



بهسازی گفتار به کمک یادگیری واژه‌نامه مبتنی بر داده

سمیرا مودّتی^{۱*} و سید محمد احمدی^۲

^۱ استادیار، دانشکده فنی و مهندسی، دانشگاه مازندران، بابلسر، ایران

^۲ استاد، دانشکده مهندسی برق، دانشگاه صنعتی امیرکبیر، تهران، ایران

چکیده

بهسازی گفتار یکی از پرکاربردترین حوزه‌ها در زمینه پردازش گفتار است. در این مقاله، یکی از روش‌های بهسازی گفتار مبتنی بر اصول بازنمایی تُنک بررسی می‌شود. بازنمایی تُنک این امکان را فراهم می‌سازد که عمدۀ اطلاعات لازم برای بازنمایی سیگنال، براساس بُعد بسیار کمتری از پایه‌های فضایی اصلی قابل مدل‌سازی باشد. روش یادگیری در این مقاله براساس تصحیح الگوریتم تطبیقی حریصانه مبتنی بر داده خواهد بود که واژه‌نامه در آن، به طور مستقیم از روی سیگنال داده و براساس شاخص تُنکی مبتنی بر نرم به منظور تطابق بیشتر میان اتم‌ها و ساختار داده آموزش می‌یابد. در این مقاله شاخص تُنکی جدیدی براساس معیار جینی پیشنهاد می‌شود. همچنین محدوده پارامتر تُنکی بخش‌های نویه‌ای با توجه به فریم‌های ابتدایی گفتار تعیین و طی یک روال پیشنهادی در تشکیل واژه‌نامه مورد استفاده قرار می‌گیرد. نتایج بهسازی نشان می‌دهد که عملکرد روش پیشنهادی در انتخاب قاب‌های داده براساس معیار معرفی شده در شرایط نویه‌ای مختلف بهتر از شاخص تُنکی مبتنی بر نرم و سایر الگوریتم‌های پایه در این راستا است.

واژگان کلیدی: بهسازی گفتار، بازنمایی تُنک، یادگیری واژه‌نامه، مبتنی بر داده، تطبیقی حریصانه، شاخص تُنکی جینی

Speech Enhancement using Adaptive Data-Based Dictionary Learning

Samira Mavaddati^{1*} & Seyyed Mohammad Ahadi²

¹ Electrical Department, Faculty of Technology and Engineering, University of Mazandaran, Babolsar, Iran

² Electrical Engineering Department, Amirkabir University of Technology, Tehran, Iran

Abstract

In this paper, a speech enhancement method based on sparse representation of data frames has been presented. Speech enhancement is one of the most applicable areas in different signal processing fields. The objective of a speech enhancement system is improvement of either intelligibility or quality of the speech signals. This process is carried out using the speech signal processing techniques to attenuate the background noise without causing any distortion in the speech signal. In this paper, we focus on the single channel speech enhancement corrupted by the additive Gaussian noise. In recent years, there has been an increasing interest in employing sparse representation techniques for speech enhancement. Sparse representation technique makes it possible to show the major information about the speech signal based on a smaller dimension of the original spatial bases. The capability of a sparse decomposition method depends on the learned dictionary and matching between the dictionary atoms and the signal features. An over complete dictionary is yielded based on two main steps: dictionary learning process and sparse coding technique. In dictionary selection step, a pre-defined dictionary such as the Fourier basis, wavelet basis or discrete cosine basis is employed. Also, a redundant dictionary can be constructed after a learning process

* Corresponding author

*نویسنده عهده‌دار مکاتبات

that is often based on the alternating optimization strategies. In sparse coding step, the dictionary is fixed and a sparse coefficient matrix with the low approximation error has been earned. The goal of this paper is to investigate the role of data-based dictionary learning technique in the speech enhancement process in the presence of white Gaussian noise. The dictionary learning method in this paper is based on the greedy adaptive algorithm as a data-based technique for dictionary learning. The dictionary atoms are learned using the proposed algorithm according to the data frames taken from the speech signals, so the atoms contain the structure of the input frames. The atoms in this approach are learned directly from the training data using the norm-based sparsity measure to earn more matching between the data frames and the dictionary atoms. The proposed sparsity measure in this paper is based on Gini parameter. We present a new sparsity index using Gini coefficients in the greedy adaptive dictionary learning algorithm. These coefficients are set to find the atoms with more sparsity in the comparison with the other sparsity indices defined based on the norm of speech frames. The proposed learning method iteratively extracts the speech frames with minimum sparsity index according to the mentioned measures and adds the extracted atoms to the dictionary matrix. Also, the range of the sparsity parameter is selected based on the initial silent frames of speech signal in order to make a desired dictionary. It means that a speech frame of input data matrix can add to the first columns of the over complete dictionary when it has not a similar structure with the noise frames. The data-based dictionary learning process makes the algorithm faster than the other dictionary learning methods for example K-singular value decomposition (K-SVD), method of optimal directions (MOD) and other optimization-based strategies. The sparsity of an input frame is measured using Gini-based index that includes smaller measured values for speech frames because of their sparse content. On the other hand, high values of this parameter can be yielded for a frame involved the Gaussian noise structure. The performance of the proposed method is evaluated using different measures such as improvement in signal-to-noise ratio (ISNR), the time-frequency representation of atoms and PESQ scores. The proposed approach results in a significant reduction of the background noise in comparison with other dictionary learning methods such as principal component analysis (PCA) and the norm-based learning method that are traditional procedures in this context. We have found good results about the reconstruction error in the signal approximations for the proposed speech enhancement method. Also, the proposed approach leads to the proper computation time that is a prominent factor in dictionary learning methods.

Keywords: Speech enhancement, Sparse representation, Dictionary learning, Data-Based learning, Greedy adaptive

کدگذاری منجر شوند. در مرجع [6] کاربرد روش سنجش فشرده^۵ در زمینه کدگذاری گفتار براساس پیش‌گویی خطی تنک مورد بررسی قرار گرفت. از این روش بهمنظور محاسبه تقریب تنک گفتار با استفاده از سیگنال تحریک در پیش‌گویی خطی به کار گرفته شده که دارای نمونه‌های غیرصفر بسیار کمی است. در [7] از روش سنجش فشرده بهمنظور کاهش نویه سیگنال صوت و گفتار بهره گرفته شده و مسئله کاهش نویه براساس تئوری سنجش فشرده بهصورت مسئله کمینه‌سازی نرم ۱/۱ با ترکیب خطی از قیود و عمل گر تبدیل فوریه جزئی تصادفی بیان شد. در ادامه، تنکی سیگنال در فضای تبدیل کسینوسی گستته^۶ نیز مورد بررسی قرار گرفته و معیارهایی برای تعیین میزان تنکی سیگنال ارائه شده است [8]. یکی از این معیارها، قید تنکی اکید سیگنال^۷ است که تنکی سیگنال را با شمارش محض تعداد درایه‌های غیرصفر بردار سیگنال در حوزه زمان یا هر حوزه تبدیل دیگر در نظر می‌گیرد [9-10]. در [11] از نمونه‌برداری فشرده بهمنظور بهسازی سیگنال گفتار^۸ در

۱- مقدمه

بحث نمونه‌برداری فشرده^۹ و مفاهیم مرتبط با آن مانند بازنمایی تنک^{۱۰} و یادگیری واژه‌نامه^{۱۱} مدت زمان زیادی در دست بررسی بوده و پژوهش‌های گسترده‌ای بهخصوص در زمینه پردازش سیگنال تصویر بر روی آن انجام گرفته است. این پژوهش‌ها شامل فشرده‌سازی، حذف نویه، قطعه‌بندی، طبقه‌بندی و بازسازی است [۱-۴]. مطالعات به کمک این روش‌ها بهمنظور بهسازی سیگنال گفتار تنها به سال‌های اخیر باز می‌گردد و پژوهش‌های زیادی در این زمینه انجام نشده است. بحث پردازش گفتار با تکیه بر یادگیری واژه‌نامه و بحث کدگذاری سیگنال گفتار در ابتدا از مرجع [5] آغاز شد. در این مرجع الگوریتمی برای آموزش پایه‌های فراکامل^{۱۲} معرفی شد که مشابه با یک مدل احتمالی از داده‌های مشاهده‌ای است. در ادامه نشان داده شد که پایه‌های فراکامل می‌توانند تقریب بهتری از توزیع آماری داده‌های مشاهده‌ای نمایش داده و به کارایی بیشتری در زمینه

⁵ Compressive sensing

⁶ Discrete cosine transform

⁷ Strict sparsity

⁸ Speech enhancement

¹ Compressive sampling

² Sparse representation

³ Dictionary learning

⁴ Overcomplete bases

شد [24]. تُنکی مضاعف به این معنی است که علاوه بر نمایش تُنک سیگنال، خود واژه‌نامه $D = AB$ نیز تُنک باشد؛ یعنی اتمه‌های A بر روی واژه‌نامه ثابت B بازنمایی تُنک داشته باشند که در آن B می‌تواند هر واژه‌نامه ثابت موجک با تبدیل کسینوسی گستته باشد. تبدیل موجک هنگامی که توابع پایه آن به خوبی محلی شده باشند، می‌تواند ارتباط مناسی را میان تُنکی در تجزیه و تُنکی در واژه‌نامه به منظور آنالیز سیگنال‌های طبیعی مانند صوت، تصویر و سیگنال‌های پزشکی فراهم کند. همچنین در [25] نیز از سیگنال گفتار به صورت مستقیم به عنوان اتمه‌های واژه‌نامه به کمک روش کدگذاری تُنک مبتنی بر آنالیز مؤلفه‌های مستقل¹ استفاده شده است؛ بنابراین می‌توان مفهوم تُنکی را برای نوع خاصی از سیگنال‌ها مانند گفتار یا موسیقی هم در بخش تجزیه و هم در واژه‌نامه استفاده کرد. نوع نوفه در روال بهسازی این روش به علت وابستگی کم قاب‌های زمانی به منظور تعیین شاخص تُنکی مناسب قاب‌ها، نوفه گوسی جمع‌شونده در نظر گرفته شده است. در این روش، استفاده از شاخص تُنکی دقیق که بتواند چینش مناسبی از قاب‌ها را بر حسب میزان تُنکی‌شان در واژه‌نامه مبتنی بر داده مورد نظر فراهم آورد، بسیار حائز اهمیت است. کارایی این عملکرد در حضور انواع نوفه دیگر پایین‌تر خواهد بود؛ زیرا هر گونه شباهت میان محتوای سیگنال گفتار و نوفه، موجب مرتبسازی نامناسب قاب‌ها در واژه‌نامه شده و ممکن است، قاب‌های نوفه به اشتباہ در اتمه‌های ابتدایی و قاب‌های گفتاری نیز در انتهایی واژه‌نامه قرار گرفته و حذف اتمه‌های انتهایی واژه‌نامه، اعوجاج سیگنال گفتار را نتیجه دهد. در روال یادگیری واژه‌نامه GAD، این مسئله در نظر گرفته شده است که با توجه به ساختار سیگنال گفتار، بیش‌تر قاب‌های گفتاری تُنکی مشخص و بالا و قاب‌های نوفه‌ای تُنکی پایینی دارند، بنابراین ستون‌های واژه‌نامه در هنگام طراحی، به ترتیب با قاب‌هایی از سیگنال مشاهده‌ای که بیشترین تا کمترین میزان تُنکی را دارند، جایگزین می‌شوند؛ سپس برای حذف بخش نوفه‌ای سیگنال تنها تعداد مشخص و محدودی از ستون‌های اولیه سیگنال در نظر گرفته و تخمین سیگنال تمیز حاصل می‌شود. واژه‌نامه وابسته به داده در این الگوریتم به صورت وفقی و به کمک استخراج پی‌درپی ستون‌های ماتریس الگوی ورودی X آموزش می‌یابند. استخراج ستون جدید در هر تکرار متوط به یافتن ستونی از X است که شرط زیر را برآورده کند:

$$\xi = \max_k \frac{\|x_k\|_1}{\|x_k\|_2} \quad (3)$$

¹ Independent component analysis(ICA)

می‌شود. نکته مهمی که در این روش وجود دارد و مورد بررسی قرار نگرفته است، این است که در انتخاب اتم‌ها برای ساخت واژه‌نامه تنها از معیار تُنکی بهره گرفته می‌شود و هیچ معیار دیگری به منظور تعیین مناسب‌بودن قاب انتخابی به منظور بازنمایی کردن محتوای سیگنال وجود ندارد. از آنجایی که هدف در این روش حذف نوفه است و اتم‌های انتهایی واژه‌نامه می‌باشد ساختار نوفه را به درستی بازنمایی کنند، بنابراین دخیل کردن پارامتر دیگری در انتخاب بهترین اتم‌ها با ماهیت نوفه‌ای مهم به نظر می‌رسد. در روش پیشنهادی، راه حلی به منظور مرتفع کردن این مشکل ارائه می‌شود. همچنین یک معیار جدید تُنکی معرفی می‌شود که می‌تواند بیان درستی از تُنکی سیگنال را به دست دهد تا روال حذف نوفه با دقت بالاتری حاصل شود.

در این مقاله، روال یادگیری واژه‌نامه توسط الگوریتم تطبیقی حریصانه بیان شده و سپس روش پیشنهادی ارائه می‌شود. در ادامه نتایج به دست‌آمده از به کارگیری روش پیشنهادی به منظور بهسازی گفتار مورد ارزیابی و مقایسه قرار می‌گیرد. در بخش آخر نیز به نتیجه گیری در مورد پژوهش انجام‌شده خواهیم پرداخت.

۲- الگوریتم یادگیری واژه‌نامه

تطبیقی حریصانه

سیگنال گفتار در مواجهه با نوفه سفید گوسی به صورت خطی زیر قابل مدل‌شدن است:

$$X(k) = S(k) + N(k) \quad (1)$$

که $X(k)$ ، $S(k)$ و $N(k)$ به ترتیب سیگنال نوفه‌ای، گفتار و نوفه با شماره قاب k هستند. سیگنال نوفه‌ای $X \in \mathbb{R}^N$ می‌تواند به صورت خطی با نمایش تُنک اتم‌ها به صورت $X = D.C$ بیان شود که در آن $D \in \mathbb{R}^{N \times P}$ ، $P > N$ یک واژه‌نامه فراکامل با P اتم است که اتم‌های آن به وسیله $\{d_p\}_{p=1}^P$ با نرم واحد P و بردار $\|d_{(c,p)}\|_2 = 1, \dots, P$ کدگذاری تُنک $c \in \mathbb{R}^P$ ، $P \gg K_0$ که شامل ضرایب تُنک است نمایش داده می‌شود [5].

$$c = \arg \min_c \|X - D.c\|^2 \quad (2)$$

در این رابطه K_0 تعداد ضرایب غیرصفر در ماتریس ضرایب تُنک C یا قید تُنکی را بیان می‌کند.

ایده تُنکی مضاعف در بحث یادگیری واژه‌نامه در ابتداء به منظور بهسازی تصاویر پرتونگاری رایانه‌ای سه‌بعدی مطرح

مؤلفه‌های اساسی^۱ و روش یادگیری واژه‌نامه مبتنی بر K-SVD² و همچنین نرخ خطای بازسازی این الگوریتم‌ها در شکل (۱) نشان داده شده است [۲۲-۲۶]. نتایج شبیده‌سازی‌ها گزارش شده در شکل‌ها و جداول این بخش بر روی یک GRID جمله مربوط به یک گوینده زن از مجموعه دادگان آزمایش شده است [۲۶]. تعداد اتم‌ها و نرخ تُنکی برای الگوریتم K-SVD به ترتیب برابر با چهارصد و ده تنظیم شده است. همچنین تعداد اتم‌های GAD چهارصد و تعداد اتم‌ها در PCA، سیصد در نظر گرفته شده است.

همانطور که مشاهده می‌شود، تُنکی ۴ حاصل از الگوریتم GAD با افزایش تعداد اتم‌ها با شبکه ملایمی افزایش می‌یابد. نرخ تُنکی برای الگوریتم‌های PCA و K-SVD وابستگی زیادی به تعداد اتم‌ها در واژه‌نامه ندارد و در ابتدای گام یادگیری قابل تنظیم و ثابت است. بنابراین می‌توان بیان کرد که با توجه به کمبودن شاخص تُنکی ۳ بر طبق رابطه (۱)، سیگنال بازنمایی شده توسط الگوریتم GAD نمایش تُنکتری نسبت به حوزه زمان و در مقایسه با سایر الگوریتم‌های بیان شده دارد. همچنین خطای بازسازی مطابق با قسمت (b) برای این الگوریتم مقدار مناسبی دارد. خطای تقریب حاصل نیز در جدول (۱) بیان شده که نشان می‌دهد بر طبق انتظار با افزایش تعداد اتم‌ها خطای تقریب کاهش می‌یابد. همان‌طور که بیان شد در این روش اتم‌ها به‌طور مستقیم از روی سیگنال گفتار یادگیری شده و به‌منظور داشتن یک الگوریتم سریع از یک تبدیل متعامد حاصل از مجموعه قاب‌های محلی که از قاب‌بندی سیگنال گفتار به‌دست می‌آید، استفاده می‌شود. این الگوریتم، تُنکترین قاب‌های داده ورودی را تشخیص داده و از روش متعامدسازی به‌منظور کاهش پیچیدگی‌های مسئله بهره می‌گیرد [۲۵].

(جدول-۱): خطای تقریب $\epsilon \times 10^{-3}$ بر حسب دسیبل برای الگوریتم GAD و PCA

(Table-1): The reconstruction error value ($\epsilon \times 10^{-3}$) based on decibel for GAD, PCA and K-SVD algorithms.

		تعداد اتم‌ها در واژه‌نامه					
		512	400	300	200	100	50
GAD		0.00	0.35	0.89	1.63	2.45	4.30
PCA		0.00	0.23	0.81	1.52	2.38	4.12
K-SVD		0.00	0.21	0.97	1.71	2.65	4.67

در گام نهایی الگوریتم، دو قانون به‌منظور رسیدن به تعداد اتم‌های موردنظر یا کاهش خطای تقریب از مقدار

¹ Principal component analysis(PCA)

² K-Singular value decomposition

بیشینه‌سازی رابطه بالا بربطق روال زیر خواهد بود [۲۳]

- مقداردهی اولیه: تضمین می‌شود که ستون‌های ماتریس X یعنی (x_k) که نرم l_1 واحد دارند، به عنوان ماتریس مانده اولیه در نظر گرفته شود:

$$\tilde{x}_k = \frac{x_k}{\|x_k\|_1} \rightarrow R^0 = \tilde{X} \quad (4)$$

- نرم l_2 هر فریم به صورت زیر محاسبه می‌شود:

$$E_k = \|r_k^j\|_2 = \sum_{n=1}^L |r_k^j(n)|^2 \quad (5)$$

در این رابطه $[r_k^j, r_k^j, \dots, r_k^j] = [r_1^j, r_2^j, \dots, r_L^j]$ بدار سوتونی مانده K بعدی متناظر با k امین ستون R^j است.

- شاخص \hat{k} متناظر با فریمی از سیگنال با بزرگترین نرم l_2 به دست می‌آید:

$$\hat{k} = \arg \max_{k \in \mathbb{K}} (E_k) \quad (6)$$

- ضریب توسعه با حاصل ضرب داخلی بدار مانده و اتم به دست آمده محاسبه می‌شود:

$$\alpha_k^j = \langle r_k^j, d_j \rangle \quad (7)$$

- مانده جدید با حذف مؤلفه‌های متناظر با اتم انتخاب شده برای هر k در r_k^{j+1} محاسبه می‌شود:

$$r_k^{j+1} = r_k^j - \frac{\alpha_k^j}{\langle d_j, d_j \rangle} d_j \quad (8)$$

عبارت $\langle d_j, d_j \rangle$ در مخرج رابطه بالا تضمین می‌کند که ضریب توسعه نرم‌البیز شده خواهد بود. بنابراین ستون متناظر در ماتریس مانده R^j به صفر مقداردهی شده و در نتیجه موجب متعامدشدن تبدیل می‌شود. تعامد تبدیل ویژگی مهم این الگوریتم است، زیرا با داشتن C=X=D.C، بدار ضرایب تُنک از رابطه $C=D^T X$ قابل محاسبه است.

- بازگشت به مرحله دوم تا هنگامی که تعداد اتم‌های مورد نیاز به دست آید [۲۳].

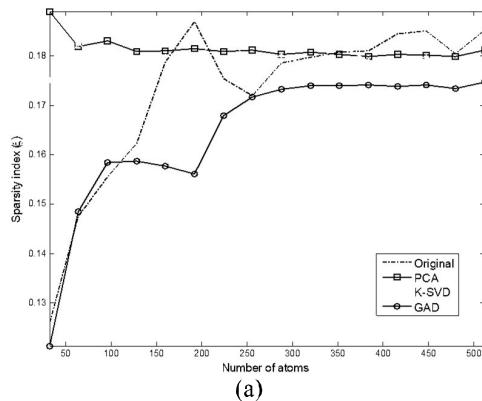
به‌منظور تعیین میزان تُنکی حاصل از یک الگوریتم یادگیری واژه‌نامه، شاخص تُنکی مبتنی بر نرم ۳ برای ضرایب تبدیل حاصل به صورت زیر تعریف می‌شود:

$$\xi = \|\mathbf{C}\|_1 / \|\mathbf{C}\|_2 \quad (9)$$

هر چقدر میزان این شاخص تُنکی کمتر باشد، به معنای آن است که سیگنال تُنک‌تر خواهد بود و بالعکس. مقایسه شاخص تُنکی در الگوریتم‌های GAD، آنالیز

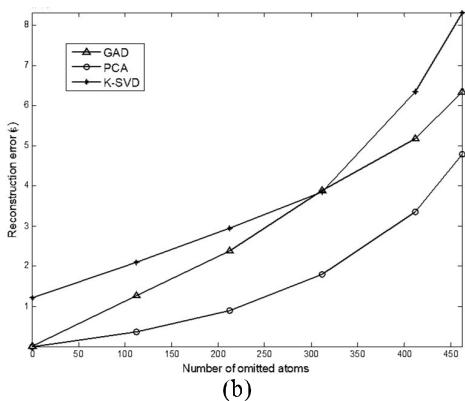
ا تم‌های مورد استفاده در واژه‌نامه مبتنی بر داده از نظر تُنکی در یک سطح هستند و ویژگی بارزی از سیگنال گفتار را بیان نمی‌کنند.

ا تم‌های حاصل از یادگیری با واژه‌نامه K-SVD نیز ساختار مشابه با داده دارند؛ اما در الگوریتم GAD، ا تم‌های ابتدایی به دست آمده برخلاف PCA نسبت به ا تم‌های انتها بیشتر کامل خصوصیات محلی سیگنال را به تصویر کشیده و دارای بیشترین تُنکی هستند.



(a)

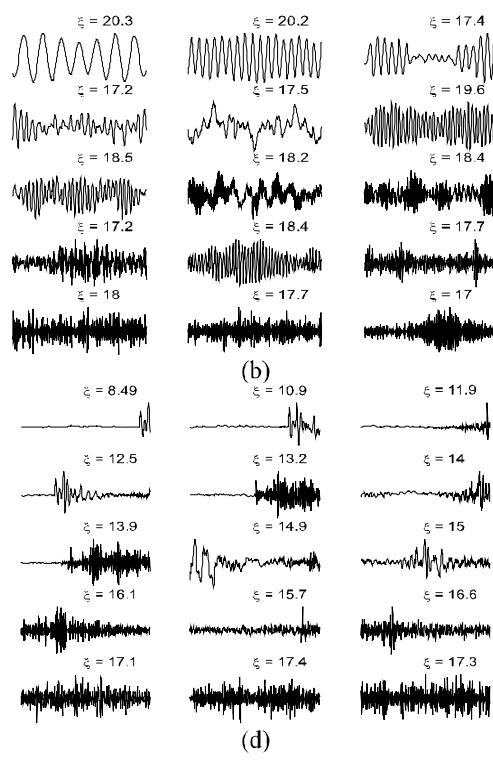
پیش فرض برای خاتمه الگوریتم در نظر گرفته شده است. در شبیه‌سازی‌های انجام شده، نتایج حاصل از این الگوریتم با روش‌های یادگیری واژه‌نامه دیگر مانند PCA و K-SVD مقایسه شده است. شاخص تُنکی نمونه‌هایی از ا تم‌های یادگرفته شده از روش‌های مختلف در شکل (۲) نشان داده شده است. همان‌طور که دیده می‌شود، ا تم‌های استخراج شده از روش PCA، خصوصیات محلی داده را به تصویر نمی‌کشند؛ یعنی تمامی قاب‌های برگزیده شده به عنوان



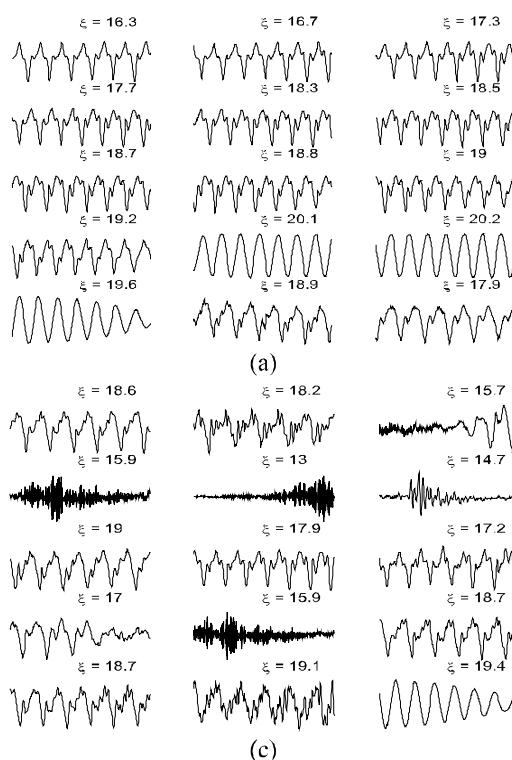
(b)

(شکل-۱): (a) مقایسه شاخص تُنکی حاصل از ا تم‌های استخراجی از الگوریتم GAD، PCA و K-SVD. (b) مقایسه خطای بازسازی در الگوریتم K-SVD و PCA، GAD

(Figure-1): a) Comparison of sparsity index of speech frames extracted from GAD, PCA and K-SVD algorithms. b) Comparison of reconstruction error for GAD, PCA and K-SVD algorithms.



(b)



(c)

(شکل-۲): (a) نمونه‌هایی از قاب‌های سیگنال گفتار. نمونه‌هایی از ا تم‌های آموخته دیده با: (b) PCA. (c) K-SVD. (d) GAD algorithms. (Figure-2): a) The selected frames of speech signal. The examples of the learned atoms using: b) PCA. c) K-SVD. d) GAD algorithms.

مطابق شکل (۲)، اتمهای یادگیری شده از قابهای سیگنال گفتار با الگوریتم GAD، در ابتدا شاخص تُنکی کمی دارند و با یافتن اتمهای جدید در هر مرحله مقدار این شاخص افزایش پیدا کرده و اتمهای با تُنکی کمتر (نوفه‌گونه) به دست می‌آیند. پس با انتخاب تعداد مناسبی از اتمهای اولیه می‌توان به نتایج بهسازی مناسبی دست یافت. نتایج حاصل برتری الگوریتم مبتنی بر داده GAD نسبت به روش PCA و K-SVD در بهبود نسبت سیگنال به نوفه^۱ را نشان می‌دهد.

۳- روش پیشنهادی

در روش بهسازی گفتار به‌کمک واژه‌نامه مبتنی بر داده، حذف اتمهای واژه‌نامه بدون توجه به این مسئله صورت می‌گیرد که ماهیت قابهای بی‌واک گفتار می‌تواند بسیار مشابه با قابهای نوفه‌ای و با تُنکی بسیار پایین باشد. این قابها در فرآیند بهسازی به‌کمک نمایش تُنک سیگنال گفتار نادیده گرفته شده و حذف می‌شوند؛ بنابراین سیگنال گفتار تخمینی با اعوجاج منبع حاصل می‌شود. در این صورت می‌باشد از قابهای ابتدایی که به طور معمول گفتاری نیستند و تنها می‌توانند مربوط به نوفه باشند، محدوده پارامتر تُنکی قابهای نوفه‌ای را به دست آورده و در چینش اتمهای واژه‌نامه مبتنی بر داده این مسئله را در نظر گرفت. تنها اتمهایی که تُنکی در محدوده تُنکی به دست آمده دارند، در ستون‌های انتهایی واژه‌نامه قرار می‌گیرند تا در تخمین سیگنال تمیز نادیده گرفته شده و حذف شوند؛ همچنین به‌منظور بالا بردن دقت در انتخاب اتمهای واژه‌نامه می‌توان در کنار معیار تُنکی^۲ تعريف شده در روش GAD، از معیارهای تُنکی دیگری نیز بهره برد.

گروه پردازش سیگنال تُنک^۳ یک تیم پژوهشی در دانشگاه دوبلین ایرلند است که در سال ۱۹۹۷ شروع به کار کرده‌اند. این گروه بر روی کاربردهای آنالیز زمان-فرکانس و زمان-مقیاس و نمایش‌های تُنک سیگنال و جداسازی منابع با استفاده از این نمایش‌های تُنک تمرکز کرده‌اند و در نهایت به معروفی چند معیار تُنکی پرداختند [۲۹-۲۷]. برای بررسی میزان تُنکی سیگنال گفتار در حوزه‌های مختلف، ابتدا باید به بیان ویژگی‌های لازم برای یک معیار اندازه‌گیری تُنکی و سپس به معرفی معیارهای مناسب پرداخت. این ویژگی‌ها

توسط دالتون در [۳۰] مطرح شده است. نکته قابل توجه آن است که یک معیار تُنکی می‌باشد به صورت مجموع وزن یافته‌ای از همه ضرایب باشد تا تغییر در یک ضریب، در اندازه تُنکی تأثیر گذارد. به این منظور باید مؤلفه‌های کوچک نسبت به مؤلفه‌های بزرگ دارای وزن بیشتری باشند. ویژگی دیگری که یک معیار تُنکی باید داشته باشد، نرمالیزه‌بودن آن است. بدین معنا که یک مجموعه از ضرایب را تنها به این دلیل که تعداد ضرایب بیشتر یا کمتر نسبت به مجموعه دیگر دارد، نباید تُنک‌تر از دیگری دانست؛ بنابراین دو وضعیت برای نرمالیزه‌کردن می‌باشد وجود داشته باشد: نخست این که یک معیار تُنکی بایستی که به مقدار نسبی ضرایب به صورت تابعی از مقدار کلی وابسته باشد و دوم این که مقدار حاصل از معیار تُنکی مستقل از تعداد ضرایب باشد تا بتوان تُنکی مجموعه‌هایی با اندازه‌های مختلف را با هم مقایسه کرد. همچنین نشان داده شده است که تنها دو معیار جینی^۴ و هایر^۵ این دو خاصیت را برآورده می‌سازند [۳۰].

همچنین تنها معیاری که تمام ویژگی‌های شش گانه دالتون را دارد معیار جینی است. برای یک رشته N تایی از ضرایب $[C_1 C_2 C_3 \dots C_n] = \vec{C}$ این دو معیار به صورت زیر بیان می‌شوند [۳۰].

$$G = \left(\sqrt{N} - \frac{\sum_j c_j}{\sqrt{\sum_i c_i^2}} \right) (\sqrt{N} - 1)^{-1} \quad (10)$$

$$Gini = 1 - 2 \sum_{j=1}^N \frac{c_j}{\|\vec{C}\|_1} \left(\frac{N-j+1}{N} \right) \quad (11)$$

با استفاده از نتایج گزارش شده در [۲۸-۲۷]، استفاده از این معیارها به جای معیار تُنکی می‌تواند نتایج مطلوبی در یادگیری واژه‌نامه مبتنی بر داده به دست دهد؛ بنابراین با توجه به ویژگی‌های این روش یادگیری واژه‌نامه، اعم از داشتن سرعت بالا در روال یادگیری و آموزش اتمهای واژه‌نامه که متناسب با ساختار سیگنال طراحی می‌شوند و همدوسي بالامي با رده داده آموزش دارند، بر آن شدیم تا با انتخاب یک معیار تُنکی مناسب، کارایی الگوریتم را در حذف نوفه بهبود دهیم.

هر چقدر مقدار شاخص تُنکی جینی بیشتر باشد به معنای تُنکی بیشتر در فریم داده خواهد بود و بالعکس. در حالی که در رابطه تُنکی بر مبنای نرم تعریف شده در رابطه (۳)، هرچه میزان این شاخص تُنکی بیشتر باشد

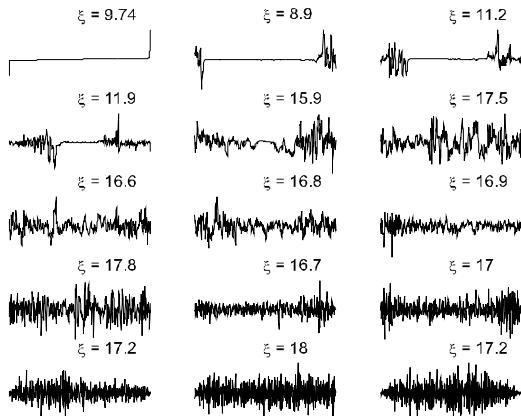
³ Gini

⁴ Hoyer

^۱ Signal to noise ratio(SNR)

^۲ Sparse signal processing team

تُنکی قاب‌هایی اولیه سیگنال مشاهده‌ای که ماهیت نویه‌ای دارند در انتخاب ستون‌های واژه‌نامه مبتنی بر داده بسیار مؤثر خواهد بود. همان‌طور که دیده می‌شود، قاب‌های استخراج شده براساس بیشینه میزان تُنکی تعیین شده از روابط (۳) و (۱۲) به ترتیب در شکل (۲) و شکل (۳) بسیار مشابه هستند؛ اما در قاب‌های استخراج شده براساس کمینه میزان تُنکی که ستون‌های ابتدایی در واژه‌نامه مبتنی بر داده را تعیین می‌کند، اندکی تفاوت دارند.



(شکل-۳): نمونه‌هایی از اتم‌های آموزش دیده با الگوریتم .GAD_Gini

(Figure-3): The examples of the learned atoms using GAD_Gini algorithms.

درواقع قاب حاصل از رابطه پیشنهادی که بر مبنای معیار جینی پایه‌ریزی شده تُنک‌تر از قابی خواهد بود که براساس تُنکی مبتنی بر نرم به دست آمده است. این نتیجه‌گیری به طور تقریبی در بیش‌تر داده‌های گفتاری متعلق به گویندگان زن و مرد مشاهده شده است.

در مورد نحوه انتخاب اتم‌های ابتدایی ماتریس واژه‌نامه نیز همان‌طور که بیان شد از اطلاعات حاصل از قاب‌های ابتدایی سیگنال گفتار نویه‌ای ورودی استفاده می‌شود. به این معنی که تنها قاب‌هایی از داده در ابتدای واژه‌نامه قرار می‌گیرند که دارای بیشترین نرخ تُنکی براساس رابطه (۱۲) باشند و این مقدار تُنکی حاصل، خارج از بازه نرخ تُنکی به دست آمده از قاب‌های ابتدایی سیگنال گفتار باشد. در هر تکرار الگوریتم برای دست‌یابی به اتم‌های جدید در واژه‌نامه نیز وضعیت بازنمایی تُنک قاب‌های سیگنال ورودی علاوه‌بر مقدار نرخ تُنکی محاسبه شده بر روی واژه‌نامه بررسی می‌شود و قاب‌هایی در ابتدای واژه‌نامه قرار می‌گیرند که علاوه‌بر داشتن نرخ تُنکی بالا دارای انرژی

به معنای سیگنال با تُنکی کمتر خواهد بود. شاخص تُنکی پیشنهادشده در این بخش براساس معیار جینی و به صورت زیر تعریف می‌شود:

$$\xi_{\text{Gini}} = \max \frac{\|x_k\|_1}{\text{Gini}(\|x_k\|)}$$
 (۱۲)

k در این رابطه، شماره قاب سیگنال ورودی است. با توجه به شاخص تُنکی معرفی شده، در صورتی که قاب داده تُنکی زیادی داشته باشد، مانند قاب‌های گفتاری، مقادیر کم و در صورتی که قاب داده تُنکی کمی داشته باشد، مانند قاب‌های نویه‌ای مقادیر بالاتری را انتخاب می‌کند. نتایج حاصل از این معیار با نام GAD_Gini در نظر گرفته می‌شود؛ همچنین مورد دیگر در روش پیشنهادی این است که با توجه به بازه شاخص تُنکی به دست آمده از قاب‌های اولیه سیگنال مشاهده‌ای که مربوط به قاب‌های نویه‌ای خواهد بود، قاب‌های با شاخص تُنکی در این محدوده در ستون‌های ابتدایی واژه‌نامه که متناظر با قاب‌های گفتاری است قرار نمی‌گیرند و به انتهای جدول هدایت می‌شوند. نمونه‌هایی از اتم‌های آموزش دیده با الگوریتم GAD_Gini در شکل (۳) نشان داده شده است. همچنین، قاب‌های داده انتخاب شده با بیشترین و کمترین مقادیر شاخص تُنکی معرفی شده در روابط (۳) و (۹) در شکل (۴) نشان داده شده است.

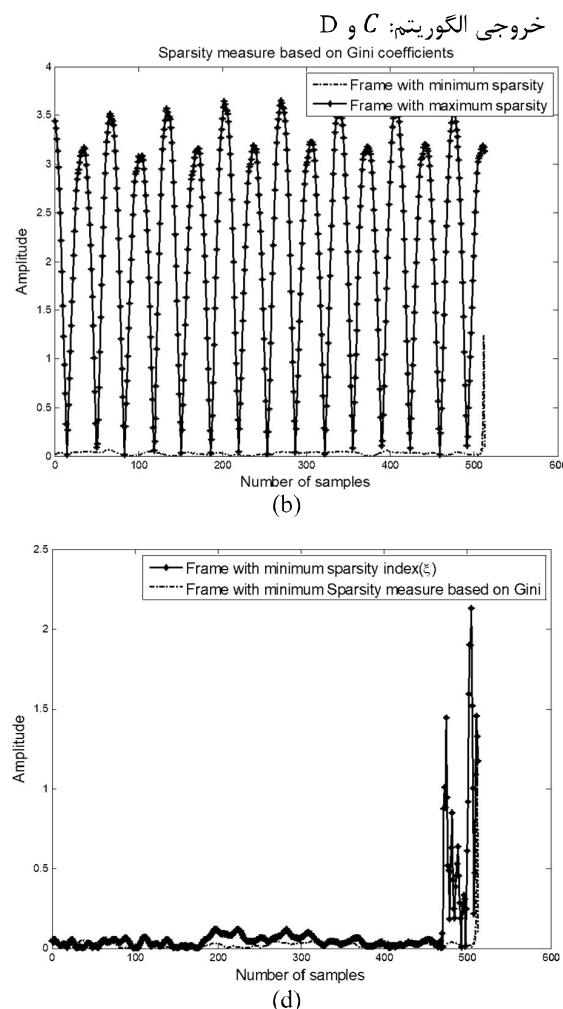
قابل ذکر است که نتایج حاصل از معیار هایر به علت مطلوب‌نبودن سیگنال بهسازی شده نتیجه در شکل‌ها و جداول آورده نشده است. با به کارگیری این شاخص تُنکی جدید مبتنی بر معیارهای دالتون در روال الگوریتم GAD، نتایج مناسبی در بهسازی گفتار حاصل می‌شود که به علت خصوصیات مهم این شاخص است که در ادامه ذکر می‌شود:

- ۱) شاخص جینی به صورت مجموع وزن یافته‌های از همه ضرایب موجود محاسبه می‌شود تا تعییر در یک ضریب حتی کوچک در اندازه تُنکی کل تأثیر بگذارد. بنابراین مؤلفه‌های کوچک در آن نسبت به مؤلفه‌های بزرگ دارای وزن بیشتری هستند.

- ۲) این شاخص نسبت به تعداد ضرایب نرمالیزه شده است تا یک مجموعه از ضرایب را تنها به این دلیل که تعداد ضرایب بیشتر یا کمتر نسبت به مجموعه دیگر دارد یا ضرایب آن خیلی بلند یا خیلی کوتاه است، تُنک‌تر از دیگری تعیین نکند.

- ۳) مقدار این شاخص برای بخش‌های نویه‌ای متماز از بخش‌های گفتاری است؛ بنابراین در نظر گرفتن شاخص

وروودی الگوریتم: X , $[.]D$ (ماتریس تهی اولیه واژه‌نامه) و محدوده تغییرات نرخ تُنکی برای قاب‌های سکوت اولیه سیگنال گفتار (ε_N)



بازنمایی تُنک مناسبی بر روی اتمهای واژه‌نامه حاصل باشند. روال بهمساری پیشنهادی در ادامه بیان شده است:

(شکل-۴): a) نمایش قاب‌های با کمینه و بیشینه مقدار تُنکی انتخاب شده براساس شاخص تُنکی مبتنی بر نرم. b) نمایش قاب‌های با کمینه و بیشینه مقدار تُنکی انتخاب شده براساس رابطه پیشنهادی. c) نمایش قاب‌های با بیشینه مقدار تُنکی انتخاب شده براساس شاخص تُنکی مبتنی بر نرم و رابطه پیشنهادی. d) نمایش قاب‌های با کمینه مقدار تُنکی انتخاب شده براساس براساس شاخص تُنکی مبتنی بر نرم و رابطه پیشنهادی.

(Figure-4): a) The representation of speech frames with minimum and maximum sparsity values selected based on the norm-based sparsity index. b) The representation of speech frames with minimum and maximum sparsity values selected based on the proposed sparsity index. c) The representation of speech frame with maximum sparsity values selected based on the norm-based sparsity index and the proposed sparsity index. d) The representation of speech frame with minimum sparsity values selected based on the norm-based sparsity index and the proposed sparsity index.

۱) محاسبه ξ_{Gini} برای تمامی $K-t$ قاب سیگنال گفتار نوشهای

۲) انتخاب قاب با بیشترین تُنکی ξ_{max}

۳) اگر ξ_{max} خارج از محدوده ε_N باشد و انرژی بازنمایی برای قاب انتخاب شده بر روی اتمهای واژه‌نامه که ماهیت گفتاری دارند، به اندازه کافی زیاد باشد، یعنی:

$$\|C\|_2 > 0.4\|X\|_2 \leftarrow (X_k)_{\xi_{\text{max}}} = D_t \cdot C$$

* در گام اول $t=1$

۱) محاسبه نرخ تُنکی ξ_{Gini} از رابطه (۱۲) برای تمامی K

قاب سیگنال گفتار نوشهای

$X_k, 1 \leq k \leq K$

۲) انتخاب قاب با بیشترین تُنکی ξ_{max} و انتقال آن به

نخستین ستون ماتریس واژه‌نامه

$D_1 \leftarrow (X_m)_{\xi_{\text{max}}}$ *** شروع حلقه یادگیری واژه‌نامه مبتنی بر داده $t \neq 1$



به منظور بررسی شاخص تُنکی، میانگین مقدار این معیار در الگوریتم‌های مورد بررسی برای اتم‌های واژه‌نامه و ماتریس ضایعات تُنک C حاصل از روال یادگیری، برای گوینده‌های زن و مرد محاسبه شده و در جدول (۲) نشان شده است. تعداد اتم‌ها در تمامی روش‌ها یکسان و برابر با ۵۱۲ و مقدار پارامتر تُنکی در الگوریتم K-SVD برابر با ۰ است. در نظر گرفته شده است. در این نتایج، نتایج حاصل از الگوریتم GAD ارائه شده در [23] با اتم‌های واژه‌نامه مبتنی بر داده انتخاب شده بر مبنای شاخص تُنکی پیشنهادی براساس معیار جینی، با نام GAD_Gini بیان شده است. این نتایج حاصل میانگین‌گیری بر روی صد سیگنال گفتاری انتخاب شده از مجموعه دادگان GRID است. نتایج به دست آمده نشان می‌دهد که اتم‌های حاصل از الگوریتم‌های GAD مبتنی بر معیار تُنکی مبتنی بر تُرم داده و معیار پیشنهادی، تُنکتر از قاب‌های سیگنال داده اصلی هستند. همچنین اتم‌های حاصل از روش GAD_Gini اندکی تُنکتر از اتم‌های حاصل از روش GAD به دست آمده است که این نتایج در مورد ماتریس ضایعات نیز برقرار خواهد بود. گفتنی است که در این بخش از بیان نتایج حاصل از دیگر روش به دلیل پایین‌بودن مقادیر حاصل خودداری شده است. شاخص تُنکی برای تمامی اتم‌ها در واژه‌نامه یادگیری شده به روش‌های مختلف در شکل (۵) نشان داده شده است. نتایج در این شکل در $\text{SNR}=0$ حاصل شده است. همان‌طور که از پیش بیان شد هر چه شاخص تُنکی در یک اتم کوچک‌تر باشد، آن اتم تُنکتر خواهد بود و به منظور به کار گیری از واژه‌نامه یادگیری شده در بهسازی گفتار می‌بایست اتم‌ها در واژه‌نامه مبتنی بر داده از قاب‌های شبه‌گفتار با تُنکی بالا (مقدار شاخص تُنکی کم) تا قاب‌های شبه‌نوغه با تُنکی کم (مقدار شاخص تُنکی بالا) مرتب شوند. نتایج حاصل از این شکل نشان می‌دهد که شاخص تُنکی در بیشتر اتم‌های یادگیری شده در روش پیشنهادی GAD_Gini پایین‌تر از سایر روش‌ها و داده اصلی به دست آمده است که نشان می‌دهد تُنکی اتم‌های حاصل بیشتر از داده گفتاری خواهد بود. همچنین می‌توان مشاهده کرد که در الگوریتم GAD و GAD_Gini، مقدار تُنکی به ترتیب بعد از تعداد اتم‌های برابر با ۲۰۰ و ۲۵۰ به مقدار تُنکی پیشنهاد نزدیک است و بنابراین این مقدار می‌تواند آستانه‌ای برای جداسازی قاب‌ها باشد که آگاهی از این مقدار آستانه در روال بهسازی گفتار بسیار مورد استفاده قرار می‌گیرد. در شکل (۶) شاخص تُنکی اتم‌ها به کمک الگوریتم یادگیری واژه‌نامه پیشنهادی و نیز به کار گیری بخش‌های آن شامل اعمال محدوده بر شاخص

قابل انتخاب شده به ستون بعدی ماتریس واژه‌نامه اختصاص می‌یابد:

$$D_t \leftarrow (X_k)_{\epsilon_{max}}$$

(۴) اگر تعداد ستون‌های کافی برای واژه‌نامه حاصل نشده است، حلقه یادگیری تکرار شود و $t=t+1$.

به این صورت در هر گام یک اتم به ماتریس واژه‌نامه اضافه شده و در نهایت چینش اتم‌ها تکمیل می‌شود. مقدار آستانه به منظور بررسی انرژی بازنمایی قاب ورودی بر روی واژه‌نامه و قرار گرفتن آن قاب به عنوان ستون جدید واژه‌نامه براساس مقایسه آن با مقدار چهل درصد از نرم دوم ماتریس قاب‌های ورودی X در گام سوم خواهد بود که این میزان به صورتی تجربی انتخاب شده است. در گام بهسازی و پس از تکمیل واژه‌نامه، تنها بخشی از اتم‌های چیده شده در ستون‌های ابتدایی در نظر گرفته می‌شوند و ماقبی اتم‌ها که ساختار نوغه را بازنمایی می‌کنند در روال دستیابی به سیگنال بهسازی شده نادیده گرفته می‌شوند و ماتریس واژه‌نامه \widehat{D} حاصل می‌شود. به صورت تجربی هفتاد درصد ستون‌ها به منظور بازنمایی سیگنال گفتار مورد استفاده قرار می‌گیرند و از ضرب این درصد اتم‌ها در ماتریس بازنمایی تُنک مختصر شده $\widehat{\mathbf{t}}$ ، سیگنال گفتار حذف نوغه شده \widehat{X} حاصل می‌شود.

۴- نتایج شبیه‌سازی

داده گفتاری مورد نیاز از مجموعه دادگان GRID برای گویندگان زن و مرد انتخاب شده است [26]. نرخ نمونه‌برداری این داده برابر با شانزده کیلوهرتز تنظیم شده است. در ابتدا سیگنال گفتار (t) به قاب‌های با طول L که به اندازه G نمونه همپوشانی دارند، تقسیم‌بندی می‌شوند. **K**امین قاب به صورت زیر به دست می‌آید:

$$x_k = [x((k-1)(L-G)+1), \dots, x(kl-(k-1)G)]^T \quad (13)$$

که در آن $\{1, \dots, K\}$ شماره قاب مورد بررسی است. سپس از قاب‌های ساخته شده ماتریس $X \in R^{L \times K}$ تشکیل می‌شود که **K**امین ستون با قاب داده $x_{l,k}$ مشخص و **I**, **k**امین ستون به صورت زیر خواهد بود:

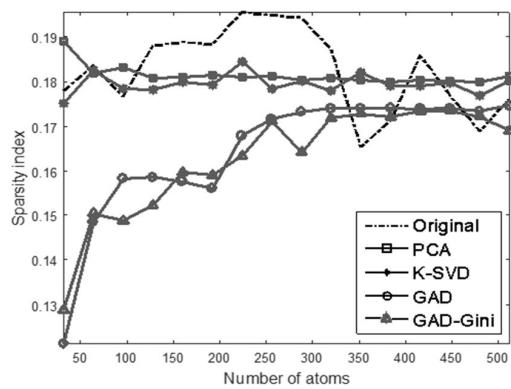
$$[X]_{l,k} = X(l+(k-1)(L-G)) \quad (14)$$

که در آن $\{1, \dots, L\}$ و $l \in \{1, \dots, K\}$ است. در این بخش، مقدار L و G به ترتیب برابر با ۵۱۲ و ۱۲ در نظر گرفته شده که بیان گر قاب‌های ورودی با طول ۳۲ میلی ثانیه و با همپوشانی ۷۵٪ میلی ثانیه است.

بهازی SNR=0 نمایش داده شده است. همان‌طور که دیده می‌شود اتم‌های یادگیری شده براساس الگوریتم پیشنهادی GAD_Gini به خوبی توانسته است، فضای زمان-فرکانس و مشخصات آن را بازنمایی کند.

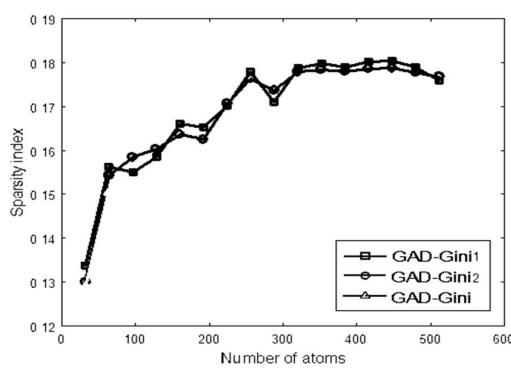
(جدول-۲): میانگین شاخص تُنکی در الگوریتم‌های مختلف برای اتم‌ها و ماتریس ضرایب تُنک گوینده‌های مختلف (Table-2): The average values of atom sparsity index and sparsity for coefficient matrix of different algorithms

Method	Atom sparsity index		Sparsity of coefficient matrix	
	Female	Male	Female	Male
Speech data	0.1741	0.1673	0.1741	0.1673
PCA	0.1722	0.1731	0.1527	0.1597
K-SVD	0.1643	0.1724	0.1336	0.1387
GAD	0.1531	0.1641	0.1371	0.1357
GAD_Gini	0.1476	0.1431	0.1107	0.1202



(شکل-۵): بررسی شاخص تُنکی برای اتم‌ها در واژه‌نامه‌های یادگیری شده با روش‌های مختلف

(Figure-5): Plot of sparsity index of atoms in the learned dictionaries of different algorithms



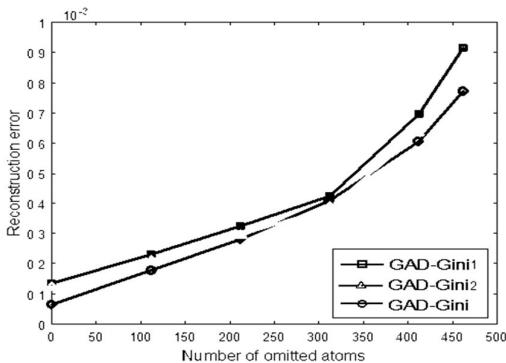
(شکل-۶): بررسی شاخص تُنکی اتم‌ها در واژه‌نامه‌های یادگیری شده با روش پیشنهادی و به کارگیری بخش‌های آن

(Figure-6): Plot of sparsity index of atoms based on the proposed dictionary learning algorithm and its different steps

تُنکی به کمک قاب‌های ابتدایی و نیز به کارگیری شاخص Gini نشان داده شده است. الگوریتم پیشنهادی به کمک اعمال محدوده بر شاخص تُنکی با نام GAD_Gini1 و الگوریتم پیشنهادی به کمک به کارگیری شاخص تُنکی Gini با نام GAD_Gini2 درنظر گرفته شده است. نتایج به دست آمده نشان می‌دهد که هر یک از این دو گام سهمی در دقت انتخاب تُنکترین قاب‌های وروودی داشته‌اند و در کنار یکدیگر توانسته‌اند، نتیجه مطلوبی در این راستا بدست دهند. بهمنظور بررسی بیشتر عملکرد، خطای بازسازی بهازی حذف تعداد مختلفی از اتم‌های در هر یک از روش‌ها محاسبه شده و در شکل (۷) نمایش داده شده است. نتایج این شکل در SNR=0 به دست آمده است. محدوده این تعداد اتم حذف شده از صفر تا ۴۶۰ اتم در نظر گرفته شده که به معنی به کارگیری ۵۱۲ تا ۵۲ اتم است. همان‌طور که مشاهده می‌شود نرخ تغییر خطای بازسازی تمامی الگوریتم‌ها با بالارفتن تعداد اتم‌های حذف شده افزایشی است. همچنین، خطای بازسازی در الگوریتم K-SVD حتی بدون حذف هیچ یک از اتم‌ها غیرصفر است. الگوریتم PCA کمترین نرخ خطای بازسازی را در استفاده از هر تعداد دلخواه اتم بدست آورده است، زیرا مؤلفه‌های سیگنال در این تبدیل به گونه‌ای مرتباً می‌شوند که بیشترین انرژی در تعداد بسیار کمی از مؤلفه‌ها متمرکز شود و این مؤلفه‌ها همان اتم‌هایی هستند که زودتر استخراج می‌شوند. همچنین مشاهده می‌شود که نتایج حاصل از الگوریتم GAD_Gini با تعداد اتم‌های حذف شده کمتر مناسب‌تر از الگوریتم بر مبنای GAD است و از آنجایی که مواردی با تعداد اتم‌های حذف شده زیاد به‌طور معمول رخ نمی‌دهد، ابتدای این نمودار اهمیت بیشتری داشته و کارایی روش مبتنی بر معیار GAD_Gini مناسب‌تر خواهد بود. نتایج حاصل از خطای بازسازی بهازی حذف تعداد مختلفی از اتم‌ها به کمک الگوریتم یادگیری واژه‌نامه پیشنهادشده و همچنین به کارگیری گام‌های مختلف این روش در شکل (۸) نشان داده شده است. الگوریتم پیشنهادی با به کارگیری محدوده شاخص تُنکی با نام GAD_Gini1 و الگوریتم پیشنهادی با GAD_Gini2 به کارگیری شاخص تُنکی Gini با نام GAD_Gini به این دو گام توانسته است، خطای بازسازی را کاهش دهد و اعمال هم‌زمان این دو گام و در روش پیشنهادی منجر به کاهش خطای بازسازی محسوسی شده است.

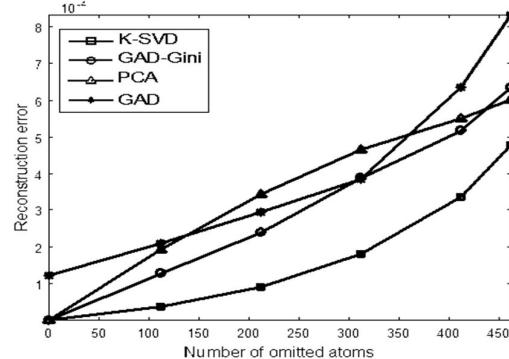
نمودار مشخصه زمان-فرکانس اتم‌های آموزش دیده به کمک روش‌های مختلف یادگیری واژه‌نامه در شکل (۹) و





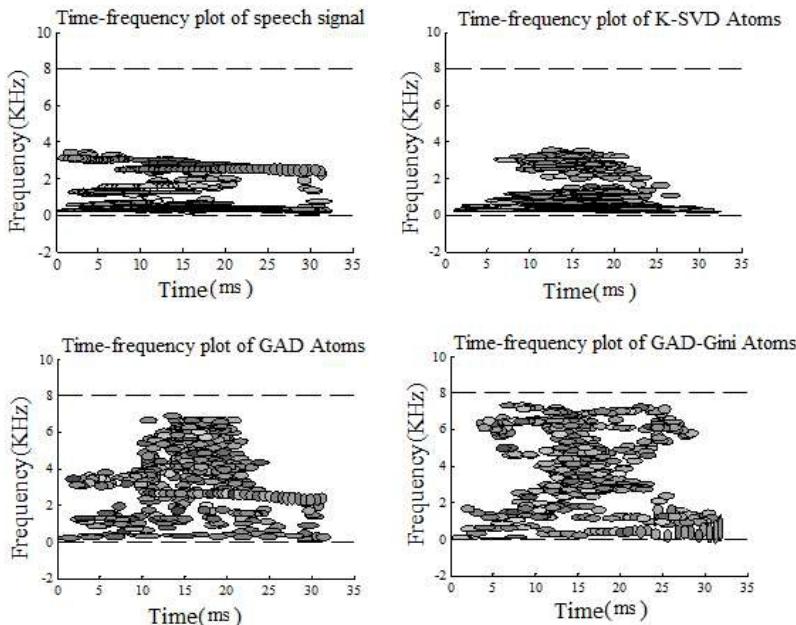
(شکل-۸): مقایسه خطای بازسازی بهازای حذف تعداد مختلفی برای واژه‌نامه‌های یادگیری شده با روش پیشنهادی و به کار گیری بخش‌های آن

(Figure-8): Comparison of the reconstruction error of sparse representation by removing different number of atoms based on the proposed dictionary learning algorithm and its different steps



(شکل-۷): مقایسه خطای بازسازی بهازای حذف تعداد مختلفی از اتم‌های واژه‌نامه در روش‌های مختلف

(Figure-7): Comparison of the reconstruction error of sparse representation by removing different number of atoms



(شکل-۹): نمایش بازنمایی مشخصات زمان-فرکانس توسط اتم‌های واژه‌نامه آموزش دیده به کمک الگوریتم‌های مختلف
(Figure-9): Plot of time-frequency structure of the atoms learned using different methods

$$ISNR = 10 \log \frac{E\{(x(t)-y(t))^2\}}{E\{(x(t)-\hat{x}(t))^2\}} \quad (15)$$

که در آن $x(t)$ و $y(t)$ به ترتیب سیگنال داده اصلی، سیگنال نوفه‌ای مشاهده شده و سیگنال بهسازی شده هستند. همان‌طور که قابل نتیجه‌گیری است، هرچه سیگنال بهسازی شده به سیگنال اصلی شبیه‌تر باشد، مقدار این معیار بزرگ‌تر خواهد بود؛ همچنین مقدار این معیار در تمامی روش‌های مورد بررسی به جز K-SVD، هنگامی که از تمامی اتم‌ها در روال بازسازی استفاده شود، مقدار صفر خواهد داشت و بازسازی کامل خواهد بود؛ اما در

امهای یادگیری شده توسط GAD نیز تا حد مطلوبی قادر به بازنمایی کل فضای زمان-فرکانس خواهد بود. در جدول (۳) نتایج حاصل از بهسازی گفتار در حضور نوفه سفید گوسی برای روش‌های مختلف آورده شده است. نتایج گزارش شده حاصل میانگین گیری بر روی صد سیگنال گفتاری منتخب از مجموعه دادگان است.

این نتایج به کمک معیار عینی بهبود نسبت سیگنال به نوفه^۱ در سطوح نوفه ۰ dB، +۱۰ dB و -۱۰ dB به دست آمده است. این معیار به صورت زیر بیان می‌شود:

^۱ Improvement in signal to noise ratio(ISNR)

سایر شرایط که تعدادی از اتمها در نظر گرفته نمی‌شوند، SVD بهترین نتایج را در همه سطوح نویه به دست آورده است. در تمامی روش‌ها می‌توان نتیجه گرفت که هرچه

(جدول-۳): نتایج حاصل از معیار ISNR در حضور نویه سفید گوسی در SNR‌های مختلف

(Table-3): The results of ISNR value in the presence of white Gaussian noise in different SNR values

SNR	Method	Number of atoms in dictionary				
		512	400	300	100	50
-10dB	PCA	0	0.48	1.44	2.37	4.69
	K-SVD	5.41	6.08	6.52	6.41	5.29
	GAD	0	1.27	2.94	4.33	4.98
	GAD Gini	0	2.29	3.97	4.87	5.01
	OMLSA			4.71		
0dB	PCA	0	0.46	1.31	2.71	5.41
	K-SVD	5.69	6.91	7.18	7.57	8.53
	GAD	0	1.31	2.69	5.22	7.24
	GAD Gini	0	2.43	3.56	6.42	8.07
	OMLSA			7.24		
-10dB	PCA	0	0.51	1.31	2.72	5.48
	K-SVD	5.82	6.75	7.56	9.91	12.51
	GAD	0	1.31	2.84	5.31	7.69
	GAD Gini	0	1.54	3.12	6.57	8.24
	OMLSA			7.93		

محدوده قاب‌های انتخابی با توجه به محدوده تُنکی قاب‌های اولیه نویه‌ای سیگنال مشاهده شده است. همچنین همان‌طور که انتظار می‌رود تعیین تعداد اتم‌های تشکیل‌دهنده واژه‌نامه و حذف اتم‌های بازنمایی‌کننده ویژگی‌های نویه‌ای بسیار حائز اهمیت است و بهمین دلیل با کم‌شدن یا زیادشدن بیش از حد مطلوب تعداد اتم‌ها، نتایج بهسازی بدشت مورد تأثیر قرار می‌گیرد. گفتنی است که نتایج بهسازی الگوریتم پیشنهادی در حضور نویه صورتی اندکی بالاتر از نتایج حاصل از کاهش نویه در مقابل نویه قوه‌های است.

نتایج گزارش شده نشان می‌دهد که بهسازی حاصل از الگوریتم GAD و GAD_Gini به هم و مناسب بوده و در بیشتر موارد بالاتر از الگوریتم PCA است که در آن فضای مسأله به دو زیرفضای گفتار و نویه تجزیه می‌شود. در برخی شرایط بسته به تعداد اتم به کار گرفته شده، نتایج حاصل از الگوریتم GAD_Gini بالاتر از GAD به دست آمده است. گفتنی است که کارایی الگوریتم GAD_Gini در SNR پایین مناسب‌تر خواهد بود. نتایج حاصل از بهسازی در حضور نویه‌های صورتی و قوه‌های اندکی پایین‌تر از نتایج گزارش شده از نویه سفید گوسی بوده است که می‌تواند به این علت باشد که معیار جینی پیشنهادی بهترین انطباق را به منظور اکتساب قاب‌های تُنک براساس نویه سفید گوسی به دست می‌دهد.

در این جدول، نتایج حاصل از روش OMLSA نیز اضافه شده که یک تخمین‌گر گفتار است.تابع بهره طیفی بهینه در این روش براساس یک میانگین هندسی وزن دار مرتبط با عدم حضور گفتار به دست می‌آید [31]. این روش بهمنظور حذف نویه‌های نایاستا ارائه شده است. طیف نویه در این روش به کمک میانگین‌گیری بازگشتی توان طیف با استفاده از یک پارامتر هموارسازی که بر طبق احتمال حضور گفتار تنظیم می‌شود، تخمین زده می‌شود؛ همچنین نتایج بهسازی گفتار حاصل از روش پیشنهادی در حضور نویه‌های غیر گوسی مانند نویه صورتی و نویه قوه‌های بهترتب در جداول (۴) و (۵) گزارش شده است. انجام این شبیه‌سازی بهمنظور بررسی کارایی روش پیشنهادی در حضور نویه‌های غیر گوسی صورت گرفته است. همان‌طور که نتایج گزارش می‌دهند، می‌توان مشاهده کرد الگوریتم پیشنهادی توائسته است به مقادیر ISNR بالاتری در شرایط نویه‌ای مختلف نسبت به روش‌های پایه آموزش واژه‌نامه PCA و K-SVD، روش پایه یادگیری واژه‌نامه منطبق بر داده به کمک معیار نرم و نیز روش OMLSA که یک تخمین‌گر گفتاری بهمنظور حذف نویه‌های نایاستا است، دست پیدا کند. برتری روش پیشنهادی می‌تواند به علت اصلاحات دقیقی باشد که در روش پایه یادگیری واژه‌نامه مبتنی بر داده صورت گرفته است که شامل معرفی یک معیار تعیین تُنکی قاب‌ها براساس شاخص Gini و بررسی

^۱ Optimally-modified log-spectral amplitude



(جدول-۴): نتایج حاصل از معیار ISNR در حضور نویه صورتی در SNRهای مختلف

(Table-4): The results of ISNR value in the presence of Pink noise in different SNR values

SNR	Method	Number of atoms in dictionary				
		512	400	300	200	100
-10dB	PCA	0	0.41	1.32	2.21	4.49
	K-SVD	5.12	5.82	6.21	6.03	5.10
	GAD	0	1.14	2.81	4.20	4.82
	GAD_Gini	0	2.17	3.81	4.72	4.89
	OMLSA			4.52		
0dB	PCA	0	0.59	1.22	2.65	5.32
	K-SVD	5.03	6.83	7.02	7.69	8.46
	GAD	0	1.22	2.59	5.14	7.16
	GAD_Gini	0	2.36	3.47	6.35	7.95
	OMLSA			6.84		
-10dB	PCA	0	0.44	1.23	2.65	5.37
	K-SVD	4.92	6.68	7.47	9.85	12.39
	GAD	0	1.22	2.71	5.19	7.52
	GAD_Gini	0	1.43	2.96	6.43	8.16
	OMLSA			7.18		

(جدول-۵): نتایج حاصل از معیار ISNR در حضور نویه قهوه‌ای در SNRهای مختلف

(Table-5): The results of ISNR value in the presence of Brown noise in different SNR values

SNR	Method	Number of atoms in dictionary				
		512	400	300	200	100
-10dB	PCA	0	0.38	1.29	2.17	4.42
	K-SVD	5.08	5.73	6.15	5.89	5.03
	GAD	0	1.08	2.75	4.15	4.81
	GAD_Gini	0	2.14	3.76	4.65	4.81
	OMLSA			4.42		
0dB	PCA	0	0.52	1.18	2.60	5.31
	K-SVD	4.95	6.75	6.92	7.61	8.45
	GAD	0	1.18	2.51	5.06	7.08
	GAD_Gini	0	2.29	3.41	6.28	7.91
	OMLSA			6.77		
-10dB	PCA	0	0.39	1.14	2.72	5.34
	K-SVD	4.88	6.64	7.42	9.81	12.36
	GAD	0	1.20	2.67	5.12	7.49
	GAD_Gini	0	1.38	2.91	6.59	8.11
	OMLSA			7.14		

پیشنهادی در حضور نویه‌های غیرگوسی گزارش شده است. نتایج حاصل از این شبیه‌سازی نیز بر توانایی روش پیشنهادی در حذف نویه‌های غیرگوسی از سیگنال گفتار تأکید کرده و مؤید نتایج قبلی است. روش پیشنهادی توانسته در شرایط نویه‌ای مختلف، مقادیر PESQ بالاتری نسبت به روش‌های پایه آموزش واژه‌نامه، روال پایه یادگیری واژه‌نامه منطبق بر داده و نیز روش OMLSA بدت دهد که می‌تواند به دلیل تصحیحات صورت‌گرفته در روش پایه یادگیری واژه‌نامه مبتنی بر داده باشد که در بخش ۳ به آن‌ها اشاره شد. همان‌طور که انتظار می‌رود، تعداد اتم‌های انتخابی واژه‌نامه در این روال بهسازی نیز تأثیرگذار بوده و با افزایش تعداد اتم‌ها تا حد مناسب، نتایج بهتری حاصل می‌شود. این نتایج در حضور نویه صورتی اندکی بالاتر از نتایج حاصل از کاهش نویه در مقابل نویه قهوه‌ای حاصل شده است؛ همچنین نتایج الگوریتم پیشنهادی به کمک اعمال محدوده بر شاخص تنکی با نام GAD_Gini^۱ و الگوریتم پیشنهادی

همچنین نتایج بهسازی به دست آمده با استفاده از معیار PESQ^۲، از آنجاییکه همدوسی بالایی با امتیازهای حاصل از معیارهای قابلیت فهم و کیفیت گفتار دارد، محاسبه شده است [32]. در این آزمایش از برنامه‌های آماده در نرم‌افزار متلب^۳ که در [32] ارائه شده است، نتایج حاصل از این معیار به منظور بهسازی سیگنال در حضور نویه سفید گوسی با SNRهای مختلف در جدول (۶) نشان داده شده که حاصل میانگین‌گیری از نتایج صد سیگنال گفتاری است. همان‌طور که انتظار می‌رود با کاهش تعداد اتم‌ها در واژه‌نامه، خطای بازسازی بیشتر و معیار PESQ کاهش می‌یابد؛ همچنین نتایج معیار PESQ حاصل از بهسازی گفتار به کمک روش پیشنهادی و سایر الگوریتم‌های مورد بررسی در حضور نویه‌های صورتی و قهوه‌ای بهتر ترتیب در جداول (۷) و (۸) گزارش شده است. نتایج حاصل از معیار PESQ به منظور بررسی تکمیلی کارایی روش

¹ Perceptual evaluation of speech quality² MATLAB



به کمک روش پیشنهادی در SNR های مختلف در شکل (۱۰) نشان داده شده است. همان‌طور که انتظار می‌رود با افزایش میزان SNR، نتایج بهتری به دست می‌آید؛ همچنین در شکل (۱۱)، اسپکتروگرام سیگنال‌های بهسازی شده در SNR=+5dB برای الگوریتم‌های GAD، PCA، K-SVD و GAD_Gini به نمایش گذاشته شده است. نتایج بهسازی حاصل از این الگوریتم‌ها در SNR=0dB در شکل (۱۲) نشان داده شده است. نتایج حاصل نشان می‌دهد که الگوریتم بهسازی مبتنی بر آموزش واژه‌نامه بر مبنای داده با شاخص تُنکی مشکل از معیار Gini قادر است، نتایج بهتری را نسبت به سایر الگوریتم‌های نامبرده به دست دهد و بر نتایج گزارش شده در جداول پیشین مبنی بر برتری روش بهسازی پیشنهادی تأکید دارد.

به کمک به کارگیری شاخص تُنکی Gini با نام GAD_Gini2 در این جدول بیان شده است. نتایج حاصل نشان می‌دهد که هر یک از این دو گام به تنها ی قادر است، نتایج بهسازی را افزایش دهد و به کارگیری این دو گام در کنار یکدیگر توانایی دستیابی به نتایج مناسب در این حوزه برداشی را فراهم می‌سازد. مقدار PESQ برای سیگنال نویه‌ای نیز در آخرین سطر این جدول گزارش شده است؛ این افزایش خطای بازنمایی قاب‌های داده با کاهش میزان SNR، تشدید می‌شود؛ همچنین مشاهده می‌شود که نتایج حاصل از الگوریتم پیشنهادی بهتر از سایر روش‌های بیان شده به‌منظور حذف نویه سفید گوسی از سیگنال گفتار است. این نتایج حاصل میانگین گیری بر روی پنجاه داده آزمایش است. به‌منظور بررسی بهتر عملکرد الگوریتم بهسازی گفتار پیشنهادی، نمایش اسپکتروگرام سیگنال‌های بهسازی شده

(جدول-۶): نتایج حاصل از معیار PESQ در حضور نویه سفید گوسی و در SNR های مختلف برای تعداد متفاوت اتم‌ها

SNR →	Methods ↓	-10dB			0dB			+10dB		
		512	300	100	512	300	100	512	300	100
PCA		0.76	0.42	0.23	2.61	1.98	0.93	3.56	2.88	2.04
K-SVD		1.06	0.79	0.73	2.78	2.37	1.12	3.70	2.96	2.26
OMLSA		0.15				0.76			2.53	
GAD		1.59	1.28	0.80	3.03	2.54	1.67	3.75	3.09	2.39
GAD_Gini1		1.61	1.32	0.83	3.09	2.57	1.70	3.77	3.14	2.41
GAD_Gini2		1.63	1.35	0.85	3.12	2.59	1.73	2.81	3.17	2.45
GAD_Gini		1.65	1.39	0.87	3.17	2.61	1.78	3.84	3.21	2.48
Noisy signal			0.18			0.88			1.86	

(جدول-۷): نتایج حاصل از معیار PESQ در حضور نویه صورتی و در SNR های مختلف برای تعداد متفاوت اتم‌ها

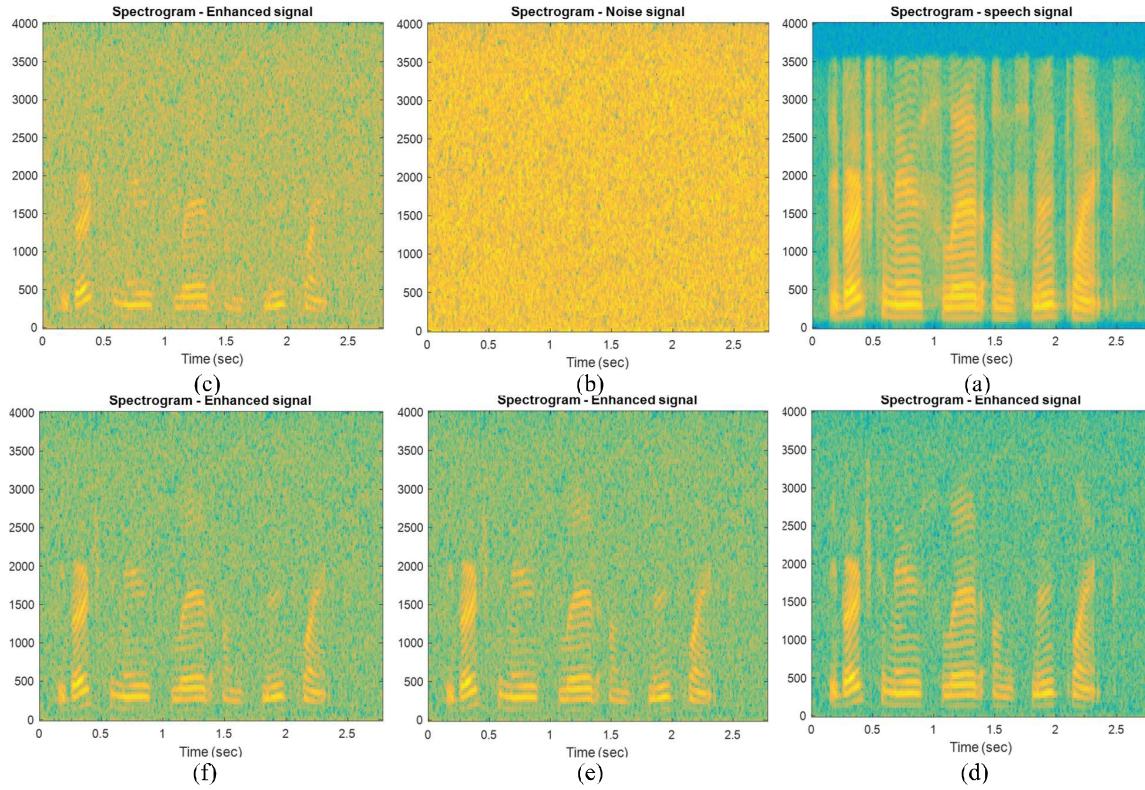
(Table-7): The results of PESQ values in the presence of Pink noise in different SNR values for various numbers of atoms

SNR →	Methods ↓	-10dB			0dB			+10dB		
		512	300	100	512	300	100	512	300	100
PCA		0.73	0.40	0.22	2.59	1.95	0.91	3.52	2.81	1.98
K-SVD		1.02	0.78	0.71	2.75	2.38	1.09	3.66	2.93	2.21
OMLSA		0.13				0.73			2.49	
GAD		1.53	1.25	0.77	3.05	2.51	1.70	3.72	3.01	2.34
GAD_Gini1		1.57	1.29	0.79	3.04	2.58	1.73	3.75	3.12	2.38
GAD_Gini2		1.61	1.32	0.80	3.09	2.55	1.71	2.78	3.15	2.39
GAD_Gini		1.66	1.37	0.85	3.15	2.63	1.76	3.82	3.19	2.45
Noisy signal			0.17			0.83			1.85	

(جدول-۸): نتایج حاصل از معیار PESQ در حضور نویه قهوه‌ای و در SNR های مختلف برای تعداد متفاوت اتم‌ها

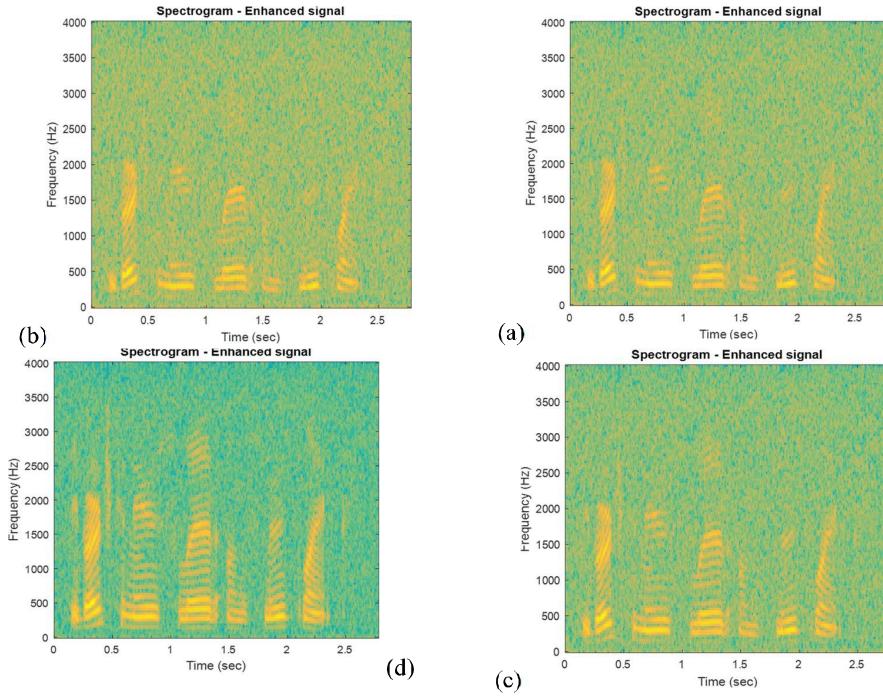
(Table-8): The results of PESQ values in the presence of Brown noise in different SNR values for various numbers of atoms

SNR →	Methods ↓	-10dB			0dB			+10dB		
		512	300	100	512	300	100	512	300	100
PCA		0.68	0.35	0.18	2.50	1.89	0.87	3.46	2.77	1.91
K-SVD		0.96	0.71	0.65	2.71	2.31	0.98	3.59	2.88	2.18
OMLSA		0.11				0.67			2.42	
GAD		1.49	1.21	0.72	2.98	2.47	1.66	3.67	2.94	2.28
GAD_Gini1		1.51	1.23	0.73	2.96	2.49	1.67	3.69	3.04	2.31
GAD_Gini2		1.55	1.27	0.76	3.01	2.49	1.68	2.73	3.07	2.31
GAD_Gini		1.61	1.31	0.79	3.09	2.58	1.70	3.78	3.14	2.39
Noisy signal			0.19			0.81			1.79	



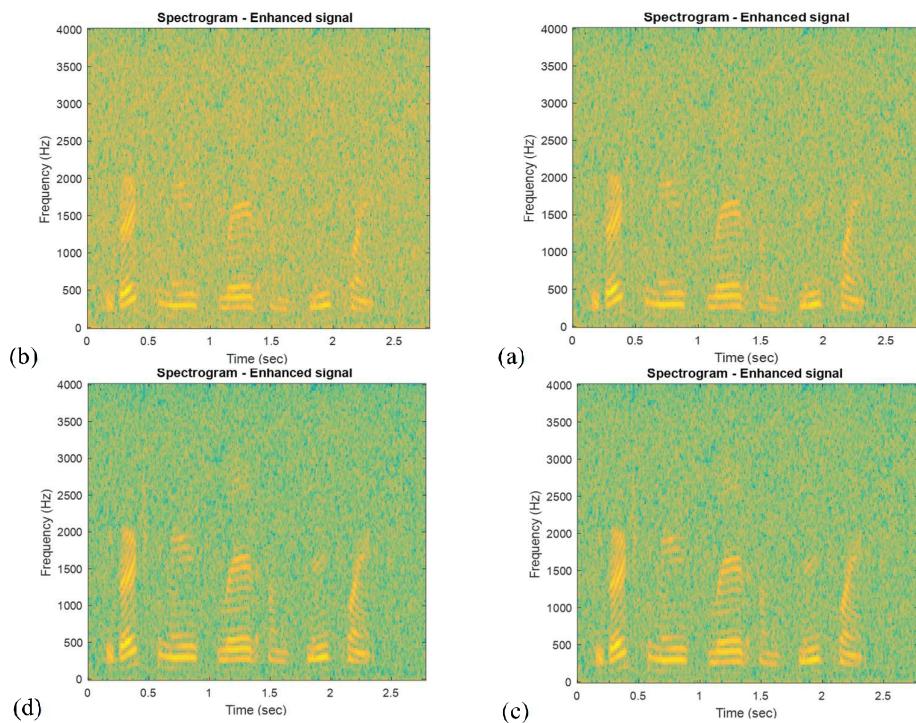
(شکل-۱۰): نمایش اسپکتروگرام برای: a) سیگنال گفتار. b) سیگنال نویه سفید گوسی. c) سیگنال گفتار نویه ای با SNR=0dB. d) سیگنال بهسازی شده در SNR=+5dB. e) سیگنال بهسازی شده در SNR=0dB. f) سیگنال بهسازی شده در SNR=-5dB.

(Figure-10): Spectrogram representation for: a) Speech signal. b) Gaussian white noise signal. c) Noisy signal. d) Enhanced signal at SNR=+5dB. e) Enhanced signal at SNR=0dB. f) Enhanced signal at SNR=-5dB.



(شکل-۱۱): نمایش اسپکتروگرام برای سیگنال بهسازی شده در SNR=+5dB برواساس: a) الگوریتم K-SVD .b) الگوریتم PCA .c) الگوریتم GAD_Gini .d) الگوریتم GAD

(Figure-11): Spectrogram representation for the enhanced signal at SNR=+5dB using: a) K-SVD algorithm. b) PCA algorithm. c) GAD algorithm. d) GAD_Gini algorithm.



(شکل-۱۲): نمایش اسپکتروگرام برای سیگنال بهسازی شده در SNR=0dB براساس: a) الگوریتم K-SVD .b) الگوریتم PCA .c) الگوریتم GAD_Gini .d) الگوریتم GAD

(Figure-12): Spectrogram representation for the enhanced signal at SNR=0dB using: a) K-SVD algorithm. b) PCA algorithm. c) GAD algorithm. d) GAD_Gini.

۵- نتیجه‌گیری

در این مقاله روش‌های یادگیری و ازنامه مبتنی بر داده معرفی شد. از این میان روش یادگیری و ازنامه تطبیقی حریصانه GAD بهمنظور بهسازی سیگنال گفتار در حضور نویه سفید گوسی مورد توجه قرار گرفت. و ازنامه در این روش بهطور مستقیم از روی سیگنال گفتار یادگیری می‌شود تا تطابق بیشتری میان اتمها و ساختار سیگنال ایجاد شده و همدوسي اتمها و رده داده افزایش یابد. انتخاب این اتم‌ها به کمک شاخص تُنکی مبتنی بر نرم از میان قاب‌های سیگنال مشاهده‌ای است. از آنجایی که این نحوه انتخاب اتم‌ها، یادگیری را درگیر روال پیچیده نکرده و سرعت هم‌گرایی بالایی به دست می‌دهد، بر آن شدید تا با بررسی این الگوریتم‌ها به نتایج بهتری در زمینه بهسازی گفتار دست پیدا کنیم. بر این اساس ایده تُنکی مورد بررسی قرار گرفت و بر مبنای معیارهای مطرح شده برای یک شاخص تُنکی مطلوب، شاخص تُنکی جینی در نظر گرفته شد. با بررسی عملکرد این شاخص در انتخاب قاب‌های داده به این نتیجه رسیدیم که عملکرد این شاخص در برخی جنبه‌ها بهتر از شاخص تُنکی مبتنی بر نرم

همچنین، زمان محاسباتی برای اجرای روش‌های مختلف مورد بررسی با میانگین‌گیری بر روی اجرای هر یک از این الگوریتم‌ها بر روی ده سیگنال ورودی در جدول (۹) گزارش شده است. تمامی برنامه‌ها بر روی نرمافزار متلب براساس ویندوز ۶۴ بیتی با مشخصات Core i5 3.2GHz CPU اجرا شده است. همان‌طور که دیده می‌شود، زمان روال اجرای روش پیشنهادی به علت مرتب‌کردن ضرایب و محاسبه معیار جینی اندکی بیشتر از الگوریتم GAD خواهد بود. الگوریتم K-SVD به علت به کارگیری روال تکرارشونده بهمنظور هم‌گرایی به کمترین خطای تقریب، زمان اجرای طولانی دارد و روش PCA نیز که تبدیل کامل است به علت نداشتن روال پیچیده محاسباتی کمترین زمان اجرا را به خود اختصاص می‌دهد.

(جدول-۹): میانگین زمان محاسباتی برای اجرای روش‌های یادگیری و ازنامه مختلف

(Table-9): The average of computation time for dictionary learning of different methods

روش	میانگین زمان اجرا (ثانیه)
PCA	10
K-SVD	6744
GAD	173
GAD-Gini	178

- information”, IEEE Trans. Inform. Theory, vol. 52, No. 2, pp. 489-509, 2006.
- [10] E. J. Candes, T. Tao, “Near-optimal signal recovery from random projections and universal encoding strategies”, IEEE Trans. Inform. Theory, vol. 52, No. 12, pp. 5406-5425, 2006.
- [11] H. Xu, “Speech enhancement based on compressed sensing technology”, Sensors & Transducers Journal, vol. 181, pp. 141-145, 2014.
- [12] G. S. Sivaram, S. K. Nemala, M. Elhilali, H. Hermansky, “Sparse coding for speech recognition”, IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 4346-4349, 2010.
- [13] D. D. Lee, H. S. Seung, “Learning the parts of objects by non-negative matrix factorization”, Nature 401, 788- 791, 1999.
- [14] N. Mohammadiha, P. Smaragdis, A. Leijon, “Supervised and unsupervised speech enhancement using nonnegative matrix factorization”, IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, No. 10, pp.2140-2151, 2013.
- [15] J. T. Geiger, J. F. Gemmeke, B. Schuller, and G. Rigoll, “Investigating NMF speech enhancement for neural network based acoustic models”, Proceedings of the Annual Conference of International Speech Communication Association (INTERSPEECH), 2014.
- [16] P. O. Hoyer, “Non-negative matrix factorization with sparseness constraints”, J. Mach. Learn. Res., vol. 5, pp. 1457-1469, 2004.
- [17] C. D. Sigg, T. Dikk, J. M. Buhmann, “Speech enhancement with sparse coding in learned dictionaries”, In Acoustics Speech and Signal Processing (ICASSP), IEEE International Conference on, pp. 4758-4761, 2010.
- [18] C. D. Sigg, T. Dikk, J. M. Buhmann, “Speech enhancement using generative dictionary learning”, IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, No. 6, pp. 1698-1712, 2012.
- [19] M. Aharon, M. Elad, A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation”, IEEE Trans. Signal Process, vol. 54, No. 11, pp. 4311-4322, 2006.
- [20] H. Yongjun, J. Han, S. Deng, T. Zheng, G. Zheng, “A solution to residual noise in speech denoising with sparse representation”, In Acoustics, Speech and Signal Processing

معرفی شده، خواهد بود. بنابراین این شاخص در روال الگوریتم یادگیری واژه‌نامه تطبیقی حریصانه مورد استفاده قرار گرفت. نتایج حاصل از شبیه‌سازی‌های مختلف نشان می‌دهد که اگرچه شاخص پیشنهادی در تمامی شرایط در نظر گرفته شده بهتر از شاخص مبتنی بر نرم عمل نمی‌کند؛ اما نتایج حاصل از آن مطلوب بوده و می‌تواند با دقت مناسبی در بسازی گفتار مورد استفاده قرار گیرد.

6- References

- [1] H. Guo, B. Zhao, G. Zhou, “Image compression based on compressed sensing theory and wavelet packet analysis”, Cross Strait Quad-Regional Radio Science and Wireless Technology Conference, vol. 2, pp. 1426-1429, 2011.
- [2] J. L. Starck, M. Elad, D. L. Donoho, “Image decomposition via the combination of sparse representation and a variational approach”, IEEE Trans. on Image Proces, vol. 14, No. 10, pp. 1570-1582 , 2005.
- [3] F. Rodriguez, G. Sapiro, “Sparse representation for image classification: Learning discriminative and reconstructive non-parametric dictionaries”, IMA Preprint, 2007.
- [4] S. Zhang, J. Huang, D. Metaxas, W. Wang, X. Huang, “Discriminative sparse representations for cervigram image segmentation”, IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp.133-136, 2010.
- [5] M. S. Lewiki, T. J. Sejnowski, “Learning overcomplete representations”, Neural Computing, vol. 12, No. 2, pp. 337-365, 2000.
- [6] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen, M. Moonen, “Retrieving sparse patterns using a compressed sensing framework: applications to speech coding based on sparse linear prediction”, IEEE Signal Processing Letters, vol. 17, No. 1, pp.103-106, 2010.
- [7] D. Wu, Z. Ping, M. N. S. Swamy, “A compressive sensing method for noise reduction of speech and audio signals”, IEEE 54th International Midwest Symposium on Circuits and Systems, pp. 1-4, 2011.
- [8] D. Wu, Z.W. Ping, M. N. S. Swamy, “On sparsity issues in compressive sensing based speech enhancement”, IEEE International Symposium on Circuits and Systems, pp. 285-288, 2012.
- [9] E. Candès, J. Romberg, T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency





سمیرا مودّتی مدارک کارشناسی و کارشناسی ارشد خود را بهترتب در سال‌های ۱۳۸۶ و ۱۳۸۹ از دانشگاه مازندران در رشته مهندسی-برق-الکترونیک دریافت کرد. ایشان درجه

دکترای خود را در سال ۱۳۹۵ از دانشگاه صنعتی امیرکبیر در رشته مهندسی برق-الکترونیک اخذ کرد. وی هم‌اکنون استادیار گروه مهندسی برق دانشکده فنی و مهندسی دانشگاه مازندران است. زمینه‌های پژوهشی مورد علاقه ایشان عبارت است از: پردازش سیگنال گفتار، پردازش سیگنال تصویر، بهینه‌سازی و هوش مصنوعی.

نشانی رایانمه ایشان عبارت است از:

s.mavadati@umz.ac.ir

سید محمد احمدی مدارک کارشناسی و کارشناسی ارشد خود را بهترتب در سال‌های ۱۲۶۳ و ۱۳۶۶ از دانشگاه صنعتی امیرکبیر در رشته مهندسی-برق-الکترونیک دریافت کرد. ایشان درجه دکترای خود را در سال ۱۳۷۵ از دانشگاه کمبریج در رشته مهندسی برق-الکترونیک اخذ نمود. وی

هم‌اکنون استاد گروه مهندسی برق دانشگاه صنعتی امیرکبیر است. زمینه‌های پژوهشی مورد علاقه ایشان عبارت است از: بازناسی سیگنال گفتار، بهسازی سیگنال تصویر.

نشانی رایانمه ایشان عبارت است از:

sma@aut.ac.ir



(ICASSP), IEEE International Conference on, pp. 4653-4656, 2012.

- [21] S. Mavadatty, S. M. Ahadi, S. Scycedin, "A novel speech enhancement method by learnable sparse and low-rank decomposition and domain adaptation", Speech Communication, vol. 76, pp. 42-60, 2016.
- [22] M. G. Jafari, M. D. Plumbley, "Speech denoising based on a greedy adaptive dictionary algorithm", 17th European Signal Processing Conference (EUSIPCO), pp. 1423-1426, 2009.
- [23] M. G. Jafari, M. D. Plumbley, "Fast dictionary learning for sparse representations of speech signals", IEEE Journal of selected topics in signal processing, vol. 5, pp. 1025-1031, 2011.
- [24] R. Rubinstein, M. Zibulevsky, M. Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation", IEEE Trans. on Signal Processing, Vol. 58, pp. 1553-1564, 2010.
- [25] M. G. Jafari, E. Vincent, S. A. Abdallah, M. D. Plumbley, M. E. Davies, "An adaptive stereo basis method for convolutive blind audio source separation", Neuro computing, vol. 71, pp. 2087-2097, 2008.
- [26] <http://www.dcs.shef.ac.uk/spandh/gridcorpus>.
- [27] N. Hurley, S. Rickard, "Comparing measures of sparsity", IEEE Trans. on Information Theory, vol. 55, pp. 4723-4741, 2009.
- [28] S. Rickard, M. Fallon, "The gini index of speech", In Proc. Of Information Sciences and Systems conference, Princeton, NJ, 2004.
- [29] P. J. Wolfe, "Sparse time-frequency representations in audio processing, as studied through a symmetrized lognormal model," In Proc. of the European Signal Processing Conference (EUSIPCO), pp. 355-359, 2007.
- [30] H. Dalton, "The measurement of the inequity of incomes", Economic Journal, Vol. 30, pp. 348-361, 1920.
- [31] I. Cohen, B. Berdugo, "Speech enhancement for non-stationary noise environments", Signal processing, vol. 81, No. 11, pp. 2403-2418, 2001.
- [32] A. Rix, J. Beerends, M. Hollier, A. Heikstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs", In Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process., pp. 749-752, 2001.