Semnan University

Journal of Modeling in Engineering

Journal homepage: https://modelling.semnan.ac.ir/

ISSN: 2783-2538

Applying Dictionary Learning Algorithms In Sparse Representation of Speech Signals

Naser Sharafi [1], Salman Karimi [1,*], Samira Mavaddati [2]

[1] *Student, Faculty of Engineering, Lorestan University, Khorramabad, Iran.*
[2] *Faculty of Engineering and Technology, University of Mazandaran, Babolsar, Iran*

**Corresponding author: karimi.salman@lu.ac.ir*

Received: 2024-05-07 Revised: 2025-03-29 Accepted: 2025-06-24

Abstract

Sparse representation is one of the most widely used techniques in the area of signal processing, and receives a great deal of attention in various applications such as data compression, speech and image denoising, pattern recognition, and other signal processing tasks. Sparse representation describes a signal as a linear combination of only a few atoms selected over a redundant dictionary, so that the data dimensions are reduced and the data processing becomes more efficient. Good representation of speech signals heavily relies on the ability of the dictionary used to fit the essential nature of speech data. In this work, several dictionary learning algorithms such as K-SVD, MOD, RAMC, UD4-MOD and OMP are exploited in different representation domains such as time domain, wavelet transform domain and short-time Fourier transform domain. They are assessed by different objective measures such as the reconstruction error (RE), the mean square error (MSE), the frequency-weighted segmental signal-to-noise ratio (fwSegSNR), the segmental signal-to-noise ratio (Seg SNR), the perceptual evaluation of speech quality (PESQ) and short-time objective intelligibility (STOI). The results show that the combination of K-SVD in STFT domain with OMP sparse representation method achieves better results with an efficient speech signal reconstruction.

Keywords :
Sparse Representation, Dictionary Learning, Speech Processing, K-SVD, OMP, STFT.

1. Introduction

Dictionary learning and sparse representation are two prominent techniques in the areas of machine learning and signal processing [2, 3], using which we can analyze and process the high-dimensional data intelligently. The goal of dictionary learning is to find a set of basis functions, i.e. the dictionary atoms, by learning from training samples, which can be used to represent the training samples sparsely. Sparse representation is to represent a data as the linear combination of a small number of basis vectors selected from an overcomplete dictionary, thereby reducing the dimensionality of the signal and enabling powerful signal processing by simple coding and matching pursuits. They are subject to increased interests since the birth of compressed sensing theory in 2006 [1].

Sparse representation has been successfully used in many fields, such as image processing, speech enhancement and denoising, pattern recognition, medical image analysis and speech processing applications [2–5]. In speech processing applications, sparse representation performance is almost directly impacted by dictionary learning and the representation domain. For this reason, relevant researches have been dedicated to enhance dictionary learning schemes and sparse coding algorithms for speech reconstruction and enhancement.

Recently, sparse representation techniques have been applied in the domain of speech enhancement in various transform domains. Several recent researches are mostly focused on using wavelet-based sparse representation techniques [6] for speech denoising and enhancement. Several time-frequency sparse representation approaches have also been explored for voice activity detection and speech enhancement in noisy environments [7–10]. Apart from this, K-SVD dictionary learning algorithms integrated with the orthogonal matching pursuit (OMP) approach have been effectively used for speech applications due to its efficiency of dealing with high dimensional speech data [11].

In this paper, different dictionary learning and sparse representation algorithms for speech signal in the time, wavelet transform and short-time Fourier transform (STFT) domains are explored, which includes MOD, K-SVD, RAMC, UD4-MOD, and OMP. Different objective assessment measures are used to evaluate the trained dictionaries to find the best representation domain and dictionary learning algorithm for speech signal reconstruction.

2. Methodology

In the presented approach, the efficiency of various dictionary learning algorithms will be studied in terms of sparse representation and reconstruction of speech signals in various feature domains. In the study, four algorithms for learning dictionaries: MOD (Method of Optimal Directions), K-SVD, RAMC (Re- weighted Alternating Minimization of the Coefficients) and UD4-MOD, were tested with the Orthogonal Matching Pursuit (OMP) sparse coding algorithm. The block diagram of the procedure is shown in the figure 1. Here, the speech signal is transformed to the corresponding feature domain and then processing by the dictionary learning and sparse coding algorithms for the signal representation and reconstruction.

Speech samples from the NOIZEUS database processed at 8kHz that were used for all experiments. In the pre processing stage, speech signal was frame-blocked by 12.5 msec Hamming windows with 50% overlapping. Dictionary training and sparse representation were performed within three various feature domains:

- Time domain
- Wavelet Packet Transform (WPT) domain
- The STFT domain

A redundancy of 4, namely overcomplete dictionary was used which was experimentally optimal for this application trade-offs. The following problem was used to get the sparse representation:

$$X^* = \underset{X}{\operatorname{argmin}} \|Y - DX\|_F^2 , \|X\|_0 \leq K \tag{1}$$

where (Y) denotes the speech feature matrix, (D) represents the learned dictionary, and (X) is the sparse coefficient matrix.

Were used to evaluate the quality of the reconstructed speech: Reconstruction Error (RE), Mean Square Error (MSE), Segmental Signal-to-Noise Ratio (SegSNR), Frequency Weighted Segmental SNR (fwSeg SNR), Perceptual Evaluation of Speech Quality (PESQ), and Short Time Objective Intelligibility (STOI).

For the Speech Enhancement experiments, the STFT domain was used due to better reconstruction quality. Different dictionaries were trained for the clean speech and noise signals (white, car and street). Finally the entire dictionary was used to estimate sparse coefficients and reconstructed the enhanced speech signals under noisy conditions (0, 5 and 10dB SNR).

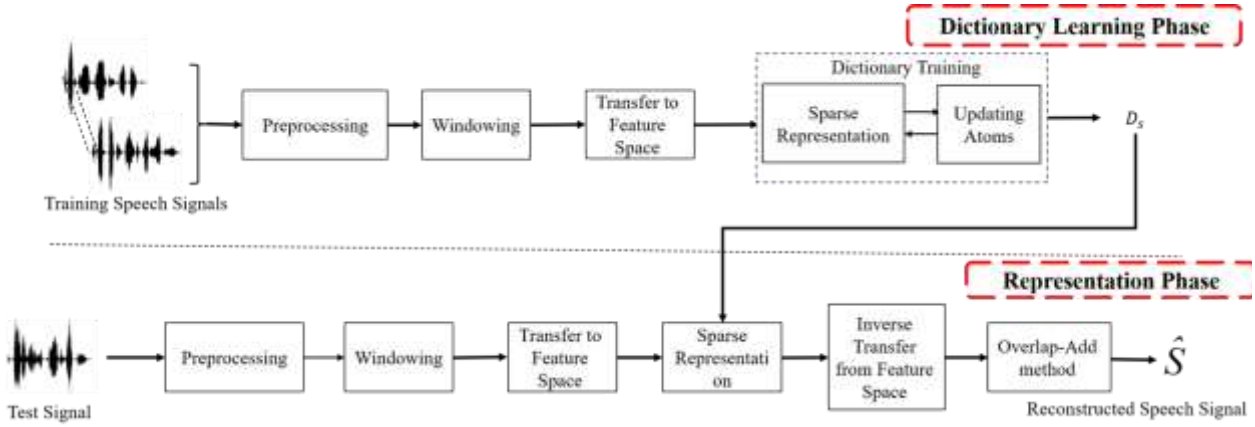


Figure 1. Framework of speech representation based on dictionary learning and sparse coding

3. Discussion and Results

The dictionary learning algorithms in the time,WPT and STFT domains were tested. The best reconstruction performance reached by the algorithms in each domain is summarized on table 1.

Table 1. Best reconstruction results obtained in different feature domains.

Feature Domain	Best Algorithm	PESQ	STOI	SegSNR (dB)
Time Domain	UD4-MOD	3.93	0.94	17.61
WPT Domain	MOD (Method 1)	4.11	0.98	19.83
STFT Domain	K-SVD	4.43	0.99	33.56

The results show that the use of dictionary learning is very beneficial for speech representation quality. In time domain, UD4-MOD had the best Reconstruction error among all the algorithms considered for all methods. The performance of Method 1 in WPT domain is slightly better than Method 2 which emphasizes on the importance of

segmentation being carried out before wavelet packet decomposition.

Figure 2 presents the speech reconstructions from different feature domains. The optimal reconstruction was in the STFT domain, especially with the application of the K-SVD learning algorithm. It yielded the highest PESQ and STOI values as well as SegSNR, indicating the best perceived speech quality and intelligibility.

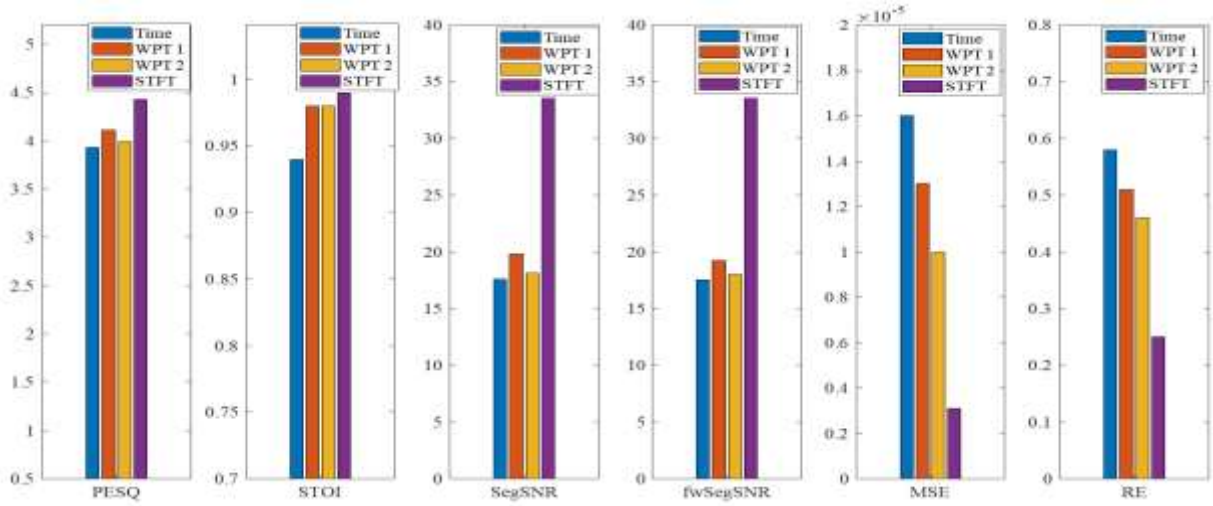
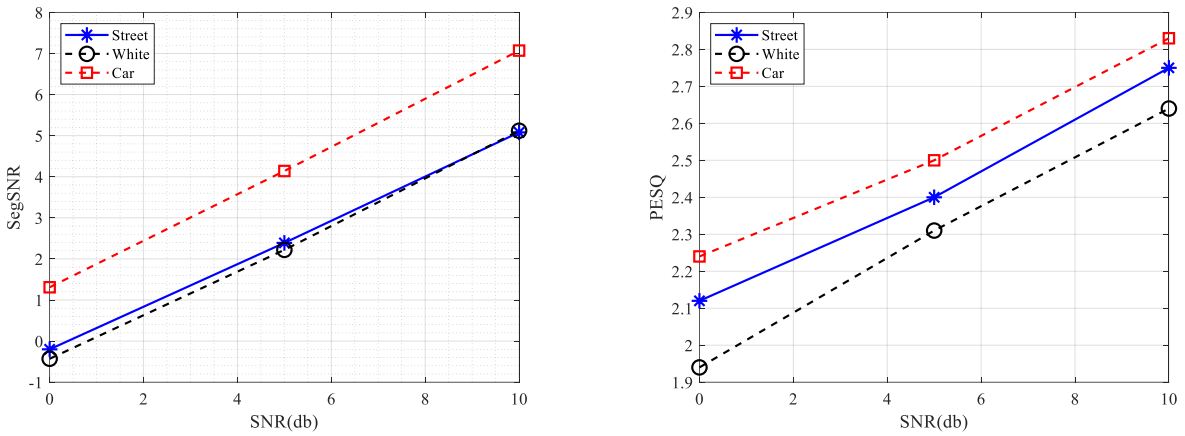


Figure 2. Comparison of speech reconstruction performance in Time, WPT, and STFT domains.

In order to test the practical usefulness of the approach, speech enhancement experiments for white, car and street noise were performed. The STFT based K-SVD dictionaries captured the spectral features of both speech and noise, providing a better reconstruction of the signal after sparse decomposition. The enhancement performance at different input SNRs was evaluated in fig. 3. For all three types of noises, the proposed method gains steady improvement of PESQ and SegSNR when the input SNR increased from 0 d B to 10 d B, which indicate the approach is a robust and effective method for speech enhancement.



B

Figure 3. Speech enhancement results in terms of PESQ and SegSNR for different noise types and input SNR levels.

4. Conclusions

We studied how dictionary learning algorithms can be used for the sparse coding and reconstruction of speech signals in various feature domains. We assessed the performance of the algorithms MOD, K-SVD, RAMC, and UD4-MOD by means of objective measures of speech quality and intelligibility.

The experimental results indicated that the feature domain has a big influence to the reconstruct quality. The best results in the time domain were obtained by UD4-MOD, it was found that the best results in the wavelet packet transform domain are achieved by the MOD, and the best results in the short-time Fourier transform domain was found to be the STFT.

Of all the approaches tested, the joint application of K-SVD dictionary learning and OMP sparse coding achieved the highest standard of performance , with the minimum reconstruction error and the maximum PESQ, ST0I and

SegSNR values.

In addition, the proposed sparse representation approach has been further evaluated with a speech enhancement experiment in the presence of several different noise types. The learned speech and noise dictionaries enabled a fast separation of the speech components and produced significant gains in speech quality and intelligibility.

5. References

[1] Tsaig, Yaakov, and David L. Donoho. "Extensions of Compressed Sensing." Signal Processing 86, no. 3 (2006): 549–571.

[2] Zhao, Yongqiang, and Jingxiang Yang. "Hyperspectral Image Denoising via Sparse Representation and Low-Rank Constraint." IEEE Transactions on Geoscience and Remote Sensing 53, no. 1 (2015): 296–308.

[3] Yang, Meng, Dengxin Dai, Linlin Shen, and Luc Van Gool. "Latent Dictionary Learning for Sparse Representation Based Classification." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4138–4145. 2014.

[4] iu, Yu, Xun Chen, Aiping Liu, Rabab K. Ward, and Z. Jane Wang. "Recent Advances in Sparse Representation Based Medical Image Fusion." IEEE Instrumentation & Measurement Magazine 24, no. 2 (2021): 45–53.

[5] Sharma, Pulkit, Vinayak Abrol, and Anil Kumar Sao. "Deep-Sparse-Representation-Based Features for Speech Recognition." IEEE/ACM Transactions on Audio, Speech, and Language Processing 25, no. 11 (2017): 2162–2175.

[6] Mavaddaty, Samira, Seyed Mohammad Ahadi, and Sanaz Seyedin. "Speech Enhancement Using Sparse Dictionary Learning in Wavelet Packet Transform Domain." Computer Speech & Language 44 (2017): 22–47.

[7] Eshaghi, Mohadese, Farbod Razzazi, and Alireza Behrad. "A Voice Activity Detection Algorithm in Spectro-Temporal Domain Using Sparse Representation." International Journal of Machine Learning and Cybernetics 10, no. 7 (2019): 1791–1803.

[8] Sugiura, Yosuke, and Tetsuya Shimamura. "Speech Enhancement Based on Sparse Representation in Logarithmic Frequency Scale." In 2018 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), 252–257. 2018.

[9] Shaheen, Dima, Oumayma Al Dakkak, and Mohiedin Wainakh. "Incoherent Discriminative Dictionary Learning for Speech Enhancement." Journal of Telecommunications and Information Technology 3 (2018): 42–54.

[10] Ji, Yunyun, Wei-Ping Zhu, and Benoit Champagne. "Speech Enhancement Based on Dictionary Learning and Low-Rank Matrix Decomposition." IEEE Access 7 (2019): 4936–4947.

[11] Wang, Lianzi, Nikos Mastorakis, and Xiaodong Zhuang. "Voiced/Unvoiced Pronunciation Judgement Based on Sparse Representation and Learning Dictionary." In MATEC Web of Conferences, vol. 292, 04012. 2019.