The priori SNR estimation based on the post frame information and self-adaptive averaging factor

Xiao-Qun Yi, Ru-Wei Li and Shuang Zhang
School of Information and Communications Engineering, Beijing University of Technology, Beijing,100124, China
E-mail: yixiaoqun@emails.bjut.edu.cn, liruwei@bjut.edu.cn, zhangshuang@emails.bjut.edu.cn

In view of the estimation problem of the priori signal-to-noise ratio (SNR) estimation in single-channel speech enhancement algorithm, a novel method based on the further information is proposed. In this paper, the information from the post frame is considered in obtaining priori SNR of the current frame. Firstly, the original priori SNR is achieved by the decision directed (DD) algorithm. secondly, in order to solve the smoothness drawback of Two-Step Noise Reduction (TSNR) method, we utilize the correlation of the inter-frame and introduce speech information from the post frame to refine the estimation priori SNR. Lastly, to track noisy speech change more quickly, a self-adaptive averaging factor is introduced by the minimum mean-squared error (MMSE) estimator. The proposed algorithm has good performance in reducing the speech distortion while the advantages in noise suppression are kept. The simulation experiment results show that the performance of proposed algorithm is superior than TSNR algorithm under various noise conditions.

Keywords: A Priori SNR; The Post Frame Speech Information; Speech Enhancement.

1. Guidelines

In real life, the speech signal is always disturbed by a large amount of noise. These noises degrade the quality of service provided by the speech processing system. Speech enhancement based on noise reduction technology has become a key factor of digital signal processing. At present, a widely used single-channel speech enhancement method to determine the a priori SNR from noisy speech is the DD[1] approach. The best known approaches are spectral subtraction, minimum mean square error and Wiener filtering approaches[2-4]. The DD approach can eliminate the musical noise efficiently. However, a priori SNR estimation of DD method tracks the instantaneous SNR with one frame delay[5]. To solve this problem, the TSNR[6] method has been proposed by Plapous refine the a priori SNR estimation based on result of DD. But the smoothness of a priori SNR estimation is ignored. the TSNR method often leads to the transient

energy distortion and speech distortion while reducing the continuity of the speech signal [7].

We propose a method to solve this drawback of TSNR method. Firstly, the original priori SNR is achieved by the DD algorithm. Secondly, we refine a priori SNR estimation by integrating the post frame speech information and utilizing the correlation of the inter-frame. Lastly, a self-adaptive averaging factor is introduced by the MMSE estimator in order to track noisy speech change more quickly. Simulation results demonstrate that the proposed algorithm reduces the transient energy distortion and the speech distortion while improving the quality of the enhanced speech.

2. The Two-Step Noise Reduction (TSNR) Technique

Let y(t), x(t) and n(t) represent the noisy speech signal, the clean speech signal and noise signal respectively.

$$y(t) = x(t) + n(t). \tag{1}$$

The TSNR method refines the a priori SNR by (2).

$$SNR_{prio}^{TSNR}(p,k) = \frac{|G_{DD}(p,k)Y(p,k)|^2}{\sigma_d'^2(p,k)}. \tag{2}$$

The TSNR algorithm is actually the result of the combination of two successive DD algorithms and it suppresses the frame delay bias and limits the level of residual musical noise, however, the smoothness of the priori SNR is not considered in (2), which can result in the transient energy distortion and the distortion of speech [7] .

3. The Proposed Method

To solve this drawback of TSNR approach, the information from the post frame is considered in obtaining priori SNR of the current frame, as in Eq. (3).

$$\tilde{\varepsilon}(p,k) = SNR_{prio}^{DD}(p+1,k) = \beta(p,k)\frac{|G_{DD}(p,k)Y(p,k)|^2}{\sigma_d'^2(p,k)} + (1-\beta(p,k))P[SNR_{post}(p+1,k)-1]. \tag{3}$$

In order to make the estimated priori SNR respond to the transients of speech quickly, a self-adaptive averaging factor[1] is introduced by the MMSE estimator, as in Eq. (4).

$$J_\varepsilon = E[(\tilde{\varepsilon}(p,k)-\varepsilon(p,k))^2 \mid \tilde{\varepsilon}(p-1,k)]. \tag{4}$$

Equating $\partial J_{\varepsilon} / \partial \varepsilon$ to zero, we obtain an expression of optimum $\beta(p,k)$,as in Eq. (5).

$$\beta(p,k) = \frac{1}{1+(\frac{\varepsilon(p,k)-\tilde{\varepsilon}(p-1,k)}{1+\varepsilon(p,k)})^2}. \tag{5}$$

An approximate value of $\varepsilon(p,k)$ can be obtained by (6).

$$\varepsilon(p,k) \cong P[SNR_{post}(p+1,k)-1]. \tag{6}$$

An expression for optimum $\beta(p,k)$ is obtained by (7) in the proposed approach.

$$\bar{\beta}(p,k) = \frac{1}{1+(\frac{\varepsilon(p,k)-\tilde{\varepsilon}(p-1,k)}{1+\bar{\lambda}})^2}. \tag{7}$$

In the above relationship, when the change of a priori SNR is very small, $\bar{\beta}(p,k)$ will be close to 1 to restrain the music noise efficiently. Meanwhile, for any abrupt change, $\bar{\beta}(p,k)$ attains a lower value enabling $\ddot{\varepsilon}(p,k)$ to respond to the speech transients suitably[7].

Containing the only white noise, the noisy signal with the SNR of 10dB has been processed by DD, TSNR and proposed algorithms. In Fig. 1 where the time varying SNRs at frequency 469Hz are displayed. The solid line represents the time varying instantaneous $SNR_{inst}(p,k)$ defined as (8), which is used to denote the transient characteristics of the speech signal.

$$SNR_{inst}(p,k) = \frac{|Y(p,k)|^2 - \sigma^2(p,k)}{\sigma^2(p,k)} = SNR_{post}(p,k) - 1. \tag{8}$$

In Fig. 1, compared with the DD algorithm, the values and fluctuations of the a priori SNR estimation by TSNR and proposed algorithm are both relatively small in non-speech segments(i.e., the about first 4 frames).These characteristics are very conducive to eliminate music noise. During the speech segments, the DD algorithm can track the shape of the instantaneous SNR with one frame delay evidently. The delay effect can be limited by TSNR algorithm efficiently. However, during the speech onset , it shows the obvious deficiency in tracking the fast change of the instantaneous SNR. The proposed algorithm can not only solve the delay problem, but also can greatly improve the tracking ability in the transitional areas. This will enable the output speech to reserve speech and reduce the distortion of speech.

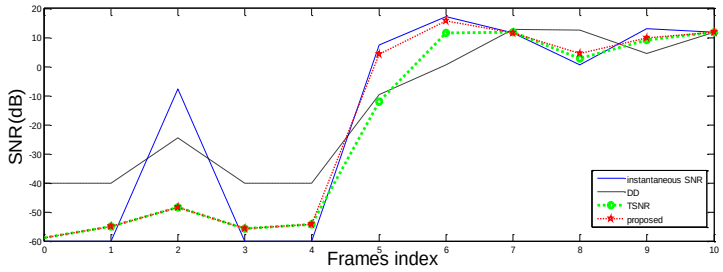

Fig. 1a priori SNR estimations derived from three algorithms

4. Experimental Result and Discussion

In the following simulation experiments, the performances of the DD, TSNR and proposed approach are evaluated. The noise is White noise and SNR is 10dB. After the processing of the DD algorithm, there is still a small amount of residual music noise. Although the level of musical noise is reduced by TSNR algorithm, the transient energy distortion is introduced. In Fig. 2(e), the background noise is effectively suppressed by the proposed algorithm. what's more, the transient energy distortion introduced by the TSNR algorithm can be reduced while improving the continuity of the speech.

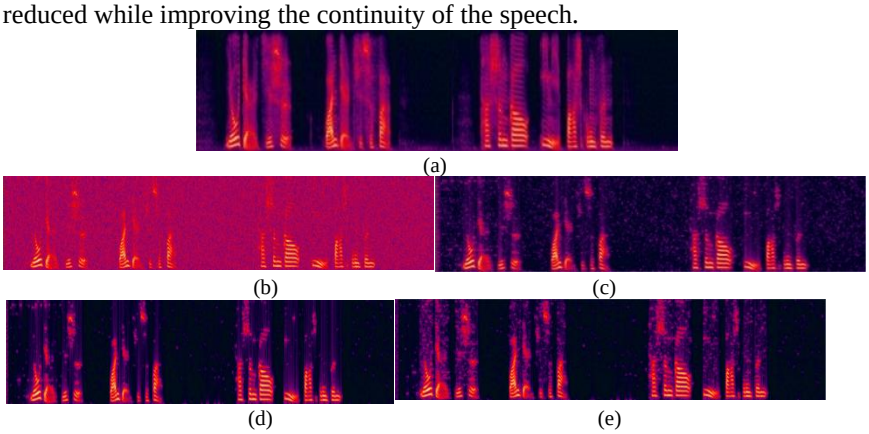

Fig. 2 Speech spectrograms. (a) Clean speech. (b) Noisy speech corrupted by white noise at 10dB. (c) enhanced speech by DD approach. (d) enhanced speech by TSNR approach. (e) enhanced speech by proposed approach

The mean opinion score (MOS) was used as subjective quality test to evaluate the performance of TSNR and proposed approach. Table 1 shows the proposed approach provides the higher score than the TSNR approach, which illustrates that speech distortion degree is reduced.

Tab. 1 MOS values for the two methods

Noise type	Noisy speech	TSNR	Proposed
MOS	1.72	3.11	3.23

The perceptual evaluation of speech quality (PESQ) and log spectral distance (LSD) are chosen as objective quality test standards. The greater value of PESQ denotes the speech intelligibility is higher while the lower value of LSD indicates the degree of clean speech damage by algorithms is lesser. It can be seen from table 2, compared with the TSNR method, the test performance of proposed algorithm improved by the proposed method.

Tab. 2 Comparison on PESQ and LSD of the enhanced speeches

Noise type	Input SNR	PESQ		LSD	
		TSNR approach	Proposed approach	TSNR approach	Proposed approach
White	0dB	2.254	2.291	9.247	8.912
	5dB	2.589	2.623	7.868	7.614
	10dB	2.813	3.852	6.657	6.487
Babble	0dB	2.012	2.054	6.596	6.286
	5dB	2.422	2.464	8.404	8.251
	10dB	2.755	2.817	10.023	9.878
Street	0dB	2.809	2.835	6.834	6.473
	5dB	3.068	3.105	5.610	5.471
	10dB	3.278	3.312	4.827	4.682
Office	0dB	2.267	2.296	8.505	8.216
	5dB	2.606	2.624	6.967	6.681
	10dB	2.866	2.887	5.807	5.621

5. Conclusion

In this paper, we provide a novel method based on the future information to estimate the a priori SNR. We refine the a priori SNR by integrating the future information and introducing the self-adaptive averaging factor. The problem of the delay and smoothness can be solved efficiently in our approach. Compared with the TSNR algorithm, the simulation experiment results show that the performance of proposed algorithm is superior under various noise conditions and SNRs.

Acknowledgments

This work was supported by the Scientific Research Program of Beijing Municipal Commission of education 2015 (No. KM201510005007).

References

1. Y. Ephraim and D. Malah, Speech enhancement using a minimum mean square error short-time spectral amplitude estimator, IEEE Trans on Audio, Speech, and Language Processing. vol. ASSP-32, pp. 1109–1121, Dec. 1984.
2. T. Inoue and H. Saruwatari, Theoretical analysis of musical noise in generalized spectral subtraction based on higher order statistics, IEEE Trans on Audio, Speech, and Language Processing. vol. 19, no. 6, 2011, pp. 1770-1779.
3. L. A. Dalton and E. R. Dougherty, Bayesian Minimum Mean-Square Error Estimation for Classification Error—Part I: Definition and the Bayesian MMSE Error Estimator for Discrete Classification, IEEE Trans on Signal Processing. Year: 2011, Volume: 59, Issue: 1. Pages: 115 – 129.
4. H. Ding and I. Soon, A spectral filtering method based on hybrid wiener filters for speech enhancement, Speech Communication, vol. 51, no. 3, 2009, pp. 259-267.
5. Z. Chen and V. Hohmann, Online Monaural Speech Enhancement Based on Periodicity Analysis and A Priori SNR Estimation. IEEE/ACM Trans on Audio, Speech, and Language Processing. Year: 2015, Volume: 23, Issue: 11. Pages: 1904 – 1916.
6. C. Plapous, Improved Signal-to-Noise Ratio Estimation for Speech Enhancement. IEEE Transactions on Audio Speech and Language Processing. IEEE Trans on Audio, Speech, and Language Processing. Year: 2006, Volume: 14, Issue: 6. Pages: 2098 - 2108.
7. Y. Fang and G. Liu, Speech enhancement based on modified a priori SNR estimation, Frontiers of Electrical and Electronic Engineering in China. vol. 6,no.4,2011,pp.542-546.
8. M. K. Hasan and S. Salahuddin. M. R. Khan, A modified a priori SNR for speech enhancement using spectral subtraction rules, IEEE Signal Processing Letters. IEEE Signal Processing Letters, pp.450-453.