Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor
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 (2) was preferred because it allows a simpler interpretation of \( R_{\text{post}}(p, \omega_k) \). The so-called a priori signal-to-noise ratio (or a priori SNR) \( R_{\text{prio}}(p, \omega_k) \) represents the information on the unknown spectrum magnitude gathered from previous frames and is evaluated in the "decision-directed" approach [5] by

\[
R_{\text{prio}}(p, \omega_k) = (1 - \alpha) P[R_{\text{post}}(p, \omega_k)] + \alpha \frac{|G(p-1, \omega_k)X(p-1, \omega_k)|^2}{v(\omega_k)} \tag{3}
\]

where \( P[x] = x \) if \( x \geq 0 \), and \( P[x] = 0 \) otherwise. As \( R_{\text{post}}(p, \omega_k) \) defined by (2) is not necessarily positive, the operator \( P \) guarantees that \( R_{\text{prio}}(p, \omega_k) \) 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 \( p-1 \). \( R_{\text{prio}}(p, \omega_k) \) is therefore an estimate of the SNR that takes into account the information from the previous frame and the current one.

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 \( R_{\text{prio}}(p, \omega_k) \) is low (left half of Fig. 2), and low attenuations are obtained only if \( R_{\text{prio}}(p, \omega_k) \) 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 \( R_{\text{post}} \) in Fig. 1 and to \( R_{\text{prio}} \) 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 opposite of what is intuitively expected: The larger \( R_{\text{post}}(p, \omega_k) \), the stronger the attenuation. This overattenuation 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.
III. ELIMINATION OF THE MUSICAL NOISE

A. The Smoothing Effect in the EMSR

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

1) When $R_{\text{post}}(p, \omega_k)$ stays below or is sufficiently 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 $R_{\text{aprio}}(p, \omega_k)$ is much smaller than that of $R_{\text{apost}}(p, \omega_k)$.

2) On the contrary, when $R_{\text{apost}}(p, \omega_k)$ 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

$$R_{\text{aprio}}(p, \omega_k) \approx (1-\alpha)R_{\text{apost}}(p, \omega_k) + \alpha X(p-1, \omega_k)^2/v(\omega_k).$$

As $R_{\text{apost}}(p, \omega_k) \gg 1$, this can be written as

$$R_{\text{aprio}}(p, \omega_k) \approx (1-\alpha)R_{\text{apost}}(p, \omega_k) + \alpha R_{\text{post}}(p-1, \omega_k).$$

Finally, because the parameter $\alpha$ is generally chosen very close to 1, we can make the following approximation

$$R_{\text{aprio}}(p, \omega_k) \approx \alpha R_{\text{post}}(p-1, \omega_k).$$

These two different behaviors of $R_{\text{aprio}}(p, \omega_k)$ are visible on the example of Fig. 3. Notice how in the left-hand part of the figure, the variance of $R_{\text{aprio}}(p, \omega_k)$ is much lower than that of $R_{\text{apost}}(p, \omega_k)$, whereas on the right-hand part, $R_{\text{aprio}}(p, \omega_k)$ follows $R_{\text{post}}(p, \omega_k)$ 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 nonlinearity of the averaging procedure. When the signal level is well above the noise level, (3) becomes equivalent to a mere one-frame delay, and $R_{\text{aprio}}(p, \omega_k)$ is no longer a smoothed SNR estimate, which is important in the case of nonstationary signals.

B. 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 $R_{\text{aprio}}(p, \omega_k)$ [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 in 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 overattenuation is all the more important because $R_{\text{aprio}}(p, \omega_k)$ 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 nonstationary (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.

IV. INFLUENCE OF THE PARAMETERS

A. Influence of $\alpha$

The choice of the value of parameter $\alpha$ is guided by a trade-off between the degree of smoothing of parameter $R_{\text{aprio}}(p, \omega_k)$ 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 $\alpha$ is sufficiently close to one (above 0.9). As a result, in order to counter the musical noise effect, one will choose values of $\alpha$ 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 $R_{\text{prior}}(p, \omega_k)$ 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 that equals zero before frame index $p_0$ and has a fixed value of $R$ 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)

$$R_{\text{prior}}(p_0, \omega_k) = (1 - \alpha)R.$$  

If this first value satisfies $R_{\text{prior}}(p_0, \omega_k) \gg 1$, 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 $(1 - \alpha)/R \gg 1$.

The influence of parameter $\alpha$ appears clearly when comparing Figs. 3 and 4. In Fig. 4, the factor $(1 - \alpha)$ is divided by 10, compared with the case of Fig. 3. The average value of $R_{\text{prior}}(p, \omega_k)$ when noise is present drops from approximately $-15$ dB for the case of Fig. 3 to $-25$ dB for Fig. 4. The variance of $R_{\text{prior}}(p, \omega_k)$ 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 $R_{\text{prior}}(p, \omega_k)$ 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 $\alpha$ 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 66% 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 $\alpha$ are used; for example, with $\alpha = 0.98$, the limit of $1/(1 - \alpha)$ results in a SNR value of 15 dB.

B. Residual Noise Level

In the original paper by Ephraim and Malah, the gain function of (1) is tabulated for values of both SNR's between $-15$ 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), $R_{\text{prior}}(p, \omega_k)$ 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 $R_{(min)}$. In practice, the value of $R_{(min)}$ 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 $R_{(min)}$. 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 $R_{(min)}$. $1/R_{(min)}$ can therefore be interpreted as the average noise power reduction.

When $\alpha$ equals 0.98, the average value of $R_{\text{prior}}(p, \omega_k)$ is of $-15$ dB, and a value of $R_{(min)}$ around $-15$ dB is sufficient to eliminate the musical noise phenomenon, but $R_{(min)}$ could be set to a larger value as well, 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.

V. 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 nonlinear smoothing procedure used to obtain a more consistent estimate of the SNR. With an appropriate choice of parameter $\alpha$, the use of the smoothing procedure does not generate audible distortion in the signal. However, low-level signal components actually undergo a measurable overattenuation during abrupt transients. This transient distortion is hardly perceptible, and more precise listening tests would be necessary to decide whether 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


Olivier Cappé was born in Villeurbanne, France, in 1968. He received the M.S. degree from the Ecole Supérieure d’Electricité (ESE), Paris, in 1990 and the Ph.D degree in signal processing from TELECOM Paris (ENST) in 1993.

He is currently with the Laboratoire de Police Scientifique, Paris, France. His research interests are in signal processing applied to audio and acoustics and speech processing.

He is a member of the Société Française d’Acoustique and the IEEE Signal Processing Society.