ISCA Archive
http://www.isca-speech.org/archive

International Symposium on Chinese Spoken Language Processing (ISCSLP 2008)
Kunming, China
December 16-19, 2008

NOISE REDUCTION BASED RANDOM MATRIX THEORY

X. Lu [1,2], *S. Matsuda* [1,2], *T. Shimizu* [1,2] *and S. Nakamura* [1,2]

1. National Institute of Information and Communications Technology, Kyoto
2. ATR Spoken Language Communication Research Labs, Kyoto

ABSTRACT

In speech enhancement literature, the signal subspace based method gains a lot of attention because of its simplicity in analytical formulations. The original idea in this method is based on the assumption that clean speech signal occupies a certain low dimensional space, while the noise signal which is a white additive noise spread the whole observation space. In this method, accurate estimation of the noise power (or variance) is required. However, in real applications, the noise power can only be estimated with some degree of uncertainty. This uncertainty will degrade the signal subspace based speech enhancement algorithms, especially in heavy noisy situations since it does not take this uncertainty into consideration. In this study, we took the uncertainty of the estimation of noise power into consideration by using the statistical property of noise based on random matrix theory. The noise statistical property (eigenvalue distribution) was analytically formulated based on the maximum and minimum eigenvalues of the noise random matrix. Based on the statistical property of the eigenvalues of noise, we reduced the part contributed by noise from the covariance matrix of noisy speech. We tested our method for speech enhancement using AURORA-2J speech corpus. Our initial experiments showed that the proposed method performed better than the traditional signal subspace based speech enhancement method.

Index Terms— Speech enhancement, signal subspace, eigenvalue decomposition, random matrix theory.

1. INTRODUCTION

The background noise smears the time-frequency structure of speech. The smeared speech has low quality and intelligibility which degrades the performance of some speech application systems, e.g., speech coding, hearing aids, and automatic speech recognition (ASR), etc. In order to reduce the effect of noise, many speech enhancement methods are developed. For examples, spectral subtraction [1], Wiener filtering [2], minimum mean squared error (MMSE) based estimator [3], signal subspace based speech enhancement [4], etc. The final goal of those methods are to achieve maximum noise reduction with minimum speech distortion. However, as researches show that more noise reduction will result in more speech distortion [5]. In order to make a tradeoff between noise reduction and speech distortion, two criterions are often adopted, one is to minimize speech distortion with a given residual noise level, another is to maximize noise reduction with a given speech distortion threshold. Among the speech enhancement methods, the signal subspace method is widely used as it can use a linear transform to integrate the noise reduction and speech distortion factors in a simple analytical formula, and control the tradeoff between noise reduction and speech distortion [4]. The original principle of the signal subspace based noise reduction is based on the assumption that speech signal occupy a low dimensional space, while additive white noise span the whole observation space. The noisy speech can be decomposed into signal subspace and noise subspace. By nulling the noise subspace, and removing the noise part from the signal subspace, we can get the estimation of clean speech [4]. During the processing, the estimation of noise power (or variance) is needed. However, the noise power can only be estimated with uncertainty because of the fluctuations of the noise amplitude level. Especially in heavy noisy conditions, the uncertainty is high. This estimation uncertainty of noise power is not considered in signal subspace based noise reduction methods. As studied in random matrix theory, the statistical property of the eigenvalues of the noise random matrix can be represented by using the maximum and minimum eigenvalues, i.e., the noise power distribution can be beautifully represented by using the eigenvalue and eigenvector analysis of the random matrix [6, 7]. In this case, the noise uncertainty can be predicted by using the analysis of the noise random matrix. In this paper, we propose a noise reduction method based on random matrix theory by considering the uncertainty of noise power estimation.

The paper is organized as follows. Section 2 describes the basic principle of the eigen filtering for noise reduction based on signal subspace method. We will also discuss the problems in the signal subspace method for speech enhancement in this section. Section 3 introduces the random matrix theory based on which the noise reduction method is derived. Based on Sections 2 and 3, we carry out speech enhancement experiments. Lastly, the discussion and conclusion is given in Section 5.

xugang.lu@atr.jp

978-1-4244-2942-4/08/$25.00 (c) 2008 IEEE

2. EIGEN FILTERING FOR NOISE REDUCTION

For a clean speech signal in additive noise environment, the noisy speech is represented as

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{x}, \mathbf{n}$, and $\mathbf{y}$ are k dimensional vectors of the clean speech signal, noise signal and observed noisy speech signal, respectively. Under the assumption of uncorrelation between speech and noise, the covariance matrix in eq. (1) can be written as

$$\mathbf{C_y} = \mathbf{C_x} + \mathbf{C_n}. \tag{2}$$

Based on eigenvalue decomposition, the noisy speech matrix $\mathbf{C_y}$ is represented as:

$$\mathbf{C_y} = \mathbf{U}\Lambda_\mathbf{y}\mathbf{U}^T, \tag{3}$$

where $\Lambda_\mathbf{y} = diag(\lambda_\mathbf{y}(1), \lambda_\mathbf{y}(2), ..., \lambda_\mathbf{y}(k))$ with eigenvalues $\lambda_\mathbf{y}(i), i = 1, 2, ..., k$, the $\mathbf{U}$ is the eigen matrix, and T is the transpose operator of the matrix. The noise covariance matrix can be written as:

$$\mathbf{C_n} = \mathbf{U}\Lambda_\mathbf{n}\mathbf{U}^T. \tag{4}$$

Suppose the noise variance is $\sigma_\mathbf{n}^2$, then the diagonal matrix of noise is

$$\Lambda_\mathbf{n} = \sigma_\mathbf{n}^2 I_k, \tag{5}$$

where the I_k is the k dimensional identity matrix.

The speech covariance matrix is supposed to be confined in a low dimensional space p ($p < k$) as:

$$\mathbf{C_x} = \mathbf{U}\Lambda_\mathbf{x}\mathbf{U}^T, \tag{6}$$

where $\Lambda_\mathbf{x} = diag(\lambda_\mathbf{x}(1), ..., \lambda_\mathbf{x}(p), 0, 0, ...)$ is a diagonal matrix with only p non-zero eigenvalues. Based on eqs. (3), (4), and (6), the noisy speech covariance matrix can be decomposed as:

$$\mathbf{C_y} = [\mathbf{U}_p\mathbf{U}_{k-p}] \begin{bmatrix} \Lambda_x(1:p) + \sigma_n^2 I_p & 0 \\ 0 & \sigma_n^2 I_{k-p} \end{bmatrix} [\mathbf{U}_p\mathbf{U}_{k-p}]^T. \tag{7}$$

In signal subspace based speech enhancement, based on eq. 7, the matrix for eigen filtering is:

$$\mathbf{F} = \mathbf{U}_p[\Lambda_\mathbf{x}(1:p)]\mathbf{U}_p^T, \tag{8}$$

i.e., nulling the subspace occupied by noise only, and removing the noise contribution in the signal subspace. In eq. (8), the estimation of $\Lambda_\mathbf{x}(1:p)$ is done as:

$$\tilde{\Lambda}_\mathbf{x}(1:p) = \Lambda_\mathbf{y}(1:p) - \sigma_\mathbf{n}^2 I_p. \tag{9}$$

Traditionally, as used in eq. (5) for white noise assumption, the eigenvalues of the noise covariance matrix are supposed to be the same as variance $\sigma_\mathbf{n}^2$ which is estimated from finite noise samples as the averaged noise power energy. Actually, the eigenvalues of the noise covariance matrix should have a certain distribution, i.e., there exists fluctuation or uncertainty of the noise power energy. However, this fluctuation or uncertainty is not considered in signal subspace based noise reduction. Underestimate of this noise power energy will result in much residual noise in the enhanced speech while overestimate of it will bring much speech distortion. Another problem in this subspace based speech enhancement is to estimate the speech subspace dimension p as used in eqs. (7), (8), and (9) which can not be obtained exactly [4], [10]. In next section, we will discuss the noise reduction by considering the distribution of eigenvalues of noise covariance matrix without the estimation of the speech subspace dimension.

3. NOISE POWER ENERGY DISTRIBUTION AND RANDOM MATRIX THEORY

In the estimation of noise variance, if we know the noise power energy distribution, i.e., if the uncertainty of noise variance can be quantified when doing noise reduction based on eigen filtering, we can get a better tradeoff of noise reduction and speech distortion for speech enhancement than using the signal subspace based method. In many real implementations, the covariance matrices are estimated via Hankel or Toeplitz matrix of signals, i.e., $C_\mathbf{y} = H_\mathbf{y}H_\mathbf{y}^T$, $H_\mathbf{y} = H_\mathbf{x} + H_\mathbf{n}$. Suppose the Toeplitz or Hankel matrix is a $m * q$ matrix formed by signals. In this case, the noise covariance matrix $C_\mathbf{n} = H_\mathbf{n}H_\mathbf{n}^T$ is a random matrix since the noise signal is supposed as Gaussian white noise. The statistical properties of random matrices have been investigated and applied in many fields for many years, for examples, in high energy physics, machine learning, etc. [6]. The random matrices show many interesting characteristics. Especially, random matrices show asymptotic analytical representation of the eigenvalue distributions which is preferred in this study. Theoretically, for the noise covariance matrix, in the limit sense $m \to \infty$, $q \to \infty$, and $Q = \frac{q}{m}$, the distribution of eigenvalues of the random matrix can be analytically written as [6]

$$p(\lambda) = \frac{Q}{2\pi}\frac{\sqrt{(\lambda - \lambda_{\min})(\lambda_{\max} - \lambda)}}{\sigma_n^2\lambda}, \tag{10}$$

for eigenvalue λ within the bound $\lambda_{min} \le \lambda \le \lambda_{\max}$. Where the σ_n^2 is the variance of the elements of the random matrix (noise variance), and the minimum and maximum bounds are:

$$\begin{aligned} \lambda_{\min} &= \sigma_n^2\left(1 - \tfrac{1}{\sqrt{Q}}\right)^2 \\ \lambda_{\max} &= \sigma_n^2\left(1 + \tfrac{1}{\sqrt{Q}}\right)^2. \end{aligned} \tag{11}$$

An example of the eigenvalue distribution of the random matrix is shown in Fig. 1.

Fig. 1 shows the theoretical eigenvalue distribution of two random matrices with different Q values. From Fig. 1, one can see that the eigenvalue is distributed in the range between

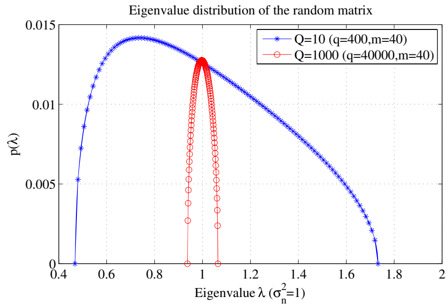

Fig. 1. Eigenvalue distribution of the random matrix (see text for more details).

λ_{min} and λ_{max} around the variance value σ_n^2. The higher the Q is, the more compact range of the distribution of $p(\lambda)$ will be. When $Q \to \infty$, the minimum and maximum bounds go to the same value σ_n^2 as represented in eq. (11). Only under this condition, all the eigenvalues of the noise covariance matrix are the same as σ_n^2, i.e, the distribution of the eigenvalues is the impulse response function at σ_n^2 which is widely used in traditional eigenvalue analysis of noise covariance matrix (such as used in eq. (5)).

In real applications, the m, q, and Q can only be fixed values. We must estimate the noise power by considering the eigenvalue distribution of the noise covariance matrix. Eq. (10) is valid only when $m \to \infty$, $q \to \infty$. However, even for finite values of m and q, the distribution of the eigenvalues of the random matrix fits the eq. (10) well [8]. Since the eigenvalues of the noise random matrix has theoretical trackable distribution, we use this property to separate or reduce the contributions of the noise from the eigenvalues of the observed noisy matrix, i.e., the eigenvalues lying in the interval of $[\lambda_{\min}, \lambda_{\max}]$ are regarded as contributed by noise, while the eigenvalues lying outside of the interval are contributed by speech. In this study, for reducing noise, rather than using eq. (9) to reduce the effect of noise, we remove the eigenspace corresponding to the eigenvalues of the random matrix distributed between the theoretic trackable $\lambda_{\min}$ and $\lambda_{\max}$ from the observation eigenspace.

4. SPEECH ENHANCEMENT EXPERIMENTS

Based on the discussion in Sections (2) and (3), we carry out speech enhancement experiments to test the algorithms. 100 clean speech utterances with sampling rate 8000Hz were chosen from AURORA2J data corpus [9]. The noisy speech is produced by adding gaussian white noise to the clean speech with signal to noise ratio (SNR) levels at -5dB, 0dB, 5dB, 10dB, 15dB, and 20dB, respectively. The covariance matrix of noisy speech was estimated from Toeplitz matrix (40*400 matrix) formed using the noisy speech samples (about 50ms duration with 5ms shift). For signal subspace method, the dimension of speech subspace was estimated using the approach proposed in [10]. For both the signal subspace and random matrix based methods, the noise variance used in eqs. (9) and (11) is estimated from the initial 100ms of speech utterances which is supposed as noise only. One example of the noise reduction effect based on signal subspace and random matrix methods is shown in Fig. (2).

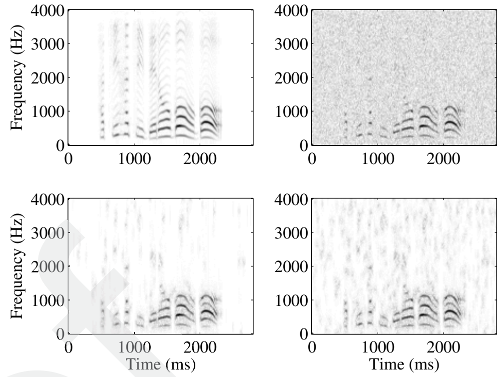

Fig. 2. Speech enhancement based on signal subspace and random matrix theory (see text for more details).

Fig. (2) shows the the power spectrum of a speech utterance of the clean (upper-left panel), the noisy with SNR=0dB (upper-right panel), the enhanced based on random matrix theory (lower-left), and the enhanced based on signal subspace method (lower-right), respectively. From Fig. (2), one can see that the white noise distorted the speech in full time-frequency bands. The signal subspace and random matrix based methods both reduce the noise. However, comparing the random matrix based and signal subspace based noise reduction methods, we find that the enhanced speech based on signal subspace method has more spot-like noise, which is widely known as musical noise [4]. This musical noise brings low quality and intelligibility for speech perception. The reason behind the more spot-like residual noise is that it is caused by the assumption of the fixed noise power (noise variance) without considering the uncertainty of the noise fluctuations. By considering the uncertainty of the noise variance (eigenvalue distribution of the noise random matrix), the noise can be excluded since the statistical distribution of noise and speech is different.

For quantitative comparison of signal subspace and random matrix based methods, subjective and objective evaluations can be adopted. For objective evaluations, many criterions can be used, e.g., segmental SNR, cepstral distance, etc. [4]. For subjective evaluations, the mean opinion score (MOS) can be used based on subjective listening test. In this study, we adopted the perceptual evaluation of speech quality (PESQ) measure. The PESQ is the ITU standard for evaluation of speech which has high correlation with subjective evaluations [10], the higher the PESQ score, the better quality the speech is. We calculated the PESQ for the noisy speech and the enhanced speech utterances using each method, and averaged for the 100 utterances for each SNR condition. The result is shown in Fig. (3). In Fig. (3), the three curves with

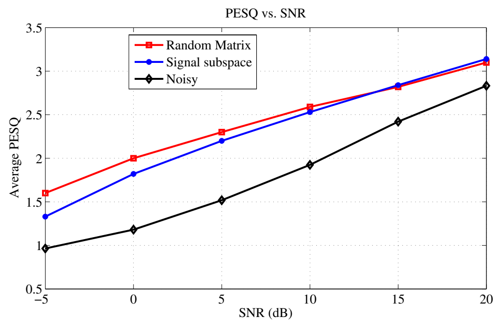

Fig. 3. Averaged PESQ score for the enhanced speech based on signal subspace and random matrix methods (see text for more details).

legends "Random matrix", "Signal subspace", and "Noisy" represent the PESQ curves for the speech processed using random matrix based, traditional signal subspace based methods and the noisy speech without any processing. From Fig. (3), we can see that both the enhanced methods improve the PESQ compared with the original noisy speech. In addition, we can see that in SNR less than 15dB, the random matrix based enhancement shows higher PESQ score compared with that of signal subspace based enhancement. Especially in low SNR condition, since the noise variance uncertainty is higher compared with high SNR condition, the advantage of random matrix based enhancement is more obvious. For high SNR conditions ($>$ 15dB), the random matrix and signal subspace based enhancements perform almost the same.

5. DISCUSSION AND CONCLUSION

In traditional signal subspace based speech enhancement, the noise variance uncertainty is not taken into account. In this paper, we proposed a speech enhancement method based on random matrix theory. This method takes the noise variance uncertainty into consideration based on the eigenvalue distribution of noise random matrix. Our experiments showed that the enhanced speech has a high quality compared with that of signal subspace based method. Although, in some advanced signal subspace algorithms, the noise reduction and speech distortion can be dealt well by giving a residual noise level or a speech distortion threshold. The level or threshold usually is not easily decided. In addition, in the random matrix based enhancement, we do not need to estimate the speech subspace dimension.

In this study, the white noise is assumed. In the original signal subspace method, this assumption is also used. However, in real situation, the noise can be other types of noise, e.g., car noise, factory noise, etc. For non-white noise, a pre-whitening processing is often adopted in signal subspace based method [4]. Similarly, the random matrix based speech enhancement can also adopt the same processing procedure. In addition, in this paper, we did not touch the problem of the nonstationarity of noise. Traditionally, in nonstationary noise condition, only the average noise energy is tracked. The uncertainty is also ignored. Tracking both the average energy and uncertainty is also a problem needing to be solved. The investigation and generalization of the random matrix based speech enhancement method to non-white and nonstationary conditions remains as our future work.

6. REFERENCES

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoustic., speech, signal processing*, ASSP(27), pp.113-120, 1979.

[2] Simon Haykin, *Adaptive filtering theory*, Englewood Cliffs, NJ: Prentice Hall, 2002.

[3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error Log-Spectral Amplitude Estimator," *IEEE Trans. on Acoustics, speech and signal processing*, 33(2), pp.443-445, 1985.

[4] P. C. Loizou, *Speech enhancement: Theory and practice*, CRC Press, 2007.

[5] J. D. Chen, et al, "New insights into the noise reduction Wiener filter," *IEEE Transactions on Audio, Speech & Language Processing*, 14(4), pp. 1218-1234, 2006.

[6] M. L. Mehta, *Random matrices*, Third edition, Elsevier/Academic Press, Amsterdam, 2004.

[7] A.Boutet de Monvel, et al, "On the statistical mechanics approach in the random matrix theory: integrated density of states," *Journal of Statistical Physics*, Vol.79, pp. 585-611, 1995.

[8] Szilard Pafkaa, Imre Kondor,"Noisy covariance matrices and portfolio optimization II," *Physica A*, 319, pp. 487-494, 2003.

[9] http://sp.shinshu-u.ac.jp/CENSREC/, AURORA-2J database.

[10] N. Merhav, "The estimation of model order in exponential families," *IEEE Trans. Inf. Theory*, 35(5), pp. 1109-1113, 1989.

[11] ITU-T Rec. P.862, "Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs," International Telecommunications Union, Geneva, Switzerland, 2001.