

Fractional Eigen Based MVDR Beamformer for Speech Enhancement

Seyed Mostafa Alaviyan Shahri, Hamid Reza Abutalebi*

Department of Electrical Engineering, Yazd University, Yazd, Iran.
E-mails: mostafa.alaviyan@stu.yazd.ac.ir; habutalebi@yazd.ac.ir

Short Abstract

One of the most widely used beamforming algorithms for the application of speech enhancement is the Minimum Variance Distortionless Response (MVDR) technique. The optimal coefficients of the MVDR beamformer are calculated based on the incoherence assumption of environmental interferences and the desired signal. Due to the nature of noise and speech signals, this assumption is not valid in many practical situations. This, in turn, results in inaccuracy of derived coefficients of the MVDR. In this paper, as the first change in the MVDR beamformer, by applying the eigenvalue analysis to the desired signal covariance matrix and removing small eigenvalues, the accuracy of the beamformer coefficients is improved. As the second contribution, we use a generalized version of the Short-Time Fourier Transform (STFT), namely the Short-Time Fractional Fourier Transform (STFrFT), to calculate the MVDR beamformer weights. In this research, after obtaining the optimal value of STFrFT parameter experimentally, the effect of each of the above two changes on the performance is investigated and compared with the basic methods. The results show that the proposed methods, while being stable to the changes of parameters and environmental conditions, achieve signal-to-noise ratio (SNR) values between $-0.57dB$ and $-0.22dB$, while the performance of the baseline method is in the range of $[-1.12dB, -0.40dB]$. Although each of the above changes alone improves the performance, it is noted that the superior performance is obtained when both changes are applied together on the beamformer.

Keywords

Speech Enhancement, Microphone Array, Eigen Value Analysis, Short-Time Fractional Fourier Transform.

1- Short Introduction

MVDR beamforming is one of the most applicable and attractive algorithms for multi-channel speech enhancement. MVDR weights are calculated under the assumption that the desired speech signal and the environmental interferences are incoherent. Although this simplifies the calculation of the beamformer weights, the assumption is not established in many practical situations. Furthermore, the optimal MVDR weights are commonly obtained in the STFT domain. The STFrFT, as a generalization of STFT, can be used to reduce the noise more effectively considering its one more degree of freedom.

2- Proposed Work and Methodology

In the MVDR beamformer, the output power is minimized under a single linear constraint on the response of the array in the STFT domain. In speech enhancement, it is usually assumed that the desired speech and the environmental noise are incoherent. Therefore, the covariance matrices of noise and desired signal can be easily calculated and used to determine the optimal MVDR coefficients. However, this assumption is not established since the covariance matrix of the desired source signal may contain some noise. In this paper, we propose two extensions to the standard MVDR beamformer: 1- To enhance the optimal MVDR coefficients, we apply eigenvalue decomposition on the covariance matrix of the desired speech. The eigenvalues are sorted in descending order and the biggest one is retained. Then the covariance matrix is reconstructed. 2- The optimal weights of the MVDR beamformer are calculated in STFrFT domain, which is a generalization of STFT. The parameter of the STFrFT (α) can be tuned so that the transform captures the signal's time-varying characteristics more effectively.

To evaluate the performance, we first determine the optimum value of α (STFrFT parameter) in the simulated acoustic room considering two performance measures, i.e. Perceptual Evaluation of Speech Quality (PESQ), and Segmental Signal to Noise Ratio (SSNR). Based on the obtained performance evaluation metrics, $\alpha = 0.94$ is chosen as the optimum STFrFT parameter. In the following, we consider three different scenarios: 1) noisy, 2) reverberant, and 3) noisy-reverberant environments and compare the performance of the proposed beamformers (fractional MVDR (FracMVDR), eigen-based MVDR (EigenMVDR), and fractional eigen-based MVDR (FracEigenMVDR)) with those of the baseline methods (standard MVDR and RICMWF). The simulation results show that three proposed beamformers can achieve more noise reduction and quality enhancement in comparison with standard MVDR. As an example of the results, the input PESQ is equal to 1.56 and SSNR = $-1.63dB$ in the noisy-reverberant scenario. In this scenario, the FracEigenMVDR can improve the PESQ to 2.18 and the SSNR to -0.57 ; the EigenMVDR improves the PESQ to 2.10 and the SSNR to -0.66 , and the FracMVDR output the PESQ is 1.98 and the output SSNR is $-1.06dB$, while the standard MVDR reaches PESQ = 1.91 and SSNR = $-1.12dB$. Comparing FracMVDR and EigenMVDR, we observe that in the whole proposed structure, the effect of eigenvalue analysis on the noise cancellation is more significant rather than that of applying STFrFT transform. Furthermore, our analysis shows the robustness of the proposed method against the changes in parameters and environmental situations.

In this paper, some waveforms and spectrograms of the proposed and baseline methods have been also presented. The depicted results are consistent with the performance evaluations done in the previous experiments.

3- Conclusion

In this research, we proposed two extensions to the standard MVDR beamformer to reduce the noise more and improve the speech quality of signals received by a microphone array. As the first contribution, we performed an eigenvalue analysis on the covariance matrix to improve the accuracy of the optimal MVDR coefficients. Secondly, the optimal weights of the MVDR beamformer are calculated in STFrFT domain (instead of the standard STFT). Simulated results demonstrate that the proposed FracEigenMVDR beamformer improves the speech quality and reduces the noise more effectively compared to the standard MVDR.

4- References

- 1- J. Shi, J. Zheng, X. Liu, W. Xiang, and Q. Zhang, "Novel Short-Time Fractional Fourier Transform: Theory, Implementation, and Applications," *IEEE Trans. Signal Process.*, vol. 68, pp. 3280–3295, 2020.
- 2- E. A. P. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, "New insights into the MVDR beamformer in room acoustics," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 18, no. 1, pp. 158–170, 2009.
- 3- V. W. Neo, C. Evers, and P. A. Naylor, "Enhancement of noisy reverberant speech using polynomial matrix eigenvalue decomposition," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 3255–3266, 2021.
- 4- D. Ying and Y. Yan, "Robust and fast localization of single speech source using a planar array," *IEEE Signal Process. Lett.*, vol. 20, no. 9, pp. 909–912, 2013.

شکل دهنده پرتو MVDR مرتبه کسری و مبتنی بر آنالیز مقادیر ویژه برای بهسازی گفتار

سید مصطفی علویان شهری

دانشجوی دکتری، دانشکده مهندسی برق، دانشگاه یزد، یزد، ایران

حمیدرضا ابوطالبی

استاد، دانشکده مهندسی برق، دانشگاه یزد، یزد، ایران

چکیده

یکی از پرکاربردترین الگوریتم‌های شکل‌دهی پرتو در بهسازی گفتار، الگوریتم حداقل واریانس بدون اعوجاج است. در محاسبه ضرائب این شکل‌دهنده فرض ناهمبسته بودن نویز و سیگنال مطلوب با توجه به وجود پژواک و نویزهای با طیفی مشابه سیگنال مطلوب برقرار نیست. در نتیجه، ماتریس‌های کواریانس و به تبع آن، ضرائب شکل‌دهنده دارای دقت کافی نخواهد بود. در این مقاله، به‌عنوان اولین تغییر در شکل‌دهنده پرتو حداقل واریانس، با اعمال آنالیز مقادیر ویژه به ماتریس کواریانس سیگنال مطلوب و حذف مقادیر ویژه کوچک، دقت ضرائب شکل‌دهنده بهبود داده می‌شود. در دومین تغییر، از تبدیل فوریه مرتبه کسری به‌عنوان حالت تعمیم‌یافته تبدیل فوریه استاندارد استفاده شده است. در ادامه تغییرات پیشنهادی فرمول‌بندی شده و با بدست آوردن پارامتر بهینه تبدیل فوریه کسری به‌صورت تجربی، اثر هریک از دو تغییر با روشهای پایه مقایسه می‌شود. نتایج نشان می‌دهد که روش‌های پیشنهادی ضمن پایداری نسبت به تغییر پارامترها و شرایط محیطی، در سناریوهای مختلف به‌نسبت توان سیگنال به نویزی بین $-0.57dB$ تا $-0.22dB$ دست می‌یابند. در مقابل، روش پایه عملکردی در محدوده $-1.12dB$ تا $-0.40dB$ دارد. همچنین اگرچه هریک از تغییرات فوق به‌تنهایی نیز منجر به بهبود کارایی می‌شوند، اما بهترین عملکرد در حالتی بدست می‌آید که هر دو تغییر پیشنهادی با هم اعمال شود.

کلمات کلیدی

بهسازی گفتار، آرایه میکروفونی، آنالیز مقادیر ویژه، تبدیل فوریه زمان-کوتاه کسری.

نام نویسنده مسئول: دکتر حمیدرضا ابوطالبی

ایمیل نویسنده مسئول: habutalebi@yazd.ac.ir

تاریخ ارسال مقاله: ۱۴۰۱/۰۶/۰۸

تاریخ(های) اصلاح مقاله: ۱۴۰۱/۰۸/۱۳

تاریخ پذیرش مقاله: ۱۴۰۱/۱۱/۱۱

۱- مقدمه

روش‌های دسته دوم آن است که تنها در صورت یکسان بودن آرایه میکروفونی در مراحل آموزش و آزمایش، عملکرد قابل قبولی از خود نشان می‌دهند. از طرفی در دسته روش‌های اول، نه الگوریتم‌های شکل‌دهی پرتو و نه شبکه آموزش داده شده وابستگی به هندسه آرایه ندارند. از این رو، دسته روش‌های اول بسیار مورد توجه قرار گرفته‌اند. همانطور که بیان شد یکی از اجزای مهم روش‌های دسته اول بهسازی مبتنی بر شبکه‌های عصبی عمیق، الگوریتم‌های شکل‌دهی پرتو هستند. از این رو با وجود کارهای انجام شده بر روی بهسازی گفتار مبتنی بر DNN، کماکان موضوع طراحی شکل‌دهنده‌های مناسب پرتو حائز اهمیت می‌باشد.

شکل‌دهنده پرتو حداقل واریانس با پاسخ بدون اعوجاج ($MVDR^4$) یکی از پرکاربردترین الگوریتم‌های پردازش آرایه‌ای سیگنال گفتار با هدف حذف عوامل مخرب محیطی است. ضرائب این شکل‌دهنده از طریق کمینه کردن توان خروجی شکل‌دهنده همزمان با اعمال یک شرط خطی روی پاسخ آرایه (شرط بهره واحد) در راستای منبع مطلوب بدست می‌آیند. شرط خطی در نظر گرفته شده بدین صورت است که هیچ گونه اعوجاج سیگنالی در راستای منبع مطلوب

در بسیاری از کاربردهای پردازش گفتار نظیر پل صوتی^۱ و کنفرانس تلفنی نیاز به دریافت دست-آزاد^۲ وجود دارد. در این مواقع معمولاً از یک آرایه میکروفونی به منظور دریافت سیگنال گفتار منبع مطلوب استفاده می‌شود. با هدف بدست آوردن سیگنال منبع مطلوب، سیگنال‌های دریافتی پردازش و با حذف یا کاهش اثر عوامل مخرب محیطی، سیگنال‌های دریافتی تا حد ممکن به سیگنال منبع مطلوب شبیه خواهد شد [۱-۳].

در سال‌های اخیر، شبکه‌های عصبی عمیق (DNN^3) نیز در زمینه بهسازی چندکاناله گفتار مورد استفاده قرار گرفته‌اند [۴،۱]. الگوریتم‌های بهسازی چندکاناله مبتنی بر DNN را می‌توان به دو دسته کلی تقسیم کرد: ۱- روش‌هایی که ابتدا با استفاده از یک DNN طیف سیگنال مطلوب و عوامل مخرب محیطی را بدست آورده و سپس با اعمال یک شکل‌دهنده پرتو، نویز دریافت شده همراه سیگنال مطلوب را حذف می‌کنند. ۲- روش‌هایی که در آنها شبکه عصبی عمیق مستقیماً از روی سیگنال دریافتی از آرایه آموزش دیده و سپس از طریق این شبکه، حذف نویز به انجام می‌رسد [۲]. یک محدودیت عمده

^۴ Minimum Variance Distortionless Response

^۱ Audio bridging

^۲ Hands-free

^۳ Deep Neural Network

ایجاد نشود [۵].

محققین شکل‌دهنده‌های پرتو مختلفی از جمله $LCMV^5$ ، GSC^6 ، $AMNOR^7$ و ... را با اضافه کردن قیود خطی متفاوت به مسئله بهینه‌سازی در نظر گرفته شده برای $MVDR$ پیشنهاد داده‌اند. یکی از این شکل‌دهنده‌ها تحت عنوان شکل‌دهنده حداقل واریانس با قیود خطی ($LCMV$) شناخته می‌شود [۵]. شکل‌دهنده پرتو $MVDR$ را می‌توان یک حالت خاص از شکل‌دهنده‌ی $LCMV$ دانست.

در [۶] با اعمال یک نگاهت به توزیع احتمال سیگنال ورودی، و سپس استفاده از داده‌های نگاهت‌یافته برای بدست آوردن ضرائب شکل‌دهنده $MVDR$ ، شکل‌دهنده جدیدی با نام $Measured-Transform MVDR$ به منظور مقابله با نویز ضربه‌ای ارائه شده است.

همان‌طور که بیان شد، بسیاری از الگوریتم‌های شکل‌دهنده پرتو به نوعی با اضافه کردن قیود و یا تعمیم شکل‌دهنده پرتو $MVDR$ مبادرت به بهبود عملکرد حذف نویز و عوامل مخرب محیطی از سیگنال‌های رسیده به آرایه کرده‌اند [۵]. اگرچه شکل‌دهنده پرتو $MVDR$ توانسته توجه زیادی را مخصوصاً در زمینه پردازش آرایه‌ای سیگنال‌های صوتی به خود جلب کند، اما همچنان ظرفیت افزایش کارایی و بهبود عملکرد مشتقات آن وجود دارد. یکی از مفروضاتی که معمولاً در بدست آوردن ضرائب شکل‌دهنده پرتو $MVDR$ استفاده می‌شود، فرض ناهمبسته بودن سیگنال مطلوب و نویز محیطی است [۹-۷]. این فرض اگرچه باعث ساده شدن محاسبات می‌شود، اما در عمل با توجه به وجود عوامل مخرب محیطی نظیر پژواک و نویزهای با طیفی مشابه سیگنال مطلوب ممکن است برقرار نباشد [۹-۷]؛ در نتیجه ماتریس کواریانس سیگنال مطلوب که در محاسبه ضرائب شکل‌دهنده پرتو $MVDR$ مورد استفاده قرار می‌گیرد، از دقت کافی برخوردار نخواهد بود.

در مقاله حاضر به منظور حل این مشکل، از آنالیز مقادیر ویژه برای افزایش دقت ماتریس کواریانس سیگنال مطلوب استفاده شده است. با فرض اینکه در مسئله بهسازی گفتار یک منبع سیگنال مطلوب وجود دارد، می‌توان بزرگترین مقدار ویژه ماتریس کواریانس را مرتبط با منبع مطلوب در نظر گرفت. در نتیجه با برابر با صفر قرار دادن دیگر مقادیر ویژه می‌توان دقت ماتریس کواریانس سیگنال مطلوب را بهبود بخشید. در [۷] با فرض ناهمبسته بودن سیگنال منبع مطلوب و عوامل مخرب محیطی، از آنالیز مقادیر ویژه ماتریس کواریانس سیگنال دریافتی استفاده شده و الگوریتم مکان‌یابی گفتار مقاوم در مقابل عوامل مخرب پیشنهاد شده است. همچنین [۸] برای حذف پژواک، با استفاده از فرض ناهمبسته بودن سیگنال مطلوب و پژواک دیر (late reverberation)، ابتدا ماتریسی تحت عنوان ماتریس چندجمله‌ای که نشان‌دهنده همبستگی زمانی، محلی و فرکانسی سیگنال‌های رسیده به آرایه میکروفونی است را بدست می‌آورد. سپس با اعمال تجزیه مقادیر ویژه به ماتریس مذکور اطلاعات زمانی، محلی و فرکانسی سیگنال‌های رسیده به میکروفون‌ها را تجزیه کرده و بر اساس آن مبادرت به حذف پژواک نموده است. سپس در [۹] با ایده و فرضی مشابه، علاوه بر نویز جمع‌شونده، پژواک دیر نیز ناهمبسته با سیگنال منبع مطلوب فرض شده و با تجزیه مقادیر ویژه ماتریس چندجمله‌ای مبادرت به حذف توأم نویز و پژواک گردیده است.

در حالت کلی، در پردازش آرایه‌ای سیگنال گفتار برای بدست آوردن ضرائب شکل‌دهنده پرتو از تبدیل فوریه زمان کوتاه ($STFT^8$) استفاده می‌شود. به عنوان دومین تغییر و نوآوری تحقیق حاضر برای بهبود عملکرد $MVDR$ ، به جای محاسبه ضرائب در حوزه $STFT$ از تبدیل فوریه مرتبه کسری زمان کوتاه

($STFrFT^9$) استفاده کرده‌ایم. این تبدیل با توجه به این که دارای یک درجه آزادی است، امکان پردازش بهتر مشخصات زمان-فرکانس سیگنال‌های غیرایستانت نظیر گفتار را فراهم می‌آورد [۱۰-۱۲]. به عنوان مثال، در [۱۳] به منظور تخمین زاویه ورود سیگنال در محیط زیر آب، با استفاده از تبدیل فوریه مرتبه کسری چندین شکل‌دهنده پرتو پیشنهاد شده است. نتایج حاصل از تست و ارزیابی‌های انجام شده نشان می‌دهد که بیم‌ترین بدست آمده در پردازش آرایه‌ای مبتنی بر تبدیل فوریه کسری علاوه بر دارا بودن لوب‌های اصلی باریکتر، نسبت لوب اصلی به لوب‌های کناری بالاتری ویژه در محیط‌های با شدت نویز بالا دارد.

در این تحقیق، تغییراتی به شکل‌دهنده پرتو $MVDR$ اعمال و سیستمی بهبودیافته برای کاربرد حذف نویز و پژواک از سیگنال گفتار ارائه می‌شود. به طور خلاصه، هدف این مقاله، پیشنهاد و بررسی فرم جدیدی از شکل‌دهنده پرتو $MVDR$ از طریق اقدامات زیر می‌باشد:

- ۱- اعمال آنالیز مقادیر ویژه به منظور به دست آوردن ماتریس کواریانس سیگنال مطلوب با دقت بالا
- ۲- استفاده از تبدیل $STFrFT$ به جای $STFT$ استاندارد در محاسبه ضرائب شکل‌دهنده پرتو $MVDR$
- ۳- مقایسه عملکرد شکل‌دهنده پرتو پیشنهادی در حضور توأم نویز و پژواک به عنوان دو عامل اصلی مخرب محیطی
- ۴- تحلیل پایداری روش پیشنهادی در مقابل تغییر پارامترهای مختلف از جمله: تغییر در راستای ورود سیگنال، تغییر در فاصله بین منبع و آرایه، تغییر در شرایط نویزی محیط، و تغییر در پارامتر تبدیل فوریه زمان کوتاه کسری. ادامه این مقاله شامل بخش‌های زیر می‌باشد. در بخش دوم، مدل سیگنال در محیط نویزی-پژواکی و روابط کلی حاکم بر شکل‌دهی پرتو بررسی و مسئله فرمول بندی خواهد شد. در بخش سوم، به ترتیب تعمیم آنالیز مقادیر ویژه و تبدیل $STFrFT$ ارائه و همچنین نحوه اعمال و فرمول‌بندی هر یک نیز بررسی می‌شود. در ادامه، نتایج شبیه‌سازی، مقایسه و بررسی پایداری روش‌های پیشنهادی در بخش چهارم گزارش خواهد شد. در نهایت در بخش پنجم، جمع‌بندی و نتیجه‌گیری کلی از مقاله به عمل خواهد آمد.

۲- مدل سیگنال و شرح مسئله

در مسئله پردازش آرایه‌ای سیگنال گفتار، مدل سیگنال در نظر گرفته شده شامل یک آرایه میکروفونی با N حسگر است. هر یک از حسگرها (میکروفون‌ها) سیگنال مطلوب کانال‌لوشده در یک میدان نویزی را دریافت می‌کند. سیگنال‌های دریافتی توسط هر یک از میکروفون‌ها به بیان ریاضی به فرم زیر هستند [۱-۲]:

$$\begin{aligned} y_n(k) &= a_n(k) * s(k) + v_n(k) \\ &= x_n(k) + v_n(k) \end{aligned} \quad (1)$$

که در آن $s(k)$ سیگنال منبع مطلوب و $a_n(k) = r_n \delta(k - \psi_n)$ پاسخ ضربه کانال بین منبع مطلوب و n امین میکروفون است. r_n و ψ_n نیز به ترتیب بیان‌کننده بهره تضعیف و زمان انتشار کانال می‌باشد. همچنین مولفه مسیر مستقیم به صورت $x_n(k) = a_n(k) * s(k)$ تعریف می‌شود. مولفه همه‌جهته $v_n(k)$ نیز ترکیبی از نویز جمعی و پژواک است که از جهت‌های مختلف به هر یک از میکروفون‌ها می‌رسند. معمولاً فرض می‌شود که $x_n(k)$ و $v_n(k)$

⁷ Adaptive Microphone array NOise Reduction

⁸ Short-Time Fourier Transform

⁹ Short Time Fractional Fourier Transform

⁵ Linear Constraints Minimum Variance

⁶ Generalized Sidelobe Canceler

$$\phi_z(k, f) = \mathbf{h}^H(k, f) \phi_x(k, f) \mathbf{h}(k, f) + \mathbf{h}^H(k, f) \phi_v(k, f) \mathbf{h}(k, f) \quad (6)$$

که در آن

$$\phi_x(k, f) = E[\mathbf{x}(k, f) \mathbf{x}^H(k, f)] \quad (7)$$

$$\phi_y(k, f) \mathbf{a}(f) \mathbf{a}^H(f)$$

ماتریس PSD سیگنال مسیر مستقیم است. در اینجا $E(\cdot)$ بیانگر امید ریاضی است و

$$\phi_v(k, f) = E[\mathbf{v}(k, f) \mathbf{v}^H(k, f)] \quad (8)$$

ماتریس PSD مولفه همه‌جهته (شامل انواع نویز و پژواک) است [۷، ۵].

یکی از پرکاربردترین شکل‌دهنده‌های پرتو مورد استفاده در زمینه بهسازی چندکاناله گفتار، شکل‌دهنده پرتو MVDR است که ضرائب این شکل‌دهنده پرتو از حل مسئله بهینه‌سازی مقید زیر به دست می‌آید [۱]:

$$\mathbf{h}_{MVDR} = \arg \min_{\mathbf{h}(k, f)} \mathbf{h}^H(k, f) \phi_v(k, f) \mathbf{h}(k, f) \quad (9)$$

$$\text{subject to } \mathbf{h}^H(k, f) \frac{E[X_1(k, f) \mathbf{x}^*(k, f)]}{E[X_1(k, f)]} = 1$$

که در آن بالانویس * بیان‌کننده‌ی مزدوج مختلط است. از حل این مسئله بهینه‌سازی مقید داریم [۱]:

$$\mathbf{h}_{MVDR}(k, f) = \frac{\phi_v^{-1}(k, f) \phi_x(k, f)}{\text{trace}(\phi_v^{-1}(k, f) \phi_x(k, f))} \mathbf{i}_{N,1} \quad (10)$$

که در آن $\mathbf{i}_{N,1}$ برابر با ستون اول ماتریس همبستگی \mathbf{I}_N (با ابعاد $N \times N$) است. حال با استفاده از فرض ناهمبسته بودن مولفه‌های همه‌جهته $\phi_v(k, m)$ و $\phi_x(k, m)$ مولفه مسیر مستقیم داریم [۱]:

$$\mathbf{x} = \mathbf{y} - \mathbf{v} \Rightarrow \phi_x(k, m) = \phi_y(k, m) - \phi_v(k, m) \quad (11)$$

با استفاده از این رابطه می‌توان PSD مولفه مسیر مستقیم را محاسبه و با جایگذاری در (۱۰)، ضرائب بهینه شکل‌دهنده پرتو MVDR را بدست آورد [۱].

۳- روش پیشنهادی

در این بخش، روش پیشنهادی این تحقیق که خود شامل اعمال دو تغییر در شکل‌دهنده پرتو MVDR استاندارد است، ارائه و فرمول‌بندی می‌شود.

تغییر اول: آنالیز مقادیر ویژه

در بخش قبل فرض بر این بود که \mathbf{x} و \mathbf{v} ناهمبسته هستند. در نتیجه می‌توان با استفاده از رابطه (۱۱) مولفه مسیر مستقیم را بدست آورد. این فرض اگرچه باعث ساده شدن روابط می‌شود، اما در کاربردهای عملی و سناریوهای مختلف ممکن است برقرار نباشد. به عبارت دیگر ممکن است مولفه‌ی مسیر مستقیم با مقداری نویز محیطی جمع شده باشد. در نتیجه PSD به دست آمده با این فرض، و به تبع آن ضرائب شکل‌دهنده حاصل از رابطه (۱۰)، دارای دقت کافی نخواهد بود. همچنین می‌توان این گونه تعبیر کرد که با استفاده از $\mathbf{x} = \mathbf{y} - \mathbf{v}$ زیرفضای نویز و سیگنال به طور کامل از هم جدا نخواهند شد. برای رفع این ایراد، از آنالیز مقادیر ویژه بهره گرفته و مقادیر ویژه ماتریس $\phi_x(k, m)$ را به

ناهمبسته‌اند و داریم [۷، ۱]:

$$Y_n(k, f) = A_n(f) S(k, f) + V_n(k, f) \quad (2)$$

$$= X_n(k, f) + V_n(k, f)$$

که در آن $X_n(k, f), S(k, f), A_n(f) = r_n e^{-j2\pi f \psi_n}, Y_n(k, f)$ و سیگنال‌های $v_n(k), x_n(k), s(k), a_n(k), y_n(k)$ هستند. در این حالت، ممکن است چنین به نظر رسد که طنین‌های موجود در مولفه همه‌جهته با مولفه‌ی مسیر مستقیم دارای همبستگی است. اما چنانچه فرض میدان-دور^{۱۰} در نظر گرفته شود، این همبستگی قابل صرف نظر کردن خواهد بود. به منظور نمایش بهتر و راحت‌تر، سیگنال‌های دریافتی توسط آرایه میکروفونی به صورت برداری بیان می‌شوند؛ داریم [۷، ۲-۱]:

$$\mathbf{y}(k, f) = \mathbf{a}(f) s(k, f) + \mathbf{v}(k, f) \quad (3)$$

که در آن

$$\mathbf{y}(k, f) = [Y_1(k, f) \ Y_2(k, f) \ \dots \ Y_N(k, f)]^T$$

$$\mathbf{x}(k, f) = [X_1(k, f) \ X_2(k, f) \ \dots \ X_N(k, f)]^T$$

$$= S(k, f) [A_1(f) \ A_2(f) \ \dots \ A_N(f)]^T$$

$$= \mathbf{a}(f) S(k, f)$$

$$\mathbf{v}(k, f) = [v_1(k, f) \ v_2(k, f) \ \dots \ v_N(k, f)]^T$$

و بالانویس T نشانگر ترانزپوز یک ماتریس یا بردار است. بر اساس فرض ناهمبسته بودن مولفه‌ی مسیر مستقیم $x_n(k)$ و مولفه‌ی همه‌جهته‌ی $v_n(k)$ و با استفاده از چگالی طیف توان (PSD^{۱۱}) داریم [۱، ۶]:

$$\phi_{y_n}(k, f) = \phi_{x_n}(k, f) + \phi_{v_n}(k, f) \quad (4)$$

$$= |A_n(f)|^2 \phi_s(k, f) + \phi_{v_n}(k, f)$$

$$n=1, 2, \dots, N$$

که در آن $\phi_{x_n}(k, f)$ و $\phi_{v_n}(k, f)$ به ترتیب PSDهای n امین سیگنال ورودی، n امین سیگنال مسیر مستقیم، سیگنال منبع مطلوب و n امین مولفه‌ی همه‌جهته هستند.

معمولاً در بهسازی چندکاناله سیگنال گفتار، به منظور حذف عوامل مخرب محیطی (بطور کلی نویز) از الگوریتم‌های شکل‌دهی پرتو استفاده می‌شود. شکل‌دهی پرتو (فیلترینگ فضایی) عبارت است از اعمال یک سری وزن‌های مختلط به هر میکروفون و جمع مقادیر به دست آمده با یکدیگر، در این صورت داریم [۵]:

$$\mathbf{z}(k, f) = \mathbf{h}^H(k, f) \mathbf{y}(k, f) \quad (5)$$

$$= \mathbf{h}^H(k, f) [\mathbf{A}(f) s(k, f) + \mathbf{v}(k, f)]$$

که در آن $\mathbf{z}(k, f)$ خروجی شکل‌دهنده پرتو و

$$\mathbf{h}(k, f) = [H_1(k, f) \ H_2(k, f) \ \dots \ H_N(k, f)]^T$$

بردار وزن شکل‌دهنده پرتو است.

PSD خروجی شکل‌دهنده پرتو به صورت زیر بدست می‌آید:

^{۱۱} Power Spectral Density

^{۱۰} Far-Field

صورت زیر به دست می آوریم [۷]:

$$\phi_x(k, m) = U(k, f) \Lambda U^H(k, f) \quad (12)$$

که در آن $U(k, f) = [u_1(k, f), \dots, u_N(k, f)]$ ماتریس بردارهای ویژه و عناصر قطری $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$ مقادیر ویژه ماتریس $\phi_x(k, m)$ است. با توجه به اینکه در مسئله بهسازی گفتار معمولاً یک منبع مطلوب داریم، می توان انتظار داشت که مقدار ویژه متناظر با مولفه مسیر مستقیم شدت و مقدار بیشتری نسبت به مولفه های هم جهته داشته باشد. به بیان ریاضی، چنانچه فرض کنیم نویز موجود در ماتریس کواریانس بصورت فضایی سفید است (یعنی $\mathbf{v}(k, f) = \eta \mathbf{I}_N$ که \mathbf{I}_N ماتریس همانی است)، در این صورت $\eta \gg \frac{|a(k, f)|^2 S(k, f)}{N}$ در نتیجه، اگر مقادیر ویژه بدست آمده بصورت نزولی مرتب شده باشند، بردار ویژه متناظر با مولفه مسیر مستقیم به صورت زیر تخمین زده می شود [۷]:

$$\begin{aligned} u_1(k, f) &\approx a(k, f) / \|a(k, f)\| \\ u_{1n}(k, f) &\approx e^{-j2\pi f \psi_n} \end{aligned} \quad (13)$$

که در آن u_{1n} برابر n امین مولفه ی $u_1(k, f)$ است. رابطه ی (۱۳) بیان کننده این مهم است که زیرفضای سیگنال توسط بزرگترین بردار ویژه افزای شده و نویز و پژواک باقی مانده در $\phi_x(k, m)$ شامل بردارهای ویژه دیگر می شود. در نتیجه با برابر با صفر قرار دادن بردارهای ویژه کوچک و بازسازی ماتریس $\phi_x(k, m)$ می توان نویز موجود در آن را به حداقل رساند. در اینجا آنالیز مقادیر ویژه به نحوی نقش محاسبه $\phi_x(k, m)$ با دقت بسیار بالاتر را ایفا می کند. در نهایت می توان رابطه ضرائب MVDR را که به واسطه آنالیز مقادیر ویژه تعمیم یافته، به فرم زیر بازنویسی کرد:

$$\begin{aligned} h(k, f) &= \frac{\phi_v^{-1}(k, f) \tilde{\phi}_x(k, m)}{\text{trace}(\phi_v^{-1}(k, f) \tilde{\phi}_x(k, m))} i_{N,1} \\ \phi_x(k, m) &= \phi_y(k, m) - \phi_\psi(k, m) \\ \phi_x(k, m) &= U(k, f) \Lambda U^H(k, f) \\ \lambda_{\max} &= \max_{1 \leq i < N} (\Lambda_{ii}) \\ \tilde{\Lambda} &= \text{diag}(\lambda_{\max}, 0, \dots, 0) \\ \tilde{\phi}_x(k, m) &= U(k, f) \tilde{\Lambda} U^H(k, f) \end{aligned} \quad (14)$$

تغییر دوم: استفاده از تبدیل فوری مرتبه کسری

تبدیل فوری استاندارد (مرتبه صحیح)، سیگنال را از حوزه زمان به حوزه فرکانس انتقال می دهد. با این کار، ویژگی های هریک از مؤلفه های فرکانسی موجود در سیگنال به دست می آید. اگرچه تبدیل فوری در بسیاری از زمینه های پردازش سیگنال و مهندسی مورد استفاده قرار می گیرد، اما این تبدیل، لزوماً تفکیک پذیری مناسبی بین سیگنال و نویزهای محیطی ایجاد ننموده و نیز اطلاعاتی در مورد مشخصات محلی زمان-فرکانس سیگنال در اختیار قرار نمی دهد [۱۰-۱۲]. مشخصات زمان-فرکانس در پردازش و آنالیز سیگنال های غیرایستنا نظیر گفتار بسیار مهم هستند. یکی از تبدیل های زمان-فرکانسی که به منظور آنالیز و پردازش سیگنال های غیرایستنا مورد استفاده قرار می گیرد، تبدیل فوری مرتبه کسری است. این تبدیل حالت تعمیم یافته تبدیل فوری استاندارد (مرتبه صحیح) است.

برای درک بهتر ارتباط بین دو تبدیل $FT^{1/2}$ و $FrFT^{1/2}$ می توان این توضیح را ارائه داد: چنانچه به سیگنال نوعی $f(t)$ دوبار تبدیل فوری استاندارد اعمال شود، یعنی $FF(f(t))$ ، سیگنال $f(-t)$ به دست می آید. همچنین اگر چهار بار از سیگنال $f(t)$ به طور متوالی تبدیل فوری گرفته شود $FFFF(f(t))$ ، سیگنال $f(t)$ به دست خواهد آمد. تعداد دفعات اعمال تبدیل فوری به سیگنال را می توان به صورت F_α نشان داد که در اینجا α یک عدد صحیح و بیانگر تعداد دفعات اعمال تبدیل است. حال چنانچه مقدار α به جای مقدار صحیح، یک مقدار غیر صحیح (کسری) انتخاب شود، تبدیل به دست آمده تبدیل فوری مرتبه کسری خواهد بود. رابطه آنالیز این تبدیل به صورت زیر است [۱۰]:

$$F_\alpha(\omega) = F^\alpha \{f(t)\}(\omega) = \int_R f(t) K_\alpha(t, \omega) dt \quad (15)$$

با کرنل تبدیل [۱۰]:

$$K_\alpha(\omega, t) = \begin{cases} A_\alpha e^{j \frac{\omega^2 + t^2}{2} \cot \alpha - j f \csc \alpha}, & \alpha \neq m\pi \\ \delta(\omega - f), & \alpha = 2m\pi \\ \delta(\omega + f), & \alpha = (2m-1)\pi \end{cases} \quad (16)$$

که در آن $A_\alpha = \sqrt{(1-j \cot \alpha) / 2\pi}$ و $m \in \mathbb{Z}$ است. رابطه سنتز FrFT زاویه α برابر است با محاسبه ی FrFT با زاویه $-\alpha$ ؛ داریم [۱۰]:

$$f(t) = F^{-\alpha} \{F_\alpha(\omega)\}(t) = \int_R F_\alpha(\omega) K_\alpha^*(\omega, t) d\omega \quad (17)$$

که در آن بالانویس * بیان کننده مزدوج مختلط است.

شایان ذکر است چنانچه $\alpha = \frac{\pi}{2}$ در نظر گرفته شود، FrFT به تبدیل فوری استاندارد ساده خواهد شد. به تعبیر دیگر، اعمال فیلتر در α امین مرتبه عملگر تبدیل فوری کسری معادل با چرخش در صفحه زمان-فرکانس در خلاف جهت عقربه های ساعت است. در نتیجه، باید به دنبال یک α (حوزه فوری کسری) بود که در آن، نسبت به تبدیل فوری استاندارد، سیگنال و نویز با یکدیگر همپوشانی کمتری داشته باشند. بنابراین می توان با اعمال تبدیل فوری کسری و با زاویه چرخش مناسب، نویز را در حوزه فوری کسری در مقایسه با حوزه فوری استاندارد، به نحو مؤثرتری فیلتر نمود.

با این همه، تبدیل (FrFT) نیز مانند FT استاندارد، اطلاعاتی در مورد مشخصات محلی زمان-فرکانس سیگنال در اختیار قرار نمی دهد. همانطور که بیان شد مشخصات محلی زمان-فرکانس در پردازش سیگنال های غیرایستنا نظیر گفتار بسیار حائز اهمیت است. راه حل این مسئله، آنالیز همزمان زمان-فرکانس کسری است؛ بدین صورت که آنالیز سیگنال به طور همزمان در هر دو حوزه زمان و فرکانس انجام شود. این تبدیل به صورت زیر معرفی می شود [۱۰]:

$$STFrFT_\alpha^\omega(t, \omega) = \int f(\tau) g^*(\tau - t) K_{\alpha(\omega)}(\omega, \tau) d\tau \quad (18)$$

که در آن $f(t) \in L^2(\cdot)$ و $g(t)$ تابع پنجره است. هسته تبدیل $(K_{\alpha(\omega)}(\omega, \tau))$ همان هسته تبدیل FrFT است.

حال به منظور تعمیم شکل دهنده پرتو MVDR و بهره مندی از مزایای تبدیل فوری کسری، می توان برای به دست آوردن ضرائب شکل دهنده MVDR به جای استفاده از STFT استاندارد از STFrFT استفاده نمود. در اینجا سیگنال های دریافتی توسط هر یک از میکروفون ها به α امین حوزه تبدیل فوری منتقل می شود؛ سپس ضرائب شکل دهنده پرتو MVDR در این حوزه محاسبه شده و به

^{۱۱} Fractional Fourier Transform

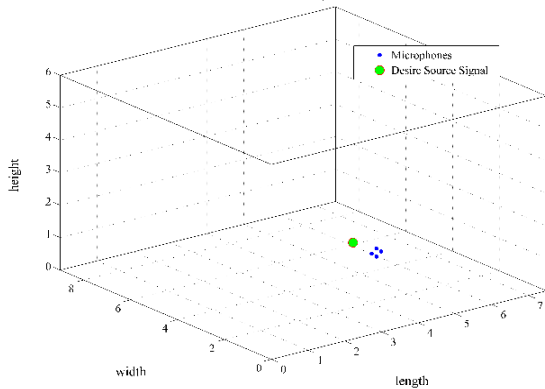
^{۱۲} Fourier Transform

در زیربخش‌های ۵-۴ تا ۷-۴ به بررسی عملکرد و پایداری شکل‌دهنده‌های پیشنهادی پرتو در مقایسه با روش‌های پایه می‌پردازیم.

۴-۱- شرایط اتاق شبیه‌سازی شده

سیگنال رسیده به میکروفون‌ها از حاصل جمع نویز پخشنده (Diffuse)، نویز همهمه (Babble)، نویز غیرهمدوس (نویز سفیدی که توسط میکروفون‌ها به سیگنال دریافتی اضافه می‌شود) و کانولوشن سیگنال منبع مطلوب در پاسخ ضربه اتاقی با ابعاد $7.5 \times 9 \times 6m$ بدست می‌آید. پاسخ ضربه اتاق با استفاده از روش تصویر [۱۴] بدست می‌آید. روش تصویر یکی از متداول‌ترین روش‌هایی است که برای مدل‌سازی پاسخ ضربه اتاق (کانال صوتی بین منبع و گیرنده) استفاده می‌شود. این روش از حل معادله موج و معادل حوزه فرکانس آن (معادله هلمهولتز) که نحوه انتشار امواج صوتی را در سیالات توصیف می‌کند، به‌دست می‌آید. در این تحقیق، از کد پیاده‌سازی شده روش تصویر [۱۴] استفاده شده که با دریافت پارامترهایی نظیر ابعاد اتاق، موقعیت هر یک از میکروفون‌ها، مکان منبع صوت، فرکانس نمونه‌برداری، نوع میکروفون مورد استفاده و همچنین سرعت صوت یک پاسخ ضربه محدود (FIR) برای اتاق بدست می‌آورد. آرایه مورد استفاده نیز یک آرایه میکروفونی دایروی با چهار میکروفون همه جهته ($N=4$) فرض شده است. در این آرایه میکروفون‌ها روی یک دایره به شعاع $25cm$ قرار گرفته‌اند. منبع مطلوب نیز در فاصله یک متری از مرکز آرایه قرار دارد. شکل ۱ شمای کلی اتاق در نظر گرفته شده را نشان می‌دهد. داده‌های گفتاری مورد استفاده از دادگان TIMIT [۱۵] انتخاب شده‌اند. فرکانس نمونه برداری $16kHz$ ، پنجره مورد استفاده از نوع همینگ و با طول $32ms$ و همپوشانی 50% است. همچنین در محاسبه‌ی ماتریس کواریانس سیگنال‌های دریافتی و نویز ضریب فراموشی 99% در نظر گرفته شده است.

Reverberant and Noisy Environment



شکل ۱: شمای کلی اتاق شبیه‌سازی شده

۴-۲- پارامترهای ارزیابی عملکرد

با توجه به اینکه روش پیشنهادی به منظور حذف نویز و پژواک از سیگنال‌های دریافتی توسط آرایه استفاده می‌شود، معیارهای ارزیابی عملکرد نیز باید متناسب با این هدف انتخاب شوند. دو پارامتر مورد استفاده در این زمینه عبارتند از:

- ۱- PESQ^{۱۴}: این پارامتر یکی از پارامترهای متداول برای بررسی میزان بهبود کیفیت شنیداری سیگنال گفتار است. مقدار این پارامتر بین 1-5 متغیر است. هرچه مقدار این پارامتر بیشتر باشد، کیفیت شنیداری گفتار نیز بهتر است [۱۶].
- ۲- SSNR^{۱۵}: در این پارامتر به جای محاسبه‌ی نسبت توان سیگنال به

سیگنال میکروفون‌ها اعمال می‌شود. در نهایت سیگنال بهبودیافته با استفاده از معکوس STFrFT به حوزه زمان منتقل خواهد شد. به‌طور خلاصه داریم:

$$z(t) = F^{-\alpha} \left\{ h^H \left(F^\alpha \{ y(t) \} \right) \right\} \quad (19)$$

که در آن $F^\alpha \{ \cdot \}$ بیان‌کننده تبدیل α امین مرتبه است. با توجه به تغییر ایجاد شده در ساختار شکل‌دهنده پرتو، ضرائب بهینه شکل‌دهنده MVDR در حوزه STFrFT باید دوباره محاسبه شود. با بازنویسی مسئله بهینه‌سازی (۹) داریم:

$$h_{EMVDR} = \arg \min_{h_\alpha(k,f)} h_\alpha^H(k,f) \phi_{v,\alpha}(k,m) h_\alpha(k,f) \quad (20)$$

$$\text{subject to } h_\alpha^H(k,f) \frac{E \left[X_{1,\alpha}(k,f) x_\alpha^*(k,f) \right]}{E \left[X_{1,\alpha}(k,f) \right]} = 1$$

با حل این مسئله بهینه‌سازی مقید بدست می‌آید:

$$h_\alpha(k,f) = \frac{\phi_{v,\alpha}^{-1}(k,f) \phi_{x,\alpha}(k,f)}{\text{trace} \left(\phi_{v,\alpha}^{-1}(k,f) \phi_{x,\alpha}(k,f) \right)} i_{N,1} \quad (21)$$

که در آن $\phi_{v,\alpha}(k,f)$ و $\phi_{x,\alpha}(k,f)$ به ترتیب ماتریس کواریانس‌های α امین مرتبه مولفه‌ی مسیر مستقیم و مولفه غیرجهتی (نویز و پژواک) هستند. این ماتریس‌ها را می‌توان در حوزه STFrFT به فرم زیر به دست آورد:

$$\phi_{x,\alpha}(k,f) = E \left\{ x_\alpha^H(k,f) x_\alpha(k,f) \right\} \quad (22)$$

و

$$\phi_{v,\alpha}(k,f) = E \left\{ v_\alpha^H(k,f) v_\alpha(k,f) \right\} \quad (23)$$

با استفاده از روابط (۲۱)، (۲۲) و (۲۳) می‌توان ضرائب شکل‌دهنده MVDR را در یک حوزه خاص از تبدیل فوریه زمان-کوتاه کسری به دست آورد. حال باید به این سوال پاسخ داد که مقدار بهینه برای α چقدر است. شایان ذکر است که مرتبه بهینه STFrFT (α بهینه) را نمی‌توان به صورت فرم بسته محاسبه نمود؛ در نتیجه برای بدست آوردن آن باید به ازای مقادیر مختلف α ، ضرائب MVDR را به دست آورد و هر کدام را که منجر به عملکرد بهتری در حذف نویز و پژواک می‌شود، به عنوان بهینه انتخاب نمود. با توجه به اینکه STFrFT حالت تعمیم‌یافته STFT است، می‌توان نتیجه گرفت که شکل‌دهنده MVDR مرتبه کسری به ازای $\alpha = 1$ معادل فیلترینگ فضایی در حوزه STFT و به ازای $\alpha = 0$ معادل MVDR استاندارد در حوزه زمان است. در بخش بعدی نشان خواهیم داد ضرائب بهینه MVDR کسری بدست آمده با فرض $\alpha \neq 0, 1$ می‌تواند منجر به میزان حذف نویز و پژواک بیشتری شود.

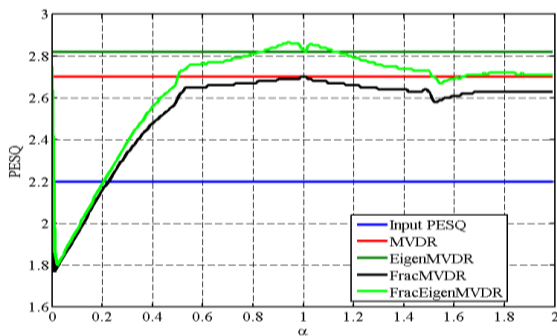
۴- ارزیابی عملکرد

در این بخش عملکرد شکل‌دهنده‌های پرتو پیشنهادی در شرایط اتاق شبیه‌سازی شده بر اساس دو معیار مطرح در زمینه بهسازی گفتار بررسی خواهد شد. ابتدا در زیربخش ۱-۴ شرایط اتاق شبیه‌سازی شده بیان می‌شود. سپس پارامترهای ارزیابی عملکرد در زیربخش ۲-۴ آورده خواهند شد. همچنین در زیربخش ۳-۴ روش‌های پایه مورد استفاده برای مقایسه عملکرد بطور مختصر توضیح داده می‌شوند. در زیربخش ۴-۴ مقدار بهینه α به‌دست آمده و در ادامه،

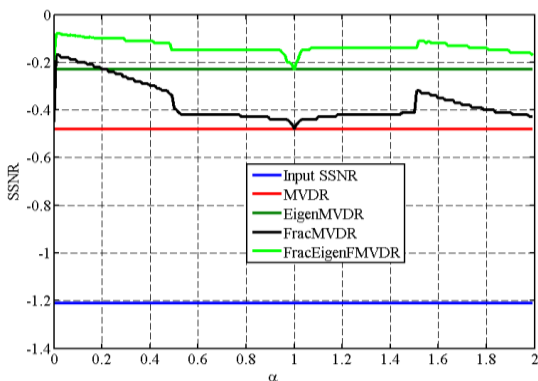
^{۱۵} Segmental Signal to Noise Ratio

^{۱۴} Perceptual Evaluation of Speech Quality

در این آزمایش زمان پژواک $RT_{60} = 300ms$ و نسبت توان سیگنال به نویز پخشنده^{۱۶} $SDNR = 12dB$ ، نسبت سیگنال به نویز همهمه $SBNR^{17} = 12dB$ و نسبت سیگنال به نویز غیرهمدوس^{۱۸} $SINR = 30dB$ در نظر گرفته شده است. در این راستا میزان بهبود کیفیت شنیداری و میزان حذف نویز به صورت تابعی از α برای شکل‌دهنده پرتو MVDR استاندارد، MVDR کسری (FracMVDR)، مبتنی بر آنالیز مقادیر ویژه (EigenMVDR) و MVDR کسری مبتنی بر آنالیز مقادیر ویژه (FracEigenMVDR) بررسی شده است. نتایج حاصل از این آزمایش در شکل‌های ۲ و ۳ نشان داده شده است. همان‌طور که ملاحظه می‌شود شکل‌دهنده پرتو FracEigenMVDR در برخی α ها توانسته به کیفیت شنیداری بیشتری نسبت به دیگر روش‌ها دست یابد. بیشترین میزان کیفیت شنیداری بدست آمده $PESQ = 2.87$ در $\alpha = 0.94$ اتفاق افتاده است. این در حالی است که کیفیت خروجی EigenMVDR در حد $PESQ = 2.82$ بوده و همچنین FracMVDR و MVDR استاندارد عملکردی حدوداً یکسان دارند و به $PESQ = 2.70$ رسیده‌اند. در $\alpha = 1$ همان‌طور که انتظار می‌رود FracMVDR و FracEigenMVDR مشابه MVDR استاندارد و EigenMVDR دارند. از این رو می‌توان نتیجه گرفت که دو شکل‌دهنده پرتو MVDR مرتبه کسری، حالت تعمیم یافته MVDR استاندارد و EigenMVDR هستند.



شکل ۲: میزان بهبود کیفیت شنیداری بر حسب تغییرات α



شکل ۳: میزان حذف نویز بر حسب تغییرات α

همان‌طور که در بخش قبل بیان شد ما دو تغییر در شکل‌دهنده پرتو MVDR پیشنهاد داده‌ایم. تغییر اول استفاده از آنالیز مقادیر ویژه و دیگری بهره‌گیری از تبدیل فوریه زمان-کوتاه کسری است. با توجه به این توضیح، نکته دیگری که از شکل ۲ قابل مشاهده و استنباط می‌باشد این است که آنالیز مقادیر ویژه در مقایسه با تعمیم به تبدیل فوریه زمان-کوتاه کسری تاثیر بیشتری بر بهبود سیگنال خروجی دارد. این مهم از مقایسه سه نمودار PESQ ورودی با PESQ بدست آمده با استفاده از MVDR استاندارد، FracMVDR و EigenMVDR قابل

نویز روی کل سیگنال، این نسبت روی هر یک از پنجره‌های مورد پردازش محاسبه شده و سپس این مقادیر، میانگین‌گیری می‌شوند. این نوع محاسبه نسبت توان سیگنال به نویز از این جهت مورد توجه است که با نحوه ادراک گوش انسان از سیگنال‌های نویزی همخوانی مناسب‌تری دارد [۱۷].

۳-۴ روش‌های پایه مبنای مقایسه

شکل‌دهنده‌های پرتو پیشنهاد شده در این تحقیق با دو روش پایه زیر مورد مقایسه قرار می‌گیرد:

۱- MVDR استاندارد [۱]:

با توجه به این که روش‌های پیشنهادی حالت تعمیم‌یافته شکل‌دهنده پرتو MVDR است، شکل‌دهنده MVDR به عنوان یکی از روش‌های پایه مبنای مقایسه انتخاب شده است. رابطه‌ی (۱۰) بیان‌کننده ضرائب این شکل‌دهنده است.

۲- فیلتر وینر چندکاناله مقید با رتبه یک (RICMWF) [۱۸]:

این روش حالت تعمیم‌یافته روش فیلتر وینر چندکاناله (MWF) است. در این روش، با هدف برقراری مصالحه‌ای بین میزان حذف نویز و اعوجاج گفتار قیودی به مسئله بهینه‌سازی طراحی شده برای بدست آوردن ضرائب فیلتر وینر چندکاناله اضافه شده است. در این روش با در نظر گرفتن این قید که شاخص حذف نویز بزرگتر یا مساوی یک باشد، میزان اعوجاج گفتار کمینه می‌شود. با حل مسئله بهینه‌سازی توصیف شده، ضرائب این فیلتر به صورت زیر بدست می‌آید [۱۸]:

$$h_{RICMWF-\mu_G}(k, f) = \frac{\phi^{-1}(k, f)\phi_x(k, f)}{\mu_G(k, f) + \lambda(k, f)} i_1 \quad (24)$$

که در آن $\lambda(k, f) = \text{trace}\{\phi_n^{-1}(l, k)\phi_x(k, f)\}$

در رابطه این فیلتر چنانچه $\mu(k, f) = 1$ در نظر گرفته شود ضرائب فیلتر وینر استاندارد بدست می‌آید. همچنین برای بدست آوردن ماتریس کواریانس نویز و سیگنال دریافتی به جای استفاده از رابطه هموارسازی بازگشتی از میانگین آماری به فرم زیر استفاده شده است [۱۸]:

$$\phi_x(k, f) = \frac{1}{L} \sum_{l=-\frac{L}{2}}^{\frac{L}{2}-1} M_x(k, f) y(k, f) y^H(k, f) \quad (25)$$

$$\phi_v(k, f) = \frac{1}{L} \sum_{l=-\frac{L}{2}}^{\frac{L}{2}-1} M_v(k, f) y(k, f) y^H(k, f)$$

که در آن M_x و M_v به ترتیب احتمال حضور گفتار و نویز هستند.

۴-۴ تعیین مقدار بهینه α

همان‌طور که در بخش قبلی بیان شد، امکان بدست آوردن مقدار بهینه α به صورت فرم بسته وجود ندارد. از این رو با تغییر α در بازه $[0-2]$ و محاسبه پارامترهای بررسی عملکرد، α به عنوان مقدار بهینه انتخاب می‌شود که به ازای آن، PESQ و SSNR سیگنال پردازش شده بیشترین مقدار را داشته باشد.

^{۱۸} Signal to Incoherence Noise Ratio

^{۱۶} Signal to Diffuse Noise Ratio

^{۱۷} Signal to Babble Noise Ratio

مشاهده است. شکل ۳ میزان حذف نویز بر حسب تغییرات α هنگام اعمال روش‌های مورد مقایسه را نشان می‌دهد. همانطور که ملاحظه می‌شود شکل‌دهنده‌پرتو FracEigenMVDR در تمامی مقادیر α توانسته نسبت به MVDR استاندارد، EigenMVDR و FracMVDR عملکرد بهتری در زمینه حذف نویز داشته باشد. ماکزیمم مقدار SSNR در $\alpha = 0.01$ ($SSNR = -0.08dB$) به دست آمده است. همچنین در $0 \leq \alpha \leq 2$ ، FracMVDR توانسته نسبت به EigenMVDR به میزان حذف نویز بیشتری دست یابد. مانند شکل ۲، در $\alpha = 1$ عملکرد حذف نویز FracEigenMVDR و FracMVDR به ترتیب برابر با MVDR استاندارد و EigenMVDR است.

برای انتخاب α ی بهینه باید مقداری را تعیین نمود که هم از لحاظ کیفیت شنیداری و هم از لحاظ میزان حذف نویز شرایط قابل قبولی برقرار باشد. در $\alpha = 0.01$ اگر چه میزان بهبود SSNR قابل توجه است اما از لحاظ کیفیت شنیداری عملکرد روش FracEigenMVDR مناسب نیست. اما در $\alpha = 0.94$ علاوه بر رسیدن به بیشترین میزان PESQ، با عنایت به $SSNR = -0.15dB$ عملکرد حذف نویز نیز قابل توجه است. در نتیجه می‌توان $\alpha = 0.94$ را به عنوان α بهینه انتخاب نمود.

سناریو ۲: محیط پژواکی:

در این سناریو $RT_{60} = 600ms$ در نظر گرفته می‌شود. نتایج حاصل از ارزیابی مقایسه‌ای روش‌ها در این سناریو در جدول ۲ آورده شده است. با توجه به این که در این سناریو هدف بررسی عملکرد در راستای حذف پژواک است، فقط نویز پخشنده در محیط با $SDNR = 20dB$ در نظر گرفته شده است.

نتایج نشان می‌دهد که هر سه روش پیشنهادی نسبت به روش پایه عملکرد بهتری دارند. در عین حال به دلیل کم بودن نویز محیط (ورودی $SDNR = 20dB$) میزان بهبود پارامترهای ارزیابی عملکرد نسبت به نتایج حاصل از سناریوی قبل کمتر است. به‌طور مثال روش FracEigenMVDR این آزمایش مقدار SSNR را $0.78dB$ بهبود داده، اما در سناریوی اول این میزان بهبود برابر با $1.14dB$ بود. این وضعیت در مورد پارامتر PESQ نیز برقرار است.

جدول ۲- میزان بهبود PESQ و SSNR در محیط پژواکی با

$SDNR = 20dB$ و $RT_{60} = 600ms$

Method \ Parameter	SSNR (dB)	PESQ
Input	-۱/۰۰	۲/۲۱
MVDR [1]	-۰/۴۰	۲/۵۰
EigenMVDR	-۰/۳۳	۲/۵۳
FracMVDR	-۰/۳۲	۲/۵۱
FracEigenMVDR	-۰/۲۲	۲/۵۵
MWF [18]	-۰/۳۴	۲/۴۹
RICMWF [18]	-۰/۳۳	۲/۴۹

سناریو ۳: محیط نویزی-پژواکی:

در اتاق شبیه‌سازی شده در این سناریو، نویز و پژواک به صورت توأم دارای شدت زیادی هستند. در این حالت، $RT_{60} = 600ms$ ، $SBNR = 5dB$ و $SDNR = 5dB$ در نظر گرفته شده است. جدول ۳ عملکرد روش‌ها در این سناریو را نشان می‌دهد. در این سناریو هم روش FracEigenMVDR توانسته در زمینه حذف نویز $0.55dB$ نسبت به روش MVDR استاندارد، $0.42dB$ نسبت به R1CMWF، $0.49dB$ نسبت به FracMVDR و $0.09dB$ نسبت به EigenMVDR عملکرد بهتری در حذف نویز پخشنده داشته باشد. این حالت در مورد میزان بهبود کیفیت شنیداری و مقابله با نویز همهمه نیز برقرار است. در این سناریو نیز دیده می‌شود که تاثیر آنالیز مقادیر ویژه در راستای حذف نویز و نیز در بهبود کیفیت شنیداری بیشتر است. در یک دید کلی، نتایج حاصل

شکل ۳ میزان حذف نویز بر حسب تغییرات α هنگام اعمال روش‌های مورد مقایسه را نشان می‌دهد. همانطور که ملاحظه می‌شود شکل‌دهنده‌پرتو FracEigenMVDR در تمامی مقادیر α توانسته نسبت به MVDR استاندارد، EigenMVDR و FracMVDR عملکرد بهتری در زمینه حذف نویز داشته باشد. ماکزیمم مقدار SSNR در $\alpha = 0.01$ ($SSNR = -0.08dB$) به دست آمده است. همچنین در $0 \leq \alpha \leq 2$ ، FracMVDR توانسته نسبت به EigenMVDR به میزان حذف نویز بیشتری دست یابد. مانند شکل ۲، در $\alpha = 1$ عملکرد حذف نویز FracEigenMVDR و FracMVDR به ترتیب برابر با MVDR استاندارد و EigenMVDR است.

برای انتخاب α ی بهینه باید مقداری را تعیین نمود که هم از لحاظ کیفیت شنیداری و هم از لحاظ میزان حذف نویز شرایط قابل قبولی برقرار باشد. در $\alpha = 0.01$ اگر چه میزان بهبود SSNR قابل توجه است اما از لحاظ کیفیت شنیداری عملکرد روش FracEigenMVDR مناسب نیست. اما در $\alpha = 0.94$ علاوه بر رسیدن به بیشترین میزان PESQ، با عنایت به $SSNR = -0.15dB$ عملکرد حذف نویز نیز قابل توجه است. در نتیجه می‌توان $\alpha = 0.94$ را به عنوان α بهینه انتخاب نمود.

۴-۵- بررسی عملکرد در شرایط مختلف نویزی-پژواکی

در ادامه، عملکرد روشهای پیشنهادی و پایه در محیط‌های با شرایط نویزی-پژواکی مختلف مورد بررسی قرار می‌گیرد. در این راستا سه سناریوی مختلف در نظر گرفته شده است. با توجه به نتایج آزمایش قبل، در هر سه سناریو $\alpha = 0.94$ قرار داده شده است.

سناریو ۱: محیط نویزی:

در این سناریو در محیطی با اثر غالب نویز، عملکرد روش‌های FracEigenMVDR، FracMVDR و EigenMVDR با MVDR استاندارد [۱] و MWF و R1CMWF [۱۸] مقایسه می‌شود. در این سناریو علاوه بر نویز پخشنده و سفید، نویز همهمه هم حضور داشته و عملکرد روش‌های پیشنهادی و پایه در مقابله با این نویز نیز بررسی می‌شود. بدین منظور $RT_{60} = 200ms$ ، $SBNR = 5dB$ و $SDNR = 5dB$ در نظر گرفته شده است. نتایج حاصل از این ارزیابی در جدول ۱ آورده شده است.

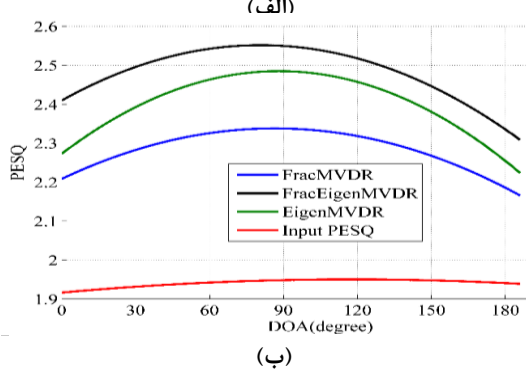
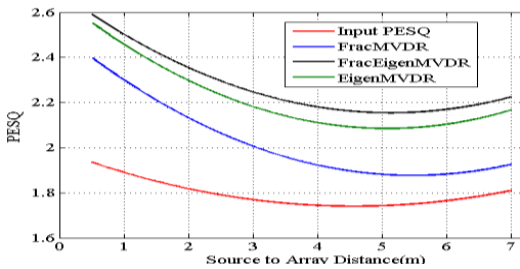
جدول ۱- میزان بهبود PESQ و SSNR در محیط نویزی با

$SDNR = 5dB$ و $SBNR = 5dB$ ، $RT_{60} = 200ms$

Method \ Parameter	SSNR (dB)		PESQ	
	Diffuse	Babble	Diffuse	Babble
Input	-۱/۴۶	-۱/۴۵	۱/۶۸	۱/۷۲
MVDR [1]	-۰/۹۷	-۰/۹۱	۲/۲۵	۲/۲۹
EigenMVDR	-۰/۴۵	-۰/۴۸	۲/۳۷	۲/۳۱
FracMVDR	-۰/۹۵	-۰/۹۰	۲/۲۹	۲/۲۴
FracEigenMVDR	-۰/۳۳	-۰/۴۷	۲/۳۹	۲/۳۴
MWF [18]	-۰/۹۶	-۰/۹۱	۲/۱۳	۲/۲۹
RICMWF [18]	-۰/۹۰	-۰/۹۱	۲/۳۲	۲/۲۹

نتایج حاصل از اعمال روشهای پیشنهادی و پایه به سیگنال‌های دریافتی توسط آرایه در این سناریو نشان می‌دهد روش FracEigenMVDR هم در محیط با نویز پخشنده و هم در محیطی با نویز همهمه توانسته نسبت به روش‌های پایه

در ادامه، به منظور بررسی پایداری روش‌های پیشنهادی در مواجهه با تغییر زاویه ورود سیگنال در بازه $[0^\circ - 180^\circ]$ و تغییر در فاصله بین منبع تا آرایه در بازه $[0.5 - 7m]$ دو آزمایش با شرایط اتاقی مشابه بخش ۴-۴ صورت گرفته و خروجی روش‌های پیشنهادی بدست آمده و پارامتر PESQ برای این خروجی‌ها محاسبه شده است. نتایج حاصل از این دو آزمایش در شکل ۴ آمده است. همانطور که دیده می‌شود روش‌های پیشنهادی در زوایا و فواصل مختلف منبع دارای عملکرد قابل قبول و پایداری بوده و به بیان دیگر، روش‌های پیشنهادی نسبت به تغییر زاویه ورود و فاصله منبع تا آرایه مقاوم است.



شکل ۴: کیفیت شنیداری روش‌های پیشنهادی بر حسب تغییرات: (الف) فاصله منبع تا آرایه و (ب) زاویه ورود سیگنال

۴-۷- شکل موج و طیف سیگنال‌های خروجی

در این زیربخش، نمونه‌هایی از طیف‌نگاشت و شکل‌موج زمانی سیگنال ورودی به میکروفون اول و سیگنال خروجی هر یک از روش‌های پایه و پیشنهادی نشان داده می‌شود. طول سیگنال پردازش شده برابر 1sec بوده و $\alpha = 0.94$ قرار داده شده است. در اینجا شکل ۵(الف) سیگنال ورودی در بافتی توسط میکروفون اول (که دارای $RT_{60} = 300ms$, $SDNR = 12dB$ و $SINR = 30dB$ است) را نشان می‌دهد. شکل‌های ۵(ب) تا (ه) به ترتیب شکل موج زمانی و طیف‌نگاشت خروجی روش MVDR استاندارد، شکل‌دهنده پرتو EigenMVDR، شکل‌دهنده پرتو FracMVDR و شکل‌دهنده پرتو FracEigenMVDR را نشان می‌دهد. با مقایسه طیف‌نگاشت سیگنال‌های بهبودیافته و ورودی (به طور خاص در حدود زمان 0.3sec و فرکانس 6-8kHz) دیده می‌شود پهن‌شدگی طیفی ایجاد شده بر اثر پژواک محیط مخصوصاً در خروجی روش‌های پیشنهادی کاهش یافته است. کاهش پهن‌شدگی طیفی در FracEigenMVDR نسبت به دیگر روش‌ها بیشتر است.

با در نظر گرفتن طیف‌نگاشت و شکل‌موج زمانی در بازه زمانی حدود 0-0.125sec نیز می‌توان به صورتی شهودی نشان داد که روش FracEigenMVDR نسبت به دیگر روش‌ها توانسته به میزان حذف نویز بیشتری دست یابد. مشاهدات حاصل از این مقایسه، تاییدکننده آزمایش‌های قبلی است.

از این سناریو نیز تاییدی بر نتایج دو سناریوی قبلی بوده و روش FracEigenMVDR در همه سناریوهای فوق نسبت به بقیه روش‌ها عملکرد بالاتری دارد.

جدول ۳- میزان بهبود PESQ و SSNR در محیط نویزی-پژواکی با

$$SDNR = 5dB \text{ و } SBNR=5dB, RT_{60} = 600ms$$

Method	SSNR (dB)		PESQ	
	Diffuse	Babble	Diffuse	Babble
Input	-۱/۶۳	-۱/۶۰	۱/۵۵	۲/۳۳
MVDR [1]	-۱/۱۲	-۱/۱۲	۱/۹۱	۱/۹۱
EigenMVDR	-۰/۶۶	-۰/۶۳	۲/۱۰	۲/۱۴
FracMVDR	-۱/۰۶	-۱/۰۳	۱/۹۸	۱/۹۹
FracEigenMVDR	-۰/۵۷	-۰/۶۲	۲/۱۸	۲/۱۵
MWF [18]	-۱/۰۹	-۱/۰۴	۱/۹۶	۲/۰۵
RICMWF [18]	-۱/۰۸	-۱/۰۳	۱/۹۷	۲/۰۵

به عنوان جمع‌بندی نتایج شبیه‌سازی و ارتباط آنها با تغییرات اعمالی در شکل‌دهنده پرتو MVDR، می‌توان این موارد را ذکر نمود:

- در تغییر اول با توجه به اینکه سیگنال منبع مطلوب معادل بزرگترین مقدار ویژه در ماتریس کواریانس است. می‌توان با تجزیه مقادیر ویژه و حذف مقادیر ویژه کوچک دقت ماتریس کواریانس را افزایش داد. در نتیجه ضرایب $h(k, f)$ (رابطه ۱۴) بهبود قابل توجهی می‌یابد. به عنوان مثال در سناریو سوم، روش EigenMVDR توانسته نسبت به MVDR استاندارد، $SSNR$ را $0.46dB = -(-1.12) - (-0.66)$ بیشتر بهبود بخشد.
- در تغییر دوم، با استفاده از تبدیل فوریه زمان کوتاه کسری با $\alpha = 0.94$ به جای تبدیل فوریه زمان کوتاه استاندارد ($\alpha=1$) میزان حذف نویز به اندازه $0.06dB = -1.06 - (-1.12)$ بیشتر بهبود یافته است. این میزان بهبود را می‌توان این گونه تعبیر نمود که در $\alpha = 0.94$ نسبت به $\alpha = 1$ همپوشانی زیرفضای نویز و سیگنال کمتر شده است.

۴-۶- تحلیل پایداری روش در برابر تغییر پارامترها و عوامل محیطی

در این بخش، پایداری روش‌های پیشنهادی در برابر تغییر پارامترها و عوامل محیطی مورد بررسی قرار می‌گیرد. عوامل و پارامترهای مؤثر بر پایداری شکل‌دهنده‌های پیشنهادی پرتو عبارتند از:

- پارامتر تبدیل فوریه کسری (α)
- انواع نویزهای موجود در محیط با توان‌های متفاوت
- وجود پژواک در محیط با زمان پژواک متفاوت
- تعبیر در راستای ورود سیگنال
- تغییر در فاصله بین منبع سیگنال تا آرایه

در بخش ۴-۴ به منظور تعیین مقدار بهینه پارامتر α ، این پارامتر در بازه $[0-2]$ تغییر داده شد و خروجی هر یک از روش‌های پیشنهادی محاسبه گردید. با توجه به نتایج بدست آمده در این آزمایش (شکل‌های ۲ و ۳) می‌توان نتیجه گرفت که روش‌های پیشنهادی نسبت به تغییرات α مقاوم است. همچنین در بخش ۴-۵ دیده شد که روش‌های پیشنهادی در شرایط متنوع نویزی و پژواکی (با میزان و انواع متغیر نویز و پژواک) رفتار پایداری را از خود نشان می‌دهند.

۵- جمع‌بندی و نتیجه‌گیری

در این مقاله به منظور حذف عوامل مخرب محیطی از سیگنال دریافتی توسط آرایه، دو تعمیم به الگوریتم شکل‌دهنده پرتو MVDR پیشنهاد و اعمال شد. در بحث بهسازی گفتار معمولاً فرض می‌شود سیگنال گفتار مطلوب و نویز ناهمدوس هستند. با توجه به این فرض، ماتریس کواریانس نویز و سیگنال مطلوب محاسبه شده و بر اساس این ماتریس‌ها ضرائب شکل‌دهنده پرتو به دست می‌آید. به دلیل برقرار نبودن این فرض در بسیاری از محیط‌های واقعی، ماتریس کواریانس‌های بدست آمده و ضرائب شکل‌دهنده حاصل، از دقت کافی برخوردار نیست. به عنوان راه‌حلی برای این مساله، با استفاده از تجزیه مقادیر ویژه، ماتریس کواریانس سیگنال مطلوب تجزیه شده و با حذف مقادیر ویژه کوچک و سپس بازسازی آن ماتریس کواریانس سیگنال مطلوب با دقت بالا به دست می‌آید. در تعمیم دوم، با بهره‌گیری از تبدیل STFrFT به جای STFT و تعیین پارامتر بهینه آن، میزان حذف نویز و پژواک افزایش یافته است. پس از فرمول‌بندی هر یک از تعمیم‌های مذکور، با طراحی چندین آزمایش پارامتر بهینه تبدیل STFrFT از طریق تجربی برابر با $\alpha = 0.94$ بدست آمده و سپس عملکرد روش‌های پیشنهادی با روش پایه مقایسه شده است. نتایج حاصل از آزمایش‌ها نشان می‌دهد که روش پیشنهادی FracEigenMVDR نسبت به روش پایه و دیگر روش‌های پیشنهادی عملکرد بالاتری در راستای حذف نویز و پژواک دارد. به طور مثال در محیط با میزان زیاد نویز و پژواک ($SDNR = 5dB$) و روش پیشنهادی FracEigenMVDR کیفیت شنیداری ($PESQ$) را به مقدار 0.62 واحد و مقدار SSNR را $1.06dB$ بهبود داده است. این نتایج در حالی است که در خروجی روش پایه (MVDR)، مقدار PESQ و SSNR به ترتیب 0.35 واحد و $0.51dB$ بهبود یافته است.

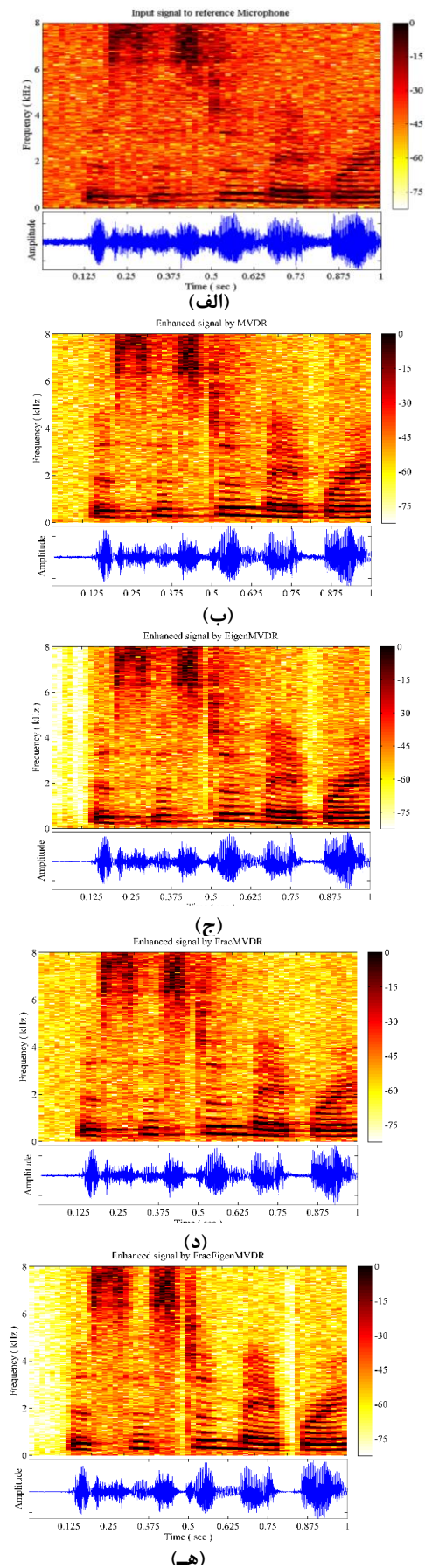
در مجموع، بهبود عملکرد ایجاد شده بواسطه‌ی هر یک از تعمیم‌های پیشنهاد شده را می‌توان بدین صورت تعبیر کرد: ۱- در تغییر اول با توجه به اینکه معمولاً یک منبع مطلوب داریم، ماتریس کواریانس سیگنال دریافتی دارای یک مقدار ویژه بزرگ (معادل منبع مطلوب) و تعدادی مقدار ویژه کوچک (معادل عوامل مخرب محیطی) خواهد بود. بنابراین چنانچه مقادیر ویژه کوچک حذف شوند، مقدار قابل توجهی از نویز دریافتی توسط آرایه حذف خواهد شد. ۲- در تغییر دوم نیز، استفاده از تبدیل فوریه زمان-کوتاه کسری به جای تبدیل فوریه زمان-کوتاه استاندارد، به دلیل توانایی بیشتر تبدیل کسری در پردازش سیگنال‌های غیرایستنا (به‌دست آوردن بهتر مشخصات محلی زمان-فرکانس سیگنال)، امکان حذف بیشتر نویز توسط شکل‌دهنده پیشنهادی فراهم شده است.

مراجع

- [1] E. A. P. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, "New insights into the MVDR beamformer in room acoustics," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 18, no. 1, pp. 158-170, 2009.
- [2] S. Zhang and X. Li, "Microphone array generalization for multichannel narrowband deep speech enhancement," in *Proceedings of INTERSPEECH*, pp. 667-670, 2021.

[۳] آوید آوخ، حمیدرضا ابوطالبی «بهسازی وقتی سیگنال گفتار در محیط‌های واقعی با استفاده از ساختار ترکیبی مبتنی بر شکل‌دهنده‌های پرتو و فیلتر پسینه»، *مجله مهندسی برق دانشگاه تبریز*، جلد ۴۸، شماره ۲، صفحات ۴۹۵-۴۸۱، ۱۳۹۷.

[۴] امیرحسین حاج‌احمدی، محمد مهدی همایون‌پور، «بهسازی گفتار دومرحله‌ای توسط خودرمنگار عمیق کاهنده نویز»، *مجله مهندسی برق دانشگاه تبریز*، جلد ۵۰، شماره ۴، صفحات ۱۵۴۰-۱۵۳۳، ۱۳۹۹.



شکل ۵: نمونه شکل موج زمانی و طیف نگاشت (الف): ورودی، و خروجی‌های (ب): MVDR استاندارد (ج): EigenMVDR، (د): FracMVDR و (ه): FracEigenMVDR

- [12] S. Das and I. Pan, *Fractional order signal processing: introductory concepts and applications*. Springer Science & Business Media, 2011.
- [13] J. Yin, K. Guo, X. Han, and G. Yu, "Fractional Fourier transform based underwater multi-targets direction-of-arrival estimation using wideband linear chirps," *Appl. Acoust.*, vol. 169, no. 1, 107477, 2020.
- [14] E. A. P. Habets, "Room impulse response (RIR) generator," May 2008. [Online]. Available: <https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator>.
- [15] J. S. Garofolo *et al.*, "TIMIT acoustic phonetic continuous speech corpus," *Linguist. Data Consortium*, 1993, Accessed: May 24, 2022. [Online]. Available: <http://ci.nii.ac.jp/naid/20000921365/en/>.
- [16] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, pp. 749–752, 2001.
- [17] S. Yook *et al.*, "Modified segmental signal-to-noise ratio reflecting spectral masking effect for evaluating the performance of hearing aid algorithms," *Speech Commun.*, vol. 55, no. 10, pp. 1003–1010, 2013.
- [18] Z. Wang, E. Vicent, R. Serizel, and Y. Yan, "Rank-1 constrained multichannel Wiener filter for speech recognition in noisy environments," *Comput. Speech & Lang.*, vol. 49, pp. 37–51, 2018.
- [5] J. Benesty, J. Chen, and E. A. P. Habets, *Speech enhancement in the STFT domain*. Springer Science & Business Media, 2011.
- [6] N. Yazdi and K. Todros, "Measure-transformed MVDR beamforming," *IEEE Signal Process. Lett.*, vol. 27, pp. 1959–1963, 2020.
- [7] D. Ying and Y. Yan, "Robust and fast localization of single speech source using a planar array," *IEEE Signal Process. Lett.*, vol. 20, no. 9, pp. 909–912, 2013.
- [8] V. W. Neo, C. Evers, and P. A. Naylor, "PEVD-based speech enhancement in reverberant environments," in *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 186–190, 2020.
- [9] V. W. Neo, C. Evers, and P. A. Naylor, "Enhancement of noisy reverberant speech using polynomial matrix eigenvalue decomposition," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 3255–3266, 2021.
- [10] J. Shi, J. Zheng, X. Liu, W. Xiang, and Q. Zhang, "Novel short-time fractional Fourier transform: theory, implementation, and applications," *IEEE Trans. Signal Process.*, vol. 68, pp. 3280–3295, 2020.
- [11] I. S. Yetik and A. Nehorai, "Beamforming using the fractional Fourier transform," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1663–1668, 2003.