Optimal design of minimum mean-square error noise reduction algorithms using the simulated annealing technique

Mingsian R. Bai,[a] Ping-Ju Hsieh, and Kur-Nan Hur
Department of Mechanical Engineering, National Chiao-Tung University, 1001 Ta-Hsueh Road, Hsin-Chu 300, Taiwan

(Received 8 August 2008; revised 20 November 2008; accepted 21 November 2008)

The performance of the minimum mean-square error noise reduction (MMSE-NR) algorithm in conjunction with time-recursive averaging (TRA) for noise estimation is found to be very sensitive to the choice of two recursion parameters. To address this problem in a more systematic manner, this paper proposes an optimization method to efficiently search the optimal parameters of the MMSE-TRA-NR algorithms. The objective function is based on a regression model, whereas the optimization process is carried out with the simulated annealing algorithm that is well suited for problems with many local optima. Another NR algorithm proposed in the paper employs linear prediction coding as a preprocessor for extracting the correlated portion of human speech. Objective and subjective tests were undertaken to compare the optimized MMSE-TRA-NR algorithm with several conventional NR algorithms. The results of subjective tests were processed by using analysis of variance to justify the statistic significance. A *post hoc* test, Tukey's Honestly Significant Difference, was conducted to further assess the pairwise difference between the NR algorithms.

© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3050292]

PACS number(s): 43.60.Dh, 43.60.Np, 43.60.Uv, 43.72.Kb [EJS] Pages: 934–943

I. INTRODUCTION

In recent years, applications of mobile communication, video conferencing, and peer-to-peer internet telephony networks, such as SKYPE®, hands-free car kits, etc., are rapidly advancing in modern daily life. In these applications, effective communication in noisy environments has been one of the pressing problems. To enhance speech quality, noise reduction (NR) technology has been extensively studied in the communication community. The main problem with most NR algorithms is that sheer NR does not necessarily lead to the general preference of the users. Overly aggressive NR schemes often result in processing artifacts and degradation of speech quality. How to effectively reduce background noise without impairing speech quality has become an imminent issue for NR algorithm design.

NR algorithms fall into three categories: spectral-subtraction algorithms, statistical-model-based algorithms, and subspace algorithms. Spectral-subtraction algorithms[1–6] subtract directly the estimated noise spectrum from the spectrum of the noisy speech. Statistical-model-based algorithms estimate Fourier coefficients using statistically optimal linear or nonlinear estimators of clean signals. The Wiener algorithm[7–10] and the minimum mean-square error (MMSE)[1,11] algorithm belong to this class. Subspace algorithms are based on the principle that the vector space of the noisy signal can be decomposed into the "signal" and "noise" subspaces. Noise is suppressed by projecting the noisy signals onto the signal subspace and nullifying the components in the noise subspace. The decomposition of these two orthogonal subspaces can be done by using the singular value decomposition or the eigenvalue decomposition. The Karhunen–Loéve transform (KLT) algorithm[11,12] falls into this category. All NR algorithms require the information of noise spectra or noise covariance matrices, which must be estimated and updated from frame to frame. Noise estimation can be carried out either during speech pauses, which requires a voice activity detector (VAD), or continuously using time-recursive averaging (TRA) algorithms. A more comprehensive review of speech enhancement and NR methods can be found in the monograph by Loizou.[11]

In this paper, a MMSE-NR algorithm based on TRA[11,13] noise estimation (denoted as MMSE-TRA-NR) is investigated. This algorithm is found to be very sensitive to the choice of two recursion parameters. To address this problem in a more systematic manner, this paper proposes an optimization method to efficiently search the optimal parameters of the MMSE-TRA-NR algorithms. A global optimization technique, simulated annealing (SA)[14–16] algorithm, is exploited for locating the optimal parameters. The objective function is a combined objective measure for NR and the incurred distortion of processed signals. Sensitivity analysis of the TRA parameters obtained using the SA optimization was undertaken for nine types of background noise. In addition to the optimized MMSE-TRA-NR, the possibility of using linear prediction coding (LPC)[6,17–19] as a preprocessor to the NR algorithm is also explored.

In order to evaluate the proposed optimized algorithm and the other NR algorithms, objective and subjective tests were carried out. The objective tests were conducted according to ITU-T P.862.[20] The subjective listening tests were conducted according to ITU-T P.835.[21] The test data were processed by using analysis of variance (ANOVA) to justify the statistic significance of the difference among the NR algorithms. A *post hoc* test, Tukey's HSD, was also employed in

[a] Author to whom correspondence should be addressed. Electronic mail: msbai@mail.nctu.edu.tw

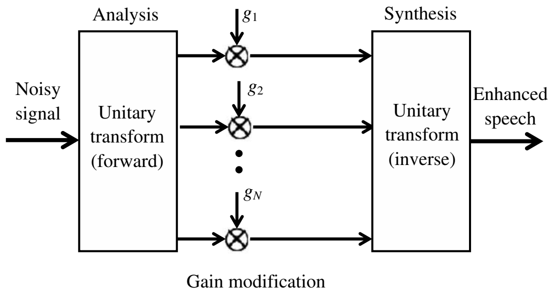

FIG. 1. General structure of NR algorithms (adapted from Ref. 11).

the paired comparison between the NR algorithms.

II. NOISE REDUCTION ALGORITHMS

Figure 1 illustrates the general three-step structure of NR algorithms.[11] The noisy signal is forward transformed using unitary transformations, e.g., Fourier transform, discrete cosine transform, and KLT transform. Next, gain modification, the major NR operation, takes place in the transformed domain. Finally, the time-domain signal of the enhanced speech is recovered by an overlap-and-add procedure. In this section, the MMSE-NR algorithm will be reviewed. The other traditional NR algorithms, such as the spectral subtraction, the Wiener filtering, and the KLT, to be compared in this paper are only mentioned in Sec. I with references.

A. Statistical-model-based noise reduction algorithm

The MMSE-NR algorithm is also based on a statistical model. Instead of the complex spectrum as in the Wiener filter method, a nonlinear estimator of the magnitude spectrum is optimized in the MMSE-NR algorithm. It is assumed that the discrete Fourier transform (DFT) coefficients are statistically independent and follow the Gaussian distribution. The mean-square error between the estimated $(\hat{S}_k)$ and the true (S_k) magnitudes of the clean speech signal is

$$E_{\text{mse}} = E\{(\hat{S}_k - S_k)^2\}. \tag{1}$$

This expectation can be estimated using the following Bayesian mean-square error approach:

$$B_{\text{mse}}(\hat{S}_k) = \int\int (S_k - \hat{S}_k)^2 p(\mathbf{Y},S_k)d\mathbf{Y}dS_k, \tag{2}$$

where $\mathbf{Y}=[Y(\omega_0)Y(\omega_1)\cdots Y(\omega_{N-1})]$ is the noisy speech spectrum and $p(\mathbf{Y},S_k)$ is the joint probability density function (pdf). The posterior pdf of S_k can be determined by using Bayes' rule. Minimization of the Bayesian MSE with respect to $\hat{S}_k$ leads to the optimal MMSE estimator,

$$\hat{S}_k = E[S_k|Y(\omega_k)] = \int_0^\infty s_k p(s_k|Y(\omega_k))ds_k$$

$$= \frac{\int_0^\infty s_k p(Y(\omega_k)|s_k)p(s_k)ds_k}{\int_0^\infty p(Y(\omega_k)|s_k)p(s_k)ds_k}, \tag{3}$$

where s_k is a realization of the random variable S_k and $p(s_k|Y(\omega_k))$ is the conditional posterior pdf of s_k under the observation $Y(\omega_k)$. Assuming that the pdf of the noise Fourier coefficients is Gaussian, it was shown by Ephraim and Malah that the statistically optimal MMSE magnitude estimator takes the form[1]

$$\hat{S}_k = \frac{\sqrt{\pi}}{2}\frac{\sqrt{v_k}}{\gamma_k}\exp\left(-\frac{v_k}{2}\right)\left[(1+v_k)I_0\left(\frac{v_k}{2}\right)+v_kI_1\left(\frac{v_k}{2}\right)\right]Y_k, \tag{4}$$

where $I_0(\cdot)$ and $I_1(\cdot)$ are the modified Bessel functions of the zero and the first order, respectively, Y_k is the spectral magnitude of the noisy signal, and v_k is defined by

$$v_k = \frac{\xi_k}{1+\xi_k}\gamma_k, \tag{5}$$

where γ_k denotes the *a posteriori* signal-to-noise ratio (SNR) given by

$$\gamma_k \triangleq \frac{Y_k^{\,2}}{P_{vv}(\omega_k)} = \frac{Y_k^{\,2}}{E\{|V(\omega_k)^2|\}}. \tag{6}$$

In practice, the noise variance and hence the *a priori* SNR ξ_k are unknown, given the noisy signal $y(n)$. Thus, noise spectrum must be estimated prior to NR processing. First, the noise variance is estimated during speech pauses with the aid of a VAD (Ref. 22) provided the noise is stationary. For example, the following statistical-model-based VAD can be used:

$$\frac{1}{N}\sum_{k=1}^{N-1}\log\left(\frac{1}{1+\xi_k}\exp\left(\frac{\gamma_k\xi_k}{1+\xi_k}\right)\right)\underset{H_0}{\overset{H_1}{\gtrless}}\Delta, \tag{7}$$

where N is the Fast Fourier transform size, H_0 and H_1 denote the hypotheses of speech absence and speech presence, respectively, and the threshold Δ is usually set to 0.15. Here, the MMSE-NR algorithm used in conjunction with VAD for noise estimation is denoted as "MMSE-VAD-NR." Next, the *a priori* SNR ξ_k is estimated with a "decision-directed" approach using the recursive formula

$$\hat{\xi}_k(m) = a\frac{\hat{S}_k^{\,2}(m-1)}{P_{vv}(\omega_k,m-1)} + (1-a)\max(\gamma_k(m)-1,0), \tag{8}$$

where m is the frame number and $0<a<1$ is a weighting factor commonly chosen to be $a=0.98$.

As mentioned above, Eq. (4) is only a spectral magnitude estimator. To recover the enhanced signal, one needs to estimate the phase of the clean speech signal. It was shown by Ephraim and Malah[1] that the optimal phase estimate is simply the noisy phase. Thus, the enhanced complex signal spectrum is calculated by combing the preceding estimated magnitude spectrum $\hat{S}_k$ and the noisy signal phase spectrum $j\theta_y(k)$, i.e., $\hat{S}(\omega_k)=\hat{S}_k\exp(j\theta_y(k))$.

III. ENHANCMENT OF MMSE-NR ALGORITHMS

In this section, three approaches of technical refinement are exploited to enhance the aforementioned MMSE-NR algorithm.

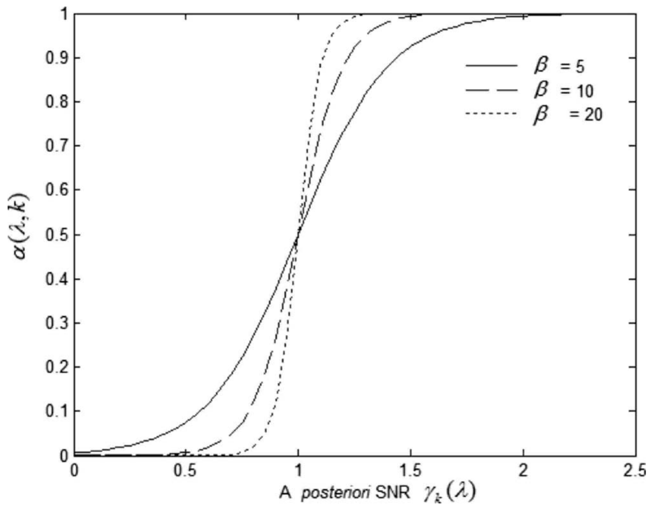

FIG. 2. The smoothing factor $\alpha(\lambda,k)$ calculated according to Eq. (10)for different values of the parameter β with $\delta=1$. (Solid line: $\beta=5$; dashed line: $\beta=10$; dotted line: $\beta=20$).

A. MMSE-time recursive averaging noise reduction

As mentioned earlier in the MMSE-VAD-NR algorithm, the noise variance can be estimated and updated during speech pauses via a VAD provided the noise is stationary. In practice, however, many background noises are often transient and nonstationary. For background noise of this kind, a more practical noise estimation algorithm called the TRA algorithm[13] can be used.

In the TRA algorithm, noise variance $\hat{\sigma}_v^2(\lambda,k)$ at the frame λ and the frequency k is estimated with the following recursive formula:

$$\hat{\sigma}_v^2(\lambda,k)=\alpha(\lambda,k)\hat{\sigma}_v^2(\lambda-1,k)+(1-\alpha(\lambda,k))|Y(\lambda,k)|^2, \tag{9}$$

where $|Y(\lambda,k)|$ is the noisy speech magnitude spectrum and $\alpha(\lambda,k)$ is a time and frequency dependent smoothing factor. The smoothing factor α in the one-pole recursive formula was used to avoid the excessive fluctuations during the process of noise estimation. Various algorithms were proposed to determine the smoothing factor $\alpha(\lambda,k)$ on the basis of the estimated SNR or the probability of speech presence. In this paper, a SNR-based smoothing factor $\alpha(\lambda,k)$ is selected to follow a sigmoid function,

$$\alpha(\lambda,k)=\frac{1}{1+e^{-\beta[\gamma_k(\lambda)-\delta]}}, \tag{10}$$

where β and δ are constants and the *a posteriori* SNR $\gamma_k(\lambda)$ is calculated by averaging the estimated noise variance in the past ten frames,

$$\gamma_k(\lambda)=\frac{|Y(\lambda,k)|^2}{\frac{1}{10}\Sigma_{m=1}^{10}\hat{\sigma}_v^2(\lambda-m,k)}. \tag{11}$$

Figure 2 plots the smoothing factor α for different values of the parameter β and $\delta=1$. Equations (10) and (11) can be interpreted as follows. If the speech is present, the *a posteriori* SNR $\gamma_k(\lambda)$ will be large, and therefore $\alpha(\lambda,k)\approx 1$. In this case, the noise update will cease and the noise estimate will remain the same as that of the previous frame [the first term of Eq. (9)]. Conversely, if the speech is absent, the *a posteriori* SNR $\gamma_k(\lambda)$ will be small, and therefore $\alpha(\lambda,k)\approx 0$. That is, the noise estimate will follow the power spectral density of the noisy spectrum [the second term of Eq. (10)]. In a long stationary noise period, α would stay at a very small value. As a consequence, $\hat{\sigma}_v^2(\lambda,k)\approx|Y(\lambda,k)|^2$. This ensures an accurate and robust estimation of noise level, which gives rise to good reduction performance. Thus, α is strongly dependent on the *a posteriori* SNR $\gamma_k(\lambda)$. The choice of parameters β and δ dictates the slope and the location of the transition of the sigmoid function. This transition can be considered as a "soft switch" between the bistates of speech presence and absence. How to select these two parameters to maximize the NR performance is crucial to the resulting NR performance, as will be explored in the subsequent sections.

When noise is strong and the SNR becomes rather low, the distinction of speech and noise segments could be difficult. Moreover, the noise is estimated intermittently and updated only during the speech silent periods. This may cause problems if the noise is nonstationary, which is the case in many applications. The recursive nature of the TRA algorithm enables estimating noise variance continuously, even during speech activities, which is advantageous in dealing with nonstationary noises. Figure 3 compares NR performance between the VAD and the TRA algorithms. The test signal is a speech signal corrupted by random noise (solid line) varied with three different levels (low-high-medium). The noise (dotted line) estimated using the VAD and the TRA algorithms are also superimposed in the left side of the top and the bottom panels in Fig. 3, respectively. Unlike the VAD algorithm that fails to respond to the noise level variation, the TRA is capable of estimating the noise with drastically transient fluctuation. In other words, VAD and TRA deal with different noise scenarios. VAD is suited for the estimation of stationary noise during speech absence, while TRA is preferred for estimating transient noise, where synchronization of noise estimation is crucial. As a result, a marked difference in NR performance is observed in the enhanced signals using these two noise estimation methods. The right side of the top and the bottom panels in Fig. 3 shows the signals (dotted lines) processed by the MMSE-NR using VAD and TRA, respectively, for noise estimation. The noisy signals (solid lines) are also superimposed to ease comparison. Obviously, the TRA is more superior to the VAD in estimating nonstationary background noise. Thus, the MMSE with TRA noise estimation (denoted as MMSE-TRA-NR) will be employed in the following presentation.

B. Intelligent tuning of the parameters for the MMSE-TRA-NR algorithm

As mentioned previously, the parameters β and δ are used in the sigmoid function of the TRA algorithm for noise estimation. Conventionally, choices such as $\delta=1.5$, $15\leqslant\beta\leqslant 30$ are recommended in the literature.[11] To our surprise, we found that these two parameters β and δ have profound impact on noise estimation and hence on the NR performance

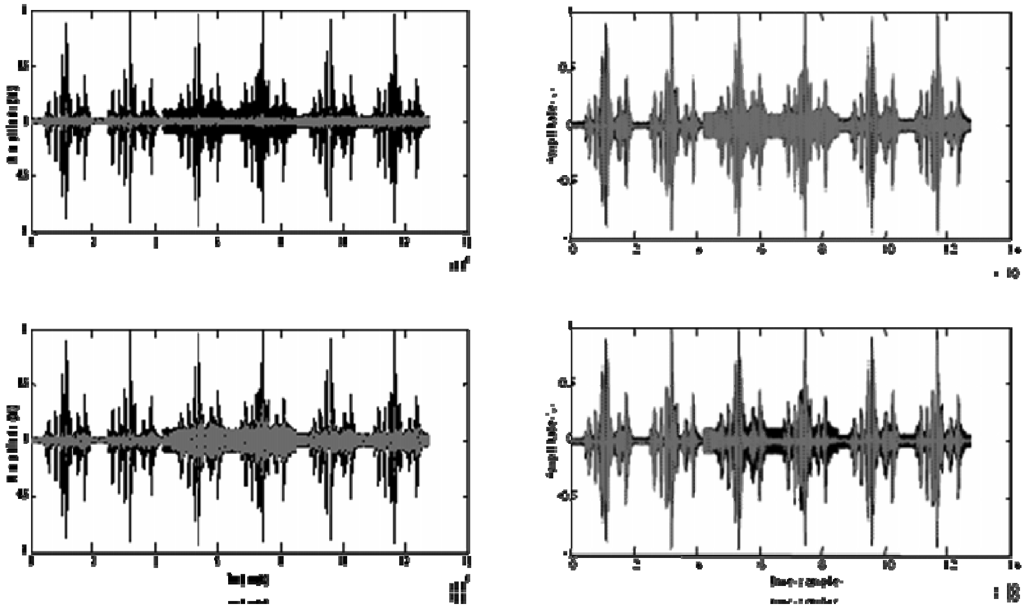

FIG. 3. Comparison of the VAD and TRA algorithms. The noise estimated using the VAD and the TRA algorithms are superimposed in the left side of the top and the bottom panels. The processed speech signals using the MMSE-VAD-NR and MMSE-TRA-NR algorithms are superimposed in the right side of the top and the bottom panels.

of the MMSE-TRA-NR algorithm. Therefore, it is worth exploring how to adjust these two parameters such that NR performance can be maximized without too much speech quality degradation. In the following, a procedure based on the SA optimization method is presented for automated tuning of the TRA parameters.

1. Simulated annealing algorithm

SA is a generic probabilistic meta-algorithm for the global optimization problem, namely, locating a good approximation to the global optimum of a given function in a large search space.[14–16] SA is a technique well suited for solving global optimization problems with many local optima. The flowchart of the SA is illustrated in Fig. 4. In the SA method, each point in the search space is analogous to the thermal state of the annealing process in metallurgy. The objective function Q to be maximized is analogous to the internal energy of the system in that state. The goal of search is to bring the system from an initial state to a randomly generated state with the maximum possible objective function. Two conditions are used to determine whether or not to accept an improved solution. If the objective function is increased, the new state is always accepted. Conversely, if the objective function is decreased and the following condition holds, the new state is accepted:

$$p_{\mathrm{SA}} = \exp(\Delta Q/T) > \varphi, \tag{12}$$

where p_{SA} is the acceptance probability function, ΔQ denotes the increment of the objective function, T is the temperature that follows a certain annealing schedule, and φ is a random number generated subject to the uniform distribution on the interval $[0, 1]$. It follows that the system may possibly move to a new state that is "worse" than the present one. It is this mechanism that prevents the search from being trapped in a local maximum.

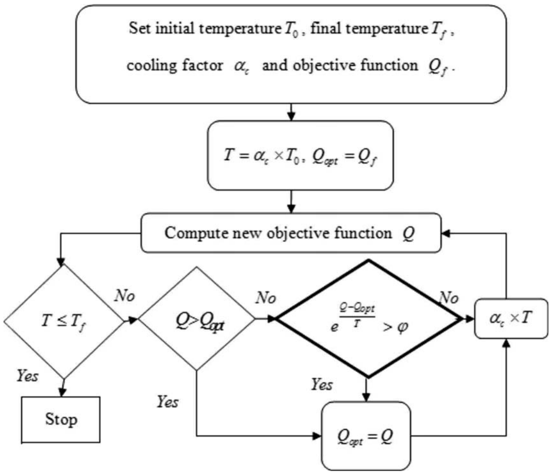
Set initial temperature T_0, final temperature T_f,
cooling factor α_c and objective function Q_f.
$T = \alpha_c \times T_0$, $Q_{opt} = Q_f$
Compute new objective function Q
$T \le T_f$
No
$Q > Q_{opt}$
No
$e^{\frac{Q-Q_{opt}}{T}} > \varphi$
No
$\alpha_c \times T$
Yes
Stop
Yes
Yes
$Q_{opt} = Q$

FIG. 4. The flowchart of the SA optimization algorithm.

Initially, the high temperature T results in the high probability of accepting a move that decreases the objective function, which is analogous to a steel piece whose thermal state is highly active at high temperatures. As the annealing process goes on and T decreases, the probability of accepting a move becomes increasingly small until it finally converges to a stable solution.

A simple annealing schedule is the exponential cooling, which begins at some initial temperature T_0 and decreases temperature in steps according to

$$T_{k+1} = \alpha_c T_k, \tag{13}$$

where $0 < \alpha_c < 1$ is a cooling factor. It is likely that a number of moves are accepted at each temperature before proceeding to the new state. SA search is terminated at some final value T_f. An empirical choice for α_c is 0.95, and T_0 should be chosen such that the initial acceptance probability is higher than 0.8.

2. Objective function Q

Two objective measures, the segmental SNR (denoted as SNRseg) and the perceptual evaluation of sound quality (PESQ),[21] are considered in constructing the objective function for optimizing the performance in the MMSE-TRA-NR algorithm. The index SNRseg calculates SNR based on the noisy signals and the processed signals averaged over frames,

$$\mathrm{SNRseg} = \frac{10}{M_s}\sum_{m=0}^{M_s-1}\log_{10}\frac{\sum_{n=N_s m}^{N_s m+N_s-1} s^2(n)}{(s(n)-\hat{s}(n))^2}, \tag{14}$$

where N_s is the frame length and M_s is the number of frames. The SNRseg is a widely used objective measure for assessing NR performance in the telephony industry. The index PESQ is a more sophisticated objective measure for assessing speech quality, which takes into account psychoacoustic aspects of human hearing. The original and the processed signals are first level—equalized to a standard listening level and filtered by a filter, with a response similar to a standard telephone handset. The signals are aligned in time to correct for time delays and then processed through an auditory transform to obtain the loudness spectra. A more detailed information of the PESQ can be found in ITU-T P. 862.[21]

SNRseg and PESQ reflect the NR performance and the sound quality, respectively, of the processed signals. Hence, an objective function Q is constructed by combining the SNRseg and the PESQ using a weighting factor r, i.e.,

$$Q = r \times \mathrm{SNRseg} + \mathrm{PESQ}. \tag{15}$$

The weighting factor r will be found from a subjective listening test. Two kinds of background noise at the SNR level of 5 dB, white noise and car noise, were processed using five NR algorithms including spectral subtraction, Wiener filtering, MMSE-VAD-NR, MMSE-TRA-NR, and KLT-NR. The TRA parameters in MMSE-TRA-NR are chosen to be β=0.6 and δ=1.5. Figure 6 shows the clean speech signal used in the simulation. The test signal is a male speech sentence sampled at 8 kHz and separated into 25 ms frames with 50% overlap. The test signals last for 2 s in duration. All test signals were adjusted to the same level of loudness. A headset was used as the means of audio rendering.

Owing to the space limitation, we show only the results processed using the MMSE-TRA-NR algorithm. Figures 5(a) and 5(b) show the spectrograms of the noise and the signal processed by the MMSE-TRA-NR algorithm for the white noise case. Figures 6(a) and 6(b) show the spectrograms of the noise and the signal processed by the MMSE-TRA-NR algorithm for the car noise case. Thirty-two experienced listeners participated in the listening test. Three subjective indices including NR, *sound quality*, and *total preference* were employed in the listening test. The grading scale is set to be −3 to 3. A multiple regression analysis based on five NR algorithms and two background noises was utilized to establish a linear model between the NR, sound quality, and total preference. The results of the multiple regression analysis determine the weighting factors between the SNRseg and the PESQ for the objective function. This gives the weighting factor in Eq. (27), r=1.867, which will be used in the objective function in optimizing the MMSE-TRA-NR algorithm using the SA method next.

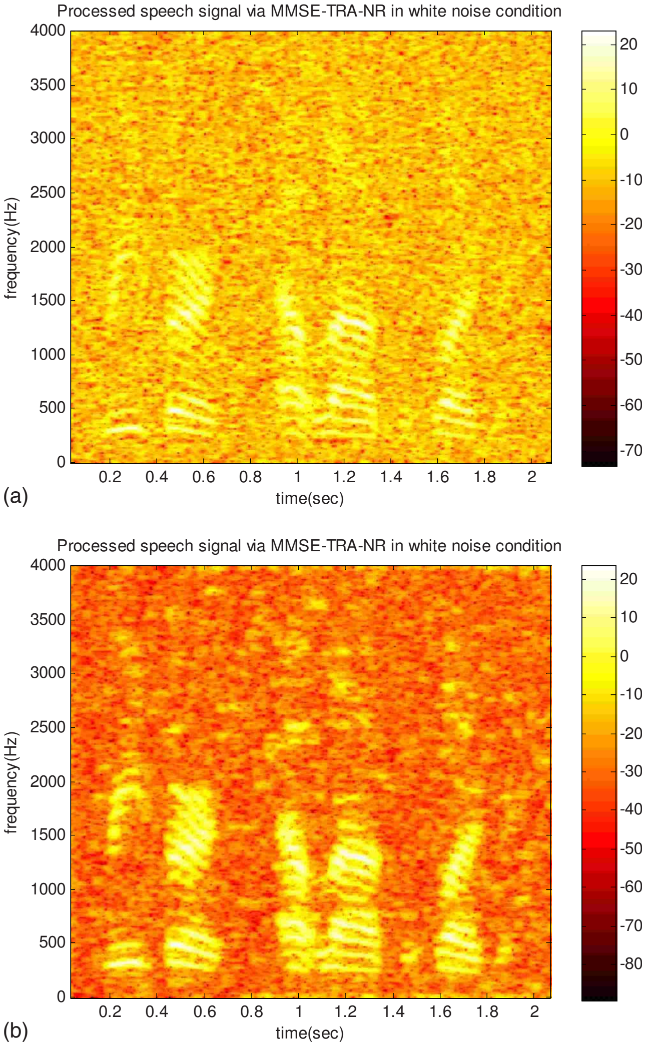

FIG. 5. (Color online) The spectrograms of a male speech sentence in the white noise scenario. (a) Speech corrupted by the white noise. (b) Enhanced speech signal processed by the MMSE-TRA-NR algorithm.

3. SA optimization of the MMSE-TRA-NR algorithm

The objective function with r=1.867 is employed in the SA optimization of the MMSE-TRA-NR algorithm. Initially, the TRA parameters are arbitrarily chosen to be β=1.6 and δ=1. The parameters of SA are chosen as T_0=1 K, T_f $=10^{-9}$ K, and α_c=0.95. With the SA optimization, the optimal parameters are obtained for the white noise (β=0.6117 and δ=0.5214) and the car noise (β=0.7128 and δ=0.5265). Figure 7 shows the "learning curve" of the SA for the car noise scenario, where the objective function Q settles to a constant value after about 500 iterations. To see the effect of optimization, NR performances in terms of the SNRseg and PESQ attained using the initial and the optimal parameters β and δ are compared in Table I. In comparison with the initial nonoptimal setting, a marked improvement in performance is obtained using the optimal TRA parameters.

To further justify the optimized NR algorithm, a subjective listening test was conducted. The test speech signal and

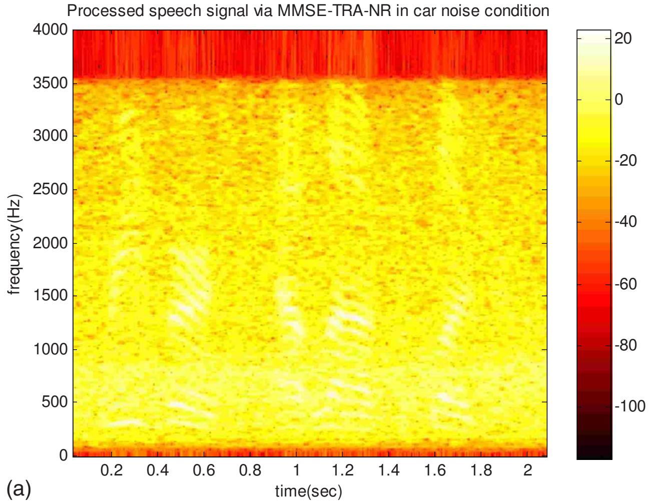

FIG. 6. (Color online) The spectrograms of a male speech sentence in the car noise scenario. (a) Speech corrupted by the car noise. (b) Enhanced speech signal processed by the MMSE-TRA-NR algorithm.

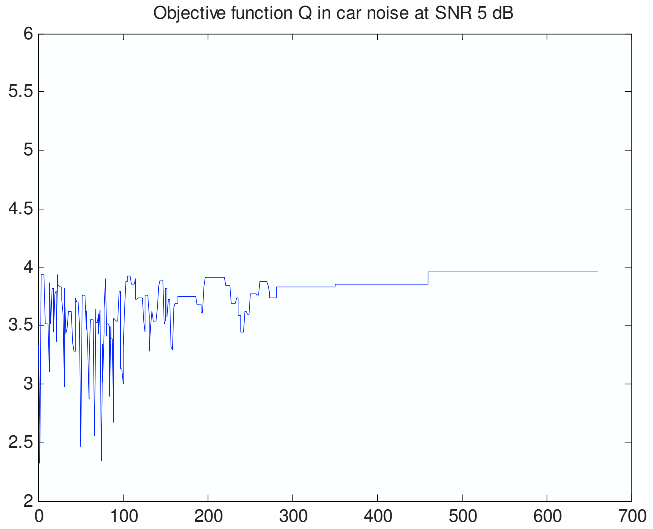

FIG. 7. (Color online) The learning curve of the SA optimization algorithm applied to the car noise.

the test conditions are the same as those used in the listening test for the preceding regression analysis. The grading scale is set to be 1–5, as recommended by ITU-T P.835.[22] Three subjective indices, including *scale of signal distortion (SIG)*, *scale of background intrusiveness (BAK)*, and *scale of overall quality (OVL)*, were employed in the listening test. Every subject participating in the test was instructed with the definitions of the subjective indices prior to the listening test. Figures 8(a) and 8(b) show the results of the listening test for the white noise and car noise, respectively. The grades were also processed by using the Multivariate Analysis of Variance (MANOVA) (Ref. 23) to justify the statistical significance of the test results. The average—a 5%–95% bracket is shown in the figure—and the significance level of the grades were summarized in Table II. Cases with significance levels below 0.05 indicate that a statistically significant difference exists among methods. Although there is no significant difference in OVL, the difference in SIG and BAK between the initial and optimal results is significant. The trade-off between NR (BAK) and signal distortion (SIG) is clearly visible—the optimized algorithm has attained remarkable NR performance at some expense of speech quality. Thus, we choose the optimized MMSE-TRA-NR algorithm for the following objective and subjective comparison with several other NR algorithms.

C. Linear prediction coding preprocessor

Another possibility of enhancing NR algorithms is to use LPC as the preprocessor. The underlying idea is that the highly correlated portion of human speech can be extracted by using the LPC approach. The timbral quality of voice is preserved as the spectral envelope is captured using the LPC. Figure 10(a) illustrates the one-step forward linear prediction problem.[17–19] The current input $x(n)$ is predicted by a linear combination of past input samples,

TABLE I. The NR performance of the MMSE-TRA-NR algorithm in terms of the objective measures SNRseg and PESQ for the initial and the optimized parameters β and δ (the optimal parameters are marked with *).

	MMSE-TRA-NR parameters				
Noise type	β	δ	SNRseg	PESQ	Q
White noise	1.6	1	−1.0942	1.9639	−0.1984
	0.6117*	0.5214*	1.5155	2.1619	4.8106
Car noise	1.6	1	−1.5609	2.2168	−0.2998
	0.7128*	0.5265*	0.7061	2.3145	3.9544

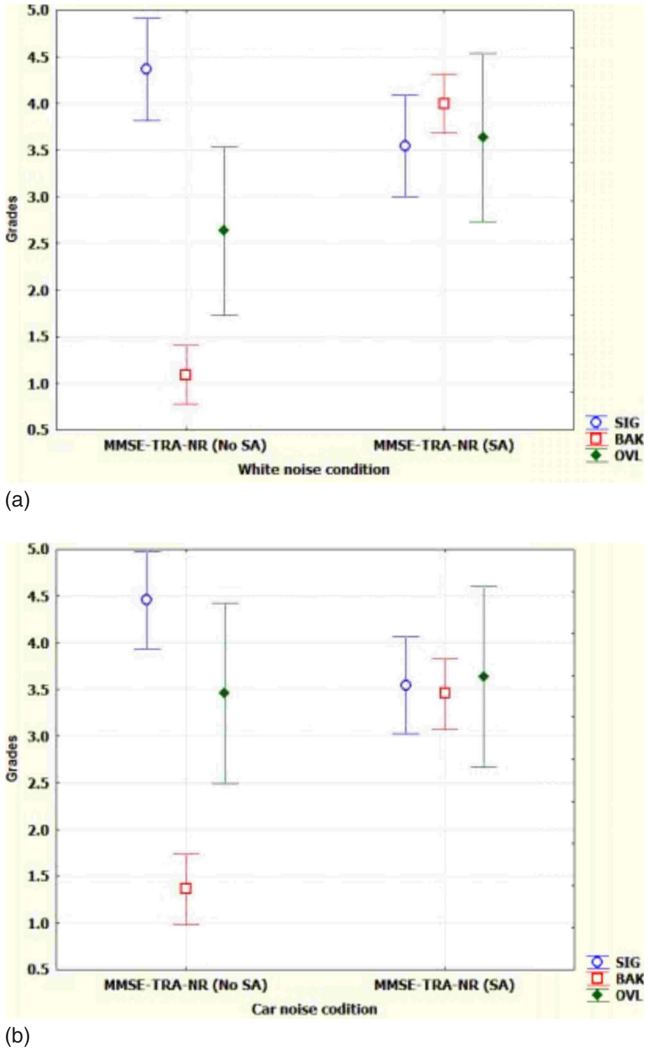

FIG. 8. (Color online) Comparison of the MMSE-TRA-NR algorithm with and without SA optimization. The results of the listening test are processed by using the MANOVA. (a) White noise. (b) Car noise.

$$\hat{x}(n)=\sum_{k=1}^{p}A_k x(n-k), \tag{16}$$

where p is the prediction order and A_k are the prediction coefficients. The associated prediction finite impulse response (FIR) filter is

$$P(z)=\sum_{k=1}^{p}A_k z^{-k}. \tag{17}$$

TABLE II. The MANOVA output of the subjective listening test to compare the MMSE-TRA-NR algorithm with and without optimization. The background noises are the white noise and the car noise. Cases with significance value p below 0.05 indicate that statistically significant difference exists among all methods.

	Significance value		
Noise type	SIG	BAK	OVL
White noise	0.040	0.000	0.117
Car noise	0.017	0.000	0.784

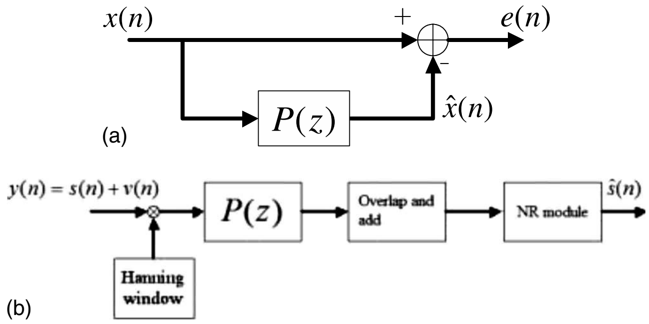

FIG. 9. The NR algorithm cascaded with a LPC preprocessor. (a) Feedforward linear prediction structure. (b) The cascaded LPC-NR system.

By minimizing the mean squares of the one-step forward prediction error, $E_p=E\{e^2(n)\}=E\{[x(n)-\hat{x}(n)]^2\}$, the following equation for the linear prediction problem can be derived:

$$\sum_{k=0}^{p}A_k\gamma_{xx}(l-k)=\begin{cases}E_p^f, & l=0\\ 0, & l=1,2,\ldots,p,\end{cases} \tag{18}$$

where E_p^f is the mean of the forward prediction error of order p and

$$\begin{aligned}\gamma_{xx}(m)&=E\{x^*(n)x(n+m)\}\\ &=\lim_{N\to\infty}\frac{1}{2N+1}\sum_{n=-N}^{N}x^*(n)x(n+m)\end{aligned} \tag{19}$$

is the autocorrelation sequence. The optimal LPC coefficients of the prediction filter can be efficiently calculated by using the Levinson–Durbin algorithm. According to the LPC coefficients, The noisy input can be preprocessed by using the prediction filter $P(z)$ in Eq. (17) to extract the correlated input with minimal timbral distortion for the MMSE-TRA-NR module. Figure 9(b) illustrates a MMSE algorithm concatenated with the LPC as its preprocessor (denoted as LPC-MMSE-TRA-NR).

IV. OBJECTIVE AND SUBJECTIVE EVALUATIONS OF NR ALGORITMS

Objective and subjective experiments were undertaken to compare the proposed optimized LPC-MMSE-TRA-NR algorithm with a number of other widely used NR algorithms.

A. Performance evaluation of NR algorithms by objective measures

The preceding objective measures SNRseg and the PESQ are employed to assess the performance of six NR algorithms (spectral subtraction, Wiener filtering, MMSE-VAD-NR, MMSE-TRA-NR, LPC-MMSE-TRA-NR, and KLT-NR algorithms) for the speech signal corrupted by two kinds of background noise (white noise and car noise). All test signals and conditions are similar to those used in the previous test.

According to Table III, the Wiener filtering algorithm tends to underestimate noise level and yield high residual noise (or low SNRseg). The KLT-NR algorithm attains the highest SNRseg. In addition, LPC seems to slightly improve the speech quality over the MMSE-TRA-NR algorithm for

TABLE III. Comparison of processing time and objective NR measures for six NR algorithms.

Noise type	SNRseg		PESQ	
	Noise type			
NR algorithm	White	Car	White	Car
Spectral subtraction	2.115	1.450	2.224	2.118
Wiener filtering	0.878	0.073	2.162	2.322
MMSE-VAD-NR	2.215	1.224	2.250	2.394
MMSE-TRA-NR	1.515	0.7061	2.161	2.314
LPC-MMSE-TRA-NR	1.439	0.3110	2.234	2.162
KLT-NR	3.177	1.856	2.400	2.367

the white noise case. As for the PESQ objective evaluation, there seems to be no significant difference in speech quality resulting from these NR algorithms.

B. Performance evaluation of NR algorithms by subjective measures

In order to further compare the preceding NR algorithms, subjective listening tests were conducted according to the ITU-T P.835.[22] Thirty-two experienced listeners participated in the subjective tests. The six NR algorithms used in the objective test are compared again in this subjective test. The test signals and conditions remain the same as the preceding listening tests (Table IV). The mean and spread of the listening test results are shown in Figs. 10(a) and 10(b). The test results were processed using MANOVA (Ref. 23) with significance levels summarized in Table V. Cases with significance levels below 0.05 indicate that a statistically significant difference exists among methods. From Table V, the difference of the indices SIG, BAK, and OVL among the NR methods was found to be statistically significant (except for OVL in the car noise scenario).

Next, a *post hoc* Tukey HSD test[23] was employed to perform multiple paired comparisons of the NR algorithms. *Post hoc* tests are generally performed after ANOVA, which is able to determine whether or not significant difference is present in the data of a number of cases. Tukey's HSD test is one of the commonly used *post hoc* tests for the assessment of differences in the means between pairs of populations following the ANOVA test. Table VI summarizes the results of the test in terms of the subjective indices SIG, BAK, and OVL. To facilitate the comparison, the NR algorithms that have attained good subjective performance (with no statistical difference) are marked with asterisks in the table. In Figs. 10(a) and 10(b), surprisingly, in contrast to the results of objective evaluation, the KLT-NR algorithm performed quite poorly in SIG for all noise conditions. The price paid for high NR using the KLT-NR algorithm is obviously the signal distortion, which was noticed by many subjects. Despite the excellent performance in SIG, the Wiener filtering algorithm received the lowest scores in BAK for all noise conditions,

TABLE IV. The optimal parameters β and δ obtained using the SA search for nine types of background noise (babble, station, car, airport, street, train, exhibition, restaurant, and white noise).

Background noise	Optimal β	Optimal δ
White noise	0.6117	0.5214
Babble	0.7178	0.8710
Station	0.6889	0.5350
Car	0.7128	0.5265
Airport	0.6259	0.5016
Street	0.5266	0.5016
Train	0.4609	0.5043
Exhibition	0.5440	0.5026
Restaurant	0.5103	0.5310

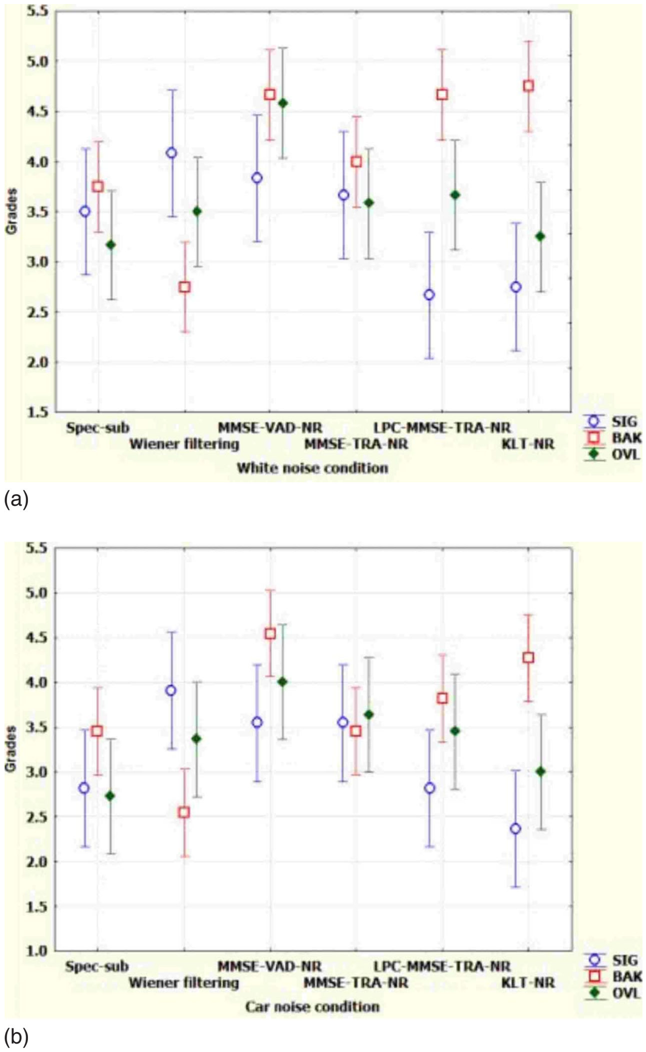

FIG. 10. (Color online) Comparison of six NR algorithms in time-domain waveforms. (a) The noisy and processed signals in the white noise condition. (b) The noisy and processed signals in the car noise condition (dotted line: noisy speech signals; solid line: processed speech signals).

TABLE V. The MANOVA output of the listening test of the six NR algorithms. Cases with significance value p below 0.05 indicate that statistically significant difference exists among all methods.

	Significance value p		
Noise type	SIG	BAK	OVL
White noise	0.008	0.000	0.008
Car noise	0.011	0.000	0.093

TABLE VI. The *post hoc* Tukey HSD test of the subjective measures SIG, BAK, and OVL obtained using six NR algorithms. The NR algorithms that have attained good subjective performance (with no statistical difference) are marked with asterisks.

	SIG		BAK		OVL	
	Noise condition					
NR algorithms	White	Car	White	Car	White	Car
Spectral subtraction	*	*				*
Wiener filtering	*	*			*	*
MMSE-VAD-NR	*	*	*	*	*	*
MMSE-TRA-NR	*	*	*		*	*
LPC-MMSE-TRA-NR		*	*	*	*	*
KLT-NR			*	*		*

which is consistent with the observation in the objective evaluation. The spectral-subtraction algorithm received the lowest grade in BAK for all noise conditions because of the "musical noise"[11] problem, which is quite disturbing to the listeners. There is no significant difference in OVL among all NR algorithms for the car noise scenario. The spectral-subtraction and KLT-NR algorithms received lower scores in OVL than the other algorithms in the white noise case. It can be concluded that the MMSE-VAD-NR, MMSE-TRA-NR, and LPC-MMSE-TRA-NR algorithms are superior to the other algorithms.

Overall, these three algorithms performed equally well in terms of all subjective indices in the two noise scenarios. For background noise with rapidly varying levels, however, the MMSE-TRA-NR algorithm should be more practical than the MMSE-VAD-NR. The LPC preprocessor may contribute to enhancing the NR algorithms, albeit this observation is not statistically significant.

C. Sensitivity analysis in the MMSE-TRA-NR algorithm

In this section, a sensitivity analysis is presented to demonstrate the effect of the choice of TRA parameters. The SA method is employed to search for the optimal parameters β and δ of the aforementioned MMSE-TRA-NR algorithm in dealing with nine types of background noise at the SNR level of 5 dB. These nine types of noise include babble, station, car, airport, street, train, exhibition, restaurant, and white noise, which were taken from the database of Ref. 11. The results of the optimal parameters β and δ summarized in Table IV are plotted in a scatter diagram in Fig. 11 for each noise condition. It is worth noting that the NR performance of the MMSE-TRA-NR algorithm is very sensitive to the choice of the parameter β. The optimal parameter β falls in the range of $0 \leqslant \beta \leqslant 1$ for all noise conditions, which is quite different from the values of $15 \leqslant \beta \leqslant 30$ recommended in Ref. 11. By contrast, the optimal parameter δ is relatively constant (≈ 0.5) for all types of background noise except for "babble" ($\delta = 0.871$), which is also different from the value $\delta = 1.5$ recommended in Ref. 11. The recommended parameter δ should be in the range of $0.5 \leqslant \delta \leqslant 1.5$ because δ decides the transition point of the sigmoid function in the previous TRA algorithm. The transition point can be considered as a threshold to discriminate speech presence from speech absence according to the *a posteriori* SNR. In the present study, a judicious but more reasonable range of $0.5 \leqslant \delta \leqslant 1.5$ is recommended.

V. CONCLUSIONS

An optimized MMSE-TRA-NR algorithm has been presented. The SA optimization technique is exploited to search for optimal TRA parameters, especially the parameter β, which has a profound impact on the estimation of the noise spectrum and hence the resulting NR performance of the algorithm. The optimal parameter β generally falls in the range of $0 \leqslant \beta \leqslant 1$, whereas the optimal parameter δ stays at a relatively constant value of 0.5 for many types of background noise. In addition, a LPC preprocessor has been presented to enhance the MMSE-TRA-NR algorithm.

The proposed NR algorithms have been compared with several other widely used algorithms via extensive objective and subjective tests. These methods exhibit different degrees in trading off reduction performance and speech quality. It can be concluded that the MMSE-VAD-NR, MMSE-TRA-NR, and LPC-MMSE-TRA-NR algorithms are more superior

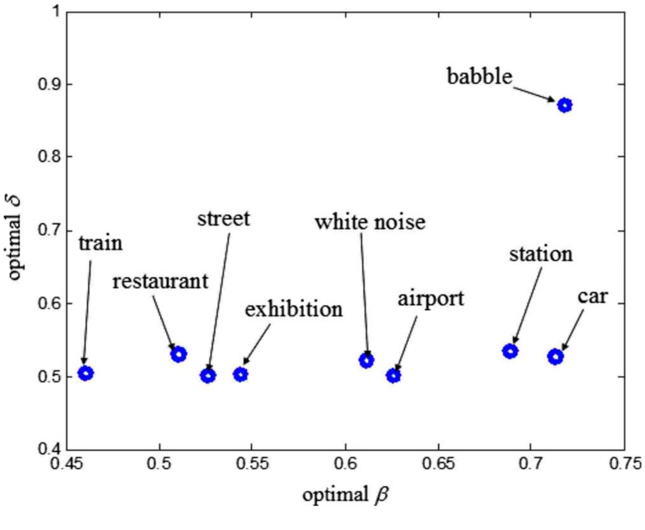

FIG. 11. (Color online) Sensitivity analysis of the optimal parameters β and δ of the MMSE-TRA-NR algorithm for nine kinds of background noise.

to the other algorithms. Overall, these three algorithms performed equally well in terms of all subjective indices in the white and car noise scenarios. For background noise with rapidly varying levels, however, the MMSE-TRA-NR algorithm is more practical than the MMSE-VAD-NR.

ACKNOWLEDGMENTS

This work was supported by the National Science Council of Republic of China, under Project No. NSC 95-2221-E-009-179.

[1] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **32**, 1109–1121 (1984).

[2] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," IEEE Trans. Acoust., Speech, Signal Process. **28**, 137–145 (1980).

[3] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach* (Wiley, New York, 2004).

[4] R. E. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/synthesis," IEEE Trans. Acoust., Speech, Signal Process. **281**, 99–102 (1980).

[5] M. R. Portnoff, "Implementation of the digital phase vocoder using the fast Fourier transform," IEEE Trans. Acoust., Speech, Signal Process. **24**, 243–248 (1976).

[6] U. Zölzer, *DAFX—Digital Audio Effects* (Wiley, New York, 2002).

[7] S. L. Gay and J. Benesty, *Acoustic Signal Processing for Telecommunication* (Kluwer Academic, Norwell, MA, 2000).

[8] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications* (Wiley, New York, 1949).

[9] B. Farhang-Boroujeny, *Adaptive Filters Theory and Application* (Wiley, New York, 2000).

[10] S. V. Vaseghi, *Advanced Signal Processing and Digital Noise Reduction* (Wiley, New York, 1996).

[11] P. C. Loizou, *Speech Enhancement Theory and Practice* (CRC, New York, 2007).

[12] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," IEEE Trans. Acoust., Speech, Signal Process. **11**, 334–341 (2003).

[13] L. Lin, W. Holmes, and E. Ambikairajah, "Adaptive noise estimation algorithm for speech enhancement," Electron. Lett. **39**, 754–755 (2003).

[14] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," J. Chem. Phys. **21**, 1087–1092 (1953).

[15] *Quantum Annealing and Related Optimization Methods*, edited by A. Das and B. K. Chakrabarti (Springer, Heidelberg, 2005).

[16] J. De Vicente, J. Lanchares, and R. Hermida, "Placement by thermodynamic simulated annealing," Phys. Lett. A **317**, 415–423 (2003).

[17] J. Makhoul, "Linear prediction: A tutorial review," Proc. IEEE **63**, 561–580 (1975).

[18] J. D. Markel and A. H. Gray, *Linear Prediction of Speech* (Springer-Verlag, Berlin, 1976).

[19] S. J. Orfanidis, *Optimum Signal Processing: An Introduction* (McGraw-Hill, New York, 1996).

[20] ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," International Telecommunications Union, Geneva, Switzerland, 2000.

[21] ITU-T Rec. P.835, "Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm," International Telecommunications Union, Geneva, Switzerland, 2003.

[22] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Trans. Acoust., Speech, Signal Process. **9**, 504–512 (2001).

[23] G. Keppel and S. Zedeck, *Data Analysis for Research Designs* (Freeman, New York, 1989).