A NEW APPROACH FOR REDUCTION OF SUPERGAUSSIAN NOISE USING AUTOREGRESSIVE INTERPOLATION AND TIME-FREQUENCY MASKING

[1]*Marco Ruhland,* [1]*Stefan Goetze,* [2]*Matthias Brandt,* [1,3]*Simon Doclo,* [1,2]*Joerg Bitzer*

[1]Fraunhofer IDMT, Project group Hearing, Speech and Audio Technology, Oldenburg, Germany
[2]Jade University of Applied Sciences, Institute for Hearing Technology and Audiology, Oldenburg, Germany
[3]Carl von Ossietzky University, Signal Processing Group, Oldenburg, Germany
{marco.ruhland, s.goetze}@idmt.fraunhofer.de, {matthias.brandt, joerg.bitzer}@jade-hs.de, simon.doclo@uni-oldenburg.de

ABSTRACT

A new approach for noise reduction is presented. The method is capable of reducing noise of Gaussian, supergaussian and impulsive characteristics in degraded high-quality audio signals. The approach is based on classical autoregressive (AR) detection and interpolation, applied to the residual signal of a binary time-frequency (T-F) masking process. Analytic inspection allows for predicting the noise reduction level for white noise types and shows good accordance to simulation results. High reduction levels are achieved especially for supergaussian and impulsive disturbances having a higher sample kurtosis than Gaussian noise. The approach ensures high preservation of the underlying desired signal, satisfying the needs of high quality audio restoration. Furthermore, the approach is capable of reducing optical soundtrack noise of celluloid movie footage.

Index Terms— Interpolation, Noise Reduction, Optical Soundtrack Noise, Supergaussian, Time-Frequency Masking

1. INTRODUCTION

The term "noise reduction" covers several types of disturbances. For suppression of hiss noise in audio recordings, spectral attenuation methods are used, like the well-known Wiener filter or the Ephraim-Malah method [1]. Since the Gaussian assumption of these methods does not hold for most desired signals [2], as well as for the noise disturbances, more sophisticated methods have been developed [3, 4]. For the removal of impulsive disturbances, like e.g. clicks caused by dust and scratches on a gramophone disc, pioneering work has been done by Vaseghi [5] and others [6, 7]. The hiss reduction methods have in common to work in the frequency domain, while the methods for impulse reduction are mostly based on interpolation in the time domain, due to the sparse, localized occurrence of the disturbances in the time signal. But the border between those two types of disturbances is not strict. For example, the noise of a heavy rainfall could be imagined as a vast number of small clicks per time instant, giving a grainy noise somewhere between hiss and impulses. Optical soundtrack noise, caused by dust, mould, or bad exposure of celluloid film footage [8] can offer similar characteristics. Motivated by the latter, this contribution investigates the effect of time-domain interpolation techniques on noise signals with supergaussian properties.

The remainder of this paper is organized as follows. Section 2 presents the idea behind the new approach, being a combination of T-F masking and AR detection and interpolation. In Section 3, an analytic examination is elaborated to predict the amount of noise reduction. The analytic solution is compared to simulation results. Conclusions are drawn in Section 4.

This work was supported in part by the German Ministry of Education and Research (BMBF) and the European Commission under grant no. 16SV5490, Project ALIAS (M. Ruhland, S. Goetze), and in part under BMBF grant no. 17N3008 (M. Brandt, J. Bitzer). The views and conclusions contained in this document, however, are those of the authors.

2. THEORY

The idea behind the new denoising approach is, to first separate a desired audio signal from the noise signal as good as possible, then to apply AR detection and interpolation on the extracted noise, and then to add the estimated desired signal back to the interpolated noise signal to form the restored signal. The separation task can be done using T-F masking. Wang gives in his article [9] an excellent overview on this topic. In [10], the *ideal binary mask* (IBM) is proposed, calculated from the local signal-to-noise ratio. Of course, in real-world applications, the instantaneous SNR is not known, so that the IBM has to be estimated. Several binary mask estimation techniques are compared in [11]. The imperfectness of the estimation justifies the use of AR detection and interpolation for the reduction of the noisy residual signal. Since the binary mask will not be perfect in real applications, components of the desired signal will fall into the residual signal erroneously. These components will be preserved by the AR interpolation process, while the noisy components will be suppressed. If the binary mask would work perfectly, the detection and interpolation process would not be necessary.

2.1. Detection Algorithm

The task of a detection algorithm is to figure out the location of damaged samples as exactly as possible. A low false alarm rate is required for the preservation of unaffected desired signal parts. Kauppinen [12] shows that the AR method works best in terms of missing detection rate and lowest false alarm rate compared to other methods. The AR detection method originally has been introduced in [13, 14] and is also recommended in common audio restoration literature [6, 15]. A clean audio signal $x(n)$ is modelled by an autoregressive process,

$$x(n) = \sum_{m=1}^{P} a_m x(n-m) + \epsilon(n) \quad , \tag{1}$$

with a_m being the AR coefficients, n the discrete time index, P the AR model order and $\epsilon(n)$ being a white Gaussian excitation signal. This corresponds to filtering the excitation $\epsilon(n)$ by an all-pole filter $A(z)$:

$$A(z) = \frac{1}{1 - \sum_{m=1}^{P} a_m z^{-m}} \quad . \tag{2}$$

The AR coefficients are estimated from a signal block using the Yule-Walker method or the Burg method (see e.g. [16]). Eq. (1) can also serve as a predictor for a disturbed signal $y(n)$, being the sum of the desired signal $x(n)$ and a noise disturbance $d(n)$. Then, $\epsilon(n)$ is replaced by the prediction error $e(n)$:

$$y(n) = \sum_{m=1}^{P} a_m y(n-m) + e(n) \quad . \tag{3}$$

Vice versa, the prediction error $e(n)$ can be calculated from the disturbed signal $y(n)$:

$$e(n) = y(n) - \sum_{m=1}^{P} a_m y(n-m) \quad . \tag{4}$$

As long as there is no impulse disturbance in $y(n)$, the predicition error $e(n)$ will be low. But if an impulse occurs, the misfit in prediction to the underlying signal $x(n)$ will yield a high error signal. Usually, clicks are detected by thresholding the squared error $e^2(n)$. AR model orders can be quite low for good detection results, e.g. $P = 10$ (see [12]). Since the underlying audio signal $x(n)$ is unknown, the AR coefficient have to be estimated using the corrupted signal $y(n)$. Thanks to the Yule-Walker and Burg estimation techniques being robust towards impulsive disturbances, this is not a problem.

2.2. Interpolation Algorithm

Several interpolation methods exist to replace the corrupted samples identified by the detection stage. The least-squares AR-based (LSAR) interpolator, proposed in [17] and [18], is a very successful tool for audio restoration (cf. [19] and others). By expressing Eq. (4) in a matrix/vector form, and minimizing the sum of squared errors, the following interpolator equation is obtained (cf. [6, 18] and others):

$$\mathbf{y}_\mathrm{u}^\mathrm{LS} = -(\mathbf{A}_\mathrm{u}^T \mathbf{A}_\mathrm{u})^{-1} \mathbf{A}_\mathrm{u}^T \mathbf{A}_\mathrm{k} \mathbf{y}_\mathrm{k} \quad . \tag{5}$$

In Eq. (5), $\mathbf{y}_\mathrm{k}$ denotes a column vector containing the "known" samples of the audio signal, i.e. the samples within a column vector $\mathbf{y}$ of length N that are identified as not disturbed by the detection algorithm, in order of appearance, and $\mathbf{y}_\mathrm{u}^\mathrm{LS}$ is the solution for the "unknown" or defective samples of $\mathbf{y}$ in the least-squares sense. Fig. 1 shows how the matrices $\mathbf{A}_\mathrm{k}$ and $\mathbf{A}_\mathrm{u}$ are made up by column-wise partitioning of a $(N-P) \times N$ matrix $\mathbf{A}$, holding the AR coefficients, according to the positions of known and unknown samples within a signal block.

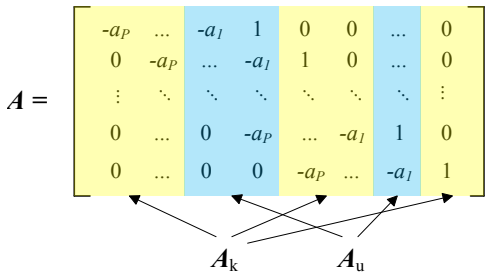

Fig. 1. Partitioning of the AR coefficient matrix $\mathbf{A}$ into matrices $\mathbf{A}_\mathrm{k}$ and $\mathbf{A}_\mathrm{u}$, according to the positions of known and unknown samples in a signal block.

2.3. AR Detection and Interpolation within a Noise Signal having a White Spectrum

If AR coefficients are drawn from a white process, the corresponding all-pole filter transfer function $A(z)$ of Eq. (2) will be equal to one, since the assumed excitation signal $\epsilon(n)$ is also white (The gain is inherent to the excitation signal $\epsilon(n)$, respectively $e(n)$, see [20]). So all the coefficients $a_1, ..., a_P$ remain zero. Since common AR detection works by thresholding the squared error signal $e^2(n)$ (see Section 2.1), a look at Eq. (4) reveals that for the white case $e(n)$ is equal to the signal $y(n)$. That means that the noise samples having the highest magnitude will be detected as disturbance samples.

For the LSAR interpolator the following holds. Inside the AR coefficient matrix shown in Fig. 1, all the coefficients are zero, and only the secondary diagonal, containing the ones, remains. So the product $\mathbf{A}_\mathrm{u}^T \mathbf{A}_\mathrm{k}$ in Eq. (5) will always end up as a zero matrix, and the term $(\mathbf{A}_\mathrm{u}^T \mathbf{A}_\mathrm{u})^{-1}$ as a unity matrix (except for some special cases, like e.g. if all unknown samples are within the first $P-1$ samples of the block [21]). So for the ideal white case, the interpolator result $\mathbf{y}_\mathrm{u}^\mathrm{LS}$ will be a vector of zeroes.

2.4. Noise Manipulation

Let us consider a white noise signal as the sum of a vast number of small impulses. Then it would be interesting to know, how the detection and interpolation algorithm would change statistical properties of the noise signal. Fig. 2 (top) shows a histogram of a white noise

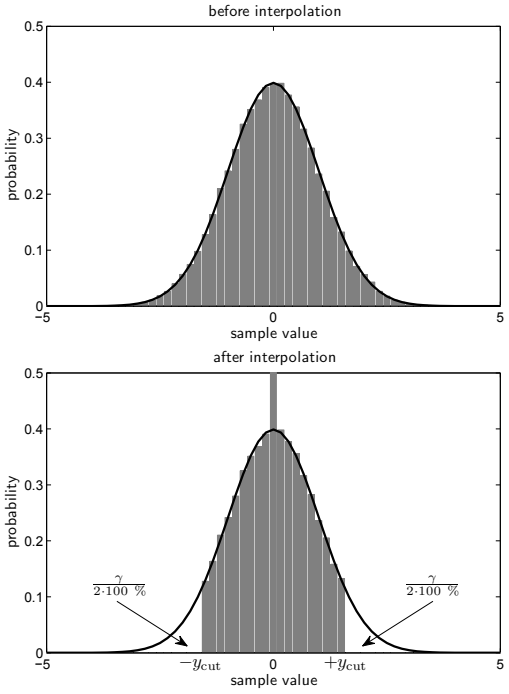

Fig. 2. Histogram of a Gaussian white residual signal before (top) and after (bottom) AR detection and interpolation. The grey bars represent simulated data, the black curves show the theoretical Gaussian probabilty density function.

signal before detection and interpolation. After detection and interpolation to a certain amount, the histogram is concatenated from both sides (same figure, bottom). This is a result of the behaviour

described in Section 2.3. The concatenation of the histogram expresses a reduction of the noise level, since the statistical variance measure represents the level of energy of a random signal. The amount of noise reduction can be predicted analytically, as will be shown in Section 3. We can have direct influence on the amount of noise reduction, by introducing a *detection percentage* γ rather than the state-of-the-art detection threshold described in Section 2.1. The percentage γ expresses how many percent of the samples of a signal block shall be detected as impulsive, and thereby be interpolated. Detection in this case means, taking the γ % samples of the signal block having the highest squared error $e^2(n)$. A value of $\gamma = 0$ % would leave all samples unaffected (no concatenation of the histogram), while a value of $\gamma = 100$ % would cause all the samples of a block to be interpolated, i.e. setting all the samples to zero, and therefore leaving the histogram as a single high peak at sample value zero.

3. ANALYTIC REVIEW

In the case of a white residual signal, it is possible to predict the amount of noise reduction [21]. As mentioned before, the power of a zero-mean random process is given by its variance. To determine the noise power after interpolation, we have to calculate the variance of the concatenated histogram of Fig. 2 (bottom), in dependency of the percentage γ and the underlying probability density function (PDF) of the noise signal. The concatenation borders $-y_{\mathrm{cut}}$ and $+y_{\mathrm{cut}}$ are given by

$$y_{\mathrm{cut}} = -\Phi^{-1}\left(\frac{\gamma}{2 \cdot 100\,\%}\right) \quad , \tag{6}$$

with $\Phi^{-1}(\cdot)$ being the inverse cumulative density function (ICDF) of the noise, and γ being the percentage parameter [21]. Then, the variance σ^2 of the signal can be calculated as

$$\sigma^2 = \int_{-y_{\mathrm{cut}}}^{y_{\mathrm{cut}}} y^2 \varphi(y) dy \quad , \tag{7}$$

with $\varphi(\cdot)$ being the PDF of the noise process. Finally, the corresponding noise level L_{noise} in dB is given by

$$L_{\mathrm{noise}} = 10 \log_{10}(\sigma^2) \quad . \tag{8}$$

The analytic prediction was verified by a simulation, using four different noise types, each having a flat spectrum:

- Gaussian
- Laplace
- modified Cauchy with density parameter $\eta = 0.02$
- modified Cauchy with density parameter $\eta = 0.002$.

The perceived sound of these noise types ranges from "smooth" (Gaussian) over "sharp" (Laplace) up to "grainy" or "impulsive" (modified Cauchy types). The modified Cauchy noise was generated by a script based on the α-stable random number generator from [22]. See the appendix for the PDFs and ICDFs. The noise signals shall be regarded as residual signals after a binary masking process. They were interpolated using the proposed detection and interpolation scheme, and the detection percentage γ was increased from 0 % up to 100 %. Fig. 3 shows the simulation results and the analytic solution. Deviations towards higher percentage levels γ result from the low block length of $N = 2048$ samples. A good agreement between analytic and simulation is given within the preferred working range of γ of 0 % to 30 %. Values above $20 - 30$ % exhibit too much musical noise with state-of-the-art BM techniques, so that there is no need to increase the block length, which is of advantage for low-latency implementations.

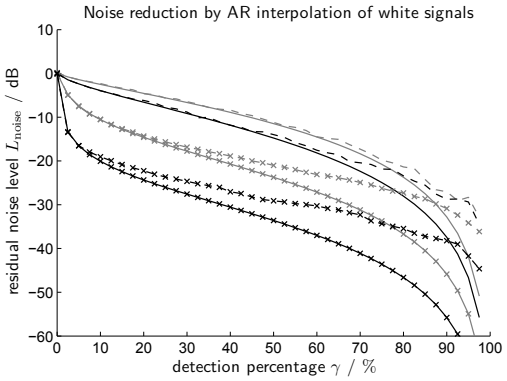

Fig. 3. Simulated results (dashed lines) vs. analytic solution (solid lines) of the noise level after interpolation, over the detection percentage γ, for four different noise types: Gaussian (grey, no markers), Laplace (black, no markers), mod. Cauchy with $\eta = 0.02$ (grey, crosses), and mod. Cauchy with $\eta = 0.002$ (black, crosses). All noise signals have flat spectra. Simulated block length $N = 2048$ samples.

Another point of view can be obtained by the sample kurtosis. It is defined as the fourth central moment of a signal $y(n)$, divided by the fourth power of the standard deviation of $y(n)$. It expresses the "peakedness" of a signal, and thereby can serve as a measure of how "impulsive" a noise signal is. The kurtosis w is defined as

$$w = \frac{\frac{1}{N}\sum_{n=1}^{N}(y(n) - \bar{y})^4}{\left(\frac{1}{N}\sum_{n=1}^{N}(y(n) - \bar{y})^2\right)^2} \quad , \tag{9}$$

where $\bar{y}$ stands for the arithmetic mean value of $y(n)$. White standard Gaussian noise always has a kurtosis of $w = 3$, whereas standard Laplace noise has a kurtosis of $w = 6$. For modified Cauchy noise, the kurtosis is dependent of the density parameter η. A higher kurtosis means a steeper slope of the histogram. Setting the detection percentage γ to a fixed value, it would be interesting to know how the residual noise level drops for a growing kurtosis. Fig. 4 shows this dependency for two fixed detection percentages, $\gamma = 10$ % and $\gamma = 20$ %. The results indicate a steadily growing noise reduction as the kurtosis increases.

4. CONCLUSIONS

The proposed approach works for noise disturbances of Gaussian, supergaussian and impulsive characteristics. The reduction of the noise level is achieved by manipulation of the PDF of the noisy BM residual, using AR detection and interpolation in the time-domain. The method is especially successful for noise types that exhibit high energy on the left and right edge of their PDFs, like supergaussian types do. T-F masking and AR interpolation are the favoured tools for a practical implementation, but other separation and interpolation techniques might also be considered. The level of noise reduction is rather moderate, but high conservation of the desired signal is guaranteed. The properties of the approach make it attractive for

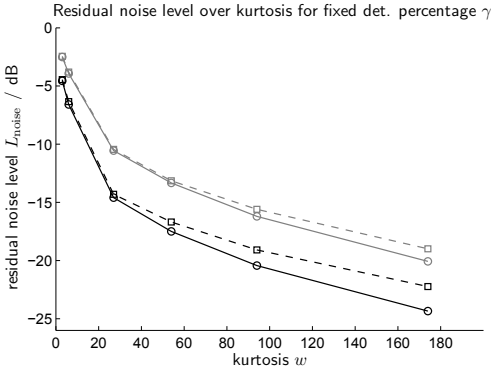

Fig. 4. Residual noise level L_{noise} after interpolation, plotted over the kurtosis w, for fixed detection percentages $\gamma = 10\ \%$ (grey) and $\gamma = 20\ \%$ (black). Dashed lines indicate simulation results, solid lines are the analytic solution. Data points from left to right: Gaussian ($w = 3$), Laplace ($w = 6$), mod. Cauchy at $\eta = 0.02$ ($w \approx 27$), mod. Cauchy at $\eta = 0.01$ ($w \approx 54$), mod. Cauchy at $\eta = 0.005$ ($w \approx 94$), and mod. Cauchy at $\eta = 0.002$ ($w \approx 174$).

high quality audio restoration tasks, like e.g. the reduction of optical soundtrack noise of old celluloid movie footage. However, single high-energy clicks will not be captured by the BM and should be removed in advance using a declicking tool. Finally, the approach is also suitable for speech enhancement under special noise conditions (e.g. rain noise). A web page with audio examples will be presented in a subsequent paper.

A. STANDARD-GAUSSIAN DISTRIBUTION

PDF $\varphi(\cdot)$ and ICDF $\Phi^{-1}(\cdot)$ of the standard Gaussian distribution, with zero mean value and variance one.

$$\varphi(y) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}y^2} \tag{10}$$

$$\Phi^{-1}(p) = \sqrt{2}\,\mathrm{erf}^{-1}(2p-1), \quad p \in (0,1) \tag{11}$$

(erf^{-1} is the inverse error function. For evaluating $\Phi^{-1}(\cdot)$, we use the MATLAB® command `norminv`.)

B. STANDARD-LAPLACE DISTRIBUTION

PDF $\varphi(\cdot)$ and ICDF $\Phi^{-1}(\cdot)$ of Laplace distribution, with zero mean value and variance one.

$$\varphi(y) = \frac{1}{2}\exp\left(-|y|\right) \tag{12}$$

$$\Phi^{-1}(p) = \begin{cases} \ln\left(2p\right), p < \frac{1}{2} \\ -\ln\left(2\left(1-p\right)\right), p \geq \frac{1}{2} \end{cases}, \quad p \in (0,1) \tag{13}$$

C. MODIFIED CAUCHY DISTRIBUTION

PDF $\varphi(\cdot)$ and ICDF $\Phi^{-1}(\cdot)$ of modified Cauchy distribution, with density parameter η.

$$\varphi(y) = \frac{1}{(\pi - 2\eta)(y^2+1)} \tag{14}$$

$$\Phi^{-1}(p) = \tan((\pi - 2\eta)(p - \frac{1}{2})), \quad p \in (0,1) \tag{15}$$

D. REFERENCES

[1] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum-Mean Square Error Short-Time Spectral Amplitude Estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, 1984.

[2] J. Porter and S. Boll, "Optimal Estimators for Spectral Restoration of Noisy Speech," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 53–56, 1984.

[3] R. Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, 2001.

[4] I. Cohen, "Speech Enhancement Using Super-Gaussian Speech Models and Noncausal a Priori SNR Estimation," *Speech communication*, vol. 47, no. 3, pp. 336–350, 2005.

[5] S. V. Vaseghi, *Algorithms for Restoration of Archived Gramophone Recordings*, Ph.D. thesis, University of Cambridge, 1988.

[6] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration*, Springer, London, Great Britain, 1998.

[7] A. Czyzewski, "Some Methods for Detection and Interpolation of Impulsive Distortions in Old Audio Recordings," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 139–142, 1995.

[8] D. Richter, I. Kurreck, and D. Poetsch, "Restoration of Optical Variable Density Sound Tracks on Motion Picture Films by Digital Image Processing," *Proceedings of the International Conference on Optimization of Electrical and Electronic Equipments*, vol. 3, pp. 793–798, 2000.

[9] D. L. Wang, "Time-Frequency Masking for Speech Separation and Its Potential for Hearing Aid Design," *Trends in Amplification*, vol. 12, no. 4, pp. 332–353, 2008.

[10] G. Hu and D. L. Wang, "Speech Segregation Based on Pitch Tracking and Amplitude Modulation," *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 79–82, 2001.

[11] Y. Hu and P. C. Loizou, "Techniques for Estimating the Ideal Binary Mask," *Proceedings of the 11th International Workshop on Acoustic Echo and Noise Control*, 2008.

[12] I. Kauppinen, "Methods for Detecting Impulsive Noise in Speech and Audio Signals," *International Conference on Digital Signal Processing*, vol. 2, pp. 967–970, 2002.

[13] S. V. Vaseghi and P. J. W. Rayner, "A New Application of Adaptive Filters for Restoration of Archived Gramophone Recordings," *International Conference on Acoustics, Speech, and Signal Processing*, pp. 2548–2551, 1988.

[14] S. V. Vaseghi and P. J. W. Rayner, "Detection and Suppression of Impulsive Noise in Speech Communication Systems," *IEEE Proceedings on Communications, Speech and Vision*, vol. 137, pp. 38–46, 1990.

[15] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, Teubner, Leipzig, Germany, 1st edition, 1996.

[16] J. G. Proakis and D. K. Manolakis, *Digital Signal Processing*, Prentice Hall, 4th edition, Apr 2006.

[17] A. J. E. M. Janssen, R. Veldhuis, and L. B. Vries, "Adaptive Interpolation of Discrete-Time Signals that can be Modelled as AR Processes," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 34, no. 2, pp. 317–330, 1986.

[18] R. Veldhuis, *Restoration of Lost Samples in Digital Signals*, Prentice-Hall, Englewood Cliffs, NJ, 1990.

[19] M. Kahrs and K. Brandenburg, *Applications of Digital Signal Processing to Audio and Acoustics*, Springer, London, Great Britain, 1st edition, 1998.

[20] K.-D. Kammeyer and K. Kroschel, *Digital Signal Processing - Filtering and Spectral Analysis with MATLAB® Exercises,* In German language: *Digitale Signalverarbeitung - Filterung und Spektralanalyse mit MATLAB®-Übungen*, Vieweg+Teubner-Verlag, Wiesbaden, Germany, 8th edition, 2012.

[21] M. Ruhland, *Reduction of Gaussian, Supergaussian and Impulsive Noise by Processing of the Binary Masking Residual*, Carl von Ossietzky University Oldenburg, 2012, Master Thesis.

[22] J. H. McCulloch, *Alpha-Stable Distributions in MATLAB®*, 1996, www.mathworks.com/matlabcentral/fileexchange/13619-toolbox-non-local-means/content/toolbox_nlmeans/toolbox/stabrnd.m, last seen on March 2012.