

SPEECH ENHANCEMENT BASED ON LABEL CONSISTENT K-SVD UNDER NOISY ENVIRONMENT

Ching-Tang Hsieh^{1, a}, Cheng-Yuan Chiang^{1, b}, Ting-Wen Chen^{1, c}

¹Department of Electrical Engineering, Tamkang University, New Taipei, 25137, Taiwan

^aemail:hsieh@ee.tku.edu.tw, ^bemail:603440263@s03.tku.edu.tw, ^cemail:alucardoom@hotmail.com

Keywords: Speech enhancement, sparse representations, K-SVD, Label Consistent K-SVD(LCKSVD).

Abstract. The sparse algorithm for sparse enhancement is more and more popular issues, recently. In previous research, the sparse algorithm for sparse enhancement will spend much time, so we propose LC K-SVD(Label Consistent K-SVD) to reduce spending time. We focus on the White Gaussian Noise. The experiments show that denoising performance of our proposed method is very closed to sparse algorithm in SNR, LLR, SNRseg and PESQ, even better then it. Our method only need half time then sparse algorithm.

Introduction

Speech is the most important tool for the people who communicate with each other. If we need to use machine for communicate, we need speech processing to help us. For example speech recognition system. When the speech is in noisy environment, noise will make speech recognition rate decreased. So speech enhancement processing is necessary. Speech enhancement research methods such as Kalman filter [1], spectral subtraction [2], Wiener filter [2] have been proposed. These methods have some effect on speech enhancement.

Recently, more and more researchers concerned about the sparse representations issue. The primary thing is dictionary learning. Michal Aharon, Michael Elad and Alfred Bruckstein proposed the K-SVD method [3]. In the dictionary updating steps, we can update the dictionary and its coefficients together. Ching-Tang Hsieh and Yan-heng Chen apply sparse theory to speech enhancement.[4] The experimental results show that their proposed method is superior than methods above mentioned, but spending time is lengthy.

We purpose LC K-SVD(Label Consistent K-SVD)[5] [6] to reduce spending time. In Section II, we will introduce the LC K-SVD and how it works. Section III is experimental result, we compare our proposed method with Hsieh's sparse theory. We discuss conclusion and future research works in Section IV.

LC K-SVD

First, we use SBAV (Sub-Band Amplitude Variance) algorithm[7] to classify two label, unvoiced and voiced of speech. Then, we slide a window to divide the sequence of noisy speech signal into N frames, the window length is K, shift K/2 per slide, and then stored in an matrix Y. Second, we put label and matrix Y into LC K-SVD, then we obtain the updated dictionary D and the sparse coefficients representations of matrix X, that belonging to dictionary D. The training process is given in Fig.1.

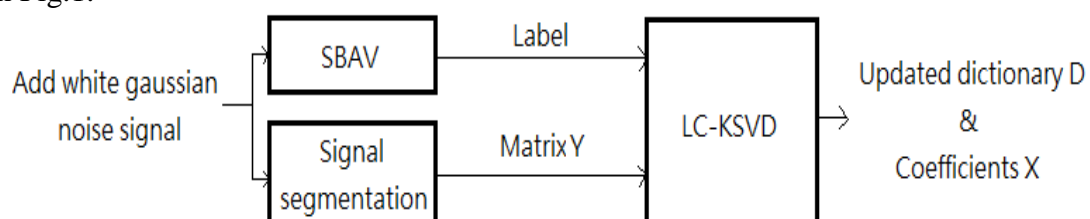


Fig.1 The training process of updated dictionary D and coefficients X.

Finally, we multiply two matrixes D and X to reconstruct the clean speech signal. The reconstruction process of clean speech signal is given in Fig.2.

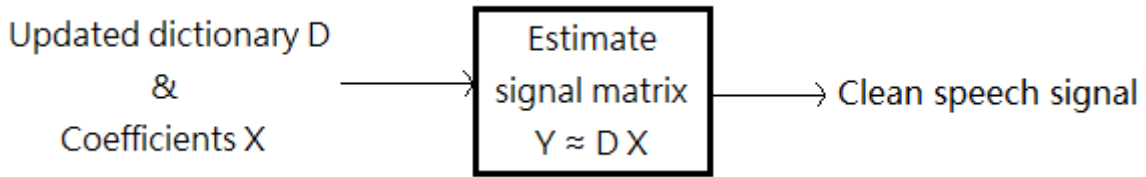


Fig.2 The reconstruction process of clean speech signal.

Each input signal will obtain their learning dictionary and a set of coefficients. Then we use the trained dictionary D and reconstructed coefficients X to estimate the clean speech signal.

Test results

We use four kind of objective quality measures [8]-[9] to evaluate the effect of denoising signal, such as the SNR, Log-Likelihood Ratio (LLR), segmental SNR (SNRseg) and Perceptual Evaluation of Speech Quality (PESQ).

The clean speeches are taken from CHIME data [10] which includes 600 speeches by 34 speakers reading 6 sequences of the command-color-preposition-letter-number-adverb. All data have a 16kHz sampling rate. Input signal will be limited to the amplitude range between -1 to 1. The speech signal will pass a high-pass filter to eliminate the effect of lips and vocal cords during phonation.

We add with white gaussian noise at SNR levels of -10, -5, 0, 5 and 10 dB into 600 spechs. Then all spechs use LC K-SVD and sparse KSVD to denoise. The average results of four objective quality measures show in Fig.3-6. Spending time is tabulated in Table.1.

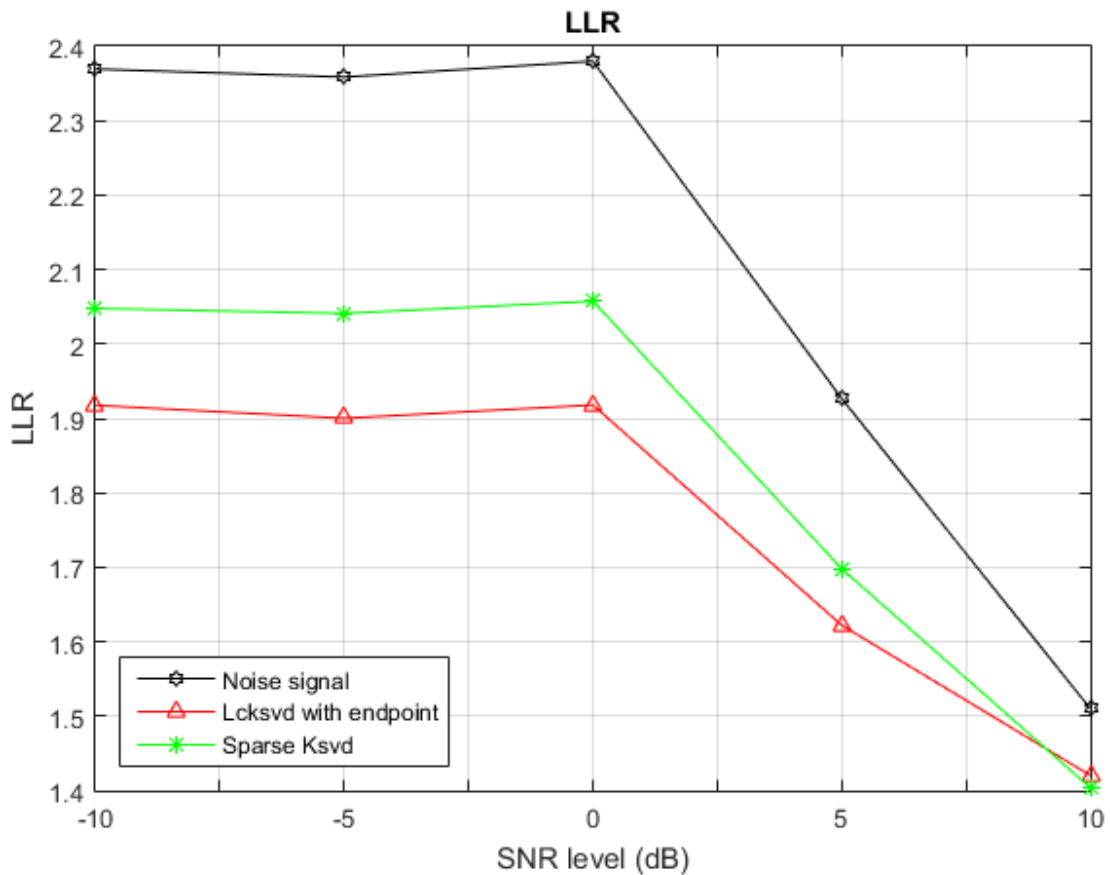


Fig.3 The average results under LLR quality measures.

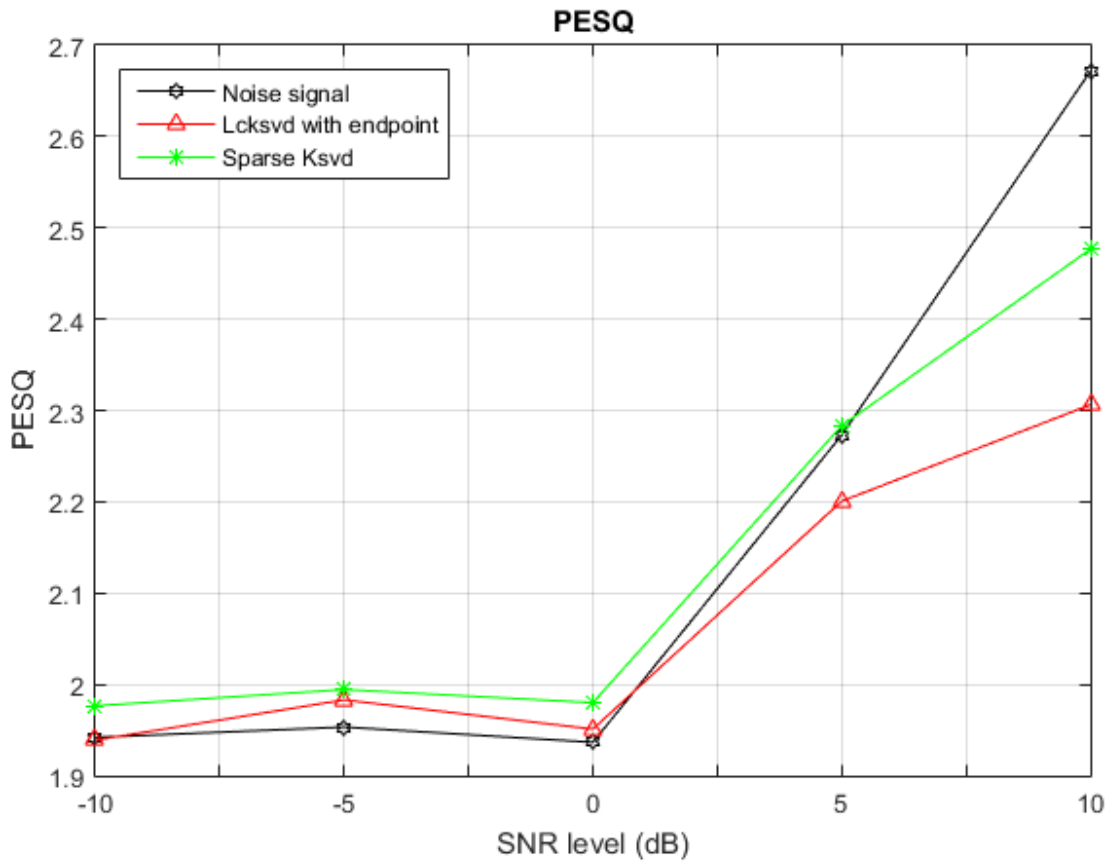


Fig.4 The average results under PESQ quality measures.

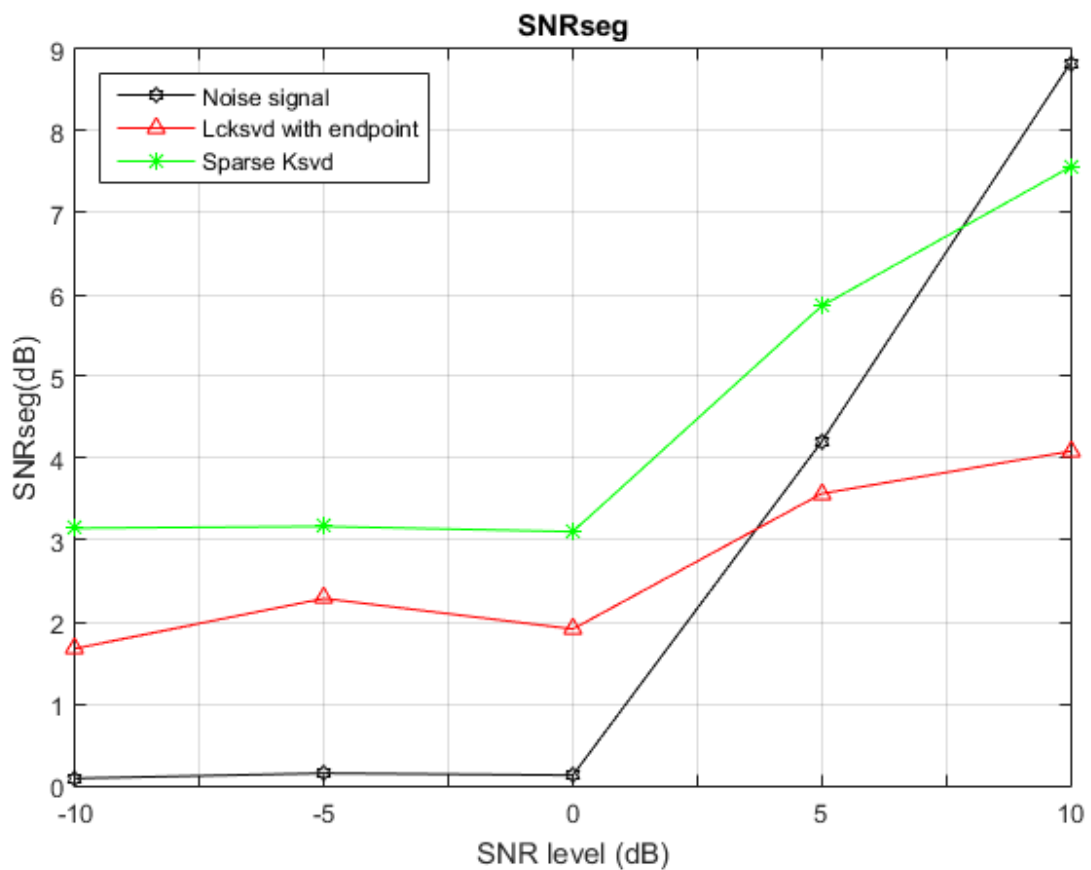


Fig.5 The average results under SNRseg quality measures.

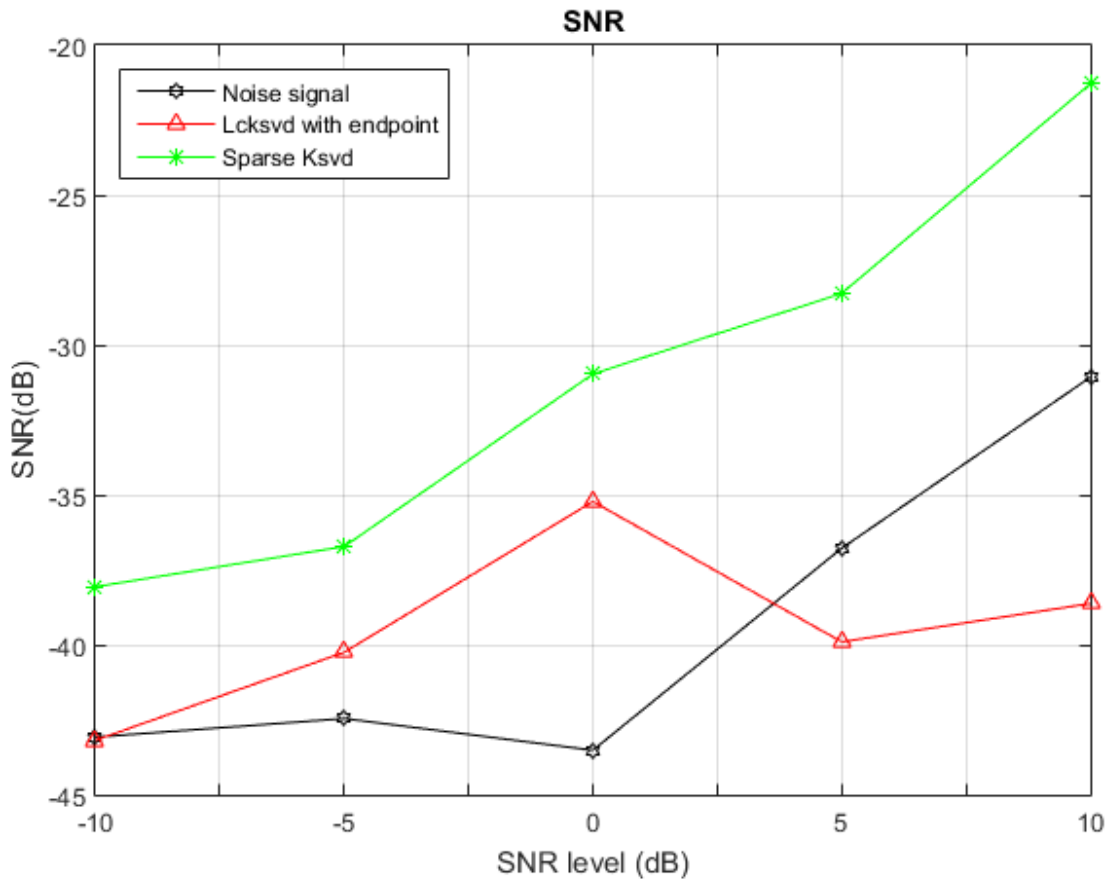


Fig.6 The average results under SNR quality measures.

Table.1 The spending time of LC K-SVD and Sparse K-SVD.

	LC K-SVD	Sparse K-SVD
Time(second per frame)	0.0237	0.0430

According to Fig.3-6 and Table.1, we get the conclusion. In LLR assessment measure, our method is the best. Although our method is lose than Sparse K-SVD a little at SNR levels of -10, -5 and 0, it has a certain effect on denoising. For reducing spending time, we sacrifice some effect.

Conclusion

In this paper, we proposed LC K-SVD to reduce spending time on sparse algorithm for speech signal denoising. Training clean dictionary of each input noise signal based on LC K-SVD algorithm, then reconstruct the clean speech signal. In Simulation results, our proposed method is only lose than Sparse K-SVD a little, but it saves half time even more. In the future work, we will use other classifying way of label to upgrade effect of LC K-SVD. Our proposed method will be applied to color noise environment and the speech recognition.

Acknowledgement

This work was supported by the National Science Council under grant number MOST 103-2632-E-032-001-MY3, and the Tamkang University under grant number FDRX10-2321.

References

[1] Tanabe, N., urukawa, T., Matsue, H. and Tsujii, S. "Kalman Filter for Robust Noise Suppression in White and Colored Noises," IEEE International Symposium on Circuits and Systems. ISCAS. pp.

1172-1175, 2008.

[2] Marwa, A. Abd, El-Fattah., Moawad, I., Dessouky, Alaa, M., Abbas, Salaheldin, M., Diab, El-Sayed, M., El-Rabaie, Waleed, Al-Nuaimy, Saleh, A., Alshebeili, Fathi, E., Abd and El-samie. "Speech enhancement with an adaptive Wiener filter," International Journal of Speech Technology, 17(1), pp. 53-64, 2014.

[3] Aharon, M., Elad, M. and Bruckstein, A. "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," IEEE Transactions on Signal Processing, 54(11), pp. 4311-4322, 2006.

[4] Ching-Tang Hsieh, Yan-heng Chen, Ting-Wen Chen, Li-Ming Chen. "Speech Enhancement Based on Sparse Theory under Noisy Environment," International Conference on Internet Studies, 2015, Tokyo, Japan.

[5] Zhuolin Jiang, Zhe Lin, Larry S. Davis. " Learning a Discriminative Dictionary for Sparse Coding via Label Consistent K-SVD". IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[6] Zhuolin Jiang, Zhe Lin, Larry S. Davis. "Label Consistent K-SVD: Learning A Discriminative Dictionary for Recognition". IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(11): 2651-2664.

[7] Xueying Zhang, Zhefeng Zhao, Gaofeng Zhao. "A Speech Endpoint Detection Method Based on Wavelet Coefficient Variance and Sub-Band Amplitude Variance," Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC'06)0-7695-2616-0/06 2006

[8] Hu, Y. and Loizou, P. "Evaluation of objective quality measures for speech enhancement," IEEE Transactions on Speech and Audio Processing, 16(1), pp. 229-238, 2008.

[9] Ma, J., Hu, Y. and Loizou, P. "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions", Journal of the Acoustical Society of America, 125(5), pp. 3387-3405, 2009.

[10] Barker, J., Vincent, E., Ma, N., Christensen, C. and Green, P. The PASCAL CHiME speech separation and recognition challenge. Computer Speech and Language.