Speech enhancement using Rao–Blackwellized particle filtering of complex DFT coefficients

Mounir Meddah, Abderrahmane Amrouche, Abdelmalik Taleb-Ahmed

▶ **To cite this version:**

Mounir Meddah, Abderrahmane Amrouche, Abdelmalik Taleb-Ahmed. Speech enhancement using Rao–Blackwellized particle filtering of complex DFT coefficients. Computers and Electrical Engineering, 2018, 71, pp.847-861. 10.1016/j.compeleceng.2017.07.024 . hal-03428142

HAL Id: hal-03428142
https://uphf.hal.science/hal-03428142v1

Submitted on 3 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Computers and Electrical Engineering 71 (2018) 847–861

Speech enhancement using Rao–Blackwellized particle filtering of complex DFT coefficients

Mounir Meddah [a], Abderrahmane Amrouche [a, *], Abdelmalik Taleb-Ahmed [b]

[a] USTHB, FEI, LCPTS, Speech Communication & Signal Processing Lab., USTHB, Bab Ezzouar 16111, Algeria
[b] LAMIH, UMR CNRS 8201, Université de Valenciennes et du Hainaut Cambrésis, Le Mont Houy, 59 304 Cedex, France

Abstract :

An improved method for speech enhancement, which is based on particle filtering, is pre- sented in this paper. The Rao-Blackwellized particle filter (RBPF) is used to estimate the model parameters and recover a clean speech signal. The proposed method (named DFT- RBPF) enhances the complex discrete Fourier transform (DFT) coefficients of a noisy speech signal. The real and imaginary parts are filtered separately, under the assumption of mu- tual independence, using a low-order time-varying auto-regressive (TVAR) process with a linear Gaussian model. The obtained results, in terms of the coherence speech intelligi- bility index (CSII), perceptual evaluation of speech quality (PESQ), segmental and overall signal-to-noise ratios (SNRseg, SNR), demonstrate the improved performance of the pro- posed method, when compared with the recent methods based on particle filters and the existing algorithms for speech enhancement.

Keywords :
Speech enhancement
Sequential Monte Carlo
Rao-Blackwellized particle
filter Complex DFT
coefficients
DFT-RBPF

1. Introduction

Single-channel speech enhancement algorithms attempt to recover a clean speech signal from a signal that has been corrupted by additive noise. Statistical model-based approaches are one of the most commonly used classes of speech enhancement algorithms. The short-time spectral amplitude is estimated using the maximum likelihood (ML) [1] and minimum mean square error estimators (MMSE) [2]. The Kalman filter (KF) can be considered to be a recursive Bayesian estimator [3,4]. An auto-regressive (AR) model that exploits the time correlation between the component models of speech signals has been used in [5] to deduce the state-space equations, and the Kalman filter has been exploited to estimate the states under a linear Gaussian model. However, in practice, the AR model parameters are not known, and the segmentation does not consider the variations in the speech signal. Furthermore, the vocal tract is continuously changing. Thus, a time-varying auto-regressive (TVAR) model is more appropriate for modelling speech signals [6], and the assumption of quasi-stationary speech signals can be avoided. The resulting non-linear model estimation issues do not have analytical solutions, and approximation methods have to be employed for these computations.

The sequential Monte Carlo (SMC) algorithm, also called the particle filter (PF) in the filtering context, has been used in [6] and associated with the Rao-Blackwellization method (RB) to form the RBPF [7], to filter a noisy speech signal [6]. In [8], the performances of the PF and the RBPF for speech enhancement in the time domain were analyzed. The authors found

* Corresponding author.
E-mail addresses: mounir.meddah@outlook.com (M. Meddah), namrouche@usthb.dz, abd_amrouche@yahoo.fr (A. Amrouche), taleb@univ-valenciennes.fr

DOI : 10.1016/j.compeleceng.2017.07.024

that the residual noise level was modulated by the speech signal power. Many works have tried to reduce this residual noise, and several hybrid approaches have been proposed [9]. In [10], the RBPF was operated in the discrete cosine domain (DCT) to enhance the noisy speech signal. The PF was also used to estimate the amplitudes and phases of the speech signal coefficients in the DFT domain [11]. An alternative to the estimation of the short-time spectral amplitude was the estimation of the real and imaginary parts of the clean-speech DFT components [12]. In [13], the complex AR parameters were estimated using the autocorrelation vector obtained from the past restored samples.

Based on the fact that noise does not affect the speech signal uniformly over the entire spectrum [14], in this work, we propose the use of the RBPF method, to enhance the complex DFT coefficients of the noisy speech signal. The real and imaginary parts are processed separately, under the linear Gaussian model, and a low-order TVAR process is adopted over each frequency bin. The remainder of this paper is organized as follows. In Section 2, the state-space model is presented. Section 3 gives an overview of the SMC and RBPF approaches. Subsequently, in Section 4, we describe the proposed DFT-RBPF method. Section 5 provides the experimental results. Finally, Section 6 lists the concluding remarks.

2. State-space model

Let us consider the observed noisy speech signal y_n at time n, resulting from the linear combination of the clean speech x_n and the noise v_n. After windowing with overlapping, the noisy signal Fourier transform $Y_{m,k}$ at the kth frequency bin, derived from the frame with index m, is modeled as a linear sum of the spectral components of the clean signal $X_{m,k}$ and the noise $V_{m,k}$:

$$Y_{m,k} = X_{m,k} + V_{m,k.} \tag{1}$$

For the above modelling, the correlation over the time–frequency trajectory is maintained for the speech TVAR model order under consideration, and the additive noise is assumed to be uncorrelated with the speech signal over this trajectory [15]. Based on the study of the scatter plots of the real and imaginary parts of the DFT coefficients of clean speech [13,15], the real and imaginary parts are assumed to be independent. Hence, (1) can be written as follows [12]:

$$Y_{m,k}(real) = X_{m,k}(real) + V_{m,k}(real) \tag{2}$$

$$Y_{m,k}(imag) = X_{m,k}(imag) + V_{m,k} \tag{3}$$

Subsequently, equations hold for the both parts over all frequency bins. The real and imaginary parts of the spectral components of the clean speech signal are assumed to follow a TVAR model, as in:

$$X_{m,k} = \sum_{b=1}^{p} a_{b,m,k} X_{m-b,k} + U_{m,k} \tag{4}$$

where $a_{b,m,k}$ are the TVAR model coefficients, p is the TVAR order, and $U_{m,k}$ is the prediction error assumed white Gaussian process with zero mean and variance $\sigma^2_{U_{m,k}}$, which is uncorrelated with all previous values of $X_{m,k}$. We note $U_{m,k} \sim \mathcal{N}(0,\ \sigma^2_{U_{m,k}})$. The state-space transition equation can be described as follows:

$$\begin{bmatrix} X_{m-p+1,k} \\ \vdots \\ X_{m,k} \end{bmatrix} = \begin{bmatrix} 0_{p-1\times 1} & \boldsymbol{I}_{p-1\times p-1} \\ a_{p,m,k} & \cdots \quad a_{1,m,k} \end{bmatrix} \begin{bmatrix} X_{m-p,k} \\ \vdots \\ X_{m-1,k} \end{bmatrix} + \begin{bmatrix} 0_{p-1\times 1} \\ \sigma_{U_{m,k}} \end{bmatrix} e_{m,k} \tag{5}$$

with $e_{m,k} \sim \mathcal{N}(0,\ 1)$. We can write the corresponding matrix form as:

$$\boldsymbol{\alpha}_{m,k} = \boldsymbol{A}_{m,k}\boldsymbol{\alpha}_{m-1,k} + \boldsymbol{B}_{m,k} e_{m,k} \tag{6}$$

where $\boldsymbol{\alpha}_{m,k} = \begin{bmatrix} X_{m-p+1,k} \\ \vdots \\ X_{m,k} \end{bmatrix}$, $\boldsymbol{A}_{m,k} = \begin{bmatrix} 0_{p-1\times 1} & \boldsymbol{I}_{p-1\times p-1} \\ a_{p,m,k} & \cdots \quad a_{1,m,k} \end{bmatrix}$, and $\boldsymbol{B}_{m,k} = \begin{bmatrix} 0_{p-1\times 1} \\ \sigma_{U_{m,k}} \end{bmatrix}$.

Assuming, that $V_{m,k}$ is additive white Gaussian noise (AWGN) with zero mean and variance $\sigma^2_{V_{m,k}}$, we note $V_{m,k} \sim \mathcal{N}(0,\ \sigma^2_{V_{m,k}})$. The state-space measurement equation is:

$$Y_{m,k} = \begin{bmatrix} 0_{1\times p-1} & 1 \end{bmatrix} \begin{bmatrix} X_{m-p+1,k} \\ \vdots \\ X_{m,k} \end{bmatrix} + \begin{bmatrix} \sigma_{V_{m,k}} \end{bmatrix} g_{m,k} \tag{7}$$

with $g_{m,k} \sim \mathcal{N}(0,\ 1)$. We can write the corresponding matrix form as:

$$Y_{m,k} = \boldsymbol{C}_{m,k}\boldsymbol{\alpha}_{m,k} + \boldsymbol{D}_{m,k}\, g_{m,k} \tag{8}$$

where $\boldsymbol{C}_{m,k} = [0_{1\times p-1} \quad 1]$ and $\boldsymbol{D}_{m,k} = [\sigma_{V_{m,k}}]$.

3. Sequential Monte Carlo method

The a posteriori probability density without channel index can be written as follows [16]:

$$p(\boldsymbol{\alpha}_{m_1:m_2}/Y_{1:m}) = \frac{p(\boldsymbol{\alpha}_{m_1:m_2},\ Y_{1:m})}{p(\ Y_{1:m})} \tag{9}$$

Once (9) defines the estimate $\hat{\boldsymbol{\alpha}}_{m_1:m_2}$ of the state $\boldsymbol{\alpha}_{m_1:m_2}$, conditionally to the observations $Y_{1:\,m}$, is derived according to an optimization criterion (absolute error value, error square value, etc.).

3.1. Importance sampling (IS)

If (9) cannot be derived analytically, the approximation with Monte Carlo (MC) method is used [17,16]; for $m_1 = 0$ and $m_2 = m$:

$$\hat{p}(\boldsymbol{\alpha}_{0:m}/Y_{1:m}) \cong \frac{1}{N}\sum_{i=1}^{N}\delta\left(\boldsymbol{\alpha}_{0:m} - \boldsymbol{\alpha}_{0:m}^{i}\right) \tag{10}$$

where $\delta(.)$ denotes the Dirac function and $\boldsymbol{\alpha}_{0:m}^{i}$ are the independent identically distributed particles according to $p(\boldsymbol{\alpha}_{0:\,m}/Y_{1:\,m})$. We note that $\boldsymbol{\alpha}_{0:m}^{i} \sim\ p(\boldsymbol{\alpha}_{0:m}/Y_{1:m})$, and N is the total number of used particles.

Using (10), a posteriori expectation of the function $f(\boldsymbol{\alpha}_{0:\,m})$ becomes:

$$E_{\hat{p}(\boldsymbol{\alpha}_{0:m}/Y_{1:m})}[f(\boldsymbol{\alpha}_{0:m})] = \frac{1}{N}\sum_{i=1}^{N} f\left(\boldsymbol{\alpha}_{0:m}^{i}\right) \tag{11}$$

However, the a posteriori density is unknown. Let us consider $q(\boldsymbol{\alpha}_{0:\,m}/Y_{1:\,m})$, the importance density, which is similar to the probability density of interest, where:

$p(\boldsymbol{\alpha}_{0:\,m}/Y_{1:\,m}) > 0 \Rightarrow q(\boldsymbol{\alpha}_{0:\,m}/Y_{1:\,m}) > 0$. Thus, (9) can be written as follows:

$$p(\boldsymbol{\alpha}_{0:m}/Y_{1:m}) = \frac{\dfrac{p(\boldsymbol{\alpha}_{0:m}, Y_{1:m})}{q(\boldsymbol{\alpha}_{0:m}/Y_{1:m})}q(\boldsymbol{\alpha}_{0:m}/Y_{1:m})}{\int \dfrac{p(\boldsymbol{\alpha}_{0:m}, Y_{1:m})}{q(\boldsymbol{\alpha}_{0:m}/Y_{1:m})}q(\boldsymbol{\alpha}_{0:m}/Y_{1:m})\ d\boldsymbol{\alpha}_{0:m}} \tag{12}$$

$$p(\boldsymbol{\alpha}_{0:m}/Y_{1:m}) = \frac{w(\boldsymbol{\alpha}_{0:m}, Y_{1:m})q(\boldsymbol{\alpha}_{0:m}/Y_{1:m})}{\int w(\boldsymbol{\alpha}_{0:m}, Y_{1:m})q(\boldsymbol{\alpha}_{0:m}/Y_{1:m})\ d\boldsymbol{\alpha}_{0:m}} \tag{13}$$

Using the MC approximation of the importance density, (13) becomes:

$$\hat{p}_N(\boldsymbol{\alpha}_{0:m}/Y_{1:m}) \cong \frac{\sum_{i=1}^{N} w\ \left(\boldsymbol{\alpha}_{0:m}^{i},\ Y_{1:m}\right)\delta\left(\boldsymbol{\alpha}_{0:m} - \boldsymbol{\alpha}_{0:m}^{i}\right)}{\sum_{i=1}^{N} w\left(\boldsymbol{\alpha}_{0:m}^{i},\ Y_{1:m}\right)} \tag{14}$$

$$\hat{p}_N(\boldsymbol{\alpha}_{0:m}/Y_{1:m}) \cong \sum_{i=1}^{N} W_{0:m}^{i}\delta\left(\boldsymbol{\alpha}_{0:m} - \boldsymbol{\alpha}_{0:m}^{i}\right) \tag{15}$$

where

$$W_{0:m}^{i} = \frac{w_{0:m}^{i}}{\sum_{j=1}^{N} w_{0:m}^{j}} \tag{16}$$

$$w_{0:m}^{i} = w\ \left(\boldsymbol{\alpha}_{0:m}^{i}, Y_{1:m}\right) = \frac{p\left(\boldsymbol{\alpha}_{0:m}^{i}, Y_{1:m}\right)}{q\left(\boldsymbol{\alpha}_{0:m}^{i}/Y_{1:m}\right)} \tag{17}$$

$W_{0:m}^{i}$ is the normalized importance weight, which measures the importance of particle i at the frame with the index m. Hence, (11) can be written as follows:

$$E_{\hat{p}_N(\boldsymbol{\alpha}_{0:m}/Y_{1:m})}[f(\boldsymbol{\alpha}_{0:m})] = \sum_{i=1}^{N} W_{0:m}^{i}\ f\left(\boldsymbol{\alpha}_{0:m}^{i}\right) \tag{18}$$

3.2. Sequential importance sampling (SIS)

If the importance density is chosen to be factorized such that [17,16]:

$$q(\boldsymbol{\alpha}_{0:m}/Y_{1:m}) = q(\boldsymbol{\alpha}_{0:m-1}/Y_{1:m-1})\; q(\boldsymbol{\alpha}_m/\boldsymbol{\alpha}_{0:m-1}, Y_{1:m}) \tag{19}$$

we can write

$$w_{0:m} = \frac{p(\boldsymbol{\alpha}_{0:m-1}, Y_{1:m-1})\; p(\boldsymbol{\alpha}_m/\boldsymbol{\alpha}_{m-1})p(Y_m/\boldsymbol{\alpha}_m)}{q(\boldsymbol{\alpha}_{0:m-1}/Y_{1:m-1})\; q(\boldsymbol{\alpha}_m/\boldsymbol{\alpha}_{0:m-1}, Y_{1:m})} \tag{20}$$

$$w_{0:m} = w_{0:m-1}\; \frac{p(\boldsymbol{\alpha}_m/\boldsymbol{\alpha}_{m-1})p(Y_m/\boldsymbol{\alpha}_m)}{q(\boldsymbol{\alpha}_m/\boldsymbol{\alpha}_{0:m-1}, Y_{1:m})} \tag{21}$$

Thus, $w^i_{0:m}$ in (17) becomes:

$$w^i_{0:m} = w^i_{0:m-1}\; \frac{p\left(\boldsymbol{\alpha}^i_m/\boldsymbol{\alpha}^i_{m-1}\right)p\left(Y_m/\boldsymbol{\alpha}^i_m\right)}{q\left(\boldsymbol{\alpha}^i_m/\boldsymbol{\alpha}^i_{0:m-1}, Y_{1:m}\right)} \tag{22}$$

To minimize the variance of the importance weight [7], a resampling step is applied, where particles with large normalised importance weights are duplicated with replacements, and the other particles are not propagated again. Then, all normalised weights are set to $(1/N)$, which defines the sequential importance sampling with resampling. The resampling step is usually implemented depending on the effective sample size N_{eff} compared to a prefixed level.

$$N_{eff} \cong \frac{1}{\sum_{i=1}^{N}\left(w^i_m\right)^2} \tag{23}$$

3.3. Rao–Blackwellization for sequential importance sampling

In order to reduce the variance of the MC estimator, the Rao–Blackwellization approach is used, where a hybrid filter is obtained such that a part of the calculation is obtained analytically and the other part is realized by MC [16]. Assuming that we can decompose the state $\boldsymbol{\alpha}_{0:\,m}$ as $(\boldsymbol{\alpha}_{0:\,m},\ \boldsymbol{\theta}_{0:\,m})$, the a posteriori expectation of the function $f(\boldsymbol{\alpha}_{0:\,m},\ \boldsymbol{\theta}_{0:\,m})$ becomes:

$$E_{p\left(\boldsymbol{\alpha}_{0:m},\ \boldsymbol{\theta}_{0:m}/Y_{1:m}\right)}\left[f\left(\boldsymbol{\alpha}_{0:m},\ \boldsymbol{\theta}_{0:m}\right)\right] = \frac{\iint f\left(\boldsymbol{\alpha}_{0:m},\ \boldsymbol{\theta}_{0:m}\right)p\left(\boldsymbol{\alpha}_{0:m},\ \boldsymbol{\theta}_{0:m}, Y_{1:m}\right)d\boldsymbol{\alpha}_{0:m}d\boldsymbol{\theta}_{0:m}}{\iint p\left(\boldsymbol{\alpha}_{0:m},\ \boldsymbol{\theta}_{0:m}, Y_{1:m}\right)d\boldsymbol{\alpha}_{0:m}d\boldsymbol{\theta}_{0:m}} \tag{24}$$

This a posteriori expectation can be deduced using either the PF algorithm or the RBPF algorithm, where (24) is developed as follows:

$$E_{p\left(\boldsymbol{\alpha}_{0:m},\ \boldsymbol{\theta}_{0:m}/Y_{1:m}\right)}\left[f\left(\boldsymbol{\alpha}_{0:m},\ \boldsymbol{\theta}_{0:m}\right)\right] = \frac{\iint f\left(\boldsymbol{\alpha}_{0:m},\ \boldsymbol{\theta}_{0:m}\right)p\left(\boldsymbol{\alpha}_{0:m}/\boldsymbol{\theta}_{0:m}, Y_{1:m}\right)p\left(\boldsymbol{\theta}_{0:m}, Y_{1:m}\right)d\boldsymbol{\alpha}_{0:m}d\boldsymbol{\theta}_{0:m}}{\iint p\left(\boldsymbol{\alpha}_{0:m}/\boldsymbol{\theta}_{0:m}, Y_{1:m}\right)p\left(\boldsymbol{\theta}_{0:m}, Y_{1:m}\right)d\boldsymbol{\alpha}_{0:m}d\boldsymbol{\theta}_{0:m}} \tag{25}$$

Considering the importance density $q(\boldsymbol{\theta}_{0:\,m}/Y_{1:\,m})$, (25) becomes:

$$E_{p\left(\boldsymbol{\alpha}_{0:m},\ \boldsymbol{\theta}_{0:m}/Y_{1:m}\right)}\left[f\left(\boldsymbol{\alpha}_{0:m},\ \boldsymbol{\theta}_{0:m}\right)\right] = \frac{\iint f\left(\boldsymbol{\alpha}_{0:m},\ \boldsymbol{\theta}_{0:m}\right)p\left(\boldsymbol{\alpha}_{0:m}/\boldsymbol{\theta}_{0:m}, Y_{1:m}\right)\frac{p\left(\boldsymbol{\theta}_{0:m}, Y_{1:m}\right)}{q\left(\boldsymbol{\theta}_{0:m}/Y_{1:m}\right)}q\left(\boldsymbol{\theta}_{0:m}/Y_{1:m}\right)d\boldsymbol{\alpha}_{0:m}d\boldsymbol{\theta}_{0:m}}{\int \frac{p\left(\boldsymbol{\theta}_{0:m}, Y_{1:m}\right)}{q\left(\boldsymbol{\theta}_{0:m}/Y_{1:m}\right)}q\left(\boldsymbol{\theta}_{0:m}/Y_{1:m}\right)d\boldsymbol{\theta}_{0:m}} \tag{26}$$

Using the MC approximation of the considered importance density, (26) becomes:

$$E_{\hat{p}_N\left(\boldsymbol{\alpha}_{0:m},\ \boldsymbol{\theta}_{0:m}/Y_{1:m}\right)}\left[f\left(\boldsymbol{\alpha}_{0:m},\ \boldsymbol{\theta}_{0:m}\right)\right] \cong \frac{\sum_{i=1}^{N} w\left(\boldsymbol{\theta}^i_{0:m}\right)\int f\left(\boldsymbol{\alpha}_{0:m},\ \boldsymbol{\theta}^i_{0:m}\right)p\left(\boldsymbol{\alpha}_{0:m}/\boldsymbol{\theta}^i_{0:m}, Y_{1:m}\right)d\boldsymbol{\alpha}_{0:m}}{\sum_{i=1}^{N} w\left(\boldsymbol{\theta}^i_{0:m}\right)} \tag{27}$$

with $\int f(\boldsymbol{\alpha}_{0:m},\ \boldsymbol{\theta}^i_{0:m})p(\boldsymbol{\alpha}_{0:m}/\boldsymbol{\theta}^i_{0:m}, Y_{1:m})d\boldsymbol{\alpha}_{0:m}$ being calculated analytically.

4. Proposed speech enhancement methods using particle filters

Conceptually, the PF algorithm acts as a basic framework for the proposed DFT-RBPF. It offers flexibility and circumvents the difficulty of measuring the parameters from the noise-corrupted signal. The DFT-RBPF method can be described as follows: candidate states are estimated using the Kalman filter from the corresponding estimated TVAR model parameters (particles). The predictive densities of the resulting TVAR model parameters are evaluated using only one step of the Kalman filter. These predictive density weights are then used to compute the estimate of the clean signal coefficients (either for the real or imaginary part) as a weighted sum of the Kalman filter outputs. The resampling is performed to select the most likely candidate to be propagated over the frequency trajectory to the next frame.

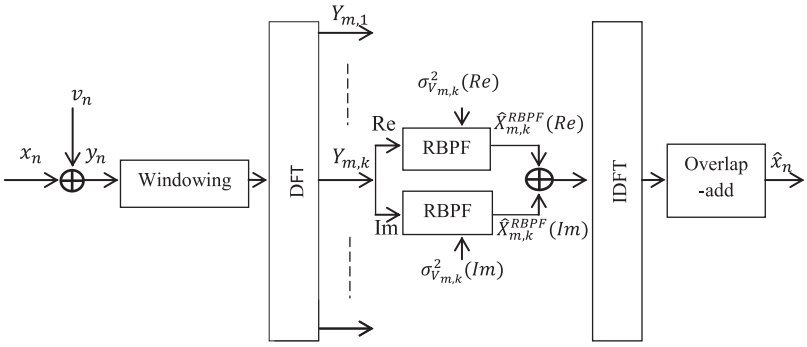

Fig. 1. Block diagram of the proposed DFT-RBPF speech enhancement method.

4.1. Speech enhancement using RBPF for the noisy speech complex DFT coefficients (DFT-RBPF)

The block diagram of the proposed filtering method for the complex DFT-RBPF coefficients of the noisy speech signal, is depicted in Fig. 1.

Depending on the adopted state-space model, the estimation of the hidden states requires the definition of a vector of parameters $\boldsymbol{\theta}_{m,k} = [\boldsymbol{a}_{m,k} \quad \sigma^2_{U_{m,k}}]$, where $\boldsymbol{a}_{m,k} = [a_{p,m,k} \quad \dots \quad a_{1,m,k}]$ is the vector of prediction coefficients. For each frequency bin, we have:

$$p\left(\boldsymbol{\alpha}_m,\ \boldsymbol{\theta}_{0:m}/Y_{1:m}\right) = p\left(\boldsymbol{\alpha}_m/\ \boldsymbol{\theta}_{0:m}, Y_{1:m}\right) p\left(\ \boldsymbol{\theta}_{0:m}/Y_{1:m}\right) \tag{28}$$

From Section 3, if the a posteriori density function $p(\ \boldsymbol{\theta}_{0:\ m}/Y_{1:\ m})$ is defined using SMC, the resulting model is linear and Gaussian, conditional to each parameter vector, and for $f(\boldsymbol{\alpha}_m, \boldsymbol{\theta}_m) = \boldsymbol{\alpha}_m(\boldsymbol{\theta}_m)$,(27) becomes:

$$E_{\hat{p}_N\left(\boldsymbol{\alpha}_m,\ \boldsymbol{\theta}_{0:m}/Y_{1:m}\right)}\left[\boldsymbol{\alpha}_m\left(\boldsymbol{\theta}_m\right)\right] \cong \frac{\sum_{i=1}^{N} w\left(\boldsymbol{\theta}^i_{0:m}\right) \int \boldsymbol{\alpha}_m\left(\boldsymbol{\theta}^i_m\right) p\left(\boldsymbol{\alpha}_m/\boldsymbol{\theta}^i_{0:m}, Y_{1:m}\ \right) d\boldsymbol{\alpha}_m}{\sum_{i=1}^{N} w\left(\boldsymbol{\theta}^i_{0:m}\right)} \tag{29}$$

It is equivalent to the modulation of the recursive Kalman filter output by the normalized weight of the corresponding parameter vector. We can write:

$$\hat{\boldsymbol{\alpha}}^{RBPF}_m = \sum_{i=1}^{N} W^i_{0:m} \int \boldsymbol{\alpha}_m\left(\boldsymbol{\theta}^i_m\right) p\left(\boldsymbol{\alpha}_m/\boldsymbol{\theta}^i_{0:m}, Y_{1:m}\ \right) d\boldsymbol{\alpha}_m \tag{30}$$

$$p\left(\boldsymbol{\theta}_{0:m}, Y_{1:m}\right) = p\left(\boldsymbol{\theta}_{0:m-1}, Y_{1:m-1}\right) p\left(Y_m/Y_{1:m-1}, \boldsymbol{\theta}_{0:m}\ \right) p\left(\boldsymbol{\theta}_m/\boldsymbol{\theta}_{m-1}\right) \tag{31}$$

The vector of parameters is assumed to evolve according a first-order Markov model, and its prior density is adopted as the importance density for the SIS, i.e.,

$$q\left(\boldsymbol{\theta}_m/\boldsymbol{\theta}_{0:m-1}, Y_{1:m}\right) = p\left(\ \boldsymbol{\theta}_m/\boldsymbol{\theta}_{m-1}\right) \tag{32}$$

Thus, $\boldsymbol{\theta}^i_m \sim p(\boldsymbol{\theta}_m/\boldsymbol{\theta}^i_{m-1})$, and the weights of the particles in (29) become:

$$w^i_{0:m} = w^i_{0:m-1}\ p\left(Y_m/Y_{1:m-1}, \boldsymbol{\theta}_{0:m}\right) \tag{33}$$

For each particle $\boldsymbol{\theta}^i_m$, we propagate the posterior mean and covariance using exact computations with a Kalman filter as follows:

$$\boldsymbol{\alpha}^i_{m/m-1} = \boldsymbol{A}^i_m\ \boldsymbol{\alpha}^i_{m-1/m-1} \tag{34}$$

$$\boldsymbol{P}^i_{m/m-1} = \boldsymbol{A}^i_m \boldsymbol{P}^i_{m-1/m-1} \boldsymbol{A}^{i\ T}_m + \boldsymbol{B}^i_m \boldsymbol{B}^{i\ T}_m \tag{35}$$

$$\boldsymbol{K}^i_m = \ \boldsymbol{P}^i_{m/m-1} \boldsymbol{C}^T_m \left[\boldsymbol{C}_m \boldsymbol{P}^i_{m/m-1} \boldsymbol{C}^T_m + \boldsymbol{D}_m \boldsymbol{D}^T_m\right]^{-1} \tag{36}$$

$$\boldsymbol{\alpha}^i_{m/m} = \boldsymbol{\alpha}^i_{m/m-1} + \ \boldsymbol{K}^i_m \left(Y_m - \boldsymbol{C_m} \boldsymbol{\alpha}^i_{m/m-1}\right) \tag{37}$$

$$\boldsymbol{P}^i_{m/m} = \left(\boldsymbol{I}_{p\times p} - \boldsymbol{K}^i_m \boldsymbol{C}_m\right) \boldsymbol{P}^i_{m/m-1} \tag{38}$$

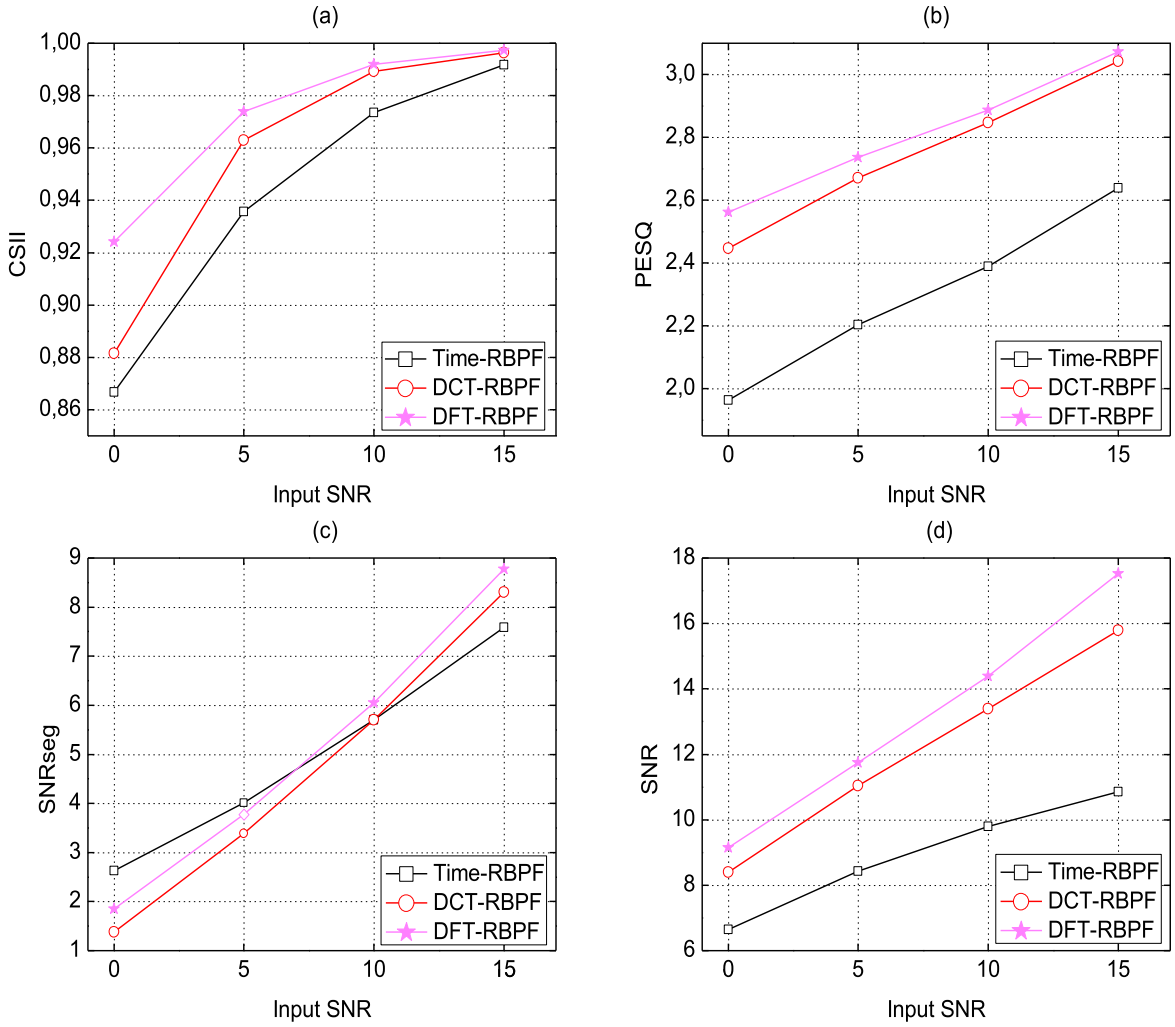

Fig. 2. Performance comparison of the proposed DFT-RBPF and speech enhancement algorithms based on PF, for AWGN and various input SNR levels, in terms of: (a) CSII, (b) PESQ, (c) SNRseg, and (d) overall SNR.

where $\boldsymbol{\alpha}^i_{m/m-1} = E[\boldsymbol{\alpha}^i_m/Y_{1:m-1}]$, $\boldsymbol{\alpha}^i_{m/m} = E[\boldsymbol{\alpha}^i_m/Y_{1:m}]$, $\boldsymbol{P}^i_{m/m-1} = Cov[\boldsymbol{\alpha}^i_m/Y_{1:m-1}]$, $\boldsymbol{P}^i_{m/m} = Cov[\boldsymbol{\alpha}^i_m/Y_{1:m}]$, and $\boldsymbol{K}^i_m$ is the Kalman filter gain. Hence, we have [6,18]:

$$p\left(Y_m/Y_{1:m-1}, \boldsymbol{\theta}^i_{1:m}\right) = \mathcal{N}\left(Y_m; \boldsymbol{C_m}\boldsymbol{\alpha}^i_{m/m-1},\ \boldsymbol{C_m}\boldsymbol{P}^i_{m/m-1}\boldsymbol{C}^T_m + \boldsymbol{D_m}\boldsymbol{D}^T_m\right) \tag{39}$$

The parameters and the Kalman filter output (a posteriori covariance, a posteriori mean) are then resampled before the transition step. Therefore, (33) becomes

$$w^i_m \ \propto p\left(Y_m/Y_{1:m-1},\ \boldsymbol{\theta}^i_{0:m}\right) \tag{40}$$

The components of the parameter vector are assumed to be independent and to evolve randomly according to the Gaussian random walk with the first-order Markov model. The stability of the TVAR model is maintained by keeping the instantaneous poles of the model, or the roots of the polynomial $(1 - \sum_{b=1}^{p} a_{b,m} z^{-b})$, strictly within the unit circle [6]. Similarly, the variance of the excitation is assumed to evolve according to a Gaussian random walk, and to keep it positive, the propagation is done over its logarithm. i.e.,:

$$p\left(\boldsymbol{\theta}_0\right) = p(\boldsymbol{a}_0)\ p\left(\phi_{U_0}\right) \tag{41}$$

$$p\left(\boldsymbol{\theta}_m/\boldsymbol{\theta}_{m-1}\right) = \ p(\boldsymbol{a}_m/\boldsymbol{a}_{m-1})\ p\left(\phi_{U_m}/\phi_{U_{m-1}}\right) \tag{42}$$

where, $p(\boldsymbol{a}_m/\boldsymbol{a}_{m-1}) = \ \mathcal{N}(\boldsymbol{a}_{m-1},\ \sigma^2_{\boldsymbol{a}}\boldsymbol{I}_{p\times p})$ under the condition of stability, and $p(\phi_{U_m}/\phi_{U_{m-1}}) = \ \mathcal{N}(\phi_{U_{m-1}},\ \sigma^2_{\phi_U})$, with $\phi_{U_m} = log\ (\sigma^2_{U_m})$.

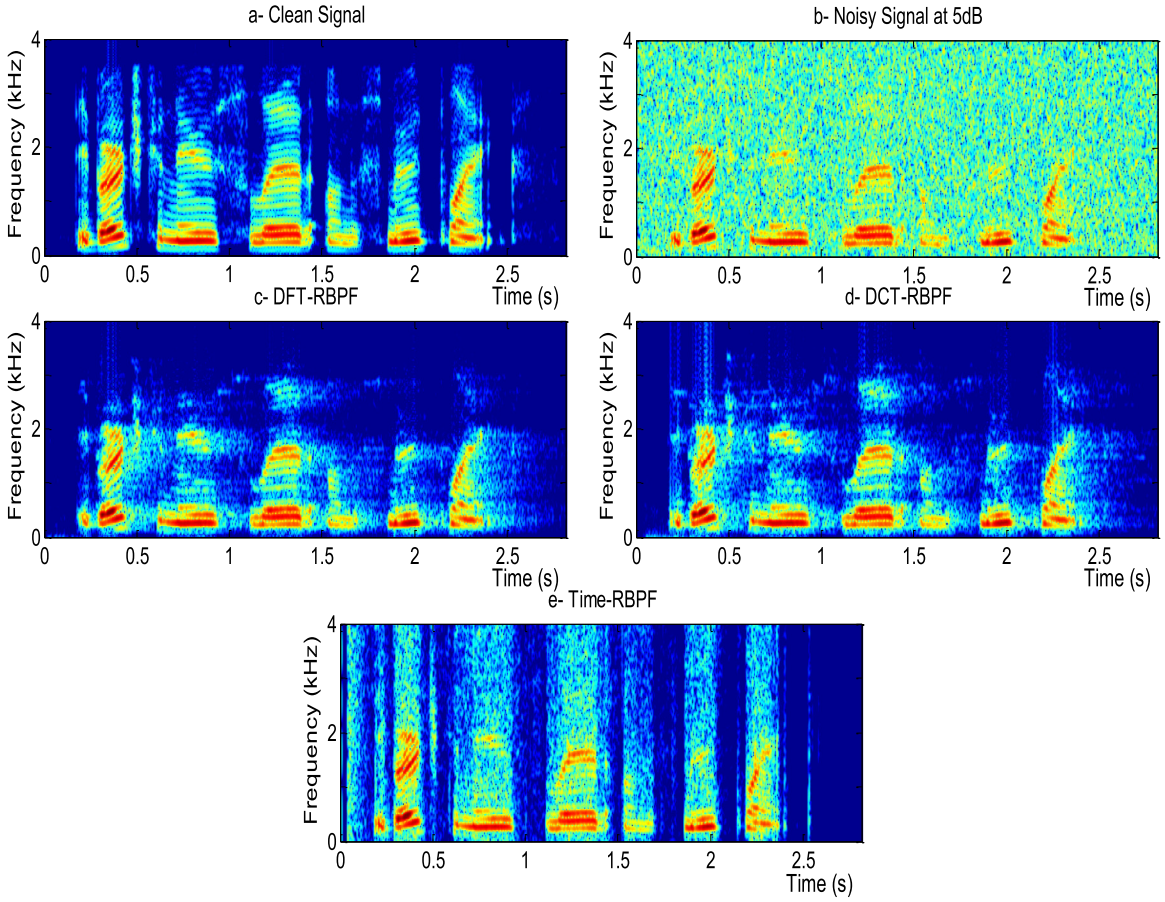

Fig. 3. Spectrograms of (a) the clean sentence (sp01.wav) 'The birch canoe slid on the smooth planks', (b) the sentence corrupted by AWGN at 5 dB SNR, the corrupted sentence enhanced by: (c) DFT-RBPF, (d) DCT-RBPF, and (e) Time-RBPF.

Table 1
RBPF algorithm for filtering real and imaginary parts of the noisy-speech complex DFT coefficients.

1: for $m \geq 1$
2: for $i = 1: N$
3: $\boldsymbol{a}^i_{m,k} \sim p(\boldsymbol{a}_{m,k}/\boldsymbol{a}^i_{m-1,k})$ (with stability condition)
4: $\phi^i_{U_{m,k}} \sim p(\phi_{U_{m,k}}/\phi^i_{U_{m-1,k}})$
5: Deduce the matrices in (5)
6: Compute the prediction step in the Kalman filter
7: Compute the weight according to (39)
8: Compute the correction step in the Kalman filter
9: end for i
10: Compute the sum of the weights
11: Normalize weight
12: Compute the sum of the weighted outputs $\hat{X}^{RBPF}_{m,k} = \sum_{i=1}^{N} W^i_{m,k}\, \boldsymbol{C}_{m,k}\, \hat{\boldsymbol{\alpha}}^{\,i}_{m,k}$
13: Resample according to the normalized weight: the parameters and the a posteriori covariance matrix and a posteriori mean for the next time step
14: end for m

The variance values $\{\sigma^2_{\boldsymbol{a}},\ \sigma^2_{\phi_U}\}$ are prefixed, and are the same for both parts of the DFT coefficient. Table 1 presents the RBPF algorithm for filtering the real and imaginary parts of the noisy-speech complex DFT coefficients.

5. Results and discussion

5.1. Experimental setup

For the proposed method evaluation, we use the NOIZEUS speech corpus [14]. The database contains 30 sentences, produced by three male and three female speakers, corrupted by eight different real-world noises at different SNR levels. The

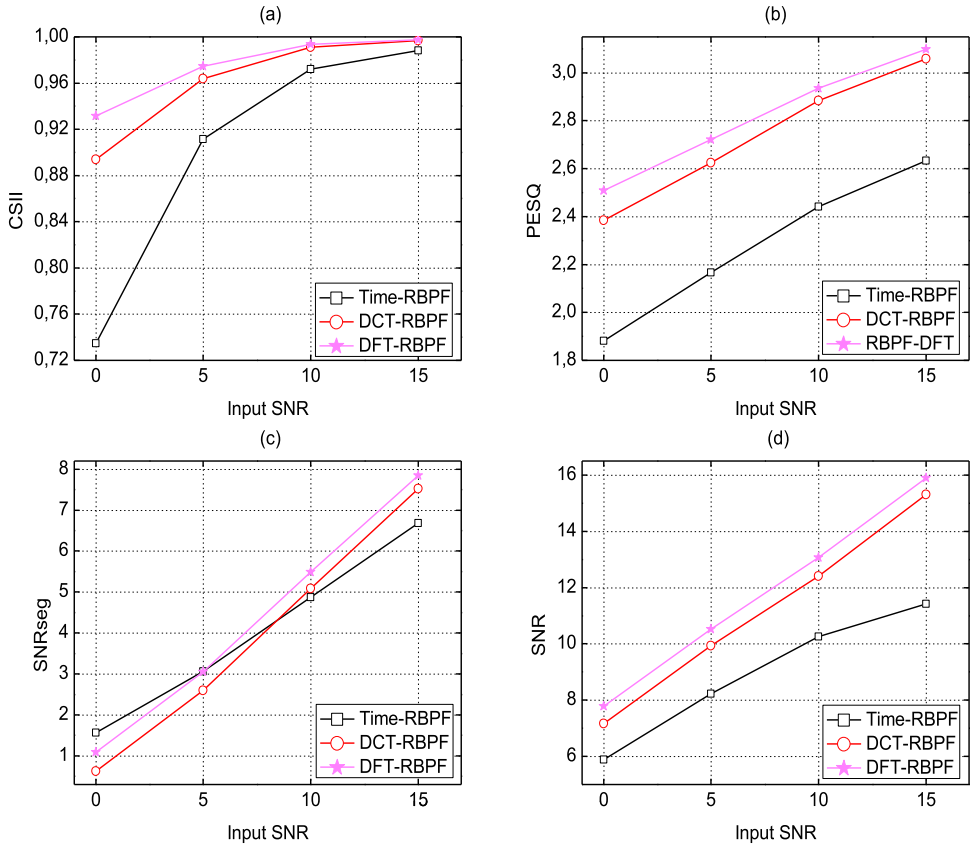

Fig. 4. Performance comparison of the proposed DFT-RBPF and enhancement algorithms based on PF, for Babble noise and various input SNR levels, in terms of: (a) CSII, (b) PESQ, (c) SNRseg, and (d) overall SNR.

phonetically balanced sentences were originally sampled at 25 kHz and downsampled to 8 kHz. From the database, we use the sentences corrupted with the Babble noise; and we also generate a corresponding synthetic set affected by AWGN at four SNR levels: 0 dB, 5 dB, 10 dB, and 15 dB

The algorithms are evaluated using the coherence speech intelligibility index (CSII), perceptual evaluation of speech quality (PESQ), segmental SNR (SNRseg) and overall SNR (SNR) in dB. The CSII [19] is based on the speech intelligibility index (SII) measure. It estimates the amount of speech information reaching a listener, by computing a weighted sum of the signal-to-distortion ratios (SDRs) instead of the SNRs used in the SII. The PESQ, ITU-T standard P.862, uses a perceptual model to convert the input and degraded speech into internal representations. The difference is used to estimate the mean opinion score (MOS). It was shown in [20] that the PESQ measure yielded the highest correlation in terms of the overall quality and signal distortion. The SNRseg values are evaluated in the time domain. For this measure, it is important that the original and processed signals are aligned in time and that any phase error present is corrected. The SNRseg is calculated by splitting the two signals into smaller segments and calculating the SNR value for each segment. The final SNRseg value is obtained by averaging the resulting per-segment SNR values. An extensive development of the used objective metric can be found in [14].

The performance results are obtained by averaging the resulting objective measure values from the used sentences. Owing to the stochastic nature of the proposed algorithm, the same noisy speech signals are enhanced several times using the proposed method.

The evaluation is achieved in two parts. In the first part, the DFT-RBPF algorithm is compared with the algorithms based on SMC. For this part, we assume perfect knowledge of the noise variance. Subsequently, the proposed DFT-RBPF algorithm is compared with the existing speech enhancement algorithms, and the noise variance is estimated during a silent frame, using a simple voice activity detector (VAD) based on the posterior SNR evaluated over each frame separately. For the noise variance initiation, the first five frames are considered as silence frames, which matches the used sentences. Furthermore, we introduce an overestimate factor that allows additional tuning for the estimated noise variances over each frame [21].

In the first part of the simulation, the following methods are compared:

- The proposed **DFT-RBPF** speech enhancement method with: the first order TVAR (1) model; 100 particles for each DFT coefficient part $\{\sigma^2_{\boldsymbol{a}} = 0.001,\ \sigma^2_{\phi_U} = 0.1\}$, , a Hanning window of length 20 ms, half overlapping, and 512 fast Fourier

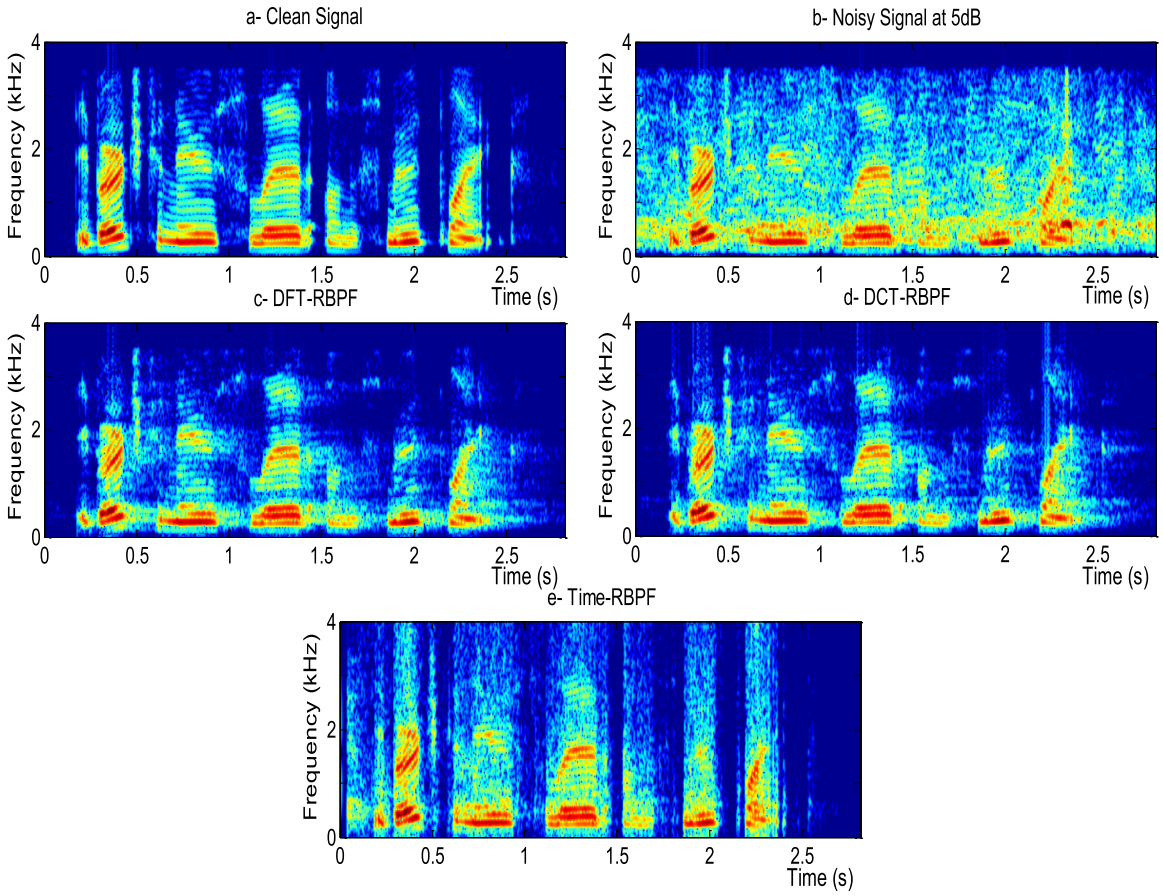

Fig. 5. Spectrograms of: (a) the clean sentence (sp01.wav) 'The birch canoe slid on the smooth planks', (b) the sentence corrupted by Babble noise at 5 dB SNR, the corrupted sentence enhanced by: (c) DFT-RBPF, (d) DCT-RBPF, and (e) Time-RBPF.

transform (FFT) samples. The used Gaussian random walk variances are fixed based on careful manual tuning with the objective of achieving the best performance.

- **DCT-RBPF** [10], with TVAR (2) and 100 particles.
- **Time-RBPF** [6], with the TVAR (10) model and 1000 particles.

In the second part of the simulation, the following methods are compared:

- The proposed **DFT-RBPF** speech enhancement method with : the first order TVAR (1) model, 100 particles for each DFT coefficient part, $\{\sigma_{\boldsymbol{a}}^2 = 0.001,\ \sigma_{\phi_U}^2 = 0.1\}$, a Hanning window of length 20 ms, half overlapping, and 512 FFT samples. The used Gaussian random walk variances are fixed based on careful manual tuning with the objective of achieving the best performance.
- **LSA:** minimum mean square error log spectral amplitude estimator [22].
- **MAP:** maximum a posteriori estimator of magnitude-squared spectrum [23].
- **GSS:** spectral amplitude estimators based on a parametric generalized spectral subtraction method [24]
- **NC-LSE:** log spectral amplitude estimation based on a non-causal a priori SNR estimator [25]
- **Wiener:** Wiener filter based on a priori signal to noise estimation [26]

5.2. Results and discussion

The performance comparisons between the proposed methods and the existing particle-filter speech enhancement algorithms, in terms of the CSII, PESQ, SNRseg, and the overall SNR, for the case of AWGN, are depicted in Fig. 2. The speech enhancement using DFT-RBPF results in better improvement of the objective measurement in terms of intelligibility (Fig. 2(a)). In terms of the speech-quality score, the proposed method outperforms the DCT-RBPF and the Time-RBPF (Fig. 2(b)) methods. In terms of the SNRseg (Fig. 2(c)), the Time-RBPF presents the best improvement for low input SNRs, as it performs better during the silent periods of speech segments [8]. However, the DFT-RBPF outperforms this method when the input

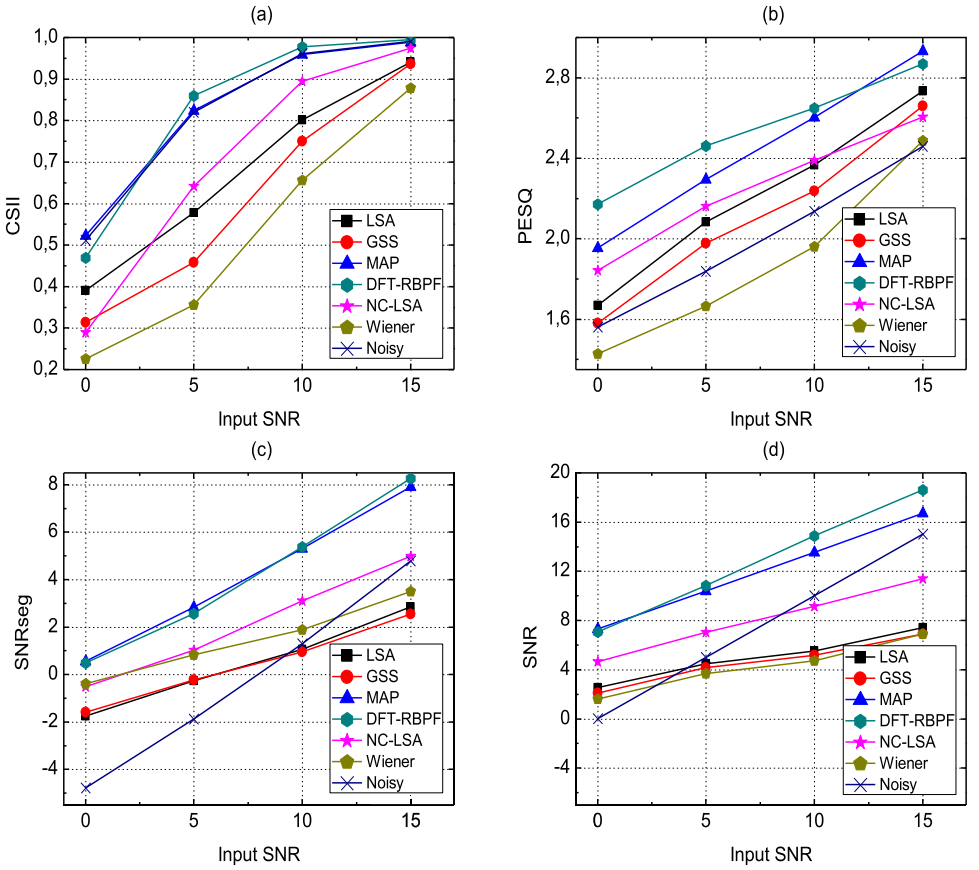

Fig. 6. Performance comparison of the proposed DFT-RBPF with other speech enhancement algorithms for AWGN and various input SNR levels in terms of: (a) CSII, (b) PESQ, (c) SNRseg, and (d) overall SNR.

SNR exceeds 5 dB. In terms of the overall SNR, the proposed method presents the highest SNR (Fig. 2(d)), when compared to the reference methods.

Fig. 3 shows the spectrograms for the speech processed by the proposed method and the existing methods based on the PF algorithm. A sample sentence was corrupted by AWGN at 5 dB SNR (Fig. 3(b)). The Time-RBPF (Fig. 3(e)) presents the lowest residual noise during the silent period. However, unlike the Time-RBPF, the DFT-RBPF (Fig. 3(c)) and DCT-RBPF (Fig. 3(d)) do not present a residual noise, that is modulated by speech power during the speech activity period, which confirms the performances presented in Fig. 2. In the cases of the DFT-RBPF and DCT-RBPF, the modulation occurs over each frequency channel; therefore, the residual noise is modified. From Fig. 3, we observe that the DFT-RBPF presents less speech distortion and residual noise when compared with the other methods.

Fig. 4 shows the performance comparisons between the proposed DFT-RBPF method and the existing particle-filter speech enhancement algorithms, in terms of the CSII, PESQ, SNRseg, and the overall SNR, for the case of Babble noise. The speech enhancement using the proposed DFT-RBPF gives the best intelligibility index (Fig. 4(a)), followed by that using the DCT-RBPF. In terms of speech-quality scores, the DFT-RBPF outperforms the DCT-RBPF and the Time-RBPF (Fig. 4(b)). For the SNRseg (Fig. 4(c)), the Time-RBPF exhibits the best improvement for low input SNRs, when compared with the other methods, nevertheless, this performance degrades rapidly at high SNRs because of the modulated residual noise that characterizes this method. In terms of the overall SNR (Fig. 4(d)), the DFT-RBPF shows the best improvement, followed by the DCT-RBPF.

Fig. 5 shows the spectrograms of speech processed by the proposed and existing methods based on the PF algorithm. The sample sentence was corrupted by Babble noise at 5 dB SNR (Fig. 5(b)). The DFT-RBPF (Fig. 5(c)) presents less residual noise and speech distortion, when compared with the existing speech enhancement methods, which is in accordance with the Fig 4(d). The resulting spectrograms from the DFT-RBPF and DCT-RBPF appear similar, but the DFT-RBPF contains more peaked formants, compared to the DCT-RBPF.

Fig. 6 shows the performance comparisons between the proposed method and the MAP, GSS, LSA, Wiener, and NC-LSA algorithms, in terms of the CSII, PESQ, SNRseg, and overall SNR, for the case of AWGN. In terms of the intelligibility index, the proposed method outperforms the reference methods (Fig. 6(a)). In terms of the PESQ score (Fig. 6(b)), the DFT-RBPF outperforms the reference methods for input SNRs ranging from 0 dB to 10 dB. However, the MAP presents the highest

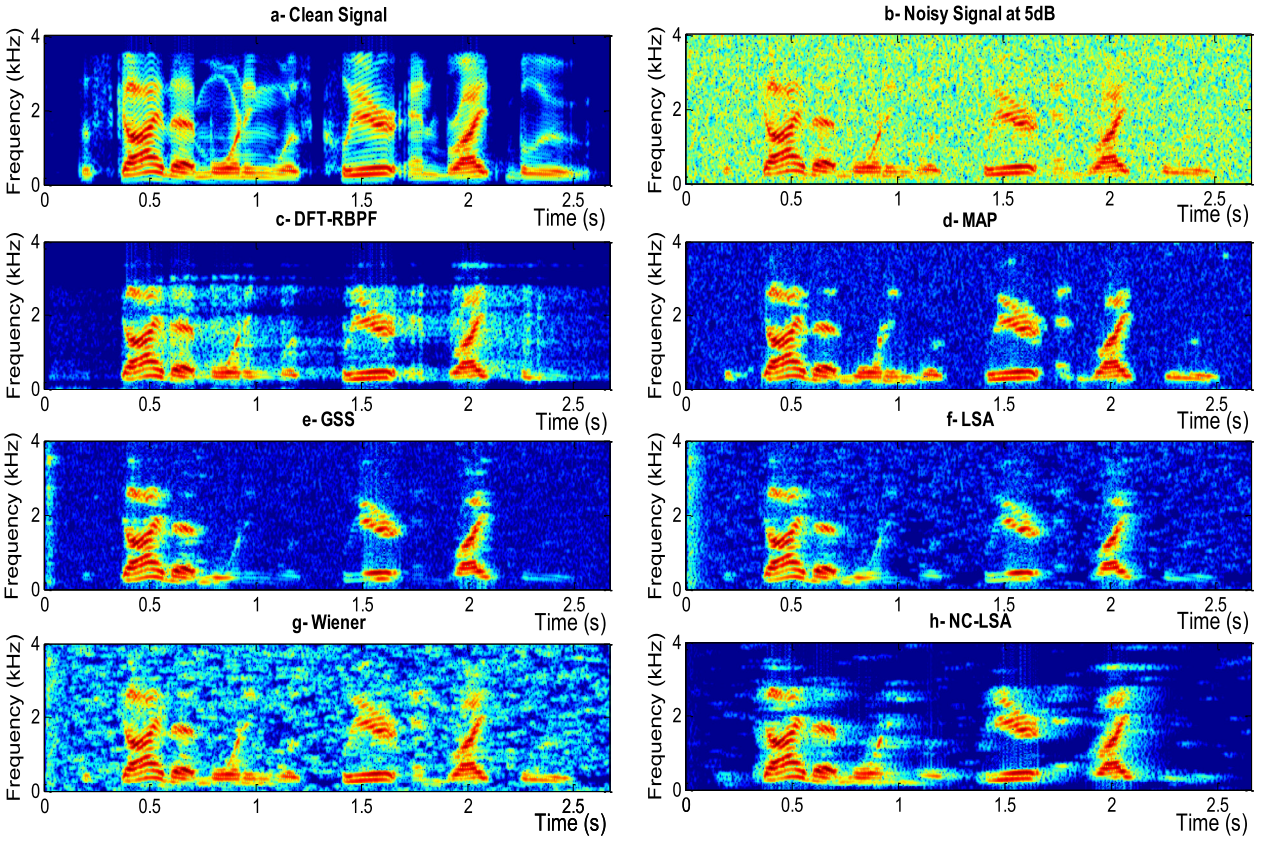

Fig. 7. Spectrograms of: (a) the clean speech sentence (sp10.wav) 'The sky that morning was clear and bright blue', (b) the sentence corrupted by AWGN at 5 dB SNR, the corrupted sentence enhanced by: (c) DFT-RBPF, (d) MAP, (e) GSS, (f) LSA, (g) Wiener filter, and (h) NC-LSA. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

speech quality score when the input SNR exceeds 10 dB. The DFT-RBPF (Fig. 6(c)) demonstrates a SNRseg performance that is close to that of the MAP. In terms of the overall SNR, the proposed method outperforms the reference methods used (Fig. 6(d)).

Fig. 7 shows the spectrograms of speech processed by the proposed method and the MAP, GSS, LSA, Wiener, and NC-LSA methods. The sample sentence was corrupted by AWGN at 5 dB SNR (Fig. 7(b)). The proposed DFT-RBPF (Fig. 7(c)) presents less residual noise and less speech distortion, and no musical noises are observed in the informal listening tests. The resulting spectrogram using the DFT-RBPF presents more apparent formants, compared to the other methods, which confirms the obtained objective results for this input SNR.

Fig. 8 shows the performance comparisons between the proposed method and the MAP, GSS, LSA, Wiener, and NC-LSA algorithms, in terms of the CSII, PESQ, SNRseg, and overall SNR, for the Babble noise. The intelligibility performances presented by the DFT-RBPF (Fig. 8(a)) are higher than those resulting from the other methods. The NC-LSA gives an intelligibility performance close to that of the MAP algorithm. In terms of the speech-quality scores, the DFT-RBPF gives the best score compared to the reference methods used (Fig. 8(b)). In Fig. 8(c), the evaluation using the SNRseg depicts that the Wiener filter based on a priori signal to noise estimation presents the best SNRseg for low input SNRs. However, the proposed DFT-RBPF gives the best SNRseg when the input SNR increases. In terms of the overall SNR, the DFT-RBPF presents the highest SNR improvement (Fig. 8(d)).

Fig. 9 shows the spectrograms of speech processed by the proposed method and the MAP, GSS, LSA, Wiener, and NC-LSA methods. The sample sentence was corrupted by Babble noise at 5 dB SNR (Fig. 9(b)). The proposed DFT-RBPF method (Fig. 9(c)) presents more residual noise, compared to the case of the AWGN, that matches linear Gaussian model. Likewise, the other methods also present more residual noise. Nevertheless, the proposed DFT-RBPF method presents the least residual noise and less speech distortion.

Table 2 presents the results of multiple comparisons of the statistical tests based on paired sample t-tests, which were conducted to assess the performance differences between the proposed method and the other methods. Differences between the scores are considered to be statistically significant if the obtained p -value (level of significance) is smaller than 0.05. In other words, the null hypothesis H0 (if the proposed method reaches the same or lower performance than the other methods) can be rejected if $p \leq 0.05$ [27]. From Table 2, it can be inferred that the tests are statistically significant ($p <$

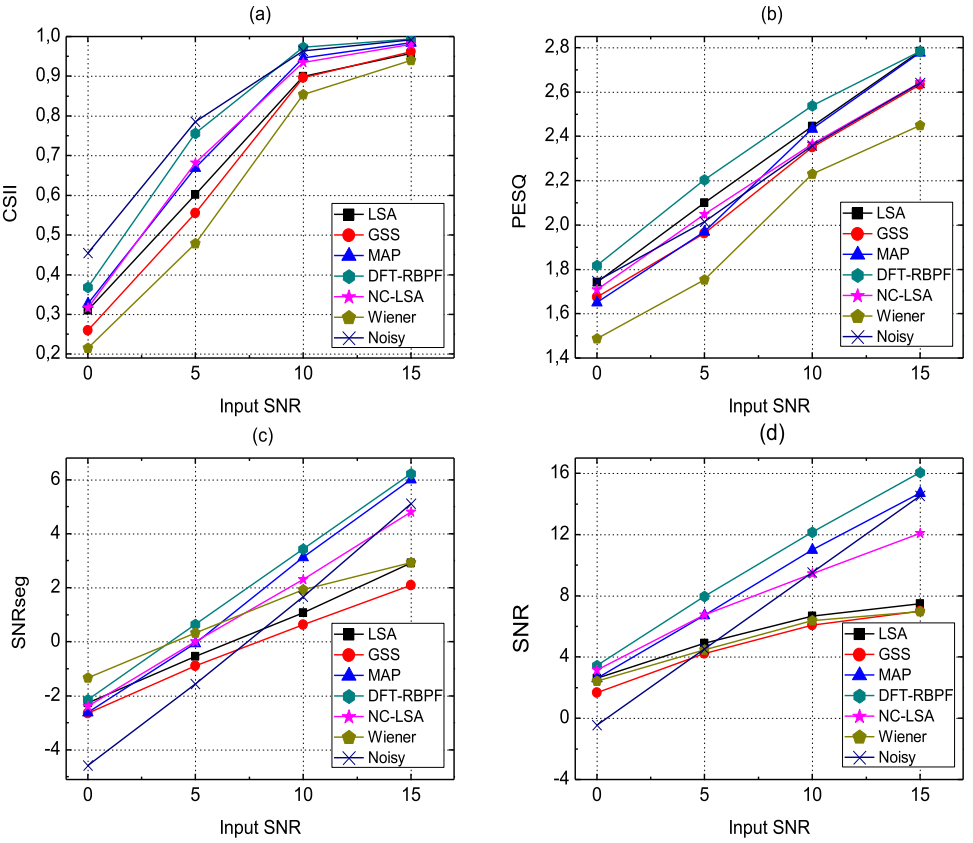

Fig. 8. Performance comparison of the proposed DFT-RBPF with other speech enhancement algorithms, for Babble noise and various SNR levels, in terms of: (a) CSII, (b) PESQ, (c) SNRseg, and (d) overall SNR.

Table 2
P -values for objective CSII, PESQ, SNRseg, and SNR evaluations of the comparisons with Time-RBPF, DFT-RBPF, DCT-RBPF, MAP, GSS, LSA, Wiener, and NC-LSA for AWG and additive Babble noise with 5 dB input SNR.

Noise type	Metric	p -value of the DFT-RBPF speech enhancement methods						
		DCT-RBPF	Time-RBPF	MAP	LSA	GSS	NC-LSA	Wiener
AWGN	CSII	0,0004039	0,030398	0,0063927	0,0000094	0,0000046	0,000001	0,000007
Babble noise		0,0017260	0,0000436	0,0016514	0,0003087	0,0001975	0,000812	0,000039
AWGN	PESQ	0,0003688	0,0000995	0,0007121	0,0017784	0,0008905	0,002873	0,000323
Babble noise		0,0021755	0,0000186	0,0000800	0,0256265	0,0012899	0,011358	0,000148
AWGN	SNR	0,0047343	0,0007641	0,0092629	0,0000351	0,0000349	0,002965	0,000023
Babble noise		0,0058705	0,0017841	0,0000399	0,0000005	0,0000007	0,008020	0,000001
AWGN	SNRseg	0,0000162	> 0.05	0,0323207	0,0001497	0,0001706	0,005272	0,005174
Babble noise		0,0001307	> 0.05	0,0015258	0,0000062	0,0000003	0,001803	0,034870

0.05) in terms of the CSII, PESQ score, and overall SNR. In terms of segmental SNR, the tests are statistically significant ($p < 0.05$) in all cases, except for the case of Time-RBPF, for AWG and Babble noises. The resulting statistical t-tests confirm the performance of the proposed speech enhancement method for the considered input SNR.

5.3. DFT-RBPF complexity

The Kalman filter speech enhancement complexity is $\mathcal{O}(p)$ if fast Kalman filter techniques are applied [5]. For L-points DFT, the complexity is $\mathcal{O}(L\log(L))$. The overall complexity of an N-particle Time-RBPF is approximately $\mathcal{O}(Np)$ computations per unit time [10]. Hence, considering the half overlapping decimated rate of $(L/2)$ times, compared to Time-RBPF, and taking into account the symmetric properties of the FFT, the DFT-RBPF complexity for (N) particles used with (L) frequency bins, is approximately $(\frac{(2(\frac{L}{2}-1)+1\ +\ 1)}{L/2}\ \mathcal{O}(Np)+\mathcal{O}(L\log(L))=\ \mathcal{O}(2Np)+\mathcal{O}(L\log(L)))$.

The complexities of the proposed DFT-RBPF, Time-RBPF and DCT-RBPF speech enhancement algorithms based on PF, depicted in Table 3, depend on the number of used particles and the adopted TVAR model order. The AR vectors are generated

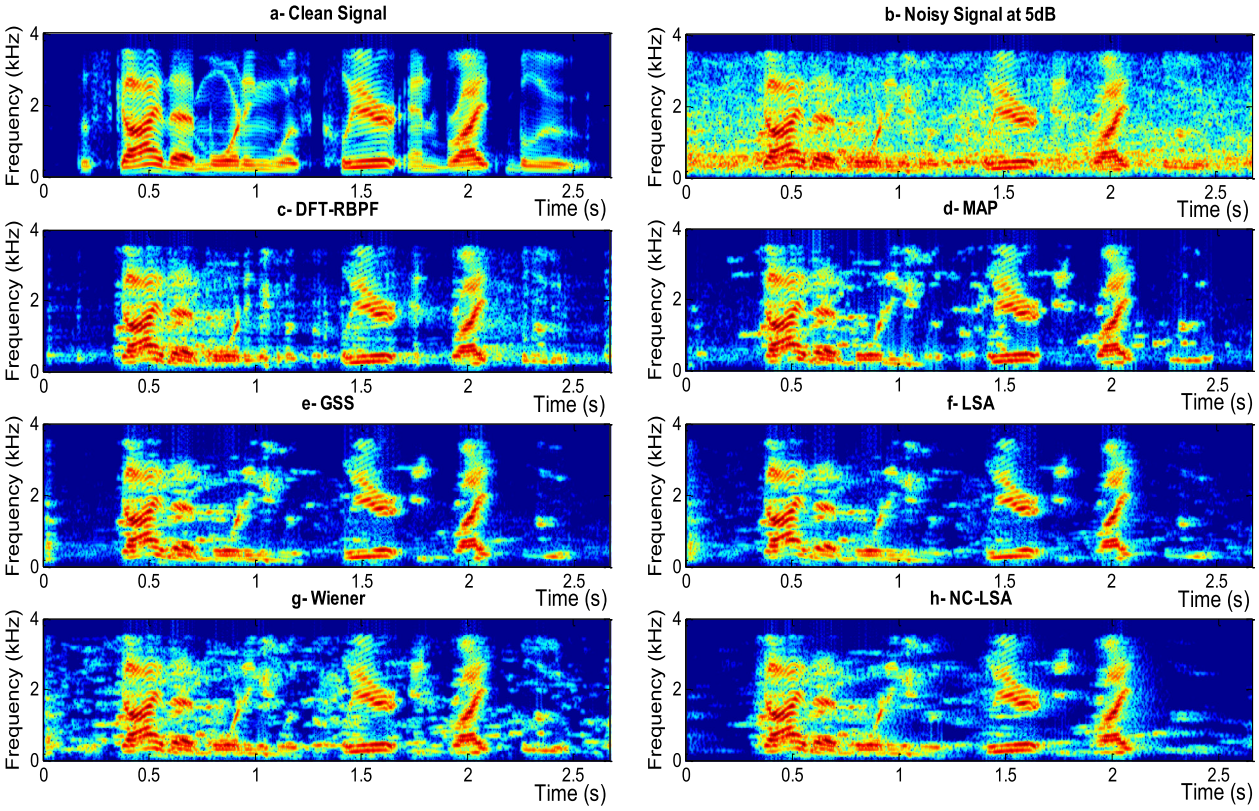

Fig. 9. Spectrograms of: (a) the clean sentence (sp10.wav) 'The sky that morning was clear and bright blue', (b) the sentence corrupted by Babble noise at 5 dB SNR, the corrupted sentence enhanced by: (c) DFT-RBPF, (d) MAP, (e) GSS, (f) LSA, (g) Wiener filter, and (h) NC-LSA. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Complexity of the Time-RBPF, DCT-RBPF, and the proposed DFT-RBPF.

Algorithms	Time-RBPF	DCT-RBPF	DFT-RBPF
Complexity	$\mathcal{O}(Np)$ [10]	$\mathcal{O}(Np) + \mathcal{O}(L\log(L))$	$\mathcal{O}(2Np) + \mathcal{O}(L\log(L))$.

by sampling until a stable draw is achieved, and the stability of the entire AR vector is affected by every coefficient. The number of vectors that are rejected as unstable will increase as the AR model order increases [10].

6. Conclusion

In this work, a new speech enhancement method was proposed: the DFT-RBPF, which was based on the assumption of independence between the real and imaginary parts of the noisy speech spectral components. The RBPF was used to estimate the vector parameters and the states using a low-order TVAR model that allowed us to account for the inter-frame correlation. The adopted low-order TVAR models and the used decimated frames significantly reduced the complexity of the proposed algorithm, compared to the Time-RBPF speech enhancement algorithm. The proposed DFT-RBPF algorithm was compared with the Time-RBPF and DCT-RBPF algorithms, under the same conditions, when the noise variance was assumed to be known. Results in terms of speech intelligibility, speech quality, SNRseg, and overall SNR indicated that the proposed DFT-RBPF method outperformed the reference methods based on the SMC, in the presence of AWGN and Babble noise.

Subsequently, the proposed DFT-RBPF method was compared with the MAP, GSS, LSA, Wiener, and NC-LSA methods, for the AWGN and additive Babble noise cases. The proposed method demonstrated more improvements in the adopted objective measures, compared to the reference methods. The obtained results were also reinforced by using paired sample t-tests, where the resulting p-values matched the obtained objective results. The proposed method can be improved by using a model approach that takes into account the transitions between active speech and silent intervals, and when the additive colored noise model is considered.

References

[1] McAulay RJ, Malpass ML. Speech enhancement using a soft-decision noise suppression filter. IEEE Trans Acoust Speech Sig Process 1980;28:137–44.
[2] Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Trans Acoust Speech Sig Process 1984;32:1109–21.
[3] Ho YC, Lee RCK. A Bayesian approach to problems in stochastic estimation and control. IEEE Trans Automat Control 1964;9:333–9.
[4] Dahia K. Nouvelles méthodes en filtrage particulaire-Application au recalage de navigation inertielle par mesures altimétriques *PhD thesis*. Grenoble, France: Univ Joseph Fourier; 2005.
[5] Paliwal KK, Basu A. A speech enhancement method based on Kalman filtering. In: Proc IEEE int conf acoust speech signal processing; 1987. p. 177–80.
[6] Vermaak J, Andrieu C, Doucet A, Godsill SJ. Particle methods for Bayesian modeling and enhancement of speech signals. IEEE Trans Speech Audio Process 2002;10:173–85.
[7] Doucet A, Godsill S, Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering. Stat Comput 2000;10:197–208.
[8] Mustiere F, Bouchard M, Boli M. Quality assessment of speech enhanced using particle filters. In: Proc IEEE int conf acoust speech signal processing; 2007. p. 1197–200.
[9] Mustiere F, Bouchard M, Bolic M. Low-cost modifications of Rao–Blackwellized particle filters for improved speech denoising. Sig Process 2008;88:2678–92.
[10] Laska B, Bolić M, Goubran R. Discrete cosine transform particle filter speech enhancement. Speech Commun 2010;52:762–75.
[11] Laska BNM, Bolic M, Goubran RA. Particle filter enhancement of speech spectral amplitudes. IEEE Trans Audio Speech Lang Process 2010;18:2155–67.
[12] Martin R. Speech enhancement using MMSE short time spectral estimation with gamma distributed speech priors. In: Proc IEEE int conf acoust speech signal processing, vol. 1; 2002. p. 253–6.
[13] Zavarehei E, Vaseghi S, Yan Q. Speech enhancement using Kalman filters for restoration of short-time DFT trajectories. In: Proc IEEE *int* workshop automatic speech recognition and understanding; 2005. p. 219–24.
[14] Loizou PC. Speech enhancement: theory and practice. CRC Press; 2013.
[15] Cohen I. Relaxed statistical model for speech enhancement and a priori SNR estimation. IEEE Trans Speech Audio Process 2005;13:870–81.
[16] Doucet A. Algorithmes monte Carlo pour l'estimation Bayésienne de modèle markovienne, application au traitement de signaux rayonnent *PhD thesis*. France: Univ Paris 11; 1997.
[17] Doucet A, Johansen AM. A tutorial on particle filtering and smoothing: fifteen years later. The oxford handbook of nonlinear filtering. New York: Oxford Univ. Press; 2009.
[18] de Freitas N. Rao–Blackwellised particle filtering for fault diagnosis. In: Proc IEEE int conf aerosp; 2002. p. 1767–72.
[19] Kates JM, Arehart KH. Coherence and the speech intelligibility index. J Acoust Soc Am 2005;117:2224–37.
[20] Ma J, Hu Y, Loizou PC. Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. J Acoust Soc Am 2009;125:3387.
[21] Berouti M, Schwartz R, Makhoul J. Enhancement of speech corrupted by acoustic noise. In: Proc IEEE int conf acoust speech signal processing; 1979. p. 208–11.
[22] Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans Acoust Speech Sig Process 1985;33:443–5.
[23] Lu Y, Loizou P. Estimators of the magnitude-squared spectrum and methods for incorporating SNR uncertainty. IEEE Trans Audio Speech and Lang Process 2011;19:1123–37.
[24] Sim BL, Tong YC, Chang JS, Tan CT. A parametric formulation of the generalized spectral subtraction method. IEEE Trans Speech Audio Process 1998;6:328–37.
[25] Cohen I. Speech enhancement using a noncausal a priori SNR estimator. IEEE Sig Process Lett 2004;11:725–8.
[26] Scalart P, Viera Filho J. Speech enhancement based on a priori signal-to-noise ratio estimation. In: Proc IEEE int conf acoust speech signal processing; 1996. p. 629–32.
[27] Saadoune A, Amrouche A, Selouani SA. Perceptual subspace speech enhancement using variance of the reconstruction error. Digital Sig Process 2014;24:187–96.

Mounir Meddah received the "Magister" degree (M.Sc.) in Telecommunication and Signal Processing from the Electronic department of the Sidi Bel Abbes University, Algeria in 2007. He is a member of the Speech Communication and Signal Processing Laboratory (LCPTS) at USTHB. His research interests include signal processing, speech enhancement, and communication systems.

Abderrahmane Amrouche received the "Magister" degree (M.Sc.) in 1995 and the "Doctorat d'Etat" (PhD) in real-time systems in 2007, both from the USTHB Algiers, where he is currently a full Professor of communication systems and an active member of the Speech Communication and Signal Processing Laboratory (LCPTS). His research interests include speech processing, robust speech and speaker recognition, and communication systems.

Abdelmalik Taleb-Ahmed received a PhD in Electronics and Microwaves from the University de Lille 1 in 1992. From 1992 to 2004, he was an associate professor at the University du Littoral, Calais. Since 2004, he has been a professor at the University of Valenciennes and conducts research at the LAMIH UMR CNRS 8201 UVHC; his research interests include signal and image processing.