CRITERIA TO MEASURE THE QUALITY OF TVAR ESTIMATION FOR AUDIO SIGNALS

Imen Samaali[1], *Gaël Mahé*[2], *Monia Turki-Hadj Alouane*[1]

[1]Unité Signaux et Systèmes (U2S), Ecole Nationale d'Ingénieurs de Tunis (ENIT)
BP 37, 1002 Tunis, Tunisia
email: imen_samaali@yahoo.fr, m.turki@enit.rnu.tn
[2] CRIP5 / UFR de mathématiques et informatique , Université Paris Descartes
12, rue de l'Ecole de Médecine, 75006 Paris, France
email: Gael.Mahe@math-info.univ-paris5.fr

ABSTRACT

An audio signal can be represented by a Time-Varying Auto-Regressive (TVAR) model, whose parameters can be estimated by a particle filter. Since the original parameters are unavailable for real signals, an evaluation of the estimation may be traditionally performed through indirect criteria such as the SNR of the signal denoised by a Kalman filter based on the TVAR estimated model or through a statistical analysis based on the observation. We propose a new evaluation method based on the statistical characterization of the output of the inverse TVAR estimated model. The proposed criteria are much more suitable and coherent when correlated to the direct criterion (cepstral distance), which is related to the estimated TVAR parameters.

1. INTRODUCTION

Non-stationary signals, like audio signals, may be represented through Time-Varying Auto-Regressive (TVAR) processes, where the AR coefficients evolve continuously in time. The current tendency is to estimate such models through Monte Carlo methods: the principle of these algorithms is to explore the space of solutions thanks to a population of particles, each of them corresponding to a candidate model [2].

When the original model of the signal is known, the performances of these algorithms can be easily evaluated by comparing the true and the estimated model. However, for natural signals, the original model is rarely available. Consequently, the quality of the TVAR estimation is traditionally evaluated through the reduction of the observation noise obtained by a Kalman filter using the estimated model [1].

The estimated model is also validated by statistical tests on the series u_t defined by:

$$u_t = p(Y_t \leq y_t | y_{1:t-1}), \tag{1}$$

where $Y_1, \ldots, Y_N$ are the random variables associated to the observations $y_1, \ldots, y_t$. This method has the drawback of being computationaly complex.

After a brief presentation of the TVAR model and its estimation by particle filtering in Section 2, the classical criteria for model validation will be presented in Section 3. We will propose a new method for the evaluation of the TVAR estimation in Section 4. In Section 5, the proposed method will be evaluated and compared to the classical one.

2. TVAR MODEL AND PARAMETERS ESTIMATION

2.1 TVAR model

The output, x_t, of a TVAR process of order p is modeled at time $t > 0$ as follows:

$$x_t = \sum_{i=1}^{p} a_{i,t} x_{t-i} + \sigma_{e_t} e_t, \tag{2}$$

where e_t is a white gaussian noise ($e_t \sim \mathcal{N}(0,1)$), $a_t \equiv (a_{1,t}, \ldots, a_{p,t})$ is the vector of TVAR coefficients and $\sigma^2_{e_t}$ is the variance of the TVAR innovation sequence, all depending on t.

The signal, x_t, is assumed to be corrupted by an additive white gaussian noise, so the observation at time $t > 0$ is given by

$$y_t = x_t + \sigma_{n_t} n_t, \tag{3}$$

where n_t is a white gaussian noise ($n_t \sim \mathcal{N}(0,1)$) and $\sigma^2_{n_t}$ is the time varying variance of the observation noise.

The variances of observation and innovation noise are defined by their corresponding logarithms, *i.e.* $\phi_{e_t} \equiv \log \sigma^2_{e_t}$ and $\phi_{n_t} \equiv \log \sigma^2_{n_t}$.

The order p is assumed to be fixed and known. The unknown parameters are then the TVAR coefficients, the variances of the excitation and the observation noise. Each of the three components of the unknown parameter vector $\theta_t = (a_t, \phi_{e_t}, \phi_{n_t})$ is supposed to evolve according to a first-order Markov process, which can be defined by its initial state and by the distribution of its state transition. Those distributions are defined by:

$$p(a_t | a_{t-1}) \equiv \mathcal{N}(a_t; a_{t-1}; \Delta_a); \tag{4}$$

$$p(\phi_{e_t} | \phi_{e_{t-1}}) \equiv \mathcal{N}(\phi_{e_t}; \phi_{e_{t-1}}; \delta_e^2); \tag{5}$$

$$p(\phi_{n_t} | \phi_{n_{t-1}}) \equiv \mathcal{N}(\phi_{n_t}; \phi_{n_{t-1}}; \delta_n^2), \tag{6}$$

where $\mathcal{N}(x; m; P)$ denotes a gaussian density with argument x, mean m, and covariance P.

An estimate of the parameter vector θ_t is given by a conventional particle filter method.

2.2 Particle filter

The principle of the particle filtering is to generate a set of N particles $\theta^{(i)}_{|i=1:N}$, each of them representing a state $\theta_{0:t}$ of the system which is likely to occur.

According to the SIS particle filter algorithm [2], the particles evolve according to the stage of prediction, $p(\theta_t|\theta_{t-1}) = p(a_t|a_{t-1})p(\phi_{e_t}|\phi_{e_{t-1}})p(\phi_{n_t}|\phi_{n_{t-1}})$, and the weights are updated according to the observation (stage of correction: $\tilde{w}_t^{(i)} = p(y_t|\theta_t^{(i)})$). However, the particle filter as described presents a major drawback. The increase of the scattering of the weights has ominous effects on the quality of the estimation and induces a long-term divergence of the filter: this phenomenon is known as "degeneration" of the weights. In order to avoid this phenomenon, a new stage called re-sampling has been introduced. It consists in duplicating the particles of strong weight and eliminating the particles of weak weight.

The TVAR parameters vector is then estimated by a Monte Carlo method: $\hat{\theta}_t \approx \sum_{i=1}^{N} \theta_t^{(i)} \tilde{w}_t^{(i)}$. An estimate $\hat{x}$ of the original signal x is performed by a Kalman filter based on the estimated TVAR parameters [1].

3. VALIDATION OF THE TVAR ESTIMATION: DIRECT AND INDIRECT CRITERIA

This part presents classical criteria used to assess the quality of the TVAR estimation:

- a direct criterion relying on the comparison between the original and estimated parameters. For audio signals, this can be performed through the cepstral distance.
- indirect criteria based either on the SNR improvement or on statistical tests carried out on the observation.

3.1 Cepstral distance measurement

For signals with known TVAR model (synthetic TVAR signals for example), the quality of the TVAR estimation obtained by the particle filtering can be evaluated by the comparison between the trajectories of the original and the estimated TVAR coefficients.

The AR coefficients are related to the cepstrum by the relation [3]:

$$c_t(i) = -a_t(i) - \sum_{n=1}^{i-1}(1-\frac{n}{i})a_t(n)c_t(i-n). \tag{7}$$

One can compare the true and the estimated models by the cepstral distance, which is the euclidian distance between the cepstra,

$$d_t^{\mathscr{C}} = \sqrt{\sum_{i=1}^{p}(c_t(i) - \hat{c}_t(i))^2}. \tag{8}$$

The cepstral distance, as shown here, allows to aggregate the evaluation of all TVAR estimated coefficients in a single and perceptually significant criterion. It will be taken as reference in the following.

3.2 SNR criteria

When the estimated TVAR parameters are used to denoise the observed signal y by a Kalman filtering, a classical measure of performance of the particle filter algorithm is the Signal to Noise Ratio improvement (SNR improvement), *i.e.* the difference between the SNRs after and before denoising. Referring to the experimental results of Vermaak [1], the SNR improvement increases as the number of particles, N, increases up to 100.

However, in another experiment, we notice that the criterion of the SNR improvement is not relevant to measure the quality of the TVAR estimation. Figure 1 shows 400 samples of the test signal generated by a second-order TVAR process. The corresponding TVAR parameters, depicted in Figure 2, follow a first-order Markov process as defined in section 2.1 with fixed parameters $\Delta_{a_1} = \Delta_{a_2} = 10^{-2}$ for the TVAR coefficients and $\delta_e^2 = \delta_n^2 = 10^{-3}$ for the log-variance.

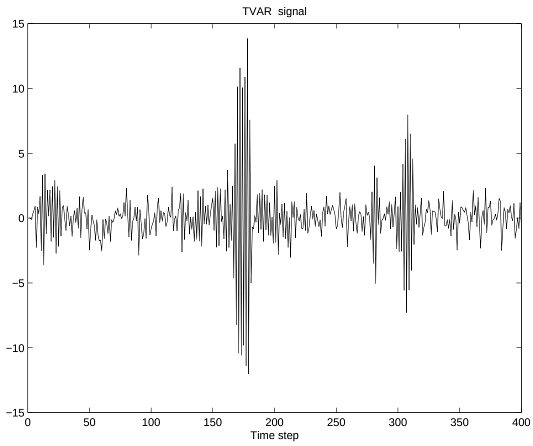

Figure 1: Time variations of the synthetic second-order TVAR signal

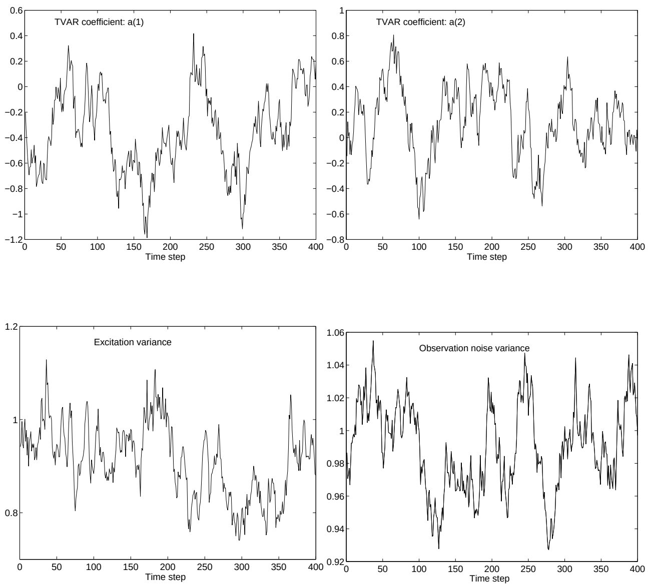

Figure 2: Time variations of the TVAR parameters for the signal in Figure 1

The variations of the SNR improvement according to the input SNR are shown in Figure 3. They were obtained by a single run of the algorithm for each value of the input SNR, using 500 particles. The use of a single run is allowed by the small variance of the results when using 500 particles, as shown in [1].

The SNR improvement decreases as the input SNR increases, whereas the cepstral distance decreases, indicating a better model estimation. One could suggest to take simply

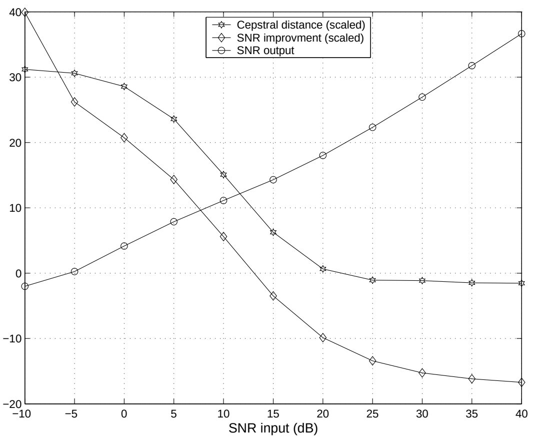

Figure 3: *SNR improvement, output SNR and cepstral distance.*

the SNR output as a measure of the quality but its growth does not match the decrease of the cepstral distance: the latter stops decreasing from input SNR = 20 dB while the output SNR goes on increasing.

Another drawback of the SNR criterion is its dependence on the denoising phase: what is evaluated is not the quality of the TVAR estimation but the quality of the denoising using this estimation.

At last, such a criterion implies that the original signal x is available, which is not realistic for a natural signal.

3.3 Conventional statistical approach for model adequacy

Let $Y_1,...Y_N$ be the random variables associated to the observations $y_1,..y_t$. If the model is correct, the sequence

$$u_t = p(Y_t \leq y_t | y_{1:t-1}) \tag{9}$$

is a realization of an independent random variable uniformly distributed on $[0,1]$. By letting the time series $v_t = \phi^{-1}(u_t)$, where ϕ is the standard normal cumulative distribution function, the TVAR model is correct if v_t is i.i.d according to $\mathcal{N}(0,1)$ [1].

According to [1], the computing of the u_t requires integration over the model parameters. The latter is approximated by a Monte Carlo estimation using the particle filter. Therefore, an estimate of u_t is given by:

$$\hat{u}_t \triangleq \frac{1}{N}\sum_{i=1}^{N} p(Y_t \leq y_t | \theta_{1:t}^{(i)}, y_{1:t-1}), \tag{10}$$

where $\theta_t^{(i)}$ is the i-th particle at time t.

The statistical tests employed here aim at testing the normality and the whiteness of the time series v_t and are briefly described below:

Whiteness

We measure the correlation by a *whiteness index* given by the Ljung-Box test. This index is defined by:

$$q_K^{LB} = N(N+2)\sum_{i=1}^{K}\frac{\hat{r}_i^2}{(N-i)}, \tag{11}$$

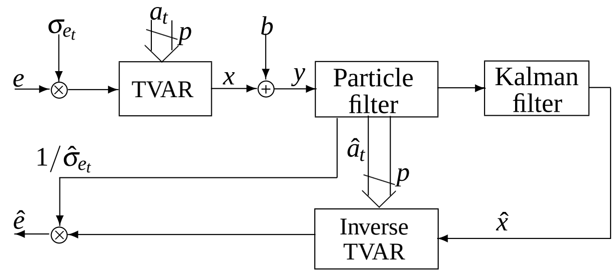

Figure 4: *Diagram of the proposed method.*

where N = sample size, K = number of autocorrelation lags and $\hat{r}_i$ denotes the i^{th} autocorrelation coefficient of the time series. For a white signal, q_k^{LB} is asymptotically Chi-Square distributed.

Normality

The Bowman-Shenton test evaluates the hypothesis that the time series has a normal distribution. The test is based on the skewness $\gamma_1 = \frac{\mu_3}{\mu_2^{(3/2)}}$ and kurtosis $\gamma_2 = \frac{\mu_4}{\mu_2^{(2)}} - 3$ of the time series, with μ_i the i^{th} central moment of the random variable associated with the time series around its mean μ. The statistic value associated to this test is given by:

$$q^{BS} = \gamma_1^2 + \gamma_2^2. \tag{12}$$

For a true gaussian distribution, the statistic q^{BS} should be closer to 0.

4. PROPOSED APPROACH

Assuming that a signal x is produced by a TVAR system excited by a stationary gaussian white noise, the estimation is good if there exists a stationary gaussian white noise that can produce x by exciting the estimated TVAR model. The proposed method is then based on the evaluation of the statistical properties of the estimated excitation, $\hat{e}(t)$, that can generates the TVAR signal $x(n)$. Indeed $\hat{e}(t)$, which is defined by

$$\hat{e}(t) = \frac{1}{\hat{\sigma}_{e_t}}\left(x(t) - \sum_{i=1}^{p}\hat{a}_{i,t}x(t-i)\right), \tag{13}$$

must be stationary, white and gaussian. Note that $\hat{a} = (\hat{a}_{i,t},...,\hat{a}_{p,t})$ is the vector of estimated TVAR coefficients and $\hat{\sigma}_{e_t}$ corresponds to the estimated variance of the TVAR innovation sequence.

Since the original signal x is not available for natural signals under noisy observation, we replace it by the estimated signal $\hat{x}$, given by the Kalman filter. With this approximation, an estimation for the excitation may be defined as follows:

$$\hat{e}(t) = \frac{1}{\hat{\sigma}_{e_t}}\left(\hat{x}(t) - \sum_{i=1}^{p}\hat{a}_{i,t}\hat{x}(t-i)\right). \tag{14}$$

According to (14), the principle of the method can be then illustrated with a diagram (see Figure 4).

Besides the normality and whiteness that are classically evaluated, we have introduced the assessment of the estimated excitation stationarity. Indeed, the stationarity criterion is all the more crucial since the original TVAR signal is non stationary.

Stationarity

The method we choose to detect non stationarity is based on the stationarity index used in [5], which is the Kolmogorov distance between the time frequency representations (TFR) of the signal at different times. The latter is given by:

$$SI(n) = \int_{\tau=0}^{p} \int_{f=-\infty}^{+\infty} |NI_1(n;\tau,f) - NI_2(n;\tau,f)| df d\tau. \quad (15)$$

$NI_1(n;\tau,f)$ and $NI_2(n;\tau,f)$ represent a normalization of respectively subimages $I_1(n;\tau,f)$ and $I_2(n;\tau,f)$:

$$NI_k(n;\tau,f) = \frac{|I_k(n;\tau,f)|}{\int_{\tau=0}^{p}\int_{f=-\infty}^{+\infty} |I_k(n;\tau,f)| df d\tau}, \quad (16)$$

where the two subimages $I_1(n;\tau;f)$ and $I_1(n;\tau;f)$ with equal duration p are extracted from the global TFR on both sides of instant n:

$$I_1(n;\tau,f) = TFR(n-p+\tau,f); \quad (17)$$

$$I_2(n;\tau,f) = TFR(n+\tau,f). \quad (18)$$

The parameter p delimits the considered analysis duration at each instant n and allows the selectivity/sensivity control of the SIs: higher p lead to smoother SIs. As used in [4], we fixed p to 20.

We propose to measure the stationarity of the estimated excitation $\hat{e}$ by the variance of its stationarity index. To evaluate the normality and the whiteness of $\hat{e}$, we use the respective criteria presented in section 3.3.

5. EXPERIMENTAL RESULTS

The experiments aim at validating our approach and compare it to the one presented in subsection 3.3. Thus, we propose to study the correlation between the cepstral distance and the statistical criteria performed on both residual time series v as defined in subsection 3.3 and the estimated excitation $\hat{e}$.

The test signal is a synthetic TVAR signal of order 2 (see Figure 1), with $\Delta_a = 10^{-2} I_2$ for the TVAR coefficients and $\delta_e^2 = \delta_n^2 = 10^{-3}$ for the log-variance (see Figure 2). From Figure 1, one can see that the time variations of the synthetic TVAR signal are similar to those of a natural speech signal.

At a first stage, we investigated the quality of TVAR estimation according to the particles' number N. For each experiment, the time series v and the estimated excitation $\hat{e}$ were computed and analyzed using the three statistical tests presented in sections 3.3 and 4. Since the subsection 3.2 stressed the need of a new criterion especially for input SNR > 20 dB, we fixed SNR to 30 dB. The TVAR estimation was performed for different values of N.

Figure 5 compares the variations, over N, of the standardized statistical criteria and the cepstral distance for both time series v and $\hat{e}$. These reported results were obtained by an averaging over 30 independent runs for each value of N. As expected, the estimation quality is improved by the increase of N: both statistical criteria and cepstral distance decrease as the particles' number increases up to 100. As a preliminary conclusion, the whiteness, the normality and the stationarity of $\hat{e}$ and v appear as good indicators of the TVAR estimation quality, since they follow the same evolution as the cepstral distance.

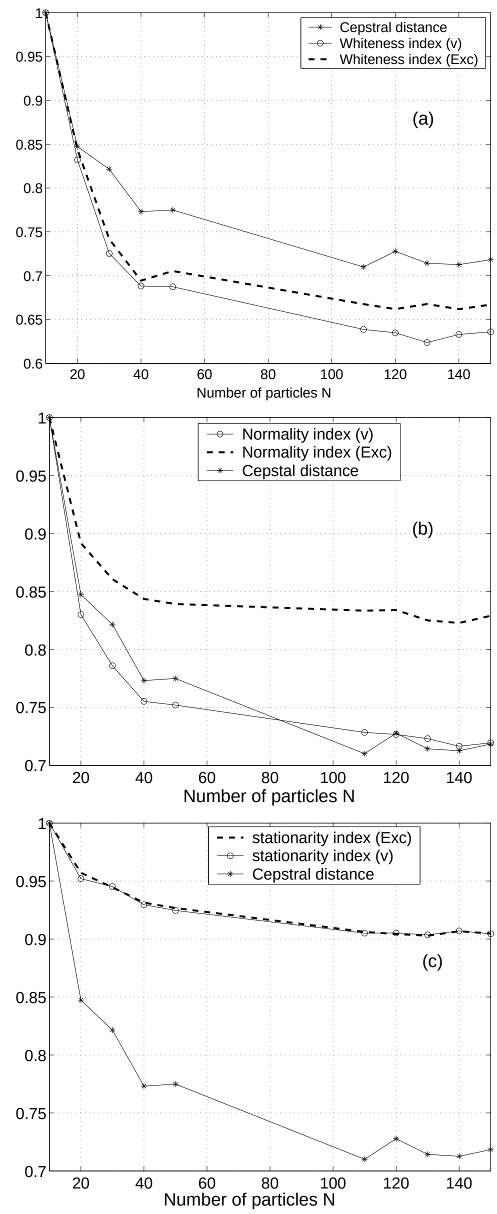

Figure 5: Cepstral distance compared to statistical criteria (whiteness (a), normality (b) and stationarity (c)) for both $\hat{e}$ and v, according to the particle's number.

At a second stage, the estimation of the TVAR parameters through particle filter was performed for various values of the input SNR ($-10:2:40$ dB) and for various values of the particles' number N ($10:10:150$ particles). This simulation was repeated 30 times for each pair (SNR, N). Figure 6 shows the correlation of the statistical criteria with the cepstral distance for both time series v and $\hat{e}$. Each experiment is represented by a point of coordinates $(d^{\mathscr{C}}, I)$ where $d^{\mathscr{C}}$ is the cepstral distance and I refers to one of the three indices (whiteness / normality / stationarity) for v or $\hat{e}$.

These reported results demonstrate that the statistical indices related to the estimated excitation $\hat{e}$ are well correlated to the cepstral distance. The correlation degrees depends on the estimation quality . Indeed, for cepstral distances below 0.6, the whiteness, the normality and the stationarity are al-

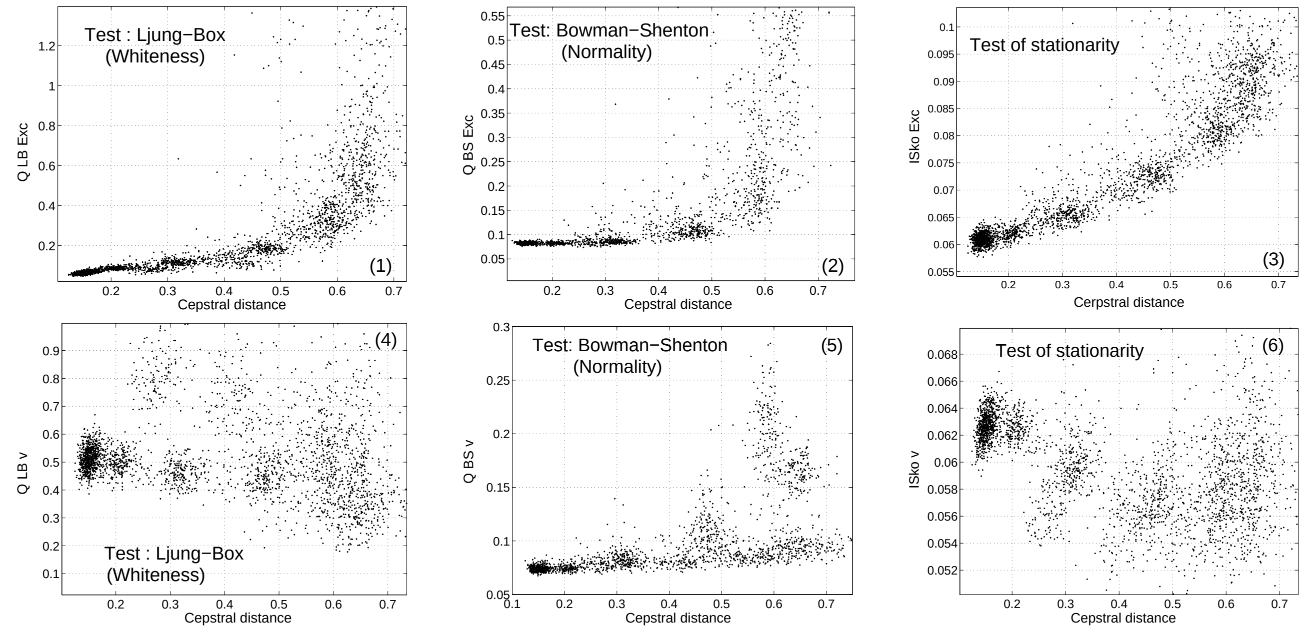

Figure 6: Correlation between statistic criteria and cepstral distance for both the estimate excitation $\hat{e}$ ((1), (2), (3)), and the residual time series v ((4), (5), (6)).

most linearly related to the cepstral distance. Such significant correlation is not observable for v.

To support these results, we computed the correlation coefficients between the cepstral distance and each criterion. The obtained results, summarized in Table 1, confirm that the proposed approach is more coherent with the cepstral distance than the conventional assessing method (time series v).

	Whiteness	Normality	Stationarity
$\hat{e}$	0.93	0.6	0.95
v	0.06	0.80	-0.7

Table 1: *Correlation coefficients between the statistical criteria and the cepstral distance for both the estimated excitation $\hat{e}$ and the time series v.*

The low complexity is another advantage of the proposed method. Whereas the classical validation method based on v, requires the computation of an $erfc$ for each particle, our method requires a simple FIR filtering based on the estimated TVAR model.

6. CONCLUSION

We have shown that the indirect criteria based on the SNR are not relevant for the evaluation of a TVAR estimation. We have proposed a new evaluation method based on the measure of the normality, the stationarity and the whiteness of the estimated excitation. The latter is the output of the inverse estimated TVAR system.

With a lower complexity than the classical method based on residual time series v, the proposed method leads to better results: the whiteness, normality and stationarity indices are strongly correlated to the cepstral distance between the true and the estimated TVAR models.

Note that the proposed method does not aim at performing a binary validation (valid / not valid). It is thought as a quantitative measure of the quality of the TVAR model estimation, which could be used for example to compare two TVAR estimations. In addition, the great advantage of this method is that it does not need any knowledge of the original model, neither of the original signal, contrary to the evaluation through SNR criteria.

REFERENCES

[1] J. Vermaak, C. Andrieu, A. Doucet, SJ. Godsill. *Particle Methods For Bayesian Modeling And Enhancement Of Speech Signals*, IEEE Trans. Speech and Audio Proc, Vol. 10 (3), pp. 173–185, 2002.

[2] M. Sanjeev, S. Maskell, N. Gordon, T. Clapp, *A Tutorial On Particle Filters For Online Nonlinear/Non-Gaussian Bayesian Tracking*, IEEE Trans. Signal Processing, Vol. 50 (2), pp. 174–188, 2002.

[3] G. Faucon, R. Le Bouquin, A. Akbari, *Mesures Objectives De La Réduction De Bruit*, GRETSI, Juan les Pins, pp. 587–590, 1993

[4] S. Larbi, M. Jaidane, *Audio Watermarking: A Way To Stationarize Audio Signals*, IEEE Trans. Signal Processing, Vol. 53 (2), pp. 816–823, 2005.

[5] H. Laurent, C. Doncarli *Stationarity index for abrupt changes detection in the time-frequency plane*, IEEE Signal Processing Letters, Vol. 5, no 2, pp. 43-45, 1998.