



# بهسازی گفتار به کمک یادگیری واژه‌نامه

## مبتنی بر داده

سمیرا مودّتی<sup>۱\*</sup> و سید محمد احدی<sup>۲</sup>

<sup>۱</sup>استادیار، دانشکده فنی و مهندسی، دانشگاه مازندران، بابلسر، ایران

<sup>۲</sup>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 عمده 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 حریمانه مبتنی بر داده خواهد بود که واژه‌نامه در آن، به‌طور مستقیم از روی سیگنال داده و براساس شاخص تُنکی مبتنی بر نُرم به منظور تطابق بیشتر میان اتم‌ها و ساختار داده آموزش می‌بیند. در این مقاله شاخص تُنکی جدیدی براساس معیار جینی پیشنهاد می‌شود. همچنین محدوده پارامتر تُنکی بخش‌های نوفه‌ای 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 نوفه‌ای مختلف بهتر از شاخص تُنکی مبتنی بر نُرم و سایر الگوریتم‌های پایه در این راستا است.

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

## Speech Enhancement using Adaptive Data-Based Dictionary Learning

Samira Mavaddati<sup>1\*</sup> & Seyyed Mohammad Ahadi<sup>2</sup>

<sup>1</sup>Electrical Department, Faculty of Technology and Engineering, University of Mazandaran, Babolsar, Iran

<sup>2</sup>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<sup>5</sup> 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 نوفه سیگنال صوت و گفتار بهره گرفته شده و مسأله کاهش نوفه براساس تئوری سنجش فشرده به‌صورت مسأله کمینه‌سازی نرم  $l_1$  با ترکیب خطی از قیود و عمل‌گر تبدیل فوریه جزئی تصادفی بیان شد. در ادامه، تنکی سیگنال در فضای تبدیل کسینوسی گسسته<sup>6</sup> نیز مورد بررسی قرار گرفته و معیارهایی برای تعیین میزان تنکی سیگنال ارائه شده است [8]. یکی از این معیارها، قید تنکی اکید سیگنال<sup>7</sup> 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<sup>8</sup> در

## ۱- مقدمه

بحث نمونه‌برداری فشرده<sup>1</sup> و مفاهیم مرتبط با آن مانند بازنمایی تنک<sup>2</sup> و یادگیری واژه‌نامه<sup>3</sup> 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 نوفه، قطع‌بندی، طبقه‌بندی و بازسازی است [4-1]. مطالعات به‌کمک 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<sup>4</sup> 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

<sup>5</sup> Compressive sensing

<sup>6</sup> Discrete cosine transform

<sup>7</sup> Strict sparsity

<sup>8</sup> Speech enhancement

<sup>1</sup> Compressive sampling

<sup>2</sup> Sparse representation

<sup>3</sup> Dictionary learning

<sup>4</sup> Overcomplete bases

شد [24]. تُنکی مضاعف به این معنی است که علاوه بر نمایش تُنک سیگنال، خود واژه‌نامه  $D = AB$  نیز تُنک باشد؛ یعنی اتم‌های  $A$  بر روی واژه‌نامه ثابت  $B$  بازنمایی تُنک داشته باشند که در آن  $B$  می‌تواند هر واژه‌نامه ثابت موجک یا تبدیل کسینوسی گسسته باشد. تبدیل موجک هنگامی که توابع پایه آن به‌خوبی محلی شده باشند، می‌تواند ارتباط مناسبی را میان تُنکی در تجزیه و تُنکی در واژه‌نامه به‌منظور آنالیز سیگنال‌های طبیعی مانند صوت، تصویر و سیگنال‌های پزشکی فراهم کند. همچنین در [25] نیز از سیگنال گفتار به‌صورت مستقیم به‌عنوان اتم‌های واژه‌نامه به‌کمک روش کدگذاری تُنک مبتنی بر آنالیز مؤلفه‌های مستقل<sup>۱</sup> 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 به‌علت وابستگی کم قاب‌های زمانی به‌منظور 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 برحسب 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 به‌اشتباه در اتم‌های ابتدایی و قاب‌های گفتاری نیز در انتهای واژه‌نامه قرار گرفته و حذف اتم‌های انتهایی واژه‌نامه، اعوجاج سیگنال گفتار را نتیجه دهد. در روال یادگیری واژه‌نامه GAD، این مسأله 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، به‌ترتیب 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 به‌صورت وقتی و به‌کمک استخراج پی‌درپی ستون‌های ماتریس الگوی ورودی  $X$  آموزش می‌بیند. استخراج ستون جدید در هر تکرار منوط به یافتن ستونی از  $X$  است که شرط زیر را برآورده کند:

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

<sup>۱</sup> 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 به‌منظور تعیین مناسب‌بودن قاب انتخابی به‌منظور بازنمایی کردن محتوای سیگنال وجود ندارد. از آنجایی که هدف در این روش حذف نوفه است و اتم‌های انتهایی واژه‌نامه می‌بایست ساختار نوفه را به‌درستی بازنمایی کنند، بنابراین دخیل کردن پارامتر دیگری در انتخاب بهترین اتم‌ها با ماهیت نوفه‌ای مهم به نظر می‌رسد. در روش پیشنهادی، راه‌حلی به‌منظور مرتفع‌کردن این مشکل ارائه می‌شود. همچنین یک معیار جدید تُنکی معرفی می‌شود که می‌تواند بیان درستی از تُنکی سیگنال را به‌دست 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 به‌دست‌آمده از به‌کارگیری روش پیشنهادی به‌منظور بهسازی گفتار مورد ارزیابی و مقایسه قرار می‌گیرد. در بخش آخر نیز به نتیجه‌گیری در مورد پژوهش انجام‌شده خواهیم پرداخت.

## ۲- الگوریتم یادگیری واژه‌نامه تطبیقی حریصانه

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

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

که  $X(k)$ ،  $S(k)$  و  $N(k)$  به‌ترتیب سیگنال نوفه‌ای، گفتار و نوفه با شماره قاب  $k$  هستند. سیگنال نوفه‌ای  $X \in \mathbb{R}^N$  می‌تواند به‌صورت خطی با نمایش تُنک اتم‌ها به‌صورت  $X = D.C$  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 نُرم واحد  $\|d_{(i,p)}\|_2 = 1, \dots, P$  و بردار کدگذاری تُنک  $c \in \mathbb{R}^P$ ،  $P \gg K_0$  که شامل ضرایب تُنک  $X$  است نمایش داده می‌شود [5].

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

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 ابتدا به‌منظور بهسازی تصاویر پرتونگاری رایانه‌ای سه‌بعدی مطرح

بیشینه‌سازی رابطه بالا برطبق روال زیر خواهد بود [۲۳]:

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

$$\tilde{x}_k = \frac{x_k}{\|x_k\|_1} \rightarrow R^0 = \tilde{X} \quad (۴)$$

۲- نرم  $l_2$  هر فریم به‌صورت زیر محاسبه می‌شود:

$$E_k = \|\tilde{r}_k\|_2 = \sum_{n=1}^l |r_k^j(n)|^2 \quad (۵)$$

در این رابطه  $R^j = [r_1^j, \dots, r_k^j]$  و بردار ستونی مانده  $K$  بعدی متناظر با  $k$  امین ستون  $R^j$  است.

۳- شاخص  $\hat{k}$  متناظر با فریمی از سیگنال با بزرگترین نرم  $l_2$  به‌دست می‌آید:

$$\hat{k} = \arg \max_{k \in K} (E_k) \quad (۶)$$

۴- ضریب توسعه با حاصل ضرب داخلی بردار مانده و اتم به‌دست آمده محاسبه می‌شود:

$$\alpha_k^j = \langle r_k^j, d_j \rangle \quad (۷)$$

۵- مانده جدید با حذف مؤلفه‌های متناظر با اتم انتخاب‌شده برای هر  $k$  در  $r_k^j$  محاسبه می‌شود:

$$r_k^{j+1} = r_k^j - \frac{\alpha_k^j}{\langle d_j, d_j \rangle} d_j \quad (۸)$$

عبارت  $\langle d_j, d_j \rangle$  در مخرج رابطه بالا تضمین می‌کند که ضریب توسعه نرمالیزه‌شده 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  $X=D.C$ ، بردار ضرایب  $\hat{C}$  از رابطه  $C=D^T X$  قابل محاسبه است.

۶- بازگشت به مرحله دوم تا هنگامی که تعداد اتم‌های مورد نیاز به‌دست آید [23].

به‌منظور تعیین میزان  $\xi$  حاصل از یک الگوریتم یادگیری واژه‌نامه، شاخص  $\xi$  مبتنی بر نرم  $\xi$  برای ضرایب تبدیل حاصل به‌صورت زیر تعریف می‌شود:

$$\xi = \frac{\|C\|_1}{\|C\|_2} \quad (۹)$$

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

مؤلفه‌های اساسی<sup>۱</sup> و روش یادگیری واژه‌نامه مبتنی بر K-SVD<sup>۲</sup> و همچنین نرخ خطای بازسازی این الگوریتم‌ها در شکل (۱) نشان داده شده است [19-22]. نتایج شبیه‌سازی‌ها گزارش‌شده در شکل‌ها و جداول این بخش بر روی یک جمله مربوط به یک گوینده زن از مجموعه دادگان GRID آزمایش شده است [26]. تعداد اتم‌ها و نرخ  $\xi$  برای الگوریتم K-SVD به‌ترتیب 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،  $\xi$  حاصل از الگوریتم GAD با افزایش تعداد اتم‌ها با شیب ملایمی افزایش می‌یابد. نرخ  $\xi$  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  $\xi$  بر طبق رابطه (۱)، سیگنال بازنمایی‌شده توسط الگوریتم GAD نمایش  $\xi$  تنک‌تری نسبت به حوزه زمان و در مقایسه با سایر الگوریتم‌های بیان‌شده 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 می‌آیند، استفاده می‌شود. این الگوریتم،  $\xi$  تنک‌ترین قاب‌های داده ورودی را تشخیص داده و از روش متعامدسازی به‌منظور کاهش پیچیدگی‌های مسأله بهره می‌گیرد [25].

(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

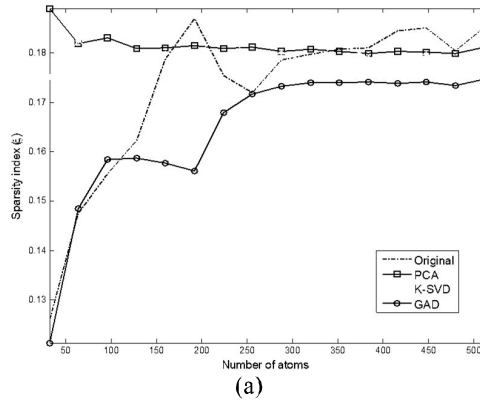
در گام نهایی الگوریتم، دو قانون به‌منظور رسیدن به تعداد اتم‌های موردنظر یا کاهش خطای تقریب از مقدار

<sup>1</sup> Principal component analysis (PCA)

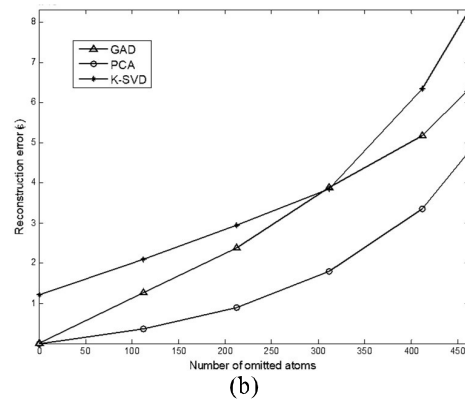
<sup>2</sup> K-Singular value decomposition

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

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

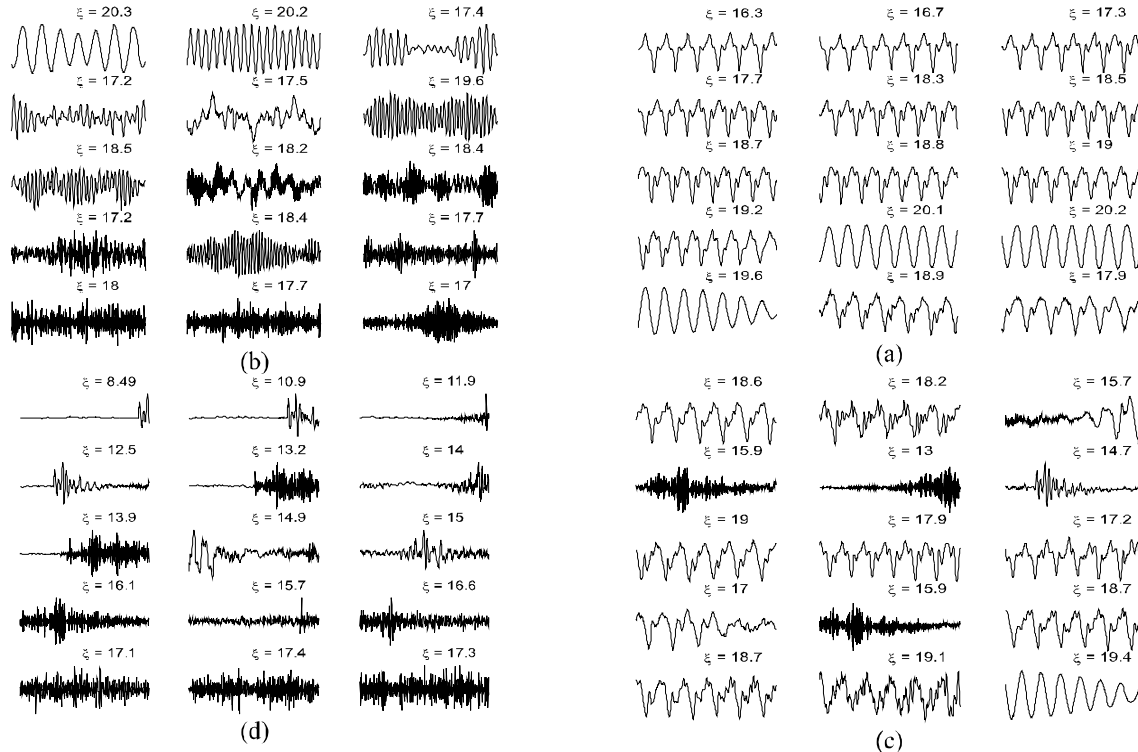


پیش فرض برای خاتمه الگوریتم در نظر گرفته شده است. در شبیه‌سازی‌های انجام شده، نتایج حاصل از این الگوریتم با روش‌های یادگیری واژه‌نامه دیگر مانند PCA و K-SVD مقایسه شده است. شاخص تُنکی نمونه‌هایی از اتم‌های یادگرفته شده از روش‌های مختلف در شکل (۲) نشان داده شده است. همان‌طور که دیده می‌شود، اتم‌های استخراج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



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

(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.



(شکل-۲): (a) نمونه‌هایی از قاب‌های سیگنال گفتار. نمونه‌هایی از اتم‌های آموزش دیده با: (b) PCA. (c) K-SVD. (d) GAD.

(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<sup>۱</sup> را نشان می‌دهد.

### ۳- روش پیشنهادی

در روش بهسازی گفتار به کمک واژه‌نامه مبتنی بر داده، حذف اتم‌های واژه‌نامه بدون توجه به این مسأله صورت می‌گیرد که ماهیت قاب‌های بی‌واک گفتار می‌تواند بسیار مشابه با قاب‌های نوفه‌ای و با تُنکی بسیار پایین باشد. این قاب‌ها در فرآیند بهسازی به کمک نمایش تُنک سیگنال گفتار نادیده گرفته شده و حذف می‌شوند؛ بنابراین سیگنال گفتار تخمینی با اعوجاج منبع حاصل می‌شود.

در این صورت می‌بایست از قاب‌های ابتدایی که به‌طور 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 مسأله 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 به منظور بالا بردن دقت در انتخاب اتم‌های واژه‌نامه می‌توان در کنار معیار تُنکی  $\xi$  تعریف شده در روش GAD، از معیارهای تُنکی دیگری نیز بهره برد.

گروه پردازش سیگنال تُنک<sup>۱</sup> 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 معرفی چند معیار تنکی پرداختند [27-29]. 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 [30] مطرح شده است. نکته قابل توجه آن است که یک معیار تنکی می‌بایست 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، نرمالیزه بودن آن است. بدین معنا که یک مجموعه از ضرایب را تنها به این دلیل که تعداد ضرایب بیشتر یا کمتری نسبت به مجموعه دیگر دارد، نباید تُنک‌تر از دیگری دانست؛ بنابراین دو وضعیت برای نرمالیزه کردن می‌بایست وجود داشته باشد: نخست این که یک معیار تنکی 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<sup>۲</sup> و هایر<sup>۳</sup> این دو خاصیت را برآورده می‌سازند [30].

همچنین تنها معیاری که تمام ویژگی‌های شش‌گانه دالتون را دارد معیار جینی است. برای یک رشته N تایی از ضرایب  $\vec{C} = [C_1 C_2 C_3 \dots C_n]$  این دو معیار به صورت زیر بیان می‌شوند [30].

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

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

با استفاده از نتایج گزارش شده در [27-28]، 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 هم‌دوسی بالایی 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 نُرم تعریف شده در رابطه (۳)، هرچه میزان این شاخص تُنکی بیشتر باشد

<sup>3</sup> Gini

<sup>4</sup> Hoyer

<sup>1</sup> Signal to noise ratio (SNR)

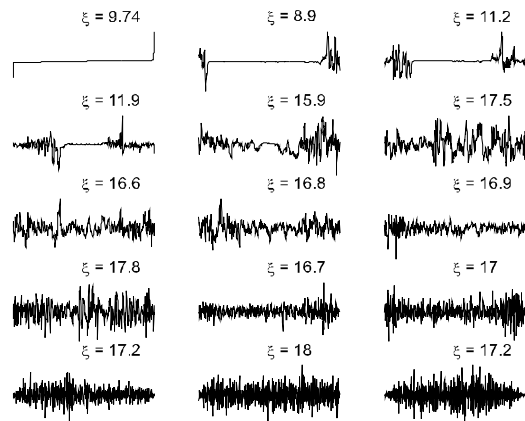
<sup>2</sup> Sparse signal processing team

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

$$\xi_{Gini} = \max \frac{\|x_k\|_1}{Gini(|x_k|)} \quad (12)$$

تُنکی قاب‌هایی اولیه سیگنال مشاهده‌ای که ماهیت نوفه‌ای 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) - (d) 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.

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 نوفه‌ای مقادیر بالاتری را اتخاذ 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 نوفه‌ای 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.



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

GAD\_Gini

(Figure-3): The examples of the learned atoms using GAD\_Gini algorithms.

قابل ذکر است که نتایج حاصل از معیار هابیر به علت مطلوب نبودن 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 یافته‌ای از همه ضرایب موجود محاسبه می‌شود تا تغییر در یک ضریب حتی کوچک در اندازه تنکی 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 کمتری نسبت به مجموعه دیگر دارد یا ضرایب آن خیلی بلند یا خیلی کوتاه است، تنک‌تر از دیگری تعیین نکنند.

(۳) مقدار این شاخص برای بخش‌های نوفه‌ای متمایز از بخش‌های گفتاری است؛ بنابراین در نظر گرفتن شاخص

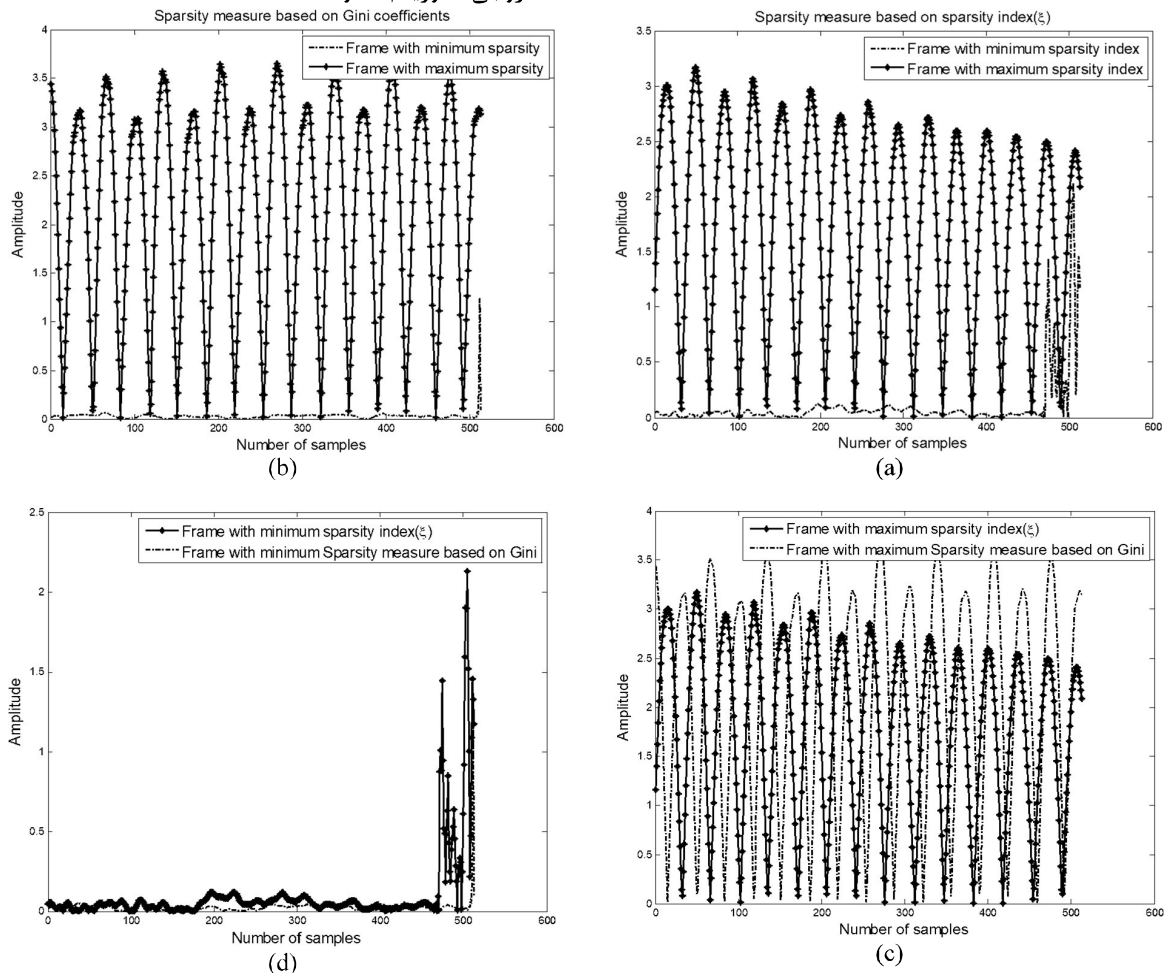
در واقع قاب حاصل از رابطه پیشنهادی که بر مبنای معیار جینی پایه‌ریزی شده تُنک‌تر از قابی خواهد بود که براساس تُنکی مبتنی بر نُرم به دست آمده است. این نتیجه‌گیری به طور تقریبی در بیش‌تر داده‌های گفتاری متعلق به گویندگان زن و مرد مشاهده شده است.

در مورد نحوه انتخاب اتم‌های ابتدایی ماتریس واژه‌نامه نیز همان‌طور که بیان شد از اطلاعات حاصل از قاب‌های ابتدایی سیگنال گفتار نوفه‌ای 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 علاوه بر مقدار نرخ تُنکی محاسبه شده بر روی واژه‌نامه بررسی می‌شود و قاب‌هایی در ابتدای واژه‌نامه قرار می‌گیرند که علاوه بر داشتن نرخ تُنکی بالا دارای انرژی

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

ورودی الگوریتم:  $X$ ,  $D$  (ماتریس تهی اولیه واژه‌نامه) و محدوده تغییرات نرخ تُنکی برای قاب‌های سکوت اولیه سیگنال گفتار ( $\epsilon_N$ )

خروجی الگوریتم:  $C$  و  $D$



(شکل-۴): (a) نمایش قاب‌های با کمینه و بیشینه مقدار تُنکی انتخاب شده براساس شاخص مُبتنی بر نُرم. (b) نمایش قاب‌های با کمینه و بیشینه مقدار تُنکی انتخاب شده براساس رابطه پیشنهادی. (c) نمایش قاب‌های با کمینه مقدار تُنکی انتخاب شده براساس شاخص مُبتنی بر نُرم و رابطه پیشنهادی. (d) نمایش قاب‌های با کمینه مقدار تُنکی انتخاب شده براساس شاخص مُبتنی بر نُرم 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  $\xi_{Gini}$  برای تمامی  $K-t$  قاب سیگنال گفتار نوفه‌ای

۲) انتخاب قاب با بیشترین تُنکی  $\xi_{max}$

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

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

\* در گام اول  $t=1$

۱) محاسبه نرخ تُنکی  $\xi_{Gini}$  از رابطه (۱۲) برای تمامی  $K$

قاب سیگنال گفتار نوفه‌ای  $X_k, 1 \leq k \leq K$

۲) انتخاب قاب با بیشترین تُنکی  $\xi_{max}$  و انتقال آن به

نخستین ستون ماتریس واژه‌نامه  $D_1 \leftarrow (X_m)_{\xi_{max}}$

\*\* شروع حلقه یادگیری واژه‌نامه مبتنی بر داده  $t \neq 1$

به‌منظور بررسی شاخص تُنکی، میانگین مقدار این معیار در الگوریتم‌های مورد بررسی برای اتم‌های واژه‌نامه و ماتریس ضرایب 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 ده 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 به‌دست 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 به‌دست آمده است که این نتایج در مورد ماتریس ضرایب نیز برقرار خواهد بود. گفتنی است که در این بخش از بیان نتایج حاصل از دیگر روش به‌دلیل پایین بودن 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 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 به‌منظور به‌کارگیری از واژه‌نامه یادگیری شده در بهسازی گفتار می‌بایست اتم‌ها در واژه‌نامه مبتنی بر داده از قاب‌های شبه‌گفتار با تُنکی بالا (مقدار شاخص تنکی کم) تا قاب‌های شبه‌نوفه با تُنکی کم (مقدار شاخص تنکی 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 به‌دست 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 به‌ترتیب بعد از تعداد اتم‌های برابر با ۲۰۰ و ۲۵۰ به مقدار تُنکی بیشینه 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 به‌کمک الگوریتم یادگیری واژه‌نامه پیشنهادی و نیز به‌کارگیری بخش‌های آن شامل اعمال محدوده بر شاخص

قاب انتخاب شده به ستون بعدی ماتریس واژه‌نامه اختصاص می‌یابد:

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

(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 به‌منظور بررسی انرژی بازنمایی قاب ورودی بر روی واژه‌نامه و قرار گرفتن آن قاب به‌عنوان 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 به‌صورتی 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 مابقی اتم‌ها که ساختار نوفه را بازنمایی می‌کنند در روال دست‌یابی به سیگنال بهسازی شده نادیده گرفته می‌شوند و ماتریس واژه‌نامه  $\tilde{D}$  حاصل می‌شود. به‌صورت تجربی هفتاد درصد ستون‌ها به‌منظور بازسازی سیگنال گفتار مورد استفاده قرار می‌گیرند و از ضرب این درصد اتم‌ها در ماتریس بازنمایی تُنک مختصر شده  $\tilde{C}$ ، سیگنال گفتار حذف نوفه شده  $\tilde{X}$  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  $X(t)$ ، به قاب‌های با طول L که به اندازه G نمونه هم‌پوشانی دارند، تقسیم‌بندی می‌شوند. K آمین قاب به‌صورت زیر به‌دست می‌آید:

$$X_k = [X((k-1)(L-G)+1), \dots, X(kL-(k-1)G)]^T \quad (13)$$

که در آن  $k \in \{1, \dots, K\}$  شماره قاب مورد بررسی است. سپس از قاب‌های ساخته شده ماتریس  $X \in R^{L \times K}$  تشکیل می‌شود که K آمین ستون با قاب داده  $X_k$  مشخص و  $(l, k)$  آمین ستون به‌صورت زیر خواهد بود:

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

که در آن  $l \in \{1, \dots, L\}$  و  $K > L$  است. در این بخش، مقدار L و G به‌ترتیب برابر با ۵۱۲ و ۱۲ در نظر گرفته شده که بیان‌گر قاب‌های ورودی با طول ۳۲ میلی‌ثانیه و با هم‌پوشانی ۰/۷۵ میلی‌ثانیه است.

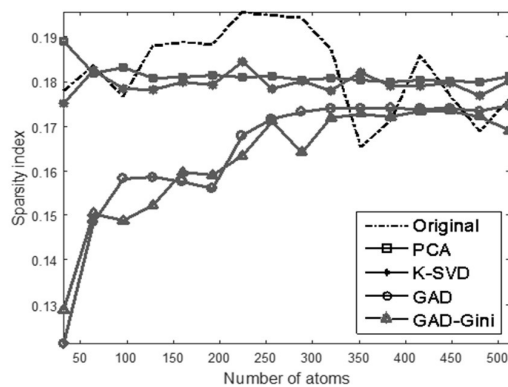
به‌ازای  $SNR=0$  نمایش داده شده است. همان‌طور که دیده می‌شود اتم‌های یادگیری‌شده براساس الگوریتم پیشنهادی  $GAD\_Gini$  به‌خوبی توانسته است، فضای زمان-فرکانس و مشخصات آن را بازنمایی کند.

(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

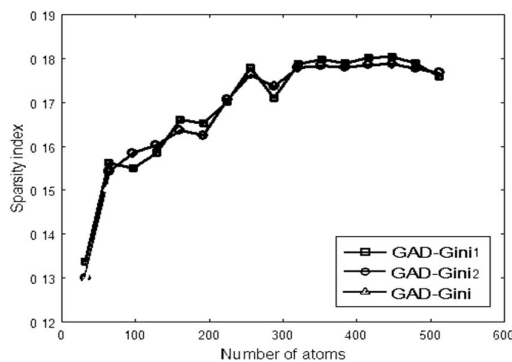
Method	Atom sparsity index		Sparsity of coefficient matrix	
	Female	Male	Female	Male
Speech data	0.1741	0.1673	0.1741	0.1673
PCA	0.1722	0.1731	0.1527	0.1597
K-SVD	0.1643	0.1724	0.1336	0.1387
GAD	0.1531	0.1641	0.1371	0.1357
GAD Gini	<b>0.1476</b>	<b>0.1431</b>	<b>0.1107</b>	<b>0.1202</b>



(شکل-۵): بررسی شاخص تُنکی برای اتم‌ها در واژه‌نامه‌های

یادگیری‌شده با روش‌های مختلف

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



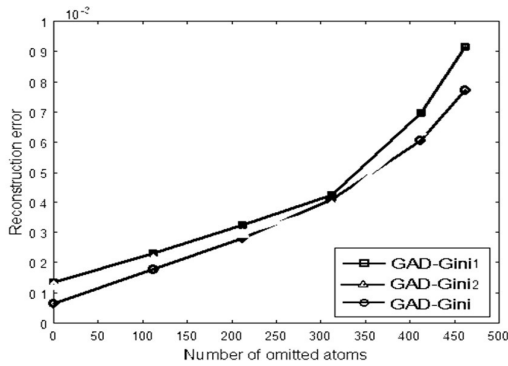
(شکل-۶): بررسی شاخص تُنکی اتم‌ها در واژه‌نامه‌های

یادگیری‌شده با روش پیشنهادی و به‌کارگیری بخش‌های آن

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

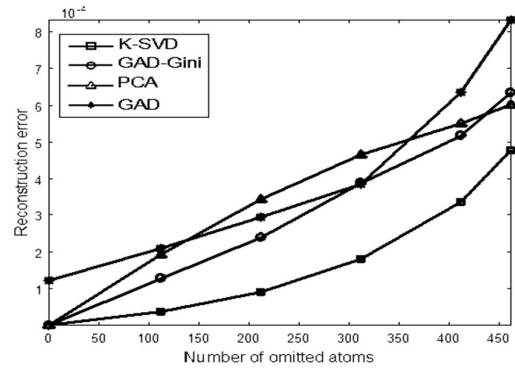
تُنکی به‌کمک قاب‌های ابتدایی و نیز به‌کارگیری شاخص Gini نشان داده شده است. الگوریتم پیشنهادی به‌کمک اعمال محدوده بر شاخص تُنکی با نام  $GAD\_Gini1$  و الگوریتم پیشنهادی به‌کمک به‌کارگیری شاخص تُنکی  $Gini$  با نام  $GAD\_Gini2$  در نظر گرفته شده است. نتایج به‌دست‌آمده نشان می‌دهد که هر یک از این دو گام سهمی در دقت انتخاب تُنک‌ترین قاب‌های ورودی داشته‌اند و در کنار یکدیگر توانسته‌اند، نتیجه مطلوبی در این راستا به‌دست دهند. به‌منظور بررسی بیشتر عملکرد، خطای بازسازی به‌ازای حذف تعداد مختلفی از اتم‌های در هر یک از روش‌ها محاسبه شده و در شکل (۷) نمایش داده شده است. نتایج این شکل در  $SNR=0$  به‌دست آمده است. محدوده این تعداد اتم حذف‌شده از صفر تا ۴۶۰ اتم در نظر گرفته شده که به معنی به‌کارگیری ۵۱۲ تا ۵۲ اتم است. همان‌طور که مشاهده می‌شود نرخ تغییر خطای بازسازی تمامی الگوریتم‌ها با بالا رفتن تعداد اتم‌های حذف‌شده 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 به‌دست آورده است؛ زیرا مؤلفه‌های سیگنال در این تبدیل به‌گونه‌ای مرتب 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 حذف‌شده کمتر مناسب‌تر از الگوریتم بر مبنای  $GAD$  است و از آنجایی که مواردی با تعداد اتم‌های حذف‌شده 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 به‌ازای حذف تعداد مختلفی از اتم‌ها به‌کمک الگوریتم یادگیری واژه‌نامه پیشنهادی و همچنین به‌کارگیری گام‌های مختلف این روش در شکل (۸) نشان داده شده است. الگوریتم پیشنهادی با به‌کارگیری محدوده شاخص تُنکی با نام  $GAD\_Gini1$  و الگوریتم پیشنهادی با به‌کارگیری شاخص تُنکی  $Gini$  با نام  $GAD\_Gini2$  نام‌گذاری شده است. نتایج نشان می‌دهد که اعمال هر یک از 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 به‌کمک روش‌های مختلف یادگیری واژه‌نامه در شکل (۹) و



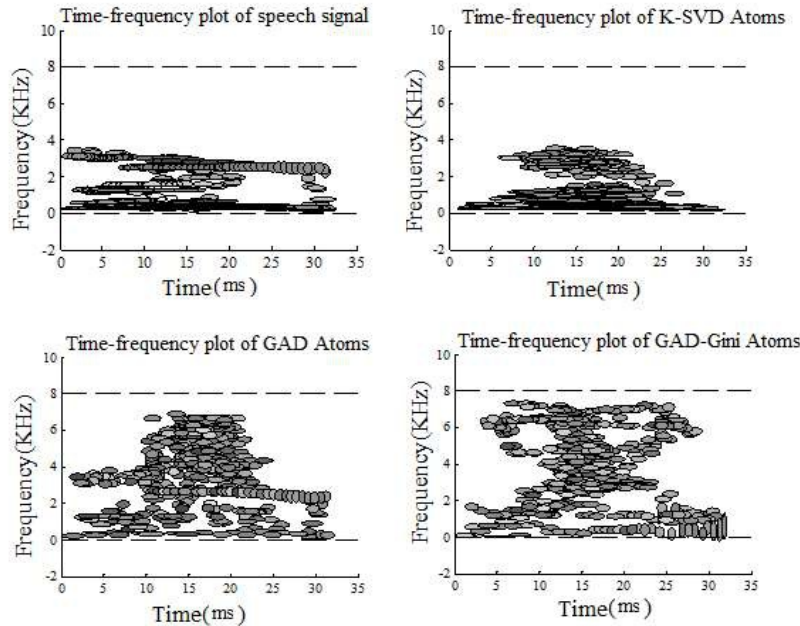
(شکل-۸): مقایسه خطای بازسازی به‌ازای حذف تعداد مختلفی برای واژه‌نامه‌های یادگیری‌شده با روش پیشنهادی و به‌کارگیری بخش‌های آن

(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  $x(t)$ ،  $y(t)$  و  $\hat{x}(t)$  به‌ترتیب سیگنال داده اصلی، سیگنال نوفه‌ای مشاهده شده و سیگنال بهسازی شده هستند. همان‌طور که قابل نتیجه‌گیری است، هرچه سیگنال بهسازی‌شده 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

اتم‌های یادگیری‌شده 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 میانگین‌گیری بر روی صد سیگنال گفتاری منتخب از مجموعه دادگان است.

این نتایج به‌کمک معیار عینی بهبود نسبت سیگنال به نوفه<sup>۱</sup> در سطوح نوفه +۱۰ dB، ۰ dB و -۱۰ dB به‌دست آمده است. این معیار به‌صورت زیر بیان می‌شود:

<sup>۱</sup> Improvement in signal to noise ratio (ISNR)

میزان اتم‌های حذف شده بیشتر باشد، نتایج بهسازی بهتری به دست خواهد آمد.

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

(جدول-۳): نتایج حاصل از معیار ISNR در حضور نوفه 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	200	100	50
-10dB	PCA	0	0.48	1.44	2.37	4.69	5.73
	K-SVD	5.41	<b>6.08</b>	<b>6.52</b>	<b>6.41</b>	<b>5.29</b>	<b>4.11</b>
	GAD	0	1.27	2.94	4.33	4.98	4.07
	GAD_Gini	0	<b>2.29</b>	<b>3.97</b>	<b>4.87</b>	<b>5.01</b>	<b>4.26</b>
	OMLSA	4.71					
0dB	PCA	0	0.46	1.31	2.71	5.41	8.20
	K-SVD	5.69	<b>6.91</b>	<b>7.18</b>	<b>7.57</b>	<b>8.53</b>	<b>10.05</b>
	GAD	0	1.31	2.69	5.22	7.24	9.14
	GAD_Gini	0	<b>2.43</b>	<b>3.56</b>	<b>6.42</b>	<b>8.07</b>	7.18
	OMLSA	7.24					
-10dB	PCA	0	0.51	1.31	2.72	5.48	8.29
	K-SVD	5.82	<b>6.75</b>	<b>7.56</b>	<b>9.91</b>	<b>12.51</b>	<b>13.68</b>
	GAD	0	1.31	2.84	5.31	7.69	9.28
	GAD_Gini	0	<b>1.54</b>	<b>3.12</b>	<b>6.57</b>	<b>8.24</b>	<b>10.17</b>
	OMLSA	7.93					

محدوده قاب‌های انتخابی با توجه به محدوده تُنکی قاب‌های اولیه نوفه‌ای سیگنال مشاهده شده است. همچنین همان‌طور که انتظار می‌رود تعیین تعداد اتم‌های تشکیل‌دهنده واژه‌نامه و حذف اتم‌های بازنمایی‌کننده ویژگی‌های نوفه‌ای بسیار حائز اهمیت است و به همین دلیل با کم شدن یا زیاد شدن بیش از حد مطلوب تعداد اتم‌ها، نتایج بهسازی به شدت مورد تأثیر قرار می‌گیرد. گفتنی است که نتایج بهسازی الگوریتم پیشنهادی در حضور نوفه صورتی اندکی بالاتر از نتایج حاصل از کاهش نوفه در مقابل نوفه قهوه‌ای است.

نتایج گزارش شده نشان می‌دهد که بهسازی حاصل از الگوریتم GAD و GAD\_Gini نزدیک به هم و مناسب بوده و در بیشتر موارد بالاتر از الگوریتم PCA است که در آن فضای مسأله به دو زیرفضای گفتار و نوفه 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 نوفه‌های صورتی و قهوه‌ای اندکی پایین‌تر از نتایج گزارش شده از نوفه سفید گوسی بوده است که می‌تواند به این علت باشد که معیار جینی پیشنهادی بهترین انطباق را به منظور اکتساب قاب‌های تُنک بر اساس نوفه سفید گوسی به دست می‌دهد.

در این جدول، نتایج حاصل از روش OMLSA<sup>۱</sup> اضافه شده که یک تخمین‌گر گفتار است. تابع بهره طیفی بهینه در این روش بر اساس یک میانگین هندسی وزن‌دار مرتبط با عدم حضور گفتار به دست می‌آید [31]. این روش به منظور حذف نوفه‌های نالیستا ارائه شده است. طیف نوفه 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 نوفه‌های غیرگوسی مانند نوفه صورتی و نوفه قهوه‌ای به ترتیب در جداول (۴ و ۵) گزارش شده است. انجام این شبیه‌سازی به منظور بررسی کارایی روش پیشنهادی در حضور نوفه‌های غیرگوسی صورت گرفته است. همان‌طور که نتایج گزارش می‌دهند، می‌توان مشاهده کرد الگوریتم پیشنهادی توانسته است به مقادیر ISNR بالاتری در شرایط نوفه‌ای مختلف نسبت به روش‌های پایه آموزش واژه‌نامه مانند K-SVD و PCA، روش پایه یادگیری واژه‌نامه منطبق بر داده به کمک معیار نرم و نیز روش OMLSA که یک تخمین‌گر گفتاری به منظور حذف نوفه‌های نالیستا 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 بر اساس شاخص Gini و بررسی

<sup>۱</sup> Optimally-modified log-spectral amplitude

(جدول-۴): نتایج حاصل از معیار ISNR در حضور نوفه 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	50
-10dB	PCA	0	0.41	1.32	2.21	4.49	5.63
	K-SVD	5.12	<b>5.82</b>	<b>6.21</b>	<b>6.03</b>	<b>5.10</b>	<b>3.96</b>
	GAD	0	1.14	2.81	4.20	4.82	3.85
	GAD_Gini	0	<b>2.17</b>	<b>3.81</b>	<b>4.72</b>	<b>4.89</b>	<b>4.11</b>
	OMLSA	4.52					
0dB	PCA	0	0.59	1.22	2.65	5.32	8.04
	K-SVD	5.03	<b>6.83</b>	<b>7.02</b>	<b>7.69</b>	<b>8.46</b>	<b>9.88</b>
	GAD	0	1.22	2.59	5.14	7.16	<b>9.01</b>
	GAD_Gini	0	<b>2.36</b>	<b>3.47</b>	<b>6.35</b>	<b>7.95</b>	6.93
	OMLSA	6.84					
-10dB	PCA	0	0.44	1.23	2.65	5.37	8.19
	K-SVD	4.92	<b>6.68</b>	<b>7.47</b>	<b>9.85</b>	<b>12.39</b>	<b>13.41</b>
	GAD	0	1.22	2.71	5.19	7.52	9.18
	GAD_Gini	0	<b>1.43</b>	<b>2.96</b>	<b>6.43</b>	<b>8.16</b>	<b>9.86</b>
	OMLSA	7.18					

(جدول-۵): نتایج حاصل از معیار ISNR در حضور نوفه 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	50
-10dB	PCA	0	0.38	1.29	2.17	4.42	5.55
	K-SVD	5.08	<b>5.73</b>	<b>6.15</b>	<b>5.89</b>	<b>5.03</b>	<b>3.89</b>
	GAD	0	1.08	2.75	4.15	4.81	3.80
	GAD_Gini	0	<b>2.14</b>	<b>3.76</b>	<b>4.65</b>	<b>4.81</b>	<b>4.05</b>
	OMLSA	4.42					
0dB	PCA	0	0.52	1.18	2.60	5.31	7.96
	K-SVD	4.95	<b>6.75</b>	<b>6.92</b>	<b>7.61</b>	<b>8.45</b>	<b>9.82</b>
	GAD	0	1.18	2.51	5.06	7.08	<b>8.92</b>
	GAD_Gini	0	<b>2.29</b>	<b>3.41</b>	<b>6.28</b>	<b>7.91</b>	6.90
	OMLSA	6.77					
-10dB	PCA	0	0.39	1.14	2.72	5.34	8.16
	K-SVD	4.88	<b>6.64</b>	<b>7.42</b>	<b>9.81</b>	<b>12.36</b>	<b>13.38</b>
	GAD	0	1.20	2.67	5.12	7.49	9.15
	GAD_Gini	0	<b>1.38</b>	<b>2.91</b>	<b>6.59</b>	<b>8.11</b>	<b>9.81</b>
	OMLSA	7.14					

پیشنهادی در حضور نوفه‌های غیرگوسی گزارش شده است. نتایج حاصل از این شبیه‌سازی نیز بر توانایی روش پیشنهادی در حذف نوفه‌های غیرگوسی از سیگنال گفتار تأکید کرده و مؤید نتایج قبلی است. روش پیشنهادی توانسته در شرایط نوفه‌های مختلف، مقادیر PESQ بالاتری نسبت به روش‌های پایه آموزش واژه‌نامه، روال پایه یادگیری واژه‌نامه منطبق بر داده و نیز روش OMLSA بدست 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 نوفه صورتی اندکی بالاتر از نتایج حاصل از کاهش نوفه در مقابل نوفه قهوه‌ای حاصل شده است؛ همچنین نتایج الگوریتم پیشنهادی به کمک اعمال محدوده بر شاخص تُنکی با نام GAD\_Gini و الگوریتم پیشنهادی

همچنین نتایج بهسازی به دست آمده با استفاده از معیار<sup>۱</sup> PESQ، 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<sup>۲</sup> که در [32] ارائه شده، استفاده شده است. نتایج حاصل از این معیار به منظور بهسازی سیگنال در حضور نوفه سفیدگوسی 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 نوفه‌های صورتی و قهوه‌ای به ترتیب در جداول (۷ و ۸) گزارش شده است. نتایج حاصل از معیار PESQ به منظور بررسی تکمیلی کارایی روش

<sup>1</sup> Perceptual evaluation of speech quality

<sup>2</sup> MATLAB

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 تنهایی قادر است، نتایج بهسازی را افزایش دهد و به کارگیری این دو گام در کنار یکدیگر توانایی دستیابی به نتایج مناسب در این حوزه پردازشی را فراهم می‌سازد. مقدار PESQ برای سیگنال نوفه‌ای 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 به منظور حذف نوفه سفید گوسی از سیگنال گفتار است. این نتایج، حاصل میانگین‌گیری بر روی پنجاه داده آزمایش است. به منظور بررسی بهتر عملکرد الگوریتم بهسازی گفتار پیشنهادی، نمایش اسپکتروگرام سیگنال‌های بهسازی شده

به کمک روش پیشنهادی در SNRهای 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 K-SVD، PCA، GAD 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 متشکل 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.

(جدول ۶): نتایج حاصل از معیار PESQ در حضور نوفه سفید گوسی و در SNRهای مختلف برای تعداد متفاوت اتم‌ها

(Table-6): The results of PESQ values in the presence of white Gaussian 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	0.76	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	3.81	3.17	2.45
	GAD_Gini	<b>1.65</b>	<b>1.39</b>	<b>0.87</b>	<b>3.17</b>	<b>2.61</b>	<b>1.78</b>	<b>3.84</b>	<b>3.21</b>	<b>2.48</b>
	Noisy signal	0.18			0.88			1.86		

(جدول ۷): نتایج حاصل از معیار PESQ در حضور نوفه صورتی و در SNRهای مختلف برای تعداد متفاوت اتم‌ها

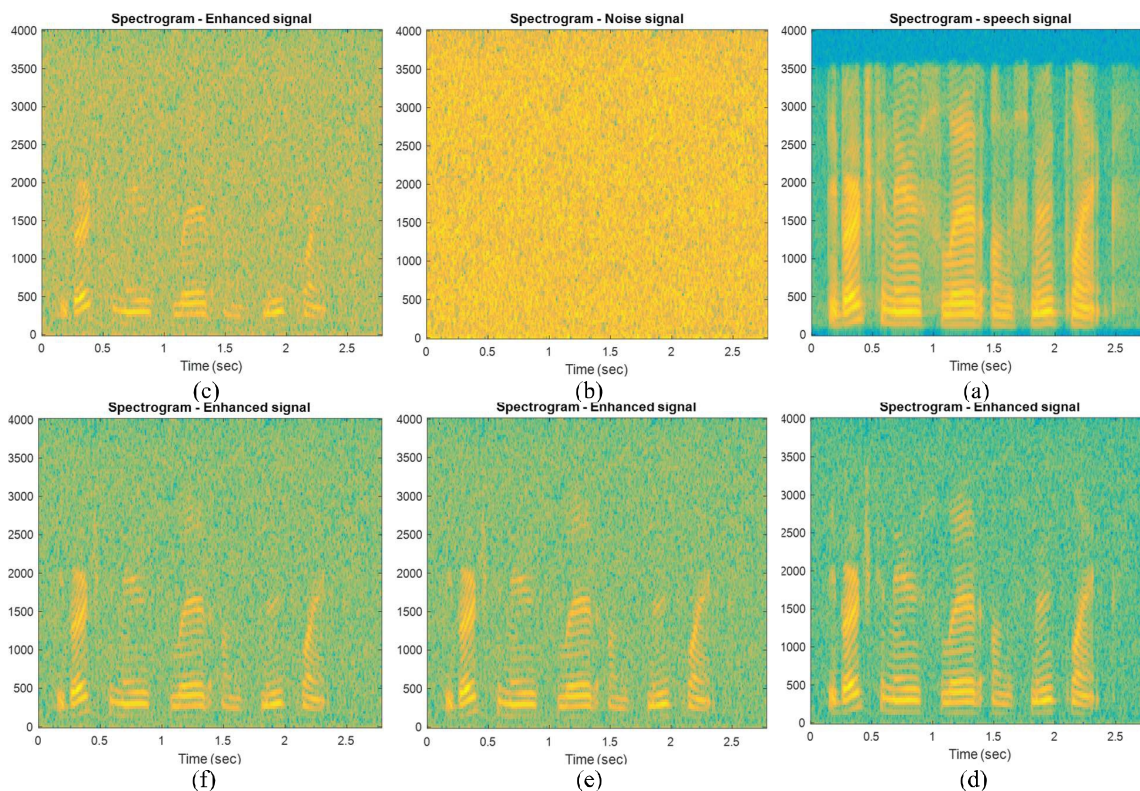
(Table-7): The results of PESQ values in the presence of Pink noise in different SNR values for various numbers of atoms

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	3.78	3.15	2.39
	GAD_Gini	<b>1.66</b>	<b>1.37</b>	<b>0.85</b>	<b>3.15</b>	<b>2.63</b>	<b>1.76</b>	<b>3.82</b>	<b>3.19</b>	<b>2.45</b>
	Noisy signal	0.17			0.83			1.85		

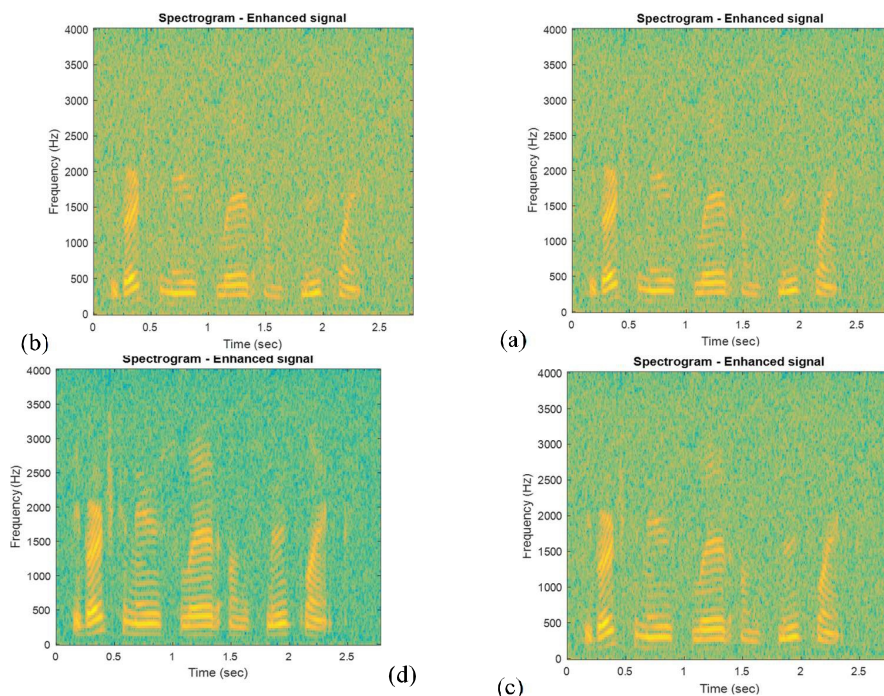
(جدول ۸): نتایج حاصل از معیار PESQ در حضور نوفه قهوه‌ای و در SNRهای مختلف برای تعداد متفاوت اتم‌ها

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

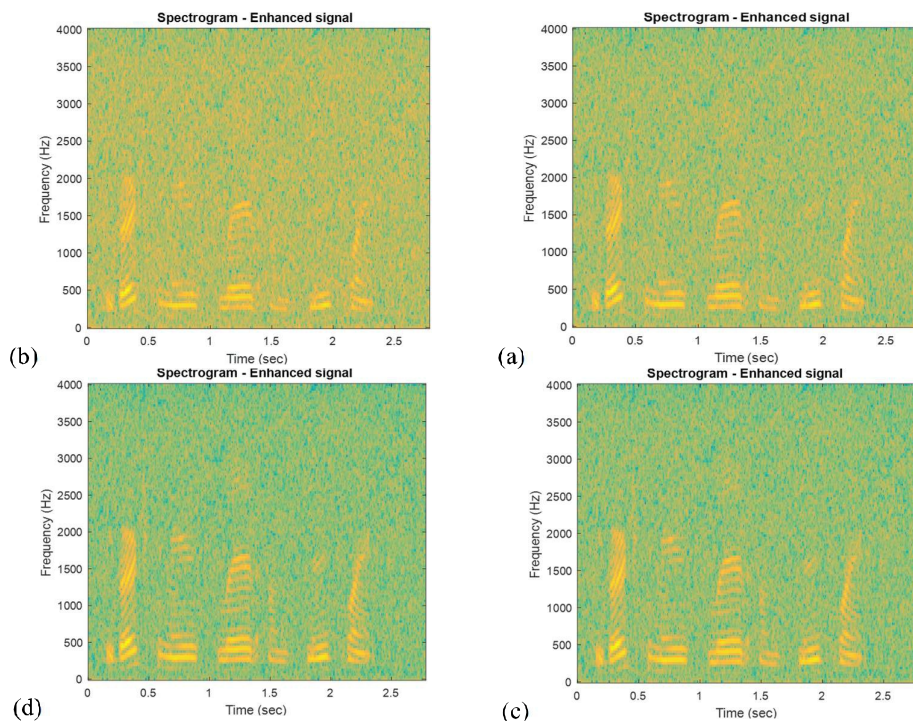
SNR →	Methods ↓	-10dB			0dB			+10dB		
		512	300	100	512	300	100	512	300	100
	PCA	0.68	0.35	0.18	2.50	1.89	0.87	3.46	2.77	1.91
	K-SVD	0.96	0.71	0.65	2.71	2.31	0.98	3.59	2.88	2.18
	OMLSA	0.11			0.67			2.42		
	GAD	1.49	1.21	0.72	2.98	2.47	1.66	3.67	2.94	2.28
	GAD_Gini1	1.51	1.23	0.73	2.96	2.49	1.67	3.69	3.04	2.31
	GAD_Gini2	1.55	1.27	0.76	3.01	2.49	1.68	3.73	3.07	2.31
	GAD_Gini	<b>1.61</b>	<b>1.31</b>	<b>0.79</b>	<b>3.09</b>	<b>2.58</b>	<b>1.70</b>	<b>3.78</b>	<b>3.14</b>	<b>2.39</b>
	Noisy signal	0.19			0.81			1.79		



(شکل-۱۰): نمایش اسپکتروگرام برای: (a) سیگنال گفتار. (b) سیگنال نوفه سفید گوسی. (c) سیگنال گفتار نوفه‌ای با  $SNR=0dB$ . (d) سیگنال بهبودی‌شده در  $SNR=+5dB$ . (e) سیگنال بهبودی‌شده در  $SNR=0dB$ . (f) سیگنال بهبودی‌شده 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 بهبودی‌شده در  $SNR=+5dB$  بر اساس: (a) الگوریتم K-SVD. (b) الگوریتم PCA. (c) الگوریتم GAD. (d) الگوریتم GAD\_Gini.  
 (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، (d) الگوریتم GAD\_Gini.

(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- نتیجه گیری

در این مقاله روش‌های یادگیری واژه‌نامه مبتنی بر داده معرفی شد. از این میان روش یادگیری واژه‌نامه تطبیقی حریصانه GAD به‌منظور بهسازی سیگنال گفتار در حضور نوفه سفید گوسی مورد توجه قرار گرفت. واژه‌نامه در این روش به‌طور مستقیم از روی سیگنال گفتار یادگیری می‌شود تا تطابق بیشتری میان اتم‌ها و ساختار سیگنال ایجاد شده و همدوسی اتم‌ها و رده داده افزایش یابد. انتخاب این اتم‌ها به‌کمک شاخص تنگی 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 به‌دست می‌دهد، بر آن شدیم تا با بررسی این الگوریتم‌ها به نتایج بهتری در زمینه بهسازی گفتار دست پیدا کنیم. بر این اساس ایده تنگی مورد بررسی قرار گرفت و بر مبنای معیارهای مطرح‌شده برای یک شاخص تنگی مطلوب، شاخص تنگی 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 تنگی 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 به‌علت مرتب‌کردن ضرایب و محاسبه معیار جینی اندکی بیشتر از الگوریتم GAD خواهد بود. الگوریتم K-SVD به‌علت به‌کارگیری روال تکرارشونده به‌منظور هم‌گرایی به کمترین خطای تقریب، زمان اجرای طولانی دارد و روش PCA نیز که تبدیل کامل است به‌علت نداشتن روال پیچیده محاسباتی کمترین زمان اجرا را به خود اختصاص می‌دهد.

(جدول-۹): میانگین زمان محاسباتی برای اجرای روش‌های یادگیری واژه‌نامه 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. Scung, “Learning the parts of objects by non-negative matrix factorization”, Nature 401, 788- 791, 1999.

[14] N. Mohammadiha, P. Smaragdīs, 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 بهسازی گفتار مورد استفاده قرار گیرد.

## 6- References

## ۶- مراجع

[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 به ترتیب 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 رایانامه ایشان عبارت است از:

**s.mavaddati@umz.ac.ir**

**سید محمد احدی** مدارک کارشناسی و کارشناسی ارشد خود را به ترتیب در سال‌های ۱۳۶۳ 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: بازشناسی سیگنال گفتار، بهسازی سیگنال تصویر.

نشانی رایانامه ایشان عبارت است از:

**sma@aut.ac.ir**

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

- [21] S. Mavaddaty, S. M. Ahadi, S. Scycdin, "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. Heckstra, "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.