Audio Engineering Society

Convention Paper

Presented at the 128th Convention
2010 May 22–25 London, UK

Adaptive Noise Reduction for Real-time Applications

Constantin Wiesener[1], Tim Flohrer[2], Alexander Lerch[2], Stefan Weinzierl[1]

[1] *Audio Communication Group, TU Berlin, Berlin, Germany*

[2] *zplane.development, Berlin, Germany*

Correspondence should be addressed to Constantin Wiesener (c_wiesener@gmx.de)

ABSTRACT

We present a new algorithm for real-time noise reduction of audio signals. In order to derive the noise reduction function, the proposed method adaptively estimates the instantaneous noise spectrum from an autoregressive signal model as opposed to the widely-used approach of using a constant noise spectrum fingerprint. In conjunction with the Ephraim and Malah suppression rule a significant reduction of both stationary and non-stationary noise can be obtained. The adaptive algorithm is able to work without user interaction and is capable of real-time processing. Furthermore, quality improvements are easily possible by integration of additional processing blocks such as transient preservation.

1. INTRODUCTION

Background noise is a degradation common to all analogue measurement, storage and recording systems for speech or music. It originates from irregularities in the storage medium, ambient noise from the recording environment and electrical circuit noise. Analogue recordings typically show noise characteristics which can be assumed to be stationary and white. However, many early 78rpm shellac recordings as well as tape recordings may have a perceptible non-stationary coloured noise characteristic. The noise can vary considerably within each revolution of the playback system [1]. Hence, a wide range of applications such as restoration of old recordings, teleconferencing, in-car cabin communication systems or automated speech recognition services benefit from efficient noise reduction.

Noisy audio signals are usually described using the following additive noise model

$$y(k) = x(k) + n(k) \,, \tag{1}$$

where $y(k)$ is the noisy signal, $x(k)$ the clean signal, $n(k)$ the noise signal and k the time index. Using the linearity of the short-time Fourier transform (STFT), Eq. (1) can equivalently be formulated as

$$Y(m,n) = X(m,n) + N(m,n), \tag{2}$$

where m represents the frequency index and n the block index. The spectrum of common audio signals can be assumed to be quasi-stationary over a time window of at least $20\,ms$ [1]. By applying a special noise reduction function to the noisy signal block, the clean STFT-block $\hat{X}(m,n)$ can be estimated with

$$\hat{X}(m,n) = f(Y(m,n)), \tag{3}$$

where $f(\cdot)$ is often replaced by a linear filter function $H(m,n)$

$$\hat{X}(m,n) = H(m,n)\, Y(m,n). \tag{4}$$

2. RELATED WORK

2.1. Noise reduction functions

Many possible variants to derive function $f(\cdot)$ have been proposed in the literature. The most prominent noise reduction methods are Wiener filtering [2], spectral subtraction [3] and the Ephraim and Malah filter [4], where typically only the magnitude of $Y(m,n)$ is treated and the phase spectrum is left untouched.

Wiener filter

The Wiener filter modifies the magnitude spectrum of the noisy input signal according to an estimate of the signal-to-noise ratio at each frequency. It requires estimates of the power spectra (or equivalently the correlation matrices) of signal and noise [5]. The Wiener noise reduction function can be written as

$$H = \begin{cases} \frac{|Y|^2 - S_N}{|Y|^2}, & |Y|^2 > S_N \\ 0, & otherwise, \end{cases} \tag{5}$$

where S_N is the current power spectrum of the so-called noise fingerprint (see Sect. 2.2) and the block and frequency index n and m were dropped for convenience.

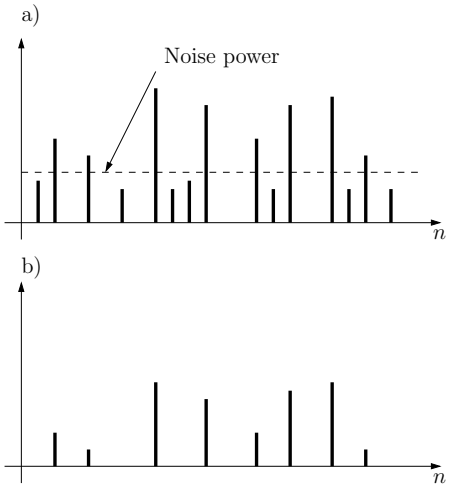


Fig. 1: a) Signal power in bin m before spectral subtraction plotted over time-block index n; b) Signal power in bin m after spectral subtraction plotted over time-block index n [6, p. 208]

Spectral subtraction

The spectral subtraction was first proposed in [3]. By means of this method an amount of noise, equal to the root mean-squared noise level, is subtracted from the spectral amplitude in each frequency bin as followed

$$H = \begin{cases} \frac{|Y| - \sqrt{S_N}}{|Y|}, & |Y|^2 > S_N \\ 0, & otherwise. \end{cases} \tag{6}$$

Although these methods can provide a significant reduction of background noise for audio signals, there are several unfavorable properties in practical applications. The main drawback is the appearance of so-called musical or tonal noise. Due to the above mentioned noise reduction functions randomly spaced spectral residuals can remain after application of the noise reduction function during strong fluctuations of the STFT magnitude $|X(m,n)|$ of the audio signal in noisy areas (Fig. 1). These residual components, occurring at random frequencies, comprise a perceptually annoying musical noise [6, p. 208].

Ephraim and Malah filter

The Ephraim and Malah method is known to produce a lesser amount of musical noise since the noise reduction function depends to a lower extent on time

variations of the short-time spectrum of the audio signal [6, 1].

The calculation of the suppression rule uses an a-priori SNR

$$\xi(m,n) = \frac{S_Y(m,n)}{S_N(m,n)} \tag{7}$$

and an a posteriori SNR

$$\gamma(m,n) = \frac{|Y(m,n)|^2}{S_N(m,n)}. \tag{8}$$

Applying an auxiliary vector $\rho(m,n)$ with

$$\rho(m,n) = \frac{\xi(m,n)}{1+\xi(m,n)}\gamma(m,n) \tag{9}$$

the Ephraim and Malah noise suppresion rule is obtained by

$$H(m,n) = \frac{\sqrt{\pi\rho(m,n)}}{2\gamma(m,n)} \left[(1+\rho(m,n))\, I_0\left(\frac{\rho(m,n)}{2}\right) \cdots \right.$$
$$\left. + \rho(m,n) I_1\left(\frac{\rho(m,n)}{2}\right) \right] e^{\frac{-\rho(m,n)}{2}}, \tag{10}$$

where $I_i(\cdot)$ describes the i-th modified Bessel function.

2.2. Noise spectrum estimation

All described methods assume a precalculated fingerprint. This noise fingerprint usually has to be manually chosen from "silent" or noise-only sections where no music is playing [14, 1, 12]. These architectures are not capable of real-time processing because they need a user interaction. They perform well as long as the noise can be assumed to be stationary and segments of the audio signal with noise-only characteristics can be reliably identified and pre-selected. However, non-stationarities of the noise process or inaccuracies during the noise estimation can cause both insufficient noise reduction and unintended suppression of signal components.

There are several methods described in the literature to estimate the noise spectrum adaptively without user interaction.
Noise spectrum can be determined or updated during pauses without user interaction by using voice activity detection algorithms [7, 8].

Martin et al. described an estimation of the power spectral density of non-stationary noise without using a voice activity detection [9]. Instead, spectral minima in each frequency band are tracked without any distinction between speech and non-speech phases and used for the estimation of an adaptive noise spectrum estimation. However those estimates are not suitable for audio signals with dense spectra.

Yeh et al. presented an algorithm for estimation of colored noise level in audio signals based on the assumptions a) that the noise envelope varies slowly with frequency and that b) the magnitude of the noise peaks obeys a Rayleigh distribution [10]. The derived noise level can then be used for the noise reduction.

3. PROPOSED DENOISING METHOD

To reduce both stationary and non-stationary noise the noise reduction method shown in Fig. 2 is proposed. After a short overview, the individual processing block will be discussed in detail in the following section.

In a first step the STFT is calculated and the instantaneous noise spectrum is estimated by using an autoregressive (AR) model. Therefore, only the current estimation of the spectral envelope of the noise will be considered during the calculation of the current noise reduction function. No pre-selected fingerprint is required. In a second step the noise reduction function is calculated using an approximation of the Ephraim and Malah suppression rule introduced by Wolfe et al.

$$H(m,n) = \sqrt{\frac{\xi(m,n)}{1+\xi(m,n)}\left(\frac{1+\rho(m,n)}{\gamma(m,n)}\right)}. \tag{11}$$

While yielding similar noise reduction gains, this approximation is much more efficient to compute [11].

Preliminary listening test results indicate that additional smoothing over time of the noise reduction function with

$$H(m,n) = (1-\alpha)H(m,n) + \alpha H(m,n-1), \tag{12}$$

where $0 < \alpha < 1$, can reduce the perceived amount of musical noise.

To preserve transient signal components a transient detection is carried out in combination with a preservation rule to control the noise reduction function. Finally, the denoised amplitude spectrum and the untreated phase spectrum are recomposed.

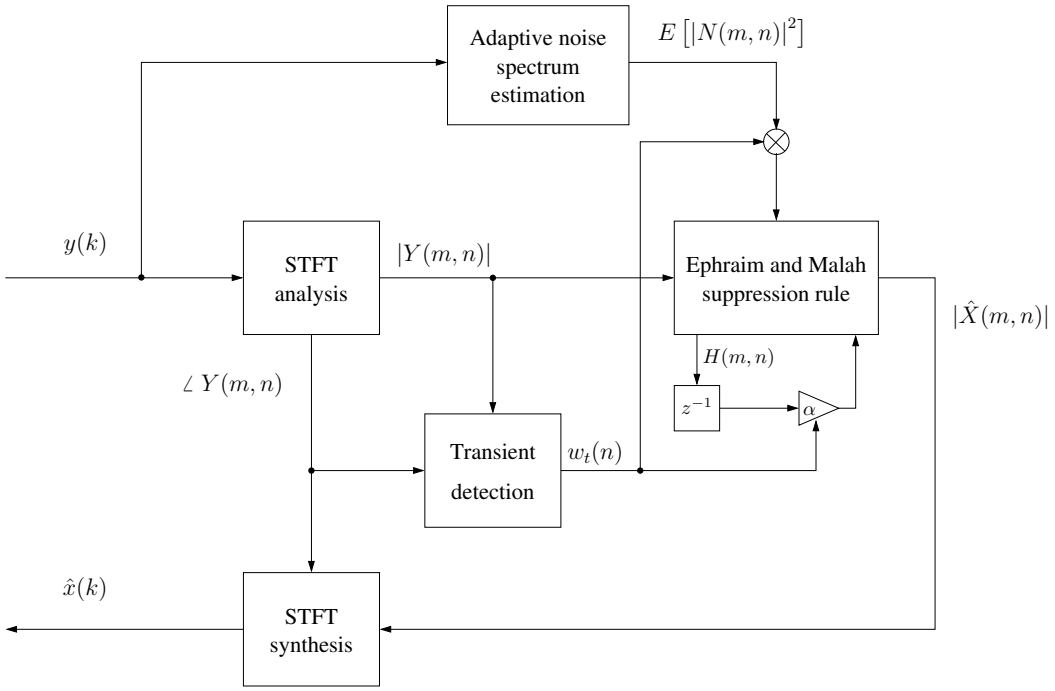


Fig. 2: Block diagram of the noise reduction scheme.

3.1. Adaptive Noise Spectrum Estimation

As pointed out in Sect. 2.2 an adaptive noise spectrum estimation is desirable. Using a signal model such as the autoregressive model (AR) the noise spectrum of the current frame can be estimated. The AR-model describes a time-discrete signal $x(k)$ as the output of an all-pole filter

$$x(k) = \sum_{i=1}^{M} a_i x(k-i) + e(k), \tag{13}$$

whose input $e(k)$ is white noise. Solving the equation for $e(k)$ and using the z-transform yields

$$X(z)A(z) = E(z), \tag{14}$$

where $A(z) = 1 - a_1 z^{-1} - \cdots - a_M z^{-M}$ is an all-zero filter of M-th order. By considering $x(k)$ as input, Eq. (14) can be interpreted as a FIR-filter as presented in Fig. 3.

A standard approach to estimate the AR parameters is the Levinson-Durbin recursion (see e.g. [13]). The Fourier transform of this impulse response can then

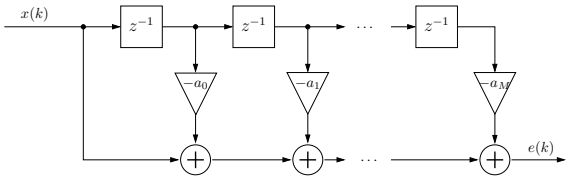


Fig. 3: AR-model as a FIR-filter

be interpreted as an estimation of the noise spectrum of the current block. By means of a frame-to-frame processing only the current noise spectrum estimation will be used for the update of the noise reduction function.

3.2. Transient Preservation

Transients are short-duration signal components that occur during attack phases of musical sounds or spoken words. They contain a high degree of non-periodic components and an increased magnitude of higher frequencies. Linear prediction models such as the AR-model are well suited for approximation of harmonic audio signals [14]. However,

it is not possible to predict transient signal components with fast attacks and decays. In the context of the proposed noise reduction method this can cause an unintended suppression of transients. To preserve those signal components a transient detection in combination with a preservation rule is formulated to control the noise reduction function.

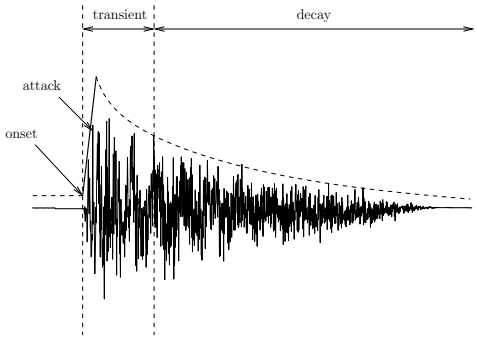


Fig. 4: Example of a transient

Widely used yet comparably simple models for the detection of transients have been summarized by Bello et al. [15]. Many of them use the difference between adjacent STFT frames using magnitude and/or phase spectrum. At the beginning of a transient an increase of signal energy and a sudden change of the phase spectrum can be expected. Therefore, the euclidean difference between adjacent spectra can be used for detecting transient signal components with

$$\Gamma(m,n) = \Big\{ |\hat{X}(m,n)|^2 + |X(m,n)|^2 \cdots$$

$$-2|\hat{X}(m,n)||X(m,n)|\cos(d\tilde{\phi}(m,n)) \Big\}^{1/2}, \qquad (15)$$

where

$$\hat{X}(m,n) = |X(m,n-1)|e^{d\tilde{\phi}(m,n)} \qquad (16)$$

describes the predicted spectrum of the current frame and

$$d\tilde{\phi}(m,n) = princarg[\tilde{\phi}(m,n) \quad \cdots$$

$$-2\tilde{\phi}(m,n-1) + \tilde{\phi}(m,k-2)], \qquad (17)$$

describes the unwrapped phase and $princarg[\cdot]$ the *principal argument*-function, which maps the phase on the interval $[-\pi, \pi]$. By calculating the median of $\Gamma(m,n)$ over all bins and normalising by the block size a transient preservation factor $w_t(n)$ is calculated by applying a certain function

$$w_t(n) = f\left(median\left(\Gamma(m,n)\right)\right), \qquad (18)$$

where $f(\cdot)$ could be $w_t(n) = 1 - median\left(\Gamma(m,n)\right)$. Using this factor the smoothing coefficient α in Eq. (12) can be controlled. Furthermore the current noise spectrum estimation $S_N(m,n)$ can e.g. be scaled by $w_t(n)$ to reduce the amount of noise reduction during the presence of a transient.

3.3. Algorithm summary

Combining all presented methods the proposed noise reduction method is obtained as shown in Fig. 2. The algorithm includes the following steps:

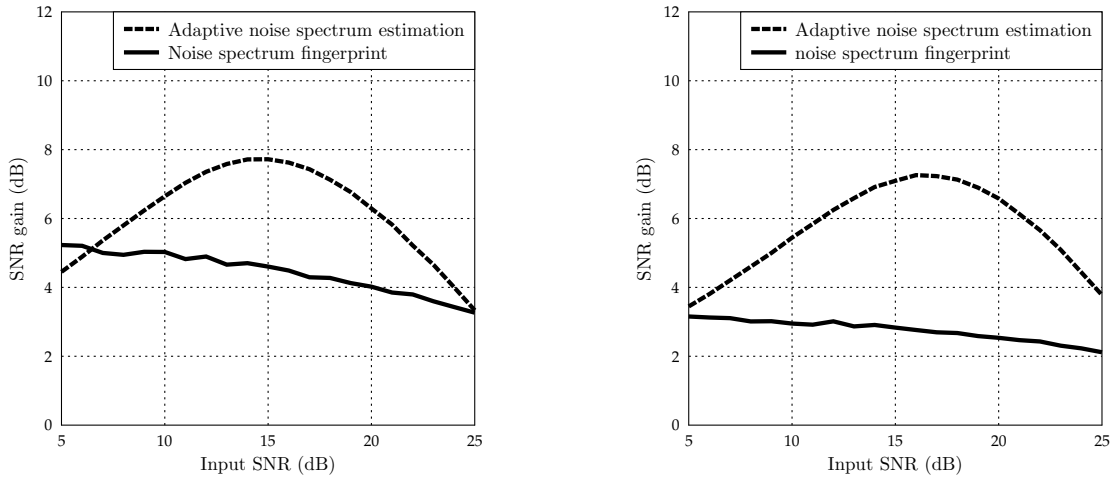
1. Calculate STFT of current block
2. Noise spectrum estimation using the Levinson-Durbin recursion
3. Calculation of transient factor $w_t(n)$ and control of noise reduction function
4. Noise reduction using the Ephraim and Malah filter
5. Combine denoised magnitude and untreated phase

4. EVALUATION

The proposed algorithm was compared to the standard Ephraim and Malah noise reduction function using a preselected noise fingerprint. To evaluate the performance of the noise reduction methods, technical as well as perceptive evaluation criteria have to be considered.

4.1. Technical evaluation

An easy and robust technical measure is the signal-to-noise ratio (SNR). Depending on the input-SNR a certain SNR-gain can be measured after applying the denoising function to the degraded signal. For this technical evaluation measure a test set-up was prepared by artificially degrading diverse wide-band music signals at 44.1 kHz and 16 bit with a length of about 10 sec.



(a) stationary noise

(b) amplitude modulated noise

Fig. 5: Averaged SNR-gains for both experiments; x-axis input-SNR, y-axis SNR after noise reduction

In a first experiment gaussian white noise at different signal-to-noise ratios was added to the audio signals. In a second experiment the typical influence of historical recording system on the stationarity of the noise process was simulated by applying an amplitude modulated white noise with modulation frequency of $^{78}/_{60}$ Hz to the audio signal. The input-SNR varied in both experiments from 5 to 25 dB and the processing was done with a block size of 2048 samples, a Hann window and an overlap of about 1024 samples.

Figure 5(a) shows the SNR-gain for the first experiment. For almost the whole input-SNR range an increased SNR-gain can be observed for the proposed algorithm, with the maximum difference of about 3 dB at an input-SNR of 15 dB. For amplitude modulated noise (Fig. 5(b)) a decrease of about 2-4 dB of the SNR-gain for the standard Ephraim and Malah noise reduction function based on a precalculated noise fingerprint can be observed, while the SNR-gain for the proposed method is only slightly lower than in the stationary case.

4.2. Perceptive evaluation

The described technical evaluation method does not necessarily correlate with the perceived amount of noise reduction and the perceived annoyance of artifacts. Therefore an informal listener test was conducted to compare the proposed algorithm to well established noise reduction software solutions. 31 expert listeners were asked to rate the perceived amount of noise after noise reduction as well as the amount of musical noise on a 5-point scale with 5 being the highest and 1 the lowest achieved quality.

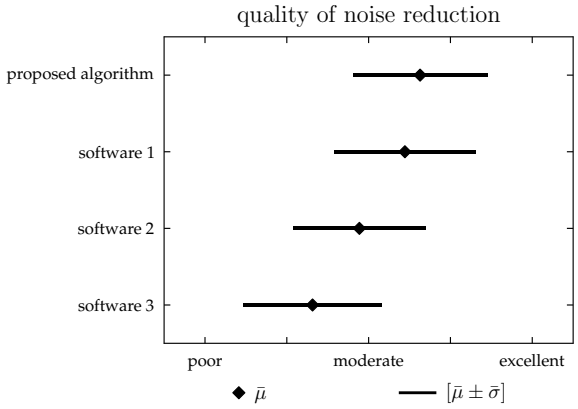


Fig. 6: Boxplot of the rated quality of noise reduction.

The perceived quality of noise reduction (Fig. 4.2) was rated to be comparable to the evaluated denoising software, all of them requiring a fingerprint. Furthermore the results indicate with statistical significance that the perceptible amount of musical noise produced by the test systems is negligible.

5. CONCLUSION

We presented a new algorithm for real-time noise reduction of audio signals using the Ephraim and Malah noise suppression rule. Due to the novel usage of an adaptive noise spectrum estimation based on an AR-model a preselected noise fingerprint is no longer necessary. The adaptive estimation of the noise spectrum enables the algorithm to deal with non-stationarities of the noise process and makes the algorithm capable of real-time processing. In addition quality improvements are easily possible by integration of additional processing blocks such as transient preservation.

6. REFERENCES

[1] S. J. Godsill, P. J. W. Rayner, *Digital Audio Restoration*, Springer Verlag, London, 1. edition, 1998.

[2] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications*, 1. edition. The Technology Press of the Massachusetts Institute of Technology, Cambridge, Mass; John Wiley & Sons, Inc., August 15th 1949.

[3] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 27, pp. 8, April 1979.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error log-spectral amplitude estimator," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-32, pp. 12, December 1984.

[5] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, John Wiley & Sons, Chichester, 3. edition, 2006.

[6] B. U. Köhler, *Konzepte der statistischen Signalverarbeitung*, Springer Verlag, Berlin, 1. edition, 2005.

[7] M. Marzinzik and B. Kollmeier, "Speech pause detection for noise spectrum estimation by tracking power envelope dynamics," *IEEE Transactions on Speech and Audio Processing*, vol. 10, pp. 10, february 2002.

[8] N. Derakhshan et al., "Noise power spectrum estimation using constrained variance spectral smoothing and minima tracking," *Elsevier Science Publishers B. V. Speech Communication*, vol. 51, pp. 5, 2009.

[9] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 9, pp. 9, july 2001.

[10] C. Yeh and A. Röbel, "Adaptive noise level estimation," *Proc. of the 9th Int. Conference on Digital Audio Effects*, p. 4, september 2006.

[11] P. J. Wolfe and S. J. Godsill, "Efficient alternatives to the ephraim and malah suppression rule for audio signal enhancement," *EURASIP Journal on Applied Signal Processing*, vol. 2003, pp. 9, February 2003.

[12] M. Lorber and R. Hoeldrich, "A combined approach for broadband noise reduction," *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, p. 4, October 1997.

[13] P. M. T. Broersen, *Automatic Autocorrelation and Spectral Analysis*, Springer Verlag, London, 1. edition, 2006.

[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. 8, February 1990.

[15] J. P. Bello et al., "A tutorial on onset detection in music signals," *IEEE Transactions on Speech and Audio Processing*, vol. 13, pp. 13, September 2005.