DFT Domain Subspace Based Noise Tracking for Speech Enhancement

Richard C. Hendriks, Jesper Jensen and Richard Heusdens

Department of Mediamatics, Delft University of Technology, 2628 CD Delft, The Netherlands
{R.C.Hendriks, J.Jensen, R.Heusdens}@TUDelft.nl

Abstract

Most DFT domain based speech enhancement methods are dependent on an estimate of the noise power spectral density (PSD). For non-stationary noise sources it is desirable to estimate the noise PSD also in spectral regions where speech is present. In this paper a new method for noise tracking is presented, based on eigenvalue decompositions of correlation matrices that are constructed from time series of noisy DFT coefficients. The presented method can estimate the noise PSD at time-frequency points where both speech and noise are present. In comparison to state-of-the-art noise tracking algorithms the proposed algorithm reduces the estimation error between the estimated and the true noise PSD and improves segmental SNR when combined with an enhancement system with several dB.

Index Terms: Speech enhancement, noise tracking, DFT domain subspace decompositions.

1. Introduction

The increasing interest for mobile speech processors to work well in noisy environments have led to an increased focus on single-channel speech enhancement methods. Many of these methods work in the discrete Fourier transform (DFT) domain, where clean DFT coefficients are estimated by applying either a gain function to the noisy DFT coefficients or to the magnitude of the noisy DFT coefficients. These gain functions are derived under an error criterion, e.g. minimum mean square error (MMSE), and an assumption on the distribution of the speech and noise DFT coefficients. Gain function have been derived assuming a Gaussian distribution [1] for the speech DFT coefficients. More recently somewhat more advanced estimators under supergaussian densities have been proposed, see e.g. [2].

All gain function have in common that they rely on knowledge of the noise power spectral density (PSD). Since this is generally unknown, it must be estimated from the noisy speech signal. An overestimation of the noise PSD will lead to oversuppression and, as a consequence, to a potential loss of speech quality, while an underestimation will lead to an unnecessary high level of residual noise. Accurate tracking of the noise PSD is therefore essential, especially in non-stationary noise conditions.

A well-known noise tracking algorithm is minimum statistics [3] (MS). This method exploits the property that the minimum power level in a particular frequency bin seen across a sufficiently long time interval is due to the noise process. From this minimum the average noise power can be estimated by applying a bias compensation. The size of this time interval should be such that there is at least one noise-only observation within the window. The minimum size of the time window is therefore dependent on the duration of speech presence in a frequency bin. If the time window is chosen too short and speech energy is constantly present in the search window, MS will track the PSD of the noisy speech instead of the noise PSD and consequently overestimate the noise level. If, on the other hand, the time window is chosen too long, changes in the noise power level are not tracked or can only be tracked with a large delay.

In this paper we present a new noise tracking method that allows to update the noise PSD for each DFT coefficient when both speech and noise are present. This method is based on correlation matrices that are constructed from time series of noisy DFT coefficients. We exploit the fact that these correlation matrices can be decomposed using an eigenvalue decomposition into two matrices of which the columns span two mutually orthogonal vector spaces, namely a signal (+ noise) subspace and a noise-only subspace. The noise-only subspace is used to update the noise PSD.

2. Signal Model and DFT Domain Subspace Decompositions

We assume an additive noise model, i.e. $Y(k,i) = X(k,i) + D(k,i)$, where Y is a noisy speech DFT coefficient, X a clean speech DFT coefficient, D a noise DFT coefficient, k the frequency index and i the time frame index. The DFT coefficients Y, X and D are assumed to be complex zero-mean random variables and X and D are assumed uncorrelated, i.e. $E[X(k,i)D(k,i)] = 0, \forall\,(k,i)$.

The proposed method is based on subspace decompositions of noisy speech correlation matrices in the DFT domain. Per frequency bin, DFT coefficients are collected from time frames $i-p_1$ up to frame $i+p_2$ and form a vector $\mathbf{Y}(k,i) \in \mathbb{C}^M$ with $M = p_1 + p_2 + 1$. That is,

$$\mathbf{Y}(k,i) = [Y(k,i-p_1), ..., Y(k,i+p_2)]^T. \quad (1)$$

Let $\mathbf{R}_Y(k,i) \in \mathbb{C}^{M \times M}$ be the noisy speech correlation matrix related to frequency bin k and time frame i defined as

$$\mathbf{R}_Y(k,i) = E\left[\mathbf{Y}(k,i)\mathbf{Y}^H(k,i)\right], \quad (2)$$

where H indicates Hermitian transposition. Similarly we define the speech correlation matrix $\mathbf{R}_X(k,i) \in \mathbb{C}^{M \times M}$ and the noise correlation matrix $\mathbf{R}_D(k,i) \in \mathbb{C}^{M \times M}$. Using the assumption that speech and noise are uncorrelated we can write the noisy speech correlation matrix $\mathbf{R}_Y(k,i)$ as

$$\mathbf{R}_Y(k,i) = \mathbf{R}_X(k,i) + \mathbf{R}_D(k,i).$$

Let us assume that $\mathbf{R}_D(k,i) = \sigma_D^2(k,i) I_M$, that is, the noise DFT coefficients in $\mathbf{D}(k,i)$ are uncorrelated. Strictly speaking, this is only true for white noise signals with non-overlapping frames. However, this is approximately true for non-white noise

The research is supported by Philips Research and the Technology Foundation STW.

signals in case frames do not overlap and the correlation time of the noise is small enough [4]. In case of overlapping frames this assumption will be violated. This violation can be overcome by applying a pre-whitening transform, as we describe in Section 3.

The clean speech correlation matrix $\mathbf{R}_X(k,i)$ is assumed to be of low rank. In particular this is true when speech sounds can be modelled by a sum of complex exponentials, e.g. voiced speech sounds [5]. When the speech can be described using such a low-rank signal subspace and the noise-only subspace is of full rank, as would be the case when $\mathbf{R}_D(k,i) = \sigma_D^2(k,i)I_M$, the eigenvalues that describe the energy in the noise-only subspace allow for an update of the noise statistics, even when speech is constantly present.

Let $\mathbf{R}_X(k,i) = \mathbf{U}\mathbf{\Lambda}_X\mathbf{U}^H$ denote the eigenvalue decomposition of the clean speech correlation matrix related to a frequency bin k and time frame i. Here, $\mathbf{U} \in \mathbb{C}^{M\times M}$ is a unitary matrix and contains the eigenvectors as columns and $\mathbf{\Lambda}_X = \text{diag}\left(\lambda_{X_1}, ..., \lambda_{X_Q}, 0, ..., 0\right)$ $Q \leq M$, a diagonal matrix with the eigenvalues $\lambda_{X_1} \geq \lambda_{X_2} \geq ... \geq \lambda_{X_Q} \geq 0$ on the main diagonal. Using the assumption that $\mathbf{R}_D(k,i) = \sigma_D^2(k,i)I_M$ and that $X(k,i)$ and $D(k,i)$ are uncorrelated we can write the eigenvalue decomposition of $\mathbf{R}_Y(k,i)$ as $\mathbf{R}_Y(k,i) = \mathbf{U}\left(\mathbf{\Lambda}_X + \sigma_D^2 I_M\right)\mathbf{U}^H$, i.e. $\mathbf{R}_Y(k,i)$, $\mathbf{R}_X(k,i)$ and $\mathbf{R}_D(k,i)$ have the same eigenvectors and the eigenvalues add up.

The eigenvector matrix $\mathbf{U}$ can be partitioned as $\mathbf{U} = [\mathbf{U}_1\mathbf{U}_2]$, where the columns of $\mathbf{U}_1 \in \mathbb{C}^{M\times Q}$ form a Q-dimensional basis for the signal (+ noise) subspace and the columns of $\mathbf{U}_2 \in \mathbb{C}^{M\times M-Q}$ form a basis for the noise-only subspace. Assuming that there indeed exists a low dimensional signal subspace, i.e. $Q < M$, the eigenvalues in the noise-only subspace can be used to determine the noise PSD $\sigma_D^2(k,i)$ as the noise-only subspace eigenvalue matrix equals $I_{(M-Q)}\sigma_D^2(k,i)$.

2.1. Estimating $\sigma_D^2(k,i)$

In practice $\mathbf{R}_Y(k,i)$ in (2) is unknown and estimated from a limited number of samples by $\hat{\mathbf{R}}_Y(k,i) = \mathcal{Y}(k,i)\mathcal{Y}^H(k,i)$, with $\mathcal{Y} \in \mathbb{C}^{M\times L}$ a Hankel-structured datamatrix with $[y(k,i-n_1), ..., y(k,i-n_1+M-1)]^T$ and $[y(k,i-n_1), ..., y(k,i-n_1+L-1)]$ its first column and first row, respectively, where the small letters y indicate realizations of the random variable Y.

Let $\hat{\lambda}_l$ be an eigenvalue of an estimated covariance matrix. Given the eigenvalue decomposition of $\hat{\mathbf{R}}_Y(k,i)$, it can then be shown [6] that under certain assumptions a maximum likelihood estimate of the spectral noise variance is given by $\hat{\sigma}_D^2(k,i) = \frac{1}{M-Q}\sum_{l=Q+1}^{M}\hat{\lambda}_{Y_l}$. Estimation of the model-order Q is a well-known problem for large data-records, see e.g. [7]. However, when $\mathbf{R}_Y(k,i)$ is estimated based on a few data samples, which is the case here, existing model order estimators lead to inaccurate estimates of Q. Moreover, due to the inaccurate model order estimation and not always clear distinction between the noise and signal subspace, the resulting noise PSD estimate may be biased dependent on the dimension of the signal subspace.

2.1.1. Model order estimation

To reduce the inaccuracy in the estimation of the signal subspace dimension we consider an alternative approach for signal subspace model order estimation, where we exploit the fact that some *a priori* information of the noise PSD is present. In this paper we use the noise PSD estimate of the previous frame. This implicitly assumes relatively slowly varying noise. However, as will be shown in simulation experiments in Section 4 this does not limit the practical performance; a change in the noise level up to 15 dB per second can successfully be tracked.

In order to determine the model order Q, we assume that the eigenvalues of $\hat{\mathbf{R}}_Y(k,i)$ that correspond to the noise-only subspace have an exponential distribution. A noisy eigenvalue $\hat{\lambda}_{Y_l}$ is decided to belong to the signal subspace when the probability of observing an eigenvalue equal or larger than $\hat{\lambda}_{Y_l}$ is smaller than a pre-chosen minimum probability P_{min}, we can write this as

$$\int_{\hat{\lambda}_{Y_l}}^{+\infty} f_{\Lambda_D}(\lambda_D)d\lambda_D < P_{min}, \tag{3}$$

with $f_{\Lambda_D}(\lambda_D)$ the pdf of the noise eigenvalues with its first moment equal to the *a priori* known noise PSD, i.e. the noise PSD estimate of the previous frame.

This approach can be seen within a hypothesis testing framework where H_0 and H_1 are defined as

$$\begin{array}{ll} H_0: & \hat{\lambda}_{Y_l} \text{ belongs to the noise-only subspace} \\ H_1: & \hat{\lambda}_{Y_l} \text{ belongs to the signal subspace.} \end{array} \tag{4}$$

Given a threshold λ_{th}, H_1 is decided when $\hat{\lambda}_{Y_l} > \lambda_{th}$. When $\hat{\lambda}_{Y_l} \leq \lambda_{th}$, $\hat{\lambda}_{Y_l}$ is decided to belong the the noise-only subspace. The hypothesis is evaluated for all eigenvalues in increasing order until H_0 is rejected, which determines then the dimension of the noise and the signal subspace, $M-Q$ and Q, respectively. The threshold λ_{th} can be expressed in terms of the false alarm probability $P_{fa} = P_{min}$ and is given by $\lambda_{th} = -\sigma_D^2 \ln P_{fa}$.

While the MDL estimator [7] does not assume knowledge of the noise PSD, it is possible to modify [7], such that it uses *a priori* information on the noise PSD as well [6]. Comparisons to the existing MDL based model order estimator [7] and the modified MDL estimator [6] have shown that the proposed model order estimator leads to a smaller mean square error of the estimated model order.

2.1.2. Bias compensation

Because there is not always a clear separation between the noise-only and the signal subspace, a bias can sometimes be introduced in the noise PSD estimate. To remove a possible bias in the noise PSD estimate, we introduce a signal subspace dimension dependent bias compensation factor $B(Q)$ and compute $\hat{\sigma}_D^2(k,i)$ as

$$\hat{\sigma}_D^2(k,i) = \frac{1}{B(Q)}\frac{1}{M-Q}\sum_{l=Q+1}^{M}\hat{\lambda}_{Y_l}(k,i). \tag{5}$$

To compute the bias compensation factor $B(Q)$, $Q = 0, 1, ..., M$, we make use of a training procedure based on speech data degraded by white noise with a known variance $\sigma_D^2(k,i) = 1 \ \forall \ (k,i)$. Let the local bias at a time-frequency point (k,i) be denoted by $\tilde{B}(k,i)$. We can then write

$$\tilde{B}(k,i) = \frac{\frac{1}{M-Q}\sum_{l=Q+1}^{M}\hat{\lambda}_{Y_l}(k,i)}{\sigma_D^2} \tag{6}$$

Let $\mathcal{Q}(Q)$ be the set of time-frequency points in the training data for which signal subspace dimension Q is estimated.

$B(Q), Q = 0, 1, ..., M$, is then computed by averaging $\tilde{B}(k,i)$ over the set $\mathcal{Q}(Q)$ leading to

$$B(Q) = \frac{1}{|\mathcal{Q}(Q)|} \sum_{(k,i)\in\mathcal{Q}} \left[\tilde{B}(k,i)\right],$$

where $|\mathcal{Q}(Q)|$ is the cardinality of the set $\mathcal{Q}(Q)$. Notice that computing the bias compensation factor in the training phase using the same signal subspace dimension estimator as when used in practice has the advantage that it can help to overcome systematic errors due to the dimension estimator.

3. Pre-Whitening

In Section 2 the assumption was made that $\mathbf{R}_D(k,i) = \sigma_D^2(k,i)\mathbf{I}_M$. Although this assumption holds as long as the DFT coefficients in $\mathbf{D}(k,i)$ are computed from time frames that are not overlapping, it becomes less valid when an overlap is introduced. In this Section we indicate how to determine a whitening transformation, such that the aforementioned assumption is valid.

Let $D_t(m)$ denote a time domain noise sample considered as a random variable and let P denote the frame shift in samples. Let $R_D(k,i;p)$ denote the correlation between a noise DFT coefficient $D(k,i+p)$ and $D(k,i)$ with frame lag p and let $\overline{D(k,i)}$ denote complex conjugation of $D(k,i)$. It can then be shown [6] that the correlation $R_D(k,i;p)$ between noise DFT coefficients $D(k,i+p)$ and $D(k,i)$ can be written as

$$\begin{aligned} R_D(k,i;p) &= E[D(k,i+p)\overline{D(k,i)}] \\ &= E\left[\left(\sum_{m=0}^{K-1} D_t(m+(i+p)P)e^{-j2\pi km/K}\right)\right. \\ &\quad \times \left.\overline{\sum_{n=0}^{K-1} D_t(n+iP)e^{-j2\pi kn/K}}\right] \\ &= \underbrace{e^{j2\pi kpP/K}\sum_{m=pP}^{K-1} E\left[|D_t(m+iP)|^2\right]}_{\tilde{R}_D(k,i;p)} + R_C(k,i;p) \end{aligned}$$

We conclude that the correlation $R_D(k,i;p)$ consists of two components; a term $\tilde{R}_D(k,i;p)$ and a term $R_C(k,i;p)$. $R_C(k,i;p)$ contains all the cross-terms and is dependent on the cross-correlation between the time samples. In general it holds that $R_C(k,i;p)$ decreases for increasing P. Also, the shorter the correlation time in the noise, the smaller $R_C(k,i;p)$ becomes. For $R_D(k,i;p)$ with $p > 0$ it follows from Eq. (3) that that even if the time domain process $D_t(\cdot)$ is completely uncorrelated $R_D(k,i;p) \neq 0$, unless $P > K-1$, which means no overlap between consecutive frames.

Let $\mathbf{R}_D(k,i)$ denote a Toeplitz matrix with $[R_D(k,i;0), ..., R_D(k,i;p)]^T$ its first column and $[R_D(k,i;0), ..., R_D(k,i;-p)]$ its first row. Whitening of $\mathbf{Y}(k,i)$ is performed by computing $\mathbf{Y}_{pre}(k,i) = \mathbf{R}_D^{-\frac{1}{2}}(k,i)\mathbf{Y}(k,i)$. $\mathbf{Y}_{pre}(k,i)$ is then used in Eq. (2).

Unfortunately, in practice, matrix $\mathbf{R}_D(k,i)$ is unknown. Using simulations with white noise training data we can estimate $\tilde{R}_D(k,i;p)$ for a given overlap. The term $R_C(k,i;p)$ is dependent on the correlation in the noise signal itself and is in general unknown. Simulations on various noise sources have shown [6] that the influence of $R_C(k,i;p)$ is rather small. In practice we will therefore neglect the correlation term $R_C(k,i;p)$ and use $\tilde{R}_D(k,i;p)$ computed on white noise with a given overlap to whiten colored noise in $\mathbf{Y}(k,i)$. We thus use the approximated correlation matrix $\tilde{\mathbf{R}}_D(k,i)$ and compute

$$\mathbf{Y}_{pre}(k,i) = \tilde{\mathbf{R}}_D^{-\frac{1}{2}}(k,i)\mathbf{Y}(k,i). \quad (7)$$

Notice, that when $\hat{\sigma}_D^2(k,i)$ is estimated in the whitened domain, it has to be corrected with a scaling factor $\frac{tr\left[\tilde{\mathbf{R}}_D(k,i)\right]}{M}$, which is due to the whitening transform.

For some highly correlated noise types, i.e. with long correlation time, the aforementioned assumption of neglecting the correlation term $R_C(k,i;p)$ might be less valid. In that case Eq. (7) is not sufficient to whiten the noise process. A possible solution is to use the signal subspace dimension estimator and update the estimated correlation matrix when the estimated noise-only subspace is full rank, i.e. $Q = 0$. However, all results presented in Section 4 are obtained using Eq. (7).

4. Experimental Results

For performance evaluation we compare the proposed method with the MS noise tracking algorithm [3]. The speech and noise signals originate from the Noizeus [8] database. This database was extended with non-stationary white Gaussian noise. For the non-stationary white Gaussian noise, the initial noise level is 0, 5, 10 and 15 dB, respectively, and then gradually increases in one second by 15 dB where it stays at that level for 2 seconds after which it decreases again by 15 dB in one second. Noisy signals are constructed synthetically at global input SNRs of 0, 5, 10 and 15 dB. All signals are sampled at 8 kHz and filtered at telephone bandwidth. The noisy time domain signals are divided in frames of 256 samples with 50 % overlap. For both analysis and synthesis a square root Hann window is used. The DFT coefficients that are used to form the data-matrix $\mathcal{Y}$ originate from time frames taken with an overlap of 87.5 %. The dimensions of $\mathcal{Y}$ were chosen as $M = L = 7$ and $n_1 = n_2 = 6$. The estimated noise PSDs $\hat{\sigma}_D^2(k,i)$ are smoothed using an exponential smoother with adaptive smoothing factors [3].

To illustrate the noise tracking performance of the proposed approach within a typical example, we concatenated four speech signals and degraded this with the non-stationary white Gaussian noise with initial noise level at an SNR of 10 dB. In Fig. 1 the estimated noise PSDs are shown for the proposed approach and the MS approach together with the true noise variance for a single frequency bin $k = 20$, which corresponds to 625 Hz and thus contains speech energy most of the time. The true noise variance was measured using noise periodograms that are smoothed over time using an exponential window, i.e.

$$\sigma_D^2(k,i) = \alpha\sigma_D^2(k,i-1) + (1-\alpha)|D(k,i)|^2, \quad (8)$$

with $\alpha = 0.9$ [3].

We see that the proposed approach follows the increase in the noise level much better than the MS approach. This is due to the fact that the proposed approach can track changes in the noise level during speech presence. The MS approach on the other hand is limited in its update rate due to its search window and because it can not track the noise when speech is continuously present in a bin. This results for MS in the delayed tracking of a rising noise level in Fig. 1.

For further objective performance evaluation we propose the use of a symmetric segmental logarithmic estimation error, defined as

$$\text{LOG-Err}_{\text{seg}} = \frac{1}{IK}\sum_k^I \sum_i^I \left|10\log\left[\frac{\sigma_D^2(k,i)}{\hat{\sigma}_D^2(k,i)}\right]\right|,$$

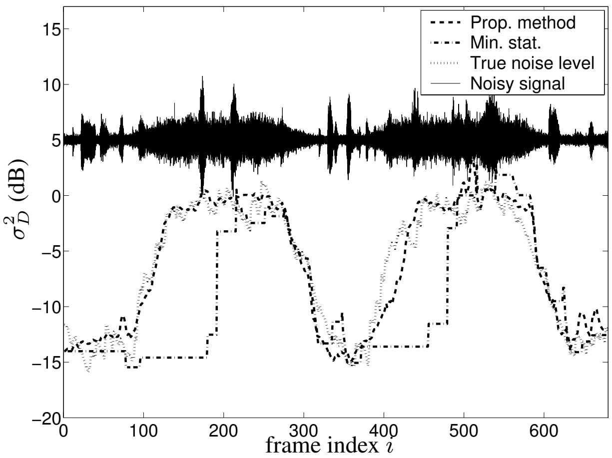

Figure 1: *Comparison between proposed method and minimum statistics.*

where I is the number of frames and where $\sigma_D^2(k,i)$ is the ideal noise PSD measured using Eq. (8). We prefer the use of LOG-Err$_{\text{seg}}$ above the segmental relative estimation error defined in [9] because the latter is non-symmetric and is more sensitive to overestimates than to underestimates.

In order to evaluate the influence of the proposed noise tracking algorithm on speech enhancement performance we use the estimated noise PSDs within a DFT domain based speech enhancement algorithm. Here, the gain function is the MMSE amplitude estimator under the generalized Gamma model with $\gamma = 2$ and $\nu = 0.1$ as presented in [2]. For *a priori* SNR estimation we use the decision directed approach [1] with a smoothing factor $\alpha = 0.98$. For performance comparison we use segmental SNR, i.e.,

$$\text{SNR}_{\text{seg}} = \frac{1}{I}\sum_{i}^{I} 10\log_{10}\frac{\sum_k |x(k,i)|^2}{\sum_k |x(k,i)-\hat{x}(k,i)|^2}, \qquad (9)$$

where $x(k,i)$ is a realization of a clean speech DFT and $\hat{x}(k,i)$ is its clean speech DFT estimate, respectively. Notice that unlike the LOG-Err$_{\text{seg}}$ measure, the performance measured using SNR_{seg} is not only influenced by the noise tracking algorithm, but also by the gain function and *a priori* SNR estimator.

In Table 1 we show performance evaluations for several noise types averaged over speech signals originating from the Noizeus database [8]. We compare noise tracking using MS and the proposed approach. We see that in general the proposed approach improves on both objective measures. Especially for noise sources that are characterized by a gradual change in the noise power (e.g. non-stationary white gaussian noise) the proposed approach outperforms MS. This is mainly due to the fact that the proposed method allows for a continuous update of the noise PSD, leading to a faster update of changes in the noise power than MS.

noise source	input SNR (dB)	LOG-Err$_{\text{seg}}$ MS	LOG-Err$_{\text{seg}}$ prop. method	SNR_{seg} (dB) MS	SNR_{seg} (dB) prop method.
train	0	2.6	1.9	-4.1	-2.3
	5	2.9	1.9	-0.5	1.1
	10	2.6	2.0	3.6	4.7
	15	2.7	2.1	7.4	8.0
street	0	2.3	2.0	-4.1	-3.5
	5	2.8	2.0	-0.9	0.3
	10	3.1	2.2	3.2	4.1
	15	2.7	2.5	7.0	7.1
restaurant	0	3.3	2.5	-6.1	-5.7
	5	3.4	2.6	-2.2	-1.2
	10	3.1	2.6	2.3	2.7
	15	2.9	2.9	6.1	6.2
non-stationary White Gaussian	0	4.0	0.9	-14.5	-9.4
	5	4.1	1.0	-10.5	-5.8
	10	4.1	1.1	-6.5	-2.5
	15	4.0	1.4	-2.5	0.6

Table 1: *Performance in terms of* LOG-Err$_{\text{seg}}$ and SNR_{seg} (dB)

5. Concluding Remarks

In this paper we presented a new approach for noise tracking. The method is based on construction of correlation matrices in the DFT domain per time frequency point. These correlation matrices are decomposed into a signal subspace and a noise-only subspace. The noise PSD is updated whenever the signal subspace is not of full rank. This approach makes it possible to update the noise PSD, even when speech is present. Experimental results have shown that the error between the true and estimated noise PSD is decreased and segmental SNR is increased with several dB.

6. References

[1] 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, no. 6, pp. 1109–1121, December 1984.

[2] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum mean-square error estimation of discrete Fourier coefficients with generalized gamma priors," *Accepted for publication in IEEE Trans. Audio Speech and Language Processing*.

[3] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Processing*, vol. 9, no. 5, pp. 504–512, July 2001.

[4] D. R. Brillinger, *Time Series: Data Analysis and Theory*, SIAM, Philadelphia, 2001.

[5] R. J. McAulay and T. F. Quarieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 4, pp. 744–754, Aug. 1986.

[6] R. C. Hendriks, J. Jensen, and R. Heusdens, "Noise tracking using DFT domain subspace decompositions," *submitted to IEEE Trans. Audio Speech and Language Processing*.

[7] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 387–392, 1985.

[8] Y. Hu and P. Loizou, "Subjective comparison of speech enhancement algorithms," in *IEEE Int. Conf. Acoust., Speech,Signal Processing*, May 2006, vol. 1, pp. 153–156.

[9] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Transactions on speech and audio processing*, vol. 11, no. 5, pp. 446–475, September 2003.