RESEARCH

Quadratic Approach for Single-channel Noise Reduction

Gal Itzhak[1*], Jacob Benesty[2] and Israel Cohen[1]

Abstract

In this paper, we introduce a quadratic approach for single-channel noise reduction. The desired signal magnitude is estimated by applying a linear filter to a modified version of the observations' vector. The modified version is constructed from a Kronecker product of the observations' vector with its complex conjugate. The estimated signal magnitude is multiplied by a complex exponential whose phase is obtained using a conventional linear filtering approach. We focus on the linear and quadratic maximum signal-to-noise ratio (SNR) filters, and demonstrate that the quadratic filter is superior in terms of subband SNR gains. In addition, in the context of speech enhancement, we show that the quadratic filter is ideally preferable in terms of perceptual evaluation of speech quality (PESQ) and short-time objective intelligibility (STOI) scores. The advantages, compared to the conventional linear filtering approach, are particularly significant for low input SNRs, at the expanse of a higher computational complexity. The results are verified in practical scenarios with nonstationary noise and in comparison to well-known speech enhancement methods. We demonstrate that the quadratic maximum SNR filter may be superior, depending on the nonstationary noise type.

Keywords: Quadratic filtering; Maximum SNR filter; Frequency-domain filtering; Optimal filters; Nonlinear processing; Kronecker product

1 Introduction

Communications and signal processing systems are very likely to operate in adverse environments, which are characterized by the presence of background noise that might severely degrade the quality of desired signals. Noise reduction methods are designed and applied to noisy signals with the objective of improving their quality and attenuating the background noise. Single-channel noise reduction (SCNR) methods are often implemented in physically small or low cost systems. SCNR filters are usually derived by minimizing a given distortion function between the clean signal and its estimate, or by minimizing the energy of the residual noise under some constraints.

Frequency-domain methods, e.g., [1–6], are typically formulated on a frame basis, that is, a frame of noisy observations is transformed into the frequency (or time-frequency) domain using the short-time Fourier transform (STFT). Then, the optimal filter is derived in the chosen domain and applied to the transformed observations. Finally, the filtered observations are transformed back to the time domain using the inverse STFT.

*Correspondence: sgalitz007@gmail.com
[1]Technion, Israel Institute of Technology, 32000 Haifa, Israel
Full list of author information is available at the end of the article

It is clear by construction that signals in the frequency domain are complex. Nonetheless, in many cases most of the information in a desired signal is stored in its spectral magnitude. Indeed, this property is well-known for speech signals, whose spectral magnitude has received special attention in the context of statistical models and optimal estimators, e.g., a maximum-likelihood spectral magnitude estimator [1], short-time spectral [2], log-spectral [3] and optimally modified log-spectral [7] magnitude estimators, and a maximum *a posteriori* spectral magnitude estimator [8]. These celebrated estimators assume that time trajectories in the STFT domain of clean speech and noise signals are independent complex Gaussian random processes. Other statistical models, e.g., super-Gaussian [9–11], Gamma [12, 13] or Laplace [14, 15] distributions were also investigated, and were demonstrated to be potentially more effective, depending on the desired speech spectral magnitude estimator and the speech conditional variance evolution model. While all the foregoing estimators rely on the strong correlation between magnitudes of successive coefficients (in a fixed frequency) [5, 16, 17], their derivation is typically cumbersome and requires one to numerically evaluate non-analytical functions following the assumed statistical speech and noise models. Moreover, with

the aforementioned spectral magnitude correlation hidden behind first-order recursive temporal processes, additional parameters and lower boundaries must carefully be set to guarantee the model tracking over time.

Recently, it has been proposed to exploit the self-correlation property of STFT domain coefficients in a linear manner. That is, instead of explicitly assuming statistical models which depend on unobserved measures, e.g., the *a priori* SNR, it was suggested to employ linear filters which require the second-order statistics of the desired signal and noise. These linear filters are derived within a multi-frame framework that takes into account the interframe correlation of the STFT coefficients from successive time frames and adjacent frequencies [5,18,19]. The multi-frame formulation highly resembles a sensor array formulation, which implies that conventional array filters may be modified for the single-channel case, but with an interframe correlation interpretation rather than spatial sensing. Examples of such filters are the Wiener filter, the minimum variance distortionless response (MVDR) filter [5,18], the linearly constrained minimum variance (LCMV) filter [5], and the maximum SNR filter [19].

In this paper, we present a quadratic approach for SCNR which extends the multi-frame approach suggested in [18]. The interframe correlation property is taken into account in the same manner as in [18], but the noise reduction filters are not applied to the observations' vector directly, but rather to its modified version. The modified version is obtained from the Kronecker product of the observations' vector and its complex conjugate. In its mathematical formulation, this approach is similar to the approach presented in [20] in the context of multichannel noise reduction. On the contrary, while in [20] the essence of the innovation is the direct utilization of higher-order statistics, the key idea in this work is a generalization of the single-channel linear filtering approach. We demonstrate that by focusing on the estimation of the desired signal magnitude in the transform domain, we are able to achieve further reduction of the background noise. More specifically, we propose the quadratic maximum SNR filter, which may potentially achieve a theoretically unbounded subband output SNR. We compare the quadratic and the linear maximum SNR filters and demonstrate that the quadratic filter is superior, in particular in low input SNR environments.

The rest of the paper is organized as follows. In Section 2, we present the signal model and formulate the SCNR problem. In Section 3, we introduce the quadratic filtering approach, from which quadratic filters may be derived. In Section 4, we propose a quadratic maximum SNR filter and derive it from two different perspectives. In Section 5, we focus on a toy example and theoretically evaluate the performances of the linear and quadratic maximum SNR filters. Finally, in Section 6, we demonstrate the noise reduction capabilities of the quadratic maximum SNR filter. We compare its performance to existing speech enhancement methods in ideal and practical conditions and in the presence of nonstationary noise.

2 Signal Model and Problem Formulation

We consider the classical single-channel noise reduction problem, where the noisy signal at time index t is given by [21,22]

$$y(t) = x(t) + v(t), \tag{1}$$

with $x(t)$ and $v(t)$ denoting the desired signal and additive noise, respectively. We assume that $x(t)$ and $v(t)$ are uncorrelated, and that all signals are real, zero mean, and broadband.

By employing the STFT or any other appropriate transform as suggested in [23], (1) can be rewritten in terms of the transform domain coefficients as

$$Y(k,n) = X(k,n) + V(k,n), \tag{2}$$

where the zero-mean complex random variables $Y(k,n)$, $X(k,n)$, and $V(k,n)$ are the analysis coefficients of $y(t)$, $x(t)$, and $v(t)$, respectively, at the frequency index $k \in \{0,1,\ldots,K-1\}$ and time-frame index n. It is well known that the same signal at different time frames is correlated [17]. Therefore, the interframe correlation should be taken into account in order to improve the performance of noise reduction algorithms. In this case, we may consider forming an observation signal vector of length N, containing the N most recent samples of $Y(k,n)$, i.e.,

$$\begin{aligned}\mathbf{y}(k,n) &= \left[\begin{array}{ccc} Y(k,n) & \cdots & Y(k,n-N+1) \end{array}\right]^T \\ &= \mathbf{x}(k,n) + \mathbf{v}(k,n),\end{aligned} \tag{3}$$

where the superscript ^{T} is the transpose operator, and $\mathbf{x}(k,n)$ and $\mathbf{v}(k,n)$ are defined similarly to $\mathbf{y}(k,n)$. Then, the objective of noise reduction is to estimate the desired signal $X(k,n)$ from the noisy observation signal vector $\mathbf{y}(k,n)$.

Since $x(t)$ and $v(t)$ are uncorrelated by assumption, the $N \times N$ correlation matrix of $\mathbf{y}(k,n)$ is

$$\begin{aligned}\mathbf{\Phi}_{\mathbf{y}}(k,n) &= E\left[\mathbf{y}(k,n)\mathbf{y}^H(k,n)\right] \\ &= \mathbf{\Phi}_{\mathbf{x}}(k,n) + \mathbf{\Phi}_{\mathbf{v}}(k,n),\end{aligned} \tag{4}$$

where the superscript ^{H} is the conjugate-transpose operator, and $\mathbf{\Phi}_{\mathbf{x}}(k,n)$ and $\mathbf{\Phi}_{\mathbf{v}}(k,n)$ are the correlation matrices of $\mathbf{x}(k,n)$ and $\mathbf{v}(k,n)$, respectively.

We end this part by defining the subband input SNR as

$$\mathrm{iSNR}(k,n) = \frac{\phi_X(k,n)}{\phi_V(k,n)}, \tag{5}$$

where $\phi_X(k,n) = E\left[|X(k,n)|^2\right]$ and $\phi_V(k,n) = E\left[|V(k,n)|^2\right]$ are the variances of $X(k,n)$ and $V(k,n)$, respectively.

3 Quadratic Filtering Approach

In the conventional linear approach [5], noise reduction is performed by applying a complex-valued filter, $\mathbf{h}(k,n)$ of length N, to the observation signal vector, $\mathbf{y}(k,n)$, i.e.,

$$\begin{aligned}\widehat{X}(k,n) &= \mathbf{h}^H(k,n)\mathbf{y}(k,n) \\ &= X_{\text{fd}}(k,n) + V_{\text{rn}}(k,n),\end{aligned} \tag{6}$$

where the filter output, $\widehat{X}(k,n)$, is an estimate of $X(k,n)$, $X_{\text{fd}}(k,n) = \mathbf{h}^H(k,n)\mathbf{x}(k,n)$ is the filtered desired signal, and $V_{\text{rn}}(k,n) = \mathbf{h}^H(k,n)\mathbf{v}(k,n)$ is the residual noise.

The two terms on the right-hand side of (6) are uncorrelated. Hence, the variance of $\widehat{X}(k,n)$ is

$$\begin{aligned}\phi_{\widehat{X}}(k,n) &= \mathbf{h}^H(k,n)\mathbf{\Phi}_{\mathbf{y}}(k,n)\mathbf{h}(k,n) \\ &= \phi_{X_{\text{fd}}}(k,n) + \phi_{V_{\text{rn}}}(k,n),\end{aligned} \tag{7}$$

where $\phi_{X_{\text{fd}}}(k,n) = \mathbf{h}^H(k,n)\mathbf{\Phi}_{\mathbf{x}}(k,n)\mathbf{h}(k,n)$ is the variance of the filtered desired signal and $\phi_{V_{\text{rn}}}(k,n) = \mathbf{h}^H(k,n)\mathbf{\Phi}_{\mathbf{v}}(k,n)\mathbf{h}(k,n)$ is the variance of the residual noise. Then, from (7), the subband output SNR is given by

$$\text{oSNR}\left[\mathbf{h}(k,n)\right] = \frac{\mathbf{h}^H(k,n)\mathbf{\Phi}_{\mathbf{x}}(k,n)\mathbf{h}(k,n)}{\mathbf{h}^H(k,n)\mathbf{\Phi}_{\mathbf{v}}(k,n)\mathbf{h}(k,n)}. \tag{8}$$

The quadratic filtering approach emerges from a different perspective. First, assuming that the desired signal is estimated with the linear approach, we find an expression for the energy of the estimated desired signal $\left|\widehat{X}(k,n)\right|^2$. We have

$$\begin{aligned}\left|\widehat{X}(k,n)\right|^2 &= \mathbf{h}^H(k,n)\mathbf{y}(k,n)\mathbf{y}^H(k,n)\mathbf{h}(k,n) \\ &= \text{tr}\left[\mathbf{y}(k,n)\mathbf{y}^H(k,n)\mathbf{h}(k,n)\mathbf{h}^H(k,n)\right] \\ &= \text{vec}^H\left[\mathbf{h}(k,n)\mathbf{h}^H(k,n)\right]\text{vec}\left[\mathbf{y}(k,n)\mathbf{y}^H(k,n)\right] \\ &= \left[\mathbf{h}^*(k,n)\otimes\mathbf{h}(k,n)\right]^H\left[\mathbf{y}^*(k,n)\otimes\mathbf{y}(k,n)\right] \\ &= \left[\mathbf{h}^*(k,n)\otimes\mathbf{h}(k,n)\right]^H\widetilde{\mathbf{y}}(k,n),\end{aligned} \tag{9}$$

where $\text{tr}[\cdot]$ is the trace of a square matrix, $\text{vec}[\cdot]$ is the vectorization operator, which consists of converting a matrix into a vector, $\otimes$ denotes the Kronecker product [24], and $\widetilde{\mathbf{y}}(k,n) = \mathbf{y}^*(k,n)\otimes\mathbf{y}(k,n)$ is a vector of length N^2.

Let $\widetilde{\mathbf{h}}(k,n)$ be a general complex-valued filter of length N^2, which is not necessarily of the form $\mathbf{h}^*(k,n)\otimes\mathbf{h}(k,n)$.

From (9) we can generate an estimate of $\left|\widehat{X}(k,n)\right|^2$ by applying the filter $\widetilde{\mathbf{h}}(k,n)$ to $\widetilde{\mathbf{y}}(k,n)$, i.e.,

$$Z(k,n) = \widetilde{\mathbf{h}}^H(k,n)\widetilde{\mathbf{y}}(k,n), \tag{10}$$

where $Z(k,n)$ is the estimate of the desired signal energy. Indeed, this approach generalizes the conventional linear approach, since (10) reduces to (9) with quadratic filters of the form $\widetilde{\mathbf{h}}(k,n) = \mathbf{h}^*(k,n)\otimes\mathbf{h}(k,n)$.

With $Z(k,n)$, we can obtain an estimate of the desired signal:

$$\widehat{X}(k,n) = e^{\jmath\psi(k,n)}\sqrt{|Z(k,n)|}, \tag{11}$$

where the phase $\psi(k,n)$ can be taken from the linear approach (6). We note that in practice this implies an additional computational complexity, as a linear filter might have to be implemented for the purpose of obtaining a desired signal phase estimate. Clearly, this approach is highly nonlinear.

Next, we would like to derive a theoretical expression for the subband output SNR with the quadratic approach. We have

$$\begin{aligned}\widetilde{\mathbf{y}}(k,n) &= \mathbf{y}^*(k,n)\otimes\mathbf{y}(k,n) \\ &= \left[\mathbf{x}^*(k,n)+\mathbf{v}^*(k,n)\right]\otimes\left[\mathbf{x}(k,n)+\mathbf{v}(k,n)\right] \\ &= \widetilde{\mathbf{x}}(k,n) + \mathbf{x}^*(k,n)\otimes\mathbf{v}(k,n) \\ &\quad + \mathbf{v}^*(k,n)\otimes\mathbf{x}(k,n) + \widetilde{\mathbf{v}}(k,n),\end{aligned} \tag{12}$$

where $\widetilde{\mathbf{x}}(k,n) = \mathbf{x}^*(k,n)\otimes\mathbf{x}(k,n)$ and $\widetilde{\mathbf{v}}(k,n) = \mathbf{v}^*(k,n)\otimes\mathbf{v}(k,n)$. Taking mathematical expectation on both sides of (12), we have

$$\begin{aligned}E\left[\widetilde{\mathbf{y}}(k,n)\right] &= E\left[\widetilde{\mathbf{x}}(k,n)\right] + E\left[\widetilde{\mathbf{v}}(k,n)\right] \\ &= \text{vec}\left[\mathbf{\Phi}_{\mathbf{x}}(k,n)\right] + \text{vec}\left[\mathbf{\Phi}_{\mathbf{v}}(k,n)\right] \\ &= \text{vec}\left[\mathbf{\Phi}_{\mathbf{y}}(k,n)\right].\end{aligned} \tag{13}$$

We deduce that

$$\begin{aligned}E\left[Z(k,n)\right] &= \widetilde{\mathbf{h}}^H(k,n)E\left[\widetilde{\mathbf{y}}(k,n)\right] \\ &= \widetilde{\mathbf{h}}^H(k,n)\text{vec}\left[\mathbf{\Phi}_{\mathbf{x}}(k,n)\right] \\ &\quad + \widetilde{\mathbf{h}}^H(k,n)\text{vec}\left[\mathbf{\Phi}_{\mathbf{v}}(k,n)\right].\end{aligned} \tag{14}$$

Consequently, the variance of $\widehat{X}(k,n)$ is

$$\begin{aligned}\phi_{\widehat{X}}(k,n) &= E\left[|Z(k,n)|\right] \\ &\approx \left|E\left[Z(k,n)\right]\right| \\ &= \left|\widetilde{\mathbf{h}}^H(k,n)E\left[\widetilde{\mathbf{y}}(k,n)\right]\right| \\ &= \Big|\widetilde{\mathbf{h}}^H(k,n)\text{vec}\left[\mathbf{\Phi}_{\mathbf{x}}(k,n)\right] \\ &\quad + \widetilde{\mathbf{h}}^H(k,n)\text{vec}\left[\mathbf{\Phi}_{\mathbf{v}}(k,n)\right]\Big|,\end{aligned} \tag{15}$$

where the approximation in the second row of (15) assumes $Z(k,n)$ to be real and positive. Thus, we can define the subband output SNR corresponding to a general quadratic filter $\widetilde{\mathbf{h}}(k,n)$ of length N^2 as

$$\mathrm{oSNR}\left[\widetilde{\mathbf{h}}(k,n)\right] = \frac{\left|\widetilde{\mathbf{h}}^H(k,n)\mathrm{vec}\left[\mathbf{\Phi}_{\mathbf{x}}(k,n)\right]\right|}{\left|\widetilde{\mathbf{h}}^H(k,n)\mathrm{vec}\left[\mathbf{\Phi}_{\mathbf{v}}(k,n)\right]\right|} \quad (16)$$
$$= \sqrt{\frac{\widetilde{\mathbf{h}}^H(k,n)\mathrm{vec}\left[\mathbf{\Phi}_{\mathbf{x}}(k,n)\right]\mathrm{vec}^H\left[\mathbf{\Phi}_{\mathbf{x}}(k,n)\right]\widetilde{\mathbf{h}}(k,n)}{\widetilde{\mathbf{h}}^H(k,n)\mathrm{vec}\left[\mathbf{\Phi}_{\mathbf{v}}(k,n)\right]\mathrm{vec}^H\left[\mathbf{\Phi}_{\mathbf{v}}(k,n)\right]\widetilde{\mathbf{h}}(k,n)}}.$$

In Sections 4 and 5, in order to simplify the notation we drop the dependence on the time and frequency indices. For example, (10) would be written as $Z = \widetilde{\mathbf{h}}^H\widetilde{\mathbf{y}}$.

4 Quadratic Maximum SNR Filter

In this section, we derive a filter $\widetilde{\mathbf{h}}$ that maximizes the output SNR given in (16). For theoretical completeness, the filter is derived from two different perspectives: by performing an eigenvalue decomposition to a rank deficient matrix defined by the noise statistics or by using an appropriate matrix projection operator.

The matrix $\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v}}\right)\mathrm{vec}^H\left[\mathbf{\Phi}_{\mathbf{v}}\right]$ may be diagonalized using the eigenvalue decomposition [25] as

$$\mathbf{U}^H\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v}}\right)\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{v}}\right)\mathbf{U} = \mathbf{\Lambda}, \quad (17)$$

where

$$\mathbf{U} = \left[\begin{array}{cc}\mathbf{u}_1 & \mathbf{U}'\end{array}\right] \quad (18)$$

is a unitary matrix, and

$$\mathbf{\Lambda} = \mathrm{diag}\left(\lambda_{\max}, 0, \ldots, 0\right) \quad (19)$$

is a diagonal matrix. The vector:

$$\mathbf{u}_1 = \frac{\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v}}\right)}{\sqrt{\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{v}}\right)\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v}}\right)}} \quad (20)$$

is the eigenvector corresponding to the only nonzero eigenvalue $\lambda_{\max} = \mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{v}}\right)\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v}}\right)$ of the matrix $\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v}}\right)\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{v}}\right)$, while $\mathbf{U}'$ contains the other N^2-1 eigenvectors of the zero eigenvalues. It is clear from (17) that

$$\mathbf{U}'^H\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v}}\right) = \mathbf{0}. \quad (21)$$

Now, let us consider filters of the form:

$$\widetilde{\mathbf{h}}_{\max} = \mathbf{U}'\widetilde{\mathbf{h}}'_{\max}, \quad (22)$$

where $\widetilde{\mathbf{h}}'_{\max} \neq \mathbf{0}$ is a filter of length N^2-1. Substituting (22) into (16), we infer that the subband output SNR with $\widetilde{\mathbf{h}}_{\max}$ may be unbounded, as opposed to the strictly bounded subband output SNR with the linear maximum SNR filter [19].

We point out the following observation. Despite achieving a potentially unbounded subband output SNR, the filter $\widetilde{\mathbf{h}}_{\max}$ is not expected to result in zero residual noise, as in practice it is applied to a vector of instantaneous analysis coefficients, while it is designed to eliminate the statistical noise PSD. Nonetheless, we recall that any linear filter may be extended to an appropriate quadratic filter but not vice versa. That is, the linear filtering approach may be regarded as a constrained version of the quadratic filtering approach. Hence, we deduce that the subband output SNR with the quadratic maximum SNR filter should be equal or larger than the subband output SNR with the linear maximum SNR filter.

With the subband output SNR maximized, it is possible to find $\widetilde{\mathbf{h}}'_{\max}$ in such a way that the desired signal distortion is minimized. Since the first term on the right-hand side of (14) corresponds to the filtered desired signal, we take this term equal to the variance of the desired signal, i.e.,

$$\widetilde{\mathbf{h}}^H_{\max}\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{x}}\right) = \phi_X. \quad (23)$$

Substituting (22) into (23) and noting that $\widetilde{\mathbf{h}}'_{\max}$ should equal the vector $\mathbf{U}'^H\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{x}}\right)$ up to appropriate scaling factors, we obtain

$$\widetilde{\mathbf{h}}'_{\max} = \frac{\mathbf{U}'^H\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{x}}\right)\phi_X}{\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{x}}\right)\mathbf{U}'\mathbf{U}'^H\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{x}}\right)}. \quad (24)$$

Therefore,

$$\widetilde{\mathbf{h}}_{\max} = \frac{\mathbf{U}'\mathbf{U}'^H\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{x}}\right)\phi_X}{\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{x}}\right)\mathbf{U}'\mathbf{U}'^H\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{x}}\right)}. \quad (25)$$

There is an alternative way to derive $\widetilde{\mathbf{h}}_{\max}$ from the first row of (16). That is, we may derive a filter $\widetilde{\mathbf{h}}_{\max,2}$ that is orthogonal to $\mathrm{vec}\left[\mathbf{\Phi}_{\mathbf{v}}\right]$, i.e., $\widetilde{\mathbf{h}}^H_{\max,2}\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v}}\right) = 0$. While the previous derivation of $\widetilde{\mathbf{h}}_{\max}$ may be considered more comparable to $\mathbf{h}_{\max}$ as both filters employ an eigenvalue decomposition, the alternative derivation of $\widetilde{\mathbf{h}}_{\max,2}$ may be more convenient to implement and analyze, and is indeed utilized for the thoretical performance analysis in Section 5. Any filter whose form is

$$\widetilde{\mathbf{h}}_{\max,2} = \widetilde{\mathbf{h}}'_{\max,2} - \frac{\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v}}\right)\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{v}}\right)}{\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{v}}\right)\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v}}\right)}\widetilde{\mathbf{h}}'_{\max,2}$$
$$= \mathbf{P}\widetilde{\mathbf{h}}'_{\max,2} \quad (26)$$

satisfies the condition, where $\widetilde{\mathbf{h}}'_{\max,2} \neq \mathbf{0}$ is an arbitrary complex-valued filter,

$$\mathbf{P} = \mathbf{I}_{N^2} - \frac{\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v}}\right)\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{v}}\right)}{\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{v}}\right)\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v}}\right)}, \tag{27}$$

and $\mathbf{I}_{N^2}$ is the identity matrix of size $N^2 \times N^2$.

Next, we wish to minimize the distortion, i.e., find $\widetilde{\mathbf{h}}_{\max,2}$ such that

$$\widetilde{\mathbf{h}}^H_{\max,2}\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{x}}\right) = \phi_X. \tag{28}$$

Substituting (26) into (28), we have

$$\widetilde{\mathbf{h}}'_{\max,2} = \frac{\mathbf{P}\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{x}}\right)\phi_X}{\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{x}}\right)\mathbf{P}\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{x}}\right)}. \tag{29}$$

Since $\mathbf{P}^2 = \mathbf{P}$, we have

$$\widetilde{\mathbf{h}}_{\max,2} = \widetilde{\mathbf{h}}'_{\max,2}. \tag{30}$$

Finally, by observing that $\mathbf{P} = \mathbf{U}'\mathbf{U}'^H$, we deduce that

$$\widetilde{\mathbf{h}}_{\max} = \widetilde{\mathbf{h}}_{\max,2}. \tag{31}$$

It should be noted that the formulation of (9) was already suggested in [20] in the context of multichannel noise reduction in the frequency domain. However, in this work the quadratic approach is applied to a single-channel observation vector in an arbitrary linear filtering domain, in which the interframe correlation is considered. Additionally, while the optimal filters suggested in [20] are designed to minimize the squared output energy and may be seen as the quadratic approach counterparts of the conventional MVDR and LCMV, this work provides a more general perspective to derive quadratic filters and proposes the quadratic maximum SNR filter $\widetilde{\mathbf{h}}_{\max}$ as a special case.

5 Performance Analysis

In this section, we analyze a toy example for which we derive the linear and quadratic maximum SNR filters. We theoretically evaluate and compare their corresponding subband SNR gains.

From Section 4, the theoretical subband SNR gain with the quadratic maximum SNR filter may be potentially unbounded. However, this would only be possible when the noise PSD matrix is precisely known. Since this assumption is never true in practice, it is important to analyze robustness to estimation errors in order to determine how practical the quadratic approach may be. Thus, our objective in this section is to evaluate the performance of the quadratic maximum SNR filter in the presence of estimation errors and compare it to the linear maximum SNR filter. This is done through a theoretical analysis of the following toy example in the STFT domain. Let us begin by assuming that the background noise is white and Gaussian, i.e., $v(t) \sim \mathcal{N}\left(0, \sigma_v^2\right)$. It can be shown that in the STFT domain with 50% overlapping rectangular analysis windows, the correlation matrix of the $N = 2$ element noise vector:

$$\mathbf{v}(k,n) = \left[\begin{array}{cc} V(k,n) & V(k,n-1) \end{array}\right]^T, \tag{32}$$

is given by

$$\mathbf{\Phi}_{\mathbf{v}} = \frac{N_{\mathrm{FFT}}\,\sigma_v^2}{2}\begin{bmatrix} 2 & (-1)^k \\ (-1)^k & 2 \end{bmatrix}, \tag{33}$$

where N_{FFT} is the number of FFT bins in a single frame. Next, we model the noise PSD matrix estimation errors as independent centralized complex Gaussian variables $\epsilon_{ij}, 1 \leq i,j \leq 2$, whose variance is denoted by σ_ϵ^2. Additionally, we use the notation $\sigma_V = N_{\mathrm{FFT}}\,\sigma_v^2/2$. Thus, the noise PSD matrix estimate with errors is given by

$$\mathbf{\Phi}_{\mathbf{v},\epsilon} = \sigma_V\begin{bmatrix} 2 & (-1)^k \\ (-1)^k & 2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{bmatrix}. \tag{34}$$

In order to derive the optimal filters, we also require the PSD matrix of the desired signal. Since our goal is to analyze the effect of the noise PSD matrix estimation errors, we assume for simplicity a fully coherent desired signal, that is

$$\mathbf{\Phi}_{\mathbf{x}} = \phi_X\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \tag{35}$$

The first step in deriving the quadratic maximum SNR filter $\widetilde{\mathbf{h}}_{\max}$ involves calculation of the projection operator. Following the simplified notation, we have

$$\mathbf{P} = \mathbf{I}_4 - \frac{\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v},\epsilon}\right)\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{v},\epsilon}\right)}{\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{v},\epsilon}\right)\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v},\epsilon}\right)}, \tag{36}$$

in which the matrix $\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v},\epsilon}\right)\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{v},\epsilon}\right)$ and the scalar $\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{v},\epsilon}\right)\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v},\epsilon}\right)$ should be computed. We have

$$\begin{aligned} &\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{v},\epsilon}\right)\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v},\epsilon}\right) = 10\sigma_V^2 \\ &+4\sigma_V\Re\{\epsilon_{11}+\epsilon_{22}\} + 2(-1)^k\sigma_V\Re\{\epsilon_{12}+\epsilon_{21}\} \\ &+|\epsilon_{11}|^2 + |\epsilon_{12}|^2 + |\epsilon_{21}|^2 + |\epsilon_{22}|^2, \end{aligned} \tag{37}$$

where $|\epsilon_{ij}|^2, 1 \leq i,j \leq 2$, are independent exponentially distributed random variables, that is, $|\epsilon_{ij}|^2 \sim \exp\left(1/2\sigma_\epsilon^2\right)$.

Next, we compute the elements of the 4×4 matrix $\mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{v},\epsilon}\right)\mathrm{vec}^H\left(\mathbf{\Phi}_{\mathbf{v},\epsilon}\right)$, by which we may approximate the expected value of $\mathbf{P}$, a key value required to approximate

the theoretical subband SNR gain. We have

$$E(\mathbf{P}) = \mathbf{I}_4 - E\left[\frac{\text{vec}(\mathbf{\Phi}_{\mathbf{v},\epsilon})\,\text{vec}^H(\mathbf{\Phi}_{\mathbf{v},\epsilon})}{\text{vec}(\mathbf{\Phi}_{\mathbf{v},\epsilon})^H\,\text{vec}(\mathbf{\Phi}_{\mathbf{v},\epsilon})}\right] \approx \mathbf{I}_4 - \frac{E\left[\text{vec}(\mathbf{\Phi}_{\mathbf{v},\epsilon})\,\text{vec}^H(\mathbf{\Phi}_{\mathbf{v},\epsilon})\right]}{E\left[\text{vec}^H(\mathbf{\Phi}_{\mathbf{v},\epsilon})\,\text{vec}(\mathbf{\Phi}_{\mathbf{v},\epsilon})\right]}, \tag{38}$$

where we used a first-order approximation [26]. Defining the error-to-noise ratio (ENR):

$$R_\epsilon = \frac{\sigma_\epsilon^2}{\sigma_V^2}, \tag{39}$$

we obtain

$$E(\mathbf{P}) \approx \frac{1}{2(4R_\epsilon+5)} \times \begin{bmatrix} 6(R_\epsilon+1) & 2(-1)^{k+1} & 2(-1)^{k+1} & -4 \\ 2(-1)^{k+1} & 6R_\epsilon+9 & -1 & 2(-1)^{k+1} \\ 2(-1)^{k+1} & -1 & 6R_\epsilon+9 & 2(-1)^{k+1} \\ -4 & 2(-1)^{k+1} & 2(-1)^{k+1} & 6(R_\epsilon+1) \end{bmatrix}. \tag{40}$$

Rewriting (15) to calculate the PSD of the estimated desired signal with the random filter $\widetilde{\mathbf{h}}$, we have

$$\begin{aligned} \phi_{\widehat{X}} &= E(|Z|) \\ &\approx |E(Z)| \\ &= \left|E\left[E\left[\widetilde{\mathbf{h}}^H\widetilde{\mathbf{y}}|\{\epsilon_{ij}\}\right]\right]\right| \\ &= \left|E\left(\widetilde{\mathbf{h}}^H\right)E(\widetilde{\mathbf{y}})\right| \\ &= \left|E\left(\widetilde{\mathbf{h}}^H\right)\text{vec}(\mathbf{\Phi}_{\mathbf{x}}) + E\left(\widetilde{\mathbf{h}}^H\right)\text{vec}(\mathbf{\Phi}_{\mathbf{v}})\right|, \end{aligned} \tag{41}$$

which implies that the subband output SNR is

$$\text{oSNR}\left(\widetilde{\mathbf{h}}\right) = \frac{\left|E\left(\widetilde{\mathbf{h}}^H\right)\text{vec}(\mathbf{\Phi}_{\mathbf{x}})\right|}{\left|E\left(\widetilde{\mathbf{h}}^H\right)\text{vec}(\mathbf{\Phi}_{\mathbf{v}})\right|}, \tag{42}$$

and its corresponding subband SNR gain is

$$\mathcal{G}\left(\widetilde{\mathbf{h}}\right) = \frac{\left|E\left(\widetilde{\mathbf{h}}^H\right)\text{vec}(\mathbf{\Phi}_{\mathbf{x}})\right|}{\left|E\left(\widetilde{\mathbf{h}}^H\right)\text{vec}(\mathbf{\Phi}_{\mathbf{v}})\right|} \times \frac{\phi_V}{\phi_X}. \tag{43}$$

Thus, in order to evaluate the subband SNR gain, we must first compute the expected value of the random filter $\widetilde{\mathbf{h}}_{\max}$. We have

$$\begin{aligned} E\left(\widetilde{\mathbf{h}}_{\max}\right) &= E\left[\frac{\mathbf{P}\text{vec}(\mathbf{\Phi}_{\mathbf{x}})\,\phi_X}{\text{vec}^H(\mathbf{\Phi}_{\mathbf{x}})\,\mathbf{P}\text{vec}(\mathbf{\Phi}_{\mathbf{x}})}\right] \\ &\approx \frac{E\left[\mathbf{P}\text{vec}(\mathbf{\Phi}_{\mathbf{x}})\,\phi_X\right]}{E\left[\text{vec}^H(\mathbf{\Phi}_{\mathbf{x}})\,\mathbf{P}\text{vec}(\mathbf{\Phi}_{\mathbf{x}})\right]} \\ &= \frac{E(\mathbf{P})\,\text{vec}(\mathbf{\Phi}_{\mathbf{x}})\,\phi_X}{\text{vec}^H(\mathbf{\Phi}_{\mathbf{x}})\,E(\mathbf{P})\,\text{vec}(\mathbf{\Phi}_{\mathbf{x}})}, \end{aligned} \tag{44}$$

where we used a first-order approximation in the second row of (44). Substituting (35) and (44) into (43), the subband SNR gain reduces to

$$\begin{aligned} &\mathcal{G}\left(\widetilde{\mathbf{h}}_{\max}\right) \\ &= \frac{\text{vec}^H(\mathbf{\Phi}_{\mathbf{x}})\,E(\mathbf{P})\,\text{vec}(\mathbf{\Phi}_{\mathbf{x}})}{\text{vec}^H(\mathbf{\Phi}_{\mathbf{x}})\,E(\mathbf{P})\,\text{vec}(\mathbf{\Phi}_{\mathbf{v}})} \\ &= \frac{1}{R_\epsilon}\frac{2\left[4(-1)^{k+1}+5\right]}{3\left[2+(-1)^k\right]} + \frac{4}{2+(-1)^k} + O\left(\frac{1}{R_\epsilon^2}\right). \end{aligned} \tag{45}$$

We deduce that when the ENR approaches zero, the theoretical subband SNR gain goes to infinity, and when the ENR is large the subband SNR gain is finite and frequency dependent.

The derivation of the linear filter $\mathbf{h}_{\max}$ of [19], which is used as a baseline for performance evaluation, begins by assessing the eigenvector corresponding to the maximum eigenvalue of the matrix $\mathbf{\Phi}_{\mathbf{v},\epsilon}^{-1}\mathbf{\Phi}_{\mathbf{x}}$. We have

$$\mathbf{\Phi}_{\mathbf{v},\epsilon}^{-1}\mathbf{\Phi}_{\mathbf{x}} = \frac{\phi_X}{|\mathbf{\Phi}_{\mathbf{v},\epsilon}|} \times \begin{bmatrix} \left[2+(-1)^{k+1}\right]\sigma_V+\epsilon_{22}-\epsilon_{12}, & \left[2+(-1)^{k+1}\right]\sigma_V+\epsilon_{22}-\epsilon_{12} \\ \left[2+(-1)^{k+1}\right]\sigma_V+\epsilon_{11}-\epsilon_{21}, & \left[2+(-1)^{k+1}\right]\sigma_V+\epsilon_{11}-\epsilon_{21} \end{bmatrix}, \tag{46}$$

whose eigenvalues are

$$\lambda_{\min} = 0, \tag{47}$$

$$\lambda_{\max} = \frac{\phi_X}{|\mathbf{\Phi}_{\mathbf{v},\epsilon}|} \times \left[2\sigma_V\left[2+(-1)^{k+1}\right] + \epsilon_{11} + \epsilon_{22} - \epsilon_{12} - \epsilon_{21}\right]. \tag{48}$$

It is easily verified that the (unnormalized) eigenvector $\mathbf{b}_{\max}$ that corresponds to $\lambda_{\max}$ is given by

$$\mathbf{b}_{\max} = \begin{bmatrix} \left[2+(-1)^{k+1}\right]\sigma_V+\epsilon_{22}-\epsilon_{12} \\ \left[2+(-1)^{k+1}\right]\sigma_V+\epsilon_{11}-\epsilon_{21} \end{bmatrix}, \tag{49}$$

which implies that

$$E\left(\mathbf{b}_{\max}\mathbf{b}_{\max}^H\right) = \begin{bmatrix} \left[2+(-1)^{k+1}\right]^2\sigma_V^2+4\sigma_\epsilon^2 & \left[2+(-1)^{k+1}\right]^2\sigma_V^2 \\ \left[2+(-1)^{k+1}\right]^2\sigma_V^2 & \left[2+(-1)^{k+1}\right]^2\sigma_V^2+4\sigma_\epsilon^2 \end{bmatrix}. \tag{50}$$

Formulating the PSD expression of the estimated desired signal with the random linear filter $\mathbf{h}_{\max}$ in a similar manner to (41), its subband SNR gain is

$$\mathcal{G}\left(\mathbf{h}_{\max}\right) = \frac{\phi_V}{\phi_X} \times \frac{E\left(\mathbf{h}_{\max}^H\right)\mathbf{\Phi}_{\mathbf{x}} E\left(\mathbf{h}_{\max}\right)}{E\left(\mathbf{h}_{\max}^H\right)\mathbf{\Phi}_{\mathbf{v}} E\left(\mathbf{h}_{\max}\right)}, \tag{51}$$

where the expected value of $\mathbf{h}_{\max}$ is given by

$$\begin{aligned} E\left(\mathbf{h}_{\max}\right) &= E\left(\frac{\mathbf{b}_{\max}\mathbf{b}_{\max}^H\mathbf{\Phi}_{\mathbf{x}}\mathbf{i}_1}{\mathbf{b}_{\max}^H\mathbf{\Phi}_{\mathbf{x}}\mathbf{b}_{\max}}\right) \\ &\approx \frac{E\left(\mathbf{b}_{\max}\mathbf{b}_{\max}^H\mathbf{\Phi}_{\mathbf{x}}\mathbf{i}_1\right)}{E\left(\mathbf{b}_{\max}^H\mathbf{\Phi}_{\mathbf{x}}\mathbf{b}_{\max}\right)} \\ &= \frac{E\left(\mathbf{b}_{\max}\mathbf{b}_{\max}^H\right)\mathbf{\Phi}_{\mathbf{x}}\mathbf{i}_1}{E\left(\mathbf{b}_{\max}^H\right)\mathbf{\Phi}_{\mathbf{x}} E\left(\mathbf{b}_{\max}\right)} \\ &= [0.5\ 0.5]^T, \end{aligned} \tag{52}$$

where we used a first-order approximation in the second row of (52). Substituting (35) and (52) into (51), the subband SNR gain is finally

$$\mathcal{G}\left(\mathbf{h}_{\max}\right) = \frac{4}{2+(-1)^k}, \tag{53}$$

which is ENR independent, but frequency dependent.

We infer that when the ENR is low, i.e., when the relative noise PSD estimation error is negligible, the quadratic approach achieves a highly preferable subband SNR gain. However, when the estimation error is in the same order of the noise energy, the two approaches exhibit a similar subband SNR gain. To illustrate the latter result we return to (5) in the limit of ENR that approaches infinity. We have

$$\lim_{R_\epsilon \to \infty} E\left(\mathbf{P}\right) \propto \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{54}$$

and hence

$$\lim_{R_\epsilon \to \infty} E\left(\widetilde{\mathbf{h}}_{\max}\right) = [0.25\ 0.25\ 0.25\ 0.25]^T. \tag{55}$$

This implies that in the high ENR limit, the quadratic max SNR filter converges to a version of the linear max SNR filter of (52), in which case both filters are simple averaging filters. While this result is explicitly derived for the toy example, we would expect such a behavior in any high ENR scenario in which the errors are modelled as normal identically distributed independent random variables. Additionally, we have

$$\lim_{R_\epsilon \to \infty} E\left(\widetilde{\mathbf{h}}_{\max}\right) = E\left(\mathbf{h}_{\max}\right) \otimes E\left(\mathbf{h}_{\max}\right), \tag{56}$$

which, by recalling (10) and the elaboration underneath, explains why in this limit the subband SNR gains are identical. The theoretical gain plots for odd and even values of k as a function of the ENR are illustrated in Fig. 1.

We end this part by addressing the computational complexity issue. On top of the additional complexity required with the quadratic maximum SNR filter in order to generate a desired signal phase estimate, the computational costs of the two filters are not straightforward to theoretically compare. That is, while deriving the quadratic maximum SNR filter typically requires matrix multiplications of a squared dimension, with the linear maximum SNR filter derivation a matrix inversion and an eigenvalue decomposition are computed. In practice, running the toy example with MATLAB software on an ordinary CPU takes 13 msec with the linear maximum SNR filter and 22 msec with the quadratic maximum SNR filter. Increasing the observation signal vector length to $N = 7$ yields a total runtime of 15 msec with the linear maximum SNR filter and 27 msec with the quadratic maximum SNR filter. Combining the runtime of both filters, we deduce that with a serial processor the quadratic maximum SNR filter requires about a three-time longer runtime than the linear maximum SNR filter in order to yield a desired signal amplitude and phase estimates.

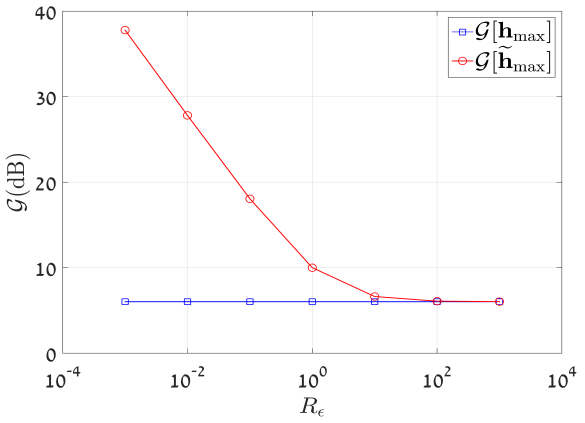
6 Experimental Results

In this section, we demonstrate the noise reduction capabilities of the quadratic maximum SNR filter in the context of speech enhancement. We perform extensive experiments in ideal and practical conditions, and compare its performance to well-known speech enhancement methods in stationary and nonstationary noise environments.

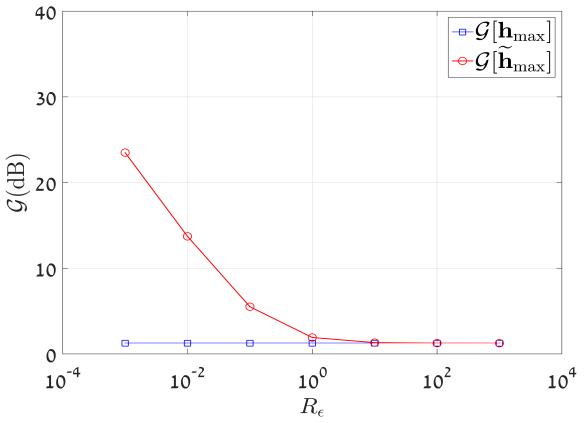
In the rest of the paper, for the sake of clarity, we return to explicit time and frequency indices notation.

6.1 Simulations in Ideal Conditions

We have shown that in the lack of estimation errors, the quadratic filter $\widetilde{\mathbf{h}}_{\max}(k,n)$ is designed to eliminate the residual noise, provided it is applied to the vector form of the additive noise correlation matrix. However, in practice, noise reduction filters are usually applied to instantaneous observation signal vectors, in which the noise term is of the form $\mathbf{v}^*(k,n)\otimes\mathbf{v}(k,n)$. Indeed, the latter may significantly differ from the statistical noise correlation matrix, which implies that the noise reduction performance might be far from optimal. It is therefore beneficial to employ a preliminary temporal smoothing step to the observation signal vector, and then apply the quadratic filtering approach to a



(a)



(b)

Figure 1: Theoretical subband SNR gain with the linear and quadratic maximum SNR filters as a function of the ENR for (a) odd k and (b) even k.

time-smoothed vector. Define

$$\begin{aligned}\widetilde{\mathbf{y}}_a(k,n;\tau_{\mathrm{y}}) &= \frac{1}{2\tau_{\mathrm{y}}+1}\sum_{n'=-\tau_{\mathrm{y}}}^{\tau_{\mathrm{y}}}\widetilde{\mathbf{y}}(k,n+n') \qquad (57)\\ &= \frac{1}{2\tau_{\mathrm{y}}+1}\sum_{n'=-\tau_{\mathrm{y}}}^{\tau_{\mathrm{y}}}\mathbf{y}^*(k,n+n')\otimes\mathbf{y}(k,n+n')\\ &= \widetilde{\mathbf{x}}_a(k,n;\tau_{\mathrm{y}})+\widetilde{\mathbf{v}}_a(k,n;\tau_{\mathrm{y}})\\ &+\frac{1}{2\tau_{\mathrm{y}}+1}\sum_{n'=-\tau_{\mathrm{y}}}^{\tau_{\mathrm{y}}}\{\mathbf{x}^*(k,n+n')\otimes\mathbf{v}(k,n+n')\\ &\qquad +\mathbf{v}^*(k,n+n')\otimes\mathbf{x}(k,n+n')\},\end{aligned}$$

where

$$\begin{aligned}&\widetilde{\mathbf{x}}_a(k,n;\tau_{\mathrm{y}}) \qquad (58)\\ &= \frac{1}{2\tau_{\mathrm{y}}+1}\sum_{n'=-\tau_{\mathrm{y}}}^{\tau_{\mathrm{y}}}\mathbf{x}^*(k,n+n')\otimes\mathbf{x}(k,n+n'),\end{aligned}$$

$$\begin{aligned}&\widetilde{\mathbf{v}}_a(k,n;\tau_{\mathrm{y}}) \qquad (59)\\ &= \frac{1}{2\tau_{\mathrm{y}}+1}\sum_{n'=-\tau_{\mathrm{y}}}^{\tau_{\mathrm{y}}}\mathbf{v}^*(k,n+n')\otimes\mathbf{v}(k,n+n'),\end{aligned}$$

and τ_{y} is the temporal smoothing preprocessing step parameter. We note that this implies a minor algorithmic delay of τ_{y} frames. Clearly, when the desired signal and noise are stationary and ergodic we should choose a high value for τ_{y}, as

$$E\left[\widetilde{\mathbf{y}}_a(k,n;\tau_{\mathrm{y}})\right] = E\left[\widetilde{\mathbf{y}}(k,n)\right], \qquad (60)$$

meaning that the temporal smoothing step does not distort the desired signal in terms of its second-order statistics. On the contrary, we have

$$\begin{aligned}&E\left[\widetilde{\mathbf{v}}_a^i(k,n;\tau_{\mathrm{y}})-\mathrm{vec}^i\left[\mathbf{\Phi}_{\mathbf{v}}(k,n)\right]\right]^2 \qquad (61)\\ &< E\left[\widetilde{\mathbf{v}}^i(k,n)-\mathrm{vec}^i\left[\mathbf{\Phi}_{\mathbf{v}}(k,n)\right]\right]^2,\end{aligned}$$

for every vector element $1 \le i \le N^2$, meaning the time-smoothed version of the noise observations' vector better resembles the theoretical noise PSD statistics than its instantaneous version. In addition, with the left-hand side of (61) being a monotonically decreasing function of τ_{y} we have

$$\begin{aligned}&\lim_{\tau_{\mathrm{y}}\to\infty}\left\{\frac{\left|\widetilde{\mathbf{h}}_{\max}^H(k,n)\widetilde{\mathbf{x}}_a(k,n;\tau_{\mathrm{y}})\right|}{\left|\widetilde{\mathbf{h}}_{\max}^H(k,n)\widetilde{\mathbf{v}}_a(k,n;\tau_{\mathrm{y}})\right|}\right\} \qquad (62)\\ &= \frac{\left|\widetilde{\mathbf{h}}_{\max}^H(k,n)\mathrm{vec}\left[\mathbf{\Phi}_{\mathbf{x}}(k,n)\right]\right|}{\left|\widetilde{\mathbf{h}}_{\max}^H(k,n)\mathrm{vec}\left[\mathbf{\Phi}_{\mathbf{v}}(k,n)\right]\right|}\\ &= \mathrm{oSNR}\left[\widetilde{\mathbf{h}}_{\max}(k,n)\right],\end{aligned}$$

which was previously shown to be potentially unbounded. On the contrary, for nonstationary desired signals there is an inherent trade-off in setting τ_{y}: as τ_{y} increases the mean-squared estimation error of the left-hand side of (61) decreases, resulting in a lower residual noise. However, by

further increasing τ_{y} the equality in (60) does not hold as the non stationary desired signal is smeared over time and hence distorted.

In order to demonstrate this trade-off, we consider a clean speech signal $x(t)$ that is sampled at a sampling rate of $f_s = 1/T_s = 16$ kHz within the signal duration T. The desired speech signal is formed by concatenating 24 speech signals (12 speech signals per gender) with varying dialects that are taken from the TIMIT database [27]. The clean speech signal is corrupted by an uncorrelated white Gaussian additive noise $v(t)$. The noisy signal is transformed into the STFT domain using 50% overlapping time frames and a Hamming analysis window of length 256 (16 msec). Next, it undergoes the foregoing temporal smoothing step, and then filtered by the two maximum SNR filters, i.e. the quadratic $\widetilde{\mathbf{h}}_{\max}(k,n)$ and the linear $\mathbf{h}_{\max}(k,n)$ of [19] to generate estimates of the desired speech signal. It is important to mention that both filters use the exact same desired speech and noise signal statistics estimates. As in this part we assume ideal conditions in which the desired speech and noise signals are known, their statistics are calculated by smoothing the corresponding signals over time. We want to compare the two approaches fairly. Hence, we allow a temporal smoothing preprocessing step for the conventional filter as well. However, we note that while with the quadratic filter $\widetilde{\mathbf{h}}_{\max}(k,n)$ the temporal smoothing step is employed over $\widetilde{\mathbf{y}}(k,n)$, with the linear $\mathbf{h}_{\max}(k,n)$ the smoothing is employed over $\mathbf{y}(k,n)$.

There is another modification that should be made with the quadratic approach in order to obtain a reliable desired signal estimation and keep the desired signal variance expression in (15) valid. While it is easy to show that with $\widetilde{\mathbf{h}}_{\max}(k,n)$ the expression in (10) is real, there is no guarantee that it is strictly positive. In practice, when a desired speech signal is present, it is very likely that the inner product is indeed positive, hence yielding a valid estimate of the desired signal spectral energy. This may be seen by applying the quadratic filter to the last equality of (12) in which the first term, that is associated with the true desired signal energy and the positive interframe correlation of adjacent time-frequency speech bins, is likely to be positive. Nevertheless, when a desired signal is absent, this positive term is approximately zero and the energy estimate may turn out negative. Clearly, such an estimate is non-physical and should be clipped to zero. Consequently, (10) is modified to

$$Z(k,n) = \max\left\{\widetilde{\mathbf{h}}^{H}_{\max}(k,n)\widetilde{\mathbf{y}}(k,n), 0\right\}. \tag{63}$$

Once the noise reduction procedure is completed, an inverse STFT transform is applied to yield the enhanced signals in the time domain. Then, it is possible to compute the PESQ [28] and STOI [29] scores, which function as a complementary performance measure to the subband SNR gain. We employ these scores to demonstrate the aforementioned trade-off in setting τ_{y} by computing them from the time-domain enhanced signals with the two maximum SNR filters. This simulation is carried out multiple times with varying values of τ_{y} with $N = 3$ and for time-domain input SNRs of -5 dB and 15 dB, where the time-domain input SNR is defined by

$$\mathrm{iSNR} = \frac{E\left[x^2(t)\right]}{E\left[v^2(t)\right]}. \tag{64}$$

The PESQ and STOI scores of the enhanced signals are shown in Fig. 2. We note that in this part the desired signal and noise are assumed to be known and are used to respectively generate their estimated statistics by performing a straightforward temporal smoothing. To begin with, it is clear that with the linear $\mathbf{h}_{\max}(k,n)$ for both time-domain input SNRs the optimal τ_{y} is zero. This is not surprising, of course, as the time-smoothed version of $\mathbf{y}(k,n)$ converges to zero according to the signal model assumption. On the contrary, while for the high input SNR a small value of τ_{y} should be used with $\widetilde{\mathbf{h}}_{\max}(k,n)$ (as the noise is very weak and the optimal filter should resemble the identity filter), for a low input SNR the convergence of the noise term $\widetilde{\mathbf{v}}_a(k,n;\tau_{\mathrm{y}})$ in $\widetilde{\mathbf{y}}_a(k,n;\tau_{\mathrm{y}})$ to the true noise correlation matrix is essential, and the optimal value of τ_{y} is found to be approximately 4. Clearly, when $\tau_{\mathrm{y}} \leq 4$ the approximation in (60) holds and the desired speech signal remains roughly distortionless. Thus, the mean-squared estimation error of the left hand side of (61) decreases as τ_{y} increases. However, we observe that while further increasing τ_{y}, i.e. when $\tau_{\mathrm{y}} > 4$, reduces the mean square estimation error of the noise, it also distorts the desired speech signal. Consequently, we infer that τ_{y} should be set to a value ranging $1-4$, with 1 being optimal for very high input SNRs and 3 or 4 being optimal for low input SNRs.

Next, in Fig. 3, we investigate the PESQ and STOI scores as a function of the input SNR for $N = 3$ and $N = 7$. We note that as a compromise between high and low input SNRs we fix $\tau_{\mathrm{y}} = 2$. We observe that in both cases the quadratic maximum SNR filter is preferable, in particular in low input SNRs where the noise reduction capabilities are stressed. As the input SNR increases, the linear and quadratic filters performances converge. This is intuitively explained as in the limit of zero additive noise the PESQ and STOI scores improvements should converge to zero and both the linear and quadratic filters should converge to a version of the identity filter. Nevertheless, we exhibit a minor STOI scores degradation in higher input SNRs. In essence, this is an artifact of the desired signal statistics estimation errors used to derive both the linear and the quadratic filters. That is, even with a stationary background noise we expect estimation errors to emerge due to the highly nonstationary nature of the speech signals. The

estimation errors inevitably result in some minor enhanced signal distortion which is more dominant in such scenarios. Finally, we note that the performance gap between the $N = 3$ and $N = 7$ cases, as exhibited in both filters, is a consequence of the stationary background noise. That is, we would not expect such a gap with an abruptly varying noise.

We return to the aforementioned subband SNR gain. In the STFT domain, it is convenient to average the subband input and output SNR expressions of (5), (8), and (16) over time, i.e.,

$$\overline{\mathrm{iSNR}}(k,:) = \frac{\sum_n \phi_X(k,n)}{\sum_n \phi_V(k,n)}, \tag{65}$$

$$\overline{\mathrm{oSNR}}\left[\mathbf{h}(k,:)\right] = \frac{\sum_n \mathbf{h}^H(k,n)\mathbf{\Phi}_{\mathbf{x}}(k,n)\mathbf{h}(k,n)}{\sum_n \mathbf{h}^H(k,n)\mathbf{\Phi}_{\mathbf{v}}(k,n)\mathbf{h}(k,n)}, \tag{66}$$

and

$$\overline{\mathrm{oSNR}}\left[\widetilde{\mathbf{h}}(k,:)\right] = \frac{\sum_n \left|\widetilde{\mathbf{h}}^H(k,n)\mathrm{vec}\left[\mathbf{\Phi}_{\mathbf{x}}(k,n)\right]\right|}{\sum_n \left|\widetilde{\mathbf{h}}^H(k,n)\mathrm{vec}\left[\mathbf{\Phi}_{\mathbf{v}}(k,n)\right]\right|}. \tag{67}$$

Consequently, the average subband SNR gains are given by

$$\overline{\mathcal{G}}\left[\mathbf{h}(k,:)\right] = \frac{\overline{\mathrm{oSNR}}\left[\mathbf{h}(k,:)\right]}{\overline{\mathrm{iSNR}}(k,:)} \tag{68}$$

and

$$\overline{\mathcal{G}}\left[\widetilde{\mathbf{h}}(k,:)\right] = \frac{\overline{\mathrm{oSNR}}\left[\widetilde{\mathbf{h}}(k,:)\right]}{\overline{\mathrm{iSNR}}(k,:)}, \tag{69}$$

respectively.

We use expressions (68) and (69), respectively, to compare $\widetilde{\mathbf{h}}_{\max}(k,n)$ and $\mathbf{h}_{\max}(k,n)$ in terms of the average subband SNR gain. The results for $\mathrm{iSNR} = 0$ dB and for $N = 3$ and 7 are depicted in Fig. 4. According to the analysis above we set $\tau_{\mathrm{y}} = 2$ with the quadratic maximum SNR filter, which is shown to result in a significantly preferable gain. This is true for both values of N. Moreover, as it is observed in Fig. 4 and in a similar fashion to the previously discussed average PESQ and STOI scores, the performance of the linear maximum SNR filter with $N = 7$ is somewhat close to the performance of the quadratic maximum SNR filter with $N = 3$. That is, the quadratic filter is demonstrated to better utilize a given noisy observation signals vector from the subband SNR gain perspective.

6.2 Experiments in Practical Scenarios

Next, we are interested in comparing the two approaches in practical scenarios and with nonstationary noise. Four scenarios are simulated with the additive noise signal being either a stationary white Gaussian noise or one of the following three nonstationary noise types: a motor crank noise, a wind noise or a traffic noise. The TIMIT set of clean desired speech signals is maintained. We set $\mathrm{iSNR} = 0$ dB and analyze the PESQ and STOI scores with the following six methods: two practical versions of the linear and quadratic maximum SNR filters, their two ideal versions (as presented in the previous part), the celebrated log-spectral amplitude estimator (LSA) [3] and the spectral subtraction in the short-time modulation domain (STSS) of [30]. We set $N = 3$ for all four maximum SNR filters, and perform the STFT transform with the same analysis window and overlap factor in all methods except the STSS. The STSS is employed in its default parameters as defined by the authors of [30], with acoustic and modulation frame lengths and overlap factors of $32\mathrm{msec}$ and 75%, and $256\mathrm{msec}$ and 87.5%, respectively. According to the previous part, we fix $\tau_{\mathrm{y}} = 2$ with $\widetilde{\mathbf{h}}_{\max}(k,n)$, whereas no smoothing is performed with $\mathbf{h}_{\max}(k,n)$.

The practical versions of the linear and quadratic maximum SNR filters, denoted, respectively, by $\mathbf{h}_{\max,\mathrm{prac}}(k,n)$ and $\widetilde{\mathbf{h}}_{\max,\mathrm{prac}}(k,n)$, require estimates of the desired speech and noise correlation matrices to be computed out of the noisy observations. In this experiment, we employ a somewhat naive estimation approach that is inspired by [31] and leave more sophisticated schemes for future research. The noisy observations correlations matrix is updated over time by a first order recursive temporal smoothing

$$\begin{aligned}\mathbf{\Phi}_{\mathbf{y}}(k,n) &= \lambda\mathbf{\Phi}_{\mathbf{y}}(k,n-1) \\ &\quad + (1-\lambda)\,\mathbf{y}(k,n)\mathbf{y}^H(k,n),\end{aligned} \tag{70}$$

with $0 < \lambda < 1$ being the smoothing parameter. We found $\lambda = 0.5$ to be an optimal choice to cope with both stationary and quickly-varying nonstationary noise. Then, the noise correlation matrix is given by

$$\mathbf{\Phi}_{\mathbf{v}}(k,n) = \min\{\mathbf{\Phi}_{\mathbf{v}}(k,n-1), \mathbf{\Phi}_{\mathbf{y}}(k,n)\}\,(1+\epsilon)\,, \tag{71}$$

with ϵ set to yield a power increase of 5 dB/s. Finally, the desired signal correlation matrix is estimated by

$$\mathbf{\Phi}_{\mathbf{x}}(k,n) = \max\{\mathbf{\Phi}_{\mathbf{y}}(k,n) - \mathbf{\Phi}_{\mathbf{v}}(k,n), 0\}. \tag{72}$$

We note the following. To begin with, the minimum and maximum operations above are considered element-wise, whereas the first 100 frames are used to generate an initial noise correlation matrix estimate, i.e., the first 808 msec are assumed to be silent. In addition, we verify that $\mathbf{\Phi}_{\mathbf{x}}(k,n)$ is

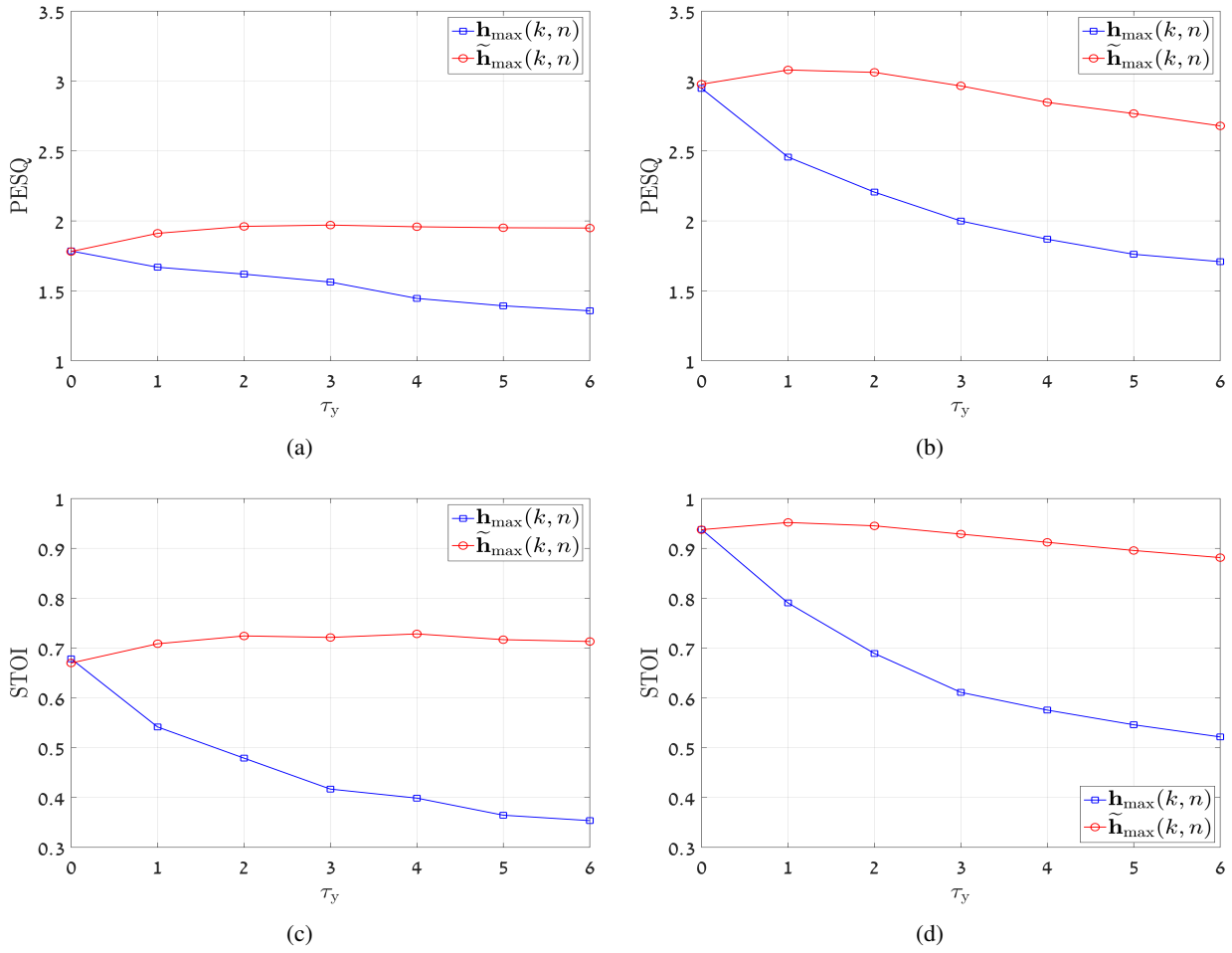


Figure 2: PESQ and STOI scores of TIMIT speech signals as a function of the temporal smoothing preprocessing step parameter τ_y for $N = 3$ in the presence of white Gaussian noise: (a) PESQ scores with iSNR $= -5$ dB, (b) PESQ scores with iSNR $= 15$ dB, (c) STOI scores with iSNR $= -5$ dB, and (d) STOI scores with iSNR $= 15$ dB. The PESQ scores of the input noisy observation signal are 1.47 and 2.78 with iSNR $= -5$ dB and iSNR $= 15$ dB, respectively; their corresponding STOI scores are 0.61 and 0.97.

obtained as a positive-definite matrix, which is the case in practically all the simulations we have performed. Finally, the presented correlation matrices estimation approach requires setting the optimal values of additional parameters in a similar manner to traditional approaches as described in Section 1.

The experimental results in terms of the average PESQ and STOI scores with their respective confidence (standard deviation) intervals computed over 24 speech utterances are described in Fig. 5. To begin with, we observe that in terms of PESQ scores, the ideal quadratic maximum SNR filter performs significantly better than the other methods in the three nonstationary noise scenarios, whereas it is slightly inferior to the STSS in the white noise scenario. In addition, the ideal quadratic maximum SNR filter is highly superior in terms of STOI scores in all the examined scenarios. In particular, the ideal quadratic maximum SNR filter outperforms its linear counterpart, which implies that the former's potential is preferable. Analyzing the practical versions of the maximum SNR filters, we note that in

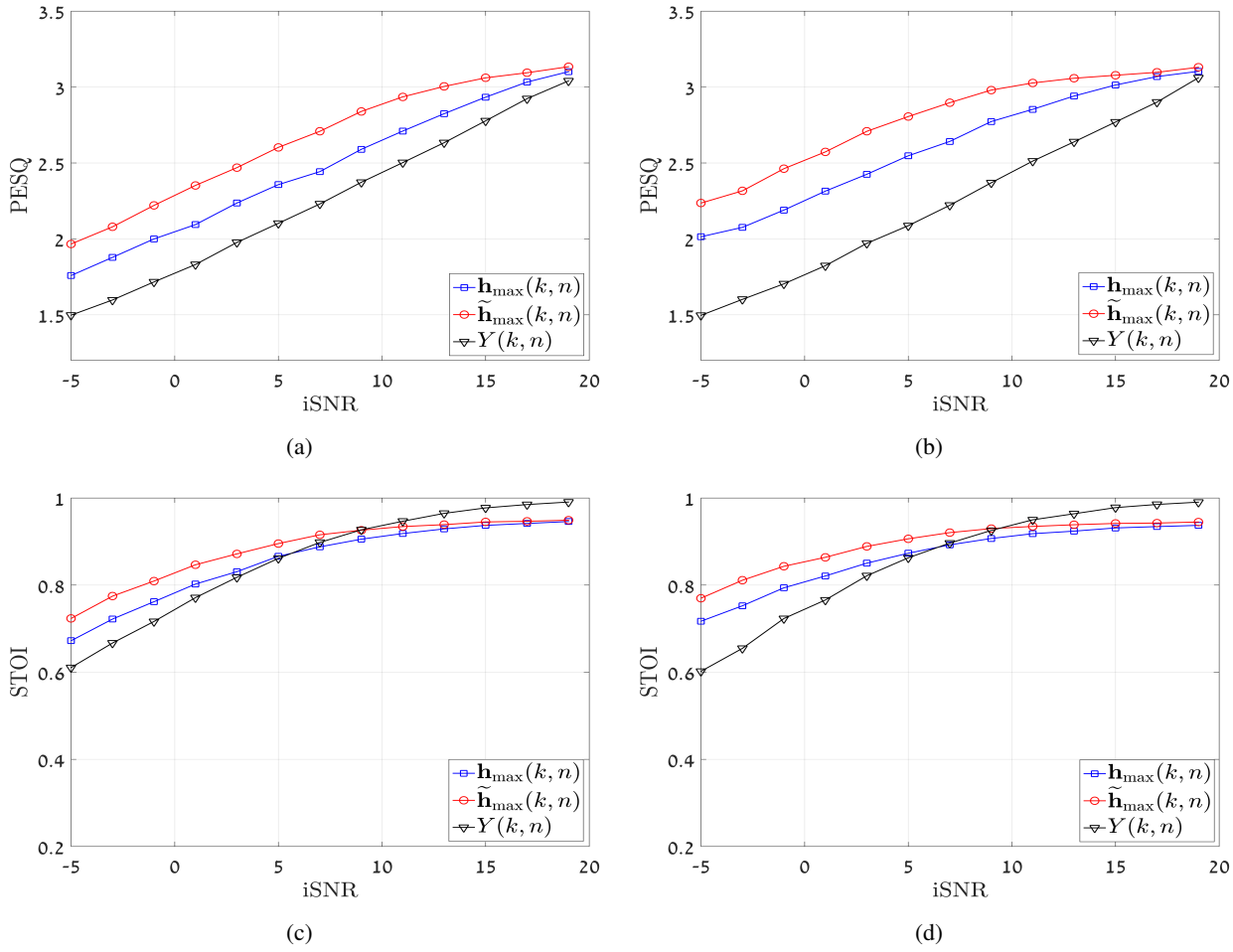


Figure 3: PESQ and STOI scores of TIMIT speech signals as a function of the iSNR for $N = 3$ and $N = 7$ in the presence of white Gaussian noise. (a) PESQ scores with $N = 3$, (b) PESQ scores with $N = 7$, (c) STOI scores with $N = 3$, and (d) STOI scores with $N = 7$. We set $\tau_{\mathrm{y}} = 2$ for the quadratic filter $\widetilde{\mathbf{h}}_{\max}(k,n)$.

general the quadratic filter is superior to the linear filter in terms of PESQ scors, whereas in terms of STOI scores the performances are overall roughly equal. A comparison to the LSA and the STSS indicates that both are significantly inferior to the practical quadratic maximum SNR filter in the motor crank noise and wind noise scenarios. On the contrary, in the white noise and traffic noise scenarios the performance gap is opposite, with the LSA and the STSS performing better than the practical quadratic maximum SNR filter, which is however preferable to the practical linear maximum SNR filter. The performances difference between noise types for the different methods is resulted in by the nature of the noise signals and the method we used to estimate and track their statistics. For example, this could be due to their level of non-stationarity, i.e., the coherence time during which the statistics of the noise remain roughly unchanged. We deduce that the quadratic maximum SNR filter is ideally of a high potential and may also be successfully applied in practice, even with naive desired signal and noise statistics estimation techniques.

We end this part by relating an informal listening experiment we conducted to verify the foregoing results. This included extensive comparisons between enhanced signals with all the presented methods in the different noise scenarios. While no musical noise nor reverberation effects were detected with any of the methods, their distinctive natures were observable. That is, while it was apparent that the four maximum SNR filters preserved the desired sig-

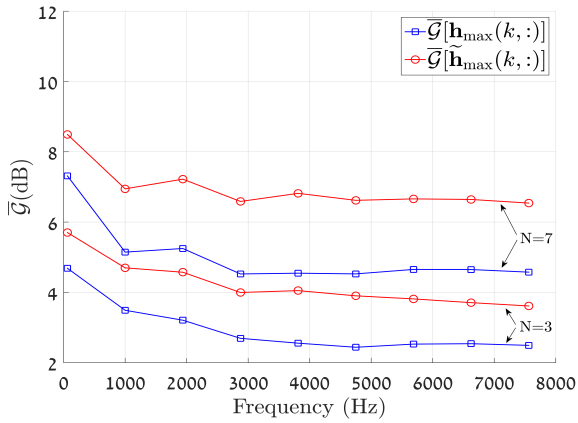


Figure 4: Average subband SNR gains of TIMIT speech signals for $\mathrm{iSNR} = 0$ dB and for $N = 3$ and $N = 7$, with white Gaussian noise. We set $\tau_{\mathrm{y}} = 2$ for the quadratic filter $\widetilde{\mathbf{h}}_{\max}(k,n)$.

nals distortionless, the noise reduction capabilities of their two practical versions were relatively limited with respect to the LSA and STSS, which featured less residual noise in the white noise and traffic noise scenarios. On the contrary, the LSA and STSS did exhibit some desired signal distortion in most cases, particularly in frequencies higher than 3 kHz. This was more stressed in the motor crank noise and the wind noise scenarios, in which their respective residual noise was significant. Considering the ideal versions of the linear and quadratic maximum SNR filters, the enhanced signals they yielded sounded considerably clearer than all other methods, with the ideal quadratic maximum SNR filter being superior to its linear counterpart particularly in the white noise and the traffic noise scenarios.

7 Conclusions

We have presented a quadratic filtering approach for single-channel noise reduction, which generalizes the conventional linear filtering approach. The advantage of the quadratic approach was demonstrated by focusing on the maximum SNR filter in the STFT domain. We have analyzed the theoretical subband SNR gain in a toy example and showed that while with the linear maximum SNR filter the subband SNR gain is strictly bounded, with the quadratic maximum SNR filter the gain is potentially unbounded and heavily depends on the ENR. We have proposed the temporal smoothing preprocessing step and verified the performance on speech signals. In ideal and practical conditions, the quadratic maximum SNR filter was compared to the linear maximum SNR filter and to two well-known speech enhancement methods in both stationary and nonstationary noise environments. We have demonstrated that the quadratic maximum SNR filter outperforms the linear maximum SNR filter, in particular in low input SNRs, at the expanse of a higher computational complexity. In addition, the former was shown to perform better than commonly-used methods in practice in some of the scenarios we examined, even with naive desired signal and noise statistics estimation techniques, whereas in other scenarios the performance gap was the opposite. In future work, we may improve these estimation techniques to reach closer to the performance of the ideal quadratic maximum SNR filter, and possibly estimate the desired signal phase directly, i.e., not through a separate linear filter.

Abbreviations

SNR: signal-to-noise ratio; PESQ: perceptual evaluation of speech quality; SCNR: single-channel noise reduction; STFT: short-time Fourier transform; HMM: hidden Markov model; MVDR: minimum variance distortionless response; LCMV: linearly constrained minimum variance

Acknowledgements

The authors thank the associate editor and the anonymous reviewers for their constructive comments and useful suggestions.

Funding

This research was supported by the Israel Science Foundation (grant No. 576/16), and the ISF-NSFC joint research program (grant No. 2514/17).

Availability of data and materials

Please contact author for data requests.

Authors' contributions

The authors' contributions are equal.

Competing interests

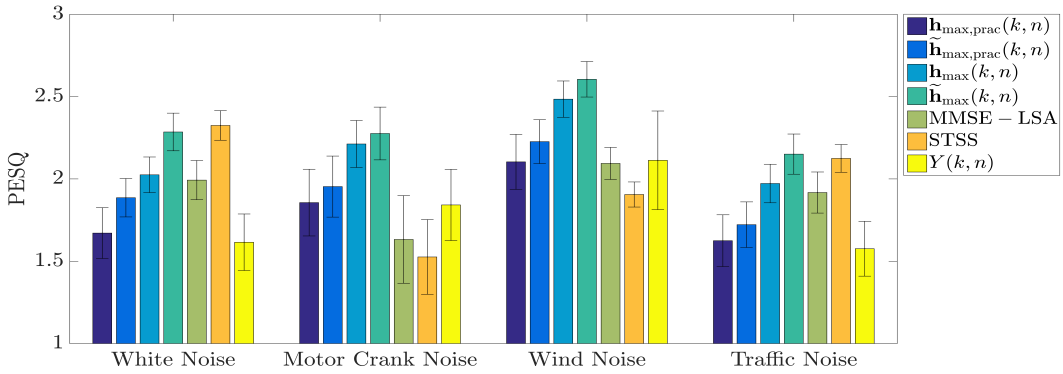
The authors declare that they have no competing interests.

Author details

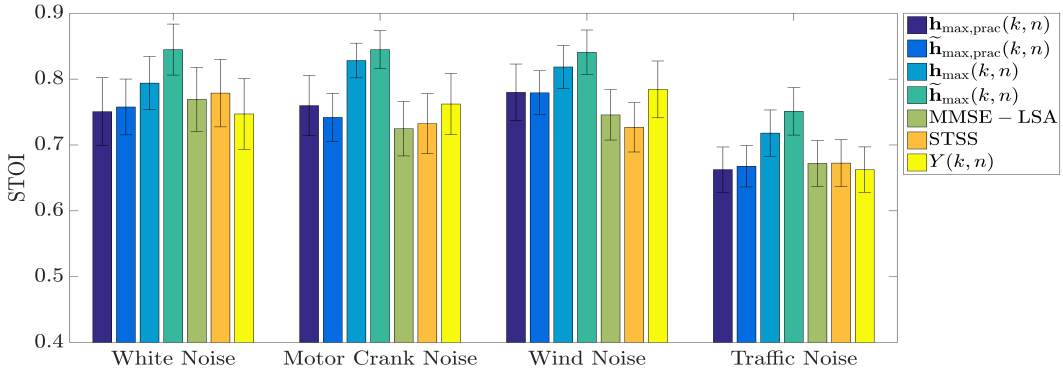
[1]Technion, Israel Institute of Technology, 32000 Haifa, Israel. [2]INRS-EMT, University of Quebec, 800 de la Gauchetiere Ouest, Suite 6900, QC H5A 1K6 Montreal, Canada.

References

1. McAulay, R., Malpass, M.: Speech enhancement using a soft-decision noise suppression filter. IEEE Transactions on Acoustics, Speech, and Signal Processing **28**(2), 137–145 (1980)
2. Ephraim, Y., Malah, D.: Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Transactions on Acoustics, Speech, and Signal Processing **32**(6), 1109–1121 (1984)
3. Ephraim, Y., Malah, D.: Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Transactions on Acoustics, Speech, and Signal Processing **33**(2), 443–445 (1985)
4. Cohen, I., Gannot, S.: In: Benesty, J., Sondhi, M., Huang, Y. (eds.) Spectral Enhancement Methods, pp. 873–902. Springer, Berlin, Heidelberg (2008)
5. Benesty, J., Chen, J., Habets, E.: Speech Enhancement in the STFT Domain. Springer-Verlag Berlin Heidelberg, Berlin (2012)
6. Benesty, J., Cohen, I., Chen, J.: Fundamentals of Signal Enhancement and Array Signal Processing. Wiley-IEEE Press, Singapore (2018)
7. Cohen, I.: Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator. IEEE Signal Processing Letters **9**(4), 113–116 (2002)
8. Wolfe, P.J., Godsill, J.S.: Efficient Alternatives to the Ephraim and Malah Suppression Rule for Audio Signal Enhancement. EURASIP Journal on Advances in Signal Processing **2003**(10) (2003)
9. Martin, R.: Speech enhancement based on minimum mean-square error estimation and supergaussian priors. Speech and Audio Processing, IEEE Transactions on **13**, 845–856 (2005). doi:10.1109/TSA.2005.851927



(a)



(b)

Figure 5: A comparison of average PESQ and STOI scores of TIMIT speech signals in practical scenarios with $\mathrm{iSNR} = 0$. The vertical margins exhibit the confidence (standard deviation) intervals computed over 24 speech utterances. $\mathbf{h}_{\max,\mathrm{prac}}(k,n)$ and $\widetilde{\mathbf{h}}_{\max,\mathrm{prac}}(k,n)$ denote, respectively, the linear and quadratic maximum SNR filters in practical conditions; $\mathbf{h}_{\max}(k,n)$ and $\widetilde{\mathbf{h}}_{\max}(k,n)$ denote, respectively, the linear and quadratic maximum SNR filters in ideal conditions; MMSE-LSA denotes the log-spectral amplitude estimator of [3]; STSS denotes the spectral subtraction in the short-time modulation domain method of [30]; $Y(k,n)$ denotes the input noisy observation: (a) PESQ scores and (b) STOI scores. We set $N = 3$ for all four maximum SNR filters and $\tau_{\mathrm{y}} = 2$ for the two quadratic maximum SNR filters.

10. Cohen, I.: Speech enhancement using super-gaussian speech models and noncausal a priori snr estimation. Speech Communication **47**(3), 336–350 (2005)
11. C. Hendriks, R., Richard, H., Jensen, J.: Log-spectral magnitude mmse estimators under super-gaussian densities., pp. 1319–1322 (2009)
12. Martin, R.: Speech enhancement using MMSE short time spectral estimation with gamma distributed speech priors. In: Proceedings of the 27th IEEE International Conference Acoustics Speech Signal Processing, ICASSP-02, vol. 1, pp. 253–256 (2002)
13. Erkelens, J.S., Hendriks, R.C., Heusdens, R., Jensen, J.: Minimum mean-square error estimation of discrete fourier coefficients with generalized gamma priors. IEEE Transactions on Audio, Speech, and Language Processing **15**(6), 1741–1752 (2007). doi:10.1109/TASL.2007.899233
14. R. Martin, R., Breithaupt, C.: Speech enhancement in the DFT domain using Laplacian speech priors. In: Proceedings of the 8th International Workshop on Acoustic Echo and Noise Control (IWAENC), pp. 87–90 (2003)

15. Cohen, I.: Speech spectral modeling and enhancement based on autoregressive conditional heteroscedasticity models. Signal Processing **86**(4), 698–709 (2006)
16. Cohen, I., Berdugo, B.: Speech enhancement for non-stationary noise environments. Signal Processing **81**(11), 2403–2418 (2001)
17. Cohen, I.: Relaxed statistical model for speech enhancement and a priori SNR estimation. IEEE Transactions on Speech and Audio Processing **13**(5), 870–881 (2005)
18. Huang, Y.A., Benesty, J.: A multi-frame approach to the frequency-domain single-channel noise reduction problem. IEEE Transactions on Audio, Speech, and Language Processing **20**(4), 1256–1269 (2012)
19. Huang, G., Benesty, J., Long, T., Chen, J.: A family of maximum SNR filters for noise reduction. IEEE/ACM Transactions on Audio, Speech, and Language Processing **22**(12), 2034–2047 (2014)
20. Itzhak, G., Benesty, J., Cohen, I.: Nonlinear kronecker product filtering for multichannel noise reduction. Speech Communication **114**, 49–59 (2019)
21. Loizou, P.C.: Speech Enhancement: Theory and Practice, 2nd edn. CRC Press, Inc., Boca Raton, FL, USA (2013)
22. Benesty, J., Chen, J., Huang, Y., Cohen, I.: Noise Reduction in Speech Processing, 1st edn. Springer-Verlag Berlin Heidelberg, Berlin (2009)
23. Benesty, J., Chen, J., Huang, Y.A.: Noise reduction algorithms in a generalized transform domain. IEEE Transactions on Audio, Speech, and Language Processing **17**(6), 1109–1123 (2009)
24. Harville, D.A.: Matrix Algebra from a Statistician's Perspective, 1st edn. Springer-Verlag New York, New York (1997)
25. Golub, G.H., Loan, C.F.V.: Matrix Computations, 3rd edn. Baltimore, Maryland: The Johns Hopkins University Press, Baltimore (1996)
26. Stuart, A., Ord, K.: Kendall's Advanced Theory of Statistics, Volume1: Distribution Theory, 6th edn. Wiley, New York (2010)
27. DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus CDROM. NIST (1993)
28. Rix, A.W., Beerends, J.G., Hollier, M.P., Hekstra, A.P.: Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. In: 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings, vol. 2, pp. 749–7522 (2001)
29. Taal, C.H., Hendriks, R.C., Heusdens, R., Jensen, J.: An algorithm for intelligibility prediction of time–frequency weighted noisy speech. IEEE Transactions on Audio, Speech, and Language Processing **19**(7), 2125–2136 (2011). doi:10.1109/TASL.2011.2114881
30. Paliwal, K., Wójcicki, K., Schwerin, B.: Single-channel speech enhancement using spectral subtraction in the short-time modulation domain. Speech Communication, 450–475 (2010)
31. Schasse, A., Martin, R.: Estimation of subband speech correlations for noise reduction via MVDR processing. IEEE/ACM Transactions on Audio, Speech, and Language Processing **22**(9), 1355–1365 (2014)