



# ارائه یک روش جدید بهسازی گفتار بر مبنای یادگیری مدل ناهمدوس به کمک ضرایب تبدیل موجک

سمیرا مودتی

استادیار، دانشکده فنی مهندسی، دانشگاه مازندران

## چکیده

بهسازی گفتار یکی از زمینه‌های پرکاربرد در پردازش سیگنال است که در حوزه‌های مختلفی مورد استفاده قرار می‌گیرد. در این مقاله از مفاهیم بازنمایی تُنک و یادگیری واژه‌نامه به‌منظور حذف نوفه از سیگنال گفتار در فضای ویژگی تبدیل موجک استفاده می‌شود. ساختار مورد نیاز جهت بازنمایی هر مؤلفه از سیگنال به کمک مفاهیم بازنمایی تُنک، براساس تعداد کمی از اتم‌های یادگیری شده امکان‌پذیر است. به‌منظور دست‌یابی به نتایج مطلوب در بهسازی گفتار، از روال یادگیری واژه‌نامه ناهمدوس بهره گرفته می‌شود. به کمک ضرایب تبدیل موجک، تجزیه سیگنال در زیرباندهای مختلف که شامل اطلاعات دقیقی از محتوای سیگنال هستند، فراهم می‌شود. در روش پیشنهادی، دو سناریوی نظارت‌شده و نیمه‌نظارت‌شده مورد بررسی قرار گرفته و یک الگوریتم آشکارساز فعالیت گفتاری در هر سناریو با توجه به شرط‌های معرفی‌شده بر اساس واژه‌نامه‌های یادگیری‌شده در گام آموزش، پیشنهاد می‌شود. با استفاده از نتایج خروجی آشکارساز پیشنهادی، سیگنال گفتار تخمینی طی یک روال بهسازی در گام بعد به‌دست خواهد آمد. نتایج گزارش‌شده براساس معیارهای مختلف ارزیابی عملکرد، بر توانایی این روش در زمینه کاهش نوفه سیگنال گفتار تأکید می‌کند. روش‌های پیشنهادی، توانایی بالایی را در خصوص کاهش نوفه‌های نایبستا به‌خصوص در مقادیر سیگنال به نوفه پایین دارد.

واژگان کلیدی: بهسازی گفتار، بازنمایی تُنک، واژه‌نامه ناهمدوس، تبدیل موجک، آشکارساز فعالیت گفتار.

## A New Method for Speech Enhancement Based on Incoherent Model Learning in Wavelet Transform Domain

Samira Mavaddati

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

### Abstract

Quality of speech signal significantly reduces in the presence of environmental noise signals and leads to the imperfect performance of hearing aid devices, automatic speech recognition systems, and mobile phones. In this paper, the single channel speech enhancement of the corrupted signals by the additive noise signals is considered. A dictionary-based algorithm is proposed to train the speech and noise models for each subband of wavelet decomposition level based on the coherence criterion. Using the presented learning method, the self-coherence measure between different atoms of each dictionary and mutual coherence between the atoms of speech and noise dictionaries are minimized and lower sparse reconstruction error is yielded. In order to reduce the computation time, a composite dictionary is utilized including only the speech dictionary and one of the noise dictionaries selected corresponding to the noise condition in the test environment. The speech enhancement algorithm is introduced in two scenarios, supervised and semi-supervised situations. In each

\* Corresponding author

\* نویسندهٔ عهده‌دار مکاتبات

scenario, a voice activity detector (VAD) scheme is employed based on the energy of sparse coefficient matrices when the observed data is coded over the related dictionary.

The presented VAD algorithms are based on the energy of the coefficient matrices in the sparse representation of the observation data over the specified dictionaries. These speech enhancement schemes are different in the mentioned scenarios. In the proposed supervised scenario, domain adaptation technique is employed to transform a learned noise dictionary into an adapted dictionary according to the noise conditions of the test environment. Using this step, the observed data is sparsely coded with low sparse approximation error based on the current situation of the noisy environment. This technique has a prominent role to obtain better enhancement results particularly when the noise signal has non-stationary characteristics. In the proposed semi-supervised scenario, adaptive thresholding of wavelet coefficients is carried out based on the variance of the estimated noise for each frame in different subbands. These implementations are carried out in two different conditions, the training and test steps, as speaker dependent and speaker independent scenarios.

Also, different measures are applied to evaluate the performance of the presented enhancement procedures. Moreover, a statistical test is used to have a more precise performance evaluation for different considered methods in the various noisy conditions. The experimental results using different measures show that the presented supervised enhancement scheme leads to much better results in comparison with the baseline enhancement methods, learning-based approaches, and earlier wavelet-based algorithms. These results have been obtained for an extensive range of noise types including the structured, unstructured, and periodic noise signals in different SNR values.

**Keywords:** Speech enhancement, Dictionary learning, Sparse representation, Domain adaptation, Voice activity detector, Wavelet transform

حوزه‌های مختلف صورت گرفته است. بسیاری از تبدیلاتی که در پردازش سیگنال مورد استفاده قرار می‌گیرند، واژه‌نامه‌های<sup>۲</sup> کاملی را فراهم می‌آورند که هر یک برای نمایش دسته‌ای از سیگنال‌ها مناسب است. از جمله حوزه یا تبدیل‌های رایج می‌توان به تبدیل فوریه، تبدیل فوریه زمان کوتاه<sup>۳</sup> (STFT)، تبدیل کسینوسی گسسته<sup>۴</sup> (DCT) و تبدیل موجک<sup>۵</sup> (WT) اشاره کرد. این نمایش‌های تُنک، کاربرد زیادی در زمینه‌های مختلف همچون جداسازی منابع، نمایش تُنک سیگنال، کدگذاری و غیره دارد. به همین جهت در سال‌های اخیر تلاش‌های زیادی برای ارائه نمایش‌های تُنک این سیگنال‌ها صورت گرفته است. فضای ویژگی در مراجع معرفی شده حوزه بهسازی گفتار به‌طور عمده تاکنون شامل فرم اسپکتروگرام<sup>۶</sup> بوده است. در این مقاله بهسازی گفتار به‌صورت تک‌کانال در حوزه تبدیل موجک با به‌کارگیری روش یادگیری واژه‌نامه و روال بازنمایی تُنک صورت می‌پذیرد و از ضرایب حاصل از آن در سطح تجزیه موردنظر و در هر یک از زیرباندهای حاصل، به‌منظور آموزش و طراحی واژه‌نامه استفاده می‌شود. استفاده از این تبدیل به چند دلیل عمده مورد توجه قرار گرفته است: این تبدیل قابلیت سازگاری را با سیگنال‌های ناپیستا مانند گفتار دارد. باند فرکانسی سیگنال ورودی جهت مدل‌کردن ویژگی

## ۱- مقدمه

بهسازی گفتار یکی از زمینه‌های مورد علاقه پژوهش‌گران جهت معرفی راه‌کارهای جدید در میان انواع حوزه‌های پردازش سیگنال است. انواع نوفه محیطی می‌توانند به‌صورت جدی کیفیت و قابلیت فهم سیگنال گفتار را با مشکل مواجه کنند. دسته روش‌های پایه به‌منظور بهسازی گفتار ارائه شده تاکنون شامل الگوریتم‌های مبتنی بر تفاضل طیفی، طراحی فیلتر، روش‌های آماری، الگوریتم‌های زیرفضا و غیره است [1-4]. در سال‌های اخیر مفهوم تُنکی<sup>۱</sup> نیز در زمینه‌های مختلف پردازش سیگنال مورد توجه قرار گرفته است. این مفهوم بیان‌گر نمایشی از سیگنال است که در آن قسمت عمده‌ای از اطلاعات اساسی سیگنال در تعداد کمی از ضرایب متمرکز شده است. تُنکی در حالت ایده‌آل به‌معنای صفربودن بیش‌تر مؤلفه‌های سیگنال بوده، اما در عمل تُنکی به‌معنای آن است که بیش‌تر ضرایب، کوچک بوده و تنها چند مؤلفه از سیگنال دارای مقادیر قابل توجهی باشند [5,6]. از این بحث در شاخه‌های مختلف علوم مهندسی مانند پردازش سیگنال گفتار و صوت، مهندسی دریا، آنتن، بازشناسی چهره، پردازش تصویر و غیره استفاده می‌شود. همان‌طور که می‌دانیم بسیاری از سیگنال‌ها از جمله سیگنال گفتار در حوزه زمان نمایش تُنک ندارند؛ اما با انتقال آن به فضای ویژگی جدید می‌توان احتمال نمایش تُنک سیگنال را افزایش داد. به همین جهت تلاش‌های زیادی در راستای نمایش تُنک این سیگنال‌ها در

<sup>1</sup> Sparsity

<sup>2</sup> Dictionary

<sup>3</sup> Short Time Fourier Transform

<sup>4</sup> Discrete Cosine Transform

<sup>5</sup> Discrete wavelet Transform

<sup>6</sup> Spectrogram

روش پیشنهادی در دو سناریوی نظارت شده<sup>۷</sup> و نیمه نظارت شده<sup>۸</sup> پیشنهاد و ارائه می شود. در هر یک از این روال ها از آشکارسازی فعالیت گفتاری دقیقی به منظور تعیین قاب های نوفه ای جهت بازآموزی واژه نامه مربوط به داده نوفه بهره گرفته می شود. به روزرسانی اتم های مرتبط با سیگنال نوفه براساس به کارگیری روش تطبیق فضا<sup>۹</sup> قاب های گفتاری و نوفه ای سیگنال مشاهده شده به کمک ضرایب موجک تعیین و برای هر یک از این دسته ضرایب، واژه نامه طراحی می شود [22,23]. از روش تطبیق فضا<sup>۱۰</sup> به منظور دستیابی به واژه نامه متناسب با محیط آزمایش و شرایط نوفه ای آن بهره گرفته می شود [23]. همچنین یک روال بهینه سازی به منظور حصول اتم های با بیشینه ناهمدوسی<sup>۱۱</sup> به کمک انتقال فضای آموزش براساس داده نوفه ای ورودی مورد استفاده قرار می گیرد.

گام بازنمایی تُنک در این روال نیز مبتنی بر یک الگوریتم طراحی شده به منظور دستیابی به اتم های با بیشترین همدوسی نسبت به قاب های داده در نظر گرفته شده خواهد بود. معیار مهمی که می بایست در یادگیری واژه نامه فراکامل<sup>۱۲</sup> مورد توجه قرار گیرد، پارامتر همدوسی<sup>۱۳</sup> میان اتم ها یا ستون های واژه نامه است. کوچکتر بودن این معیار بیان گر این مسأله است که هر اتم مستقل از سایر اتم ها در نمایش قاب داده ورودی تأثیرگذار خواهد بود؛ بنابراین به کمک اتم های آموزش دیده تا حد ممکن مستقل، کمترین خطای تقریب در نمایش تُنک سیگنال تصویر به دست می آید.

در بخش دوم این مقاله مسأله حذف نوفه از سیگنال گفتار مورد بررسی قرار می گیرد؛ سپس در بخش سوم، الگوریتم حذف نوفه پیشنهادی معرفی شده و در بخش چهارم نتایج شبیه سازی های انجام شده ارائه می شود. در ادامه نتیجه گیری در مورد روش معرفی شده بیان خواهد شد.

## ۲- تعریف مسأله

به طور معمول گام ابتدایی در یک الگوریتم بهسازی گفتار استفاده از یک حوزه تبدیل و انتقال سیگنال نوفه ای به یک فضای ویژگی جدید به منظور تحلیل مناسب تر است. این حوزه در روش پیشنهادی، فضای تبدیل موجک در نظر

شنیداری گوش، باید به باندهای فرکانسی مختلف تجزیه شود. در این راستا، تبدیل موجک از این توانایی برخوردار بوده و می تواند به نحو مناسبی آنالیز فرکانسی انجام شده توسط گوش را با تجزیه سیگنال به زیرباندهای مختلف پیاده سازی کند. این تبدیل معکوس پذیر بوده و برگشت براساس آن به فضای داده آسان خواهد بود.

استفاده از تبدیل موجک و تحلیل زیرباندهای حاصل از آن یکی از زمینه های پردازشی مهم به منظور بهسازی سیگنال گفتار است [7-21]. در [7] یک الگوریتم آستانه گذاری برای ضرایب موجک ارائه شده و مقایسه نتایج بهسازی آن با سایر روال های آستانه گذاری مورد بررسی قرار گرفته است. در [8] یک روال بهسازی گفتار در فضای موجک با اعمال ضریب وزن دهی در هر زیرباند ارائه شده است. این ضرایب در طی روال بهسازی براساس نسبت واریانس سیگنال تخمینی به واریانس سیگنال نوفه ای تخمین زده می شوند. در [9]، از آستانه گذاری ضرایب تبدیل موجک در طراحی فیلتر وینر<sup>۱</sup> در حوزه فرکانس به منظور حذف نوفه ماشین از سیگنال گفتار استفاده شده است؛ همچنین یک الگوریتم حذف نوفه<sup>۲</sup> گفتار در [10] براساس نوفه سفید گوسی و توابع آستانه گذاری مختلف پیشنهاد شده است. در [11]، آستانه گذاری به روش شونده<sup>۳</sup> ضرایب موجک به کمک روال تجزیه حالت تجربی<sup>۴</sup> و عمل گر انرژی<sup>۵</sup> Teager در فضای تبدیل موجک به منظور حذف نوفه سفید گوسی به کار گرفته شده است. همچنین از این عمل گر انرژی در [12] نیز به منظور بهسازی سیگنال گفتار در فضای موجک استفاده شده است. همچنین در [13-17]، روش هایی به منظور بهسازی گفتار براساس ضرایب فضای تبدیل موجک ارائه شده است. در [18-21] نیز به بررسی نتایج بهسازی گفتار با استفاده از ضرایب بسته موجک<sup>۴</sup> پرداخته می شود. تجزیه باند تقریب<sup>۵</sup> در این تبدیل نیز به موازات تجزیه باند جزئیات<sup>۶</sup> صورت می پذیرد. با توجه به پژوهش های انجام شده انتظار می رود که پردازش مناسب ضرایب این تبدیل، تخمین آستانه و اعمال آن به ضرایب مربوط به قاب های نوفه ای نتایج مناسبی در حوزه بهسازی گفتار به دست دهد.

در روش ارائه شده در این مقاله از مفاهیم بازنمایی تُنک و یادگیری واژه نامه در جهت حذف انواع دسته نوفه های ایستا، ناپیستا و متناوب از سیگنال گفتار استفاده می شود.

<sup>1</sup> Wiener filter  
<sup>2</sup> Empirical mode decomposition  
<sup>3</sup> Teager energy operator  
<sup>4</sup> Wavelet packet transform  
<sup>5</sup> Approximation subband  
<sup>6</sup> Detail subband

<sup>7</sup> Supervised  
<sup>8</sup> Semi-supervised  
<sup>9</sup> Domain transfer  
<sup>10</sup> Domain adaptation  
<sup>11</sup> Incoherency  
<sup>12</sup> Overcomplete  
<sup>13</sup> Coherence

مسأله کدگذاری تُنک داده آموزش بر روی واژه‌نامه مرکب به صورت زیر بیان می‌شود:

$$C_S^*, C_N^* = \arg \min_{C_S, C_N} \|Y - DC\|_2 \quad (3)$$

$$= \arg \min_{C_S, C_N} \left\| Y - [D_S \ D_N] \begin{bmatrix} C_S \\ C_N \end{bmatrix} \right\|_2$$

ماتریس‌های  $C_S$  و  $C_N$  به ترتیب شامل ضرایب بازنمایی تُنک متناظر با واژه‌نامه‌های گفتار و نوفه  $D_S$  و  $D_N$  است. در روش پیشنهادی از الگوریتم بازنمایی تُنک<sup>4</sup> LARC استفاده شده که تعمیمی از الگوریتم LARS<sup>5</sup> با شرط توقف براساس مقدار هم‌دوسی مانده<sup>6</sup> بوده و سعی به بیشینه‌کردن هم‌دوسی متقابل در گام بازنمایی تُنک را دارد [29,30]; از این‌رو رابطه (۳) می‌تواند به صورت زیر بیان شود:

$$C_S^*, C_N^* = \text{LARC}([D_S \ D_N], K, \text{Coh}) \quad (4)$$

مقدار پارامتر  $\text{Coh}$  در این رابطه بیان‌گر کمینه مقدار هم‌دوسی میان اتم‌داده لازم برای پذیرش اتم به‌عنوان ستونی از واژه‌نامه و  $K$  نرخ تُنکی است. ماتریس ضرایب تُنک به‌دست‌آمده متناظر با وزن‌های واژه‌نامه‌های گفتار و نوفه، تخمینی از طیف گفتار  $\hat{S}$  و طیف نوفه  $\hat{N}$  را به‌دست می‌دهد:

$$\hat{S} = D_S C_S^*, \hat{N} = D_N C_N^* \quad (5)$$

در نظر گرفتن معیار هم‌دوسی مانده برای خاتمه‌دادن به الگوریتم کدگذاری تُنک، قابلیت مصالحه میان انحراف منبع و اعوجاج منبع را به‌دست می‌دهد. انحراف منبع زمانی رخ می‌دهد که بازنمایی تُنک با نرخ تُنکی پایین انجام یا بازنمایی بسیار تُنک باشد؛ بنابراین تعداد اتم‌های مورد نیاز برای نمایش کافی نبوده و داده نمی‌تواند به‌درستی بر اتم‌های واژه‌نامه متناظر کد شود. اگر هر قاب تنها با واژه‌نامه متناظر خود بازنمایی شود، ERC<sup>8</sup> برقرار و کمینه انحراف رخ می‌دهد [29]:

$$\|C_S^*\|_0 + \|C_N^*\|_0 < \frac{1}{2} \left( 1 + \frac{1}{\mu(D_S, D_N)} \right) \quad (6)$$

در این رابطه،  $\mu$  مقدار هم‌دوسی میان دو واژه‌نامه گفتار و نوفه است. همچنین هم‌دوسی متقابل میان اتم‌های واژه‌نامه گفتار  $d^S$  و نوفه  $d^N$  به صورت زیر به‌دست می‌آید [22]:

$$\mu(D_S, D_N) = \max_{1 \leq i \leq L, 1 \leq j \leq L} |d_i^S \cdot d_j^N| \quad (7)$$

<sup>4</sup> Least angle regression with coherence criterion

<sup>5</sup> Least-angle regression

<sup>6</sup> Residual coherence

<sup>7</sup> Coherence

<sup>8</sup> Exact recovery condition

گرفته شده است. سیگنال گفتار نوفه‌ای در این فضا به صورت زیر مدل می‌شود:

$$Y(n) = S(n) + N(n) \quad (1)$$

در این رابطه  $S$  و  $N$  به ترتیب سیگنال گفتار و نوفه تجزیه‌شده در هر زیرباند  $n$  است. ماتریس داده یک سیگنال  $Y \in \mathbb{R}^{P \times F}$  می‌تواند با ترکیب خطی تُنکی از اتم‌ها به صورت  $Y = DC$  که  $Y \in \mathbb{R}^{P \times F}$ ،  $D \in \mathbb{R}^{P \times L}$ ،  $L > P$  یک واژه‌نامه فراکامل است، نمایش داده شود. واژه‌نامه شامل  $L$  اتم در ستون‌ها  $\{d_l\}_{l=1}^L$  با نُرم واحد  $\|d_{(:,l)}\|_2 = 1, \forall l = 1, \dots, L$  و بردار کدگذار  $-K$  تُنک<sup>1</sup>  $C \in \mathbb{R}^{L \times F}$ ،  $L \gg K$  شامل ضرایب بازنمایی سیگنال  $Y$  خواهد بود [24,25]; بنابراین مسأله بازنمایی تُنک که شامل بخش‌های خطای بازسازی و قید تُنکی است، به صورت زیر قابل بیان است [24,25]:

$$C^* = \arg \min_c \|Y - DC\|^2, \|C\|_0 \leq K \quad (2)$$

تعداد ضرایب غیرصفر در ماتریس ضرایب تُنک  $C$  به صورت  $\|C\|_0 = K$  بیان‌گر کارینالیتی سیگنال است. نُرم  $l_0$  در رابطه (۲) به یک مسأله غیرمحدب منتهی می‌شود که NP-دشووار<sup>2</sup> است و ممکن است، موجب گرفتارشدن مسأله بهینه‌سازی در کمینه‌های محلی شود. آزادسازی<sup>3</sup> این مسأله با جایگزینی نُرم  $l_0$  با نُرم  $l_1$  در [26] گزارش شده است.

### ۳- روش پیشنهادی

روندنامای روش پیشنهادشده در این مقاله که شامل دو مرحله اساسی آموزش و آزمون است، در شکل (۱) نشان داده شده است. در گام نخست که آموزش واژه‌نامه‌ها صورت می‌گیرد، داده‌های آموزش با تبدیل موجک به فضای ویژگی انتقال پیدا می‌کنند. واژه‌نامه مرتبط با هر داده گفتار و نوفه براساس روالی که در ادامه معرفی می‌شود، آموزش داده می‌شود. توضیحات درخصوص هر یک از بلوک‌های این روند در ادامه آورده شده است.

#### ۳-۱- بازنمایی تُنک

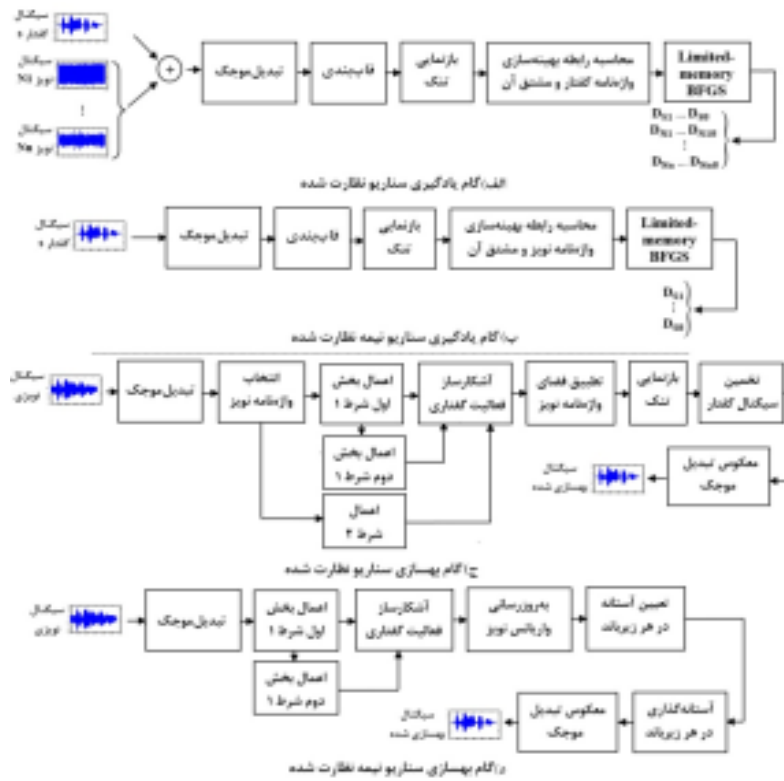
آموزش مدل و بازنمایی تُنک یکی از زمینه‌های پژوهشی جدید در حوزه‌های مختلف پردازش سیگنال است [27,28]. هدف از بازنمایی تُنک، مدل‌کردن قاب‌های سیگنال ورودی به‌عنوان ترکیب خطی از تعداد کمی از بردارهای پایه است.

<sup>1</sup> K-sparse

<sup>2</sup> Non polynomial hard

<sup>3</sup> Relaxation





(شکل-۱): روندنمای روش پیشنهادی کاهش نوفه سیگنال گفتار مبتنی بر یادگیری واژه‌نامه در حوزه تبدیل موجک. شامل: الف) گام یادگیری در سناریوی نظارت‌شده. ب) گام یادگیری در سناریوی نیمه‌نظارت‌شده. ج) گام بهسازی در سناریوی نظارت‌شده. د) گام بهسازی در سناریوی نیمه‌نظارت‌شده.

(Figure-1): The block diagram of the proposed speech enhancement method based on dictionary learning technique in wavelet transform domain included: a) Training step in the supervised scenario. b) Training step in the semi supervised scenario. c) Enhancement procedure in the supervised scenario. d) Enhancement procedure in the semi supervised scenario.

جزئیات<sup>۲</sup> و یک سطح تجزیه در باند تقریب<sup>۳</sup> محاسبه و پنج زیرباند متناظر برای هر قاب داده ورودی ایجاد می‌شود. داده حاصل در هر یک از پنج باند، قاب‌بندی و برای قاب‌های حاصل از هر زیرباند، یادگیری واژه‌نامه به صورت مجزا انجام می‌شود. این روال پیش‌پردازش در شکل (۲) نشان داده شده است. بلوک A و D در این شکل به ترتیب مربوط به زیرباند تقریب و جزئیات است. به این صورت برای هر داده ورودی پنج واژه‌نامه مرتبط با پنج زیرباند به دست می‌آید. گفتار یک سیگنال با ناپیوستایی یا تغییرات بالا است؛ بنابراین تحلیل مولفه‌های فرکانس بالا در آن بسیار حائز اهمیت است. همچنین سیگنال‌های نوفه<sup>۴</sup> مورد بررسی در این مقاله نیز به جز نویز سفید و موسیقی دارای ناپیوستایی بالا هستند که از این جهت حذف آن‌ها از سیگنال گفتار می‌بایست با دقت بیشتر انجام شود. به منظور بررسی دقیق مؤلفه‌های فرکانس بالا که تحت تأثیر شدید نوفه قرار می‌گیرند، تحلیل زیرباندهای جزئیات از اهمیت ویژه‌ای نسبت به باند تقریب

از طرف دیگر، انتخاب پارامتر تُنکی با نرخ بالا یا کدگذاری بسیار متراکم موجب اعوجاج منبع می‌شود؛ یعنی تعداد اتم‌ها در واژه‌نامه گفتار برای بازنمایی درست آن بسیار زیاد خواهد بود و از برخی اتم‌های واژه‌نامه نوفه نیز در بازنمایی سیگنال گفتار استفاده می‌شود؛ بنابراین انتخاب مقدار تُنکی باید به دقت صورت پذیرد. در روش پیشنهادی از این روش بازنمایی تُنک در گام نخست الگوریتم یادگیری واژه‌نامه به منظور دست‌یابی به واژه‌نامه‌های با هم‌دوسی متقابل بالا استفاده شده است. با توجه به شکل (۱)، روش بازنمایی تُنک<sup>۱</sup> LARC بر روی ماتریس داده در هر زیرباند اعمال و ماتریس ضرایب تُنک  $C_S$  و  $C_N$  متناظر با هر یک از زیرباندها حاصل می‌شوند [29].

## ۲-۳- روال یادگیری واژه‌نامه

در گام آموزش، در ابتدا برای هر یک از سیگنال‌های گفتار و داده‌های نوفه، ضرایب موجک با سطح تجزیه سه در باند

<sup>2</sup> Detail

<sup>3</sup> Approximation

<sup>1</sup> Least angle regression with coherence criterion



در این رابطه  $\|C\|_0$  تعیین کننده مقدار کاردینالیته با نرخ بازنمایی تُنک است. مسأله یافتن یک واژه‌نامه با هم‌دوسی متقابل کم می‌تواند به مقیدکردن ماتریس گرام<sup>۲</sup> واژه‌نامه  $G = D^T D$  منتهی شود. مقدار هم‌دوسی واژه‌نامه  $D$  بیشینه مقدار مطلق عناصر غیرقطری ماتریس گرام وقتی اتم‌ها نرمالیزه باشند، معرفی می‌شود [33].

این واژه‌نامه نرمالیزه شده که به منظور بازسازی تُنک مناسب مورد استفاده قرار می‌گیرد، قاب سخت متساوی‌الزاویه<sup>۳</sup> ETF یا قاب گرمس‌مانیان<sup>۴</sup> نامیده می‌شود [34]؛ اما مسأله اساسی این است که این ماتریس برای هر ابعاد دلخواه از واژه‌نامه وجود ندارد؛ بنابراین حل مسأله به صورت تقریبی انجام خواهد گرفت. یکی از نزدیک‌ترین راه‌حل‌های قابل قبول، پس‌پردازش واژه‌نامه مورد نظر خواهد بود. هدف در روش‌های تقریبی مانند روش به کارگرفته شده در این مقاله، کاهش مقدار خطای تقریب  $\|G - I\|_F$  برای دست‌یابی به ماتریس با کمینه هم‌دوسی میان اتم‌ها خواهد بود.

از آنجایی که روش بازنمایی تُنک OMP<sup>۵</sup> انعطاف‌پذیر است از نتایج حاصل از آن همراه با هر روش یادگیری واژه‌نامه می‌توان استفاده کرد. روال یادگیری واژه‌نامه استفاده شده در الگوریتم پیشنهادی مبتنی بر روش Limited-memory BFGS<sup>۶</sup> است که یک الگوریتم تکرار شونده برای حل مسائل بهینه‌سازی غیرخطی بدون قید خواهد بود و روش نیوتن را در حل مسأله تقریب می‌زند [35]. نحوه تخصیص داده آموزش در هر زیرباند به الگوریتم یادگیری واژه‌نامه، در شکل (۳) نمایش داده شده است. مسأله بهینه‌سازی در یادگیری واژه‌نامه با اتم‌های ناهم‌دوس برای داده گفتاری به صورت زیر خواهد بود:

$$F_{D_S} = \arg \min_{D_S} \|S - D_S C_S\|_F^2 + \gamma \|D_S' D_S - I\|_F^2 \quad (10)$$

که بخش نخست مربوط به کاهش خطای تقریب داده گفتار و بخش دوم مربوط به یافتن واژه‌نامه با ماتریس گرام نزدیک به ETF که از پیش در مورد آن صحبت شده خواهد بود و  $\gamma$  پارامتر کنترل‌کننده مصالحه میان این دو بخش است. با در نظر گرفتن رابطه کلی  $\|A\|_F^2 = \text{tr}\{A'A\}$ ، از رابطه (۱۰) خواهیم داشت:

<sup>2</sup> Gram matrix

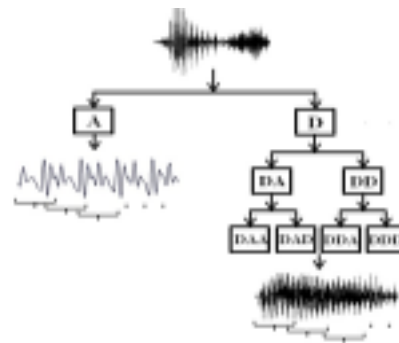
<sup>3</sup> Equiangular tight frame

<sup>4</sup> Grassmannian frame

<sup>5</sup> Orthogonal matching pursuit

<sup>6</sup> Broyden-Fletcher-Goldfarb-Shanno

برخوردار است؛ همچنین تجزیه بیشتر زیرباند تقریب اطلاعات بیشتری از سیگنال به دست نمی‌دهد.



(شکل-۲): قاب بندی داده هر زیرباند تبدیل موجک برای تجزیه سطح سه در باند جزئیات و سطح یک در باند تقریب.  
(Figure-2): Input data framing in the last subband of approximation and detail decomposition level.

K-SVD یک روش یادگیری مؤثر برای دست‌یابی به واژه‌نامه فراکامل از مجموعه‌ای از سیگنال‌های نمونه است [31]. هر قاب ورودی در این الگوریتم با ترکیب خطی از ضرایب  $K$  اتم در یک روش مبتنی بر تجزیه مقادیر منفرد، بازنمایی می‌شود. دو ویژگی اساسی در یادگیری واژه‌نامه، هم‌دوسی متقابل و هم‌دوسی بین اتم‌ها است که هم‌دوسی در اتم‌های یک واژه‌نامه و هم‌دوسی اتم‌ها به داده آموزش تعیین می‌کند. این پارامترها وقتی سیگنال‌های داده آموزش و نوفه‌های مورد نظر ساختار یکسان دارند باید بسیار مورد توجه قرار گیرد. هم‌دوسی بالاتر با رده داده و هم‌دوسی متقابل پایین‌تر میان اتم‌ها موجب می‌شود که خطای تقریب تُنک کوچک‌تری حاصل شود (رابطه ۷). واژه‌نامه‌ها باید بیشترین میزان هم‌دوسی متقابل را داشته باشند تا مانع از اعوجاج در سیگنال بازسازی شده شوند. این بدان معنی است که واژه‌نامه گفتار باید نسبت به هر یک از واژه‌نامه‌های نویز ناهم‌دوس باشد تا قاب‌های مشاهده شده از این سیگنال به درستی بر روی واژه‌نامه متناظر، کدگذاری شود. این معیار به صورت زیر محدود می‌شود:

$$\mu(D) \geq \sqrt{(L - P)/P(L - 1)} \quad (8)$$

در این رابطه  $L$  تعداد اتم‌ها و  $P$  بعد هر اتم است. همچنین، شرط بازسازی درست<sup>۱</sup> ERC، بازسازی مطلوب داده را تضمین می‌کند و منجر به نرخ سریع‌تر کاهش خطای مانده در هم‌گرایی الگوریتم یادگیری می‌شود [32]:

$$\|C\|_0 < \frac{1}{2} \left( 1 + \frac{1}{\mu(D)} \right) \quad (9)$$

<sup>1</sup> Exact recovery condition

حل با استفاده از روش Limited-memory BFGS به صورت زیر است:

$$\frac{\partial F_{D_N}}{\partial D_N} = 2(D_N C_N C_N' - N C_N') + 4\gamma_1(D_N D_N' D_N - D_N) + 2\gamma_2(D_N' D_S D_S')$$
(۱۵)

تابع  $F_{D_N}$  و  $F_{D_S}$  و مشتقات به دست آمده آن‌ها به عنوان ورودی به الگوریتم بهینه‌سازی یادشده اعمال و واژه‌نامه گفتار به همراه واژه‌نامه برای هر نوع نوفه با نرخ ناهمدوسی بالا به دست می‌آید. گفتنی است که در سناریوی بهسازی گفتار نظارت شده، هر دو مرحله یادگیری واژه‌نامه برای گفتار و نوفه و در سناریوی نیمه نظارت شده تنها گام یادگیری واژه‌نامه گفتار انجام می‌شود.

### ۳-۳- انتخاب واژه‌نامه نوفه

سیگنال نوفه‌ای در گام بهسازی پیشنهادی در ابتدا به فضای تبدیل موجک انتقال می‌یابد و سپس نوع نوفه‌ای که گفتار با آن مواجه شده تشخیص داده می‌شود؛ از این‌رو واژه‌نامه مرکب تنها شامل دو واژه‌نامه گفتار و نوفه به جای واژه‌نامه مرکب متشکل از واژه‌نامه گفتار و تمامی واژه‌نامه‌های نوفه خواهد بود. این مسأله موجب می‌شود که نتایج در بازنمایی تُنک گام بهسازی با سرعت بیشتری رخ دهد و با مشکلات پیچیدگی محاسبات مواجه نشویم. در [29]، الگوریتم بهسازی گفتار دیگری با استفاده از یادگیری واژه‌نامه مطرح شده است که در آن اتم‌های واژه‌نامه گفتار و نوفه‌های مورد بررسی به صورت برون خط آموزش می‌بینند. داده نوفه مورد استفاده از نوفه خالص بوده که به طور معمول نوفه شامل در سیگنال نوفه‌ای بر روی آن بازنمایی تُنک خیلی دقیق نخواهد داشت؛ همچنین واژه‌نامه مرکب در گام بهسازی متشکل از واژه‌نامه گفتار و تمامی واژه‌نامه‌های نوفه آموزش دیده است. این مسأله ممکن است، عملکرد الگوریتم در زمینه به دست آوردن نتایج بهسازی مناسب را با مشکل مواجه کند و منجر به کدگذاری خیلی متراکم سیگنال گفتار یا اغتشاش گفتار شود؛ زیرا ممکن است، برخی قاب‌های گفتاری با استفاده از اتم‌های نوفه مختلف به صورت تُنک بازنمایی شوند؛ همچنین زمان محاسباتی الگوریتم در گام بهسازی به علت ابعاد بالای واژه‌نامه مرکب بسیار بالا خواهد بود. مسأله دیگر این است که ساختار مؤلفه‌های برخی قاب‌ها در سیگنال‌های نوفه مختلف به صورت ذاتی مشابه هستند. از این‌رو در واژه‌نامه مرکب حاصل، اتم‌های مشابه وجود

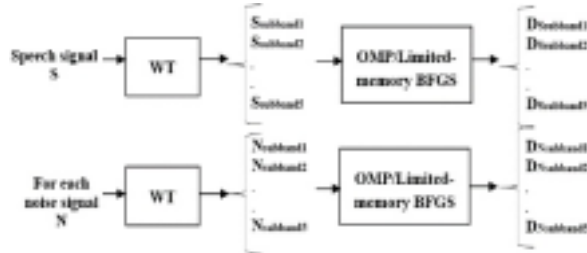
$$F_{D_S} = \text{tr}\{C_S' D_S' D_S C_S\} - 2\text{tr}\{S' D_S C_S\} + \text{tr}\{S' S\} + \gamma(\text{tr}\{D_S' D_S D_S' D_S\} - 2\text{tr}\{D_S' D_S\} + \text{tr}\{I\})$$
(۱۱)

که سه بخش نخست مربوط به خطای تقریب و مابقی بخش‌ها مربوط به ماتریس گرام ETF است. به منظور حل این مسأله به کمک روش Limited-memory BFGS، می‌بایست مشتق رابطه (۱۰) محاسبه شود:

$$\frac{\partial F_{D_S}}{\partial D_S} = 2(D_S C_S C_S' - S C_S') + 4\gamma(D_S D_S' D_S - D_S)$$
(۱۲)

مسأله بهینه‌سازی یادگیری واژه‌نامه نوفه با اتم‌های ناهمدوس نسبت به اتم‌های یادگیری شده برای داده گفتاری در مرحله قبل، به صورت زیر بیان می‌شود:

$$F_{D_N} = \arg \min_{D_N} \|N - D_N C_N\|_F^2 + \gamma_1 \|D_N' D_N - I\|_F^2 + \gamma_2 \|D_N' D_S\|_F^2$$
(۱۳)



(شکل-۳): نحوه تخصیص داده آموزش به منظور بهسازی گفتار در فضای ویژگی تبدیل موجک.

(Figure-3): Training process of speech and noise signals in different subbands of wavelet transform.

بخش نخست در این رابطه مربوط به کاهش خطای تقریب داده نوفه، بخش دوم مربوط به یافتن واژه‌نامه نوفه با ماتریس گرام نزدیک به ETF و بخش سوم مربوط به یافتن اتم‌های واژه‌نامه نوفه مستقل از اتم‌های واژه‌نامه گفتار است. پارامتر  $\gamma_1$  و  $\gamma_2$  مصالحه میان این بخش‌ها را کنترل می‌کند. با ساده کردن این رابطه خواهیم داشت:

$$F_{D_N} = \text{tr}\{C_N' D_N' D_N C_N\} - 2\text{tr}\{N' D_N C_N\} + \text{tr}\{N' N\} + \gamma_1(\text{tr}\{D_N' D_N D_N' D_N\} - 2\text{tr}\{D_N' D_N\}) + \text{tr}\{I\} + \gamma_2(\text{tr}\{D_S' D_N D_N' D_S\})$$
(۱۴)

در این رابطه سه بخش نخست مربوط به خطای تقریب، سه بخش دوم مربوط به بخش ماتریس گرام با ناهمدوسی بالا و بخش آخر مربوط به ناهمدوسی متقابل میان واژه‌نامه گفتار و نوفه است. مشتق رابطه (۱۳) به منظور

<sup>۱</sup> Speech confusion

خواهند داشت و مؤلفو نوفه سیگنال مشاهده‌ای به‌درستی نمی‌تواند بر روی واژه‌نامه نوفه متناظر بازنمایی تُنک داشته باشد. به‌منظور انتخاب نوع نوفه، قاب‌های اولیه سکوت گفتار به‌صورت تُنک بر روی واژه‌نامه مرکبی که شامل تمامی واژه‌نامه‌های نوفه است، بازنمایی داده می‌شود؛ سپس انرژی ضرایب تُنک در کدگذاری بر روی هر واژه‌نامه محاسبه می‌شود. نوع نوفه مشاهده‌شده با توجه به واژه‌نامه نوفه با بیشترین انرژی ضرایب بازنمایی حاصل می‌شود؛ زیرا هر رده داده تنها بر روی واژه‌نامه ناهمدوس طراحی‌شده متناظر خود بازنمایی خواهد داشت. از این‌رو، واژه‌نامه مرکب حاصل از واژه‌نامه گفتار و تنها واژه‌نامه نوفه انتخابی، زمان محاسباتی برای اجرای روال بهسازی را به‌طور چشم‌گیری کاهش می‌دهد.

#### ۴-۳- الگوریتم بهسازی سیگنال گفتار

در نخستین مرحله از گام بهسازی، برچسب قاب‌های گفتار/غیرگفتار به‌کمک الگوریتمی که در ادامه معرفی می‌شود، تعیین می‌شود. در این روال که برای هر زیرباند و در هر قاب انجام می‌شود، در ابتدا و در شرط نخست شباهت میان قاب داده و قاب داده باسازی‌شده که به‌کمک ضرایب تُنک حاصل از بازنمایی به روی واژه‌نامه گفتار  $D_S$  به‌دست می‌آید، بررسی می‌شود. اگر شباهت میان این دو قاب زیاد باشد؛ به معنای آن است که قاب ورودی گفتاری بوده و به همین دلیل بر روی واژه‌نامه گفتار بازنمایی مناسب داشته است؛ بنابراین به‌منظور اطمینان از تصمیم گرفته‌شده، انرژی ضرایب تُنک حاصل از بازنمایی تُنک قاب مشاهده‌شده بر روی قاب‌های ابتدایی و انتهای سیگنال گفتار که ساختار نوفه‌ای دارند و در روابط با  $D_{N0}$  مشخص شده‌اند، بررسی می‌شود. اگر انرژی ضرایب تُنک کم باشد، به معنای این است که قاب ورودی به‌علت دارابودن ماهیت گفتاری بر روی اتم‌های با ماهیت نوفه‌ای، بازنمایی مناسب نداشته و تصمیم گرفته‌شده در مورد گفتاری بودن قاب مورد بررسی درست بوده است. از طرف دیگر اگر شباهت میان این دو قاب کم باشد که به معنای خطای بازنمایی زیاد قاب ورودی بر روی واژه‌نامه گفتار است، می‌توان نتیجه گرفت که قاب غیرگفتاری است. بنابراین به‌منظور اطمینان بیشتر به تصمیم گرفته‌شده، قاب ورودی بر روی قاب‌های داده ابتدایی و انتهای سیگنال گفتار  $D_{N0}$  که ماهیت نوفه‌ای دارند، بازنمایی می‌شوند و اگر انرژی ضرایب تُنک در این بازنمایی

بالا باشد، به معنای اطمینان به تصمیم گرفته‌شده مبنی بر ماهیت غیرگفتاری داشتن قاب ورودی است [36]. این شرط مخصوص سناریوی نیمه‌نظارت‌شده است و اگر سناریو نظارت‌شده باشد شرط دیگری نیز برای اطمینان به تصمیم گرفته‌شده وجود دارد که به‌منظور تصمیم‌گیری مناسب در کنار شرط نخست قرار می‌گیرد. شرط دوم با توجه به دراختیاربودن واژه‌نامه نوفه علاوه‌بر واژه‌نامه گفتار در نظر گرفته می‌شود. در شرط دوم انرژی ماتریس ضرایب تُنک که حاصل از بازنمایی قاب داده ورودی بر روی واژه‌نامه مرکب است، بررسی می‌شود. واژه‌نامه مرکب همان‌طور که از پیش بیان شده از کنار هم قرارگرفتن واژه‌نامه گفتار و واژه‌نامه نوفه انتخاب‌شده به‌دست می‌آید. از مقایسه انرژی متناظر با هر واژه‌نامه می‌توان تصمیم مناسبی درخصوص ماهیت قاب ورودی اتخاذ کرد. اگر قاب ورودی گفتاری باشد، انرژی ضرایب تُنک آن بر روی واژه‌نامه گفتار بسیار بالاتر از انرژی ضرایب تُنک حاصل بر روی واژه‌نامه نوفه خواهد بود. اگر این شرط در هر چهار زیرباند تقریب برقرار باشد، برچسب گفتاری از شرط دوم به قاب ورودی اختصاص می‌یابد. از آنجایی که امکان شباهت در زیرباند جزئیات برای داده گفتار و نوفه زیاد است، تأثیر مقایسه انرژی ضرایب این مجموعه کم‌تر خواهد بود [36].

#### بررسی شرط نخست: تعیین میزان شباهت میان

قاب داده و قاب داده باسازی شده:

$$C = OMP(Y, D_S) \rightarrow \hat{Y} = D_S \cdot C \rightarrow \langle Y, \hat{Y} \rangle = \frac{Y \cdot \hat{Y}}{(\|Y\|_2 \cdot \|\hat{Y}\|_2)}$$

۱- اگر  $\|Y - \hat{Y}\|_2 < \varepsilon$  که به معنای گفتاری بودن قاب مشاهده‌شده است، بازنمایی بر روی قاب‌های اولیه نوفه‌ای صورت می‌گیرد:

$$C_0 = LARC(Y, D_{N0}) \rightarrow \frac{1}{L} \sum_{l=1}^L C_0^2$$

اگر انرژی ضرایب تُنک در این بازنمایی کم باشد به معنای دارابودن برچسب گفتاری برای قاب ورودی است.  $L$  در این رابطه طول قاب مورد بررسی خواهد بود.

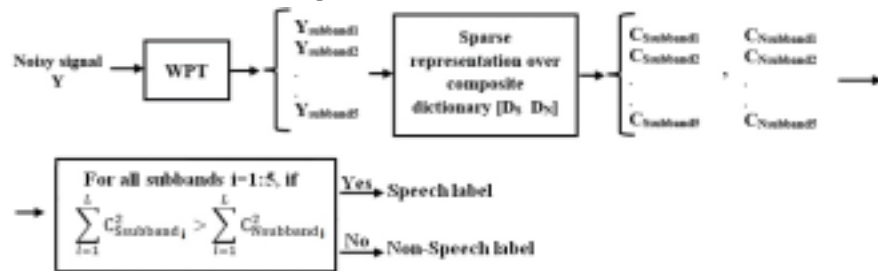
۲- به همین صورت اگر  $\|Y - \hat{Y}\|_2 > \varepsilon$  که به معنای نوفه‌ای بودن قاب مشاهده‌شده است، بازنمایی بر روی قاب‌های اولیه نوفه‌ای صورت می‌گیرد؛ اگر انرژی ضرایب تُنک در این بازنمایی زیاد باشد، به معنای اطلاق برچسب غیرگفتاری به قاب ورودی است.



واژه‌نامه مرکب حاصل از واژه‌نامه گفتار و این واژه‌نامه تطبیق‌یافته انجام می‌گیرد [36]:

$$Y = [D_S \bar{D}_N] \begin{bmatrix} C_S \\ \bar{C}_N \end{bmatrix} \rightarrow \hat{S} = D_S C_S \quad (16)$$

در سناریوی نیمه‌نظارت‌شده که واژه‌نامه نوفه در دسترس نیست، به کمک روابط به‌روز کردن تخمین نوفه در هر قاب و زیرباند، رابطه آستانه‌گذاری ضرایب موجک متناسب با سطح نوفه‌ای که قاب درگیر آن است، تغییر می‌کند. به‌طور معمول آستانه‌گذاری ضرایب در تمامی سطوح و زیرباندها با نرخ ثابتی صورت می‌گیرد و صرف‌نظر از سطح نوفه‌ای است که هر زیرباند را آلوده کرده است. این مسأله موجب عدم دستیابی به نتایج مطلوب در فرآیند بهسازی می‌شود؛ بنابراین در روش پیشنهادی با توجه به دانسته‌های حاصل در مورد قاب‌های نوفه‌ای می‌توان واریانس نوفه را در هر قاب به‌روز کرده و حذف ضرایب با توجه به آن در هر زیرباند صورت گیرد. همان‌طور که بیان شد روندنمای کلی روش پیشنهادی در حوزه تبدیل موجک به تفکیک دو سناریوی نظارت‌شده و نیمه نظارت‌شده مورد بحث در شکل (۱) نمایش داده شده است.



(شکل-۴): نحوه عملکرد آشکارساز فعالیت گفتاری پیشنهادشده در فضای تبدیل موجک در هر زیرباند با در نظر گرفتن شرط دوم.

(Figure-4): The procedure of label detection in the proposed energy-based VAD algorithm in different subbands of wavelet decomposition.

$L_j$  در این رابطه، طول قاب در زیرباند  $j$  ام است. همچنین تابع آستانه‌گذار به‌صورت زیر خواهد بود:

$$\hat{Y} = \begin{cases} Y - \frac{1}{2} \frac{T_j^2}{Y}, & |Y| \geq T_j \\ \frac{1}{2} \text{sign}(Y) \cdot \frac{|Y|^2}{T_j}, & |Y| < T_j \end{cases} \quad (20)$$

در این رابطه شیب حذف ضرایب در تابع آستانه‌گذار متناسب با سطح نویز تغییر می‌کند. هر چقدر مقدار پارامتر آستانه  $T_j$  بیشتر باشد به این معنا است که نوفه بیشتری می‌بایست از سیگنال حذف شود پس شیب منحنی در بخش  $|Y| < T_j$  کاهش پیدا کرده و امکان حذف ضرایب بیشتری در هر زیرباند فراهم می‌شود. در نهایت ضرایب آستانه‌گذاری شده حاصل با تبدیل معکوس موجک به فضای داده برگشت داده می‌شوند.

## بررسی شرط دوم: در ابتدا بازنمایی تُنک بر روی

واژه‌نامه مرکب انجام و سپس انرژی ضرایب متناظر با هر واژه‌نامه بر روی قاب محاسبه و مقایسه بر روی هر چهار زیرباند تقریب و جزئیات صورت می‌گیرد. چگونگی عملکرد این شرط در آشکارساز فعالیت گفتاری پیشنهادی برای هر زیرباند تبدیل موجک در شکل (۴) نشان داده شده است؛ در نهایت می‌توان با توجه به نتایج حاصل از شرط نخست و دوم در مورد اختصاص برچسب گفتاری یا غیرگفتاری به هر قاب ورودی تصمیم‌گیری کرد. در گام بهسازی با توجه به نتایج به‌دست‌آمده در مرحله قبل و تشخیص قاب‌های گفتار/ غیرگفتار، با توجه به سناریوی مورد بررسی دو راه‌کار پیشنهاد شده است. در سناریوی نظارت‌شده، به کمک قاب‌های نوفه‌ای به‌دست‌آمده از نتایج مرحله قبل و تکنیک تطبیق فضا، اتم‌های واژه‌نامه نوفه از پیش یادگیری شده متناسب با شرایط نوفه محیط آزمون و به‌منظور انطباق بیشتر با آن به‌روز می‌شوند؛ که این واژه‌نامه تطبیق‌یافته با  $\bar{D}_N$  معرفی می‌شود؛ سپس بازنمایی تُنک قاب ورودی بر روی

باید توجه داشت که هر یک از این مراحل برای هر زیرباند از پنج زیرباند مورد بررسی انجام می‌گیرد. رابطه به‌روزرسانی سطح نوفه در هر قاب به‌صورت زیر خواهد بود:

$$\hat{N}_{j,n} = \alpha \hat{N}_{j,n-1} + (1 - \alpha) Y_{j,n}^2 \quad (17)$$

$j$  شماره زیرباند و  $n$  شماره قاب است. انحراف معیار نویز در هر زیرباند به‌صورت زیر در نظر گرفته می‌شود [35]:

$$\hat{\sigma}_j = \text{median}(|\hat{N}_j|) / 0.6745 \quad (18)$$

عبارت  $\text{median}$  بیانگر انحراف مطلق میانه<sup>۱</sup> است. محاسبه مقدار آستانه در هر زیرباند به‌صورت زیر است [37]:

$$T_j = \hat{\sigma}_j \sqrt{2 \ln(L_j)} \quad (19)$$

<sup>۱</sup> Median absolute deviation

(جدول-۱): مقادیر همدوسی متقابل میان اتم‌ها و SNR تقریب سیگنال داده برای واژه‌نامه گفتار و نوفه بر حسب dB در روش‌های مختلف.

(Table -1): The atom coherence and SNR values in signal reconstruction for speech and noise dictionaries using different methods.

گفتار	هممه	کارخانه	سفید	ماشین	خیابان	قطار	رستوران	پیانو
همدوسی	۰/۸۷	۰/۸۹	۰/۹۲	۰/۸۹	۰/۹۰	۰/۸۸	۰/۸۹	۰/۹۴
SNR	۱۲/۴	۱۰/۶	۱۱/۲	۱۲/۱	۱۲	۱۰/۷	۱۰/۸	۱۱/۱
همدوسی	۰/۷۶	۰/۸۳	۰/۹۰	۰/۹۵	۰/۹۴	۰/۹۲	۰/۸۹	۰/۹۶
SNR	۱۲/۴	۱۰/۶	۱۰/۸	۱۱/۱	۱۱/۷	۱۱/۸	۱۰/۷	۱۱/۳
همدوسی	۰/۳۷	۰/۴۹	۰/۴۱	۰/۴۶	۰/۴۴	۰/۴۵	۰/۴۶	۰/۵۰
SNR	۱۲/۵	۱۱/۲	۱۱/۳	۱۱/۶	۱۲/۳	۱۲/۲	۱۱	۱۱/۵

(جدول-۲): مقادیر همدوسی متقابل میان اتم‌های واژه‌نامه گفتار و هر یک از واژه‌نامه‌های نوفه در روش‌های مختلف.

(Table -2): The mutual coherence between speech and each of noise dictionary atoms using different methods.

هممه	کارخانه	سفید	ماشین	خیابان	قطار	رستوران	پیانو
۰/۹۳	۰/۸۵	۰/۷۸	۰/۸۶	۰/۸۹	۰/۹۰	۰/۹۱	۰/۷۷
۰/۹۱	۰/۸۷	۰/۷۳	۰/۸۸	۰/۹۱	۰/۹۰	۰/۹۳	۰/۷۲
۰/۶۸	۰/۶۰	۰/۵۷	۰/۶۳	۰/۶۱	۰/۵۹	۰/۶۳	۰/۵۶

#### ۴- جزئیات شبیه‌سازی

شبیه‌سازی‌های انجام‌شده بر روی مجموعه‌دادگان جامع گفتاری GRID با تعداد گوینده و عبارت بیان‌شده به تعداد زیاد انجام گرفته که مناسب برای شبیه‌سازی‌های این مقاله در وضعیت وابسته به گوینده<sup>۱</sup> (SD) و مستقل از گوینده<sup>۲</sup> (SI) خواهد بود [38]. همچنین نرخ نمونه‌برداری برابر با ۸kHz در نظر گرفته شده است. گام آزمون در این مقاله در دو وضعیت مستقل از گوینده (گوینده‌های متفاوت در مرحله آزمون و آموزش) و وابسته به گوینده (گوینده‌های مشابه در مرحله آزمون و آموزش) اجرا می‌شود. مجموعه آموزش و آزمون به ترتیب شامل صدوپنجاه عبارت گفتاری خواهد بود. از این میان هشت گوینده زن و هشت گوینده مرد در گام آموزش و از سه گوینده مرد و سه گوینده زن در آزمون مستقل از گوینده استفاده شده است. طول قاب داده برابر با ۲۰ms و قاب‌ها با هم پوشانی ۵۰٪ تنظیم شده‌اند. تنظیمات در روال یادگیری برای همه منابع داده شامل گفتار و نوفه‌های مختلف یکسان است. نرخ افزونگی واژه‌نامه گفتار برابر چهار و نرخ افزونگی واژه‌نامه نوفه در سناریوی نظارت‌شده نیز برابر همین مقدار در نظر گرفته شده است. شبیه‌سازی بر روی محدوده گسترده‌ای از سیگنال‌های نوفه شامل نوفه هممه (Bab)، کارخانه (Fct)، سفید (Wht)، ماشین (Car)، خیابان (Str)، قطار (Trn)، رستوران (Rst) و نوفه موسیقی پیانو (Pno) از وبسایت جامعه پیانو صورت گرفته است [39-41]؛ همچنین با توجه به بررسی‌های انجام‌شده،

سطح تجزیه باند جزئیات برابر سه در نظر گرفته شده است. مقادیر آستانه  $\epsilon$  در شرط نخست و دوم روال آشکارسازی فعالیت گفتاری پیشنهادی براساس نتایج شبیه‌سازی‌ها و به صورت تجربی تنظیم شده‌اند. این مقدار برای تمامی نوفه‌ها، مقدار ۰/۱۲۵ در نظر گرفته شده است؛ همچنین پارامتر  $\alpha$  براساس نتایج تجربی و برای تمامی نوفه‌ها به مقدار ۰/۸۵ تنظیم شده است. در نخستین شبیه‌سازی انجام‌شده به منظور ارزیابی عملکرد در روال یادگیری واژه‌نامه ارائه‌شده، دو پارامتر همدوسی متقابل اتم‌ها و همدوسی اتم‌های واژه‌نامه‌های مختلف مورد بررسی قرار گرفت. مقدار همدوسی متقابل اتم‌ها در واژه‌نامه گفتار و واژه‌نامه‌های نوفه به دست آمده از رابطه (Y) برای روش‌های یادگیری واژه‌نامه براساس الگوریتم K-SVD با گام بازنمایی تُنک OMP، الگوریتم K-SVD با گام بازنمایی تُنک LARC، الگوریتم K-SVD با گام بازنمایی تُنک LARC دنبال‌شده با روش IPR و روش استفاده‌شده در این مقاله براساس Limited memory-BFGS، مورد بررسی قرار گرفته است. نتایج این شبیه‌سازی در جدول (۱) نشان داده شده است. در این جدول، مقدار SNR گزارش‌شده از رابطه زیر برای هر واژه‌نامه محاسبه شده است:

$$SNR(Y, DC) = 20 \log_{10}(\|Y\|_F / \|Y - DC\|_F) \quad (21)$$

در این رابطه  $Y$ ، سیگنال تقریب زده‌شده در بازنمایی تُنک است. همان‌طور که پیش از این بیان شد در روش پایه یادگیری واژه‌نامه مبتنی بر K-SVD/OMP هیچ‌گونه روش کاهش همدوسی متقابل اتم‌ها به کار گرفته نشده است. به

<sup>1</sup> Speaker dependent

<sup>2</sup> Speaker independent

[41] و روش پایه آستانه‌گذاری موجک BWT<sup>2</sup> تابع آستانه نرم معرفی شده در [3] است؛ همچنین عملکرد الگوریتم با روش ارائه شده در [43]، روش مبتنی بر یادگیری واژه‌نامه ارائه شده در [29] و نیز الگوریتم پیشنهادی در [44] مقایسه شده است.

نتایج حاصل از روش ارائه شده در [29] و روش پیشنهادی در [44] به ترتیب با نام GDL<sup>4</sup> و MDL<sup>5</sup> در شکل‌ها و جداول بیان می‌شوند؛ همچنین روش با نام AWPT<sup>6</sup> مربوط به الگوریتم ارائه شده در [43] است که بهسازی گفتار را براساس آستانه‌گذاری وفقی ضرایب تبدیل موجک انجام می‌دهد. علت ارائه این روش نواقصی بوده است که عملکرد الگوریتم آستانه‌گذاری ضرایب موجک پایه را محدود می‌کند؛ مانند: فرض نوفه گوسی سفید، عملکرد نامناسب در بخش‌های گفتاری، کیفیت شنیداری بد. در الگوریتم پیشنهادی، مقدار آستانه وفقی برای ضرایب موجک و تابع آستانه‌گذار اصلاح شده‌ای ارائه شده که دقت عملکرد بهسازی گفتار مناسبی دارد. همچنین آشکارساز فعالیت گفتاری جدیدی طراحی شده تا آمارگان نوفه را متناسب با سطح نوفه تخمین زده شده به‌روز کند. نتایج روش پیشنهادی در سناریوی نظارت شده با نماد اختصاری ((Proposed\_Sup)) و در سناریوی نیمه نظارت شده با نماد ((Proposed\_Semi)) بیان شده است. به‌منظور سهولت در بررسی نتایج گزارش شده، نام اختصاری روش‌های بیان شده که الگوریتم پیشنهادی با آن‌ها مقایسه می‌شود در جدول (۳) آورده شده است. معیارهای ارزیابی استفاده شده در این مقاله fwSegSNR<sup>7</sup> و PESQ<sup>8</sup> خواهند بود.

(جدول-۳): نام اختصاری روش‌های مقایسه شده با الگوریتم

پیشنهادی به‌کاربرده شده در شکل‌ها و جداول.

(Table -3): The abbreviation name used for each comparison method in Figures and Tables.

مرجع	نام روش	نام اختصاری
[40]	تفاضل طیفی هندسی	GSS
[2]	تفاضل طیفی چندباند	MBSS
[42]	یادگیری واژه‌نامه اصلاح شده	MDL
[27]	یادگیری واژه‌نامه مولد	GDL
[3]	روش پایه آستانه‌گذاری موجک	BWT
[41]	بسته موجک تطبیقی	AWPT <sup>1</sup>
-	روش پیشنهادی نظارت شده	Proposed_Sup
-	روش پیشنهادی نیمه نظارتی	Proposed_Semi

<sup>2</sup> Geometric spectral subtraction

<sup>3</sup> Basic wavelet thresholding algorithm(BWT)

<sup>4</sup> Generative dictionary learning(GDL)

<sup>5</sup> Modified dictionary learning(MDL)

<sup>6</sup> Adaptive wavelet packet thresholding

<sup>7</sup> Frequency-weighted segmental signal to noise ratio

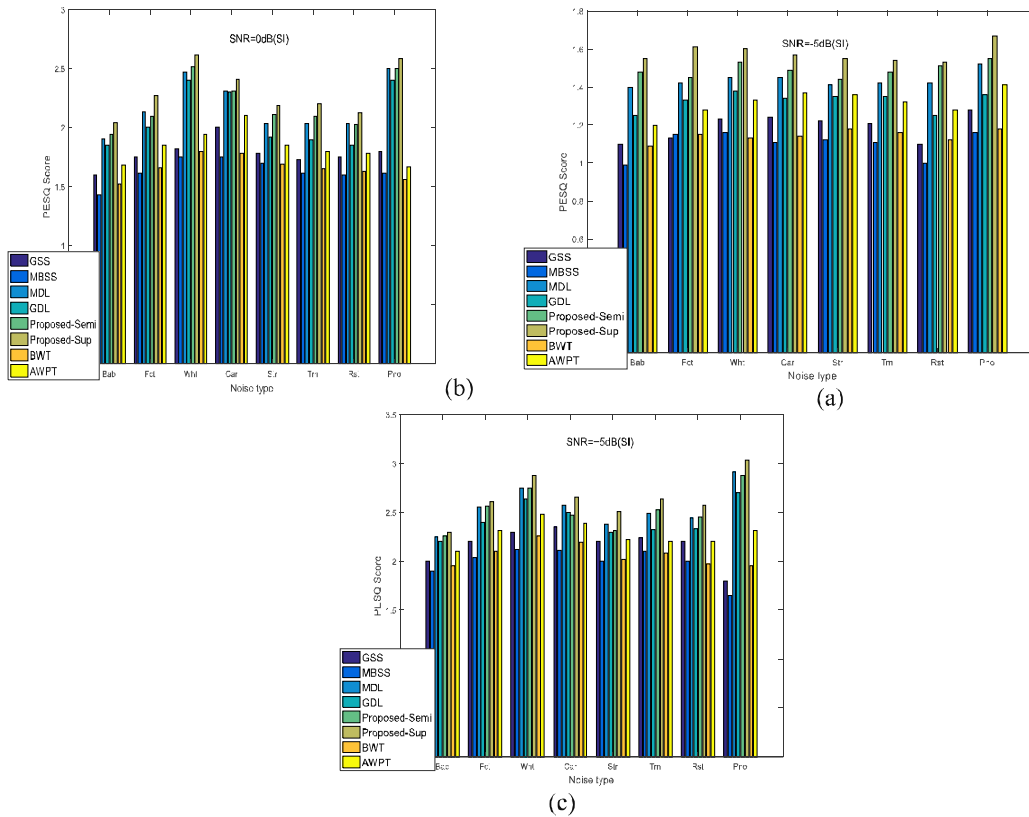
<sup>8</sup> Perceptual evaluation of speech quality

همین دلایل مقادیر همدوسی گزارش شده برای این روش بالا خواهد بود. همچنین در روال یادگیری مبتنی بر K-SVD/LARC تنها پارامتر همدوسی اتم‌ها به داده آموزش بررسی می‌شود و سعی بر این است که بیشینه همدوسی بین اتم‌ها و رده داده مورد نظر برقرار شود [29]. بنابراین مقدار SNR به دست آمده در این روش بالاتر از روال یادگیری K-SVD/OMP خواهد بود؛ همچنین، در روال یادگیری K-SVD/LARC/IPR که همان روش پیشنهادی است، از گام پس‌پردازش IPR به‌منظور کاهش همدوسی متقابل اتم‌های واژه‌نامه طی دو مرحله و به‌منظور نزدیک کردن ماتریس گرام واژه‌نامه به ساختار ETF استفاده شده، [34] به همین دلیل مقادیر همدوسی متقابل اتم‌ها در این روش یادگیری کوچک‌تر از دو مرحله یادشده قبلی است. در روش به‌کارگرفته شده در این مقاله، بخشی از تابع بهینه‌سازی مورد استفاده برای یادگیری واژه‌نامه براساس روابط (۱۰ و ۱۳) به‌منظور یافتن ساختار ETF برای واژه‌نامه در نظر گرفته شده است و بهینه‌سازی اتم‌ها بر این مینا و همراه با کاهش خطای تقریب داده صورت می‌پذیرد. همان‌طور که دیده می‌شود، روال بهینه‌سازی مبتنی بر روش Limited memory-BFGS، نتایج همدوسی کمتری را نسبت به سایر روش‌ها گزارش می‌دهد. در ادامه همدوسی متقابل میان اتم‌های واژه‌نامه گفتار و هر یک از واژه‌نامه‌های نوفه به‌کارگرفته شده در روال بهسازی مورد بررسی قرار گرفت. این مقادیر همدوسی در جدول (۲) گزارش شده است. در سه روش نخست یادگیری واژه‌نامه در نظر گرفته شده در این جدول، مقادیر همدوسی متقابل اتم‌های گفتاری و هر یک از نوفه‌ها مقادیر بالایی دارد؛ زیرا در این روش‌ها روالی برای کاهش این پارامتر در نظر گرفته نشده است؛ اما در روش پیشنهادی در این مقاله با توجه به رابطه (۱۲)، بخش از تابع بهینه‌سازی مبتنی بر کاهش مقدار همدوسی متقابل اتم‌های واژه‌نامه نوفه و گفتار در نظر گرفته شده است. مقدار همدوسی متقابل واژه‌نامه گفتار و نوفه گزارش شده در این جدول، کارایی الگوریتم پیشنهادی را تأیید و منجر به نتایج بهسازی مناسبی می‌شود. در گام بهسازی نظارت شده، در ابتدا واژه‌نامه نوفه منطبق با شرایط محیط آزمون مطابق با آنچه در بخش ۳-۳ بیان شد، انتخاب می‌شود و در روال بهسازی مورد استفاده قرار می‌گیرد. روش‌های پایه بهسازی گفتار که مقایسه با آن‌ها صورت گرفته شامل روش تفاضل طیفی چندباند<sup>1</sup> MBSS [2]، تفاضل طیفی هندسی<sup>2</sup> GSS

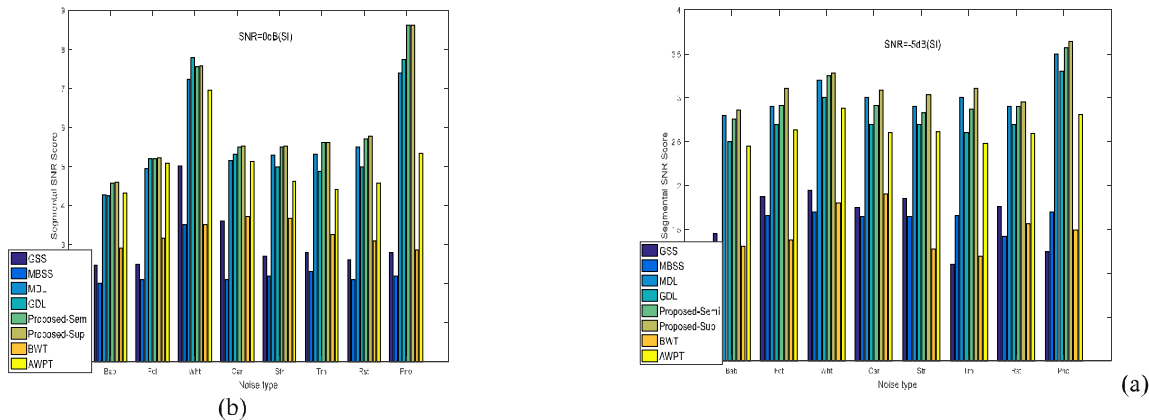
<sup>1</sup> Multi band spectral subtraction

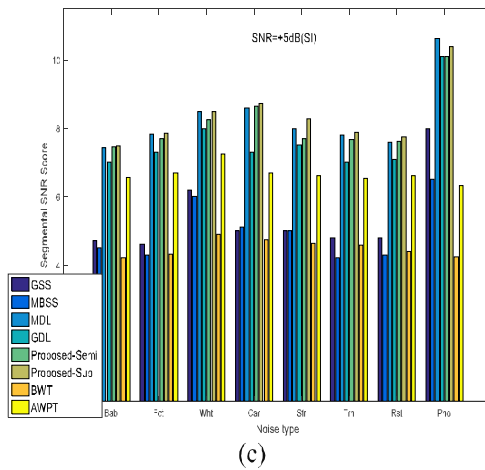
شکل‌های (۵ و ۶)، این نتایج در جداول (۴ و ۵) گزارش شده است. نتایج حاصل از این معیارها در حالت وابسته به گوینده (SD) نیز در جداول (۶ و ۷) بیان شده است. همان‌طور که در شکل‌ها و جداول مشاهده می‌شود، روش پیشنهادی نتایج مطلوبی را در ارزیابی با هر دو معیار PESQ و fwSegSNR نسبت به سایر الگوریتم‌های پایه و روش‌های پیشین مبتنی بر یادگیری واژه‌نامه در شرایط نوفه‌ای متفاوت به دست می‌دهد.

نتایج بهسازی حاصل از روش‌های یادشده برای این دو معیار در حالت مستقل از گوینده (SI) در سه مقدار SNR،  $-5\text{dB}$ ،  $0\text{dB}$  و  $+5\text{dB}$  برای هشت نوع نوفه: همهمه (Bab)، کارخانه (Fct)، سفید (Whi)، ماشین (Car)، خیابان (Str)، قطار (Tm)، رستوران (Rst) و نوفه موسیقی پیانو (Pno) در شکل‌های (۵ و ۶) نشان داده شده است. این نتایج در مقادیر  $-5\text{dB}$ ،  $0\text{dB}$  و  $+5\text{dB}$  SNR همچنین به منظور نمایش بهتر تفاوت میان دسته روش‌های مختلف در



(شکل-۵): نتایج معیار PESQ به منظور مقایسه عملکرد روش‌های مختلف در سناریوهای SI و SNRهای: (a)  $-5\text{dB}$ ، (b)  $0\text{dB}$  و (c)  $+5\text{dB}$ . (Figure -5): Performance evaluation of different methods using PESQ score in speaker independent case (SI). a) SNR =  $-5\text{dB}$ , b) SNR =  $0\text{dB}$  and c) SNR =  $+5\text{dB}$ .





(شکل ۶-): نتایج معیار SNR قطعه‌ای به منظور مقایسه عملکرد روش‌های مختلف در سناریوی SI و SNRهای: (a) -5dB (b) 0dB (c) +5dB.  
(Figure-6): Performance evaluation of different methods using frequency weighted segmental SNR value in speaker independent case (SI). a) SNR = -5dB, b) SNR = 0dB and c) SNR = +5dB.

موسیقی بهترین مقادیر و در حضور نوفه‌های با نالیستایی بیشتر مانند نوفه همهمه و رستوران نتایج پایین‌تری را به دست می‌دهند. بر طبق انتظار و شبیه‌سازی‌های به دست آمده، نتایج حاصل از بهسازی در وضعیت وابسته به گوینده به علت هم‌پوشانی ایجاد شده میان گوینده‌ها در محیط آموزش و آزمون، بهتر از نتایج به دست آمده در وضعیت مستقل از گوینده خواهد بود.

همچنین میانگین نتایج بر روی تمامی شرایط نوفه‌ای (سیگنال‌های نوفه‌ای و SNRهای مختلف) برای این دو معیار و در شرایط وابسته و مستقل از گوینده به ترتیب در جداول (۸ و ۹) گزارش شده است.

نتایج مناسب حاصل از روش پیشنهادی برای تمامی نوفه‌های دارای ساختار و بدون ساختار و حتی در مقادیر SNR کم، قابل مشاهده است. همان‌طور که انتظار می‌رفت، معیارهای بهسازی گزارش شده در حضور نوفه سفید و نوفه

(جدول ۴-): نتایج معیار PESQ به منظور مقایسه عملکرد روش‌های مختلف در سناریوی SI. نتایج روش پیشنهادی در وضعیت بانظارت و نیمه‌نظارت شده به ترتیب با نماد اختصاری ((Proposed\_Sup)) و ((Proposed\_Semi)) بیان شده است.

(Table-4): The results of PESQ measure values for performance evaluation of different methods in the presence of the mentioned noise signals in SI scenario. The proposed method is introduced using ((Proposed\_Sup)) and ((Proposed\_Semi)) in supervised and semi supervised situations, respectively.

SNR=-۵								SNR=+۵								
GSS	MBSS	MDL	GDL	Proposed_Semi	Proposed_Sup	BWT	AWPT	GSS	MBSS	MDL	GDL	Proposed_Semi	Proposed_Sup	BWT	AWPT	
۱/۱۰	۰/۹۹	۱/۳۸	۱/۲۵	۱/۴۸	۱/۵۵	۱/۰۹	۱/۴۰	۲	۱/۹۰	۲/۲۰	۲/۱۱	۲/۲۶	۲/۳۰	۱/۹۵	۲/۱۰	همهمه
۱/۱۳	۱/۱۵	۱/۴۰	۱/۳۳	۱/۴۵	۱/۶۱	۱/۱۰	۱/۲۸	۲/۲۰	۲/۰۴	۲/۴۹	۲/۲۴	۲/۵۶	۲/۶۱	۲/۱۰	۲/۳۱	کارخانه
۱/۲۳	۱/۱۶	۱/۴۳	۱/۳۸	۱/۵۳	۱/۶۰	۱/۱۳	۱/۳۳	۲/۳۰	۲/۱۲	۲/۶۹	۲/۴۴	۲/۷۵	۲/۸۸	۲/۲۶	۲/۴۸	سفید
۱/۲۴	۱/۱۴	۱/۴۳	۱/۳۴	۱/۴۹	۱/۵۷	۱/۱۴	۱/۳۷	۲/۳۵	۲/۱۱	۲/۴۵	۲/۳۸	۲/۴۷	۲/۶۶	۲/۱۹	۲/۳۹	ماشین
۱/۲۲	۱/۱۲	۱/۳۸	۱/۳۱	۱/۴۴	۱/۵۵	۱/۱۸	۱/۳۶	۲/۲۰	۲/۰۵	۲/۳۱	۲/۲۷	۲/۳۱	۲/۵۱	۲/۰۲	۲/۲۲	خیابان
۱/۲۱	۱/۱۱	۱/۴۱	۱/۳۳	۱/۴۸	۱/۵۴	۱/۱۶	۱/۳۲	۲/۲۴	۲/۱۰	۲/۴۳	۲/۲۳	۲/۵۳	۲/۶۴	۲/۰۸	۲/۲۰	قطار
۱/۱۰	۱	۱/۴۰	۱/۲۶	۱/۵۱	۱/۵۳	۱/۱۲	۱/۲۸	۲/۱۴	۲	۲/۲۹	۲/۲۲	۲/۴۵	۲/۵۷	۱/۹۷	۲/۱۸	رستوران
۱/۲۸	۱/۱۶	۱/۵۳	۱/۳۶	۱/۵۵	۱/۶۷	۱/۱۸	۱/۴۱	۱/۸۰	۱/۶۵	۲/۸۲	۲/۶۵	۲/۸۸	۳/۰۳	۱/۹۵	۲/۳۱	پیانو



(جدول-۵): نتایج معیار SNR قطعه‌ای به منظور مقایسه عملکرد روش‌های مختلف در سناریوی SI. نتایج روش پیشنهادی در وضعیت

بانظارت و نیمه‌نظارت شده به ترتیب با نماد اختصاری ((Proposed\_Sup)) و ((Proposed\_Semi)) بیان شده است.

(Table-5): The results of frequency weighted segmental SNR values for performance evaluation of different methods in the presence of the mentioned noise signals in SI scenario. The proposed method is introduced using ((Proposed\_Sup)) and ((Proposed\_Semi)) in supervised and semi supervised situations, respectively.

SNR=-۵								SNR=+۵								
GSS	MBSS	MDL	GDL	Proposed_Semi	Proposed_Sup	BWT	AWPT	GSS	MBSS	MDL	GDL	Proposed_Semi	Proposed_Sup	BWT	AWPT	
۱/۴۵	۱/۳۵	۲/۸۰	۲/۵۰	۲/۷۶	۲/۸۶	۱/۳۱	۲/۴۵	۴/۷۰	۴/۵۰	۷/۴۳	۷	۷/۴۶	۷/۴۸	۴/۲۰	۶/۵۶	همه‌مه
۱/۸۷	۱/۶۶	۲/۹۰	۲/۷۰	۲/۹۱	۳/۱۱	۱/۳۸	۲/۶۳	۴/۶۱	۴/۳۲	۷/۸۲	۷/۳۰	۷/۶۹	۷/۸۶	۴/۳۲	۶/۷۱	کارخانه
۱/۹۵	۱/۷۰	۳/۲۰	۳	۲/۲۵	۳/۲۸	۱/۸۰	۲/۸۸	۶/۲۱	۶/۰۲	۸/۴۳	۸/۰۳	۸/۲۵	۸/۵۰	۴/۸۹	۷/۲۵	سفید
۱/۷۵	۱/۶۵	۳	۲/۷۰	۲/۹۱	۳/۰۹	۱/۹۰	۲/۶۰	۵/۰۳	۵/۱۰	۸/۶۰	۷/۳۰	۸/۶۵	۸/۷۴	۴/۷۵	۶/۶۸	ماشین
۱/۸۵	۱/۶۵	۲/۹۰	۲/۷۰	۲/۸۳	۳/۰۳	۱/۲۸	۲/۶۱	۵/۰۶	۵/۰۸	۸/۰۴	۷/۵۲	۷/۶۹	۷/۲۸	۴/۶۲	۶/۶۱	خیابان
۱/۱۰	۱/۶۶	۳	۲/۶۰	۲/۸۷	۳/۱۱	۱/۱۹	۲/۴۸	۴/۸۱	۴/۲۲	۷/۸۳	۷/۰۲	۷/۶۸	۷/۸۹	۴/۵۹	۶/۵۴	قطار
۱/۷۶	۱/۴۲	۲/۹۰	۲/۷۰	۲/۹۰	۳/۹۵	۱/۵۶	۲/۵۹	۴/۸۱	۴/۳۲	۷/۶۱	۷/۱۲	۷/۶۳	۷/۷۵	۴/۴۳	۶/۶۲	رستوران
۱/۲۵	۱/۷۰	۳/۵۰	۳/۳۰	۲/۵۷	۳/۶۴	۱/۴۹	۲/۸۱	۸/۰۶	۶/۵۲	۱۰/۸	۱۰/۱	۱۰/۱	۱۰/۴	۴/۲۳	۶/۳۳	پیانو

(جدول-۶): نتایج معیار PESQ به منظور مقایسه عملکرد روش‌های مختلف در سناریوی SD. نتایج روش پیشنهادی در وضعیت بانظارت و

نیمه‌نظارت شده به ترتیب با نماد اختصاری ((Proposed\_Sup)) و ((Proposed\_Semi)) بیان شده است.

(Table-6): The results of PESQ measure values for performance evaluation of different methods in the presence of the mentioned noise signals in SD scenario. The proposed method is introduced using ((Proposed\_Sup)) and ((Proposed\_Semi)) in supervised and semi supervised situations, respectively.

SNR=-۵								SNR=+۵								
GSS	MBSS	MDL	GDL	Proposed_Semi	Proposed_Sup	BWT	AWPT	GSS	MBSS	MDL	GDL	Proposed_Semi	Proposed_Sup	BWT	AWPT	
۱/۱۰	۰/۹۹	۱/۴۶	۱/۳۳	۱/۵۱	۱/۵۸	۱/۰۹	۱/۲۰	۲	۱/۹۰	۲/۳۰	۲/۱۹	۲/۳۹	۲/۳۷	۱/۹۵	۲/۱۰	همه‌مه
۱/۱۳	۱/۱۵	۱/۴۹	۱/۴۲	۱/۵۳	۱/۶۳	۱/۱۰	۱/۲۸	۲/۲۰	۲/۰۴	۲/۵۶	۲/۳۱	۲/۵۹	۲/۶۹	۲/۱۰	۲/۳۱	کارخانه
۱/۲۳	۱/۱۶	۱/۵۴	۱/۴۷	۱/۶۰	۱/۶۹	۱/۱۳	۱/۳۳	۲/۳۰	۲/۱۲	۲/۷۵	۲/۴۹	۲/۸۰	۲/۸۸	۲/۲۶	۲/۴۸	سفید
۱/۲۴	۱/۱۴	۱/۵۲	۱/۴۵	۱/۵۸	۱/۶۴	۱/۱۴	۱/۳۷	۲/۳۵	۲/۱۱	۲/۵۶	۲/۴۶	۲/۵۹	۲/۷۲	۲/۱۹	۲/۳۹	ماشین
۱/۲۲	۱/۱۲	۱/۴۶	۱/۴۰	۱/۵۵	۱/۶۲	۱/۱۸	۱/۳۶	۲/۲۰	۲/۰۵	۲/۴۵	۲/۳۵	۲/۵۰	۲/۵۶	۲/۰۲	۲/۲۲	خیابان
۱/۲۱	۱/۱۱	۱/۵۰	۱/۴۴	۱/۵۴	۱/۵۹	۱/۱۶	۱/۳۲	۲/۲۴	۲/۱۰	۲/۵۲	۲/۴۰	۲/۵۸	۲/۶۹	۲/۰۸	۲/۲۰	قطار
۱/۱۰	۱	۱/۴۸	۱/۳۷	۱/۵۳	۱/۵۷	۱/۱۲	۱/۳۸	۲/۱۴	۲	۲/۳۵	۲/۲۷	۲/۴۸	۲/۶۴	۱/۹۷	۲/۱۸	رستوران
۱/۲۸	۱/۱۶	۱/۵۶	۱/۴۵	۱/۵۸	۱/۶۸	۱/۱۸	۱/۴۱	۱/۸۰	۱/۶۵	۲/۸۶	۲/۶۸	۲/۸۹	۳/۰۲	۱/۹۵	۲/۳۱	پیانو

(جدول-۷): نتایج معیار SNR قطعه‌ای به منظور مقایسه عملکرد روش‌های مختلف در سناریوی SD. نتایج روش پیشنهادی در وضعیت

بانظارت و نیمه‌نظارت شده به ترتیب با نماد اختصاری ((Proposed\_Sup)) و ((Proposed\_Semi)) بیان شده است.

(Table-7): The results of frequency weighted segmental SNR values for performance evaluation of different methods in the presence of the mentioned noise signals in SD scenario. The proposed method is introduced using ((Proposed\_Sup)) and ((Proposed\_Semi)) in supervised and semi supervised situations, respectively.

SNR=-۵								SNR=+۵								
GSS	MBSS	MDL	GDL	Proposed_Semi	Proposed_Sup	BWT	AWPT	GSS	MBSS	MDL	GDL	Proposed_Semi	Proposed_Sup	BWT	AWPT	
۱/۴۵	۱/۳۵	۲/۹۴	۲/۵۷	۲/۰۹	۲/۱۸	۱/۳۱	۲/۴۵	۴/۷۰	۴/۵۰	۷/۵۸	۷/۲۳	۷/۵۴	۷/۶۶	۴/۲۰	۶/۵۶	همه‌مه
۱/۸۷	۱/۶۶	۳/۱۱	۲/۷۰	۳/۱۴	۳/۲۰	۱/۳۸	۲/۶۳	۴/۶۱	۴/۳۲	۷/۹۱	۷/۴۵	۷/۹۹	۸/۰۲	۴/۳۲	۸/۷۱	کارخانه

۱/۹۵	۱/۷۰	۳/۳۱	۳	۳/۲۸	۳/۳۷	۱/۸۰	۲/۸۸	۶/۲۱	۶/۰۲	۸/۶۲	۸/۲۲	۸/۶۶	۸/۷۹	۴/۸۹	۷/۲۵	سفید
۱/۷۵	۱/۶۵	۳/۱۷	۲/۹۲	۳/۲۲	۳/۲۹	۱/۹۰	۲/۶۰	۵/۰۳	۵/۱۰	۸/۶۹	۷/۴۱	۸/۷۱	۸/۷۶	۴/۷۵	۶/۶۸	ماشین
۱/۸۵	۱/۶۵	۳/۱۶	۳/۸۸	۳/۱۸	۳/۲۴	۱/۲۸	۱/۶۲	۵/۰۶	۵/۰۸	۸/۳۱	۷/۶۲	۸/۳۶	۸/۴۸	۴/۶۲	۶/۶۱	خیابان
۱/۱۰	۱/۶۶	۳/۱۱	۲/۷۵	۳/۰۹	۳/۲۳	۱/۱۹	۲/۴۸	۴/۸۱	۴/۲۲	۷/۸۹	۷/۲۳	۷/۹۱	۸/۰۱	۴/۵۹	۶/۵۴	قطار
۱/۷۶	۱/۴۲	۳/۰۷	۲/۶۱	۳/۰۸	۳/۱۵	۱/۵۶	۲/۵۹	۴/۸۱	۴/۳۲	۷/۷۵	۷/۳۵	۷/۸۰	۷/۸۶	۴/۴۳	۶/۶۲	رستوران
۱/۲۵	۱/۷۰	۳/۶۴	۳/۵۱	۳/۶۶	۳/۷۹	۱/۴۹	۲/۸۱	۸/۰۶	۶/۵۲	۱۰/۹	۱۰/۳	۱۰/۵	۱۰/۶	۴/۲۳	۶/۳۳	پیانو

(جدول-۸): میانگین نتایج معیار PESQ در شرایط مختلف نوفه‌ای به منظور مقایسه عملکرد روش پیشنهادی در سناریوی SI و SD.  
(Table-8): The average results of PESQ score of the proposed method in SI and SD scenarios in the presence of the mentioned noise signals.

GSS	MBSS	MDL	GDL	Proposed_Semi	Proposed_Sup	BWT	AWPT		
۱/۶۷	۱/۵۵	۱/۹۵	۱/۸۶	۱/۹۶	۲/۰۹	۱/۶۰	۱/۷۹	SD	PESQ
۱/۶۷	۱/۵۵	۱/۸۹	۱/۷۹	۱/۸۸	۲/۰۱	۱/۶۰	۱/۷۹	SI	

(جدول-۹): میانگین نتایج معیار SNR قطعه‌ای در شرایط مختلف نوفه‌ای به منظور مقایسه عملکرد روش پیشنهادی در سناریوی SI و SD.  
(Table-9): The average results of frequency weighted segmental SNR values of the proposed method in SI and SD scenarios in the presence of the mentioned noise signals.

GSS	MBSS	MDL	GDL	Proposed_Semi	Proposed_Sup	BWT	AWPT		
۳/۵۱	۳/۳۰	۴/۴۴	۴/۳۱	۴/۵۳	۴/۶۵	۳/۳۰	۴/۰۶	SD	PESQ
۳/۵۱	۳/۳۰	۴/۳۰	۴/۱۶	۴/۳۵	۴/۴۷	۳/۳۰	۴/۰۶	SI	

(جدول-۱۰): نتایج معیار PESQ و SNR قطعه‌ای در سناریوی SI به منظور مقایسه عملکرد روش پیشنهادی در وضعیت نیمه‌نظارت‌شده، روش پیشنهادی نیمه‌نظارت‌شده بدون گام آشکارساز فعالیت گفتاری VAD و نیز روش پیشنهادی نیمه‌نظارت‌شده بدون گام تطبیق فضا.

(Table-10): The results of PESQ and frequency weighted segmental SNR measures in SI scenario for performance evaluation of the proposed semi supervised method, the proposed semi supervised method without VAD step, and the proposed semi supervised method without domain adaptation technique.

پیانو	رستوران	قطار	خیابان	ماشین	سفید	کارخانه	همیشه			
۲/۸۸	۲/۴۵	۲/۵۳	۲/۳۱	۲/۴۷	۲/۷۵	۲/۵۶	۲/۲۶	+۵	PESQ	Proposed_Semi
۱/۵۵	۱/۵۱	۱/۴۸	۱/۴۴	۱/۴۹	۱/۵۳	۱/۴۵	۱/۴۸	-۵		
۱۰/۱	۷/۶۳	۷/۶۸	۷/۶۹	۸/۶۵	۸/۲۵	۷/۶۹	۷/۴۶	+۵	SegSNR	
۳/۵۷	۲/۹۰	۲/۸۷	۲/۸۳	۲/۹۱	۳/۲۵	۲/۹۱	۲/۷۶	-۵		
۲/۴۹	۲/۱۶	۲/۱۲	۲/۰۷	۲/۱۳	۲/۶۰	۲/۰۸	۱/۹۱	+۵	PESQ	Proposed_Semi بدون VAD
۱/۲۹	۱/۲۱	۱/۱۸	۱/۱۵	۱/۱۷	۱/۳۷	۱/۱۳	۱/۱۸	-۵		
۹/۸۶	۷/۳۱	۷/۲۶	۷/۳۳	۷/۳۱	۷/۹۸	۷/۲۸	۶/۹۹	+۵	SegSNR	
۲/۴۱	۲/۵۱	۲/۴۴	۲/۵۱	۲/۶۲	۳/۱۴	۲/۶۸	۲/۳۶	-۵		
۲/۷۱	۲/۲۸	۲/۲۵	۲/۱۴	۲/۲۳	۲/۶۶	۲/۲۷	۲/۰۴	+۵	PESQ	Proposed_Semi بدون گام تطبیق فضا
۱/۴۸	۱/۳۳	۱/۲۳	۱/۲۲	۱/۲۴	۱/۴۸	۱/۲۲	۱/۲۶	-۵		
۹/۹۷	۷/۴۸	۷/۴۴	۷/۴۳	۷/۴۵	۸/۰۹	۷/۳۵	۷/۱۸	+۵	SegSNR	
۳/۴۹	۲/۶۳	۲/۵۶	۲/۶۱	۲/۷۴	۳/۲۰	۲/۷۷	۲/۵۰	-۵		

(۱۰) بیان شده است. همان‌طور که از این نتایج مشخص است، این دو گام و به‌خصوص استفاده از گام آشکارسازی فعالیت گفتاری پیشنهادی به‌طور محسوس در دستیابی به نتایج مطلوب مؤثر خواهد بود. همان‌طور که از نتایج حاصل مشاهده می‌شود، تأثیر روش تطبیق فضا در دستیابی به نتایج حاصل نیز محسوس است و استفاده از این روش مناسب وضعیتی خواهد بود که یک واژه‌نامه نوفه اولیه در دسترس باشد. همچنین می‌توان مشاهده کرد که تأثیر این دو گام در حضور نوفه دارای ایستایی بیشتر مانند نوفه

به‌منظور بررسی عملکرد بخش‌های مختلف الگوریتم پیشنهادی به‌صورت مجزا، آزمایشی ترتیب داده شده و در آن کارایی بلوک آشکارساز فعالیت گفتاری پیشنهادی و نیز گام تطبیق فضای به‌کارگرفته‌شده مورد ارزیابی قرار گرفته است. در این شبیه‌سازی، نتایج معیار PESQ و SNR قطعه‌ای در سناریوی SI به‌منظور مقایسه عملکرد روش پیشنهادی در وضعیت نیمه‌نظارت‌شده، روش پیشنهادی نیمه‌نظارت‌شده بدون گام آشکارساز فعالیت گفتاری VAD و نیز روش پیشنهادی نیمه‌نظارت‌شده بدون گام تطبیق فضا در جدول

بهسازی گفتار در فضای تبدیل موجک با تخمین نوفه به‌روزشونده AWPT، به‌علت استفاده از مدل‌های ناهمدوس برای مؤلفه‌های داده ورودی نتایج بهتری را به دست می‌دهد. در ادامه به‌منظور تصمیم‌گیری دقیق در مورد کارایی روش‌های مختلف مورد بررسی در شرایط متنوع یادشده، آزمون معناداری آماری به‌کارگرفته شده است. در این مقاله از آزمون Friedman به همراه آزمون Holms post hoc برای مقایسه آماری نتایج بیش از دو الگوریتم استفاده شده است [45,46]. این آزمون بر روی هر هشت روش مورد بررسی در شکل‌ها و جداول بخش شبیه‌سازی و در حضور هشت نوع سیگنال نوفه، سه مقدار SNR (-5dB، 0dB و +5dB) و دو مورد وضعیت SI و SD انجام می‌شود؛ بنابراین در این آزمون، تعداد روش‌ها و شرایط مختلف به ترتیب  $J=8$  و  $I=48$  است. گفتنی است که هر چه تعداد شرایط در مقایسه با تعداد روش‌ها بیشتر باشد، نتایج مناسب‌تری حاصل می‌شود. آزمون Friedman غیرپارامتری یکی از بهترین روش‌ها به منظور مقایسه چند روش در مجموعه شرایط گوناگون است و به فرضیات اولیه نیاز ندارد. به‌عنوان مرور کلی بر روال این آزمون فرض می‌کنیم که  $R_j$  میانگین رتبه عملکرد برای  $j$ -امین روش از میان  $I$  روش مورد بررسی باشد که هر یک از معیارهای عینی یا ذهنی به‌دست آمده و این نتایج در  $I$  شرایط مختلف صورت پذیرفته باشد:

$$R_j = \frac{1}{I} \sum_{i=1}^I r_{ij} \quad (22)$$

که در آن  $r_{ij}$  رتبه عملکرد  $j$ -امین روش در وضعیت  $i$ -ام آزمون خواهد بود. الگوریتم با بهترین عملکرد در این آزمون، کمترین مقدار رتبه را خواهد داشت. آزمون معنادار آماری با این فرضیه پوچ<sup>2</sup> که همه الگوریتم‌ها کارایی یکسان دارند شروع به کار می‌کند [47]. با استفاده از آزمون آماری Friedman، می‌توان تصمیم گرفت که این فرضیه رد یا پذیرفته می‌شود. این آزمون به‌صورت زیر بررسی می‌شود:

$$\chi_F^2 = \frac{12I}{J(J+1)} \left[ \sum_{j=1}^J R_j^2 - \frac{J(J+1)^2}{4} \right] \quad (23)$$

همچنین آمار اصلاح‌شده از آزمون Friedman که براساس توزیع F است با  $(J-1)$  و  $(I-1) \times (J-1)$  درجه آزادی تعریف می‌شود [47]:

سفید، پرلودیک و پیانو بسیار کمتر از سایر نوفه‌های دارای ناایستایی بیشتر است. این مسأله بدین دلیل است که نوفه‌های سفید و پیانو به‌علت دارابودن ایستایی مناسب خوش‌ساختار بوده و با همان واژه‌نامه آموزش‌دیده اولیه قابلیت بازنمایی مناسب را دارا هستند و دیگر نیاز به به‌روزرسانی اتم‌ها بر طبق شرایط محیط آزمون وجود ندارد زیرا شرایط محیط آموزش و نویز در مورد این دسته نوفه‌ها دارای شباهت بیشتری خواهد بود.

نتایج حاصل از الگوریتم پیشنهادی در این مقاله که مبتنی بر یادگیری واژه‌نامه مجزا برای هر یک از زیرباندهای تقریب و جزئیات است، در سناریوی نظارت‌شده بهتر از نتایج سایر روش‌ها بوده است. همچنین نتایج حاصل از سناریوی نیمه‌نظارت‌شده نیز اندکی کمتر از سناریوی نظارت‌شده است. این برتری می‌تواند به‌دلیل بازنمایی هر قاب مشاهده در زیرباندهای مختلف باشد که امکان نمایش مناسب‌تر محتوای سیگنال را به‌دست می‌دهد. این نتایج برای انواع نوفه ساختاریافته، متناوب و بدون ساختار قابل تعمیم است؛ همچنین عامل دیگری که موجب نتایج بهتر این روش نسبت به سایر روش‌ها می‌شود، تخمین نوفه‌ای است که در گام بهسازی صورت می‌پذیرد. این تخمین هم در سناریوی نظارت‌شده و هم نیمه‌نظارت‌شده انجام می‌شود. در سناریوی نظارت‌شده پس از به‌روزرسانی اتم‌های واژه‌نامه نوفه براساس قاب‌های نوفه استخراج‌شده از الگوریتم آشکارساز فعالیت گفتاری<sup>1</sup> مبتنی بر انرژی ماتریس ضرایب تُنک مرحله قبل، شرایط نوفه‌ای فضای آموزش به فضای آزمون نزدیک‌تر شده و اتم‌های نوفه بازنمایی درستی از شرایط محیط در گام انتهایی بهسازی در این سناریو را به‌دست می‌دهند. در سناریوی نیمه‌نظارت‌شده نیز با توجه به تخمین قاب به قاب نوفه با توجه به رابطه (17) و تعیین مقدار آستانه و تابع آستانه‌گذار براساس سطح واریانس نوفه فضای آزمون، حذف نوفه با توجه به شرایط نوفه‌ای محیط آزمون انجام می‌گیرد. در روش MDL که مبتنی بر یادگیری واژه‌نامه است، از واژه‌نامه‌های نوفه یادگیری‌شده در گام آموزش به‌صورت مستقیم در گام بهسازی استفاده می‌شود. در این روش به‌علت عدم انطباق شرایط نوفه‌ای در محیط آموزش و آزمون به‌خصوص در حضور نوفه‌های ناایستا، نتایج حتی در سناریوی نیمه‌نظارت‌شده، پایین‌تر از روش ارائه‌شده در این مقاله خواهد بود؛ همچنین روش پیشنهادی نسبت به روش

<sup>2</sup> Null-hypothesis

<sup>1</sup> Voice activity detector (VAD)

مقایسه این نتایج مشخص می‌شود که تمامی مقادیر  $F_F$  به‌دست آمده در هر آزمون آماری گزارش‌شده در این جداول، بیشتر از مقدار  $F$  بحرانی است؛ بنابراین بر طبق آزمون Friedman، فرضیه اولیه یکسان بودن کارایی آماری روش‌های مدنظر نادرست بوده و می‌توان ادامه آزمون  $\text{post-hoc}$  را به‌منظور تعیین رتبه کارایی روش‌ها دنبال کرد. همان‌طور که از پیش بیان شد،  $R_0$  مربوط به روشی است که بیشترین رتبه میانگین (پایین‌ترین سطح عملکرد) را در میان روش‌های مختلف دارد که در بررسی‌های انجام‌شده متعلق به روش BWT است که به الگوریتم پایه بهسازی گفتار مبتنی بر تبدیل موجک می‌پردازد. به‌منظور محاسبه پارامتر  $Z$  برای محاسبه مقدار  $\rho$  در سطح معناداری آماری  $\alpha=0.05$ ، از رابطه (۲۵) استفاده شده است؛ همچنین مقدار  $i$  در  $(\alpha/(J-i))$  (به‌منظور محاسبه پارامتر Holm، برابر با شماره ردیف روش مورد بررسی در جدول از ۱ (روش با عملکرد بهتر) تا  $J-1$  (روش با عملکرد پایین‌تر) برای روش‌های مختلف بدون در نظر گرفتن روش متناظر با  $R_0$  خواهد بود. در این جداول با توجه به اینکه مقدار  $\rho$  برای تمامی روش‌ها، کمتر از مقدار بحرانی Holm متناظر به‌دست آمده است، می‌توان نتیجه گرفت که روش‌های مورد بررسی به‌ترتیب قرار گرفته شده در جدول، از روش با بهترین عملکرد تا روش با بدترین عملکرد در زمینه بهسازی گفتار کارایی دارند. نتایج آزمون آماری بیان می‌کند که رتبه میانگین برای تمامی وضعیت‌های در نظر گرفته‌شده در شرایط آزمون برای روش پیشنهادی نظارت‌شده، نزدیک به یک بوده که موجب می‌شود این روش در جداول (۱۱ و ۱۲) در بالاترین ردیف قرار گیرد.

$$F_F = (I - 1)\chi_F^2 / (I(J - 1) - \chi_F^2) \quad (24)$$

در این آزمون آماری، اگر مقدار  $F_F$ ، بزرگ‌تر از مقدار بحرانی  $\chi_F^2$  باشد، فرضیه پوچ رد خواهد شد؛ یعنی کارایی این  $J$  الگوریتم مشابه نیست؛ بنابراین باید تعیین شود که کدام الگوریتم عملکرد بهتری دارد که این کار توسط آزمون  $\text{post-hoc}$  انجام می‌گیرد. پارامتر  $Z_j$  برای هر روش مورد ارزیابی این آزمون به‌صورت زیر محاسبه می‌شود [47]:

$$Z_j = (R_0 - R_j) / \sqrt{J(J + 1) / 6I} \quad (25)$$

$R_0$  مربوط به روشی است که بیشترین رتبه میانگین (پایین‌ترین سطح عملکرد) را دارد. مقدار  $Z$  به منظور محاسبه مقدار  $\rho$  در سطح معناداری آماری  $\alpha=0.05$  مورد استفاده قرار می‌گیرد. در چگالی احتمال توزیع نرمال استاندارد، مقدار  $\rho$  متناظر با سطح زیر این توزیع و خارج از محدوده  $(-Z, Z)$  به‌صورت زیر خواهد بود:

$$\rho = 1 - \int_{-Z}^Z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (26)$$

نتایج شبیه‌سازی با استفاده از آزمون آماری براساس معیار ارزیابی Holm برای روش‌های بهسازی گفتار معرفی‌شده در تمامی شرایط نوفه‌ای با استفاده از معیارهای PESQ و SNR قطعه‌ای، در جدول (۱۱ و ۱۲) بیان شده است. مقادیر پارامتر  $\chi_F^2$  حاصل از این جداول براساس رابطه (۲۳) به‌ترتیب برابر با  $45/81, 58/37$  است؛ همچنین مقدار پارامتر  $F_F$  با توجه به رابطه (۲۴) در این جداول به‌ترتیب  $23/62, 22/08$  خواهد بود. مقدار  $F$  بحرانی با درجه‌های آزادی  $(8-1)$  و  $(48-1) \times (8-1)$  برابر با  $2/0374$  است. از

(جدول-۱۱): نتایج آزمون آماری برای معیار PESQ در شرایط و نوفه‌های مختلف.

(Table-11): The results of the statistical Friedman and Holm's post hoc test for PESQ score over different methods and conditions.

Methods	Average Rank ( $R_j$ )	$Z$	$\rho$ -value	Holm( $\alpha/(J-i)$ )
Proposed_Sup	۱/۲۸	۷/۹۱	۰	۰/۰۰۷۱
Proposed_Semi	۱/۹۱	۶/۴۸	۰	۰/۰۰۸۳
MDL	۲/۳۵	۵/۴۱	۰	۰/۰۱۰۰
GDL	۲/۹۶	۴/۳۳	۰	۰/۰۱۲۵
$\Delta$ WPT	۳/۲۱	۴/۲۸	۰	۰/۰۱۶۷
MBSS	۳/۶۷	۳/۲۲	۰/۰۰۱۳	۰/۰۲۵۰
GSS	۴/۱۲	۲/۱۸	۰/۰۲۹۳	۰/۰۵۰۰
BWT( $R_0$ )	۲/۳۱	-	-	-

(جدول-۱۲): نتایج آزمون آماری برای معیار SNR قطعه‌ای در شرایط و نوبه‌های مختلف.

(Table-12): The results of the statistical Friedman and Holm's post hoc test for of frequency weighted segmental SNR values over different methods and conditions.

Methods	Average Rank ( $R_j$ )	Z	$\rho$ -value	Holm( $\alpha/(J-i)$ )
Proposed_Sup	۱/۲۴	۸/۲۳	۰	۰/۰۰۷۱
Proposed_Semi	۲/۰۴	۶/۹۱	۰	۰/۰۰۸۳
MDL	۲/۵۶	۵/۹۲	۰	۰/۰۱۰۰
GDL	۲/۸۳	۵/۳۸	۰	۰/۰۱۲۵
AWPT	۳/۱۲	۴/۹۲	۰	۰/۰۱۶۷
MBSS	۳/۷۵	۳/۶۱	۰/۰۰۰۳	۰/۰۲۵۰
GSS	۱/۲۶	۲/۶۷	۰/۰۰۷۶	۰/۰۵۰۰
BWT( $R_0$ )	۵/۵۲	-	-	-

تُنک حاصل، قاب ورودی بازسازی می‌شود. هر چقدر خطای این تقریب، میان قاب مشاهده شده و قاب بازسازی شده، کمتر باشد به این معنا است که قاب ورودی بر روی واژه‌نامه گفتار بازنمایی تُنک مناسب داشته و می‌توان نتیجه گرفت که قاب، ماهیت گفتاری خواهد داشت؛ سپس به منظور اطمینان بیشتر به تصمیم گرفته شده، انرژی بازنمایی تُنک قاب ورودی بر روی قاب‌های ابتدایی و انتهایی سیگنال گفتار محاسبه می‌شود. کوچک بودن انرژی محاسبه شده، صحت نتیجه حاصل از شرط قبل را تأیید می‌کند. در سناریوی نظارت شده که واژه‌نامه نوبه نیز در دسترس است، پس از انتخاب واژه‌نامه مرتبط، انرژی ضرایب تُنک حاصل از بازنمایی تُنک در زیرباندهای مختلف محاسبه، سپس از بررسی‌های انجام شده در جهت تصمیم‌گیری درست در تعیین نوع قاب ورودی استفاده می‌شود. در ادامه نتایج خروجی آشکارساز بسته به سناریوی مدنظر، طی روال بهسازی پیشنهاد شده، سیگنال گفتار تخمینی را به دست خواهد داد. نتایج گزارش شده بر مبنای معیارهای مورد بررسی، توانایی این روش نسبت به سایر روش‌های پایه ارائه شده را در فضای ویژگی موجک و دیگر روش‌های بهسازی گفتار مبتنی بر واژه‌نامه نشان می‌دهد.

بنابراین می‌توان نتیجه گرفت که روش پیشنهادی نظارت شده در فضای ویژگی تبدیل موجک برای بهسازی گفتار، بهترین رتبه میانگین در میان سایر روش‌ها را دارد و به صورت آماری بهتر از سایر الگوریتم‌ها عمل می‌کند. همچنین روش پیشنهادی نیمه نظارت شده در فضای ویژگی تبدیل موجک رتبه بعدی را به خود اختصاص می‌دهد. از آنجایی که در این جداول الگوریتم BWT با تابع آستانه‌گذاری نرم به عنوان روش با میانگین رتبه  $R_0$  معرفی شده و مقدار پارامتر Z برای آن صفر به دست می‌آید، مقدار پارامتر  $\rho$  برای آن قابل محاسبه نخواهد بود.

## ۵- نتیجه گیری

در این مقاله، یادگیری واژه‌نامه به منظور دست‌یابی به دقت بالاتر در بهسازی سیگنال گفتار و در فضای تبدیل موجک بررسی شده است. به کمک این تبدیل، امکان تجزیه سیگنال در زیرباندهای مختلف تقریب و جزئیات سیگنال امکان پذیر خواهد بود. تجزیه سیگنال به زیرباندها می‌تواند اطلاعات دقیقی از محتوای سیگنال به دست دهد. یادگیری واژه‌نامه و یافتن اتم‌هایی که با تقریب مناسبی قادر به مدل کردن این زیرباندها هستند، می‌تواند تحلیل جزئی تری را به دست داده و نتایج مطلوبی را در زمینه بهسازی گفتار فراهم کند. همچنین دو سناریوی نظارت شده و نیمه نظارت شده ارائه شد که در هر یک، الگوریتم آشکارساز فعالیت گفتاری با توجه به شرط‌های مخصوص به خود بر پایه محاسبه انرژی ضرایب تُنک حاصل از بازنمایی سیگنال مشاهده شده بر روی واژه‌نامه‌ها در گام آموزش، پیشنهاد شد. در سناریوی نیمه نظارت شده در ابتدا بازنمایی تُنک قاب مشاهده شده بر روی واژه‌نامه گفتار انجام می‌شود و از روی ماتریس ضرایب

## 6- References

## ۶- مراجع

- [1] M. Klein, P. Kabal, "Signal subspace speech enhancement with perceptual post-filtering", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process. (ICASSP)*, Vol. 1, pp. 537-540, 2002.
- [2] S. Kamath, P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise", In: *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process. (ICASSP)*, Orlando, Florida, 2002.



- [16] M.A. Messaoud, A. Bouzid, "Speech enhancement based on wavelet transform and improved subspace decomposition", *Journal of Audio Engineering society (JAES)*, Vol. 63, No.12, pp.1-11, 2015.
- [17] C.L. Wu, H.P. Hsu, S.S. Wang, J.W. Hung, Y.H. Lai, H.M. Wang, Y. Tsao, "Wavelet speech enhancement based on robust principal component analysis", *Proc. Interspeech*, 781, pp. 439-443, 2017.
- [18] T.Y. Zuo, L. He, W.D. Sheng, "A new algorithm of the wavelet packet speech denoising based on masking perception model", *7th International conference on natural computation (ICNC)*, Vol. 1, 2011, pp. 33-37.
- [19] H. Zhao, X. Peng, L. Hu, G. Wang, "An improved speech enhancement method based on teager energy operator and perceptual wavelet packet decomposition", *Journal of Multimedia*, Vol. 6, No. 3, 2011.
- [20] R. Gomez, T. Kawahara, "Optimized wavelet-based speech enhancement for speech recognition in noisy and reverberant conditions", *APSIPA ASC*, 2011.
- [21] T.F. Sanam, C. Shahnaz, "Teager energy operation on wavelet packet coefficients for enhancing noisy speech using a hard thresholding function", *Published in Signal Processing: An International Journal (SPIJ)*, Vol. 6, pp. 22-43, 2011.
- [22] G. Chen, C. Xiong, J.J Corso, "Dictionary transfer for image denoising via domain adaptation", *In Proceedings of IEEE International Conference on Image Processing*, 2012.
- [23] S. Mavaddati, S.M. Ahadi Sarkani, S. Seyedin, "A novel speech enhancement method by learnable sparse and low-rank decomposition and domain adaptation", *Speech Communication*, Vol. 76, pp. 42-60, 2016.
- [24] A. Agarwal, A. Anandkumar, P. Jain, P. Netrapalli, R. Tandon, *JMLR: Workshop and Conference Proceedings*, Vol. 35, 2014, pp. 1-15.
- [25] H. Lee, A. Battle, R. Raina, A.Y. Ng, "Efficient sparse coding algorithms", *Advances in Neural Information Processing Systems*, 2006.
- [26] J. Portilla, L. Mancera, "L<sub>0</sub>-based sparse approximation: Two alternative methods and some applications", *Proceedings of the 16th IEEE international conference on Image processing*, 2009, pp. 3865-3868.
- [27] مودتی، سمیرا، احدی، محمد، "بهسازی گفتار به کمک یادگیری واژه‌نامه مبتنی بر داده"، پردازش علائم و داده‌ها، جلد ۱۷، شماره ۱، صفحه ۹۹-۱۱۶، ۱۳۹۹.
- [3] D.L. Donoho, "De-noising by soft-thresholding", *IEEE Trans. Inf. Theory*, Vol. 41, No. 3, pp. 613-627, 1995.
- [4] N. Upadhyay, R.K. Jaiswal, "Single Chamel Speech Enhancement: Using Wiener Filtering with Recursive Noise Estimation", *Procedia Computer Science*, Vol. 84, pp. 22-30, 2016.
- [5] J. Candes, M.B. Wakin, "An introduction to compressive sampling", *IEEE Signal Processing Magazine*, pp. 21-30, 2008.
- [6] R.G. Baraniuk, "Compressive Sensing", *IEEE Signal Processing Magazine*, pp. 118-121, 2007.
- [7] S. Ayat, R. Dianat, M. Manzuri, "Wavelet Based Speech Enhancement Using a New Thresholding Algorithm", *IEEE Intl. Symposium on Intelligent Multimedia, Video & Speech Processing (ISLMP)*, Hong Kong, 2004.
- [8] C.T. Lu, H.C. Wang, "Speech enhancement using wavelet transform with constrained thresholds", *In Proc. The 3<sup>rd</sup> International Symposium on Chinese Spoken Language Processing (ISCSLP)*, Taipei, Taiwan, pp. 185-188, 2002.
- [9] E. Ambikairajah, G. Tattersall, A. Davis, "Wavelet Transform-based Speech Enhancement", *Proc. on ICSLP*, Vol. 3, 1998.
- [10] V.S.R. Kumari, D.K. Dcvarakonda, "A Wavelet Based Denoising of Speech Signal", *International Journal of Engineering Trends and Technology (IJETT)*, Vol. 5, No. 2, pp. 107-115, 2013.
- [11] K. Khaldi, A.O. Boudraa, A. Komaty, "Speech enhancement using empirical mode decomposition and the Teager-Kaiser energy operator", *J Acoust Soc Am*, Vol. 13, No. 5, pp. 451-459, 2014.
- [12] T.F. Sanam, C. Shahnaz, "A semisoft thresholding method based on Teager energy operation on wavelet packet coefficients for enhancing noisy speech", *EURASIP Journal on Audio, Speech and Music Processing*, Springer, 2013.
- [13] S. Hongo, S. Sakamoto, Y. Suzuki, "Binaural speech enhancement method by wavelet transform based on interaural level and argument differences", *International Conference on Wavelet Analysis and Pattern Recognition*, Xian, 2012, pp. 290-295.
- [14] T. V. Pham, "Wavelet analysis for robust speech processing and applications", *PHD Thesis*, Graz University of Technology, 2007.
- [15] I. Pinter, "Perceptual wavelet-representation of speech signals and its application to speech enhancement", *Computer Speech and Language*, Vol. 10, No. 1, pp. 1-22, 1996.

- [39] A. Varga, H. Steeneken, J.M. Tomlinson, D. Jones, "The Noisex-92 study on the effect of additive noise on automatic speech recognition", *Technical Report. Malvern*, U.K.: DRA Speech Res. Unit, 1992.
- [40] H.G. Hirsch, D. Pearce, "The AURORA experimental framework for the performance evaluations of speech recognition systems under noisy conditions", *Proc. ISCA ITRWASR*, pp.181-188, 2000.
- [41] <http://pianosociety.com>.
- [42] Y. Lu, P.C. Loizou, "A geometric approach to spectral subtraction", *Speech communication*, Vol. 50, No. 6, pp. 453-466, 2008.
- [43] Y. Ghanbari, M.R. Karami Mollaei, "A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets", *Speech communication*, Vol. 48, No. 40, pp. 927-940, 2006.
- [44] S. Mavaddati, S.M. Ahadi Sarkani, S. Scyedin, "Modified coherence-based dictionary learning method for speech enhancement", *Signal Processing*, IET, Vol. 9, No. 7, pp. 1-9, 2015.
- [45] J. Benesty, *Springer handbook of speech processing*, Springer's publication, pp. 843-871, 2008.
- [46] J. Demsar, "Statistical comparisons of classifiers over multiple data set", *The Journal of Machine Learning Research*, Vol. 7, pp. 1-30, 2006.
- [47] D.J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, 4nd cd. Boca Raton, FL: Chapman & Hall/CRC, 2000.
- [27] S.Mavaddati, M. Ahadi, "Speech Enhancement using Adaptive Data-Based Dictionary Learning", *JSDP*, vol. 17 (1), pp. 99-116, 2020.
- [۲۸] مظفری، رضا، مودّتی، سمیرا، "ارائه روش جدید حذف نوفه تصویر براساس یادگیری واژه‌نامه ناهمدوس و روش تطبیق فضا"، پردازش علائم و داده‌ها، جلد ۱۶، شماره ۴، صفحه ۷۳-۹۲، ۱۳۹۸.
- [28] R. Mozaffari, S. Mavaddati, "A Novel Image Denoising Method Based on Incoherent Dictionary Learning and Domain Adaptation Technique", *JSDP*, vol. 16 (4), pp.73-92. 2020.
- [29] 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.
- [30] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, "Least angle regression", *Ann. Stat.*, Vol. 32, pp. 407-499, 2004.
- [31] 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.
- [32] X. Wu, D. Yu, "Atomic Decomposition Method Based on Adaptive chirplet Dictionary", *Advances in Adaptive Data Analysis*, Vol. 4, pp. 1-19, 2012.
- [33] J. Tropp, I. Dhillon, R.J. Heath, T. Strohmer, "Designing structural tight frames via an alternating projection method", *IEEE Trans. on Information Theory*, Vol. 51, No.1, pp. 188-209, 2005.
- [34] M. Sustik, J. Tropp, I. Dhillon, R. Heath, "On the existence of equiangular tight frames", *Linear Algebra and Its Applications*, Vol. 426, No. 2, pp. 619-635, 2007.
- [35] D. Liu, J. Nocedal, "On the limited memory BFGS method for large scale optimization", *Math. Program*, Vol. 45, pp. 503-528, 1989.
- [36] S. Mavaddati, S.M. Ahadi Sarkani, S. Scyedin, "Speech enhancement using sparse dictionary learning in wavelet packet transform domain", *Computer Speech and Language*, Vol. 44, pp. 22-47, 2017.
- [37] D.L. Donoho, "De-noising by soft-thresholding", *IEEE Trans. Inf. Theory*, Vol. 4, No. 3, pp. 613-627, 1995.
- [38] <http://www.dcs.shef.ac.uk/spandh/gridcorpus>.



سمیرا مودّتی مدارک کارشناسی و

کارشناسی ارشد خود را به ترتیب در

سال‌های ۱۳۸۶ و ۱۳۸۹ از دانشگاه

مازندران در رشته مهندسی برق-

الکترونیک و درجه دکترای خود را در

سال ۱۳۹۵ از دانشگاه صنعتی امیرکبیر در رشته مهندسی

برق- الکترونیک دریافت کرد. وی هم‌اکنون استادیار گروه

مهندسی برق 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.i**