

بهبود کیفیت گفتار مبتنی بر بهینه‌سازی ازدحام ذرات با استفاده از ویژگی‌های ماسک‌گذاری سیستم شنوایی انسان

مسعود گراوانچی‌زاده^۱، دانشیار، ساناز قائمی سردرودی^۲، کارشناسی ارشد

۱- دانشکده مهندسی برق و کامپیوتر - دانشگاه تبریز - تبریز - ایران - geravanchizadeh@tabrizu.ac.ir

۲- دانشکده مهندسی برق و کامپیوتر - دانشگاه تبریز - تبریز - ایران - ghaemi. ee@gmail.com

چکیده: در این مقاله، روش‌های دوکاناله جدید زیرفضایی سیگنال با استفاده از تکنیک‌های حذف نویز ادراکی برای بهبود کیفیت سیگنال گفتار پیشنهاد شده‌اند. روش‌های پیشنهادی از مشخصات ماسک‌گذاری سیستم شنوایی انسان برای کاهش نویز باقی‌مانده قابل شنیدن بهره می‌گیرند. روش تجزیه به مقادیر منفرد کسری ادراکی، تکنیک بهینه‌سازی ذرات را برای تخمین نویز جمعی استفاده می‌کند. نتایج شبیه‌سازی دقیقی که توسط دو معیار مختلف subjective و objective ارزیابی شده است، نشان‌دهنده کیفیت بهتر سیگنال پردازش‌شده توسط روش‌های پیشنهادی نسبت به الگوریتم‌های قبلی با هر دو نوع نویز ایستا و غیرایستا و به‌خصوص نویز غیرسفید است.

واژه‌های کلیدی: آستانه‌های ماسک‌گذاری شنوایی، تخمین‌گر حداقل مربعات، تخمین‌گر حداقل واریانس، بهینه‌سازی ازدحام ذرات، تجزیه به مقادیر منفرد کسری، بهبود کیفیت گفتار

PSO-Based Speech Enhancement using Masking Properties of Human Auditory system

M. Geravanchizadeh¹, Associated professor, S. Ghaemi Sardaroudi², Msc. Student

1, 2- Faculty of Electrical & Computer Engineering, University of Tabriz, Tabriz, Iran

Emails: 1-geravanchizadeh@tabrizu.ac.ir, 2-ghaemi. ee@gmail.com

Abstract: New dual-channel perceptually motivated subspace-based approaches are proposed for enhancement of speech corrupted by noise. The proposed methods take the frequency masking properties of the human auditory system into account and reduce perceptual effects of the residual noise. The perceptually constrained quotient singular value decomposition (PCQSVD) algorithm uses the particle swarm optimization (PSO) technique to estimate the additive noise. Very carefully performed objective evaluations and subjective tests show that the proposed approaches here can offer improved speech quality, as compared to previous methods, in the case of stationary and nonstationary noises, especially when the additive noise is nonwhite.

Keywords: Auditory masking threshold, least-squares estimation, minimum-variance estimation, particle swarm optimization, quotient singular value decomposition, speech enhancement

تاریخ ارسال مقاله: ۱۳۹۱/۷/۱۸

تاریخ اصلاح مقاله: ۱۳۹۱/۱۰/۲۶

تاریخ پذیرش مقاله: ۱۳۹۳/۱/۳۰

نام نویسنده مسئول: دکتر مسعود گراوانچی‌زاده

نشانی نویسنده مسئول: ایران - تبریز - خوار ۲۹ بهمن - دانشگاه تبریز - دانشکده مهندسی برق و کامپیوتر

۱- مقدمه

کیفیت سیستم‌های مخابراتی صوتی تحت تأثیر شرایط مختلف قرار می‌گیرد. منابع نویز مختلف مانند موتور، باد، جاده و غیره با سیگنال گفتار تمیز تداخل می‌یابند و کیفیت و قابلیت فهم سیگنال را کاهش می‌دهند. هدف روش‌های بهبود کیفیت گفتار، افزایش کارایی سیستم‌های مخابراتی صوتی است.

در سال‌های اخیر، روش زیرفضایی سیگنال مبتنی بر تجزیه به مقادیر منفرد^۱ (SVD) به دلیل کارایی بالای کاهش نویز، محبوبیت زیادی کسب کرده است [۱-۳]. روش‌های بهبود کیفیت گفتار زیرفضایی سیگنال، سیگنال ورودی نویزی را به دو زیرفضای مجزای سیگنال و نویز تفکیک می‌کنند. برای بهبود کیفیت سیگنال گفتار، زیرفضای نویز حذف می‌شود. این روش برای نویز جمعی سفید بسیار مؤثر است. زمانی که نویز غیرسفید است، سفید کردن طیف نویز قبل از تجزیه به مقادیر منفرد و انجام فرآیند عکس سفیدسازی نویز بعد از آن، یک روش منطقی به نظر می‌رسد. برای اجتناب از این فرآیندهای اضافی، روش‌های مبتنی بر تجزیه به مقادیر منفرد کسری^۲ (QSVD) پیشنهاد شده‌اند [۴، ۵]. در روش‌های مبتنی بر QSVD، فرآیند سفیدسازی نویز و تجزیه به مقادیر منفرد به صورت یکپارچه انجام می‌شود و برای نویزهای رنگی بسیار مناسب است [۶-۱۰]. این الگوریتم‌ها تعمیم یافته بهبود کیفیت گفتار به روش QSVD می‌باشند [۱۱].

با وجود اینکه کارایی روش‌های مبتنی بر تجزیه به مقادیر منفرد کسری بیش‌تر از روش‌های قدیمی تجزیه به مقادیر منفرد است، ولی هنوز مقداری نویز موزیکال در سیگنال بهبود کیفیت یافته در شرایط سیگنال به نویز پائین شنیده می‌شود [۴، ۶]. به همین دلیل، آستانه‌های ماسک‌گذاری^۳ (AMT) سیستم شنوایی انسان [۱۲، ۱۳] با روش QSVD ترکیب می‌شوند تا روش بهبود کیفیت گفتار بهتری حاصل شود [۱۴]. این روش، تجزیه به مقادیر منفرد کسری ادراکی^۴ (PCQSVD) خوانده می‌شود.

روش‌های مبتنی بر QSVD، همان‌طور که از نام آن‌ها پیداست، در حوزه مقادیر منفرد تعمیم یافته^۵ (GSVD) کار می‌کنند. لذا، نگاشت حوزه فرکانس به مقادیر ویژه [۱۵] را به نگاشت حوزه فرکانس به مقادیر منفرد تعمیم یافته بسط می‌دهیم [۱۴].

در این مقاله، روش‌های نوینی مبتنی بر PCQSVD-PSO برای سیستم‌های بهبود کیفیت گفتار دو کاناله ارائه شده است. سیگنال گفتار تمیز و نویز زمینه به صورت همزمان جمع می‌شوند. الگوریتم بهینه‌سازی ذرات^۶ (PSO) برای تخمین نویز زمینه استفاده می‌شود. سپس، روش PCQSVD نویز باقی مانده را کاهش داده و زیر آستانه ماسک‌گذاری شنوایی قرار می‌دهد.

نتایج آزمایش‌های انجام شده توسط معیارهای مختلف objective و subjective نشان می‌دهند که روش پیشنهادی دو کاناله PCQSVD-PSO نویز موجود در روش PCQSVD را بیش‌تر کاهش

داده و کیفیت سیگنال پردازش یافته را بدون توجه به ایستا^۷ یا غیر ایستا^۸ بودن نویز افزایش می‌دهد. این روش برای نویزهای غیرسفید کارا تر است.

ساختار این مقاله بدین صورت است. در بخش ۲، الگوریتم تخمین نویز توسط روش PSO معرفی می‌شود. در بخش ۳، روش بهبود کیفیت گفتار مبتنی بر QSVD به طور اجمالی معرفی می‌شود. روش‌های مبتنی بر PCQSVD، شامل روش‌های پیشنهادی PCQSVD-LSE-PSO و PCSVD-MVE-PSO در بخش ۴ توضیح داده می‌شوند. نتایج شبیه‌سازی در بخش ۵ آورده شده است. در بخش ۶ مقاله نیز نتیجه‌گیری شده است.

۲- الگوریتم تخمین نویز با تکنیک بهینه‌سازی ازدحام ذرات

شکل (۱)، بلوک دیاگرام کلی سیستم فیلتر و فقی را نشان می‌دهد. در سیستم تخمین نویز و فقی دو کانال وجود دارد. کانال اول سیگنال نویزی $x(n)$ ، حاوی مجموع سیگنال گفتار تمیز $s(n)$ و نویز $b(n)$ است. لذا ورودی کانال اولیه^۹ به صورت زیر است:

$$x(n) = s(n) + b(n). \quad (1)$$

کانال دوم حاوی نویز مرجع^{۱۰} $b'(n)$ است که به صورت ناهمبسته با سیگنال گفتار تمیز $s(n)$ فرض می‌شود. تابع تبدیل صوتی مسیر بین منبع نویز و سیگنال گفتار، $F(\tau)$ فرض می‌شود. به عبارتی، سیگنال گفتار تمیز با نویز فیلتر شده توسط سیستم محیط $F(\tau)$ مخلوط می‌شود. هدف آن است که نویز موجود در کانال اول با استفاده از نویز مرجع ضبط شده در کانال دوم تخمین زده شود. برای رسیدن به این مقصود از فیلتر و فقی استفاده می‌شود. این فیلتر با استفاده از سیگنال خطا و نویز مرجع، تابع تبدیل $F(\tau)$ را مدل می‌کند. در این صورت، با تخمین ضرایب فیلتر و فقی می‌توان نویز جمع شده با سیگنال تمیز را تخمین زد. خروجی فیلتر و فقی، $y(n)$ ، تخمینی از نویز موجود در سیگنال نویزی است. سیگنال خطای $e(n)$ نیز به صورت زیر تعریف می‌شود:

$$e(n) = x(n) - y(n). \quad (2)$$

برای مدل کردن تابع تبدیل محیط، می‌توان از روش‌های بهینه‌سازی آماری بهره جست. در سال‌های اخیر، الگوریتم‌های متعددی مبتنی بر بهینه‌سازی اکتشافی^{۱۱} برای این امر استفاده می‌شود. این الگوریتم‌ها از هوش گروهی پرندگان و حشرات در طبیعت الهام گرفته‌اند. در برخی مطالعات از روش بهینه‌سازی ازدحام ذرات (PSO) به دلیل سادگی و سرعت همگرایی بالای آن استفاده شده است. یکی از محدودیت‌های اصلی الگوریتم PSO، احتمال افتادن در دام نقطه بهینه محلی است. الگوریتم جستجوی جاذبه‌ای^{۱۲} (GSA)، یک روش نوین از طبقه الگوریتم‌های بهینه‌سازی با الگوی متفاوت جستجو است. الگوریتم GSA از فاصله میان نماینده‌های^{۱۳} همسایه برای به‌هنگام‌سازی مکان نماینده‌ها استفاده می‌کند. مرجع [۱۶] از روش‌های PSO و GSA برای طراحی کنترلر به منظور ردیابی سیستم‌های غیرخطی^{۱۴} استفاده کرده است. مرجع [۱۷] از این دو

همچنین، فرض می‌کنیم $\mathbf{P}_i(t) = (p_{i1}(t), \dots, p_{id}(t), \dots, p_{id}(t))$ و $\mathbf{V}_i(t) = (v_{i1}(t), \dots, v_{id}(t), \dots, v_{id}(t))$ به ترتیب، مکان و سرعت ذره i در فضای جستجوی D بعدی در مرحله تکرار t باشد. در هر مرحله، fitness هر ذره طبق تابع fitness انتخاب شده، برآورد می‌شود. این الگوریتم مناسب‌ترین پارامترهای هر ذره $(\mathbf{PBest}_i(t))$ و مناسب‌ترین ذره کلی $(\mathbf{GBest}(t))$ را که با آن‌ها مواجه شده است، ذخیره و جایگزین می‌کنند. به عبارتی، $\mathbf{PBest}_i(t) = (pbest_{i1}(t), \dots, pbest_{id}(t), \dots, pbest_{id}(t))$ بهترین تجربه ذره i تا مرحله t و $\mathbf{GBest}(t) = (gbest_1(t), \dots, gbest_d(t), \dots, gbest_d(t))$ بهترین تجربه حاصله از $\mathbf{PBest}_i(t)$ در کل ازدحام در مرحله تکرار t است. پارامترهای هر ذره در هر مرحله t به‌هنگام می‌شوند و در این به‌هنگام شدن از روابط زیر استفاده می‌گردد [۲۰]:

$$v_{id}(t+1) = v_{id}(t) + c_1 r_1 (pbest_{id}(t) - x_{id}(t)) + c_2 r_2 (gbest_{id}(t) - x_{id}(t)), \quad (3)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1). \quad (4)$$

در این معادلات، c_1 فاکتور فراگیری^{۱۹} و c_2 فاکتور فراگیری اجتماعی^{۲۰} است. پارامترهای r_1 و r_2 دو عدد تصادفی هستند که به صورت یکنواخت در بازه $[0, 1]$ توزیع شده‌اند. به‌هنگام شدن مکان و سرعت ذره در فضای جستجوی دوبعدی در شکل (۲) نشان داده شده است. اگر تعداد مراحل تکرار الگوریتم PSO، T و M ، تعداد ذرات در نظر گرفته شود، پیچیدگی محاسباتی الگوریتم، به صورت $O(MT)$ خواهد بود [۲۰، ۲۱]. لذا، از آن جایی که پیچیدگی محاسباتی الگوریتم PSO متناسب با افزایش تعداد ذرات ازدحام زیاد می‌شود، تا حد ممکن با تعداد ذرات کم‌تر به جستجوی مکان بهینه پرداخته می‌شود.

در این مقاله، پارامترهای بهینه فیلتر و فیلتر و فقی، \mathbf{w}_{opt} ، به منظور تخمین نویز، توسط روش بهینه‌سازی زیر انجام می‌گیرد:

$$\mathbf{w}_{opt} = \arg \min_{\mathbf{w}} E[\mathbf{e}^2(n)], \quad (5)$$

که در آن تابع هزینه $\mathbf{e}(n)$ به صورت زیر تعریف می‌شود:

$$\mathbf{e}(n) = \mathbf{x}(n) - [\mathbf{w}^T(n)\mathbf{b}'(n)]. \quad (6)$$

فلوچارت الگوریتم پیشنهادی تخمین نویز توسط PSO در شکل (۳) نشان داده شده است [۲۵].

۳- بررسی اجمالی روش بهبود کیفیت گفتار مبتنی بر QSVD

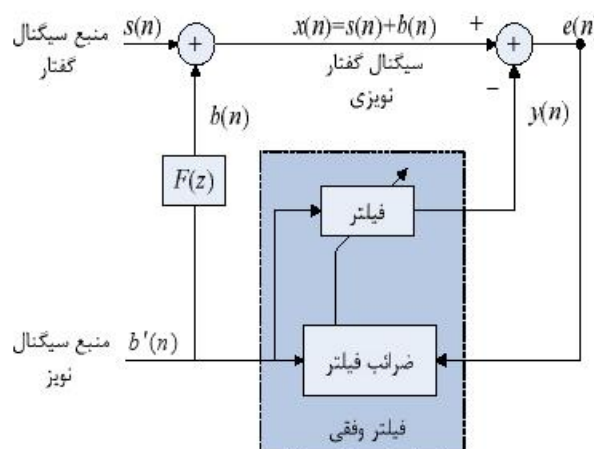
روش مبتنی بر QSVD برای بهبود کیفیت سیگنال گفتار نویزی در این قسمت بررسی می‌شود. فرض می‌شود $\mathbf{x}(n) \in \mathbb{R}^K$ بردار سیگنال نویزی ورودی است و به صورت مجموع سیگنال گفتار تمیز $\mathbf{s}(n) \in \mathbb{R}^K$ و نویز $\mathbf{b}(n) \in \mathbb{R}^K$ بیان می‌شود:

$$\mathbf{x}(n) = \mathbf{s}(n) + \mathbf{b}(n), \quad (7)$$

که در این معادله $\mathbf{x}(n) = [x_n, x_{n-1}, \dots, x_{n-K+1}]^T$ است.

روش برای تخمین توابع ریاضی بهره گرفته است. PSO و GSA در مرجع [۱۸، ۱۹]، به منظور بهبود کیفیت گفتار به کار گرفته شده‌اند. در مراجع [۱۶، ۱۷]، روش PSO نسبت به GSA عملکرد بهتری ارائه کرده است، در حالی که در مراجع [۱۸، ۱۹]، روش GSA نسبت به PSO کارایی بهتری داشته است.

به دلیل عدم توانایی حرکت سریع اجرام در روش GSA و ایستایی^{۱۵} در نقاط حداقل و حداکثر محلی و همچنین، سادگی و سرعت همگرایی بالای روش PSO، در این مقاله، از تکنیک PSO برای مدل کردن تابع تبدیل محیط، $F(z)$ ، میان منبع سیگنال گفتار و نویز استفاده می‌شود. خروجی فیلتر، $y(n)$ ، به عنوان تخمینی از نویز موجود در سیگنال نویزی $x(n)$ است. از این سیستم برای تخمین ماتریس نویز در روش‌های پیشنهادی PCQSVD-LSE-PSO و PCQSVD-MVE-PSO استفاده می‌شود.

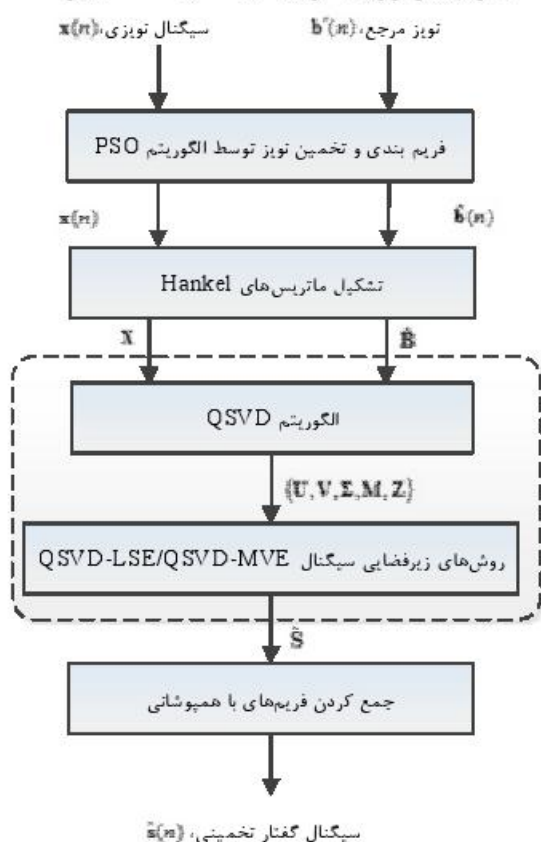


شکل ۱: سیستم تخمین نویز

روش بهینه‌سازی ازدحام ذرات یا به عبارت ساده‌تر PSO یکی از قدیمی‌ترین الگوریتم‌ها در دسته تکنیک‌های مبتنی بر بهینه‌سازی آماری است. این روش، اولین بار در سال ۱۹۹۵ توسط Eberhart و Kennedy [۲۰، ۲۱] بر پایه توجه به حرکت گروه‌هایی چون پرندگان، حشرات و غیره بنا نهاده شد. این الگوریتم سعی می‌کند که از فرآیند طبیعی مخابرات گروهی الهام بگیرد تا به یک مکان بهینه دست پیدا کند. الگوریتم بهینه‌سازی ازدحام ذرات استاندارد^{۱۶} با یک جمعیت تصادفی افراد شروع می‌شود که در اینجا به آن ازدحام ذرات^{۱۷} گفته می‌شود. اساس بحث، جستجوی کارآمد فضای حل به وسیله گروه بندی ذرات، جهت دسترسی به بهترین راه حل طبق مراحل قبلی و به منظور همگرایی به یک پاسخ با خطای حداقل منفرد^{۱۸} است. الگوریتم PSO در بسیاری از کاربردها به دلیل پیاده‌سازی آسان و سرعت همگرایی بالا موفق بوده است [۲۲-۲۴]. در روش PSO، هر ذره از تجربه خودش و دیگر ذرات هنگام جستجو سود می‌برد. فرض می‌شود که روش بهینه‌سازی ازدحام ذرات با یک ازدحام تصادفی متشکل از M ذره در فضای D بعدی شروع شود که این پارامترها می‌بایست بهینه شوند.

۳-۲- مرحله دوم: تشکیل ماتریس‌های Hankel

در این مرحله، بردار K بعدی مشاهدات و فریم نویز تخمینی $\hat{\mathbf{b}}(n)$ به فرم Hankel درمی‌آید. ماتریس Hankel، ماتریسی است که درایه‌های قطر فرعی آن یکسان و ثابت هستند. به عبارتی، بردار $\mathbf{x}(n)$ و $\hat{\mathbf{b}}(n)$ ، به ترتیب، در ماتریس مشاهدات $\mathbf{X}_{M \times N}(n)$ و ماتریس نویز تخمینی $\hat{\mathbf{B}}_{M \times N}(n)$ قرار داده می‌شود. ارتباط بین طول بردار مشاهدات و نویز و ابعاد ماتریس مشاهدات به صورت $K = M + N - 1$ است. در حالت کلی، N بسیار کوچک‌تر از M است. با الگوریتم QSVD که در قسمت بعدی توضیح داده می‌شود، می‌توان ماتریس سیگنال گفتار تمیز \mathbf{S} و فریم آن را از روی ماتریس‌های \mathbf{X} و $\hat{\mathbf{B}}$ تخمین زد.



شکل ۴: فرایند الگوریتم‌های بهبود کیفیت گفتار مبتنی بر QSVD (PCQSVD-LSE-PSO و PCQSVD-MVE-PSO).

۳-۳- مرحله سوم: الگوریتم QSVD

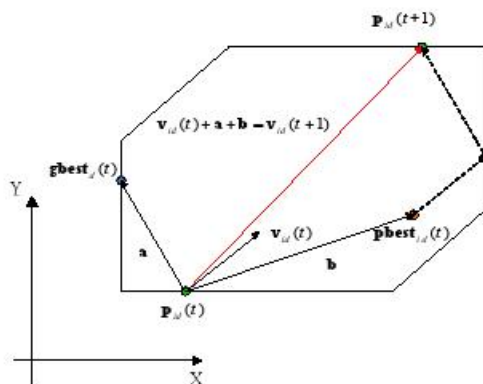
تجزیه QSVD ماتریس‌های $(\mathbf{X}, \hat{\mathbf{B}}) \in \mathbb{R}^{M \times N}$ بدین صورت است [۲۷، ۲۶]:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{Z}^T, \quad (۸)$$

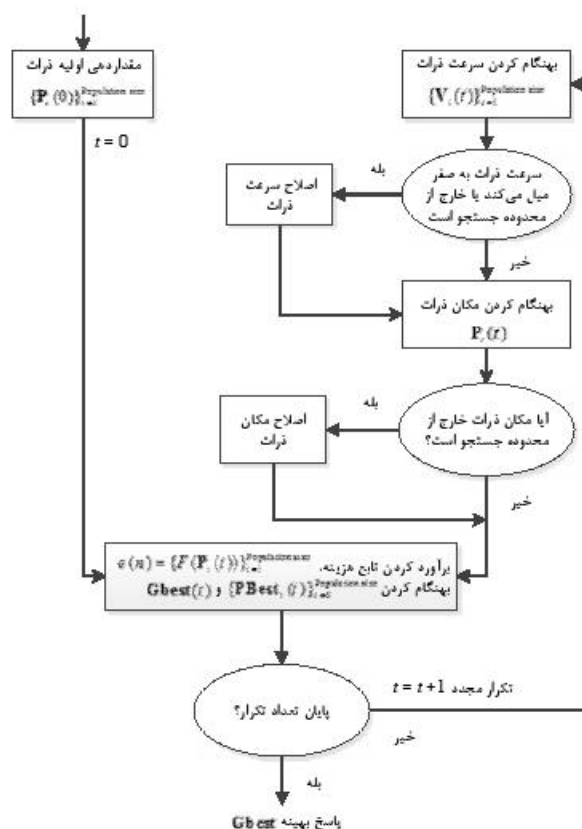
$$\hat{\mathbf{B}} = \mathbf{V}\mathbf{M}\mathbf{Z}^T, \quad (۹)$$

که در این معادله $\mathbf{U} \in \mathbb{R}^{M \times N}$ ، $\mathbf{V} \in \mathbb{R}^{M \times N}$ ، $\mathbf{Z} \in \mathbb{R}^{N \times N}$ ، $\mathbf{\Sigma} \in \mathbb{R}^{N \times N}$ ، $\mathbf{M} \in \mathbb{R}^{N \times N}$ می‌باشند.

ماتریس‌های $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_N)$ و $\mathbf{M} = \text{diag}(\mu_1, \dots, \mu_N)$ قطری هستند، به نحوی که عناصر قطری ماتریس‌های $\mathbf{\Sigma}$ و \mathbf{M} ، به ترتیب، به صورت نزولی و صعودی قرار گرفته‌اند. اگر نویز رتبه کامل و



شکل ۲: حرکت یک ذره در فضای جستجو

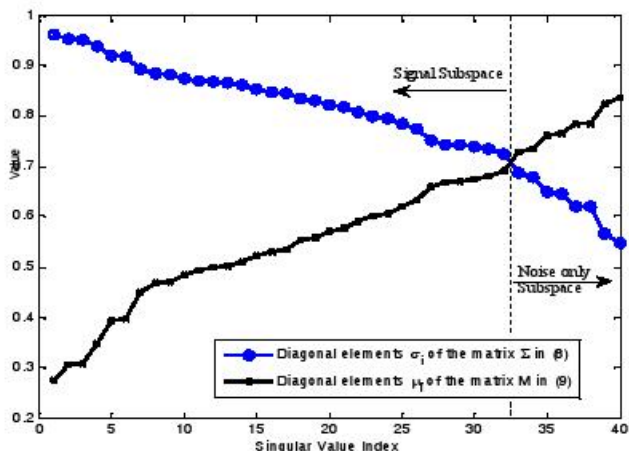


شکل ۳: فلوچارت الگوریتم پیشنهادی تخمین نویز مبتنی بر PSO.

هدف مقاله تخمین $\mathbf{s}(n)$ از روی سیگنال $\mathbf{x}(n)$ است. شکل (۴) فرایند بهبود کیفیت گفتار مبتنی بر QSVD را نشان می‌دهد که شامل مراحل زیر است:

۳-۱- مرحله اول: فریم‌بندی و تخمین نویز

ابتدا، سیگنال گفتار ورودی به فریم‌های $\mathbf{x}(n)$ با طول پنجره K که با یک‌دیگر همپوشانی دارند، تقسیم می‌شود. سپس، فرایند بهبود کیفیت گفتار برای هر فریم اجرا می‌شود. لازم به ذکر است، روش‌های پیشنهادی PCQSVD-LSE-PSO و PCQSVD-MVE-PSO برای تخمین نویز جمعی، از الگوریتم PSO استفاده می‌کنند.



شکل ۶: نمونه ای از عناصر قطری σ_i و μ_i به ترتیب، مربوط به ماتریس‌های Σ و M . حاصله از معادلات (۸) و (۹) برای یک فریم سیگنال گفتار نوعی با

$$N = 40$$

که در این معادله، ماتریس‌های U و Z از معادلات (۸) و (۹) حاصل می‌شوند و بردارهای $u_i \in \mathbb{R}^{M \times 1}$ و $z_i \in \mathbb{R}^{N \times 1}$ ، به ترتیب، i امین ستون قطری با عناصر σ'_i به شرح ذیل است:

$$\sigma'_i = \begin{cases} \sigma_i & 1 \leq i \leq k \\ 0 & k+1 \leq i \leq N \end{cases} \quad (12)$$

در روش بهبود کیفیت گفتار مبتنی بر QSVD-MVE، ماتریس سیگنال S با تخمین گر حداقل واریانس (MVE) تخمین زده می‌شود. هدف تخمین گرهای مختلف پیدا کردن ماتریس تبدیل W^* است که با ضرب آن در ماتریس مشاهدات X ، سیگنال تخمینی \hat{S} به دست می‌آید.

ماتریس تبدیل تخمینی W^* ، نرم فروبنیوس دو ماتریس S و XW را حداقل می‌کند. مسئله حداقل سازی بدین صورت بیان می‌شود:

$$W^* = \min_W \|XW - S\|_F \quad (13)$$

تخمین ماتریس S بر اساس معادله (۱۳)، با وزن دار کردن درایه‌های مربوط به زیرفضای فقط سیگنال و سیگنال به علاوه نویز (وزن زیرفضای فقط سیگنال، یک است) و صفر کردن درایه‌های مربوط به زیر فضای فقط نویز ماتریس Σ حاصله از معادله (۸) به دست می‌آید:

$$\hat{S} = XW^* = U\Sigma'Z^T = \sum_{i=1}^k \sigma'_i u_i z_i \quad (14)$$

در این معادله، ماتریس Σ' ، ماتریس قطری با عناصر σ'_i به شرح ذیل است:

$$\sigma'_i = \begin{cases} \sigma_i & 1 \leq i \leq k-r \\ \sigma_i \left(1 - \frac{\mu_i^2}{\sigma_i^2}\right) & k-r+1 \leq i \leq k \\ 0 & k+1 \leq i \leq N \end{cases} \quad (15)$$

شکل (۷) عناصر قطری σ_i ، μ_i و σ'_i مربوط به ماتریس‌های قطری Σ ، M و Σ' به کاررفته در معادلات (۸)، (۹) و (۱۵) را برای یک فریم سیگنال گفتار نوعی با $N = 40$ نشان می‌دهد.

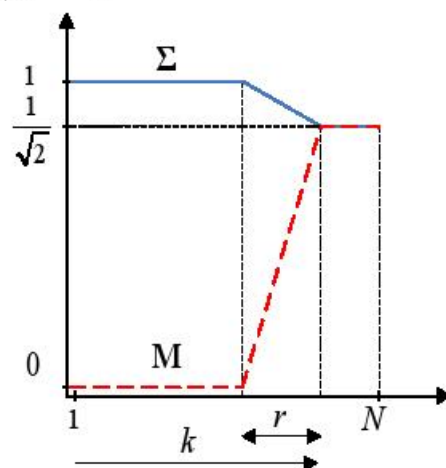
رابطه $S^T B = 0$ برقرار باشد، درایه‌های قطری $\mu_i (i=1, \dots, N)$ ماتریس M همواره در نامساوی $0 < \mu_i \leq 1/\sqrt{2}$ صدق می‌کنند، زیرا عناصر قطری این ماتریس‌ها طوری نرمالیزه شده‌اند که همواره $\sigma_i^2 + \mu_i^2 = 1$ است [۲۸]. ماتریس Z ، ماتریس منفرد تعمیم یافته^{۲۱} خوانده می‌شود و نسبت میان درایه‌های قطری (یعنی، $\delta_i = \sigma_i / \mu_i$)، مقادیر منفرد تعمیم یافته^{۲۲} ماتریس‌های X و B است [۲۹].

۳-۴ - مرحله چهارم: الگوریتم‌های مبتنی بر QSVD-LSE و QSVD-MVE

از آنجایی که در تجزیه QSVD، رابطه $\sigma_i^2 + \mu_i^2 = 1$ ($i=1, \dots, N$) در مورد مقادیر ویژه جفت ماتریس قطری Σ و M صدق می‌کند، لذا، $\sigma_i = 1$ و $\mu_i = 0$ به زیرفضای فقط سیگنال، $1/\sqrt{2} < \sigma_i < 1$ و $0 < \mu_i < 1/\sqrt{2}$ به زیرفضای سیگنال به علاوه نویز و $\sigma_i = \mu_i = 1/\sqrt{2}$ به زیرفضای فقط نویز اختصاص دارد. این مطلب در شکل (۵) نشان داده شده است. شکل (۵) برای تعیین رتبه زیرفضاهای مختلف به کار می‌رود. شکل (۶) درایه‌های قطری σ_i و μ_i ماتریس‌های قطری Σ و M به کاررفته در معادلات (۸) و (۹) را برای یک فریم نوعی $N = 40$ نشان می‌دهد.

در روش بهبود کیفیت گفتار مبتنی بر QSVD-LSE، تخمین گر حداقل مربعات (LSE) برای تخمین ماتریس سیگنال S استفاده می‌شود. این تخمین گر، مجموع مربعات خطا بین درایه‌های ماتریس سیگنال تخمینی S (تقریب رتبه پائین ماتریس مشاهدات) و ماتریس مشاهدات X را حداقل می‌کند. در حقیقت، تخمین گر LSE، نرم فروبنیوس^{۲۳} بین این دو ماتریس را حداقل می‌کند، یعنی:

$$\hat{S} = \min_{\text{rank}(S)=r} \|X - S\|_F^2 \quad (10)$$



شکل ۵: تعیین بعد زیرفضای سیگنال با k عناصر قطری ماتریس‌های M و Σ

تخمین ماتریس S بر اساس معادله (۱۰) با صفر کردن درایه‌های مربوط به زیرفضای فقط نویز ماتریس Σ حاصله از معادله (۸) به دست می‌آید، به عبارتی:

$$\hat{S} = U\Sigma'Z^T = \sum_{i=1}^k \sigma'_i u_i z_i \quad (11)$$

و سیگنال نامطلوب نویز، masker محسوب می‌شوند. معمولاً، بازه فرکانس قابل شنیدن سیستم شنوایی انسان (۲۰ Hz - ۲۰ kHz) با ۲۵ باند بحرانی مدل می‌شود.

مربع دامنه عناصر تبدیل فوریه گسسته (DFT) سیگنال گفتار تمیز در هر باند بحرانی جمع شده و سپس، با تابع spreading برای شبیه‌سازی اثر همبستگی متقابل مابین باندهای بحرانی کانال می‌شود. در مرحله بعد، این عناصر بسته به ماهیت tone-like یا noise-like باند بحرانی فریم سیگنال گفتار ورودی، بر یک سری آستانه‌ها تقسیم می‌شوند. آستانه‌های ماسک‌گذاری سیستم شنوایی انسان با نرمالیزه کردن آن‌ها، برای جبران تغییر بهره ناشی از مرحله کانولوشن و حصول اطمینان از قرار نگرفتن این آستانه‌ها زیر آستانه‌های ماسک‌گذاری مطلق، به دست می‌آیند [۳۰-۳۳].

۴-۲- روش مبتنی بر PCQSVD-LSE

روش کار الگوریتم زیرفضای سیگنال مبتنی بر PCQSVD با استفاده از تخمین گر LSE (PCQSVD-LSE)، به جز مرحله چهارم، مشابه روش مبتنی بر QSVD-LSE شرح داده شده در بخش قبل است. بدین دلیل، در اینجا فقط مرحله چهارم روش مبتنی بر PCQSVD-LSE توضیح داده می‌شود.

هدف از ترکیب ماسک‌گذاری شنوایی با روش‌های زیرفضای سیگنال، یافتن ماتریس تبدیل بهینه W^* است که فاصله فروبنیوس بین ماتریس مشاهدات و تخمین ماتریس تمیز (برای تخمین گر LS) را، تحت شرایطی که انرژی نرمالیزه شده نویز باقی‌مانده در زیرفضای سیگنال به‌علاوه نویز کم‌تر از AMT‌های نگاشت یافته و در زیرفضای فقط نویز برابر صفر باشد، حداقل کند. به عبارتی، نویز باقی‌مانده حذف نشده، بلکه با گوش انسان شنیده نمی‌شود. رابطه بهینه‌سازی برای الگوریتم PCQSVD-LSE به‌صورت زیر است:

$$W^* = \min_W \|X - S\|_F^2$$

$$\text{subject to } \begin{cases} \frac{\|v_i^T B W\|^2}{\|z_i\|^2} \leq \gamma_i, & 1 \leq i \leq k \\ \frac{\|v_i^T B W\|^2}{\|z_i\|^2} = 0, & k+1 \leq i \leq N \end{cases} \quad (16)$$

که در این معادله، بردارهای $v_i \in \mathbb{R}^{M+1}$ و $z_i \in \mathbb{R}^{1 \times N}$ ، به ترتیب، i امین ستون ($i=1, \dots, N$) ماتریس‌های V و Z می‌باشند. γ_i نیز AMT‌های نگاشت یافته به حوزه مقادیر منفرد تعمیم یافته می‌باشند که با روش توضیح داده شده در بخش ۴-۴ به دست می‌آیند.

۴-۳- روش مبتنی بر PCQSVD-MVE

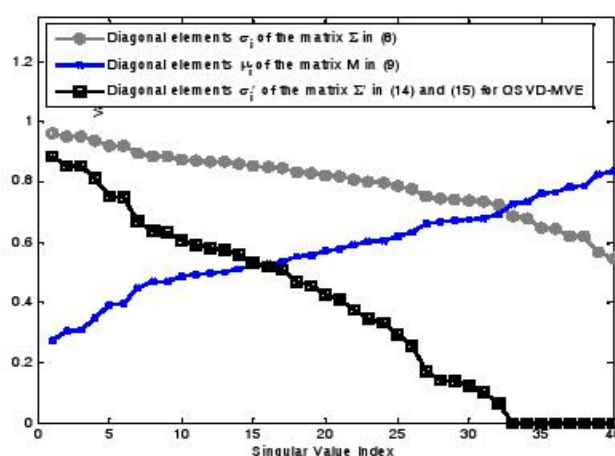
هدف در این روش، حداقل کردن فاصله فروبنیوس بین ماتریس مشاهدات فیلترشده XW و تخمین ماتریس تمیز S (برای تخمین گر MV)، تحت شرایط ادراکی و غیرقابل شنیدن کردن نویز باقی‌مانده،

۳-۵- مرحله پنجم: جمع کردن فریم‌های با همپوشانی

در نهایت، سیگنال بهبود کیفیت یافته $\hat{s}(n)$ با به هم پیوستن فریم‌های سیگنال توسط روش overlap-add به دست می‌آید.

۴- الگوریتم‌های مبتنی بر PCQSVD

کارایی الگوریتم QSVD نسبت به روش‌های تقریب طیفی جواب بهتری را در راستای بهبود کیفیت گفتار می‌دهد. با وجود این که روش‌های زیرفضایی سیگنال، کم‌تر از روش‌های تقریب طیفی، نویز موزیکال دارند، ولی در شرایط نسبت سیگنال به نویز پائین هنوز مقداری نویز موزیکال وجود دارد. برای دستیابی به کارایی بهتر صوتی، اعوجاج باقی‌مانده را با قرار دادن انرژی نویز باقی‌مانده زیر آستانه‌های ماسک‌گذاری شنوایی انسان حداقل می‌کنیم. در این بخش، آستانه‌های ماسک‌گذاری سیستم شنوایی انسان با روش‌های ذکر شده توأم استفاده می‌شود. به این روش‌ها PCQSVD می‌گویند.



شکل ۷: نمونه‌ای از عناصر قطری σ_i ، μ_i و σ_i' به ترتیب، مربوط به ماتریس‌های Σ ، M و Σ' در الگوریتم QSVD-MVE. حاصل از معادلات (۸)، (۹) و (۱۵) برای یک فریم سیگنال گفتار با $N = 40$

۴-۱- آستانه‌های ماسک‌گذاری

استفاده از خصوصیات سیستم شنوایی انسان در روش‌های حذف نویز و ظهور روش‌های "ماسک‌گذاری" پیشرفت مهمی در این زمینه محسوب می‌شود [۳۰]. در طی دهه‌های گذشته، اکثر تحقیقات روی سیستم شنوایی گوش انسان و مدل‌هایی برای شبیه‌سازی آن متمرکز گشته است. در میان این مدل‌های سیستم شنوایی انسان، می‌توان به مدل‌های یک و دو استاندارد کدینگ صوتی ISO MPEG-1/2 اشاره کرد [۳۱]. در این مقاله، از مدل MPEG-1 که در عمل بسیار معتبر است، استفاده می‌کنیم. در اینجا، خلاصه‌ای از این مدل آورده شده است [۳۱].

پدیده ماسک‌گذاری با باندهای بحرانی تعریف می‌شود. در یک باند بحرانی، یک صدا (maskee) در حضور صدای با توان بیشتر دیگر (masker) غیرقابل شنیدن می‌شود. در اینجا، سیگنال گفتار maskee

که در این معادله، ماتریس $\mathbf{G} \in \mathbb{R}^{N \times J}$ ماتریس انتقال است که تأمین سطر آن مربع دامنه تبدیل فوریه گسسته J نقطه ای تأمین ستون ماتریس \mathbf{Z} است. در نهایت، آستانه‌های ماسک‌گذاری شنوایی نگاشت یافته به حوزه مقادیر منفرد تعمیم یافته توسط رابطه زیر به دست می‌آید:

$$\gamma_i = |\gamma'_i|, \quad i = 1, 2, \dots, N. \quad (20)$$

۴-۵- فرمول بسته روش PCQSVD-LSE

تخمین ماتریس سیگنال گفتار تمیز \mathbf{S} از معادله (۱۴) حاصل می‌شود [۱۴]. در این معادله، عناصر قطری نامنفی ماتریس قطری Σ' به صورت زیر می‌باشند:

$$\sigma'_i = \begin{cases} \min \left[\sigma_i, \frac{\sigma_i \sqrt{\gamma_i}}{\mu_i} \right] & 1 \leq i \leq k \\ 0 & k+1 \leq i \leq N \end{cases} \quad (21)$$

۴-۶- فرمول بسته روش PCQSVD-MVE

تخمین ماتریس سیگنال گفتار تمیز \mathbf{S} از معادله (۱۴) حاصل می‌شود و در آن معادله عناصر قطری نامنفی ماتریس قطری Σ' بدین صورت می‌باشند:

$$\sigma'_i = \begin{cases} \sigma_i & 1 \leq i \leq k-r \\ \min \left[\sigma_i \left(1 - \frac{\mu_i^2}{\sigma_i^2} \right), \frac{\sigma_i \sqrt{\gamma_i}}{\mu_i} \right] & k-r+1 \leq i \leq k \\ 0 & k+1 \leq i \leq N \end{cases} \quad (22)$$

شکل (۸) عناصر قطری σ_i ، μ_i و σ'_i مربوط به ماتریس‌های Σ ، \mathbf{M} ، و Σ' (برای الگوریتم‌های QSVD-MVE و PCQSVD-MVE) را که، به ترتیب، در معادلات (۸)، (۹) و (۱۵) و (۲۲) به دست می‌آیند، برای یک فریم سیگنال گفتار نوعی با $N=40$ نشان می‌دهد.

۴-۷ الگوریتم‌های پیشنهادی PCQSVD-LSE-PSO و PCQSVD-MVE-PSO

همان‌طور که می‌دانیم، روش‌های PCQSVD و PCQSVD_LSE MVE نیازمند تخمینی از نویز می‌باشند. تخمین نویز امر مهمی در زمینه پردازش گفتار است. به دلیل اهمیت زیاد این مسئله، الگوریتم‌های تخمین نویز بسیاری در چند دهه گذشته پیشنهاد شده است. مطالعات متعددی در مورد تخمین نویز تک کاناله مانند روش VAD^{۲۷} [۳۷] و تخمین طیف نویز بر اساس ردیابی حداقل^{۲۸} [۳۸] انجام شده است. اخیراً، روش‌های تخمین نویز چند کاناله شهرت بسیاری یافته‌اند. این روش‌ها برای تخمین نویز بسیار مؤثر می‌باشند ولی از هزینه و حجم محاسباتی بالایی برخوردارند. سیستم حذف نویز و ففقی^{۲۹} (ANC) یک حالت خاصی از فیلترینگ بهینه است و هنگامی که اطلاعاتی در مورد نویز مرجع وجود دارد، استفاده می‌شود.

است. به طور مشابه، روش PCQSVD-MVE بدین صورت فرموله می‌شود:

$$\mathbf{W}^* = \min_{\mathbf{W}} \|\mathbf{XW} - \mathbf{S}\|_F^2 \quad (17)$$

$$\text{subject to } \begin{cases} \frac{\|\mathbf{v}_i^T \mathbf{B} \mathbf{W}\|^2}{\|\mathbf{z}_i\|^2} \leq \gamma_i & 1 \leq i \leq k \\ \frac{\|\mathbf{v}_i^T \mathbf{B} \mathbf{W}\|^2}{\|\mathbf{z}_i\|^2} = 0 & k+1 \leq i \leq N \end{cases}$$

ماتریس تمیز \mathbf{S} را با تخمین گرهای دیگری از جمله تخمین گر TDC^{۳۰} و تخمین گر SDC^{۳۱} نیز می‌توان تخمین زد [۳۵، ۳۴].

۴-۴ تخمین آستانه‌های ماسک‌گذاری شنوایی نگاشت یافته به حوزه مقادیر منفرد تعمیم یافته

روش‌های مبتنی بر PCQSVD همان‌طور که از نامشان پیداست، در حوزه مقادیر منفرد تعمیم یافته کار می‌کنند. از طرفی آستانه‌های ماسک گذاری سیستم شنوایی انسان نیز در حوزه فرکانس بیان می‌شوند. لذا، نگاشت حوزه فرکانس به مقادیر ویژه [۱۵] به نگاشت حوزه فرکانس به مقادیر منفرد تعمیم یافته بسط داده می‌شود [۱۴]. با استفاده از این نگاشت می‌توان فرمول بسته‌ای برای روش‌های PCQSVD-LSE و PCQSVD-MVE به دست آورد. هدف روش‌های مبتنی بر PCQSVD، تخمین سیگنال تمیز، محاسبه چگالی طیفی توان سیگنال تمیز، محاسبه آستانه‌های ماسک‌گذاری شنوایی در حوزه فرکانس و نگاشت آن به حوزه کاری روش‌های مبتنی بر QSVD است. چگالی طیفی توان فریم سیگنال تمیز، $\hat{\Gamma} \in \mathbb{R}^{N \times 1}$ ، توسط تخمین گر Blackman Tukey با پنجره Bartlett تخمین زده می‌شود. برای تخمین چگالی طیفی توان فریم سیگنال تمیز به مقادیر خودهمبستگی نیاز است. لذا، قدم اول در محاسبه آستانه‌های ماسک‌گذاری حوزه فرکانس، تخمین ماتریس خودهمبستگی سیگنال تمیز است. ماتریس خودهمبستگی فریم سیگنال تمیز تخمینی نیز توسط ماتریس سیگنال تمیز تخمینی، $\hat{\mathbf{S}}$ ، بدین صورت به دست می‌آید [۳۶]:

$$\hat{\Gamma} \cong \frac{1}{(K-N+1)N} \sum_{i=1}^K \sigma_i^2 \mathbf{w}_i^T, \quad (18)$$

که در این معادله k ، رتبه زیرفضای سیگنال به علاوه نویز است. بردار $\mathbf{w}_i \in \mathbb{R}^{N \times 1}$ مربع دامنه تبدیل فوریه گسسته J نقطه ای (تعداد آستانه‌های ماسک‌گذاری شنوایی است). تأمین ستون ماتریس \mathbf{Z} (یعنی در معادله (۱۱)) است. با طیف توان تخمینی $\hat{\Gamma}$ در معادله (۱۸)، بردار آستانه‌های ماسک‌گذاری شنوایی $\Theta \in \mathbb{R}^{N \times 1}$ حوزه فرکانس، با فرآیند بحث شده در بخش ۴-۱، به دست می‌آید. سپس، با وجود Θ ، آستانه‌های ماسک‌گذاری شنوایی حوزه فرکانس با روش پیشنهادی Jabloun و Champagne به حوزه مقادیر منفرد تعمیم یافته انتقال داده می‌شود [۱۵].

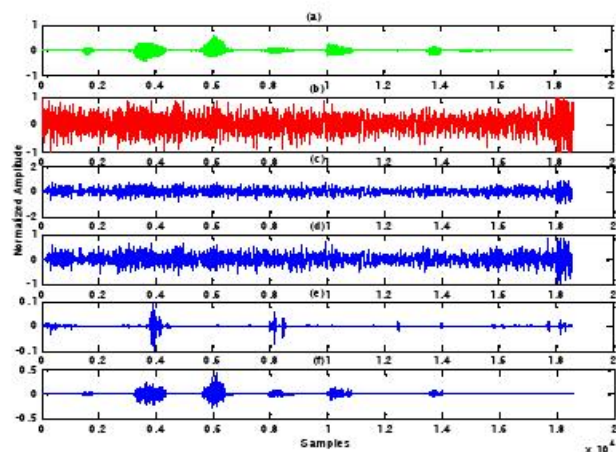
$$\gamma'_i = \frac{1}{J} \mathbf{G} \Theta, \quad (19)$$

Listening Performance Comparison [۱۴]) و یک معیار objective (SNR [۲۵]) ارزیابی می‌شوند.

شکل‌موج‌های زمانی سیگنال‌های گفتار نویزی، تمیز و بهبود کیفیت‌یافته توسط الگوریتم‌های PCQSV-D-LSE، PCQSV-D-MVE و PCQSV-D-MVE-PSO در شکل (۹) نشان داده شده‌اند.

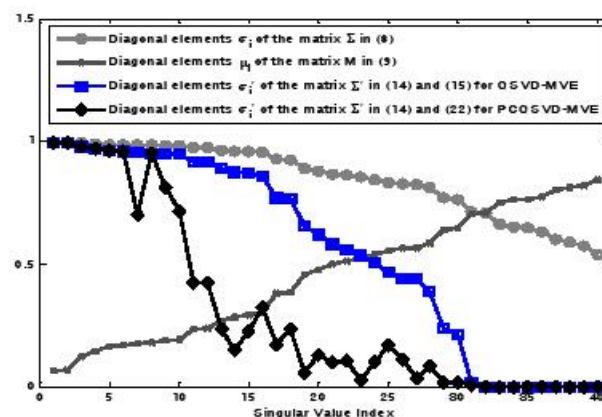
جدول (۲) و (۳) نتایج شبیه‌سازی الگوریتم‌های PCQSV-D-LSE، PCQSV-D-MVE و PCQSV-D-MVE-PSO را، با معیار ارزیابی objective برای ۳۰ جمله آلوده به ۴ نویز مختلف با نسبت سیگنال به نویز حاصله ۱۰-دسی‌بل و ۱۵-دسی‌بل، نشان می‌دهند.

همان‌طور که از نتایج جدول‌های (۲) و (۳) پیداست، روش‌های پیشنهادی مقاله نسبت به روش‌های PCQSV-D-LSE و PCQSV-D-MVE از نقطه‌نظر نسبت سیگنال به نویز عملکرد بهتری دارند. لازم به ذکر است که به دلیل کارایی بهتر تخمین‌گر حداقل واریانس (MVE) نسبت به تخمین‌گر حداقل مربعات (LSE)، روش پیشنهادی PCQSV-D-MVE-PSO در مقایسه با روش PCQSV-D-LSE-PSO قابلیت بیشتری در کاهش نویز دارد. با افزایش تعداد تکرار و ذرات الگوریتم PSO، تخمین نویز بهتری حاصل می‌شود ولی سرعت هم‌گرایی کاهش می‌یابد. شکل (۱۰) میزان متوسط بهبود نسبت سیگنال به نویز حاصله از الگوریتم‌های PCQSV-D-LSE، PCQSV-D-MVE، PCQSV-D-LSE-PSO و PCQSV-D-MVE-PSO را با چهار نوع نویز مختلف (سفید، ماشین، قطار و مهممه) نشان می‌دهد. همان‌طوری که در این شکل مشاهده می‌شود، کارایی الگوریتم‌های پیشنهادی بهتر از روش‌های موجود، به‌خصوص در نسبت‌های سیگنال به نویز پایین‌تر از صفر دسی‌بل، است.



شکل ۹: بهبود کیفیت سیگنال گفتار نوعی آلوده به نویز هواپیما با نسبت سیگنال به نویز ۱۵-دسی‌بل؛ الف) سیگنال تمیز، ب) سیگنال نویزی، پ) سیگنال بهبود کیفیت‌یافته توسط الگوریتم PCQSV-D-LSE [۱۴]، ت) سیگنال بهبود کیفیت‌یافته توسط الگوریتم PCQSV-D-MVE [۱۴]، ث) سیگنال بهبود کیفیت‌یافته توسط الگوریتم PCQSV-D-LSE-PSO، ج) سیگنال بهبود کیفیت‌یافته توسط الگوریتم PCQSV-D-MVE-PSO

در این مقاله، از الگوریتم PSO، به‌عبارتی، روش مبتنی بر بهینه‌سازی با ساختار آماری، برای تخمین نویز در روش‌های پیشنهادی PCQSV-D-MVE-PSO و PCQSV-D-LSE-PSO استفاده می‌شود. همان‌طور که نتایج شبیه‌سازی‌ها نشان می‌دهند، این الگوریتم‌ها برای بهبود کیفیت سیگنال گفتار بسیار کارا هستند.



شکل ۸: عناصر قطری σ_i ، μ_i و σ'_i مربوط به ماتریس‌های Σ ، M و Σ' در الگوریتم‌های OSVD-MVE و PCQSV-D-MVE حاصل از معادلات (۸)، (۹)، (۱۵) و (۲۲) برای یک فریم از سیگنال گفتار با $N = 40$.

۵- نتایج شبیه‌سازی

این الگوریتم‌ها توسط برنامه Matlab شبیه‌سازی شده‌اند. برای شبیه‌سازی از دادگان NOIZEUS [۳۹] و NoiseX [۴۰] استفاده شده است. سیگنال‌های گفتار تمیز توسط سه گوینده زن و سه گوینده مرد بیان شده است. جملات این سیگنال‌ها از لحاظ فوتوتیکی بالانس می‌باشند. نرخ جملات در اصل ۲۵ کیلوهرتز بوده که به ۸ کیلوهرتز کاهش داده شده است. نویزها به‌صورت مصنوعی با سیگنال‌های گفتار جمع شده و سیگنال‌های نویزی با نسبت سیگنال به نویز مختلف را ایجاد می‌کنند. انواع مختلف نویز، از قبیل نویز سفید، ماشین، قطار و مهممه^۳ در این شبیه‌سازی استفاده شده است. این نویزها از دادگان AURORA [۴۱] گرفته شده‌اند. دادگان نویز با فیلتر فرضی محیط $F(z)$ فیلتر شده و با سیگنال گفتار جمع می‌شود. در این مقاله، از یک فیلتر IIR برای مدل کردن مسیر صوتی میان دو کانال اولیه و مرجع استفاده شده است. فیلتر محیط، $F(z)$ ، بدین صورت فرض می‌شود:

$$F(z) = \frac{0.831 - 0.021z^{-1} + 1.542z^{-2} - 0.249z^{-3}}{1 - 0.436z^{-1} + 2.792z^{-2} - 0.658z^{-3}} \quad (۲۳)$$

در این شبیه‌سازی‌ها، طول فریم K ، ۲۰ میلی‌ثانیه (۱۶۰ نمونه) با میزان ۵۰ درصد همپوشانی است. شرایط آزمایشگاهی برای این الگوریتم‌ها در جدول (۱) آورده شده است.

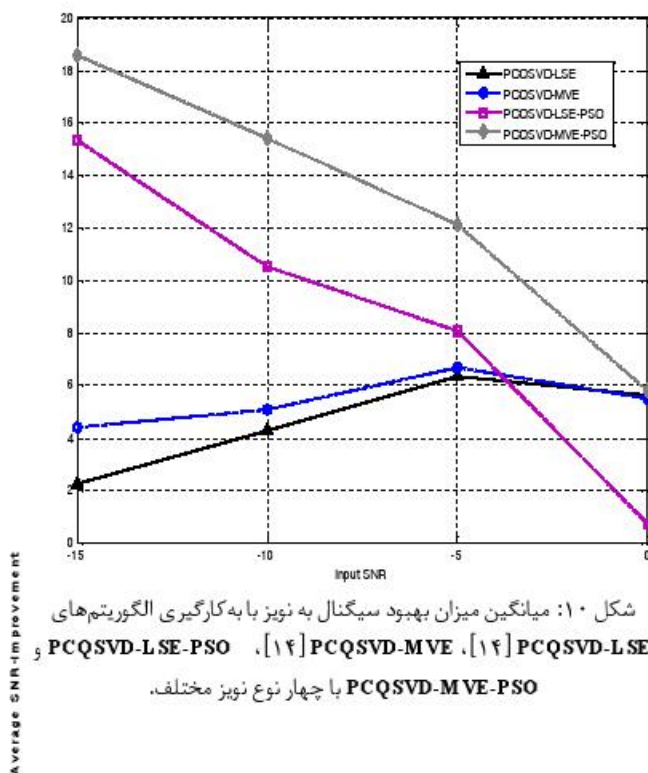
معیارهای ارزیابی به دو نوع objective و subjective تقسیم می‌شوند. کارایی الگوریتم‌های پیشنهادی با دو معیار subjective (شکل‌موج زمانی و روش تست شنوایی

جدول ۲: میزان بهبود سیگنال به نویز با به کارگیری الگوریتم‌های PCQSVD-LSE، PCQSVD-MVE، PCQSVD-MVE-PSO و PCQSVD-LSE-PSO برای سیگنال‌های نویزی با نسبت سیگنال به نویز ۱۰ dB.

الگوریتم‌ها	نویز سفید	نویز ماشین	نویز هواپیما	نویز همه‌مه
PCQSVD-LSE [۱۴]	۷/۱	۳/۷	۳/۲	۳/۱
PCQSVD-MVE [۱۴]	۶/۷	۶/۰	۳/۷	۴/۰
PCQSVD-LSE-PSO	۱۰/۶	۱۰/۶	۱۰/۳	۱۰/۶
PCQSVD-MVE-PSO	۱۴/۷	۱۵/۸	۱۵/۷	۱۵/۵

جدول ۳: میزان بهبود سیگنال به نویز با به کارگیری الگوریتم‌های PCQSVD-LSE، PCQSVD-MVE، PCQSVD-MVE-PSO و PCQSVD-LSE-PSO برای سیگنال‌های نویزی با نسبت سیگنال به نویز ۱۵ dB.

الگوریتم‌ها	نویز سفید	نویز ماشین	نویز هواپیما	نویز همه‌مه
PCQSVD-LSE [۱۴]	۲/۶	۲/۲	۲/۳	۲/۰
PCQSVD-MVE [۱۴]	۵/۴	۵/۲	۳/۴	۳/۷
PCQSVD-LSE-PSO	۱۵/۲	۱۵/۵	۱۵/۲	۱۵/۴
PCQSVD-MVE-PSO	۱۷/۲	۱۹/۱	۱۹/۱	۱۹/۰



تتبع‌ایچ تست‌ت شش‌نویابی با روش Listening Performance Comparison در جدول (۴) آورده شده است. برای مقایسه، روش زیرفضایی ادراکی قدیمی PKLT [۱۵]، که در آن نگاشت فرکانس به حوزه مقادیر ویژه^{۳۱} صورت گرفته، و نیز الگوریتم‌های تک کاناله PCQSVD-LSE و PCQSVD-MVE شبیه‌سازی شده‌اند. دو ستون اول این جدول مربوط به سیگنال‌های نویزی با نویزهای سفید، ماشین، هواپیما و همه‌مه می‌باشند. برای هر نوع نویز، پنج جفت سیگنال صوتی تولید می‌شود. در هر جفت، سیگنال بهبود-کیفیت یافته توسط PCQSVD-MVE-PSO و یکی از سیگنال‌های نویزی با بهبود کیفیت یافته توسط الگوریتم‌های PCQSVD، PKLT، LSE، PCQSVD-MVE و PCQSVD-LSE-PSO وجود دارند. در کل، ۱۵ نفر با محدوده سنی ۲۰ الی ۵۰ سال در این تست شرکت داشته‌اند که ۱۲ نفر از آن‌ها زن و ۳ نفرشان مرد می‌باشند. همان‌طور که نتایج جدول (۳) نشان می‌دهد، الگوریتم پیشنهادی PCQSVD-MVE-PSO برای تقریباً همه انواع نویز خیلی بهتر از الگوریتم‌های دیگر عمل می‌کند. به‌طور متوسط ۸۰ الی ۱۰۰ درصد از افراد، سیگنال بهبود کیفیت یافته توسط الگوریتم پیشنهادی PCQSVD-MVE-PSO را در مقایسه با سیگنال بهبود یافته توسط الگوریتم‌های دیگر ترجیح داده‌اند.

جدول ۱: شرایط آزمایشگاهی

مقادیر	پارامترها	الگوریتم‌ها
۱۲۱	M	PCQSVD-LSE [۱۴]
۴۰	N	
Blackman	نوع پنجره	PCQSVD-MVE [۱۴]
۱۲۱	M	
۴۰	N	
۳	γ	
۱۲۱	M	PCQSVD-LSE-PSO
۴۰	N	
Blackman	نوع پنجره	
۱/۵	c ₁ , c ₂	
۹	D	
۲۰۰	تعداد تکرار	
۳۰	تعداد ذرات ازدحام	
۱۲۱	M	PCQSVD-MVE-PSO
۴۰	N	
Blackman	نوع پنجره	
۳	γ	
۱/۵	c ₁ , c ₂	
۹	D	
۲۰۰	تعداد تکرار	
۳۰	تعداد ذرات ازدحام	

کار انتظار می‌رود، نویز با خطای کم‌تری تخمین زده شده و کارایی سیستم‌های بهبود کیفیت گفتار افزایش یابد.

مراجع

- [1] S. V. Huffel, "Enhanced Resolution Based on Minimum Variance Estimation and Exponential Data Modeling," *Signal Processing*, vol. 33, pp. 333-355, 1993.
- [2] B. T. Lilly and K. K. Paliwal, "Robust Speech Recognition Using Singular Value Decomposition Based Speech Enhancement," in *IEEE Speech and Image Technologies for Computing and Telecommunications*, pp. 257-260, 1997.
- [3] M. Klein and P. Kabal, "Signal Subspace Speech Enhancement with Perceptual Post-filtering," in *IEEE International Conference on Acoustics, Speech, Signal Processing*, pp. 537-540, 2002.
- [4] S. H. Jensen, P. C. Hansen, S. D. Hansen and J. A. Sørensen, "Reduction of Broad-band Noise in Speech by Truncated QSVD," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 6, pp. 439-448, 1995.
- [5] M. Dendrinos, S. Bakamidis and G. Garayannis, "Speech Enhancement from Noise: A Regenerative Approach," *Speech Communication*, vol. 10, pp. 45-57, 1991.
- [6] G. H. Ju and L. S. Lee, "Speech Enhancement Based on Generalized Singular Value Decomposition Approach," in *Proceedings of ICSLP*, pp. 1801-1804, 2002.
- [7] U. Mittal and N. Phamdo, "Signal/Noise KLT-based Approach for Enhancing Speech Degraded by Colored Noise," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 2, pp. 159-167, 2000.
- [8] Y. Hu and P. Loizou, "A Subspace Approach for Enhancing Speech Corrupted by Colored Noise," *IEEE Signal Processing Letters*, vol. 9, no. 7, pp. 204-206, 2002.
- [9] H. Lev-Ari and Y. Ephraim, "Extension of the Signal Subspace Speech Enhancement Approach to Colored Noise," *IEEE Signal Processing Letters*, vol. 10, no. 4, pp. 104-106, 2003.
- [10] A. Rezayee and S. Gazor, "An Adaptive KLT Approach for Speech Enhancement," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 2, pp. 87-95, 2001.
- [11] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd Ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
- [12] C. H. You, S. N. Koh and S. Rahardja, "Subspace Speech Enhancement for Audible Noise Reduction," in *IEEE International Conference on Acoustics, Speech, Signal Processing*, pp. 145-148, 2005.
- [13] E. Zwicker and H. Fastle, *Psychoacoustics*, 2nd Ed. New York: Springer-Verlag, 1999.
- [14] Gwo-Hwa Ju and Lin-Shan Lee, "A Perceptually Constrained GSVD-Based Approach for Enhancing Speech Corrupted by Colored Noise," *IEEE Transactions on Speech and Audio Processing*, vol. 15, no. 1, pp. 119-134, 2007.
- [15] F. Jabloun and B. Champagne, "A Perceptual Signal Subspace Approach for Speech Enhancement in Colored Noise," in *IEEE International Conference on Acoustics, Speech, Signal Processing*, pp. 569-572, 2002.
- [16] S. M. Rozali, M. F. Rahmat and A. R. Husain, "Performance Comparison of Particle Swarm Optimization and Gravitational Search Algorithm to the Design of Controller for Nonlinear System," *Journal of Applied Mathematics*, vol. 2014, pp. 1-9, 2014.
- [17] S. Mirjalili and Hashim, "A New Hybrid PSO-GSA Algorithm for Function Optimization," *International Conference on Computer and Information Application (ICCIA)*, pp. 374-377, 2010.

جدول ۴: درصد ترجیح سیگنال بهبود کیفیت یافته توسط الگوریتم

به سیگنال نویزی و بهبود کیفیت یافته توسط

الگوریتم‌های **PCQSVD-MVE-PSO**، **PCQSVD-LSE**، **PKLT** و **PCQSVD**

LSE-PSO توسط افراد مختلف

الگوریتم‌ها					سیگنال نویزی با SNR = -15 dB
PCQSVD-LSE-PSO	PCQSVD-MVE [۱۴]	PCQSVD-LSE [۱۴]	PKLT [۱۵]	سیگنال نویزی	
۱۰۰	۸۰	۸۶/۶۶	۱۰۰	۱۰۰	سفید
۱۰۰	۸۶/۶۶	۷۳/۳۳	۸۶/۶۶	۱۰۰	ماشین
۱۰۰	۸۶/۶۶	۶۶/۶۶	۹۳/۳۳	۱۰۰	هوایما
۱۰۰	۹۳/۳۳	۹۳/۳۳	۹۳/۳۳	۹۳/۳۳	همه
۱۰۰	۸۶/۶۶	۸۰	۹۳/۳۳	۹۸/۳۳	میانگین

۶- نتیجه‌گیری

در این مقاله، دو الگوریتم جدید دوکاناله بهبود کیفیت گفتار زیرفضایی ادراکی معرفی شده است. این الگوریتم‌های پیشنهادی از تکنیک بهینه‌سازی ازدحام ذرات (PSO) برای تخمین نویز موجود در کانال اولیه استفاده می‌کنند. آزمایش‌های متعددی با چهار نوع نویز و مقادیر نسبت سیگنال به نویز ورودی مختلف انجام شده است. کارایی الگوریتم‌ها، با محاسبه معیار objective نسبت سیگنال به نویز و تست شنوایی subjective ارزیابی شده‌اند. همان‌طور که نتایج شبیه‌سازی نشان می‌دهد، روش‌های پیشنهادی **PCQSVD-LSE-PSO** و **PCQSVD-MVE-PSO** در مقایسه با روش‌های **PCQSVD-LSE** و **PCQSVD-MVE** از نقطه نظر بهبود نسبت سیگنال به نویز و تست شنوایی بسیار کارا تر می‌باشند. البته به دلیل کارایی بهتر تخمین‌گر حداقل واریانس (MVE) نسبت به تخمین‌گر حداقل مربعات (LSE)، عملکرد روش پیشنهادی **PCQSVD-MVE-PSO** بهتر از روش **PCQSVD-LSE-PSO** است. ولی از آنجایی که روش‌های بهبود کیفیت گفتار دوکاناله برای تخمین نویز موجود در سیگنال گفتار نویزی از فرآیندهای تکراری استفاده می‌کنند، این روش‌ها حجم محاسباتی بالاتری نسبت به روش‌های قبلی **PCQSVD-LSE** و **PCQSVD-MVE** دارند.

همچنین، نتایج شبیه‌سازی نشان می‌دهد که تکنیک PSO نویز موجود در سیگنال گفتار نویزی را کارا تر تخمین می‌زند. همان‌طور که مطالعات قبلی نشان می‌دهد، PSO در مقایسه با تکنیک‌های بهینه‌سازی با ساختار آماری گاهی اشتباه در دام نقاط بهینه محلی می‌افتد. این مشکل باعث می‌شود که تکنیک PSO در مسائل بهینه‌سازی پیچیده تر قادر به یافتن پاسخ‌های بهینه کلی نباشد [۴۲]. در آینده و در راستای این مقاله، ما از روش‌های بهینه‌سازی با ساختار آماری بهتری، از جمله GSA، که در یافتن پاسخ‌های بهینه کلی کارا تر باشند، برای تخمین نویز موجود در کانال استفاده خواهیم کرد. با این

- International Symposium on Communications, Control and Signal Processing*, pp. 1-5, 2010.
- [35] S. Ghaemi Sardaroudi and M. Geravanchizadeh, "A Perceptual Subspace Approach for Speech Enhancement," *International Symposium on Telecommunicatios*, pp. I-569-I-572, 2010.
- [36] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, New York: Wiley, 1999.
- [37] J. Sohn and N. Kim, "Statistical Model-based Voice Activity Detection," *IEEE Signal Processing Letters*, vol. 6, no. 1, pp. 1-3, 1999.
- [38] R. Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 504-512, 2001.
- [39] <http://ecs.utdallas.edu/loizou/speech/noizeus/> (available on Nov. 9, 2014)
- [40] <http://www.speech.cs.cmu.edu/comp.speech/Section1/Data/noisex.html> (available on Nov. 9, 2014)
- [41] H. G. Hirsch and D. Pearce, "The AURORA Experimental Framework for the Performance Evaluations of Speech Recognition Systems under Noisy Conditions," *ISCA ITRW ASR2000*, 2000.
- [42] J. H. Holland, *Adaptation in Natural and Artificial Systems*, Ann Arbor, University of Michigan Press, 1975.
- [18] K. Prajna, G. S. B. Rao, K. V. V. S. Reddy and R. Uma Maheswari, "A New Approach to Dual Channel Speech Enhancement Based on Gravitational Search Algorithm (GSA)," *International Journal of Speech Technology*, vol. 17, no. 4, pp. 341-351, 2014.
- [19] K. Prajna, G. S. B. Rao, K. V. V. S. Reddy and R. Uma Maheswari, "A New Approach to Dual Channel Speech Enhancement Based on Hybrid PSO/GSA," *International Journal of Speech Technology*, vol. 18, no. 1, pp. 45-56, 2014.
- [20] J. Kennedy and R.C. Eberhart, "Particle Swarm Optimization," in *Proceedings of IEEE International Conference on Neural Networks*, pp. 39-43, 1995.
- [21] R. C. Eberhart and J. Kennedy, "A New Optimizer Using Particle Swarm Theory," in *Proceedings of Sixth International Symposium on Micromachine and Human Science*, pp. 39-43, 1995.
- [22] Y. Niu and L. Shen, "An Adaptive Multi-objective Particle Swarm Optimization for Color Image Fusion," *Lecture Notes in Computer Science, LNCS*, pp. 473-480, 2006.
- [23] W. Yi, M. Yao and Zh. Jiang, "Fuzzy Particle Swarm Optimization Clustering and its Application to Image Clustering," *Lecture Notes in Computer Science, LNCS*, pp. 459-467, 2006.
- [24] W. Zhang and Y. Liu, "Adaptive Particle Swarm Optimization for Reactive Power and Voltage Control in Power Systems," *Lecture Notes in Computer Science, LNCS*, pp. 449-452, 2006.
- [25] S. Ghaemi Sardaroudi and M. Geravanchizadeh, "Speech Enhancement Using a Perceptually Constrained TQSVD-Based Approach Incorporating the PSO Technique," *5th International Symposium on Telecommunicatios*, pp. 863-868, 2010.
- [26] F. T. Luk, "A Parallel Method for Computing the Generalized Singular Value Decomposition," *Journal of Parallel and Distributed Computing*, vol. 2, no. 3, pp. 250-260, 1985.
- [27] S. Doclo and M. Moonen, "GSVD-based Optimal Filtering for Single and Multimicrophone Speech Enhancement," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230-2244, 2002.
- [28] J. P. Kargo and K. V. Sorensen, *A Rank-independent Signal Subspace Method for Speech Enhancement*, M.Sc. Thesis, Aalborg University, Denmark, 2002.
- [29] C. C. Paige and M. A. Saunders, "Towards a Generalized Singular Value Decomposition," *SIAM Journal on Numerical Analysis*, vol. 18, pp. 398-405, 1981.
- [30] N. Virag, "Single Channel Speech Enhancement Based on Masking Properties of the Human Auditory System," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 2, pp. 126-137, 1999.
- [31] K. Brandenburg and G. Stoll, "ISO-MPEG-1 Audio: A Generic Standard for Coding of High Quality Digital Audio," *Journal of the Audio Engineering Society*, vol. 42, pp. 780-792, 1994.
- [32] J. D. Johnston, "Transform Coding of Audio Signals Using Perceptual Noise Criteria," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 2, pp. 314-323, 1988.
- [33] D. E. Tsoukalas, J. N. Mourjopoulos and G. Kokkinakis, "Speech Enhancement Based on Audible Noise Suppression," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 5, pp. 497-514, 1997.
- [34] M. Geravanchizadeh and S. Ghaemi Sardaroudi, "Incorporating the Human Hearing Properties in the GSVD-Based Approach for Speech Enhancement,"

زیرنویس‌ها

¹ Singular Value Decomposition (SVD)² Quotient Singular Value Decomposition (QSVD)³ Auditory Masking Thresholds (AMT)⁴ Perceptually Constrained QSVD (PCQSVD)⁵ Generalized Singular Value Decomposition (GSVD)⁶ Particle Swarm Optimization (PSO)⁷ Stationary⁸ Non-Stationary⁹ Primary Channel¹⁰ Reference Noise¹¹ Heuristic Optimization Algorithm¹² Gravitational Search Algorithm (GSA)¹³ Agents¹⁴ Controller Design for Tracking Purpose of Nonlinear Systems¹⁵ Stagnation¹⁶ Standard Particle Swarm Optimization (SPSO)¹⁷ Swarm of Particles¹⁸ Single Minimum Error¹⁹ Learning Factor²⁰ Social Learning Factor²¹ Generalized Singular Matrix²² Generalized Singular Values²³ Frobenius Norm²⁴ Transformation Matrix²⁵ Time Domain Constrained (TDC)²⁶ Spectral Domain Constrained (SDC)²⁷ Voice Activity Detection (VAD)²⁸ Minimum Tracking²⁹ Adaptive Noise Cancellation (ANC)³⁰ Babble³¹ Eigen Domain