Subspace Noise Estimation and Gamma Distribution Based Microphone Array Post-filter Design

CHENG Ning$^{1,2}$, LIU Wenju$^3$ and WANG Lan$^{1,2}$

(1. Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)
(2. The Chinese University of Hong Kong, Shatin, Hong Kong, China)
(3. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)

Abstract — This paper presents a novel method to design the microphone array post-filter. The key issue of post-filter is to accurately estimate the noise power spectrum, thus a subspace based noise estimation method is proposed. Furthermore, the Gamma probability density function is used to describe the signal power spectrum distribution and the signal-plus-noise subspace dimension is determined by maximizing the probability density signal to noise ratio. The noise power spectrum can be computed either with the speech or without the speech, using the eigenvalues corresponding to the noise subspace. With the same Gamma distribution assumption, a post-filter estimation method is proposed. Experiments show that the proposed noise estimation performs better than the conventional VAD based method. The post-filter can obtain a significant gain over the comparing methods in terms of quality measures of the enhanced speech.

Key words — Microphone array post-filter, Gamma distribution, Speech enhancement, Subspace.

I. Introduction

Microphone array systems are often used for high-quality hands-free communication in noisy environments$^{[1–10]}$. The main advantage of microphone arrays against single channel technique is the spatial filtering capability to suppress interfering signals coming from undesired directions. The spatial discrimination of an array is exploited by beamforming algorithm$^{[2]}$. Beamforming steers the array beam toward the arrival direction of the source to recover the desired source and attenuate the other competing sources. The local and global behaviors of the Minimum variance distortionless response (MVDR) beamformer for different noise fields such as, coherent and non-coherent noise fields, are analyzed in Ref.$^{[4]}$. Benesty pointed out that the beamforming alone does not provide sufficient noise reduction, and post-filtering techniques are required$^{[9]}$. The most commonly used criterion for speech enhancement is the Minimum mean-square error (MMSE), as the multichannel Wiener filter. This optimal multichannel MMSE filter has been shown in Ref.$^{[2]}$, which can be factorized into a MVDR beamformer, followed by a single channel Wiener post-filter. Such a post-filter can result in significant improvement in the broad Signal to noise ratio (SNR) over an MVDR beamformer used in isolation. One of the early methods for post-filter estimation was proposed by Zelinski, which was further studied by Marro et al$^{[5]}$. The generalized version of Zelinski algorithm is based on the assumption of a spatially uncorrelated noise field. McCowan and Bourlard replaced this assumption by the more general assumption of a known noise field coherence function and extended the previous method to develop a more efficient post-filtering scheme$^{[6]}$. Lefkimmiatis extended McCowan’s method by using the noise field coherence function to obtain more accurate estimation of the noise spectral density$^{[7]}$. Yousefian analyzed the dissimilarity between the powers of received signals and proposed a hybrid post-filter in which a modified Zelinski post-filter is applied to the high frequencies to suppress spatially uncorrelated noise and a Wiener post-filter is applied to the low frequencies for cancellation of spatially correlated noise$^{[8]}$. Assuming that the clean speech magnitude discrete Fourier transform coefficients are generalized-Gamma distributed, a MMSE estimator which can be decomposed into a concatenation of a linear spatial filter and a generally nonlinear filter was proposed in Ref.$^{[9]}$. Ito incorporated the imaginary parts of the inter-channel observation cross-spectra into the post-filter design and proposed a Wiener post-filter which works well with an arbitrarily arranged array under the assumption that the inter-channel noise cross-spectra are real-valued$^{[10]}$. However, the noise power spectrum at the beamformer’s output in previous methods is over-estimated and the derived filters are suboptimal. If more accurate noise estimation and better post-filter estimation could be achieved, the overall performance of the noise reduction system would be improved. For accurate...
noise estimation, Rahmani proposed an iterative approach to update noise in all speech and non-speech frames based on the coherence noise fields\footnote{111}. A noise power estimation method was proposed based on smoothed spectral minima tracking\footnote{112}. Parsi presented a binaural noise power spectral density estimator in a diffuse noise field by solving a quadratic equation\footnote{113}. Since these noise estimation methods are designed for dual channels, they can not fully exploit the potential of the microphone array signals. A more appropriate noise estimation method is needed for microphone array speech enhancement.

In this paper, we present an effective post-filter with subspace based noise estimation and Gamma distribution to deal with various real noises. First, an accurate noise estimation method is proposed based on the subspace theory\footnote{14}, which can be generalized to the non-stationary noise conditions. Then, based on the assumption that signal power spectrum distributions can be well modeled by Gamma probability density function (pdf), a more appropriate post-filter is developed to deal with various noise conditions. Experiments have shown that both the noise estimation method and post-filter can improve the speech enhancement over conventional methods.

The rest of this paper is organized as follows: Section II gives an overview of our proposed scheme, and then the main contributions are described in detail in Section III–IV, including noise estimation and post-filter design. Experimental results are presented in Section V and finally a conclusion is given in Section VI.

II. System Overview

The proposed system in this paper is the microphone array post-filtering technique\footnote{2}. Fig.1 shows the flowchart of the post-filtering method for speech enhancement. The main objective is to get the high quality enhanced speech by designing a robust post-filter using the subspace based noise estimation and under the assumption that signal power spectrums are Gamma distributed. The system can be divided into two parts, e.g. speech enhancement including beamforming and post-filtering (above the dash line) and post-filter design (below the dash line). The main contributions of this paper focus on the post-filter design part and consist of two steps: subspace based noise estimation and Gamma distribution based post-filter estimation.

![](image)

Fig. 1. Schematic diagram of the proposed system

In the noise estimation step, a subspace based analysis technique is used. An adaptive subspace selection method is proposed rather than the experience based technique in the previous works. Then, using the eigenvalues corresponding to the noise subspace, we estimate the noise power spectrum taking into account both the hypotheses of speech presence and absence. Afterwards, an effective post-filter is developed based on the assumption that the signal power spectrums obey Gamma distribution.

III. Noise Power Spectrum Estimation

According to the post-filter framework\footnote{2}, an accurate noise power spectrum estimate is important for the optimal post-filter design. Commonly used noise estimation method was based on Voice activity detection (VAD)\footnote{15, 16}. In this method, the noise power spectrum was estimated as the weighted average power spectrum of the pure noise frames. However, VAD may lead to large estimation errors since the noise varies on different frames. In order to overcome this shortcoming, this paper proposes a novel noise estimation method based on the subspace theory\footnote{14} which decomposes the noisy speech vector space into two subspaces: a signal-plus-noise subspace and a complementary noise subspace. Since the noise subspace only contains the noise information\footnote{14}, we can accurately estimate the noise power spectrum from the eigenvalues corresponding to the noise subspace in each frame, independent of other frames.

Let $X, S$ and $N$ denote the noisy speech, clean speech and noise, respectively. For an L-sensor linear microphone array, assuming that: (1) the noise power spectrum is the same on all sensors; (2) the noise is uncorrelated between sensors, the noise cross-power spectrum matrix can be expressed as a diagonal matrix in which the diagonal elements are all the noise auto-power spectrum $\phi_{N,k,t}$. $k$ and $t$ denote the frequency bin and the time frame index, respectively.

With the assumption that the speech and noise are uncorrelated, the eigen-decomposition of the cross-power spectral matrix of the noisy speech $\Phi_{X,k,t}$ can be written as\footnote{14}:

$$\Phi_{X,k,t} = U_{k,t}A_{X,k,t}\Lambda_{X,k,t}U_{k,t}^H$$

$$= U_{k,t}(A_{S,k,t} + \phi_{N,k,t}I)U_{k,t}^H$$ (1)

where

$$A_{X,k,t} = \text{diag}(\lambda_{S_1,k,t} + \phi_{N,k,t}, \ldots, \lambda_{S_Q,k,t} + \phi_{N,k,t})$$

$A_{S,k,t}$ and $U_{k,t}$ are the eigenvalue matrix and the eigenvector matrix of the clean speech, $I$ is the identity matrix and $\lambda$ denotes eigenvalue. $Q$ is the signal-plus-noise subspace dimension. From Eq.(1), the eigenvalues corresponding to the noise subspace can be used to estimate the noise power spectrum $\phi_{N,k,t}$.

1. Adaptive subspace selection

To estimate the noise, we should determine the signal-plus-noise subspace dimension $Q$ and the noise subspace dimension $L-Q$.

In the previous subspace methods\footnote{14}, $Q$ is determined as the number of the eigenvalues greater than an empirically
set threshold. Such a method can not accurately select the subspace dimension since the threshold can not be adjusted adaptively according to the signals change. Here, an adaptive subspace selection method is proposed by making the speech components remain in the signal-plus-noise subspace as much as possible in probability. We can use SNR to assess the speech proportion in noisy signal. Furthermore, a SNR similar criterion—the probability density SNR (PD-SNR) is defined:

$$PD_SNR = \frac{f(\phi_S)}{f(\phi_N)}$$

(2)

where $f(\bullet)$ is thepdf.

It has been pointed out that the distribution of speech amplitude $S_{k,t}$ is well represented by Rayleigh pdf. So, the signal power spectrum $\phi_{S,k,t} = S_{k,t}^2$ obeys Gamma pdf which is:

$$f_{\text{Gamma}}(\phi_{S,k,t}, \theta_{S,k,t}) = \frac{1}{\theta_{S,k,t}} e^{-\phi_{S,k,t}/\theta_{S,k,t}}$$

(3)

A proof of the relationship between Rayleigh and Gamma pdfs is described in Appendix I.

In addition, the amplitudes of noisy speech and noise can also be modeled by Rayleigh distribution, so, their power spectrum distributions are also represented by Gamma pdf. In order to calculate the Gamma pdf, we need to obtain $\theta_{S,k,t}$ whose maximum likelihood estimate is $\hat{\theta}_{S,k,t} = \frac{1}{Q} \sum_{i=1}^{Q} \phi_{S_i,k,t}$. A proof of this maximum likelihood estimation is included in Appendix II. We find that $\hat{\theta}_{S,k,t}$ is directly related with the dimension $Q$. It provides us with an opportunity to use this relationship to estimate the appropriate value of $Q$. For each frequency bin in each frame, $Q$ is calculated step by step as shown in Table 1.

Table 1. Subspace selection procedure

| (1) Initialize $Q = 1$ and gradually increase the value of $Q$. |
| (2) Calculate the parameters as: $\hat{\theta}_{N,k,t} = \frac{1}{L - Q} \sum_{i=1}^{L} \lambda_{X_i,k,t}$, $\hat{\theta}_{S,k,t} = \frac{1}{Q} \sum_{i=1}^{Q} (\lambda_{X_i,k,t} - \hat{\theta}_{N,k,t})$, $N_{k,t} = \left( \prod_{i=Q+1}^{L} \hat{\phi}_{S,i,k,t} \right)^{1/L}$, $\hat{S}_{k,t} = \hat{\theta}_{S,k,t}^2$, $\hat{\phi}_{S,k,t} = S_{k,t}^2$. |
| (3) Substituting $\hat{\phi}_{N,k,t}, \hat{\phi}_{S,k,t}, \hat{\theta}_{N,k,t}$ and $\hat{\theta}_{S,k,t}$ to Eq.(3) and using the definition in Eq.(2), we determine $Q$ as: $Q = \frac{\int f_{\text{Gamma}}(\hat{\phi}_{S,k,t}, \hat{\theta}_{S,k,t})/f_{\text{Gamma}}(\hat{\phi}_{N,k,t}, \hat{\theta}_{N,k,t}) \, dQ}{\int f_{\text{Gamma}}(\hat{\phi}_{S,k,t}, \hat{\theta}_{S,k,t})/f_{\text{Gamma}}(\hat{\phi}_{N,k,t}, \hat{\theta}_{N,k,t}) \, dQ}$. |
| The noise subspace dimension is given as $L - Q$. |

2. Noise estimation

The estimation of the noise power spectrum is based on the assumption that each frame of noisy speech signal is either a noise frame or a speech frame, i.e., the following hypotheses: $H_0$: only noise is present and $H_1$: both speech and noise are present.

Under the Rayleigh distribution assumption, the conditional distributions of noisy speech amplitude are given by:

$$p(X_{k,t}|H_0) = \frac{X_{k,t}}{v_{N,k,t}^2} \exp \left( \frac{-X_{k,t}^2}{2v_{N,k,t}^2} \right)$$

(4)

$$p(X_{k,t}|H_1) = \frac{X_{k,t}}{v_{N,k,t}^2} \exp \left( \frac{-X_{k,t}^2}{2v_{N,k,t}^2} \right)$$

(5)

According to the analysis in Appendix I, we derive the parameters estimation in Eq.(4) and Eq.(5) as:

$$v_{N,k,t}^2 = \frac{1}{L} \sum_{i=1}^{L} X_{i,k,t}^2,$$

$$v_{N,k,t}^2 = \frac{1}{2} \hat{\theta}_{X,k,t}$$

Using Bayes rule, the a posteriori probability for speech presence is given by:

$$p(H_1|X_{k,t}) = \frac{1}{1 + A}$$

(6)

where $A$ is defined in Ref.\cite{17}. Assuming that the speech and noise states are equally likely, we have:

$$A = \hat{\theta}_{X,k,t} \exp \left( \frac{X_{k,t}^2}{\hat{\theta}_{N,k,t} - \hat{\theta}_{X,k,t}} \right)$$

With $H_0$ and $H_1$, the noise power spectrum are respectively calculated as:

$$\hat{\phi}_{N,k,t} = \frac{1}{L} \sum_{i=1}^{L} \lambda_{X_i,k,t}$$

and

$$\hat{\phi}_{N,k,t} = \frac{1}{L - Q} \sum_{i=Q+1}^{L} \lambda_{X_i,k,t}$$

Using the conditional probability formula, the noise power spectrum estimate is given by:

$$\hat{\phi}_{N,k,t} = (1 - P(H_1|X_{k,t})) \cdot \hat{\phi}_{N,k,t} + P(H_1|X_{k,t}) \cdot \hat{\phi}_{N,k,t}$$

(7)

IV. Post-filter Estimation

An effective post-filter estimation is given by:

$$h = \frac{\hat{\phi}_S}{\hat{\phi}_S + \hat{\phi}_N} = \frac{E[S_{k,t}^2]}{E[S_{k,t}^2] + E[N_{k,t}^2]} = \frac{\xi_{k,t}}{\xi_{k,t} + 1}$$

(8)

where $\xi_{k,t}$ is a priori SNR dependent on a posteriori SNR $\gamma_{k,t}$\cite{16}.

Since $X_{k,t}^2$ is modeled by Gamma distribution, with hypothesis $H_1$, $E[X_{k,t}^2]$ can be estimated as:

$$E[X_{k,t}^2] = \int_{0}^{\infty} X_{k,t}^2 \left( \frac{1}{\theta_{X,k,t}} \right) e^{-X_{k,t}^2/\theta_{X,k,t}} dX_{k,t} = \hat{\theta}_{X,k,t}$$

(9)

Similarly, the estimate of $E[N_{k,t}^2]$ is $\hat{\theta}_{N,k,t}$. So, $\gamma_{k,t}$ can be computed as $\hat{\gamma}_{k,t} = \frac{\hat{\theta}_{X,k,t}}{\hat{\theta}_{N,k,t}}$, in which $\hat{\theta}_{X,k,t} = \frac{1}{L} \sum_{i=1}^{L} \lambda_{X_i,k,t}$. Since under $H_0$, $h$ is zero, and only under $H_1$, $h$ is non-zero, we modify the post-filter expression as in Ref.\cite{17}:

$$h_{opt} = P(H_1|X_{k,t}) \cdot h$$

(10)

V. Experimental Results

In this section, the proposed noise estimation method and post-filter will be evaluated on various noisy data. For experiments, we recorded four real noises including: electronic noise,
people noise (cheer noise), machine noise (buzz noise) and car noise (engine noise). The sampling rate is 16kHz. These noises were recorded by a linear microphone array with eight sensors. The space between sensors is 4cm. The noise source is in the front of the microphone array with the distance of 1m. Twenty clean speeches are selected from the CMU database\cite{18}. We added each noise to speech at five different SNRs of −10dB, −5dB, 0dB, 5dB and 10dB. A total of 100 noisy speech utterances are obtained. The time aligned noisy speech is divided in time into frames of 25ms, 15ms overlapped between adjacent frames. At each frame a Hamming window is applied and a STFT analysis takes place.

1. Noise estimation results

To show the performance of the noise estimation method, we define the Noise estimation error (NEE) as follows:

$$NEE = \frac{|\hat{\phi}_{N,k,t} - \phi_{true}^{\text{N},k,t}|}{\phi_{true}^{\text{N},k,t}}$$ (11)

where $\hat{\phi}_{N,k,t}$ and $\phi_{true}^{\text{N},k,t}$ are the estimated and true noise power spectrum, respectively.

We compare our method with the VAD based noise estimation method\cite{15}. Fig.2 gives the average experimental results.

![Fig. 2. The average noise estimation errors for various noises at different input SNRs. Each type of noisy data contains five utterances. (a) Electronic noise; (b) People noise; (c) Machine noise; (d) Car noise](image2)

As can be seen from Fig.2, on the NEE criterion, the proposed method is better than the competing algorithms for all the testing data. Compared to the best of the reference approaches, the average improvements of the proposed method cross the five input SNRs are 0.80dB for electronic noise, 1.71dB for people noise, 1.61dB for machine noise and 2.02dB for car noise. Furthermore, the average improvement of our method over the best of the reference approaches on the four noises is 1.54dB.

2. Speech enhancement results

To evaluate the effectiveness of the proposed speech enhancement algorithm, we compare its performance to other multichannel speech enhancement techniques, including Beamformer\cite{2}, Zelinski Post-filter\cite{5} and McCowan Post-filter\cite{6}. The objective evaluation criteria we adopted are SNR, Log-spectral distance (LSD) and Perceptual evaluation of speech quality (PESQ). High values of the SNR and PESQ, and low value of the LSD denote high speech quality.

The SNR experimental results are shown in Fig.3.

![Fig. 3. Average SNRs for various noises at different input SNRs. Each type of noisy data contains five utterances. (a) Electronic noise; (b) People noise; (c) Machine noise; (d) Car noise](image3)

As can be seen from Fig.3, on the SNR criterion, the proposed method is better than the competing algorithms for all the testing data. Compared to the best of the reference approaches, the average improvements of the proposed method cross the five input SNRs are 0.80dB for electronic noise, 1.71dB for people noise, 1.61dB for machine noise and 2.02dB for car noise. Furthermore, the average improvement of our method over the best of the reference approaches on the four noises is 1.54dB.

On the LSD criterion, the proposed method also achieves the best performance over all the comparisons. The average improvements of the proposed method across the five input SNRs, compared to the best of the reference approaches, are 2.5% for electronic noise, 3.7% for people noise, 2.5% for machine noise and 11.1% for car noise. Furthermore, the average improvement of our method over the best of the reference approaches on the four noises is 5%.

On PESQ criterion, the proposed method is better than the competing methods for electronic noise, people noise and car noise. Compared to the best of the reference approaches, the average improvements of the proposed method across the five input SNRs, compared to the best of the reference approaches, are 2.5% for electronic noise, 3.7% for people noise, 2.5% for machine noise and 11.1% for car noise. Furthermore, the average improvement of our method over the best of the reference approaches on the four noises is 5%.

On PESQ criterion, the proposed method is better than the competing methods for electronic noise, people noise and car noise. Compared to the best of the reference approaches, the average improvements of the proposed method across the five input SNRs, compared to the best of the reference approaches, are 2.5% for electronic noise, 3.7% for people noise, 2.5% for machine noise and 11.1% for car noise. Furthermore, the average improvement of our method over the best of the reference approaches on the four noises is 5%.

On PESQ criterion, the proposed method is better than the competing methods for electronic noise, people noise and car noise. Compared to the best of the reference approaches, the average improvements of the proposed method across the five input SNRs, compared to the best of the reference approaches, are 2.5% for electronic noise, 3.7% for people noise, 2.5% for machine noise and 11.1% for car noise. Furthermore, the average improvement of our method over the best of the reference approaches on the four noises is 5%.
noises, our method still outperforms the best competing algorithm 3.7%.

These results show clearly that more accurate noise estimation and post-filter design result in the improvements of speech enhancement.

Fig. 4. Average LSD for various noises at different input SNRs. Each type of noisy data contains five utterances. (a) Electronic noise; (b) People noise; (c) Machine noise; (d) Car noise

VI. Conclusion

In this paper, we have presented a multichannel post-filtering approach that is appropriate for a variety of noise conditions. First, a subspace based noise estimation method is proposed under the Gamma distribution assumption. An adaptive subspace selection method is proposed by maximizing the $P_{DSNR}$. Then, based on the same distribution assumption, we propose a post-filter estimation method. Experimental results have demonstrated the improvements of the noise estimation method and post-filter in terms of various evaluation criteria.

Appendix I

The distribution of the clean speech amplitude is assumed as:

$$f_{Rayleigh}(S_{k,t}) = \frac{S_{k,t}}{\nu_{S,k,t}^2} e^{-\frac{S_{k,t}^2}{\nu_{S,k,t}^2}}$$ (12)

To derive the distribution of the speech power spectrum, let $\phi_{S,k,t} = S_{k,t}^2$. The probability function of $\phi_{S,k,t}$ is as:

$$F(x) = P(\phi_{S,k,t} \leq x) = P(S_{k,t} \leq \sqrt{x}) = \int_0^{\sqrt{x}} \frac{S_{k,t}}{\nu_{S,k,t}^2} e^{-\frac{S_{k,t}^2}{\nu_{S,k,t}^2}} dS_{k,t} = 1 - e^{-\frac{x}{2\nu_{S,k,t}^2}}$$ (13)

To differentiate the probability function Eq.(13) with respect to $x$, the pdf of $\phi_{S,k,t}$ can be expressed as:

$$f_{Gamma}(\phi_{S,k,t}, \theta_{S,k,t}) = \frac{1}{\theta_{S,k,t}^2} e^{-\phi_{S,k,t}/\theta_{S,k,t}}$$ (14)

where, $\theta_{S,k,t} = 2\nu_{S,k,t}^2$.

Appendix II

The likelihood function of $Q$ independent and identically-distributed random variables observations $(\phi_{S1,k,t}, \ldots, \phi_{SQ,k,t})$ is given as:

$$L(\theta_{S,k,t}) = \prod_{i=1}^{Q} f(\phi_{Si,k,t}) = \frac{1}{\theta_{S,k,t}^{Q\nu_{S,k,t}^2}} e^{-\sum_{i=1}^{Q} \phi_{Si,k,t}/\theta_{S,k,t}}$$ (15)

from which we calculate the log-likelihood function:

$$l(\theta_{S,k,t}) = -Q \ln \theta_{S,k,t} - \frac{1}{\theta_{S,k,t}} \left( \sum_{i=1}^{Q} \phi_{Si,k,t} \right)$$ (16)

Finding the maximum with respect to $\theta$ by taking the derivative and setting it equal to zero yields the maximum likelihood estimator $\hat{\theta}_{S,k,t} = \frac{1}{Q} \sum_{i=1}^{Q} \phi_{Si,k,t}$.

References


CHENG Ning received the B.S. degree in applied mathematics and M.S. degree in system engineering from University of Science and Technology Beijing and Ph.D. degree in pattern recognition from Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2003, 2006 and 2009, respectively. He is currently working in Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include speech enhancement and speech recognition. (E-mail: kinchengning@hotmail.com)

LIU Wenju was born in Beijing, in 1960. He received the B.S., M.S. degrees in mathematics from Peking University and Beijing University of Posts and Telecommunications, and Ph.D. degree in computer applications from Tsinghua University, Beijing, China, in 1983, 1989 and 1993, respectively. He is currently a professor and Ph.D. supervisor at the National Key Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include speech recognition, speech synthesis, speaker recognition, voice conversion, computational auditory scene analysis, speech enhancement and noise reduction, etc. (E-mail: lwj@nlpr.ia.ac.cn)

WANG Lan is an associate professor of Ambient Intelligence and Multimodal Systems Laboratory, Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS). She received M.S. degree in the Center of Information Science, Peking University. She received Ph.D. degree from the Machine Intelligence Laboratory of Cambridge University Engineering Department in 2006, and then worked as a research associate in CUED. Her research interests are large vocabulary continuous speech recognition, speech visualization and audio information indexing. (E-mail: lan.wang@siat.ac.cn)