منجر به بهسازی بهتر شده است.

۸- نتیجه‌گیری

روشی که در این مقاله مبنای کار ما قرار گرفت، روش NNESE بود. این روش یکی از روش‌های جدید در حوزه حذف نویز است که قادر به تخمین و حذف نویزهای غیرثابت و محیطی متنوع می‌باشد. با بهره‌گیری از این روش و ترکیب آن با روش تبدیل موجک، روش Wavelet-NNESE ارائه شده است. در روش Wavelet-NNESE الگوریتم بهینه‌سازی QPSO استفاده شده است و ابرپارامترهای موردنیاز محاسبه شده است. بدین ترتیب روشی جدید ارائه شده است که نسبت به روش‌های جدید کارایی بهتری دارد و قادر به حذف انواع نویزها می‌باشد. روش Wavelet-NNESE را به‌عنوان مرحله پس‌پردازش پس از بهسازی سیگنال گفتار توسط روش Bayesian مورد استفاده قرار گرفته است و سیستم ترکیبی پیشنهادی دومرحله‌ای ارائه شده است و نتایج مورد بررسی قرار گرفته است. سیستم ترکیبی پیشنهادی نسبت به روش‌های نام‌برده در مقاله بهتر عمل کرده است و باعث افزایش کیفیت سیگنال شده است. روش پیشنهادی ارائه‌شده در کارهای آتی می‌تواند به‌عنوان مرحله پس‌پردازش و یا پیش‌پردازش با سایر روش‌های دیگر ترکیب شده و باعث ارائه روش‌های قدرتمندتری گردد. با توجه به اینکه پس از بهسازی سیگنال‌های نویزی به دلیل‌های متعددی، نویز به‌طور کامل

جدول ۱: مقایسه و ارزیابی PESQ و segSNR روش پیشنهادی با سایر روش‌ها به‌ازای تمام SNR و نویزها (پایگاه داده Noisex)

	Mean Performance for all noises and all SNRs		
	PESQ	SegSNR	RTF(xRT)
Hybrid Bayesian-Wavelet-NNESE (Proposed)	2.561	2.325	0.070
Perceptually Motivated Bayesian Estimator [22]	2.507	2.059	0.019
MSS-SMPR [26]	2.505	1.550	0.017
Spectral Subtraction [27]	2.505	1.356	0.010
MMSE-LSA [28]	2.465	0.952	0.045
MMSE-SA [29]	2.368	-0.509	0.020
Geometric Approach to Spectral Subtraction [30]	2.351	1.665	0.030
Audio De-Noising Using MATLAB Wavelets Toolbox [33]	2.245	-4.443	0.445
Modulation Spectral Subtraction [31]	2.124	-1.283	5.164
EMDH [13]	2.114	-4.503	4.088
NNESE [18]	2.024	-5.284	0.047
Teager Energy Operated Perceptual Wavelet Packet [32]	1.105	-1.101	1.422



شکل ۷: نمودار میانگین PESQ به ازای همه نویزها و همه دسی بل‌ها

- [10] M. A. A. El-Fattah, M. I. Dessouky, A. M. Abbas, S. M. Diab, E.-S. M. El-Rabaie, W. Al-Nuaimy, *et al.*, "Speech enhancement with an adaptive Wiener filter," *International Journal of Speech Technology*, vol. 17, pp. 53-64, 2014.
- [11] T. F. Sanam and C. Shahnaz, "Enhancement of noisy speech based on a custom thresholding function with a statistically determined threshold," *International Journal of Speech Technology*, vol. 15, pp. 463-475, 2012.
- [12] N. Chatlani and J. J. Soraghan, "EMD-based filtering (EMDF) of low-frequency noise for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, pp. 1158-1166, 2012.
- [13] L. Zao, R. Coelho and P. Flandrin, "Speech enhancement with EMD and hurst-based mode selection," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 22, pp. 899-911, 2014. <http://lasp.ime.eb.br/index.php?vPage=downloads>
- [14] S. Alaya, N. Zoghlami and Z. Lachiri, "Speech enhancement based on perceptual filter bank improvement," *International Journal of Speech Technology*, vol. 17, pp. 253-258, 2014.
- [15] S. Surendran and T. K. Kumar, "Perceptual Subspace Speech Enhancement with Variance Normalization," *Procedia Computer Science*, vol. 54, pp. 818-828, 2015.
- [16] W. Jin, X. Liu, M. S. Scordilis and L. Han, "Speech enhancement using harmonic emphasis and adaptive comb filtering," *IEEE transactions on audio, speech, and language processing*, vol. 18, pp. 356-368, 2010.
- [۱۷] مجتبی حاجی آبادی، عباس ابراهیمی مقدم و حسین خوش بین، «حذف نویز صوتی مبتنی بر یک الگوریتم وفقی نوین»، فصلنامه مهندسی برق دانشگاه تبریز، ۴۶(۳)، ۱۳۹۵.
- [18] R. Tavares and R. Coelho, "Speech Enhancement with Nonstationary Acoustic Noise Detection in Time Domain," *IEEE Signal Processing Letters*, vol. 23, pp. 6-10, 2016. <http://lasp.ime.eb.br/index.php?vPage=downloads>
- [19] J. Sun, W. Xu and B. Feng, "A global search strategy of quantum-behaved particle swarm optimization," in *Cybernetics and Intelligent Systems, 2004 IEEE Conference on*, 2004, pp. 111-116.
- [20] J. Sun, C.-H. Lai and X.-J. Wu, *Particle swarm optimisation: Classical and quantum perspectives*, CRC Press, 2011.
- [21] D. Pastor and F.-X. Socheleau, "Robust estimation of noise standard deviation in presence of signals with unknown distributions and occurrences," *IEEE Transactions on Signal Processing*, vol. 60, pp. 1545-1555, 2012.
- [22] P. C. Loizou, "Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum," *IEEE Transactions on Speech and Audio Processing*, vol. 13, pp. 857-869, 2005. <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>
- [۲۳] مسعود گراونچی زاده، ساناز قائمی سردودی، «بهبود کیفیت گفتار مبتنی بر بهینه سازی ازدحام ذرات با استفاده از ویژگیهای ماسک گذاری سیستم شنوایی انسان»، فصلنامه مهندسی برق دانشگاه تبریز، ۱۳۹۵.
- [24] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, *et al.*, "TIMIT acoustic-phonetic continuous speech corpus," linguistic data consortium, 1993.
- [25] A. Varga and H. J. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech communication*, vol. 12, pp. 2.1993,251-47.

حذف نمی‌شود پیشنهاد می‌شود با کمک گرفتن از سایر روش‌هایی که قادر به تخمین انواع نویزهای ایستان و غیرایستان هستند و به سیگنال گفتار اصلی آسیبی نمی‌رسانند مراحل به‌عنوان پیش‌پردازش و یا پس‌پردازش و یا هر دو بر روی سیگنال گفتار نویزی اعمال شود تا در نهایت سیگنال بهسازی شده دارای کیفیت بالایی باشد.

پیوست

اسامی فایل‌های استفاده‌شده از پایگاه داده TIMIT: (ده فایل تمیز برای قسمت آموزش و ده فایل تمیز برای قسمت تست مورد استفاده قرار گرفته است. هر کدام از جملات ۲ الی ۳ ثانیه بوده‌اند.)

ده فایل قسمت تست:

1.DR3_FDJO0_SI1565	2.DR3_FEME0_SX155
3.DR3_FGRW0_SI1990	4.DR3_FLAC0_SX181
5.DR3_MAKR0_SX272	6.DR3_MAPV0_SX213
7.DR3_MBEF0_SX381	8.DR3_MCDD0_SI1513
9.DR3_MDBB1_SX16	10.DR4_FADG0_SI1279

ده فایل قسمت آموزش:

1.DR1_FAKS0_SI943	2.DR1_FDAC1_SI844
3.DR1_FELCO_SI756	4.DR1_FJEM0_SI634
5.DR1_MDAB0_SX139	6.DR1_MJSW0_SI1010
7.DR1_MRJO0_SI734	8.DR1_MSTK0_SI2222
9.DR2_FCMR0_SX295	10.DR2_MCCS0_SI839

مراجع

- [1] J. Benesty, S. Makino and J. Chen, *Speech Enhancement*, Springer, 2005.
- [2] Apparatus for suppressing noise and distortion in communication signals. U.S. Patent 3,180,936, issued April 27, 1965.
- [3] P. C. Loizou, *Speech enhancement: Theory and practice, 2nd Ed*, CRC press, 2013.
- [4] M. Berouti, R. Schwartz and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'79.*, 1979, pp. 208-211.
- [5] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Speech and Audio processing*, vol. 9, pp. 504-512, 2001.
- [6] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *Journal of the royal statistical society: series B (statistical methodology)*, vol. 59, pp. 319-351, 1997.
- [7] S. Manikandan, "Speech enhancement based on wavelet denoising," *Academic Open Internet Journal*, vol. 17, 2006.
- [8] G. Lathoud, M. Doss, and H. Boulard, "Channel normalization for unsupervised spectral subtraction," in *Proceedings of ASRU*, 2005.
- [9] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proceedings of the IEEE*, vol. 67, pp. 1586-1604, 1979.

- short-time modulation domain", Speech Communication, 2010. <https://maxwell.ict.griffith.edu.au/spl/research/modspecspl/>
- [32] Islam, Md Tauhidul, et al. "Speech enhancement based on student t modeling of Teager energy operated perceptual wavelet packet coefficients and a custom thresholding function." IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP) 23.11 (2015): 1800-1811. <https://sourceforge.net/projects/wavelet-speech-enhancement/files/Matlab%20Codes/>
- [33] Villanueva-Luna, A. E., Jaramillo-Nuñez, A., Sanchez-Lucero, D., Ortiz-Lima, C. M., Aguilar-Soto, J. G., Flores-Gil, A., & May-Alarcon, M. (2011). De-noising audio signals using MATLAB wavelets toolbox. In *Engineering education and research using MATLAB*. InTech.
- [34] A. Rix, J. Beerends, M. Hollier and A. Hekstra, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," *ITU-T Recommendation*, vol. 862, 2001
- [26] Lu, Y. and Loizou, P. "Estimators of The Magnitude-Squared Spectrum and Methods for Incorporating SNR Uncertainty," IEEE Trans. Audio, Speech, Language Processing, 19(5), 1123-1137. <http://ecs.utdallas.edu/loizou/speech/software.htm>
- [27] M. Berouti, R. Schwartz and J. Makhoul Enhancement of speech corrupted by acoustic noise Proc IEEE ICASSP, 1979, 4, 208-211 <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>
- [28] Ephraim, Y. and Malah, D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator IEEE Trans Acoustics Speech and Signal Processing, 33(2):443-445, Apr 1985. <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>
- [29] Ephraim, Y. and Malah, D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator IEEE Trans Acoustics Speech and Signal Processing, 32(6):1109-1121, Dec 1984. <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>
- [30] Yang Lu, Philipos C. Loizou(2008) . A geometric approach to spectral subtraction Speech Communication 50 453-466. <http://ecs.utdallas.edu/loizou/speech/software.htm>
- [31] Paliwal, K.K., Wojcicki, K.K. and B. Schwerin,"Single-channel speech enhancement using spectral subtraction in the

زیر نویس‌ها

- ¹ Minimum Mean Square Error
- ² Empirical Mode Decomposition
- ³ Empirical Mode Decomposition-based Filtering
- ⁴ Empirical Mode Decomposition and Hurst-based
- ⁵ Non-stationary Noise Estimation for Speech Enhancement
- ⁶ Quantum-behaved Particle Swarm Optimization
- ⁷ Short Time Fourier Transform
- ⁸ Module
- ⁹ Particle Swarm Optimization
- ¹⁰ Soft Masking using a PRiori SNR uncertainty on Magnitude Squared Spectrum
- ¹¹ Minimum Mean Square Error Log-Spectral Amplitude estimator
- ¹² Minimum Mean Square Error short-time Spectral Amplitude estimator