Noise Tracking by Exploiting DFT-Domain Subspace Decompositions

Richard C. Hendriks[1], *Jesper Jensen*[2] *and Richard Heusdens*[1]

[1]Department of Mediamatics, Delft University of Technology, 2628 CD Delft, The Netherlands.
E-Mail: {R.C.Hendriks, R.Heusdens}@tudelft.nl
[2]Oticon A/S, 2765 Smørum, Denmark. E-Mail: jsj@oticon.dk

Abstract

DFT domain based noise reduction algorithms can be effective for noise reduction in various speech processing applications. In general these algorithms apply a noise-PSD dependent estimator to the noisy speech DFT coefficients in order to estimate the clean speech DFT coefficients. Since the noise PSD is unknown in advance, estimation is one of the crucial elements of such a noise reduction system. In this paper we discuss a method for estimation of the noise PSD while both speech and noise are present. This method is based on DFT domain based subspace decompositions of noisy correlation matrices that are estimated per frequency bin in the DFT domain. The proposed algorithm shows increased tracking performance of the noise PSD in comparison to minimum statistics. Further, listening tests show a preference for the proposed method over MS.

1 Introduction

The presence of ambient noise is an important problem for various human-to-human as well as human-to-machine speech processing applications. For reduction of ambient noise in the single-microphone context the class of discrete Fourier transform (DFT) based methods has attracted an increased interest recently. This is partly due to its low computational complexity, relatively good performance as well as its ease in conceptual understanding. A lot of attention has been paid on the derivation of clean speech DFT estimators, see e.g.[1][2][3][4]. A common property of these estimators is that they are all based on assumed knowledge of the noise power spectral density (PSD). However, in practice the noise PSD is unknown and needs to be estimated. Estimation of the noise PSD density is therefore a crucial part of a noise reduction system. An overestimation of the noise PSD will lead in general to over-suppression and, as a consequence, to a potential loss of speech quality, while an underestimation will lead to an unnecessary high level of residual noise.

An algorithm which is among state-of-the-art for noise tracking is minimum statistics [5] (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. The size of this time interval turns out to be crucial for the noise tracking performance. If the time window is chosen too short and speech energy is constantly present in the search window, MS will overestimate the noise level. If, on the other hand, the time window is chosen too long, changes in the noise power level can only be tracked with a large delay.

Recently we proposed a new method for noise tracking, which shows significant improvements over minimum statistics based approaches in terms of noise-PSD tracking performance [6]. The advantage of the method proposed in [6] over minimum statistics is that it allows to update the noise PSD for each DFT coefficient when both speech and noise are present. Obviously this is not possible with MS based methods, since these methods rely on the assumption that part of the search window consists of noise-only samples. The proposed method is based on construction of correlation matrices from time series of noisy DFT coefficients seen in a particular frequency bin. These correlation matrices can be decomposed using an eigenvalue decomposition into two sub-matrices of which the columns span two mutually orthogonal vector spaces, namely a signal (+ noise) subspace and a noise-only subspace. Because the speech signal seen in such a frequency bin across time can often be described by a low rank model, the signal subspace will be of low dimension. The eigenvalues that describe the energy in the noise-only subspace then allow for an update of the noise statistics, even when speech is constantly present. In this paper we further elaborate on the noise tracking method recently proposed in [6] and further motivate the assumptions that are made in the signal subspace dimension estimator that is used. We show by means of listening experiments that the proposed method for noise tracking is preferred over MS.

2 DFT Subspace Decompositions

The proposed noise PSD estimation algorithm aims at the situation where speech is degraded by additive noise, i.e.

$$Y(k,i) = X(k,i) + D(k,i),$$

where Y, X and D are a noisy speech, clean speech and noise DFT coefficient, respectively and k and i indicate the frequency-bin and the time-frame index, respectively. 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. DFT coefficients are collected per frequency bin for 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

$$\mathbf{R}_Y(k,i) = \mathbf{R}_X(k,i) + \mathbf{R}_D(k,i). \quad (3)$$

Let us further assume that the noise DFT coefficients in $\mathbf{D}(k,i)$ are uncorrelated, that is

$$\mathbf{R}_D(k,i) = \sigma_D^2(k,i)I_M. \quad (4)$$

Strictly speaking, this is only true when frames do not overlap and the correlation time of the noise is small enough. In case of overlapping frames this assumption will be violated, which can be overcome by applying a pre-whitening transform, as we describe in Section 3.

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 $\mathbf{\Lambda}_X = \mathrm{diag}\left(\lambda_{X_1},...,\lambda_{X_Q},0,...,0\right)$, $Q \leq M$, a diagonal matrix with the eigenvalues $\lambda_{X_1} \geq \lambda_{X_2} \geq ... \geq \lambda_{X_Q} > 0$ on the main diagonal. Using Eqs. (3) and (4) 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. When $Q < M$, i.e., for a low dimensional signal subspace, the eigenvalues in the noise-only subspace can be used to determine the noise PSD $\sigma_D^2(k,i)$. An assumption that we exploit for noise tracking is that the clean speech correlation matrix $\mathbf{R}_X(k,i)$ is in many cases of low rank, e.g., for the voiced speech sounds.

As the proposed noise tracking method is based on an eigenvalue decomposition of $\mathbf{R}_Y(k,i)$, which is unknown in practice, $\mathbf{R}_Y(k,i)$ needs to be estimated, e.g., by $\hat{\mathbf{R}}_Y(k,i) = \mathbb{Y}(k,i)\mathbb{Y}^H(k,i)$, with $\mathbb{Y} \in \mathrm{C}^{M\times L}$ a Hankel-structured data-matrix 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 row, respectively, where the small letters y indicate realizations of the random variable Y.

Let $\hat{\lambda}_{Y_l}$ be an eigenvalue of an estimated covariance matrix. Given the eigenvalue decomposition of $\hat{\mathbf{R}}_Y(k,i)$, it can be shown [6] that under certain assumptions the maximum likelihood estimate of the noise PSD is given by $\hat{\sigma}_D^2(k,i) = \frac{1}{M-Q}\sum_{l=Q+1}^{M}\hat{\lambda}_{Y_l}$.

2.1 Model order estimation

A crucial part in the estimation of $\sigma_D^2(k,i)$ from estimated correlation matrices $\hat{\mathbf{R}}_Y(k,i)$ is to determine the model-order Q. This is a well-known problem for large data-records, see e.g. [7]. However, when, as in our case, $\mathbf{R}_Y(k,i)$ is estimated based on very little data, existing model order estimators lead to inaccurate estimates of Q.

To estimate the signal subspace dimension Q based on small data records 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. At first we use the fact that the estimated noise PSD from the previous frame gives a rough estimate of the current noise PSD. Secondly we assume that the eigenvalues of $\hat{\mathbf{R}}_Y(k,i)$ that correspond to the noise-only subspace have an exponential distribution, say $f_{\Lambda_D}(\lambda_D)$ with its first moment equal to the *a priori* known noise PSD, i.e., the noise PSD estimate of the previous frame. We cannot mathematically show that the distribution of noise-only eigenvalues is truly exponential. However, the reasoning behind this assumption is that the Toeplitz matrix $\mathbf{R}_Y(k,i)$ is in its dimension asymptotically equivalent to a circulant matrix [8]. A circulant matrix is diagonalized by the DFT and the resulting sequence of eigenvalues are then under these asymptotical conditions equal to the PSD. Using the periodogram as an estimator for the PSD it is known that under the assumption that the complex noise DFT coefficients are Gaussian distributed the periodogram at a certain frequency bin is exponentially distributed. Subsequently it can be reasoned that under these asymptotic conditions the distribution of noise eigenvalues of an estimated correlation matrix can be approximated by an exponential distribution. To verify whether an exponential distribution is indeed a reasonable approximation of the distribution of noise eigenvalues we conducted an experiment where we compare measured histograms of noise eigenvalues to an exponential density. To do so, we first generate estimates of $\mathbf{R}_D(k,i)$ for a frequency bin centered around 1.56 kHz. For estimation of $\mathbf{R}_D(k,i)$ we use the same parameter settings as in the experimental results in Section 4, i.e., $M = L = 7$, $n_1 = n_2 = 6$ and DFT coefficients that are used to estimate $\mathbf{R}_D(k,i)$ are computed based on time frames taken with an overlap of 87.5 %. Then an eigenvalue decomposition is applied on the estimated noise correlation matrices $\hat{\mathbf{R}}_D(k,i)$. The resulting noise eigenvalues are collected and used to generate a histogram. This procedure is performed for three different (rather stationary) noise sources; white noise, car noise and F16 noise. In Fig. 1 the three histograms are shown and are compared to an exponential distribution. Clearly, for all three noise sources the exponential distribution shows a good fit.

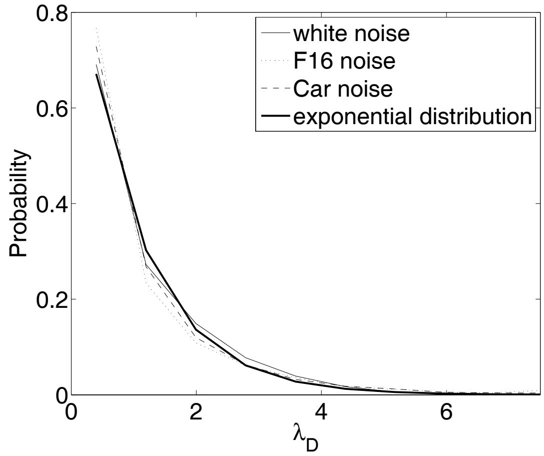

Figure 1: Comparison between exponential distribution and measured histograms of noise eigenvalues for three different noise sources.

Finally, the signal subspace dimension Q is determined. How many of the M eigenvalues $\hat{\lambda}_{Y_l}$, with $l \in \{1,...,M\}$, belong to the signal subspace is decided based on the probability that observing an eigenvalue equal or larger than $\hat{\lambda}_{Y_l}$ is smaller than a pre-chosen probability P_{min}, that is

$$\int_{\hat{\lambda}_{Y_l}}^{+\infty} f_{\Lambda_D}(\lambda_D)d\lambda_D < P_{min}. \quad (5)$$

2.2 Bias compensation of $\hat{\sigma}_D^2(k,i)$

A signal subspace dimension dependent bias compensation factor $B(Q)$ is introduced to overcome a bias in $\hat{\sigma}_D^2(k,i)$ due to consistent over or underestimates of Q, that is

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

with

$$B(Q) = \frac{E\left[\frac{1}{(M-Q)}\sum_{l=Q+1}^{M}\hat{\lambda}_{Y_l}(k,i)\right]}{\sigma_D^2(k,i)}. \quad (7)$$

The argumentation that we use to define the bias compensation factor is similar to the one introduced in [9] and based on the fact that σ_D^2 is proportional to

$$E\left[\frac{1}{(M-Q)}\sum_{l=Q+1}^{M}\hat{\lambda}_{Y_l}(k,i)\right]. \quad (8)$$

In order to compute the bias compensation factor $B(Q)$, for $Q = 0,1,...,M$, we approximate Eq. (7) by making 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 $\tilde{B}(k,i)$ be defined as

$$\tilde{B}(k,i) \quad = \quad \frac{\frac{1}{M-Q}\sum_{l=Q+1}^{M}\hat{\lambda}_{Y_l}(k,i)}{\sigma_D^2(k,i)}. \quad (9)$$

Let $\mathbb{Q}(Q)$ be the set of time-frequency points in the training data for which the signal subspace dimension is estimated to be Q. $B(Q)$, $Q = 0,1,...,M$, is then computed by averaging $\tilde{B}(k,i)$ over the set $\mathbb{Q}(Q)$ leading to

$$B(Q) = \frac{1}{|\mathbb{Q}(Q)|}\sum_{(k,i)\in\mathbb{Q}(Q)}\left[\tilde{B}(k,i)\right], \quad (10)$$

where $|\mathbb{Q}(Q)|$ is the cardinality of the set $\mathbb{Q}(Q)$.

The values for B are, dependent on Q and the exact settings of the algorithm, typically in de range from 1 to 2.

3 Pre-Whitening

Estimation of σ_D^2 according to Eq. (6) is strictly speaking only valid as long as Eq. (4) holds. In a practical setup where time frames might have an overlap this is not the case and pre-whitening of $\mathbf{R}_Y$ is necessary. Since pre-whitening implies knowledge of the yet unknown noise PSD, we indicate in this section how an approximated correlation matrix $\tilde{\mathbf{R}}_D$ can be obtained for pre-whitening.

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 frames are non-overlapping, i.e. $P > K-1$.

Let $\mathbf{R}_D(k,i)$ denote a Toeplitz matrix with $[R_D(k,i;0),\ldots,R_D(k,i;p)]^T$ its first column and $[R_D(k,i;0),\ldots,R_D(k,i;-p)]$ its first row. Whitening of $\mathbf{Y}(k,i)$ can be performed by computing $\mathbf{Y}_{pre}(k,i) = \mathbf{R}_D^{-1/2}(k,i)\mathbf{Y}(k,i)$ with $\mathbf{R}_D^{-1/2}(k,i)$ the principle square root of $\mathbf{R}_D^{-1}(k,i)$.

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 that the influence of $R_C(k,i;p)$ is rather small compared to the influence of $\tilde{R}_D(k,i;p)$ [6]. In practice we will therefore neglect the correlation term $R_C(k,i;p)$ because of its little influence and use $\tilde{R}_D(k,i;p)$ computed on white noise with a given overlap to whiten $\mathbf{Y}(k,i)$. Although this is an approximation it is sufficient for many noise types [6]. We thus use the approximated correlation matrix $\tilde{\mathbf{R}}_D(k,i)$ and compute

$$\mathbf{Y}_{pre}(k,i) = \tilde{\mathbf{R}}_D^{-1/2}(k,i)\mathbf{Y}(k,i). \tag{11}$$

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

4 Experimental Results

For performance evaluation we compare the proposed method with the minimum statistics based noise tracking algorithm implemented as described in [5]. The speech and noise signals originate from the Noizeus [10] database. This database was extended with stationary computer generated white Gaussian noise, noise originating from a passing train and non-stationary white Gaussian noise, respectively. All signals are filtered at telephone bandwidth and sampled at 8 kHz. 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 $\mathbb{Y}$ originate from time frames taken with an overlap of 87.5 %. The dimensions of $\mathbb{Y}$ 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 [5].

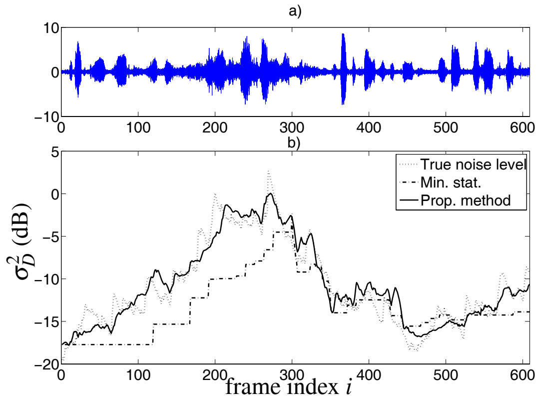

Figure 2: a) Noisy speech signal degraded by non-stationary train noise at an overall input SNR of 5 dB. b) Comparison between proposed method and minimum statistics.

To illustrate the noise tracking performance of the proposed approach within a typical example of noisy speech, we concatenated four speech signals and degraded this by noise originating from a passing train at 5 dB global SNR. In Fig. 2 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 $k = 20$. This bin index corresponds to a frequency band centered around 625 Hz. The proposed approach follows the increase in the noise level with a much smaller delay than the minimum statistics approach. This is due to the fact that the proposed approach can in contrast to MS track changes in the noise level during speech presence and continiously updates the noise PSD estimate. When the noise level decreases we see that both methods track approximately equally well. This difference in behavior of minimum statistics towards increasing and decreasing noise levels is due to the fact that MS tries to find the minimum. For a decreasing noise level the minimum will in general be found among the most recent samples in the search window resulting in a much smaller additional delay for decreasing noise levels than for increasing noise levels.

In Fig. 3 another example is shown where the same speech signal is degraded by non-stationary white noise. This noise source is created synthetically. Its initial noise level is 10 dB 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. We again see that the proposed approach tracks the increase in noise level much faster than the MS approach.

In order to evaluate the perceptual impact of the proposed noise tracking algorithm on speech enhancement performance we use the estimated noise PSDs for both the proposed approach and MS within a typical DFT domain based noise reduction algorithm and evaluate the performance by means of a listening test. The noise reduction scheme that we use is depicted in Fig. 4 and works on a frame-by-frame basis, where per frame the clean speech DFT coefficients are estimated. As estimator we use the MMSE magnitude estimator under the generalized Gamma model as presented in [4] with $\gamma = 2$ and $\nu = 0.1$. The maximum suppression was limited to 0.1 for perceptual reasons. For *a priori* SNR estimation we use the decision-directed (DD) approach [1].

The listening test that we perform is a so-called OAB test

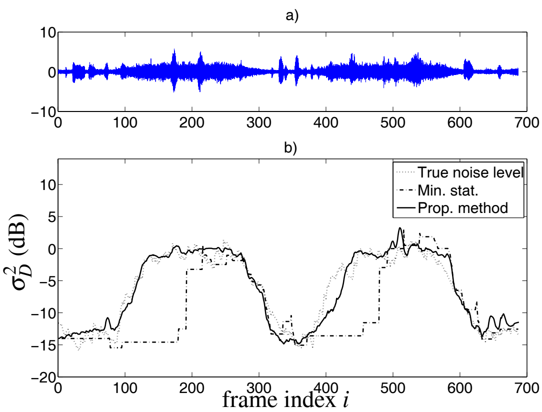

Figure 3: a) Noisy speech signal degraded by non-stationary white noise. b) Comparison between proposed method and minimum statistics. The estimated noise levels are shown for bin $k = 20$.

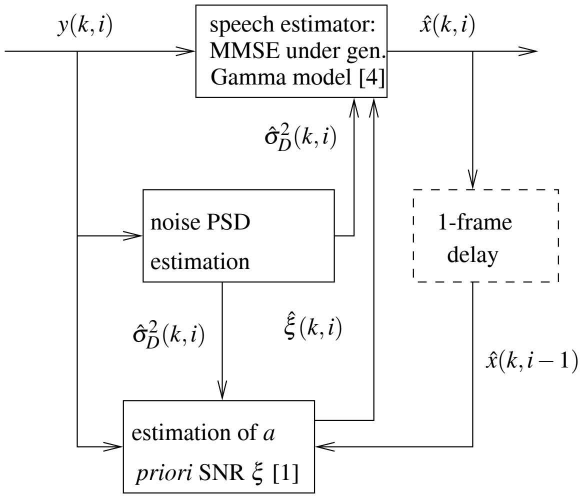

Figure 4: Blockdiagram of DFT-domain based enhancement algorithm.

with 8 participants, the authors not included. Here, O is the original clean speech signal and A and B are two noisy signals that are enhanced using the scheme in Fig. 4 with two different noise tracking algorithms. Method A uses the proposed noise tracking method, and method B uses the minimum statistics approach. The listeners were presented first the original signal followed by the two different enhanced signals A and B played in random order. The participants had to indicate their preference for excerpt A or B. Each series was repeated 4 times, with each time a randomized order of the signals A and B. In this listening test we used four different types of additive noise at two different SNRs, namely, white noise, street noise, noise originating from a passing train and non-stationary white noise at SNRs of 5 dB and 15 dB. For each noise type and noise power level we presented the listeners two female sentences and two male sentences. The average preference for method A under each test condition is shown in Table 1. Under all test conditions the proposed method for noise tracking was preferred over the minimum statistics approach.

5 Concluding Remarks

In this paper we discussed a DFT-domain based approach for noise tracking. This method exploits the fact that often speech signals observed in a frequency bin across time can be considered to be of low-rank. To estimate the noise PSD, correlation matrices are constructed in the DFT domain per time-frequency point and decomposed into a signal subspace and a noise-only subspace. The noise PSD is updated whenever the signal subspace is not of full rank. A crucial parameter in this noise tracking algorithm is the dimension of the signal subspace. To estimate this parameter certain assumptions are made on the distribution of noise eigenvalues. By means of histograms measured on several noise sources it was shown that the use of an exponential distribution is a good approximation of the noise eigenvalue distribution.

The presented noise tracking method is compared to minimum statistics and shows faster, almost immediate, tracking of changing noise statistics. In a listening test the proposed method was compared to minimum statistics, which demonstrated that under all test conditions the proposed method was preferred over minimum statistics.

noise source	input SNR	mean score for method A
white noise	5 dB	82.0 %
	15 dB	85.9 %
street noise	5 dB	77.3 %
	15 dB	67.2 %
passing train	5 dB	77.3 %
	15 dB	92.2 %
Non-stat. white noise	5 dB	96.1 %
	15 dB	89.1 %

Table 1: *Listening test results.*

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] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 5, pp. 845–856, Sept. 2005.

[3] T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-gaussian speech model," *EURASIP Journal on Applied Signal Processing*, vol. 7, pp. 1110–1126, May 2005.

[4] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum mean-square error estimation of discrete Fourier coefficients with generalized gamma priors," *IEEE Trans. Audio Speech and Language Processing*, vol. 15, no. 6, pp. 1741 – 1752, August 2007.

[5] 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.

[6] R. C. Hendriks, J. Jensen, and R. Heusdens, "Noise tracking using DFT domain subspace decompositions," *IEEE Trans. Audio Speech and Language Processing*, vol. 16, no. 3, pp. 541–553, March 2008.

[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] R. M. Gray, "Toeplitz and circulant matrices: A review," Tech. Rep., Stanford University, 2002.

[9] R. Martin, "Bias compensation methods for minimum statistics noise power spectral density estimation," *Signal Processing*, vol. 86, no. 6, pp. 1215–1229, June 2006.

[10] 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.