International Journal of Applied Engineering and Technology ISSN: 2277-212X (Online)
An Open Access, Online International Journal Available at http://www.cibtech.org/jet.htm
2014 Vol. 4 (3) July-September, pp.50-53/Sharma and Kaur.
Research Article

SPEECH ENHANCEMENT USING LINEAR PREDICTION AND DCT COEFFICIENTS

***Diksha Sharma and Rupinder Kaur**
DIET Kharar, Punjab, India
**Author for Correspondence*

ABSTRACT
One implicit assumption the speech enhancement algorithms is that the representation of speech in a transform domain or over a redundant dictionary is sparse, while that of noise is dense. Based on this assumption, clean speech can be recovered by finding the sparse representations. However, some kinds of noise are also found sparse in the above representation scenarios, which results in degradation of enhancement performance. For example, since coefficients of car interior noise are sparse in DCT domain, the speech enhancement performance for car interior noise is not as good as that in other noisy background. In addition, some features, for example, speech energy and the inter-frame correlation, are not considered sufficiently in the available speech enhancement algorithms which probably hinders the further improvement of speech quality.
In the proposed method, speech enhancement is casted to an optimization problem, where linear prediction residual and DCT coefficients are combined and adopted as the representation of speech to ensure that noise is dense in such domain. Other features, including speech energy, noise energy and correlation are also considered as constraints to improve the quality and intelligibility of recovered speech.
The proposed algorithm adopts LP residual as one of the sparse representation of speech, considering it is feasible and advantageous, as analyzed in the previous section. To make full use of the sparsity of speech, DCT coefficients are also included to contribute as a measurement. The proposed algorithm aims to recover the clean speech, whose LP residual and DCT coefficients are both sparse, via solving an optimization problem under a series of constraints.
.

Keywords: *Linear Prediction, DCT Coefficients, Speech Enhancement*

INTRODUCTION
In practical systems, such as voice communication and speech recognition, noise is almost inevitable. Speech recognition accuracy likely suffers greatly in the presence of noise. Therefore, it is essential to reduce noise effectively with signal processing techniques. Though some of them have been utilized in commercial schemes, there is still a gap between the human desire and ready-made technologies. In recent years, sparse representation is adopted to improve the quality of noise corrupted speech. However, the representation of noise is also found to be sparse in some special cases, which degrades the performance of sparsity based speech enhancement.
In recent years, sparse representation is adopted to improve the quality of noise corrupted speech. However, the representation of noise is also found to be sparse in some special cases, which degrades the performance of sparsity based speech enhancement. An adaptive speech enhancement algorithm using sparse prior information is proposed in this paper. In the proposed method, speech enhancement is casted to an optimization problem, where linear prediction (LP) residual and DCT coefficients are combined and adopted as the representation of speech to ensure that noise is dense in such domain. Other features, including speech energy, noise energy, and inter frame correlation are also considered as constraints to improve the quality and intelligibility of recovered speech.
Noise degrades the performance of these systems dramatically. For example, speech recognition accuracy likely suffers greatly in the presence of noise. Therefore, it is essential to reduce noise effectively with signal processing techniques. Over the past four decades, a number of speech enhancement algorithms

International Journal of Applied Engineering and Technology ISSN: 2277-212X (Online)
An Open Access, Online International Journal Available at http://www.cibtech.org/jet.htm
2014 Vol. 4 (3) July-September, pp.50-53/Sharma and Kaur.
Research Article

have been developed. Though some of them have been utilized in commercial schemes, there is still a gap between the human desire and ready-made technologies.

Related Works

This paper focuses on the enhancement for speech corrupted by additive noise in single channel systems. The conventional speech enhancement algorithms can be roughly divided into three categories [18]: spectral-subtractive algorithms [1–3], statistical-model-based algorithms [4–6], and subspace algorithms [7–9]. Most of them focus on the spectrum estimation of noise or speech. Recently, sparse representation is extensively investigated. Sparsity is an important feature of speech, which has been found approximately sparse in some transform domains, for example, DCT domain and wavelet domain [10], and over redundant dictionaries of clean speech exemplars [11, 12]. Linear prediction (LP) residual of speech has also been found sparse [15–17]. Sparse prior information has been introduced into speech coding and speech enhancement [10–14], which achieve good performance. In [10], the sparsity. The corresponding author of this paper is Yuantao Gu of DCT coefficients is adopted for speech enhancement. The enhancement algorithm proposed in [11] considers the sparsity over the redundant dictionary of clean speech exemplars. In [14], the prior information that the DCT-II coefficients of speech are sparse over the redundant dictionary is used to improve the speech quality.

One implicit assumption in these enhancement algorithms [10, 11, 14] is that the representation of speech in a transform domain or over a redundant dictionary is sparse, while that of noise is dense. Based on this assumption, clean speech can be recovered by finding the sparse representations. However, some kinds of noise are also found sparse in the above representation scenarios, which results in degradation of enhancement performance. For example, since coefficients of car interior noise are sparse in DCT domain, the speech enhancement performance for car interior noise is not as good as that in other noisy background [10]. In addition, some features, for example, speech energy and the interframe correlation, are not

considered sufficiently in the available speech enhancement algorithms [10–14], which probably hinders the further improvement of speech quality.

In this paper, an adaptive speech enhancement algorithm using sparse prior information (ASESI) is proposed. In this algorithm, LP residual is adopted as the representation of speech in order to keep the parameters of speech sparse while that of noise dense. The energy constraint and interframe correlation are also adopted into the proposed algorithm to improve the quality and the intelligibility of recovered speech. The clean speech is recovered by finding the component whose LP residual and DCT coefficients are both sparse under the energy and the interframe correlation constraints, i.e., by solving an optimization problem. This formalized problem can be solved with numerous existing methods. LP coefficients, speech energy, and interframe correlation, which are used to recover speech, are the distinctive features for each frame, which reveals the adaptive behaviour of the proposed algorithm. Experimental results confirm that a wide range of noise can be reduced effectively through the proposed approach.

1. Algorithm

For sparsity based speech enhancement algorithms, the enhancement performance closely depends on the sparsity distinction between the representation of speech and that of noise.

The proposed work uses the following LP model:

$$x(n) = \sum_{k=1}^{K} a_k . x(n-k) + r($$

where x(n), r(n), ak, and K denote the speech signal, the LP residual, the LP filter coefficients and order, respectively. The LP coefficients are estimated by minimizing the least squares of LP residual. A linear predictor uses observations of a signal to try to predict the next sample of the signal beyond those it can observe.

At Initial State, x(n) = 0, n 0

Therefore, LP residual vector ‘r’ is given by:

r = [r(1), r(2), (3), …. R(n)] and can be expressed using the speech vector as
x = [x(1), x(2), x(3), …. X(n)]
The LP coefficient matrix is given by:
r = A . x
For sparsity based speech enhancement algorithms, the enhancement performance closely depends on the sparsity distinction between the representation of speech and that of noise.

The proposed algorithm adopts LP residual as one of the sparse representation of speech, considering it is feasible and advantageous, as analyzed in the previous section. To make full use of the sparsity of speech, DCT coefficients are also included to contribute as a measurement. The proposed algorithm aims to recover the clean speech, whose LP residual and DCT coefficients are both sparse, via solving an optimization problem under a series of constraints.

Let y = [y(1), · · · , y(N)] be the noisy speech,
y = x + e
where 'e' denotes the additive noise and 'x' is the speech signal to be extracted.

In order to improve the accuracy, the energy of the recovered speech is constrained to be close to that of the clean speech as given below:

$$\alpha_1 \tilde{E}_{\mathbf{x}} \leq \|\mathbf{z}\|_2^2 \leq \alpha_2 \tilde{E}_{\mathbf{x}},$$

Where α1 and α2 describe the degree of approximation. Ex is the estimate of energy of speech signal as the original speech is not available.

Clean speech may be recovered, if its LP residual and DCT coefficients are both sparse under the constraints of moderate energy and correlation with previous frame, which justifies the main objective of the proposed algorithm.

Linear Predictor

Current speech sample can be closely approximated as a linear combination of past samples

$$x[n] = \sum_{k=1}^{P} a_k x[n-k] + e[n]$$

x[n-k]:	previous speech samples
p:	order of the model
a_k:	prediction coefficient
e[n]:	prediction error

RESULTS AND CONCLUSION

The proposed work finds application in enhancing the speech signal using the sparseness of the dct and LP residual matrix. The proposed work has advantage of not loosing the vital information form the speech signal on account of dct based sparse representation. The dct based decomposition of the signal into different freq. components provides the complete spectrum of the speech signal and can be selective portion of the speech could be enhanced.

REFERENCES

[1] S. Boll, “Suppression of acoustic noise in speech using spectral subtraction”, *IEEE Transaction on Acoustic, Speech and Signal Processing*, vol. 27, no. 2, pp. 113-120, Apr. 1979.

[2] S. Kamath and P. Loizou, “A multi-band spectral subtraction method for enhancing speech corrupted by colored noise”, in*Proc. ICASSP2002*, vol. 4, pp. IV-4164, May 2002.

[3] H. Gustafsson, S. E. Nordholm and I. Claesson, “Spectral subtraction using reduced delay convolution and adaptive averaging”, *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 799-807, Nov. 2001.

[4] J. Lim and A. Oppenheim, “All-pole modeling of degraded speech”, *IEEE Transaction on Acoustic, Speech and Signal Pro- cessing*, vol. 26, no. 3, pp. 197- 210, Jun. 1978.

[5] Y. Ephraim and D. Malah, “Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator”, *IEEE Transaction on Acoustic, Speech and Signal Processing*, vol. 32, no. 6, pp. 1109-1121, Dec. 1984.

[6] R. Martin, “Speech enhancement using MMSE short time spectral estimation with gamma distributed speech priors”, in *Proc.ICASSP2002*, vol. 1, pp. I-253-I-256, May 2002.

[7] Y. Ephraim and H. L. Van Trees, “A signal subspace approachfor speech enhancement”, *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 4, pp. 251-266, July 1995.

[8] H. Yi and P. C. Loizou, “A generalized subspace approach for enhancing speech corrupted by colored noise”, *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 4, pp. 334-341, July 2003.

[9] A. Rezayee and S. Gazor, “An adaptive KLT approach for speech enhancement”, *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 2, pp. 87-95, Feb. 2001.

[10] W. Dalei, Z.Wei-Ping and M. N. S. Swamy, “On sparsity issues in compressive sensing based speech enhancement”, in *Proc.ISCAS2012*, pp. 285-288, May 2012.

[11] J. F. Gemmeke, H. Van Hamme, B. Cranen and L. Boves,“Compressive sensing for missing data imputation in noise robust speech recognition”, *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 272-287, Apr. 2010.

[12] C. D. Sigg, T. Dikk and J. M. Buhmann, “Speech enhancement with sparse coding in learned dictionaries”, in *Proc. ICASSP2010*, pp. 4758-4761, March 2010.

[13] H. Yongjun, H. Jiqing, D. Shiwen, Z. Tieran and Z. Guibin, “A solution to residual noise in speech denoising with sparse representation”, in *Proc. ICASSP2012*, pp. 4653-4656, March 2012.

[14] H. Feng, L. Tan and W. B. Kleijn, “Transform-domain wiener filter for speech periodicity enhancement”,in *Proc. ICASSP2012*, pp. 4577-4580, March 2012.

[15] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen and M. Moonen, “Retrieving sparse patterns using a compressed sensing framework: applications to speech coding based on sparse linear prediction”, *IEEE Signal Processing Letters*, vol. 17, no. 1, pp. 103-106, Jan. 2010.

[16] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen and M. Moonen, “Sparse linear prediction and its applications to speech processing”, *IEEE Transaction on Acoustic, Speech and Signal Processing*, vol. 20, no. 5, pp. 1644-1657, July 2012.

[17] T. V. Sreenivas and W. B. Kleijn, “Compressive sensing for sparsely excited speech signals”, in*Proc. ICASSP2009*, pp. 4125-4128, Apr. 2009.

[18] C. L. Philipos, “Speech enhancement: theory and practice”, Taylor and Francis, 2007