

# Noise Variance Estimation In Signal Processing

David Makovoz

IPAC, California Institute of Technology, MC-220, Pasadena, CA, 91125

[davidm@ipac.caltech.edu](mailto:davidm@ipac.caltech.edu); 626-395-3521

**Abstract**—We present a new method of estimating noise variance. The method is applicable for 1D and 2D signal processing. The essence of this method is estimation of the scatter of normally distributed data with high level of outliers. The method is applicable to data with the majority of the data points having no signal present. The method is based on the shortest half sample method. The mean of the shortest half sample (shorth) and the location of the least median of squares are among the most robust measures of the location of the mode. The length of the shortest half sample has been used as the measurement of the data scatter of uncontaminated data. We show that computing the length of several sub samples of varying sizes provides the necessary information to estimate both the scatter and the number of uncontaminated data points in a sample. We derive the system of equations to solve for the data scatter and the number of uncontaminated data points for the Gaussian distribution. The data scatter is the measure of the noise variance. The method can be extended to other distributions.

**Index Terms**—Noise variance estimation, nonlinear filters, robust estimation, scatter estimation

## I. INTRODUCTION

Noise estimation is a major task in all areas of signal processing, be it speech or image processing. Signal processing algorithms for segmentation, clustering, restoration, noise reduction, statistical inference etc, depend on the knowledge of the noise variance. The literature on the noise variance estimation in speech and images abounds [1]-[7]. WE present a new algorithm that uses very few assumption about the data, namely that the noise has Gaussian distribution and that the majority of the data points in the data set have no signal present.

In signal processing in general one deals with noisy data  $z$ , where each data point  $i$  it is a combination of the clean signal  $s_i$  is the clean signal and  $v$  is the noise:  $z_i = s_i + v_i$ . In many applications the data contain a number of data points for which the signal is either not present or much smaller than the noise. In (a)

(b) **Figure 1** we show two illustrations of such data: a noisy speech waveform and an astronomical image. The distribution of the data points consists of the noise distribution and the noisy signal distribution. The noise distribution in many cases is or can be approximated by the Gaussian distribution. If one to consider this distribution from the point of view of estimating the noise power, i.e. the width of the Gaussian distribution, then the noise data points become the useful data and the noisy speech data points become outliers present in

the useful noise data. If the number of pure noise data points is greater than the number of noisy signal data points, then one can apply the method developed below to estimate the noise power without doing any explicit separation of the noise from the noisy signal.

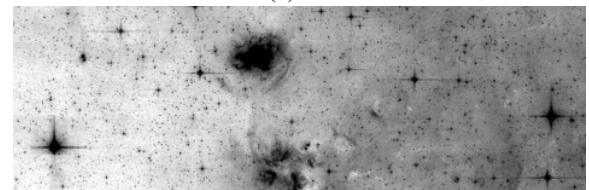
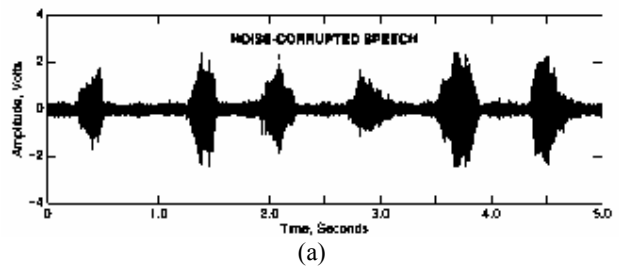


Figure 1. Examples of (a)1D data and (b) 2D data, in which the number of signal data points is smaller than the number of background noise only data points.

In Section II we present the background on data scatter estimation in outlier contaminated data. In Section III we introduce the new method and derive the algorithm for computing the scatter of normally distributed outlier contaminated data. In Section IV we present results of the simulations.

## II. BACKGROUND

Estimation of the peak and the scatter of data in a sample is a common problem encountered in many diverse areas of statistical data processing. If the data are known to have Gaussian distribution, the most common estimators of the peak and the scatter are the mean and the standard deviation of the data around the mean. If there are outliers present in the data sample, mean-based estimators break down almost immediately; even one outlier can result in a completely misguided mean. The same is true about the standard deviation from the mean as an estimator of the data scatter. A more robust estimator of the peak is the median. But even the median erodes as the number of outliers is increasing and approaches 50% of the sample size.

There exist two general approaches in dealing with outlier contaminated data. The first approach, and our method

belongs to this group, is to deal with the whole sample and devise robust estimators which are to a great extent insensitive to the presence of outliers. The second approach is to devise robust methods of identifying and excluding outliers and then to treat the uncontaminated sample with the conventional statistical methods.

Our method is built on one of the existing methods of mode estimation. The mode as an estimator of the peak of a distribution is very robust; it is mostly insensitive to up to 50% outliers in the data sample. However, whereas computing mean and median is straightforward and both estimators have a unique value for any given data sample, mode estimation is notoriously difficult and moreover for multimodal distributions there is no unique mode. There exist a whole class of mode estimators based on the notion [8,9] of the shortest half sample. The *shorth* – the mean of the shortest half sample – was proposed in [8]. In [9] it was shown that the location of the one-dimensional least median of squares, which is the mean of the minimum and maximum data points of the shortest half sample can be used as a robust estimator of the mode of a data sample. This estimator has a higher bias than the shorth. A low biased variant of mode estimator was reported in [10]. It is computed by repeatedly taking the shortest half samples within shortest half samples.

In [9] it was proposed to use the length of the shortest half sample as a robust estimator of the data scatter. However, whereas the middle point of the shortest half sample is not sensitive to the presence of outliers in the data sample, the length of the shortest half sample depends on the number of outliers. In the presence of outliers only  $N_{eff}$  data points out of the total of  $N$  actually belong to the parent distribution. The half sample is only such with respect to all points in the sample, but it is more than half-sample with respect to the points from the parent distribution with the outliers excluded. Since the scatter estimate depends critically on the fact that the shortest half sample actually encompasses half of the uncontaminated data points, it becomes meaningless in presence of outliers.

The novelty of our approach is that we derive a way of simultaneously estimating both the data scatter and the effective size of the sample  $N_{eff}$ . This allows us to estimate the data scatter for outlier contaminated data.

### III. METHOD

#### A. Multiple Shortest Subsamples

We start with the noisy data points  $z_i$  and sort them in the ascending order. We will notation  $x_i$  for the sorted data points. Let  $X_N = (x_1 \dots x_N)$  be an ordered sample of size  $N$ . In order to find the shortest sub sample consisting of  $n$  data points one finds  $i=m$  that minimizes  $(x_{n+i} - x_i)$ , where  $i = 1, \dots, N-n$ .

We introduce the fractional sub sample size  $r = n/N$ . We estimate the mode by the median of the shortest sub sample  $Mode(r) = x_{m+n/2}$ .

The mode estimator  $Mode(r)$  is applicable for any

distribution. The scatter estimator, that we derive here, is applicable only for the data that has the Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ :

$$f(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (1)$$

For the Gaussian distribution the value of  $Mode(r)$  is an unbiased estimator of  $\mu$ . If the shortest sub sample is identified as described above, then the length of the shortest sub sample is:

$$L(r) = x_{m+n} - x_m$$

The fractional sub sample size  $r$  approximates the integral of the distribution function from point  $x_m$  to point  $x_{m+n}$  (see Fig. 1). Therefore, to the extent that  $Mode(r)$  gives the correct estimate of the peak of the Gaussian, the following relation involving the error function  $erf$  holds between the sub sample fractional size  $r$  and its length  $L$ :

$$r = \frac{N_{eff}}{N} \frac{2}{\sqrt{\pi}} \int_0^{L(r)/\sigma\sqrt{2}} \exp(-t^2) dt = \frac{N_{eff}}{N} \cdot erf\left(\frac{L(r)}{2\sigma}\right) \quad (2)$$

The critical fact here is that the normalization factor depends on the size  $N_{eff}$  of the uncontaminated sample, i.e. the outliers should be excluded from this count. If one knows  $N_{eff}$ , for example if there are no outliers, then in order to compute  $\sigma$  one simply inverts the equation 2.

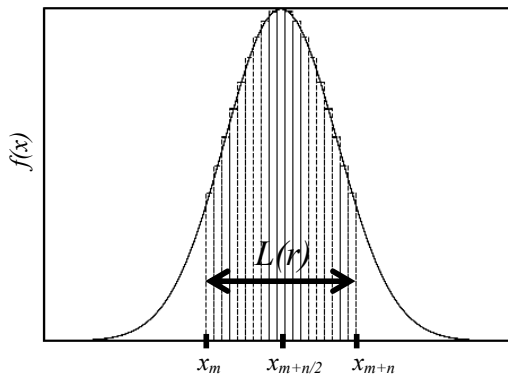


Figure 2 The shaded area, defined as the integral of  $f(x)$  from  $x_m = Mode(r) - L(r)/2$  to  $x_{m+n} = Mode(r) + L(r)/2$ , is equal to  $r$ .

If  $N_{eff}$  is not known, which is the case under consideration, one can find  $L(r_s)$  for several values of  $r_s$  and obtain a system of equations to solve for both  $N_{eff}$  and  $\sigma$ .

We introduce notation  $MSL(r_s)$  for the estimator of the scatter  $\sigma$  of the normally distributed data. ( $MSL$  stands for multiple shortest lengths.) We will call the set of fractional sizes  $r_s$  the support of the estimator  $MSL(r_s)$ .

#### B. Derivation of the Equation for Scatter Estimation

Since  $MSL(r_s)$  depends on several parameters, there is a certain degree of freedom in selecting the most effective way of computing it. The more straightforward way is to find  $L(r_s)$  for 2 values of  $r_s$  and solve the system of two equations for two unknowns:

$$\begin{aligned}
r_1 &= u \cdot \text{erf}(L(r_1)v) \\
r_2 &= u \cdot \text{erf}(L(r_2)v) \\
u &\equiv \frac{N_{\text{eff}}}{N}, v \equiv \frac{1}{2\sigma}
\end{aligned} \tag{3}$$

Factoring out  $u$  leads to the following equation for  $v$ :

$$r_2 \cdot \text{erf}(L(r_1)v) = r_1 \cdot \text{erf}(L(r_2)v). \tag{4}$$

Another approach to finding  $N_{\text{eff}}$  and  $\sigma$  is by to perform the least-square fit to the data. Quantities  $r_s$  and  $L(r_s)$  are measured for  $S$  sub samples. The following quantity is minimized:

$$\sum_{s=1}^S (r_s u - \text{erf}(L(r_s)v))^2 \tag{5}$$

Minimization with respect to  $u$  and  $v$

$$\begin{aligned}
\frac{\partial}{\partial u} \sum_{s=1}^S (r_s u - \text{erf}(L(r_s)v))^2 &= 0; \\
\frac{\partial}{\partial v} \sum_{s=1}^S (r_s u - \text{erf}(L(r_s)v))^2 &= 0,
\end{aligned} \tag{6}$$

leads to the following set of equations

$$\begin{aligned}
\sum_{s=1}^S r_s (r_s u - \text{erf}(L(r_s)v)) &= 0; \\
\sum_{s=1}^S L(r_s) \exp(-(L(r_s)v)^2) (r_s u - \text{erf}(L(r_s)v)) &= 0.
\end{aligned} \tag{7}$$

Again, factoring out  $u$  leads to the following equation for  $v$ :

$$\begin{aligned}
\sum_{s=1}^2 r_s \text{erf}(L(r_s)v) \sum_{s=1}^2 r_s L(r_s) \exp(-(L(r_s)v)^2) - \\
\sum_{s=1}^2 r_s^2 \sum_{s=1}^2 L(r_s) \exp(-(L(r_s)v)^2) \text{erf}(L(r_s)v) &= 0.
\end{aligned} \tag{8}$$

The solution of either system of equations 3 or 7 is subject to the obvious constraint:

$$u \leq 1. \tag{9}$$

The way we apply this constraint is to solve the system first, and if  $u > 1$ , then set  $u = 1$  and compute  $v$  as

$$v = \frac{1}{2} \sum_{s=1}^2 \frac{\text{erf}^{-1}(r_s)}{L(r_s)} \tag{10}$$

Based on the simulations we concluded that the only gain achieved by using the second approach is in the execution time. Below we present only results obtained by solving the equation 4.

Here is the summary of the first approach. One finds the lengths  $L(r_1)$  of the shortest subset of  $r_1 \cdot N$  data points and  $L(r_2)$  of the shortest subset of  $r_2 \cdot N$  data points. The values of  $L(r_1)$ ,  $L(r_2)$ ,  $r_1$ , and  $r_2$  are used to solve equation (4) for  $v$ , which is directly related by equation (3) to the variance  $\sigma^2$ .

### C. Further Refinement-Automated MSL

Once  $MSL(r_s)$  and  $N_{\text{eff}}$  are found a further refinement is possible. We show in Section III based on the results of simulations, that in general the greater  $r_s$  is, the more accurate are the  $MSL_S(r_s)$  and  $N_{\text{eff}}$  estimators.

We do iterative refinement using the following the strategy. We start with reasonably small  $r_s$ , measure  $L(r_s)$ , and compute  $MSL(r_s)$  and  $N_{\text{eff}}$ . The values  $r_s$  are incremented and  $MSL(r_s)$  and  $N_{\text{eff}}$  are recomputed until at least one of the three stopping criteria is met. Two empirical parameters  $R_{\text{max}}$  and  $dR_{\text{min}}$  are used to define stopping criteria of the algorithm. The iteration terminates if  $N_{\text{eff}}/N$  exceeds  $R_{\text{max}}$ , or if the change in  $N_{\text{eff}}/N$  between two iterations drops below  $dR_{\text{min}}$ . Another stopping criterion is based on the assumption that the change in the  $Mode(r)$  between two consecutive iterations cannot exceed the value of the scatter  $MSL(r_s)$ . The mode is found for the biggest  $r_s$  in the set. For example, if the support is  $r_s = (0.4; 0.5)$ , then the mode is found for  $r_2 = 0.5$ .

We use notation  $AMSL(r_s)$  (automated  $MSL(r_s)$ ) for the scatter estimator obtained using this iterative process. Support  $r_s$  in this case refers to the initial set of values of the fractional sub sample sizes. The results of the simulations presented below were obtained using the following empirical values of the parameters  $R_{\text{max}} = 0.95$ ,  $dR_{\text{min}} = 0.1$ .

## IV. SIMULATIONS

### A. Details of the Simulations

In order to test the estimators derived in the previous section we generated several sets of clean and contaminated data. The uncontaminated data are zero-mean Gaussian with the standard deviation of 1. We run our simulations for three sample sizes of 30, 100, and 1000 data points. In order to compute the average and standard deviations of the estimators  $MLS$  and  $AMLS$  we repeated the simulations 300000 times for the sample size 30, 100000 for the sample size of 100, and 10000 times for the sample size of 1000.

We generated two kinds of outliers that were added to replace the data points from the main distribution. The first kind consisted of data points uniformly distributed from 3 to 8. The second kind consisted of data points normally distributed with the same standard deviation of 1 as the main data and the mean of 4 (99.9937 percentile). The outlier fraction  $F = 1 - N_{\text{eff}}/N$  used for the uniformly distributed data was 0.2, 0.3, 0.4, and 0.5. The outlier fraction  $F$  used for the normally distributed outliers was 0.2, 0.3, and 0.4. The ratio of  $F = 0.5$  obviously could not be used for the normally distributed outliers, since in this case they are no longer outliers. In Fig. 2 we show two histograms for two extreme cases of contaminated data: one sample of data with the uniform outliers with the fraction  $F = 0.5$  and one sample of data the Gaussian outliers with the fraction  $F = 0.4$ ; for the sample size in both cases is  $N = 1000$ .

For comparison we also computed the median absolute deviation ( $MAD$ ) [1]. The value of  $MAD$  is a measure of the scale or dispersion of a distribution about the median. It is

often calculated as the median of the absolute-value distances of the points about the median:  $MAD = median\{|x_i - median(x)|\}$  and multiplied by a factor of 1.4826 to achieve consistency with the standard deviation for asymptotically normal distribution.

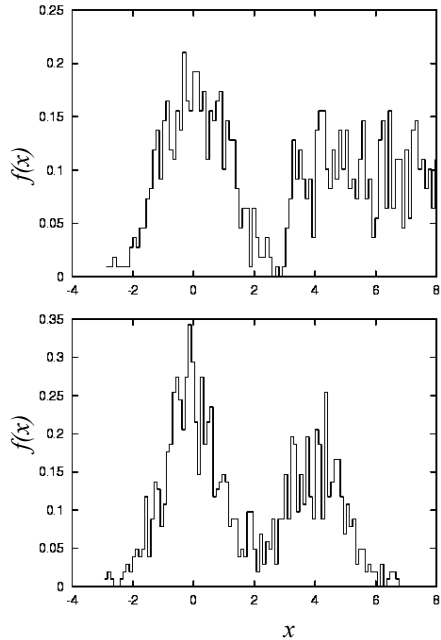


Figure 3 The top figure shows a histogram of a contaminated sample with the uniformly distributed outliers with the fraction  $F = 0.5$ ; the bottom figure shows a histogram of a contaminated sample with the normally distributed outlier with the fraction of  $F = 0.4$ . The sample size in both cases is 1000.

### B. Simulations of Contaminated Data

Now we consider outlier contaminated data. First, in Table 2 we show the results for the  $MSL$  as a function of the contamination fraction  $F$ . We also present the results for the relative uncontaminated size  $T_{eff}$ :

$$T_{eff} = \frac{N_{eff}}{N_{eff}^{true}}$$

$$N_{eff}^{true} = (1 - F) \cdot N$$

To compute these quantities we use the support  $r_s = (0.4; 0.5)$ , which is the biggest support one can use with no a priori knowledge of the outlier level, only assuming that it does not exceed 50% of the sample size. The results for lower levels of contamination display a higher bias and dispersion. The closer is the sub sample size to the size of the uncontaminated sample, the better are the results for the  $MSL$  and  $T_{eff}$ .

The refined estimator  $AMSL$  is devised with the idea of improving the accuracy of estimation for lower levels of contamination. As part of the algorithm the support is increased to come as close as possible to the size of the

uncontaminated data without overstepping that boundary and without including the outliers in the sub samples defined by the support  $r_s$ . In Table 3 we give the results of the simulations for the initial support  $r_s = (0.25; 0.35)$ . The goal is achieved, as  $AMSL$  is a better estimator for the lower outlier levels and is very close to the  $MSL$  for the highest outlier levels.

We illustrate the results for the  $MSL$  and  $AMSL$  estimators in Fig. 4. We also display the values obtained for the  $MAD$  estimator for comparison. Only the estimators computed for the uniform outlier distribution are plotted. The results for both Gaussian and uniform outliers are very similar, the only difference is that for the uniform outliers the  $MSL$  and  $AMSL$  estimators can be computed in the extreme case of 50% outliers.

### LIMITATIONS

The method developed here has certain limitations. One of them is that the equation (4) does not always have a solution for a small sample size. The failure rate becomes negligible for the sample size  $> 100$  data points.

The second limitation is the case of highly contaminated data with the outliers having a more dense distribution than the main data. In this case the mode of the outliers could be picked over that of the main data.

Further investigation is planned to establish a well defined limitations of this method and also to extend it to the other than Gaussian distributions.

### CONCLUSIONS

We derived a new method of estimating the scatter of normally distributed data with high levels of contaminations. Our method is very stable and performs well for the fraction of outliers of up to 50%. This method can be applied to estimating noise variance in noisy data, where the number of data points containing only noise is greater than the number of data points containing both signal and noise.

	Sample Size	Fraction $F$ of Uniform Outliers				Fraction $F$ of Gaussian Outliers		
		0.2	0.3	0.4	0.5	0.2	0.3	0.4
$MSL(r_s)$	30	0.74 (0.32)	0.82 (0.37)	0.89 (0.40)	0.96 (0.34)	0.74 (0.32)	0.83 (0.37)	0.90 (0.41)
	100	0.85 (0.27)	0.93 (0.30)	0.97 (0.30)	0.98 (0.15)	0.85 (0.27)	0.93 (0.30)	0.98 (0.30)
	1000	0.97 (0.16)	0.98 (0.14)	0.98 (0.09)	1.02 (0.05)	0.97 (0.16)	0.99 (0.14)	1.00 (0.10)
$T_{eff}$	30	0.90 (0.23)	0.98 (0.25)	1.03 (0.24)	1.04 (0.10)	0.90 (0.23)	0.98 (0.25)	1.04 (0.25)
	100	0.94 (0.20)	1.01 (0.21)	1.05 (0.19)	1.02 (0.03)	0.95 (0.20)	1.01 (0.21)	1.06 (0.19)
	1000	0.99 (0.12)	1.00 (0.10)	1.00 (0.05)	1.02 (0.006)	0.99 (0.12)	1.00 (0.10)	1.01 (0.05)

Table 1 . The average (standard deviation) of the scatter  $MSL(r_s)$  and relative uncontaminated size  $T_{eff}$  estimators for outlier contaminated samples. The support  $r_s = (0.4; 0.5)$ .

	Sample Size	Fraction $F$ of Uniform Outliers				Fraction $F$ of Gaussian Outliers		
		0.2	0.3	0.4	0.5	0.2	0.3	0.4
$AMSL(r_s)$	30	0.82 (0.33)	0.92 (0.39)	1.01 (0.49)	1.05 (0.65)	0.82 (0.33)	0.93 (0.39)	1.02 (0.49)
	100	0.98 (0.28)	1.06 (0.34)	1.09 (0.38)	1.08 (0.50)	0.98 (0.28)	1.06 (0.34)	1.11 (0.40)
	1000	1.01 (0.20)	1.03 (0.21)	1.01 (0.18)	0.99 (0.09)	1.02 (0.19)	1.03 (0.21)	1.01 (0.19)
$T_{eff}$	30	1.02 (0.25)	1.11 (0.28)	1.16 (0.32)	1.17 (0.37)	1.02 (0.25)	1.11 (0.28)	1.17 (0.33)
	100	1.06 (0.21)	1.12 (0.25)	1.15 (0.29)	1.09 (0.28)	1.06 (0.21)	1.12 (0.25)	1.17 (0.29)
	1000	1.03 (0.16)	1.04 (0.17)	1.02 (0.14)	1.00 (0.05)	1.03 (0.16)	1.04 (0.17)	1.03 (0.14)

Table 2. The average (standard deviation) for the scatter  $AMSL(r_s)$  and relative uncontaminated size  $T_{eff}$  estimators computed using the refined strategy. The initial support was  $r_s = (0.25; 0.35)$ .

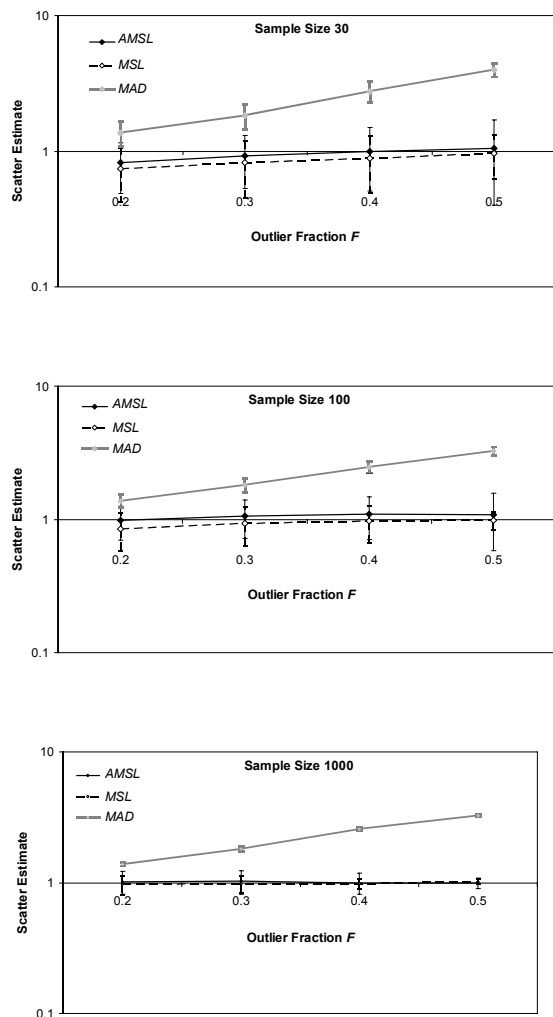


Figure 4. Three estimators of the data scatter: *AMSL*, *MSL*, and *MAD* as functions of the outlier fraction  $F$ . *MSL* is computed with the support  $r_s = (0.4; 0.5)$ , and *AMSL* is computed with the initial support  $r_s = (0.25; 0.35)$ .

#### REFERENCES

- [1] K.K. Paliwal, "Estimation of noise variance from the noisy ar signal and its application in speech enhancement", *IEEE Transactions on Acoustics, Speech and Signal Processing*, **36**, 292, 1988
- [2] H. Attias, L. Deng, A. Acero, and J. C. Platt, "A new method for speech denoising and robust speech recognition using probabilistic models for clean speech and for noise", In *Proc. Eurospeech*, **3**, 1903, 2001
- [3] K. Konstantinides, B. Natarajan, and G.S. Yovanof, "Noise Estimation and Filtering Using Block-Based Singular Value Decomposition", *IEEE Transactions of Image Processing*, **6**, 479, 1997

[4] J.L. Starck and F. Murtagh, "Automatic Noise Estimation from the Multiresolutional Support", *Publications of the Astronomical Society of the Pacific*, **110**, 193, 1998

[5] M. Jansen, *Noise Reduction by Wavelet Thresholding*, Springer-Verlag New York Inc., 2001

[6] A.B. Hamza and H. Krim, "Image Denoising: A Nonlinear Robust Statistical Approach", *IEEE Transactions on Signal Processing*, **49**, 3045, 2001

[7] L. Alparone, G. Corsini, M. Diani, "Noise modeling and estimation in image sequences from thermal infrared cameras", *Optics in Atmospheric Propagation and Adaptive Systems VII, Proceedings of the SPIE*, **5573**, 381, 2004

[8] D.F. Andrews, P.J. Bickel, F.R. Hampel, P.J. Huber, W.H. Rogers, and J.W. Tukey, *Robust Estimates of Location*, Princeton University Press, Princeton, N.J., 1972

[9] P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*, John Wiley & Sons, New York, NY, 1987

[10] D. R. Bickel, "Robust estimators of the mode and skewness of continuous data," *Computational Statistics and Data Analysis* **39**, 153-163 (2002)