Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor

Olivier Cappé

TELECOM Paris, Département SIGNAL
46 Rue Barrault, 75634 Paris Cedex 13
France

October 5, 2012

Abstract

This paper presents a study of the noise suppression technique proposed by Y. Ephraim and D. Malah. This technique has been used recently for the restoration of degraded audio recordings because it is free of the frequently encountered 'musical noise' artifact. It is demonstrated how this artifact is actually eliminated without bringing distortion to the recorded signal even if the noise is only poorly stationary.

1 Introduction

At present, the noise reduction techniques used for the restoration of degraded audio recordings are based on *short-time spectral attenuation*. In such techniques the attenuation that is to be applied to each one of the short-time Fourier transform coefficients is estimated by the *noise suppression rule* [7] [8] [11].

One artifact that has been widely reported concerning the use of short-time spectral attenuation techniques is that the noise remaining after the processing has a very unnatural disturbing quality [1] [9] [10] [12]. This comes from the fact that the magnitude of the short-time spectrum $|X(p, \omega_k)|$ exhibits strong fluctuations in noisy areas which is a well-known feature of the periodogram [2]. After application of the spectral attenuation, the short-time magnitude spectrum in the frequency bands that originally contained noise now consists in a succession of randomly spaced spectral peaks corresponding to the maxima of $|X(p, \omega_k)|$. In between these peaks the short-time spectrum values are strongly attenuated because they are close to or below the estimated average noise spectrum. As a result, the residual noise is composed of sinusoidal components with random frequencies that come and go in each short-time frame [1] [9]. This artifact is known as the 'musical noise phenomenon', the term 'musical' being a reference to the presence of pure tones in the residual noise.

Some modifications of the basic suppression rules have been proposed in order to overcome this problem [1] [12], but these techniques *only reduce* the musical noise without completely eliminating it. The complete elimination of the musical noise phenomenon is generally only obtained by a crude overestimation of the noise average spectrum. An unwanted consequence is that the short-time spectrum is attenuated much more than would be necessary, a fact which can generate audible distortions in the audio signal [3].

It has been reported that the noise suppression rule proposed by Ephraim and Malah [4] [5] (that will be referred to as the EMSR in the following) makes it possible to obtain a significant noise reduction while avoiding the musical noise phenomenon described above. This feature explains why this suppression rule is an excellent choice for the restoration of musical recordings where the musical noise artifact is to be strictly avoided [10].

In the original papers by Ephraim and Malah, this aspect of the suppression rule was only mentioned as an experimental finding. In this correspondence, we investigate the mechanisms that counter the musical noise phenomenon.

2 Description of the EMSR

The EMSR was proposed by Ephraim and Malah in [4] and developed in [5], two other suppression rules along the same principle were introduced later by the authors in [5] and [6]. Here we will focus only on the EMSR, the fundamental mechanism that counters the musical noise effect being basically the same in all these suppression rules.

The EMSR can be expressed as a spectral gain $G(p, \omega_k)$ to be applied to each short-time spectrum value $X(p, \omega_k)$, this gain is given by [4] [5]

$$G = \frac{\sqrt{\pi}}{2} \sqrt{\left(\frac{1}{1+\mathcal{R}_{post}}\right)\left(\frac{\mathcal{R}_{prio}}{1+\mathcal{R}_{prio}}\right)} \times \\ \mathbf{M}\left[(1+\mathcal{R}_{post})\left(\frac{\mathcal{R}_{prio}}{1+\mathcal{R}_{prio}}\right)\right] \quad (1)$$

where $\mathbf{M}$ stands for the function

$$\mathbf{M}\left[\theta\right] = \exp(-\frac{\theta}{2})\left[(1+\theta)\, I_0\left(\frac{\theta}{2}\right) + \theta I_1\left(\frac{\theta}{2}\right)\right]$$

I_0 and I_1 being the modified Bessel functions of zero and first order, respectively [5].

In (1), the time and frequency indexes p and ω_k have been omitted for reasons of compactness. The spectral gain depends on two parameters, $\mathcal{R}_{post}(p, \omega_k)$ and $\mathcal{R}_{prio}(p, \omega_k)$ evaluated in each short-time frame and for all spectral bins. These two parameters are interpreted as follows:

The a posteriori Signal-to-Noise Ratio (or a posteriori SNR) $\mathcal{R}_{post}(p, \omega_k)$ given by

$$\mathcal{R}_{post}(p, \omega_k) = \frac{|X(p, \omega_k)|^2}{v(\omega_k)} - 1 \quad (2)$$

Where $v(\omega_k)$ denotes the noise power at frequency ω_k. Eq. (2) indicates that $\mathcal{R}_{post}(p, \omega_k)$ is a local estimate of the SNR computed from the data in the current short-time frame. Note that in the original papers by Ephraim and Malah, the definition of the a posteriori parameter is slightly different [5]. The definition of Eq. (2) was preferred because it allows a simpler interpretation of $\mathcal{R}_{post}(p, \omega_k)$.

The so-called a priori Signal-to-Noise Ratio (or a priori SNR) $\mathcal{R}_{prio}(p, \omega_k)$ represents the information on the unknown spectrum magnitude gathered from previous frames, and is evaluated in the "decision-directed" approach [5] by

$$\mathcal{R}_{prio}(p, \omega_k) = (1-\alpha)\, P\left[\mathcal{R}_{post}(p, \omega_k)\right] + \\ \alpha\, \frac{|G(p-1, \omega_k) X(p-1, \omega_k)|^2}{v(\omega_k)} \quad (3)$$

Where $P[x] = x$ if $x \geq 0$ and $P[x] = 0$ otherwise. As $\mathcal{R}_{post}(p, \omega_k)$ defined by (2) is not necessarily positive, the operator P guarantees that $\mathcal{R}_{prio}(p, \omega_k)$ is always non-negative or equivalently, that the expression of the gain given by (1) is valid. On the second line of (3), $G(p-1, \omega_k)X(p-1, \omega_k)$ corresponds to the noiseless signal spectrum value as estimated in the previous frame. The term $|G(p-1, \omega_k)X(p-1, \omega_k)|^2 / v(\omega_k)$ thus corresponds to an estimation of the SNR in the frame of index $p-1$. $\mathcal{R}_{prio}(p, \omega_k)$ is therefore an estimate of the SNR that takes into account the current short-time frame, with weight $(1-\alpha)$, and the result of the processing in the previous frame, with weight α. On the basis of simulations, the parameter α was set by the authors to about 0.98.

For standard suppression rules, the gain applied to each short-time spectral coefficient depends only on the signal level $|X(p, \omega_k)|^2$ measured in the current frame: The gain can be expressed as a function of $\mathcal{R}_{post}(p, \omega_k)$. Fig. 1 displays such suppression characteristics for the power subtraction and the so-called Wiener suppression rules [8] [11]. The two curves of Fig. 1, although they correspond to different strategies, illustrate the same intuitive principle that those points where the SNR is close to $-\infty$ dB are the ones that

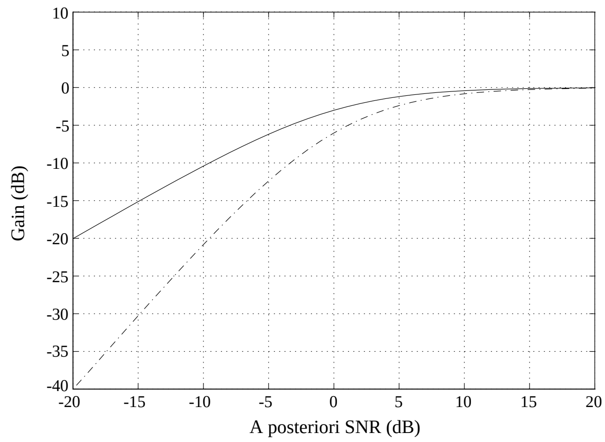

Figure 1: Gain versus a posteriori signal-to-noise ratio; solid line: Power subtraction; dashed line: Wiener.

should be attenuated. These two curves are strongly related because the Wiener gain is the square of the power subtraction gain [8].

The connection between the EMSR and more standard suppression rules is made clearer by plotting the gain of the EMSR versus the a priori SNR (in their original papers [4] [5] the authors used a reverse representation). The alternate representation of Fig. 2 highlights the respective influence of the two parameters of the EMSR:

1. The a priori SNR is the dominant parameter: Strong attenuations are obtained only if $\mathcal{R}_{prio}(p, \omega_k)$ is low (left half of Fig. 2), and low attenuations are obtained only if $\mathcal{R}_{prio}(p, \omega_k)$ is high (right half of Fig. 2). Moreover, note that the overall shape of the gain is similar in Figs. 2 and 1 (although it must be stressed that the abscissa corresponds to $\mathcal{R}_{post}$ in Fig. 1 and to $\mathcal{R}_{prio}$ in Fig. 2).

2. The a posteriori SNR acts as a correction parameter whose influence is limited to the case where the a priori SNR is low (left half of Fig. 2). The surprising point is that this correction effect acts to the opposite of what is intuitively expected: The larger $\mathcal{R}_{post}(p, \omega_k)$, the stronger the attenuation. This over-attenuation is a consequence of the disagreement between the a priori and the a posteriori SNR's. Why this counter-intuitive behavior is actually useful will be explained later.

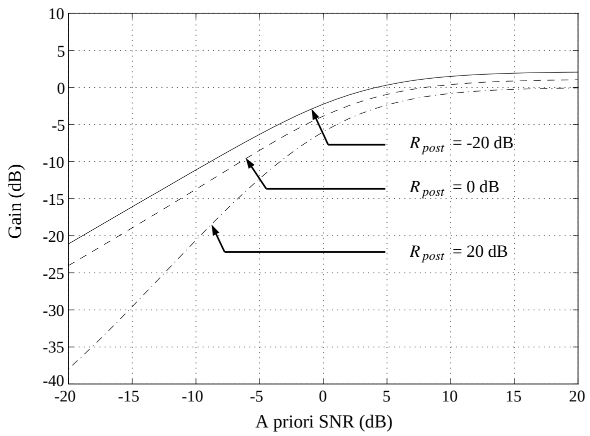

Figure 2: EMSR gain versus a priori SNR, for different values of the a posteriori SNR; topmost curve: $\mathcal{R}_{post}(p, \omega_k)$ = -20 dB; middle curve: $\mathcal{R}_{post}(p, \omega_k) = 0$ dB; bottom curve: $\mathcal{R}_{post}(p, \omega_k) = 20$ dB.

Comparison between Figs. 1 and 2 indicates that the EMSR is very close to the Wiener suppression rule, *evaluated as a function of* $\mathcal{R}_{prio}(p,\omega_k)$, when $\mathcal{R}_{post}(p,\omega_k)$ is 20 dB (bottom curves in the two figures). This remains true for values of $\mathcal{R}_{post}(p,\omega_k)$ above 20 dB. Conversely when $\mathcal{R}_{post}(p,\omega_k)$ is -20 dB, the EMSR gets very close to the power subtraction suppression rule evaluated as a function of $\mathcal{R}_{prio}(p,\omega_k)$ (top curves in the two figures). This is actually true for values of $\mathcal{R}_{post}(p,\omega_k)$ below -5 dB. In practice, it can be considered that the EMSR corresponds to a smooth transition between the two suppression rules of Fig.1, the a priori SNR $\mathcal{R}_{prio}(p,\omega_k)$ controls the x-coordinate along the suppression characteristics, while the a posteriori SNR $\mathcal{R}_{post}(p,\omega_k)$ controls the transition between the two asymptotic curves.

3 Elimination of the musical noise

3.1 The smoothing effect in the EMSR

The a priori SNR is evaluated by the non-linear recursive relation of (3). An experimental study of (3) indicates two different behaviors for the a priori SNR:

1. When $\mathcal{R}_{post}(p,\omega_k)$ stays below or is sufficient close to 0 dB, the a priori SNR corresponds to a highly smoothed version of the a posteriori SNR over successive short-time frames. As a consequence the variance of $\mathcal{R}_{prio}(p,\omega_k)$ is much smaller than that of $\mathcal{R}_{post}(p,\omega_k)$.

2. On the contrary when $\mathcal{R}_{post}(p,\omega_k)$ is much larger than 0 dB, the a priori SNR follows the a posteriori SNR with a simple delay of one short-time frame. To see that, note that when the a priori SNR is high, the attenuation brought to the spectrum is negligible (right part of Fig. 2). Then, (3) reduces to

$$\mathcal{R}_{prio}(p,\omega_k) \approx (1-\alpha)\mathcal{R}_{post}(p,\omega_k) \ + \ \alpha\frac{|X(p-1,\omega_k)|^2}{v(\omega_k)}$$

   As $\mathcal{R}_{post}(p,\omega_k) \gg 1$, this can be written as

$$\mathcal{R}_{prio}(p,\omega_k) \approx (1-\alpha)\mathcal{R}_{post}(p,\omega_k) \ + \ \alpha\mathcal{R}_{post}(p-1,\omega_k)$$

   Finally, the parameter α being generally chosen very close to 1, we can make the following approximation

$$\mathcal{R}_{prio}(p,\omega_k) \approx \alpha\mathcal{R}_{post}(p-1,\omega_k) \tag{4}$$

These two different behaviors of $\mathcal{R}_{prio}(p,\omega_k)$ are visible on the example of Fig. 3. Notice how in the left-hand part of the figure, the variance of $\mathcal{R}_{prio}(p,\omega_k)$ is much lower than that of $\mathcal{R}_{post}(p,\omega_k)$, while on the right hand part, $\mathcal{R}_{prio}(p,\omega_k)$ follows $\mathcal{R}_{post}(p,\omega_k)$ with a one frame delay.

The smoothness of the a priori SNR helps reducing the musical noise effect. In the parts of the short-time spectrum corresponding to noise only, the a posteriori SNR is $-\infty$ dB in average, which corresponds to the case 1 above: Due to the smoothing behavior, the a priori SNR has a significantly reduced variance. Because the attenuation of the EMSR depends mainly on the value of the a priori SNR, the attenuation itself does not exhibit large variations over successive frames. As a consequence, the musical noise (sinusoidal components appearing and disappearing rapidly over successive frames) is reduced.

The idea of calculating the attenuation from the short-time spectrum averaged over successive frames was also exploited in [1]. However the superiority of the EMSR lies in the non-linearity of the averaging procedure. When the signal level is well above the noise level, (3) becomes equivalent to a mere one-frame delay and $\mathcal{R}_{prio}(p,\omega_k)$ no longer is a smoothed SNR estimate which is important in the case of non-stationary signals.

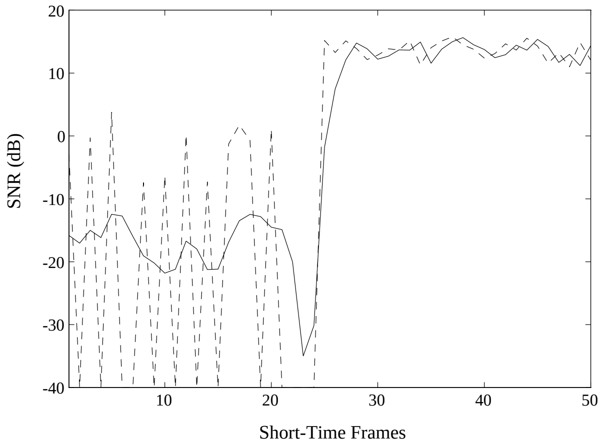

Figure 3: Signal-to-noise ratios in successive short-time frames; dashed curve: A posteriori SNR; solid curve: A priori SNR. For the first 25 short-time frames, the analyzed signal contains only noise at the displayed frequency, for the next 25 frames, a component with 15 dB SNR emerges at the displayed frequency. Parameter α is set to 0.98.

3.2 Protection from local overtaking

The preceding results remain true if the EMSR gain function G in (3) is replaced by the Wiener suppression rule, evaluated as a function of $\mathcal{R}_{prio}(p, \omega_k)$ [5]. However, simulations show that this is not the case when the power subtraction rule is used. Because the power subtraction attenuation is too small for values of the SNR around 0 dB (about -3 dB), the a priori SNR undergoes less smoothing and still exhibits important fluctuations.

In the EMSR, another effect helps eliminating the musical noise. In the frequency bands containing only noise, we have seen that the a priori SNR is about -15 dB in average (see Fig. 3). In that case, improbable high values of the a posteriori SNR are assigned an increased attenuation: In the left half of Fig. 2, the attenuation increases for high values of the a posteriori SNR (values above 0 dB). This over-attenuation is all the more important as $\mathcal{R}_{prio}(p, \omega_k)$ is small. Thus, values of the spectrum higher than the average noise level are 'pulled down'.

This feature of the EMSR is particularly important for the recordings where the background noise is non stationary (e.g. recordings of old analog disks). The use of the EMSR avoids the appearance of local bursts of musical noise whenever the noise exceeds its average characteristics.

4 Influence of the parameters

4.1 Influence of α

The choice of the value of parameter α is guided by a trade-off between the degree of smoothing of parameter $\mathcal{R}_{prio}(p, \omega_k)$ in noisy areas, and the acceptable level of transient distortion brought to the signal.

Simulations show that when the analyzed signal contains only noise at a given frequency, both the average value and the standard deviation of the a priori SNR are proportional to $(1 - \alpha)$ when α is sufficiently close to one (above 0.9). As a result, in order to counter the musical noise effect one will choose values of α as close to one as possible.

On the other hand, when a signal component appears abruptly, the EMSR reacts immediately by raising the gain from a low value to a value close to 1, only if the SNR of the signal component is larger than $1/(1 - \alpha)$. For signal components with lower SNR, simulations show that $\mathcal{R}_{prio}(p, \omega_k)$ takes a longer time to reach its final value. This results in an unwanted attenuation of low amplitude signal components during transient parts. The approximate limit of $1/(1 - \alpha)$ is found by considering the study case where the a posteriori SNR is a deterministic quantity which equals zero before frame index p_0 and has a fixed value of

$\mathcal{R}$ for short-time frames with index $p \geq p_0$. As the gain of the EMSR is null before p_0, we have from (3)

$$\mathcal{R}_{prio}(p_0, \omega_k) = (1 - \alpha)\mathcal{R}$$

if this first value satisfies $\mathcal{R}_{prio}(p_0, \omega_k) \gg 1$, the gain of the EMSR evaluated at frame index p_0 is already close to 1 (see Fig. 2). The condition that guarantees that there is no signal attenuation during the transient is thus $(1 - \alpha)\mathcal{R} \gg 1$.

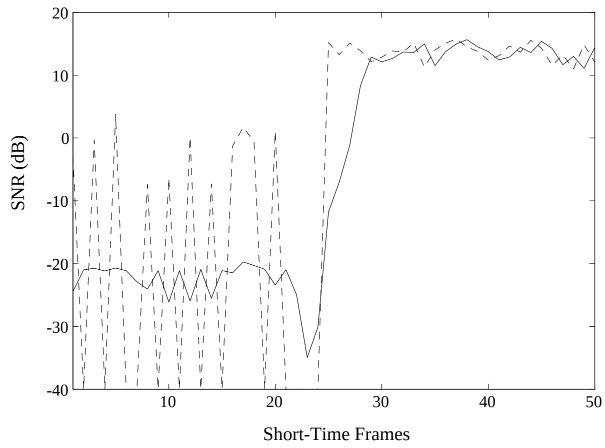

Figure 4: Signal-to-noise ratios in successive short-time frames; dashed curve: A posteriori SNR; solid curve: A priori SNR. The analyzed signal is the same as in Fig. 3. Parameter α is set to 0.998.

The influence of parameter α appears clearly when comparing Figs. 3 and 4. In Fig. 4, the factor $(1-\alpha)$ is divided by 10 compared to the case of Fig. 3. The average value of $\mathcal{R}_{prio}(p, \omega_k)$ when noise is present drops from approximately -15 dB for the case of Fig. 3 to -25 dB for Fig. 4. The variance of $\mathcal{R}_{prio}(p, \omega_k)$ is also strongly reduced in Fig. 4. But there is now an important delay between the appearance of the transient component and the time when $\mathcal{R}_{prio}(p, \omega_k)$ raises significantly above 0 dB. As a consequence, the signal component is incorrectly attenuated in the first short-time frames following the transient. In practice, the use of such a value of parameter α results in audible modifications of the signal transients.

It should be noted that a more important overlap between successive windows reduces the transient distortion as the same number of short-time frame results in a shorter time delay. As a consequence, an overlap of 2/3 or more is sometimes preferred to the standard 50 % setting [10]. However, the variation of the overlap factor gives only slight perceptual differences because only the low level transient components are distorted when reasonable values of α are used; for example with $\alpha = 0.98$, the limit of $1/(1-\alpha)$ results in a SNR value of 15 dB.

4.2 Residual noise level

In the original paper by Ephraim and Malah, the gain function of (1) is tabulated for values of both signal-to-noise ratios between -15 dB and 15 dB [5]. The lower bound of this table is in fact a key parameter for the a priori SNR. Despite the smoothing performed by the procedure of (3), $\mathcal{R}_{prio}(p, \omega_k)$ still has some irregularities that can generate a perceptible low level musical noise. A simple solution to this problem consists in constraining the a priori SNR to be larger to a threshold $\mathcal{R}_{(min)}$. In practice, the value of $\mathcal{R}_{(min)}$ is chosen to be larger than the average a priori SNR in the frequency bands containing noise only. As a consequence, in the frequency bands containing noise only, the average value of the constrained a priori SNR is close to $\mathcal{R}_{(min)}$. Furthermore, in the same frequency bands, most values of the a posteriori SNR are below 0 dB, and the gain function of the EMSR is close to the power subtraction whose squared gain can be shown to be equal to the SNR for low SNR values [8]. As a result, in the frequency bands containing noise only, the average squared gain is close to $\mathcal{R}_{(min)}$. $1/\mathcal{R}_{(min)}$ can therefore be interpreted as the average noise power reduction.

When α equals 0.98, the average value of $\mathcal{R}_{prio}(p, \omega_k)$ is of -15 dB, and a value of $\mathcal{R}_{(min)}$ around -15 dB is sufficient to eliminate the musical noise phenomenon. But $\mathcal{R}_{(min)}$ could as well be set to a larger value with the effect of raising the level of the residual noise. The possibility to control the level of the residual noise is important for old recordings where the preservation of a certain amount of background noise is often judged as a positive aspect.

5 Conclusion

We have presented an analysis of the different mechanisms that counter the musical noise effect in the suppression rule proposed by Ephraim and Malah. The major factor was found to be the non-linear smoothing procedure used to obtain a more consistent estimate of the SNR. With an appropriate choice of parameter α, the use of the smoothing procedure doesn't generate audible distortion in the signal. However, low level signal components actually undergo a measurable over-attenuation during abrupt transients. This transient distortion is hardly perceptible and more precise listening tests would be necessary to decide wether it is useful or not to use an overlap factor larger than 50%. Finally it was shown that the attenuation function proposed by Ephraim and Malah avoids the appearance of the musical noise phenomenon even when the background noise is poorly stationary.

Acknowledgment

The author would like to thank Dr. J. Laroche for his precious assistance in the writing of the paper.

References

[1] S. F. Boll. Suppression of acoustic noise in speech using spectral substraction. *IEEE Trans. Acoust., Speech, Signal Processing*, 27(2):113–120, 1979.

[2] D. R. Brillinger. *Time Series Data Analysis and Theory*. Holden-Day, expanded edition, 1981.

[3] O. Cappé and J. Laroche. Evaluation of short-time spectral attenuation techniques for the restoration of musical recordings. Submitted for publication in IEEE Trans. Speech and Audio Processing, 1993.

[4] Y. Ephraim and D. Malah. Speech enhancement using optimal non-linear spectral amplitude estimation. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pages 1118–1121, Boston, 1983.

[5] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Trans. Acoust., Speech, Signal Processing*, 32(6):1109–1121, 1984.

[6] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Trans. Acoust., Speech, Signal Processing*, 33(2):443–445, 1985.

[7] J. S. Lim and A. V. Oppenheim. Enhancement and bandwidth compression of noisy speech. *Proc. IEEE*, 67(12):1586–1604, December 1979.

[8] R. J. Mc Aulay and M. L. Malpass. Speech enhancement using a soft-decision noise suppression filter. *IEEE Trans. Acoust., Speech, Signal Processing*, 28(2):137–145, April 1980.

[9] J. A. Moorer and M. Berger. Linear-phase bandsplitting: Theory and applications. *J. Audio Eng. Soc.*, 34(3):143–152, 1986.

[10] J. C. Valière. *Restoration of old recordings by means of digital processing – Contribution to the study of some recent tecniques* (text in French). PhD thesis, Université du Maine, Le Mans, 1991.

[11] P. Vary. Noise suppression by spectral magnitude estimation – Mechanism and theoretical limits. *Signal Processing*, 8(4):387–400, 1985.

[12] S. Vaseghi and R. Frayling-Cork. Restoration of old gramophone recordings. *J. Audio Eng. Soc.*, 40(10):791–801, 1992.