NOISE POWER SPECTRAL DENSITY ESTIMATION ON HIGHLY CORRELATED DATA

Dirk Mauler and Rainer Martin

{Dirk.Mauler, Rainer.Martin}@rub.de
Institute of Communication Acoustics (IKA), Ruhr-Universität Bochum, 44780 Bochum, Germany

ABSTRACT

In this contribution the Minimum Statistics noise power spectral density estimator [1] is revised for the particular case of highly correlated data which is observed for example when framewise processing with considerable frame overlap is performed. For this special case the noise power estimator tends to underestimate the noise power. We identify the variance estimator in the Minimum Statistics approach of being the origin of the observed underestimation. The variance estimator controls the bias compensation which is necessary to infer the mean power from a minimum value. This estimator turns out to be biased when the data is correlated. We provide an expression that describes the bias and show that by exploiting this the noise power estimation can be improved.

1. INTRODUCTION

Many speech processing algorithms like single- or multi-channel noise reduction, voice activity detectors, or robust speech recognition require knowledge of the power spectral density of a disturbing background noise. Several algorithms have been proposed for estimating the noise power [2], [3], [1]. In [1] a noise power spectral density estimation based on optimal smoothing of the squared magnitude of the noisy short-time Fourier transform coefficients and Minimum Statistics has been proposed. The method requires a bias compensation which has been shown in [4] to be accurate for a variety of moderately overlapping analysis windows. If, however, analysis windows with considerable overlap are used an underestimation of the noise power spectral density can be observed. In speech enhancement applications a larger window overlap might be motivated by delay considerations. In a double buffering block processing scheme the system latency can be reduced by decreasing the frame shift which results for a given spectral resolution in a larger frame overlap.

We show that the reason for the underestimation of the noise power spectral density is a biased variance estimator that controls the bias compensation which is used to compute the noise power estimate. It turns out that the variance estimator underestimates the variance by a factor that is proportional to the data correlation. Since spectral data from consecutive and considerably overlapping windows is correlated, noise power underestimation is observed in this particular case.

In the sequel we start with the analysis of the expectation of an autoregressive variance estimator for the general case of correlated data. In Section 3.1 we apply the findings to the Minimum Statistics noise power spectral density estimator and introduce in 3.2 an extension that accounts for the effects which are observed if the data is highly correlated, e.g. due to considerably overlapping analysis windows. In Section 4 we present experimental results in terms of instrumental measures. Finally, we summarize the work in Section 5.

2. VARIANCE ESTIMATION

A short-term estimate for the variance σ_x^2 of a possibly correlated random variable $\mathbf{x}$ at time instant i can be defined as

$$\widehat{\sigma^2_{x}}_i \;=\; \widehat{\psi^2_{x}}_i - \widehat{\mu_{x}}_i^2 \tag{1}$$

where the squared mean estimate $\widehat{\psi^2_{x}}_i$ and the mean estimate $\widehat{\mu_{x}}_i$ are obtained by means of a first order recursive average over the observations x_i of the random process $\mathbf{x}$, i.e.

$$\widehat{\psi^2_{x}}_i \;=\; \beta\widehat{\psi^2_{x}}_{i-1} + (1-\beta)x_i^2 \tag{2}$$

$$\widehat{\mu_{x}}_i \;=\; \beta\widehat{\mu_{x}}_{i-1} + (1-\beta)x_i \tag{3}$$

with $0 < \beta < 1$. We assume that the smoothing parameter β is set such that the signal is stationary within 3 - 4 times the smoothing time constant. Then the expected value of the estimator (1) can be derived as

$$E[\widehat{\sigma^2_{x}}_i] \;=\; \frac{2\beta}{1+\beta}\sigma_x^2\left(1-\frac{1-\beta}{\beta}\sum_{\kappa=1}^{\infty}\beta^\kappa\rho_{xx}(\kappa)\right) \tag{4}$$

where $\rho_{xx}(\kappa)$ denotes the correlation coefficient for lag κ

$$\rho_{xx}(\kappa) = \frac{E[(x_i-\mu_x)(x_{i+\kappa}-\mu_x)]}{\sigma_x^2}.$$

Equation (4) shows that the population variance $\widehat{\sigma^2_{x}}_i$ constitutes a biased estimator that systematically underestimates the process variance σ_x^2

- by a factor that only depends on the smoothing parameter β and
- by a factor that is additionally ruled by the temporal correlation $\rho_{xx}(\kappa)$ of the observations x_i of the given random process.

The underestimation increases the smaller the smoothing parameter β and the larger the span of correlation in the input data.
A reformulation of (4) shows that both factors amount to the variance of the mean estimate

$$E[\widehat{\sigma^2_{xi}}] = \sigma_x^2 - \mathrm{v}ar\{\widehat{\mu_{xi}}\} \quad (5)$$

which suggests to compensate for the bias by adding the estimated variance of the first order recursive mean estimate, $\mathrm{v}ar\{\widehat{\mu_{xi}}\}$. Note however, that this compensation term itself constitutes a variance estimate, that may be systematically underestimated and therefore fails to completely compensate for the underestimation of $\widehat{\sigma^2_{xi}}$.

3. MINIMUM STATISTICS NOISE POWER ESTIMATOR

3.1. Variance estimation

The noise power spectral density estimator based on Minimum Statistics tracks the minima in the adaptively averaged periodograms of the noisy signal ([1], Equ. (4))

$$\begin{aligned} P(\lambda,k) &= \alpha(\lambda,k)P(\lambda-1,k) \\ &+ (1-\alpha(\lambda,k))|Y(\lambda,k)|^2 \end{aligned} \quad (6)$$

with $Y(\lambda,k)$ being the short-time Fourier transform coefficient of frame λ and frequency bin k and $\alpha(\lambda,k)$ a time and frequency dependent smoothing parameter [1],[5]. To infer the mean power from an observed minimum a bias compensation factor has to be applied. The bias compensation factor B_{min} is a function of the length of the minimum search interval and the variance of the smoothed power spectral density estimate, $\mathrm{v}ar\{P(\lambda,k)\}$. While the search interval length is usually fix for the algorithm the variance of the adaptively smoothed periodograms $P(\lambda,k)$ has to be estimated. The variance estimator used in the Minimum Statistics noise power estimator is of the same kind as defined in Equations (1) to (3) with the smoothing parameter being now a function of frame λ and frequency bin k. For convenience we keep the notation used in [1]. Then, the variance estimator at frame λ and for frequency bin k is given by ([1], Equ.(22))

$$\widehat{\mathrm{v}ar}\{P(\lambda,k)\} = \overline{P^2}(\lambda,k) - \overline{P}^2(\lambda,k) \quad (7)$$

where $\overline{P^2}(\lambda,k)$ and $\overline{P}(\lambda,k)$ denote first order recursive averages of the squared smoothed periodograms and the mean smoothed periodograms, respectively. Using the same reasoning as in the preceding section we conclude that the estimated variance of the smoothed power spectral density estimates is underestimated by a factor that is ruled by the smoothing parameter and by a factor that depends on the correlation of the random variable $P(\lambda,k)$. Since the bias compensation factor B_{min} is controlled by the variance estimate, underestimation results in a compensation factor that is too small to raise the observed minimum value to the mean power. As a consequence, the noise power becomes underestimated.

Correlation can have different origins:

- Correlation due to overlapping block processing becomes important for frame overlap of more than 50%.
- Correlation induced by lowpass filtering the squared magnitudes of the noisy short-time Fourier coefficients, cf. (6). The smoothing is necessary to obtain a low variance estimate of the power spectral density.
- Time domain signal correlation as given for example in babble noise.

While the correlation originating from frame overlap is constant for a given window shape and overlap and could therefore compensated for with a constant factor, the latter two correlation types are signal dependent, requiring an adaptive compensation.

3.2. Extension of the Algorithm

As in (5), the expected short-time variance estimate can be formulated as

$$E[\widehat{\mathrm{v}ar}\{P(\lambda,k)\}] = \mathrm{v}ar\{P(\lambda,k)\} - \mathrm{v}ar\{\overline{P}(\lambda,k)\} \quad (8)$$

which suggests to compensate for the underestimation term $\mathrm{v}ar\{\overline{P}(\lambda,k)\}$. We therefore use an estimator for the compensation term which is for reasons of computational efficiency implemented again as an autoregressive estimator

$$\begin{aligned} \widehat{\mathrm{v}ar}\{\overline{P}(\lambda,k)\} &= \beta(\lambda,k)\widehat{\mathrm{v}ar}\{\overline{P}(\lambda-1,k)\} \\ &+ (1-\beta(\lambda,k))\left(\overline{P}(\lambda,k) - \overline{\overline{P}}(\lambda,k)\right)^2 \end{aligned} \quad (9)$$

with the smoothing parameter $\beta(\lambda,k)$ being equal to the coefficient used for the first order recursive average of $\overline{P^2}(\lambda,k)$ and $\overline{P}(\lambda,k)$ and $\overline{\overline{P}}(\lambda,k)$ denotes the mean over the mean values $\overline{P}(\lambda,k)$

$$\overline{\overline{P}}(\lambda,k) = \gamma(\lambda,k)\overline{\overline{P}}(\lambda-1,k) + (1-\gamma(\lambda,k))\overline{P}(\lambda,k) \quad (10)$$

	noise power estimation error
without extension, (7)	-17.9%
with extension, (11)	-7.0%

Table 1: Noise power estimation error for framewise processing with 87.5% overlap for a stationary white Gaussian noise with and without the proposed compensation term (mean over 2200 frames and 257 frequency bins).

	AWGN	babble
input SNR	0.1 dB	5.1 dB
output SNR without extension, (7)	6.4 dB	6.8 dB
output SNR with extension, (11)	7.6 dB	7.6 dB

Table 2: Segmental SNR before and after noise reduction with and without compensation term for stationary additive white Gaussian noise and for cafeteria babble noise.

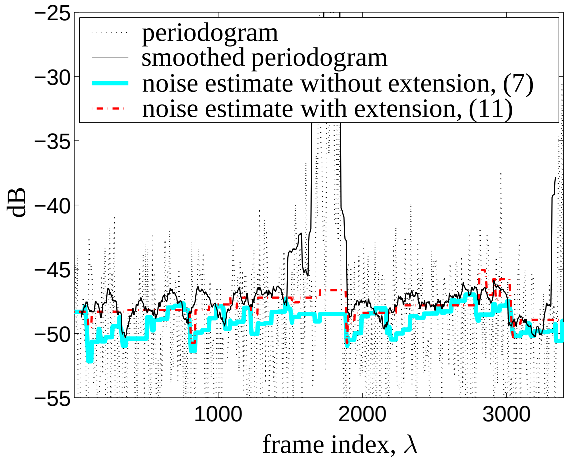

Figure 1: *Periodogram, smoothed periodogram, and noise estimate - with and without the extension - for a speech signal disturbed with stationary white Gaussian noise, regarding a single frequency bin (1340 Hz).*

with $0 < \gamma(\lambda, k) < 1$. The smoothing parameter $\gamma(\lambda, k)$ should respect the stationarity of the noisy signal. A good choice is for example $\gamma(\lambda, k) = \sqrt{\hat{\alpha}(\lambda, k)}$ where $\hat{\alpha}(\lambda, k)$ denotes the adaptive smoothing parameter in [1].
(9) constitutes a variance estimator that suffers itself from underestimation. The amount of underestimation is again determined by the smoothing parameter and the degree of correlation. While a compensation of the correlation dependent factor seems to be rather difficult we can easily compensate for the factor that depends only on the smoothing parameter. Since the variance estimator defined by (9) differs slightly from the one defined in Equations (1) to (3) the expectation of (9) turns out to have the form of (4) multiplied with the smoothing parameter, β. The modified variance estimator finally reads

$$\begin{aligned} \widehat{var}\{P(\lambda, k)\} &= \overline{P^2}(\lambda, k) - \overline{P}^2(\lambda, k) \\ &+ \frac{1 + \beta(\lambda, k)}{2\beta^2(\lambda, k)} \widehat{var}\{\overline{P}(\lambda, k)\}. \end{aligned} \tag{11}$$

Note that with Equation (9) underestimation is covered that is due to correlation not only of overlapping block processing but also of correlation that is induced by low-pass filtering the squared spectral magnitudes and correlation that may be inherent in the time domain signal.

4. EXPERIMENTAL RESULTS AND DISCUSSION

In the following we present results obtained with a single channel noise reduction system using a Wiener filter with a decision-directed *a priori* SNR estimator [6]. The analysis is performed with a Hann window of length 512 samples using a frame shift of 64 samples corresponding to 87.5% frame overlap. As shown in [4] the bias compensation function for the Minimum Statistics noise power estimator depends on the correlation of successive frames and should not use standard values if a window with greater overlap than 50% is used. Therefore the appropriate bias compensation factors B_{min} have been created prior to the subsequent investigations.

For a stationary white Gaussian noise of known variance the noise power has been estimated with the described system without and with the proposed extension (Equ. (7) and (11), respectively) while keeping all other parameters the same for both measurements. The results given in Table 1 suggest that reducing the bias of the variance estimator (7) is useful to reduce the amount of noise power underestimation as observed for correlated data. However, the proposed modification still does not result in an unbiased noise power estimator. As it has been previously argued this could be a consequence of the fact that the compensation term (9) itself constitutes a variance estimator that necessarily underestimates the true variance in presence of correlated data. To alleviate this effect the averaging length would have to be increased which is not possible for a signal with limited stationarity. Therefore, in presence of correlation, (11) improves the variance estimate but can not be expected to completely compensate for the bias.

In Table 2 we present the segmental SNR improvement for both white Gaussian noise and for cafeteria babble noise at roughly 0 dB and 5 dB input SNR, respectively. We observe that the segmental SNR after processing is in-

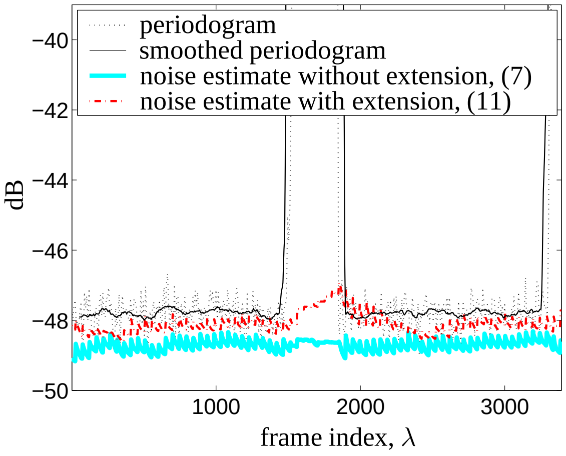

Figure 2: *Frequency mean periodogram, smoothed frequency mean periodogram, and frequency mean noise estimate - with and without the extension - for a speech signal disturbed with stationary white Gaussian noise.*

creased for the extended variance estimator (11) as compared to the case without the extension (7). This result shows that the compensation term effectively alleviates the noise power underestimation, resulting in an increased output SNR level.

Figures 1 and 2 document the temporal behavior of the noise power estimator for a noisy speech signal. The periodogram, a temporally smoothed periodogram and the noise power estimates obtained without and with the proposed extension are plotted. In Figure 1 we show the result for a single frequency bin (1340 Hz). We see that the standard noise power estimate shows a tendency to underestimate the mean noise power. Although the modified noise power estimates feature some variations which are due to the statistical nature of the compensation term, the modified noise power estimate more frequently approaches the actual noise power.
The effect becomes clearer if we observe values averaged over the frequency, Fig. 2. During speech pauses the extended noise power estimator better approximates the mean noise power. During speech activity we notice a slight increase of the modified noise power estimates.

To make the noise power estimator unbiased for a given analysis-synthesis system with significant frame overlap further measures will have to be taken. In [1] a correction factor B_c is introduced which accounts for the fact that the noise power estimator can track increasing noise power only with some delay and therefore the noise power would be underestimated for highly non-stationary noise. The factor has been empirically optimized for the algorithm in its original form and will have to be adapted for the extended algorithm described above.

5. CONCLUSIONS

We observe that the Minimum Statistics noise power spectral density estimator [1] tends to underestimate the noise power for the particular case when successive signal frames are considerably correlated. Correlation can originate from a block processing with considerable overlap, from low-pass filtering the squared spectral magnitudes or can be inherent part of the time domain noisy signal. We showed that noise power underestimation originates from an first order recursive biased variance estimator that in case of correlated data systematically underestimates the process variance. A term describing the underestimation has been identified and consequently an extension of the Minimum Statistics noise power estimator has been proposed that turned out to alleviate the underestimation effects.

Acknowledgment

This work was in part supported by grants from the European Union FP6, Project 004171 HEARCOM.

6. REFERENCES

[1] Rainer Martin, “Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics,” *IEEE Trans. on Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, July 2001.

[2] Volker Stahl, Alexander Fischer, and Rolf Bippus, “Quantile Based Noise Estimation For Spectral Subtraction and Wiener Filtering,” in *Proc. Intl. Conf. Acoustic Speech Signal Process.*, June 2000, vol. 3, pp. 1875–1878.

[3] Israel Cohen, Baruch Berdugo, “Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement,” *IEEE Signal Processing Letters*, vol. 9, no. 1, pp. 12–15, January 2002.

[4] Rainer Martin, “Bias Compensation Methods For Minimum Statistics Noise Power Spectral Density Estimation,” *Signal Processing*, vol. 86, no. 6, pp. 1215–1229, June 2006.

[5] Rainer Martin and Thomas Lotter, “Optimal Recursive Smoothing of non-stationary Periodograms,” in *Proc. Intl. Workshop Acoustic Echo and Noise Control (IWAENC)*, September 2001, pp. 167–170.

[6] Yariv Ephraim and David Malah, “Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-32, no. 6, pp. 1109–1121, December 1984.