ON THE DECISION-DIRECTED ESTIMATION APPROACH OF EPHRAIM AND MALAH

Israel Cohen

Department of Electrical Engineering, Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel
icohen@ee.technion.ac.il

ABSTRACT

The decision-directed approach of Ephraim and Malah is widely used for *a priori* SNR estimation and speech enhancement. However, it conflicts with common model assumptions. In this paper, we propose recursive estimators for the *a priori* SNR and the speech spectral components. We introduce a novel statistical model that takes into account the time-correlation between successive speech spectral components, while keeping the resulting algorithms simple. This model provides new insight into the decision-directed approach, and enables the extension of existing speech enhancement algorithms to noncausal estimation. The causal *a priori* SNR estimator degenerates, as a special case, to a "decision-directed" estimator with a *time-varying frequency-dependent* weighting factor. The noncausal estimator is capable of discriminating between speech onsets and noise irregularities, achieving lower levels of both musical noise and speech distortion.

1. INTRODUCTION

Two decades ago, Ephraim and Malah proposed a popular statistical model for speech enhancement [1]. Accordingly, the individual short-term spectral components of the speech and noise signals are modeled as statistically independent Gaussian random variables. The assumption of statistical independency is clearly unfulfilled. However, it facilitates a mathematically tractable derivation of useful estimators for various distortion measures [1–3]. Cappé [4] showed that the dominant factor in the Ephraim-Malah algorithm is the decision-directed estimation approach for the *a priori* SNR. Unfortunately, this approach conflicts with the model assumptions. On the one hand, spectral components are assumed statistically independent when deriving analytical expressions for the speech estimators. On the other hand, the *a priori* SNR estimate heavily relies on the strong time-correlation between successive speech spectral components.

In this paper, we propose estimators for the *a priori* SNR and the speech spectral components. We introduce a novel statistical model that takes into account the time-correlation between successive speech spectral components, while keeping the resulting algorithms simple. This model provides new insight into the decision-directed approach, and enables the extension of existing algorithms to noncausal estimation. In the proposed model, the sequence of speech spectral variances is a random process, which is correlated with the sequence of speech spectral components. Causal and noncausal estimators for the *a priori* SNR are derived in agreement with the model assumptions and the estimation of the speech spectral components. We show that the causal *a priori* SNR estimator degenerates, as a special case, to a "decision-directed" estimator with a *time-varying frequency-dependent* weighting factor. Furthermore, the noncausal estimator, having a few subsequent spectral measurements at hand, is capable of discriminating between speech onsets and noise irregularities. This yields lower levels of both musical noise and speech distortion.

In Sec. 2, we formulate the speech enhancement problem. In Sec. 3, we present the statistical model. In Sec. 4, we derive estimators for the speech spectral components and the *a priori* SNR. In Sec. 5, we address the relation to the decision-directed estimation approach. Finally, in Section 6, we discuss the advantages of the proposed estimation approach.

2. PROBLEM FORMULATION

Let $x(n)$ and $d(n)$ denote speech and uncorrelated additive noise signals, respectively, where n is a discrete-time index. Applying the short-time Fourier transform (STFT) to the observed signal $y(n)$, we have in the time-frequency domain

$$Y_\ell(k) = X_\ell(k) + D_\ell(k) \tag{1}$$

where k is the frequency-bin index and ℓ is the time frame index. Let $\mathcal{Y}_0^{\ell'}(k) = \{Y_0(k), \ldots, Y_{\ell'}(k)\}$ denote a set of spectral measurements, and let $d\left[X_\ell(k), \hat{X}_\ell(k)\right]$ be a given distortion measure between $X_\ell(k)$ and $\hat{X}_\ell(k)$. Our objective is to find an estimator $\hat{X}_\ell(k)$, which minimizes the conditional expected value of the distortion measure, given the set of spectral noisy measurements

$$\hat{X}_\ell(k) = \arg\min_{\hat{X}} E\left\{ d\left[X_\ell(k), \hat{X}\right] \mid \mathcal{Y}_0^{\ell'}(k) \right\} . \tag{2}$$

We consider a causal estimation of $X_\ell(k)$ (in which case $\ell' = \ell$), as well as a noncausal estimation (in which case $\ell' > \ell$)[1], while the spectral components are *not* assumed statistically independent. Therefore, in contrast to existing spectral enhancement techniques, the estimation problem is not formulated as that of estimating $X_\ell(k)$ from $Y_\ell(k)$ alone.

3. SPEECH SPECTRAL MODEL

The relation between successive spectral components of a speech signal, in comparison with a noise signal, can be investigated by evaluating the autocorrelation sequences of the STFT coefficients along time-trajectories (the frequency-bin index k is held

[1] Note that causality is defined with respect to the spectral components, rather than with respect to the samples in the time domain.

Proc. ICASSP-2004, Montreal, Canada, 17–21 May 2004.

fixed). The time-correlation between successive spectral magnitudes of speech signals is shown bo be much higher than that of white Gaussian noise [5]. To enable recursive estimators for the speech spectral components and the *a priori* SNR, while keeping the resulting algorithms simple, we propose the following statistical model:

1. The noise spectral components $D_\ell(k)$ are statistically independent zero-mean complex Gaussian random variables. The real and imaginary parts of $D_\ell(k)$ are independent and identically distributed (IID).
2. The speech spectral phases $\angle X_\ell(k)$ are IID uniform random variables on $[-\pi, \pi]$.
3. For a fixed frequency-bin index k, the sequence of speech spectral magnitudes $\{A_\ell(k) \mid \ell = 0, 1, \ldots\}$ is a random process, where $A_\ell(k) \triangleq |X_\ell(k)|$. For $k \neq k'$, the two random processes $\{A_\ell(k) \mid \ell = 0, 1, \ldots\}$ and $\{A_\ell(k') \mid \ell = 0, 1, \ldots\}$ are statistically independent.
4. Let $\lambda_{X_\ell}(k) \triangleq E\left\{|X_\ell(k)|^2\right\}$ denote the speech spectral variance. Then given $\lambda_{X_\ell}(k)$, a spectral component $X_\ell(k)$ is a zero-mean complex Gaussian random variable with IID real and imaginary parts.
5. The sequence of speech spectral variances $\{\lambda_{X_\ell}(k) \mid \ell = 0, 1, \ldots\}$ is a random process. For fixed k and ℓ, $\lambda_{X_\ell}(k)$ is correlated with the sequence of speech spectral magnitudes $\{A_{\ell'}(k) \mid \ell' = 0, 1, \ldots\}$. However, given $\lambda_{X_\ell}(k)$, $X_\ell(k)$ is statistically independent of $X_{\ell'}(k)$ for $\ell' \neq \ell$.

We note that the fundamental difference between the proposed statistical model and that of Ephraim and Malah stems from the last assumption.

4. SIGNAL ESTIMATION

In this section, we derive an estimator for $X_\ell(k)$ based on the proposed statistical model. We assume knowledge of the noise PSD, which in practice can be estimated by using the *Minima Controlled Recursive Averaging* approach [6]. For notational simplicity, the frequency-bin index k is henceforth omitted.

4.1. Spectral Enhancement

Let $p\left(X_\ell \mid \mathcal{Y}_0^{\ell'}\right)$ denote the conditional pdf of a speech spectral component X_ℓ given the noisy measurements $\mathcal{Y}_0^{\ell'}$. Then, the spectral estimator $\hat{X}_\ell$ is obtained by minimizing

$$E\left\{d\left(X_\ell, \hat{X}_\ell\right) \mid \mathcal{Y}_0^{\ell}\right\} = \int d\left(X, \hat{X}\right) p\left(X_\ell \mid \mathcal{Y}_0^{\ell'}\right) dX_\ell\,. \quad (3)$$

Since the statistical model is generally nonlinear, and there exists no simple solution for the spectral estimation, we first derive an estimate for λ_{X_ℓ} from the noisy measurements $\mathcal{Y}_0^{\ell'}$, $\hat{\lambda}_{X_\ell|\ell'} \triangleq E\left\{|X_\ell|^2 \mid \mathcal{Y}_0^{\ell'}\right\}$, and subsequently obtain the spectral estimator $\hat{X}_\ell$. The proposed statistical model implies

$$p\left(X_\ell \mid \mathcal{Y}_0^{\ell'}, \lambda_{X_\ell}\right) = p\left(X_\ell \mid Y_\ell, \lambda_{X_\ell}\right) \quad (4)$$

for $\ell' \geq \ell$. Hence, given $\hat{\lambda}_{X_\ell|\ell'}$, we may compute a spectral estimate from

$$\begin{aligned} &\min_{\hat{X}_\ell} \int d\left(X_\ell, \hat{X}_\ell\right) p\left(X_\ell \mid Y_\ell, \hat{\lambda}_{X_\ell|\ell'}\right) dX_\ell \\ &\qquad = \min_{\hat{X}_\ell} E\left\{d\left(X_\ell, \hat{X}_\ell\right) \mid Y_\ell, \hat{\lambda}_{X_\ell|\ell'}\right\}\,. \end{aligned} \quad (5)$$

This problem, when the *a priori* SNR is defined appropriately, is essentially the classical spectral enhancement problem as formulated by Ephraim and Malah [1, 2]. As a result, an estimate for X_ℓ is obtained by applying a spectral gain function to each noisy spectral component of the speech signal:

$$\hat{X}_\ell = G\left(\xi_{\ell|\ell'}, \gamma_\ell\right) Y_\ell \quad (6)$$

where the *a priori* and *a posteriori* SNR's are defined respectively by[2]

$$\xi_{\ell|\ell'} \triangleq \frac{\lambda_{X_\ell|\ell'}}{\lambda_{D_\ell}}\,; \quad \gamma_\ell \triangleq \frac{|Y_\ell|^2}{\lambda_{D_\ell}} \quad (7)$$

and where $\lambda_{D_\ell} \triangleq E\left\{|D_\ell|^2\right\}$ denotes the noise spectral variance. The specific expression for the spectral gain function $G\left(\xi_{\ell|\ell'}, \gamma_\ell\right)$ depends on the particular choice of a distortion measure $d\left(X_\ell, \hat{X}_\ell\right)$ [3].

4.2. Causal Recursive *A Priori* SNR Estimation

In this subsection, we propose a causal recursive estimator $\hat{\xi}_{\ell|\ell}$ for the *a priori* SNR. The estimator combines a "propagation" step and an "update" step, following the rational of Kalman filtering, to recursively predict and update the estimate for λ_{X_ℓ} as new data arrive.

Suppose we are given an estimate $\hat{\lambda}_{X_\ell|\ell-1}$, and a new noisy spectral component Y_ℓ is observed. Then, the estimate for λ_{X_ℓ} can be updated by computing the conditional variance of X_ℓ given Y_ℓ and $\hat{\lambda}_{X_\ell|\ell-1}$:

$$\begin{aligned} \hat{\lambda}_{X_\ell|\ell} &= E\left\{|X_\ell|^2 \mid \hat{\lambda}_{X_\ell|\ell-1}, Y_\ell\right\} \\ &= \operatorname{var}\left\{X_\ell \mid \hat{\lambda}_{X_\ell|\ell-1}, Y_\ell\right\} + \left|E\left\{X_\ell \mid \hat{\lambda}_{X_\ell|\ell-1}, Y_\ell\right\}\right|^2 \end{aligned} \quad (8)$$

Since we assume that $X_\ell|\lambda_{X_\ell}$ and D_ℓ are statistically independent Gaussian complex variables, the conditional distribution of $X_\ell|\lambda_{X_\ell}$ given Y_ℓ is Gaussian with mean and variance

$$E\left\{X_\ell \mid \lambda_{X_\ell}, Y_\ell\right\} = \frac{\lambda_{X_\ell}}{\lambda_{X_\ell} + \lambda_{D_\ell}} Y_\ell \quad (9)$$

$$\operatorname{var}\left\{X_\ell \mid \lambda_{X_\ell}, Y_\ell\right\} = \frac{\lambda_{X_\ell}}{\lambda_{X_\ell} + \lambda_{D_\ell}} \lambda_{D_\ell}\,. \quad (10)$$

Substituting (9) and (10) into (8), we have

$$\hat{\lambda}_{X_\ell|\ell} = \frac{\hat{\lambda}_{X_\ell|\ell-1}}{\hat{\lambda}_{X_\ell|\ell-1} + \lambda_{D_\ell}} \left(\lambda_{D_\ell} + \frac{\hat{\lambda}_{X_\ell|\ell-1} |Y_\ell|^2}{\hat{\lambda}_{X_\ell|\ell-1} + \lambda_{D_\ell}}\right)\,. \quad (11)$$

Using (7) and dividing both sides of (11) by λ_{D_ℓ}, we have

$$\hat{\xi}_{\ell|\ell} = \frac{\hat{\xi}_{\ell|\ell-1}}{1 + \hat{\xi}_{\ell|\ell-1}} \left(1 + \frac{\hat{\xi}_{\ell|\ell-1}\, \gamma_\ell}{1 + \hat{\xi}_{\ell|\ell-1}}\right)\,. \quad (12)$$

[2] Note that in [1], the *a priori* SNR is defined by $\xi_\ell = \lambda_{X_\ell} / \lambda_{D_\ell}$, where the variance λ_{X_ℓ} is a parameter of the prior pdf of X_ℓ.

Table 1. Summary of the Causal Recursive Speech Enhancement Algorithm.

Initialization: $\hat{A}_{-1} = 0\,,\ \hat{\xi}_{-1\mid -1} = \xi_{\min}\,.$
For all short-time frames $\ell = 0, 1, \ldots$
Obtain $\hat{\xi}_{\ell\mid\ell-1}$ by using the "propagation" step (14).
Obtain $\hat{\xi}_{\ell\mid\ell}$ by using the "update" step (12).
Estimate the speech spectral component X_ℓ by (6) and (7).

We call (12) the "update" step. Assume we are given at frame $\ell-1$ estimates for $A_{\ell-1}$ and $\lambda_{X_{\ell-1}}$, conditioned on $\mathcal{Y}_0^{\ell-1}$. Then, these estimates can be "propagated" in time to obtain an estimate for λ_{X_ℓ}. Since λ_{X_ℓ} is correlated with both $\lambda_{X_{\ell-1}}$ and $A_{\ell-1}$, we propose to use a first-order predictor of the form

$$\hat{\lambda}_{X_\ell\mid\ell-1} = \max\left\{(1-\alpha)\hat{\lambda}_{X_{\ell-1}\mid\ell-1} + \alpha\hat{A}^2_{\ell-1}\,,\ \lambda_{\min}\right\} \qquad (13)$$

where α ($0 \le \alpha \le 1$) is related to the degree of nonstationarity of the random process $\{\lambda_{X_\ell} \mid \ell = 0, 1, \ldots\}$, and $\lambda_{\min}$ is a lower bound on the variance of X_ℓ. Dividing both sides of (13) by $\lambda_{D_{\ell-1}}$, we obtain the "propagation" step

$$\hat{\xi}_{\ell\mid\ell-1} = \max\left\{(1-\alpha)\hat{\xi}_{\ell-1\mid\ell-1} + \alpha\frac{\hat{A}^2_{\ell-1}}{\lambda_{D_{\ell-1}}}\,,\ \xi_{\min}\right\} \qquad (14)$$

where $\xi_{\min}$ is a lower bound on the *a priori* SNR. The steps of the causal recursive spectral enhancement algorithm are summarized in Table 1.

4.3. Noncausal Recursive *A Priori* SNR Estimation

Now we propose a noncausal recursive estimator $\hat{\xi}_{\ell\mid\ell+L}$ for the *a priori* SNR, given the noisy measurements up to frame $\ell+L$, where $L > 0$ denotes the admissible time delay in frames.

Let $\lambda'_{X_\ell\mid\ell+L} \triangleq E\left\{A^2_\ell \mid \mathcal{Y}_0^{\ell-1}\,, \mathcal{Y}_{\ell+1}^{\ell+L}\right\}$ and $\lambda_{\ell\mid[\ell+1,\ell+L]} \triangleq E\left\{A^2_\ell \mid \mathcal{Y}_{\ell+1}^{\ell+L}\right\}$ denote conditional spectral variances of X_ℓ. An estimate for λ_ℓ given $\hat{\lambda}'_{X_\ell\mid\ell+L}$ and Y_ℓ can be updated similarly to (11) by using the "update" step

$$\hat{\xi}_{\ell\mid\ell+L} = \frac{\hat{\xi}'_{\ell\mid\ell+L}}{1+\hat{\xi}'_{\ell\mid\ell+L}}\left(1 + \frac{\hat{\xi}'_{\ell\mid\ell+L}\,\gamma_\ell}{1+\hat{\xi}'_{\ell\mid\ell+L}}\right)\,. \qquad (15)$$

To obtain an estimate for $\lambda'_{X_\ell\mid\ell+L}$, we employ the estimates $\hat{A}_{\ell-1}$ and $\hat{\lambda}_{\ell-1\mid\ell+L-1}$ from the previous frame, and derive an estimate for λ_{X_ℓ} from the measurements $\mathcal{Y}_{\ell+1}^{\ell+L}$. Suppose an estimate $\hat{\lambda}_{\ell\mid[\ell+1,\ell+L]}$ is given, we propose to propagate the estimates from frame $\ell-1$ to frame ℓ by

$$\begin{aligned}\hat{\lambda}'_{\ell\mid\ell+L} \;=\; &\max\Big\{\alpha\hat{A}^2_{\ell-1} + (1-\alpha)\Big[\alpha'\,\hat{\lambda}_{\ell-1\mid\ell+L-1} \\ &+(1-\alpha')\hat{\lambda}_{\ell\mid[\ell+1,\ell+L]}\Big]\,,\ \xi_{\min}\Big\}\end{aligned} \qquad (16)$$

where α ($0 \le \alpha \le 1$) is related to the degree of nonstationarity of the random process $\{\lambda_{X_\ell} \mid \ell = 0, 1, \ldots\}$, and α' ($0 \le \alpha' \le 1$) is associated with the reliability of the estimate $\hat{\lambda}_{\ell\mid[\ell+1,\ell+L]}$ in comparison with that of $\hat{\lambda}_{\ell-1\mid\ell+L-1}$. Dividing both sides of (16) by

Table 2. Summary of the Noncausal Recursive Speech Enhancement Algorithm.

Initialization: $\hat{A}_{-1} = 0\,,\ \hat{\xi}_{-1\mid L-1} = \xi_{\min}\,.$
For all short-time frames $\ell = 0, 1, \ldots$
Obtain $\hat{\xi}_{\ell\mid[\ell+1,\ell+L]}$ by "backward estimation" (18).
Obtain $\hat{\xi}'_{\ell\mid\ell+L}$ by "backward-forward propagation" (17).
Obtain $\hat{\xi}_{\ell\mid\ell+L}$ by the "update" step (15).
Estimate the speech spectral component X_ℓ by (6) and (7).

$\lambda_{D_{\ell-1}}$, we have the following "backward-forward propagation" step:

$$\begin{aligned}\hat{\xi}'_{\ell\mid\ell+L} \;=\; &\max\Big\{\alpha\frac{\hat{A}^2_{\ell-1}}{\lambda_{D_{\ell-1}}} + (1-\alpha)\Big[\alpha'\,\hat{\xi}_{\ell-1\mid\ell+L-1} \\ &+(1-\alpha')\hat{\xi}_{\ell\mid[\ell+1,\ell+L]}\Big]\,,\ \xi_{\min}\Big\}\,.\end{aligned} \qquad (17)$$

An estimate for the *a priori* SNR ξ_ℓ given the measurements $\mathcal{Y}_{\ell+1}^{\ell+L}$ is obtained by

$$\hat{\xi}_{\ell\mid[\ell+1,\ell+L]} = \begin{cases}\frac{1}{L}\sum_{n=1}^{L}\gamma_{\ell+n} - \beta\,, & \text{if nonnegative,}\\ 0\,, & \text{otherwise,}\end{cases} \qquad (18)$$

where β ($\beta \ge 1$) is an over-subtraction factor to compensate for a sudden increase in the noise level. This estimator is an anticausal version of the maximum-likelihood *a priori* SNR estimator suggested in [1]. The steps of the noncausal recursive spectral enhancement algorithm are summarized in Table 2.

5. RELATION TO "DECISION-DIRECTED" ESTIMATION

The proposed causal recursive estimator $\hat{\xi}_{\ell\mid\ell}$ for the *a priori* SNR is closely related to the decision-directed estimator of Ephraim and Malah [1]. The decision-directed estimator is given by

$$\hat{\xi}^{\mathrm{DD}}_{\ell\mid\ell} = w\frac{\hat{A}^2_{\ell-1}}{\lambda_{D_{\ell-1}}} + (1-w)\max\{\gamma_\ell - 1, 0\} \qquad (19)$$

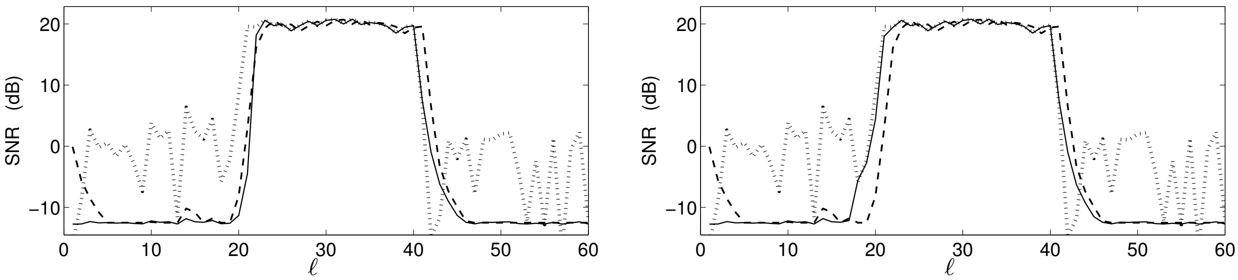
where w ($0 \le w \le 1$) is a weighting factor that controls the trade-off between the noise reduction and the transient distortion introduced into the signal [1,4]. The update step (12) of the causal recursive estimator can be written as

$$\hat{\xi}_{\ell\mid\ell} = \alpha_\ell\,\hat{\xi}_{\ell\mid\ell-1} + (1-\alpha_\ell)(\gamma_\ell - 1) \qquad (20)$$

where α_ℓ is defined by

$$\alpha_\ell \triangleq 1 - \frac{\hat{\xi}^2_{\ell\mid\ell-1}}{\left(1+\hat{\xi}_{\ell\mid\ell-1}\right)^2}\,. \qquad (21)$$

Substituting (14) into (20) and (21) with the parameter α set to 1, and applying the lower bound constraint to $\hat{\xi}_{\ell\mid\ell}$ rather than $\hat{\xi}_{\ell\mid\ell-1}$,



(a) (b)

Fig. 1. SNR's in successive short-time frames: *A posteriori* SNR γ_ℓ (dotted line), decision-directed *a priori* SNR $\hat{\xi}^{\mathrm{DD}}_{\ell|\ell}$ (dashed line), and (a) causal recursive *a priori* SNR estimate $\hat{\xi}_{\ell|\ell}$ (solid line); (b) noncausal recursive *a priori* SNR estimate $\hat{\xi}_{\ell|\ell+3}$ (solid line).

we have

$$\hat{\xi}_{\ell|\ell} = \max\left\{ \alpha_\ell \frac{\hat{A}^2_{\ell-1}}{\lambda_{D_{\ell-1}}} + (1-\alpha_\ell)(\gamma_\ell - 1)\, , \, \xi_{\min} \right\} \quad (22)$$

$$\alpha_\ell = 1 - \frac{\hat{A}^4_{\ell-1}}{\left(\lambda_{D_{\ell-1}} + \hat{A}^2_{\ell-1}\right)^2} \, . \quad (23)$$

The expression (22) with $\alpha_\ell \equiv w$ is actually a practical form of the decision-directed estimator,

$$\hat{\xi}^{\mathrm{DD}}_{\ell|\ell} = \max\left\{ w \frac{\hat{A}^2_{\ell-1}}{\lambda_{D_{\ell-1}}} + (1-w)(\gamma_\ell - 1)\, , \, \xi_{\min} \right\} , \quad (24)$$

that includes a lower bound constraint to further reduce the level of residual musical noise [4]. Accordingly, a special case of the causal recursive estimator with $\alpha \equiv 1$ degenerates to a "decision-directed" estimator with a *time-varying frequency-dependent* weighting factor α_ℓ.

6. DISCUSSION

The weighting factor α_ℓ in (23) is monotonically decreasing as a function of the instantaneous SNR, $\hat{A}^2_{\ell-1}/\lambda_{D_{\ell-1}}$. A decision-directed estimator with a larger weighting factor is indeed preferable during speech absence (to reduce musical noise phenomena), while a smaller weighting factor is more advantageous during speech presence (to reduce signal distortion) [4]. The special case of the causal recursive estimator conforms to such a desirable behavior. Moreover, the general form of the causal recursive estimator provides an additional degree of freedom for adjusting the value of α in (14) to the degree of spectral nonstationarity. This may further improve the performance.

The different behaviors of the causal recursive estimator and the decision-directed estimator are illustrated in the example of Fig. 1. The analyzed signal contains only white Gaussian noise during the first and last 20 frames, and in between it contains an additional sinusoidal component at the displayed frequency with 0 dB SNR. The signal is transformed to the STFT domain by using half overlapping Hamming windows. The *a priori* SNR estimates are obtained by using the spectral gain function, which minimizes the mean squared error of the log-spectral amplitude [2], and the parameters $\xi_{\min} = -25$ dB, $\alpha = 0.9$, $w = 0.98$. It shows that when the *a posteriori* SNR γ_ℓ is sufficiently low, the proposed *a priori* SNR estimate is smoother than the decision-directed estimate, which helps reducing the level of musical noise. When γ_ℓ increases, the response of $\hat{\xi}_{\ell|\ell}$ is initially slower than $\hat{\xi}^{\mathrm{DD}}_{\ell|\ell}$, but it then builds up faster to the *a posteriori* SNR. When γ_ℓ is sufficiently high, $\hat{\xi}^{\mathrm{DD}}_{\ell|\ell}$ follows the *a posteriori* SNR with a delay of 1 frame, whereas $\hat{\xi}_{\ell|\ell}$ follows the *a posteriori* SNR instantaneously. When γ_ℓ decreases, the response of $\hat{\xi}_{\ell|\ell}$ is immediate, while that of $\hat{\xi}^{\mathrm{DD}}_{\ell|\ell}$ is delayed by 1 frame.

Figure 1(b) demonstrates the behavior of the noncausal recursive estimator. The noncausal *a priori* SNR estimate $\hat{\xi}_{\ell|\ell+3}$ is obtained with the parameters $\xi_{\min} = -25$ dB, $\alpha = \alpha' = 0.9$, $\beta = 2$, and $L = 3$ frames delay. The differences between the causal and noncausal recursive estimators are primarily noticeable during onsets of signal components. Clearly, the *causal a priori* SNR estimator, as well as the decision-directed estimator, cannot respond too fast to an abrupt increase in γ_ℓ, since it necessarily implies an increase in the level of musical residual noise. By contrast, the *noncausal* estimator, having a few subsequent spectral measurements at hand, is capable of discriminating between speech onsets and irregularities in γ_ℓ corresponding to noise only. Experimental results [5] confirm that the advantages of the noncausal recursive estimator are particularly perceived during onsets of speech and noise only frames. Onsets of speech are better preserved, while a further reduction of musical noise is achieved.

7. REFERENCES

[1] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-32, no. 6, pp. 1109–1121, December 1984.

[2] ——, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-33, no. 2, pp. 443–445, April 1985.

[3] Y. Ephraim, "Statistical-model-based speech enhancement systems," *Proceedings of the IEEE*, vol. 80, no. 10, pp. 1526–1555, Oct. 1992.

[4] O. Cappé, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 2, no. 2, pp. 345–349, April 1994.

[5] I. Cohen, "Relaxed statistical model for speech enhancement and *a priori* SNR estimation," Technion - Israel Institute of Technology, Haifa, Israel, Technical Report, EE PUB 1384, October 2003.

[6] ——, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. Speech and Audio Processing*, vol. 11, no. 5, pp. 466–475, September 2003.