A Parametric Formulation of the Generalized Spectral Subtraction Method

Boh Lim Sim, Yit Chow Tong, Joseph S. Chang, and Chin Tuan Tan, *Member, IEEE*

***Abstract*— In this paper, two short-time spectral amplitude estimators of the speech signal are derived based on a *parametric formulation* of the original generalized spectral subtraction method. The objective is to improve the noise suppression performance of the original method while maintaining its computational simplicity. The proposed parametric formulation describes the original method and several of its modifications. Based on the formulation, the speech spectral amplitude estimator is derived and optimized by minimizing the mean-square error (MSE) of the speech spectrum. With a constraint imposed on the parameters inherent in the formulation, a second estimator is also derived and optimized. The two estimators are different from those derived in most modified spectral subtraction methods, which are predominantly nonstatistical. When tested under stationary white Gaussian noise and semistationary Jeep noise, they showed improved noise suppression results.**

***Index Terms*—Speech enhancement, speech processing.**

I. Introduction

ACOUSTIC noise suppression is useful in applications such as voice communication, automatic speech recognition, and speech processing for the hearing impaired [1]–[4]. This paper focuses on a *parametric formulation* of a conventional single-channel acoustic noise suppression method—the original generalized spectral subtraction method.

In the original generalized spectral subtraction method [5]–[7], the noise-suppressed speech spectrum is obtained by subtracting an average noise spectrum from the noisy speech spectrum. Although the original method is capable of noise suppression, it often results in *excessive* residual noise including *unpleasant* musical tone artifacts. In this paper, residual noise will be described as consisting of: i) the remaining noise that has the same perceptual characteristics as the original noise, and ii) the musical tone artifacts due to the presence of random short-duration spectral peaks in the residual noise spectrum.

Several authors have made modifications to the original method to reduce the residual noise (and musical tone artifacts). The values of the parameters inherent in the modifications were usually empirically adjusted to achieve improved performance on the basis of subjective listening tests. For example, Berouti *et al.* [9] proposed an *oversubtraction* of noise spectrum to reduce the overall residual noise level and an inclusion of a fraction of noise as a spectral floor to mask the musical tone artifacts. The method of oversubtraction and the fraction of noise were both empirically determined based on subjective listening tests. On the other hand, Kushner *et al.* [11] found that by empirically adjusting the amount of *undersubtraction* of noise and by including a fraction of noise as spectral floor, the accuracy of linear predictive coefficient (LPC) spectral estimation and the performance of automatic speech recognizers can be improved. Boll [10] proposed spectral averaging over time and several additional residual noise reduction schemes to further enhance the spectrally subtracted speech. Crozier *et al.* [13] proposed a formant filter to reduce the musical tones in the spectrally subtracted speech. Vaseghi *et al.* [12] proposed a scheme that deletes spectral components (of the spectrally subtracted speech) based on empirically determined spectral/time thresholds to reduce the musical tones. Most of the reported modifications are predominantly nonstatistical in nature.

In this paper, the approach to improve the noise suppression performance of the original method is based on a *parametric formulation* of the original method and employing *statistical optimization*. A parametric formulation of the original method is proposed to formally describe the original method and its modifications [9]–[13]. Based on the formulation, two short-time spectral amplitude estimators of the speech signal are derived and optimized by minimizing the mean-square of the error between the original (ideal) spectrum and the proposed parametric spectrum. The estimators are denoted as the unconstrained and constrained parametric estimators. The mathematical expressions of the two optimized parametric estimators are further simplified by assuming that the speech and noise spectral components are statistically independent Gaussian random variables.

The paper is organized as follows. In Section II, the unconstrained and constrained parametric estimators are derived. In Section III, the implementation and the noise suppression results of the two parametric estimators are presented. In Section IV, the results are discussed in terms of mean-square estimation error functions and spectral gain functions. Finally in Section V, this paper is concluded.

II. Derivations of the Parametric Estimators

In this section, the derivations of the unconstrained and constrained parametric estimators are presented.

Manuscript received July 26, 1995; revised September 2, 1997. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. John H. L. Hansen.

The authors are with Research Lab II, School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore 639798, Republic of Singapore (e-mail: E94007E60@ntuvax.ntu.ac.sg).

Publisher Item Identifier S 1063-6676(98)04221-7.

A. Parametric Formulation of the Generalized Spectral Subtraction Method

The noisy speech $y(n)$ is assumed to consist of the clean speech $s(n)$ additively degraded by uncorrelated random noise $d(n)$, as follows:

$$\begin{aligned} &\text{Time domain:} && y(n) = s(n) + d(n) \\ &\text{Frequency domain:} && Y(w_k) = S(\omega_k) + D(\omega_k) \\ &\text{or} && Y_k\, e^{j\phi_{y,k}} = S_k\, e^{j\phi_{s,k}} + D_k\, e^{j\phi_{d,k}} \end{aligned} \tag{1}$$

where $Y(\omega_k)$, $S(\omega_k)$, and $D(\omega_k)$ are discrete Fourier transforms (DFT's), with amplitudes Y_k, S_k, and D_k, and phases $\phi_{y,k}$, $\phi_{s,k}$, and $\phi_{d,k}$, respectively, at frequency ω_k or frequency channel k.

In the original generalized spectral subtraction method [5]–[7], S_k in (1) is estimated using the following phase-blind equation:

$$\tilde{S}^{\alpha}_{k,\alpha} = Y^{\alpha}_k - E[D^{\alpha}_k] \tag{2}$$

where i) $\tilde{S}^{\alpha}_{k,\alpha}$, Y^{α}_k, and $E[D^{\alpha}_k]$ are referred to as the noise-suppressed speech spectrum, the noisy speech spectrum and the average noise spectrum respectively in this study; ii) α is the generalized power exponent for the spectrum; and iii) $\tilde{S}_{k,\alpha}$ will be called the *original estimator* of S_k. At $\alpha = 1$, (2) is sometimes simply called the spectral subtraction method [10], and at $\alpha = 2$, the correlation [5] or the power spectral subtraction method [1].

In this study, we define a *parametric* formulation of the original generalized spectral subtraction method

$$\hat{S}^{\alpha}_{k,\alpha} = a_{k,\alpha} Y^{\alpha}_k - b_{k,\alpha} E[D^{\alpha}_k] \tag{3}$$

where i) $\hat{S}_{k,\alpha}$ will be called the *parametric estimator* of S_k; and ii) $a_{k,\alpha}$ and $b_{k,\alpha}$ are the parameters in the parametric formulation. The physical significance of $a_{k,\alpha}$ and $b_{k,\alpha}$ can be illustrated by the following examples. If we set $a_{k,\alpha} = 1$ and $b_{k,\alpha} = \nu$ where ν is the noise subtraction factor, then (3) becomes the equation considered by Berouti *et al.* [9], Kushner *et al.* [11], and Lim–Oppenheim [1, eq. (13)]. If we set $b_{k,\alpha} = a_{k,\alpha}\nu$, then $a_{k,\alpha}$ represents the weighting at frequency channel k of the formant filter proposed by Crozier *et al.* [13]. Finally, if we set $b_{k,\alpha} = a_{k,\alpha} = 1$, we have the original equation (2).

Assuming perceptual unimportance of phase (e.g., see [1] and [8]), the final noise-suppressed speech $\hat{S}(\omega_k)$ is obtained using $\hat{S}_{k,\alpha}$ in (3) and the *noisy* phase $\phi_{y,k}$:

$$\hat{S}(\omega_k) = \hat{S}_{k,\alpha}\, e^{j\phi_{y,k}}. \tag{4}$$

B. Derivation of the Parametric Estimator Without Constraint—$\hat{S}_{k,\alpha}$

The parametric estimator in (3) is optimized by minimizing the mean-square of an appropriate estimation error function—that is, by using a MMSE criterion. The error function, $\hat{\delta}_{k,\alpha}$, used in this study is defined as the error of the speech spectrum

$$\hat{\delta}_{k,\alpha} = S^{\alpha}_{k,\alpha} - \hat{S}^{\alpha}_{k,\alpha} \tag{5}$$

where $\hat{S}^{\alpha}_{k,\alpha}$ is the noise-suppressed speech spectrum from (3), and $S^{\alpha}_{k,\alpha}$ is the spectrum using an assumed "ideal" generalized spectral subtraction model. In this model, the noisy spectrum Y^{α}_k is assumed to be the sum of two independent spectra—the modeled speech spectrum $S^{\alpha}_{k,\alpha}$ and the *exact* noise spectrum D^{α}_k, for some constant α:

$$Y^{\alpha}_k = S^{\alpha}_{k,\alpha} + D^{\alpha}_k. \tag{6}$$

The above equation can be rewritten as $S^{\alpha}_{k,\alpha} = Y^{\alpha}_k - D^{\alpha}_k$, which is similar to the original equation in (2). It can thus be considered as the "ideal" equation for the original method, since the exact noise spectrum D^{α}_k is used. Note that in practice, when only the noisy speech alone is available, the exact noise spectrum D^{α}_k cannot be obtained. Also, note that when α is equal to two, the resulting estimates for $a_{k,\alpha}$ and $b_{k,\alpha}$ are optimal in the mean-square sense. However, if α is set equal to some other value, then the equal sign cannot be strictly used in (6), and the resulting $a_{k,\alpha}$ and $b_{k,\alpha}$ are not strictly optimal, but can still be useful in the enhancement of speech.

Substituting (3) and (6) into (5), we have

$$\hat{\delta}_{k,\alpha} = (1 - a_{k,\alpha}) S^{\alpha}_{k,\alpha} - a_{k,\alpha} D^{\alpha}_k + b_{k,\alpha} E[D^{\alpha}_k]. \tag{7}$$

From (7), the mean-square of the estimation error function $\hat{\delta}_{k,\alpha}$ is

$$\begin{aligned} E\left[\left\{\hat{\delta}_{k,\alpha}\right\}^2\right] = &(1 - a_{k,\alpha})^2 E[S^{2\alpha}_{k,\alpha}] + a^2_{k,\alpha} E[D^{2\alpha}_k] \\ &+ (b^2_{k,\alpha} - 2a_{k,\alpha} b_{k,\alpha}) \{E[D^{\alpha}_k]\}^2 \\ &+ 2(1 - a_{k,\alpha})(b_{k,\alpha} - a_{k,\alpha}) E[S^{\alpha}_{k,\alpha}] E[D^{\alpha}_k]. \end{aligned} \tag{8}$$

By differentiating (8) separately with respect to $a_{k,\alpha}$ and $b_{k,\alpha}$ and followed by setting the results to zero, the optimum $a_{k,\alpha}$ and $b_{k,\alpha}$ are

$$\begin{aligned} a_{k,\alpha} = &\big\{E[S^{2\alpha}_{k,\alpha}] + b_{k,\alpha}\{E[D^{\alpha}_k]\}^2 \\ &+ (1 + b_{k,\alpha}) E[S^{\alpha}_{k,\alpha}] E[D^{\alpha}_k]\big\} \\ &/\big\{E[S^{2\alpha}_{k,\alpha}] + E[D^{2\alpha}_k] + 2E[S^{\alpha}_{k,\alpha}] E[D^{\alpha}_k]\big\} \end{aligned} \tag{9}$$

$$b_{k,\alpha} = a_{k,\alpha} + (a_{k,\alpha} - 1) E[S^{\alpha}_{k,\alpha}] / E[D^{\alpha}_k]. \tag{10}$$

The optimum $a_{k,\alpha}$ and $b_{k,\alpha}$ can be simplified greatly if the expected magnitude spectra in the above equations can be written in terms of $E[D^2_k]$ and $E[S^2_{k,\alpha}]$. This can be done by further assuming that each individual spectral component of the speech and noise is a statistically independent zero-mean complex Gaussian random variable with a time-varying variance. The mean is zero ($E[S(\omega_k)] = E[D(\omega_k)] = 0$) as the speech and noise signals are assumed to be zero-mean processes ($E[s(n)] = E[d(n)] = 0$). The variance

($E[S_k^2]$ or $E[D_k^2]$) is time varying to account mainly for the nonstationarity of the signals. This statistical assumption is well elaborated in [15, pp. 1109–1110]. The use of this assumption here is largely motivated by the positive results in the studies in [15]–[17].

Using this assumption, the noise spectral amplitude D_k has a Rayleigh distribution (see [20, p. 136]) as follows:

$$f(D_k) = \frac{2D_k}{E[D_k^2]} e^{-D_k^2/E[D_k^2]}, \qquad D_k \geq 0. \tag{11}$$

By using (11) and [19, Eq. (15.77)], $E[D_k^p]$ is given by

$$E[D_k^p] = \int_0^\infty D_k^p f(D_k)\, dD_k = \Gamma(p/2+1)E[D_k^2]^{p/2} \tag{12}$$

where p is the power exponent and $\Gamma(.)$ denotes the Gamma function [19, p. 235]. Similarly, $E[S_k^p]$ is given by

$$E[S_k^p] = \Gamma(p/2+1)E[S_k^2]^{p/2}. \tag{13}$$

Following (13), the relationship between $E[S_{k,\alpha}^p]$ and $E[S_{k,\alpha}^2]$ is simply assumed to be

$$E[S_{k,\alpha}^p] = \Gamma(p/2+1)E[S_{k,\alpha}^2]^{p/2}. \tag{14}$$

Thus, by substituting (12) and (14) into (9) and (10), and further solving, we obtain

$$a_{k,\alpha} = \xi_{k,\alpha}^\alpha / \{1 + \xi_{k,\alpha}^\alpha\} \tag{15}$$

$$b_{k,\alpha} = \xi_{k,\alpha}^\alpha / \{1 + \xi_{k,\alpha}^\alpha\}\ \{1 - \xi_{k,\alpha}^{-\alpha/2}\} \tag{16}$$

where

$$\xi_{k,\alpha} = E[S_{k,\alpha}^2]/E[D_k^2]. \tag{17}$$

The parametric estimator $\hat{S}_{k,\alpha}$ in (3) with the above optimum parameters becomes

$$\hat{S}_{k,\alpha} = \left\{ \frac{\xi_{k,\alpha}^\alpha}{1+\xi_{k,\alpha}^\alpha}\ (Y_k^\alpha - (1 - \xi_{k,\alpha}^{-\alpha/2})E[D_k^\alpha]) \right\}^{1/\alpha}. \tag{18}$$

C. Derivation of the Parametric Estimator with Constraint—$\hat{S}_{k,\alpha}\big|_{a_{k,\alpha}=b_{k,\alpha}}$

The parametric estimator $\hat{S}_{k,\alpha}$ in (18) is derived without constraints in the parameters $a_{k,\alpha}$ and $b_{k,\alpha}$. In this section, a specific constraint $a_{k,\alpha} = b_{k,\alpha}$ is considered. This is because the parametric estimator with this constraint is found to have a good noise suppression performance, which will be elaborated in Sections III and IV. When the constraint is substituted into (3) and using (2), (3) becomes $\{\hat{S}_{k,\alpha}|_{a_{k,\alpha}=b_{k,\alpha}}\}^\alpha = a_{k,\alpha}\,\tilde{S}_{k,\alpha}^\alpha$. Thus, $a_{k,\alpha}$ acts as a frequency weighting function on the noise-suppressed spectrum $\tilde{S}_{k,\alpha}^\alpha$ of the original method in (2). $\hat{S}_{k,\alpha}|_{a_{k,\alpha}=b_{k,\alpha}}$ will be called the *constrained* parametric estimator, and $\hat{S}_{k,\alpha}$ in (18) will be called the *unconstrained* parametric estimator.

The derivation of the constrained parametric estimator $\hat{S}_{k,\alpha}|_{a_{k,\alpha}=b_{k,\alpha}}$ is similar to that of the unconstrained estimator $\hat{S}_{k,\alpha}$ described in the last section. First, the constraint $a_{k,\alpha} = b_{k,\alpha}$ is substituted into (8). The resulting equation is then differentiated with respect to $a_{k,\alpha}$ and set to zero. The optimum $a_{k,\alpha}$ is then found as

$$a_{k,\alpha} = E[S_{k,\alpha}^{2\alpha}]/\{E[S_{k,\alpha}^{2\alpha}] + E[D_k^{2\alpha}] - \{E[D_k^\alpha]\}^2\}. \tag{19}$$

By substituting (12) and (14) into (19), the optimum $a_{k,\alpha}$ is simplified to

$$a_{k,\alpha} = \xi_{k,\alpha}^\alpha / \{\xi_{k,\alpha}^\alpha + \beta_\alpha\} \tag{20}$$

where

$$\beta_\alpha = \{\Gamma(\alpha+1) - \Gamma^2(\alpha/2+1)\}/\Gamma(\alpha+1). \tag{21}$$

β_α is a constant for a fixed power exponent α (β_α = 0.2146, 0.5, and 0.7055 for $\alpha = 1$, 2, and 3, respectively). The constrained parametric estimator $\hat{S}_{k,\alpha}|_{a_{k,\alpha}=b_{k,\alpha}}$ with the optimum $a_{k,\alpha}$ in (20) becomes

$$\hat{S}_{k,\alpha}\Big|_{a_{k,\alpha}=b_{k,\alpha}} = \left\{ \frac{\xi_{k,\alpha}^\alpha}{\xi_{k,\alpha}^\alpha + \beta_\alpha}\ (Y_k^\alpha - E[D_k^\alpha]) \right\}^{1/\alpha}. \tag{22}$$

D. Approximation of $\xi_{k,\alpha}$ and $E[D_k^\alpha]$

To calculate $\hat{S}_{k,\alpha}$ in (18) or $\hat{S}_{k,\alpha}|_{a_{k,\alpha}=b_{k,\alpha}}$ in (22), the term $\xi_{k,\alpha}$ and the average noise spectrum $E[D_k^\alpha]$ would need to be approximated.

$\xi_{k,\alpha}$ in (17) is the ratio of the speech energy $E[S_{k,\alpha}^2]$ to the noise energy $E[D_k^2]$, and it will be called the *a priori* signal-to-noise ratio (SNR). It is equivalent to the *a priori* SNR, $\xi_k = E[S_k^2]/E[D_k^2]$, defined by Ephraim–Malah [15], [16]. Following their method of approximating ξ_k [15, eq. (51)], $\xi_{k,\alpha}$ is approximated as an average SNR to exploit the local stationarity of speech signal

$$\xi_{k,\alpha} \approx (1-\eta_S)\ \underbrace{\frac{\mathrm{Max}(Y_k^2 - E[D_k^2], 0)}{E[D_k^2]}}_{\approx\text{current SNR}} + \eta_S\ \underbrace{\frac{\{\hat{S}_{k,\alpha,prev}\}^2}{E[D_k^2]|_{prev}}}_{\text{previous SNR}} \tag{23}$$

where η_S is a smoothing constant, $(Y_k^2 - E[D_k^2])$ is the speech energy estimated in the current frame, and $\hat{S}_{k,\alpha,prev}$ is the estimator in (18) or (22) from the *previous* frame.

$E[D_k^\alpha]$ in (18) is approximated during non speech intervals as

$$E[D_k^\alpha] \approx (1-\eta_D)D_k^\alpha + \eta_D D_{k,prev}^\alpha \tag{24}$$

where η_D is a smoothing constant. $E[D_k^2]$ in (23) is calculated using (24) with $\alpha = 2$.

E. A Smoothed Lower Bound for the Constrained Parametric Estimator $\hat{S}_{k,\alpha}\big|_{a_{k,\alpha}=b_{k,\alpha}}$

For the constrained parametric estimator in (22), a smoothed lower bound is used to limit its signal attenuation. This is because the estimator results in large noise reduction and sometimes overattenuation of very low energy speech segments. This will be further discussed in Sections III and IV. The smoothed lower bound, $\overline{\mu Y_k}$, is obtained by smoothing a lower bound, μY_k, with the estimator from previous frame

[$\overline{S}_{k,\alpha,prev}$, using (26)], as follows:

$$\overline{\mu Y_k} = 0.5\,(\mu Y_k + \overline{S}_{k,\alpha,prev}). \tag{25}$$

The lower bound μY_k is a fraction (μ) of the noisy speech (Y_k). The smoothing in (25) reduces the perceptual "roughness" of μY_k. μ will be referred to as the *spectral gain floor*.

The final constrained parametric estimator, $\overline{S}_{k,\alpha}$, is implemented as follows. If the estimator $\hat{S}_{k,\alpha}|_{a_{k,\alpha}=b_{k,\alpha}}$ in (22) is greater than μY_k, then $\overline{S}_{k,\alpha}$ is set equal to $\hat{S}_{k,\alpha}|_{a_{k,\alpha}=b_{k,\alpha}}$; otherwise, $\overline{S}_{k,\alpha}$ is set equal to $\overline{\mu Y_k}$ from (25)

$$\overline{S}_{k,\alpha} = \begin{cases} \hat{S}_{k,\alpha}\Big|_{a_{k,\alpha}=b_{k,\alpha}}, & \text{if } \hat{S}_{k,\alpha}\Big|_{a_{k,\alpha}=b_{k,\alpha}} \geq \mu Y_k \\ \overline{\mu Y_k}, & \text{otherwise.} \end{cases} \tag{26}$$

In the implementation of (26), it is important to note that $\hat{S}_{k,\alpha,prev}$ in (23) and $\overline{S}_{k,\alpha,prev}$ in (25) are different. $\hat{S}_{k,\alpha,prev}$ is the previous estimated speech spectral amplitude from (22), whereas $\overline{S}_{k,\alpha,prev}$ is the previous estimated amplitude from (26).

III. Implementation and Results

The noise suppression system consists of three blocks: Fourier analysis, noise suppression algorithm, and Fourier synthesis. First, the incoming noisy speech signal is windowed by a half-overlapped Hanning window and then spectrally decomposed by the fast Fourier transform (FFT). The spectral amplitude at channel k of the noise-suppressed speech is calculated by the unconstrained parametric estimator [using (18), (23), and (24)] or the constrained parametric estimator [using (22)–(26)]. The spectral amplitude is then combined with the noisy phase. Synthesis of the enhanced signal is by the inverse FFT (IFFT) and by overlapping and adding two consecutive frames according to the weighted overlap-add method [18].

The noisy speech signals at global SNR $=-5$–10 dB were used for performance evaluation. The global SNR is defined as the ratio, $\sum s(n)^2/\sum d(n)^2$, with the summations carried out over the whole sentence [8]. The noisy signals were obtained by digitally adding noise to clean speech signals (see Appendix). The noise includes stationary white Gaussian noise and semistationary Jeep noise (mainly the engine noise, with very little wind fluttering noise). The average noise spectrum $E[D_k^\alpha]$ in (24) is calculated from the initial nonspeech interval of the noisy speech with $\eta_D = 0.90$.

For both parametric estimators, subjective listening showed that the remaining noise level and the musical tone artifacts in the noise-suppressed speech are consistently lower than those for the original estimator. The remaining noise level and the musical tone artifacts for the constrained parametric estimator are lower than that for the unconstrained parametric estimator. For the unconstrained estimator, the preferred parameter values lie in the following ranges: $\alpha = 1.0$ to 2.0 and $\eta_S = 0.96$ to 0.995. For the constrained estimator, the ranges of parameter values are $\alpha = 1.0$ to 3.0, $\eta_S = 0.96$ to 0.995, and $\mu = 0.05$ to 0.20. The effects of the parameters η_S and μ will be discussed in Sections IV-B and IV-C. For the constrained estimator, the short-time frame duration for Fourier spectral analysis is preferably in the range of 30–50 ms. (Outside this range of duration, degradation of the speech quality was sometimes observed. In this case, the degradation can be reduced by raising the spectral gain floor μ to more than 0.20.)

The two parametric estimators were also compared with two of the modifications [9], [10] of the original estimator (2) and with the Ephraim–Malah estimator [16, eq. (20)]. For both parametric estimators, subjective listening tests showed that the remaining noise level and the musical tone artifacts are perceptually less significant than those for the modified spectral subtraction methods [9], [10]. The remaining noise level and the musical tone artifacts for the constrained estimator is generally comparable to that for the Ephraim–Malah estimator.

The objective improvements in global SNR and spectral distance after processing by the parametric estimators are significantly better than those by the original estimators and its modifications (at the same value of α) for noisy speech at global SNR's $<= 5$ dB. The improvements are less significant or comparable for SNR's > 5 dB. The improvements observed for the constrained parametric estimator are generally comparable to that for the Ephraim–Malah estimator [16] for all global SNR's tested. These observations are illustrated in the Appendix.

For the modified spectral subtraction method in [9], although the remaining noise level and the musical tone artifacts can be further reduced by a larger oversubtraction of the noise spectrum and an inclusion of a fraction of noise, we find that the low-energy speech segments are usually lost, especially under high level of noise. This is largely because the fraction of noise has no relation to the speech signal. However, if a fraction of the noisy speech [see (25)] is used instead of a fraction of the noise, the overattenuation of the low-energy speech segments can be reduced.

For the constrained parametric estimator, the remaining noise level can sometimes be further reduced (lower than that for the Ephraim–Malah estimator) when the global SNR of the noisy speech is greater than 5 dB. This is done by setting the fraction $\mu < 0.10$. Increase in musical tone artifacts and signal overattenuation is usually perceptually undetectable. This feature is particularly advantageous when μ is implemented as an adjustable parameter.

IV. Discussions

The observations described in Section III are discussed in this section. The unconstrained parametric estimator, the constrained parametric estimator and the original estimator are compared in terms of mean-square error (MSE) functions, mean-error functions, and spectral gain functions. The parametric estimators are also compared with the Ephraim–Malah estimator [16] in terms of spectral gain functions.

A. Analysis of the Mean-Square Error and Mean Error Functions

The results in Section III showed that the residual noise level associated with the two parametric estimators is less than that associated with the original estimator. In this section, it will be shown that the improvement can be explained by examining both the mean-square and mean error functions of the estimators.

The mean-square of the estimation error function $\hat{\delta}_{k,\alpha}$ in (5), or the MSE function of the unconstrained parametric estimator $\hat{S}_{k,\alpha}$ in (18), can be obtained by substituting (15)—(17) into (8), as follows:

$$E[\{\hat{\delta}_{k,\alpha}\}^2] = \frac{\xi_{k,\alpha}^{\alpha}}{\xi_{k,\alpha}^{\alpha}+1}\,\beta_\alpha\,\Gamma(\alpha+1)\,\{E[D_k^2]\}^{\alpha} \tag{27}$$

where β_α is given in (21). The MSE function of the constrained parametric estimator $\hat{S}_{k,\alpha}|_{a_{k,\alpha}=b_{k,\alpha}}$ in (22) can be obtained by substituting (17) and (20) into (8), as follows:

$$E\left[\left\{\hat{\delta}_{k,\alpha}\Big|_{a_{k,\alpha}=b_{k,\alpha}}\right\}^2\right] = \frac{\alpha_{k,\alpha}^{\alpha}}{\xi_{k,\alpha}^{\alpha}+\beta_\alpha}\,\beta_\alpha\Gamma(\alpha+1)\{E[D_k^2]\}^{\alpha}. \tag{28}$$

The MSE function of the original estimator $\tilde{S}_{k,\alpha}$ in (2) can be obtained by substituting $a_{k,\alpha}=b_{k,\alpha}=1$ into (8), as follows:

$$E\left[\left\{\hat{\delta}_{k,\alpha}\Big|_{a_{k,\alpha}=b_{k,\alpha}=1}\right\}^2\right] = \beta_\alpha\,\Gamma(\alpha+1)\,\{E[D_k^2]\}^{\alpha}. \tag{29}$$

In Fig. 1(a), the MSE functions of the two parametric estimators and the original estimator are plotted against the average speech energy $E[S_{k,\alpha}^2]$ and for several values of average noise energy $E[D_k^2]$. It can be seen that for a fixed noise energy $E[D_k^2]$, the MSE functions of both parametric estimators are always lower than that of the original estimator. Since a lower MSE generally associates with a lower residual noise, the residual noise level for the two parametric estimators is therefore lower than that of the original estimator. Furthermore, the reduction in MSE functions or residual noise level for the two parametric estimators is more significant when the speech energy $E[S_{k,\alpha}^2]$ is relatively low. This is important because when the speech energy is low, the residual noise is perceptually more distinct. The reduction is less for high speech energy $E[S_{k,\alpha}^2]$; this is acceptable since the residual noise is, in this case, well masked by speech.

From Fig. 1(a), the MSE function of the constrained parametric estimator is slightly higher than that of the unconstrained parametric estimator. However, in the results in Section III, it is observed that the residual noise level for the constrained parametric estimator is lower than that for the unconstrained estimator. This can be explained by examining their mean error (ME) functions. By taking expectation of the estimation error function in (7) and by using (17) and (20), the ME function of the constrained parametric estimator is

$$E\left[\hat{\delta}_{k,\alpha}\Big|_{a_{k,\alpha}=b_{k,\alpha}}\right] = \frac{1}{\xi_{k,\alpha}^{\alpha}+\beta_\alpha}\,\beta_\alpha\,\Gamma\!\left(\frac{\alpha}{2}+1\right)\cdot\{E[S_{k,\alpha}^2]\}^{\alpha/2}. \tag{30}$$

By using (7) and (15)–(17) for the unconstrained parametric estimator, and by using (2) and (7) for the original estimator, their ME functions can be shown to be zero: $E[\hat{\delta}_{k,\alpha}] = E[\hat{\delta}_{k,\alpha}|_{a_{k,\alpha}=b_{k,\alpha}=1}] = 0$.

For the constrained parametric estimator, $E[\hat{\delta}_{k,\alpha}|_{a_{k,\alpha}=b_{k,\alpha}}]$ in (30) is positive and therefore $E[S_{k,\alpha}^{\alpha}] > E[\{\hat{S}_{k,\alpha}|_{a_{k,\alpha}=b_{k,\alpha}}\}^{\alpha}]$. The constrained parametric estimator $\hat{S}_{k,\alpha}|_{a_{k,\alpha}=b_{k,\alpha}}$ therefore theoretically *underestimates* the speech spectral amplitude. This means that the estimator overattenuates the noisy speech. For the unconstrained parametric and original estimators, there is no underestimation since their ME functions are zero. The ME functions of the constrained parametric estimator normalized with the speech energy is shown in Fig. 1(b). It can be seen that the underestimation or overattenuation is only significant when the speech energy $E[S_{k,\alpha}^2]$ is low. This explains that the residual noise level for the constrained parametric estimator is lower than that for the unconstrained parametric estimator, even though its MSE function is slightly higher.

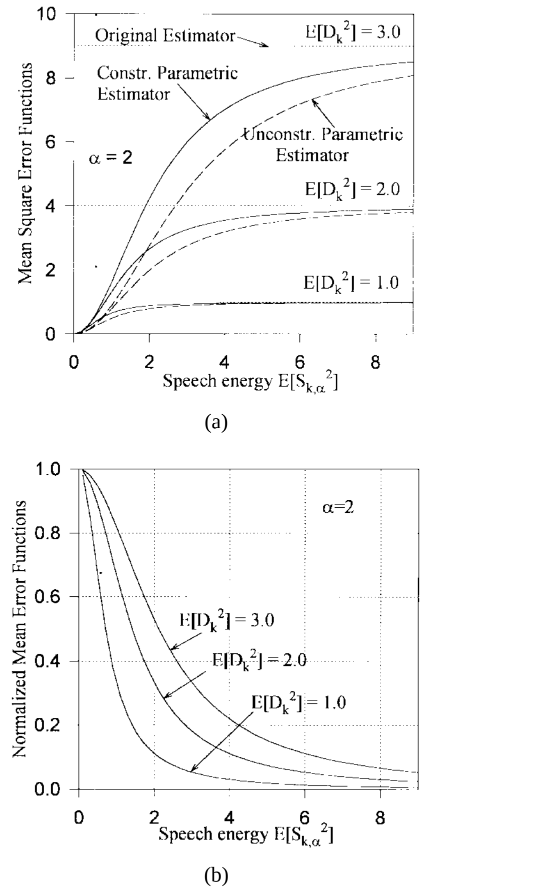

Fig. 1. (a) MSE functions of the original estimator and the unconstrained and constrained parametric estimators versus the average speech energy $E[S_{k,\alpha}^2]$. Three sets of error curves are shown here, each corresponding to a different noise variance or energy $E[D_k^2]$. (b) Normalized mean error function for the constrained parametric estimator versus the average speech energy $E[S_{k,\alpha}^2]$. The mean error function is normalized with the average speech energy $E[S_{k,\alpha}^2]$.

Note that for the constrained parametric estimator, the overattenuation under low speech energy also explains the use of the smoothed lower bound $\overline{\mu Y_k}$ in (25) to prevent overattenuation of the low speech energy segments, as described in Section II-E.

B. Analysis of the Spectral Gain Functions

For the two parametric estimators, the lower level of remaining noise and reduced musical tone artifacts can be further explained by examining their spectral gain functions.

The spectral gain functions of the original estimator $\tilde{S}_{k,\alpha}$ (2), the unconstrained parametric estimator $\hat{S}_{k,\alpha}$ (18) and the constrained parametric estimator $\hat{S}_{k,\alpha}|_{a_{k,\alpha}=b_{k,\alpha}}$ (22) are defined as in (31)–(33), shown at the bottom of the page, where $\gamma_k = \{Y_k^2 - E[D_k^2]\}/E[D_k^2]$ is the instantaneous *a posteriori* SNR as defined in [15]. γ_k can be interpreted as the *instantaneous* SNR since it is computed from the data in the current frame. $\xi_{k,\alpha}$ in (32) and (33) can be interpreted as the short-term *average* SNR since it is computed from the current and previous instantaneous SNR's [see (23)].

In Fig. 2(a) and (b), the gain functions $\tilde{G}_\alpha(\gamma_k)$, $\hat{G}_\alpha(\xi_{k,\alpha}, \gamma_k)$, and $\hat{G}_\alpha(\xi_{k,\alpha}, \gamma_k)|_{a_{k,\alpha}=b_{k,\alpha}}$ at $\alpha = 2$ are plotted against γ_k and/or several values of $\xi_{k,\alpha}$.

The single gain curve for $\tilde{G}_\alpha(\gamma_k)$ of the original estimator in Fig. 2(a) or (b) does not provide enough noise reduction as observed in the results in Section III. The characteristics of the two parametric gain functions $\hat{G}_\alpha(\xi_{k,\alpha}, \gamma_k)$ and $\hat{G}_\alpha(\xi_{k,\alpha}, \gamma_k)|_{a_{k,\alpha}=b_{k,\alpha}}$ that are responsible for an improved noise reduction can be better illustrated by dividing Fig. 2(a) and (b) into different regions corresponding to different combinations of γ_k and $\xi_{k,\alpha}$. This is shown in Figs. 3(a) and 4(a), respectively, for γ_k and $\xi_{k,\alpha}$ in the range of (-15 dB, 15 dB). The significance of the these regions can be more clearly illustrated by referring to the corresponding regions in the scatter plots in Figs. 3(b) and 4(b). The scatter plots were obtained using noisy speech (first sentence in the Appendix) with white Gaussian noise at a global SNR of 5 dB. The scatter plots are typical of all the other sentences. Each dot corresponds to a pair of ($\xi_{k,\alpha}$, γ_k) used by the unconstrained estimator (18) or the constrained estimator (22) in processing the noisy speech.

From the scatter plots, it can be seen that regions A and C are the major operating regions since more than 92% of the dots lie in these regions. Regions B and D are practically less important. Region B is a region of high $\xi_{k,\alpha}$ and low γ_k. As shown in the scatter plots, the probability of occurrence of this combination is small. By (deliberately) setting the noise-suppressed spectral amplitudes with gains in this region to zero, it was observed that the resulting speech was only slightly affected. Region D, below the "locus of minimum $\xi_{k,\alpha}$," is the region of no operation in that no combination of $(\xi_{k,\alpha}, \gamma_k)$ of the noisy speech exists in this region. The locus represents the variation of the minimum values of $\xi_{k,\alpha}$ as a function of γ_k. (Note that for other global SNR's, it can be verified that regions A and C are still the major operating regions.)

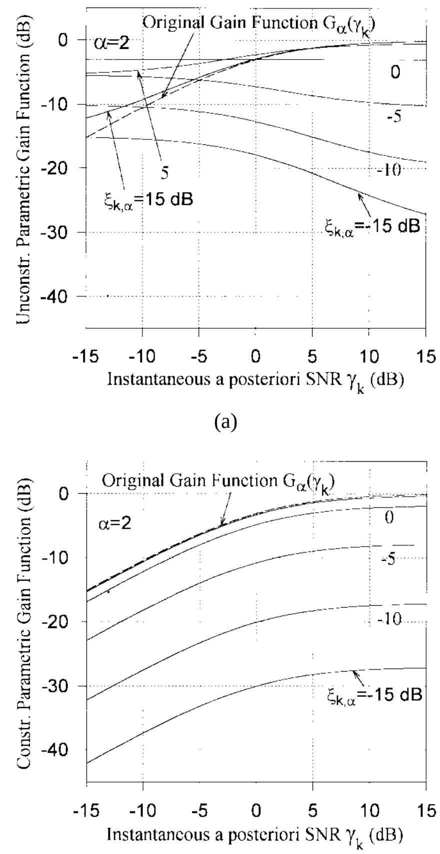

Fig. 2. (a) Gain curves of the original estimator and the unconstrained parametric estimator and (b) gain curves of the constrained parametric estimator, all at $\alpha = 2.0$.

1) Remaining Noise Level: For the two parametric estimators, the lower level of remaining noise can be explained by examining the spectral gain curves in the major operating regions.

Region A is the region of high speech energy, since the short-term average SNR $\xi_{k,\alpha}$ is relatively large ($\xi_{k,\alpha} \geq -5$

$$\begin{aligned}\tilde{G}_\alpha(\gamma_k) &= \tilde{S}_{k,\alpha}/Y_k \\ &= \left\{1 - \Gamma\left(\frac{\alpha}{2}+1\right)\left(\frac{1}{\gamma_k+1}\right)^{\alpha/2}\right\}^{1/\alpha}\end{aligned} \tag{31}$$

$$\begin{aligned}\hat{G}_\alpha(\xi_{k,\alpha}, \gamma_k) &= \hat{S}_{k,\alpha}/Y_k \\ &= \left\{\frac{\xi_{k,\alpha}^\alpha}{\xi_{k,\alpha}^\alpha+1}\right\}^{1/\alpha}\left\{1 - (1-\xi_{k,\alpha}^{-\alpha/2})\Gamma\left(\frac{\alpha}{2}+1\right)\left(\frac{1}{\gamma_k+1}\right)^{\alpha/2}\right\}^{1/\alpha}\end{aligned} \tag{32}$$

$$\begin{aligned}\hat{G}_\alpha(\xi_{k,\alpha}, \gamma_k)\Big|_{a_{k,\alpha}=b_{k,\alpha}} &= \hat{S}_{k,\alpha}\Big|_{a_{k,\alpha}=b_{k,\alpha}} \Big/ Y_k \\ &= \left\{\frac{\xi_{k,\alpha}^\alpha}{\xi_{k,\alpha}^\alpha+\beta_\alpha}\right\}^{1/\alpha}\left\{1 - \Gamma\left(\frac{\alpha}{2}+1\right)\left(\frac{1}{\gamma_k+1}\right)^{\alpha/2}\right\}^{1/\alpha}\end{aligned} \tag{33}$$

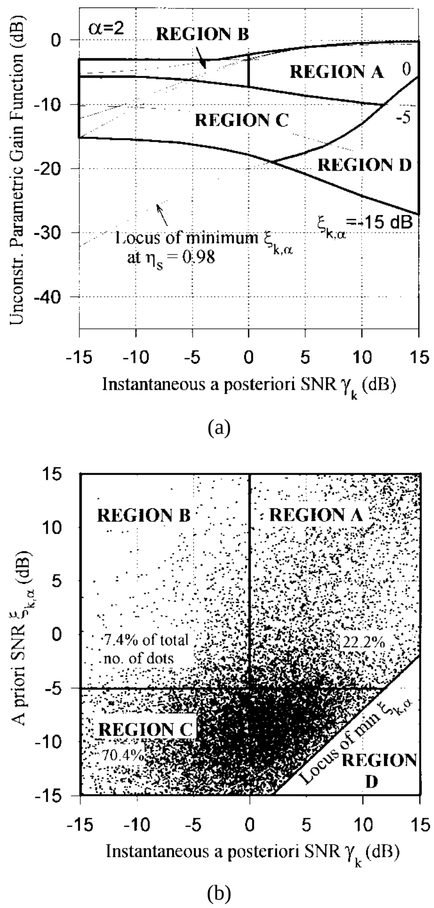

Fig. 3. (a) Four regions of the gain curves of the unconstrained parametric estimator at $\alpha = 2.0$ in Fig. 2(a). (b) Corresponding scatter plot. Regions A, B, C, and D contain 22.2%, 7.4%, 70.4%, and 0% of the total number of dots, respectively.

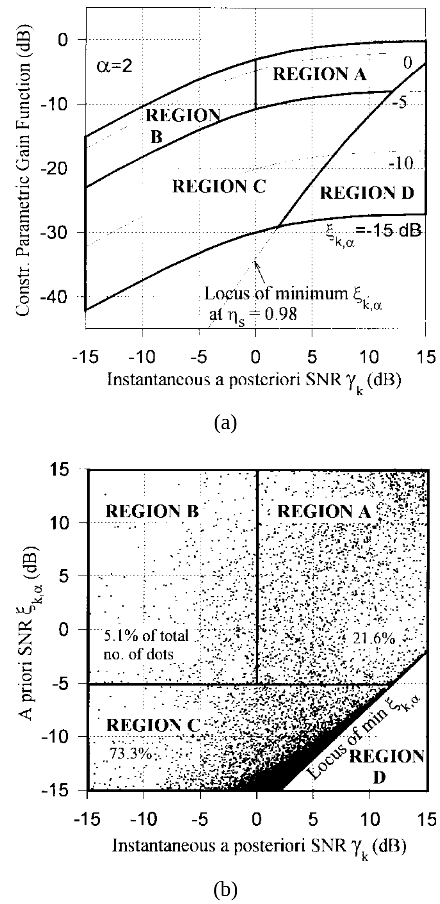

Fig. 4. (a) Four regions of the gain curves of the constrained parametric estimator at $\alpha = 2.0$ in Fig. 2(b). (b) Corresponding scatter plot.

dB). Note that the region boundary is arbitrarily chosen at $\xi_{k,\alpha} = -5$ dB for convenience of illustrations and discussions. The gain curves in this region in Figs. 3(a) and 4(a) show that the amount of additional attenuations provided by the two parametric gain functions are small and similar. In this region, the residual noise is perceptually less noticeable since it is well masked when the speech energy is high.

Region C is the region of low speech energy, since $\xi_{k,\alpha}$ is relatively small (≤ -5 dB). It corresponds to the region specified in the left part of Fig. 3 of Cappé [17]. In this region, the residual noise is perceptually more noticeable, since it is not well masked when the speech energy is low. The gain curves in this region show that the constrained parametric estimator provides more noise attenuation than the unconstrained estimator. The original estimator provides the least noise attenuation since it is independent of $\xi_{k,\alpha}$ and the four regions can be considered as "collapsed" into a single gain curve. This explains the observation that the constrained parametric estimator resulted in the lowest level of remaining noise, followed by the unconstrained estimator and then the original estimator.

To prevent the constrained parametric estimator from over-attenuation of low-energy speech, the smoothed lower bound $(\overline{\mu Y_k})$ in (25) restricts the attenuation level to about -14 to -26 dB corresponding to a spectral gain floor of $\mu = 0.20$ to 0.05, respectively.

2) Musical Tone Artifacts: The perceptually low amount of musical tone artifacts for the two parametric estimators can be explained by examining region C in Figs. 3(a) and 4(a).

For the unconstrained parametric estimator, at a fixed average SNR $\xi_{k,\alpha}$, the gain function of this region in Fig. 3(a) decreases (indicating higher attenuation) as the instantaneous SNR γ_k increases. The trend of the gain function is the same as that of the Ephraim–Malah estimator [15], [16]. Cappé [17] has explained (for the Ephraim–Malah estimator) that this type of gain function results in "protection from local-overtaking"—by using a lower gain when the instantaneous SNR γ_k ("local" SNR) becomes improbably higher than ("overtakes") the average SNR $\xi_{k,\alpha}$. This means that random spectral noise peaks, which have improbable high values of γ_k, are assigned an *increased* attenuation (lower gain). Thus, this has the effect of reducing the musical tone artifacts, since the musical tone artifacts are due to the presence of random spectral peaks of the residual noise.

It should be noted that the shape of the gain curves in Fig. 4(a) for the constrained parametric estimator is different. The explanations for the constrained parametric estimator are as follows. First, the constrained parametric estimator provides large attenuation in region C (region of low speech energy) and therefore the level of residual noise *including* the musical tone artifacts is low. Second, besides preventing overattenuation, the smoothed lower bound ($\overline{\mu Y_k}$) "fills up" the valleys between residual noise spectral peaks in the spectrum, and thus it has a perceptual masking effect on the musical tone artifacts. However, because of larger amount of attenuation, the musical tone artifacts for the constrained parametric estimator is perceptually lower than that for the unconstrained parametric estimator.

The average nature of $\xi_{k,\alpha}$ also helps to reduce the musical tone artifacts for the two parametric estimators. The additional attenuations provided by the parametric gain functions $\hat{G}_\alpha(\xi_{k,\alpha}, \gamma_k)$ and $\hat{G}_\alpha(\xi_{k,\alpha}, \gamma_k)|_{a_{k,\alpha}=b_{k,\alpha}}$ are dependent on $\xi_{k,\alpha}$, and therefore do not exhibit large variations over successive time frames. These large variations should be avoided because they generate random sinusoidal components appearing and disappearing over successive frames and would aggravate the musical tone artifacts. This can be observed by varying the smoothing constant η_S for $\xi_{k,\alpha}$ (23): *undersmoothing* of $\xi_{k,\alpha}$ ($\eta_S < 0.96$) results in insignificant reduction of musical tone artifacts, whereas *oversmoothing* ($\eta_S > 0.995$) results in almost complete reduction of musical tone artifacts but with some smearing of the speech signal. A balance point is reached with no apparent speech smearing but with some faint musical tone artifacts. This occurs somewhere at $\eta_S = 0.96$–0.995 for both parametric estimators, as described in Section III.

C. Comparison with Ephraim–Malah Estimator

For the two parametric estimators and the Ephraim–Malah estimator, the difference in the remaining noise level can be explained by considering their major operating regions (A and C). Note that the three estimators are dependent on both the instantaneous and average SNR's.

The gain curves of the Ephraim–Malah estimator are shown in Fig. 5. All the regions are partitioned in the same manner as described in the previous section. (By plotting out the scatter plot, it can be shown that regions A and C contain more than 92% of the operating dots.) An inspection of region A (region of high speech energy) in Figs. 3(a), 4(a), and 5, shows that the amount of attenuation of the three estimators are similar. Comparing region C (region of low speech energy) in Figs. 3(a), 4(a), and 5, for the same value of average SNR, the amount of attenuation for the constrained parametric estimator is the largest, followed by that for the Ephraim–Malah estimator and then the unconstrained parametric estimator. This explains that the remaining noise level for the unconstrained parametric estimator is the highest. For the constrained parametric estimator, the smoothed lower bound ($\overline{\mu Y_k}$) has the effect of raising the remaining noise level. At $\mu \approx 0.10$ corresponding to $\overline{\mu Y_k} \approx -20$ dB, subjective listening showed that the remaining noise levels for the constrained parametric estimator and the Ephraim–Malah estimator are perceptually similar.

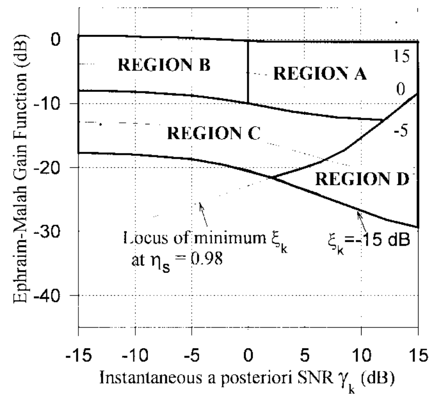

Fig. 5. Four regions of the gain curves of the Ephraim–Malah estimator.

For the Ephraim–Malah estimator, the explanation for the low amount of musical tone artifacts (see [17]) is similar to that for the unconstrained parametric estimator (described earlier). This is because the shapes of their gain curves in region C are similar. However, because of the larger amount of attenuation in region C, the amount of musical tone artifacts for the Ephraim–Malah estimator is perceptually lower than that for the unconstrained parametric estimator.

V. Conclusions

Two short-time spectral amplitude estimators of the speech signal are derived based on a parametric formulation of the original generalized spectral subtraction method and employing statistical optimization using an MMSE criterion. The estimators are derived by minimizing the mean-square of the error between the (ideal) generalized spectral subtraction spectrum and the parametric spectrum. The estimators are further simplified by using a Gaussian random assumption.

From the results of our tests, the two parametric estimators showed improved noise suppression performances under stationary white Gaussian noise and semistationary Jeep noise. Their implementations are computationally simple, and thus they are very suitable for real-time applications.

Future studies include the incorporation of a robust noise detection algorithm to determine nonspeech intervals so that the average noise spectrum can be updated during these intervals. Regular updating of the average noise spectrum is important especially for background noise that is nonstationary.

Appendix

A. Speech Material

For performance evaluation, the speech sentences used are as follows. All the sentences were spoken by different male (M) and female (F) speakers, and the sampling rate is 16 kHz.

1) She had your dark suit in greasy wash water all year (M).
2) His scalp was blistered from today's hot sun (M).
3) A leather handbag will be a suitable gift (M).
4) Ducks have webbed feet and colorful feathers (M).

TABLE I

WHITE GAUSSIAN NOISE: SNR IMPROVEMENT (dB) AND SPECTRAL DISTANCE IMPROVEMENT (%) OF THE VARIOUS NOISE SUPPRESSION SYSTEMS. CPARA-α–CONSTRAINED PARAMETRIC ESTIMATOR WITH POWER EXPONENT α; PARA-α—UNCONSTRAINED PARAMETRIC ESTIMATOR; GEN-α—THE ORIGINAL ESTIMATOR; BOLL—BY BOLL [10]; BER—METHOD BY BEROUTI *et al.* [9]; EM—EPHRAIM–MALAH ESTIMATOR [16]. NOTE THAT THE TWO BEST IMPROVEMENTS FOR EACH GLOBAL SNR ARE UNDERLINED

White Gaussian Noise								
Noisy Speech SNR (dB)	-5.0	0.0	5.0	10.0	-5.0	0.0	5.0	10.0
Noisy Speech Spect Dist	7.7	6.6	5.4	4.1	7.7	6.6	5.4	4.1
Noise Suppr. Method	SNR Improvement (dB)				Spect Dist Improvement (%)			
GEN-1.0	7.5	6.7	5.7	4.6	8.7	15.1	20.4	26.8
GEN-2.0	4.0	3.7	3.3	2.9	9.6	10.6	14.9	19.5
PARA-1.0	9.9	7.8	5.7	4.1	20.9	26.1	33.5	38.4
PARA-2.0	8.4	6.8	5.6	4.3	14.1	19.5	25.8	31.9
CPARA-1.0	10.3	8.4	6.3	4.6	23.4	29.4	37.8	41.1
CPARA-2.0	10.5	8.5	6.4	4.8	24.9	34.8	40.7	42.1
BOLL	9.6	6.9	4.5	1.3	9.8	16.7	27.8	34.1
BER	8.2	7.1	5.9	4.6	12.8	18.2	24.1	29.3
EM	10.5	8.5	6.3	4.3	22.7	31.4	38.5	43.3

TABLE II

JEEP NOISE. SEE TABLE I FOR EXPLANATIONS

Jeep Noise								
Noisy Speech SNR (dB)	-5.0	0.0	5.0	10.0	-5.0	0.0	5.0	10.0
Noisy Speech Spect Dist	5.4	4.3	3.3	2.3	5.4	4.3	3.3	2.3
Noise Suppr. Method	SNR Improvement (dB)				Spect Dist Improvement (%)			
GEN-1.0	8.7	8.3	7.9	7.2	19.9	24.9	27.5	30.0
GEN-2.0	4.8	4.7	4.6	4.4	20.6	18.9	21.4	25.8
PARA-1.0	12.5	10.8	9.5	8.2	25.6	25.9	32.4	33.5
PARA-2.0	9.8	8.4	7.8	6.8	20.6	30.3	29.7	32.2
CPARA-1.0	12.4	11.3	10.1	8.7	25.6	30.9	33.6	33.9
CPARA-2.0	12.4	11.4	10.2	8.9	25.4	30.3	32.7	34.8
BOLL	9.6	9.4	6.3	2.5	13.7	26.0	15.3	25.0
BER	8.1	7.7	7.1	6.3	15.0	21.9	26.1	30.0
EM	12.5	11.3	10.3	8.7	24.3	29.1	33.6	35.7

5) Gregory and Tom chose to watch cartoons in the afternoon (F).
6) Addition and subtraction are learned skills (F).
7) Mosquitoes exist in warm, humid climates (F).
8) The triumphant warrior exhibited naive heroism (F).

B. Objective Measures

The objective measures include improvements of the global SNR (see [8]) and the spectral distance (using Itakura's log likelihood ratio) after processing by the noise suppression system. The improvement in spectral distance in percent is defined as the improvement in spectral distance over the spectral distance of the noisy speech. The objective measures for the original estimator, the unconstrained and constrained parametric estimators at $\alpha = 1.0$ and 2.0, two modified spectral subtraction methods, and the Ephraim–Malah estimator are shown in Table I (for white Gaussian noise) and Table II (for Jeep noise). For the constrained parametric estimator, $\mu = 0.10$ is used. Note that the results are average scores over the 8 sentences described above. In the tables, the two best improvements for each global SNR are underlined.

ACKNOWLEDGMENT

This work was conducted at the Center For Signal Processing in Nanyang Technological University. The authors thank Prof. M. H. Er for his helpful comments in the first draft of this work.

REFERENCES

[1] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE,* vol. 67, pp. 1586–1604, Dec. 1979.
[2] J. S. Chang and Y. C. Tong, "A low power time-multiplexed switched capacitor speech spectrum analyzer," *IEEE J. Solid-State Circuits,* vol. 28, pp. 40–49, 1993.
[3] Y. C. Tong, R. C. Dowell, P. J. Blamey, and G. M. Clark, "Two component hearing sensations produced by two-electrode stimulation in the cochlea of a totally deaf patient," *Science,* vol. 219, pp. 993–994, 1983.
[4] Y. C. Tong, P. M. Seligman, G. M. Clark, J. F. Patrick, and J. B. Millar, "Speech processors," U.S. Patent 4 441 202, 1984.
[5] M. R. Weiss, E. Aschkenasy, and T. W. Parsons, "Study and development of the INTEL technique for improving speech intelligibility," Final Rep. NSC-FR/4023, Nicolet Scientific Corp., Dec. 1974.
[6] J. S. Lim, "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-26, pp. 471–472, Oct. 1978.
[7] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals.* New York: Maxwell Macmillian, 1993, p. 509.
[8] D. L. Wang and J. S. Lim, "The unimportance of phase in speech enhancement," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-30, pp. 679–681, Aug. 1982.
[9] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. ICASSP,* pp. 208–211, Apr. 1979.
[10] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-27, pp. 113–120, Apr. 1979.
[11] W. M. Kushner, G. Vladimir, C. Wu, V. Nguyen, and J. N. Damoulakis, "The effects of subtractive-type speech enhancement/noise reduction algorithms on parameter estimation for improved recognition and coding in high noise environments," in *Proc. ICASSP,* pp. 211–214, 1989.
[12] S. Vaseghi and R. Frayling-Cork, "Restoration of old gramaphone recordings," *J. Audio Eng.,* vol. 40, pp. 791–801, 1992.
[13] R. M. Crozier, B. M. G. Cheetham, C. Holt, and E. Munday, "Speech enhancement employing spectral subtraction and linear predictive analysis," *Electron. Lett.,* vol. 29, pp. 1094–1095, June 1993.
[14] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-28, pp. 137–145, Apr. 1980.
[15] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-32, pp. 1109–1121, Dec. 1984.
[16] ——, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-33, pp. 443–445, Apr. 1985.
[17] O. Cappé, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Trans. Speech Audio Processing,* vol. 2, pp. 345–349, Apr. 1994.
[18] R. E. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/synthesis," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-28, pp. 99–102, Feb. 1980.
[19] M. R. Spiegel, *Schaum's Outline of Theory and Problems of Mathematical Handbook of Formulas and Tables,* Int. ed. New York: McGraw-Hill, 1990.
[20] B. Picinbono, *Random Signals and Systems.* Englewood Cliffs, NJ: Prentice-Hall, 1993.
[21] G. S. Kang and L. J. Fransen, "Quality improvement of LPC-processed speech by using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 37, pp. 939–942, June 1989.

Boh Lim Sim received the B.E. degree from the National University of Singapore in 1992 and the M.E. degree from Nanyang Technological University, Singapore, Republic of Singapore, in 1996.

His research interests include real-time implementation of digital signal processing systems, acoustic noise reduction, and variable-rate speech playback.

Yit Chow Tong received the B.E. (first class honors) and Ph.D. degrees from the University of Melbourne, Australia, in 1972 and 1978, respectively.

He was a Principal Research Fellow of NH&MRC of Australia from 1988 to 1992. His research interests include signal processing, biomedical engineering, auditory physiology, biophysics, psychophysics, speech perception, and communication aids for the hearing impaired. He is currently the Associate Professor of the Electrical and Electronics Engineering Department at Nanyang Technological University, Singapore, Republic of Singapore. He is also a co-founder and director of Multitech Products Pte Ltd., a joint-venture company with Singapore Technologies.

Joseph S. Chang received the B.E. degree from Monash University, Australia, in 1983 and the Ph.D. degree from the Department of Otolaryngology, University of Melbourne, Australia, in 1990.

He worked for CSIRO and Texas Instruments from 1983 to 1985. From 1989 to 1991, he was a Senior Research Scientist at the Human Communication Research Center, University of Melbourne. He is now with the Nanyang Technological University, Singapore. He is also a co-founder and director of Multitech Products Pte Ltd., a company specializing in electroacoustics. His research interests include analog and digital signal processing, VLSI design, speech perception, psychophysics, biomedical engineering, and hearing-aid research. He holds several patents in circuit design and has several pending patents in biomedical instrumentation.

Chin Tuan Tan (M'89) received the B.E. degree from the National University of Singapore in 1992 and the M.E. degree from Nanyang Technological University, Singapore, in 1996. He is currently working toward the Ph.D. degree in Electrical and Electronics Engineering Department, Nanyang Technological University.

His research interests include signal processing, psychophysics, speech perception, speech recognition, auditory physiology, and hearing aids for the hearing impaired.