ON THE APPLICATION OF THE UNSCENTED KALMAN FILTER TO SPEECH PROCESSING

Sharon Gannot
Faculty of Electrical Engineering, Technion, Technion City, 32000 Haifa, Israel
e-mail: gannot@siglab.technion.ac.il
Marc Moonen
Dept. of Elect. Eng. (ESAT-SISTA), K.U.Leuven, B-3001 Leuven, Belgium
e-mail: Marc.Moonen@esat.kuleuven.ac.be

ABSTRACT

In a series of recent studies a new approach for applying the Kalman filter to nonlinear system, referred to as Unscented Kalman filter (UKF), was proposed. In this contribution[1] we apply the UKF to several speech processing problems, in which a model with unknown parameters is given to the measured signals. We show that the nonlinearity arises naturally in these problems. Preliminary simulation results for artificial signals manifests the potential of the method.

1 INTRODUCTION

The recently proposed *unscented transform* (UT) is a method for calculating the statistics of a random variable undergoing a nonlinear transformation that was first suggested by Julier *et al.* [1]. This method was used to generalize the Kalman filter to nonlinear systems by Julier *et al.* [1] and was further extended by Wan *et al.* [2] to problems where both signals and parameters are jointly estimated. In [2] (and other contributions) the nonlinearity arises from the parameter production model.

In this contribution we further apply the UKF to several speech processing problems, namely single microphone speech enhancement and multi-microphone speech dereverberation. We show that in these applications the nonlinearity arises naturally, due to the signals and parameters multiplication, if both are given a dynamic model. The technique is demonstrated by several simple examples.

In Section 2 the *unscented transform* and its application to nonlinear Kalman filter are reviewed. Sections 3.1 and 3.2 discuss the application of the method to the problems of single microphone speech enhancement and two microphone speech dereverberation, respectively. We draw some conclusions and discuss some further directions in Section 4.

2 PRELIMINARIES

2.1 The Unscented Transform (UT)

Let $\boldsymbol{x}$ be an L-dimensional random vector with mean $\bar{\boldsymbol{x}}$ and covariance matrix P_{xx}. Let, $\boldsymbol{y} = f(\boldsymbol{x})$ be a nonlinear transformation from the random vector $\boldsymbol{x}$ to another random vector $\boldsymbol{y}$. The first and second order statistics of the vector $\boldsymbol{y}$ should be calculated. We briefly summarize the method. The mean and covariance of $\boldsymbol{x}$ are represented by $2L+1$ points and weights

$$\begin{aligned}
\mathcal{X}_0 &= \bar{\boldsymbol{x}} \\
\mathcal{X}_l &= \bar{\boldsymbol{x}} + \left(\sqrt{(L+\lambda)P_{xx}}\right)_l ;\ l = 1,\ldots,L \\
\mathcal{X}_{l+L} &= \bar{\boldsymbol{x}} - \left(\sqrt{(L+\lambda)P_{xx}}\right)_l ;\ l = 1,\ldots,L \\
W_0^{(m)} &= \lambda/(L+\lambda) \\
W_0^{(c)} &= \lambda/(L+\lambda) + (1-\alpha^2+\beta) \\
W_l^{(m)} &= W_l^{(c)} = 1/2(L+\lambda);\ l = 1,2,\ldots,2L
\end{aligned}$$

where, $\left(\sqrt{(L+\lambda)P_{xx}}\right)_l$ is the l-th row or column of the corresponding matrix square root, and $\lambda = \alpha^2(L+\kappa) - L$. α determines the spread of the sigma points. $\alpha = 1$ was used throughout our simulations . κ is a secondary scaling parameter. The choice $\kappa = 3 - L$ maintains the kurtosis of a Gaussian vector. Throughout our simulations κ is set to 0. β is used to incorporate prior knowledge of the distribution ($\beta = 2$ for Gaussian distributions). A proper choice of these parameters and its influence on the obtainable performance is still an open topic. The mean and covariance of the vector $\boldsymbol{y}$ are calculated using the following procedure,

1. Construct the sigma points $\mathcal{X}_l,\ l = 0,\ldots,2L$.
2. Transform each point: $\mathcal{Y}_l = f\left(\mathcal{X}_l\right),\ l = 0,\ldots,2L$.
3. Mean: Use weighted averaging, $\bar{\boldsymbol{y}} \approx \sum_{l=0}^{2L} W_l^{(m)} \mathcal{Y}_l$.
4. Covariance: Use weighted outer product, $P_{yy} \approx \sum_{l=0}^{2L} W_l^{(c)} \left(\mathcal{Y}_l - \bar{\boldsymbol{y}}\right)\left(\mathcal{Y}_l - \bar{\boldsymbol{y}}\right)^T$.

The benefits of using the UT are presented in [1] and [2].

2.2 The Application of the Unscented Transform to the Nonlinear Kalman Filtering Problem

The Kalman filter is a recursive and causal solution for *minimum mean square error* (MMSE) state estimation in the Gaussian and linear case. The Kalman equations are formulated with the state-space notation and consist of two stages. A *propagation* stage in which the mean and a priori covariance of the respective state are predicted based on the system dynamics and on the previous time instant estimate, and an *update* stage in which this prediction is optimally weighted with the new measurement. The error covariance, interpreted as the amount of confidence we have in the estimate, is propagated in a similar fashion.

[1]This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of the Interuniversity Attraction Pole IUAP P4-02, *Modeling, Identification, Simulation and Control of Complex Systems*, the Concerted Research Action *Mathematical Engineering Techniques for Information and Communication Systems* (GOA-MEFISTO-666) of the Flemish Government and the IT-poject *Multi-microphone Signal Enhancement Techniques for handsfree telephony and voice controlled systems (MUSETTE-2)* of the I.W.T., and was partially sponsored by Philips-ITCL.

When the system dynamics and the measurement equation are linear, all the calculations involved are straightforward. The situation is more complex when the involved equations are nonlinear. In this case, a method for propagating mean and covariance through nonlinearities is needed.

Let $\boldsymbol{s}(t)$ and $\boldsymbol{\theta}(t)$ be a signal state space vector and a parameter vector, respectively . $\boldsymbol{u}(t)$ and $\boldsymbol{v}(t)$ are innovation and measurement noise sequences, respectively. Define also the augmented state vector $\boldsymbol{x}^T(t) = \left[\, \boldsymbol{s}^T(t) \;\; \boldsymbol{\theta}^T(t) \,\right]$. Nonlinear transition and measurement equations are given by,

$$\begin{aligned} \boldsymbol{x}(t) &= \Phi\left(\boldsymbol{x}(t-1), \boldsymbol{u}(t)\right) \\ \boldsymbol{z}(t) &= h\left(\boldsymbol{x}(t-1), \boldsymbol{v}(t)\right). \end{aligned}$$

In the past the *extended Kalman filter* (EKF), based on the linearization of the equations, was used. This method might be quite complex, as it involves the calculation of derivatives, but yet not accurate enough, as only first-order approximation is applied.

A better method, proposed in [1], is to use the previously mentioned *unscented transform* in order to propagate the mean and covariance through the nonlinearities. Fig. 1 summarizes the steps involved in Unscented Kalman filter (UKF). The method consists of calculating the mean and covariance of the augmented state vectors undergoing a known nonlinear transform by virtue of the *unscented transform*. The complexity of the suggested method is quite low as only an increase of dimensions by a factor of $2L+1$ is required.

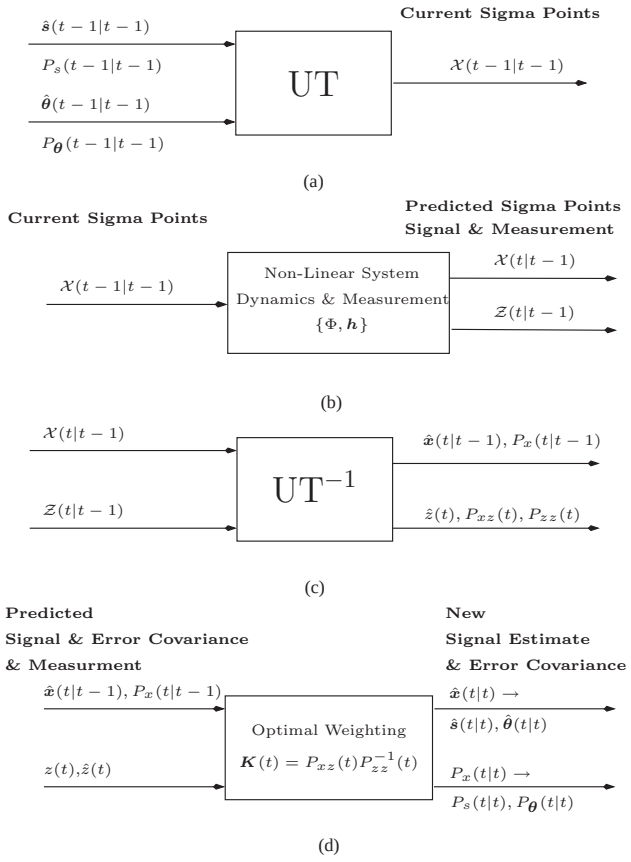

Figure 1: Unscented Kalman filter: (a) Unscented transform. (b) Propagation equations. (c) Inverse unscented transform. (d) Update equations.

3 APPLICATION TO SPEECH PROCESSING

In many model-based problems in speech processing (e.g. single microphone speech enhancement, multi-microphone speech enhancement and dereverberation) a problem of estimating both the speech signal and various parameters arises. This problem can be addressed in two ways. In the first, referred to by Wan *et al.* [2] as *dual* estimation, a two step approach is taken. In each time instant a Kalman filtering step for the signal is applied based on the current estimate of the parameters. In parallel a parameter estimate step is applied based on the current signal state estimate. The parameter estimation might be conducted using recursive methods such as RLS or LMS. Alternatively, under the Bayesian framework, the parameters can be given a dynamic model and the Kalman filter can be applied. This approach will be used throughout this work. The *dual* estimation method can be seen as a sequential variant of the *estimate-maximize* (EM) procedure, but no claims of optimality are valid. Discussion on the subject can be found in [3]. The method is summarized in Fig. 2 (top). The same problem can be reformulated into a *joint* estimation problem. Note that most operations involve parameter and state vector multiplications. Thus, the problem of *joint* estimation of the speech and the parameters becomes nonlinear if both are modelled as stochastic processes. We remark that as this nonlinearity is *separable* this formulation might lead to the same performance as in the *dual* scheme. This subject is still under investigation. The approach of jointly estimating speech signal and its parameters is summarized in Fig. 2 (bottom).

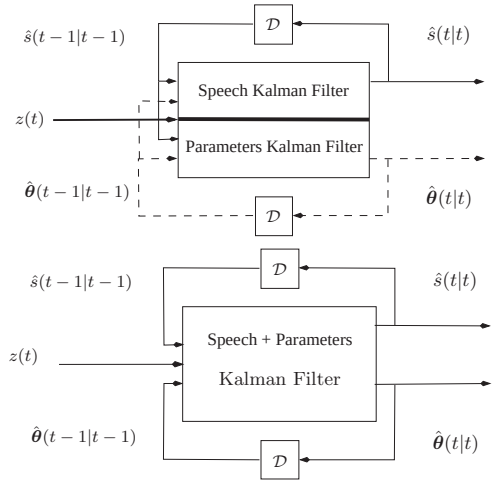

Figure 2: *Dual* (top) and *joint* (bottom) estimation procedures.

3.1 Single Microphone Speech Enhancement

The problem of single-microphone speech enhancement was extensively studied. Specifically, the use of Kalman filter for estimating both the signal and the parameters is presented by Gannot *et al.* [3]. By assuming AR model to the speech signal and giving dynamic model to the AR parameters both *dual* and *joint* schemes can be formulated. Each of the two steps comprising the *dual* scheme is linear, while the *joint* scheme consists of a single nonlinear step.

3.1.1 Signals Model

Let the signal measured by the microphone be given by $z(t) = s(t) + v(t)$, where $s(t)$ represents the sampled speech

signal and $v(t)$ represents an additive background noise. We shall assume a time varying AR model for the speech signal, i.e.

$$s(t) = -\sum_{k=1}^{p} \alpha_k(t) s(t-k) + g_s(t) u_s(t) \tag{1}$$

where the excitation $u_s(t)$ is a normalized (zero mean unit variance) white noise. $g_s(t)$ represents the innovation gain, and $\alpha_1(t), \alpha_2(t), \ldots, \alpha_p(t)$ are the AR coefficients. The additive noise $v(t)$ is assumed to be a realization from a zero mean white Gaussian stochastic with variance g_v^2. Define, $\boldsymbol{g}_s^T(t) = [g_s(t)\, 0 \ldots 0]$ and $\boldsymbol{h}_s^T = [1\, 0 \ldots 0]$. Then a state-space form is given by,

$$\begin{aligned} \boldsymbol{s}_p(t) &= \Phi_s(t)\boldsymbol{s}_p(t-1) + \boldsymbol{g}_s(t)u_s(t) \\ z(t) &= \boldsymbol{h}_s^T \boldsymbol{s}_p(t) + v(t) \end{aligned} \tag{2}$$

where $\boldsymbol{s}_p^T(t) = \big[s(t)\, s(t-1) \ldots s(t-p)\big]$. The signal state transition matrix $\Phi_s(t)$ is given by:

$$\Phi_s(t) = \begin{bmatrix} -\alpha_1(t) & -\alpha_2(t) & \cdots & \cdots & -\alpha_p(t) & 0 \\ 1 & 0 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 1 & 0 \end{bmatrix}. \tag{3}$$

3.1.2 Parameter model

Define the parameter vector $\boldsymbol{\alpha}^T(t) = [\alpha_1(t)\, \alpha_2(t) \ldots \alpha_p(t)]$ and the innovation vector $\boldsymbol{u}_{\boldsymbol{\alpha}}^T(t) = \big[u_{\alpha_1}(t)\, u_{\alpha_2}(t) \ldots u_{\alpha_p}(t)\big]$ with the respective covariance matrix $Q_{\boldsymbol{\alpha}}(t) = E\{\boldsymbol{u}_{\boldsymbol{\alpha}}(t)\boldsymbol{u}_{\boldsymbol{\alpha}}^T(t)\}$. The parameter state-space equations are,

$$\begin{aligned} \boldsymbol{\alpha}(t) &= \Phi_{\boldsymbol{\alpha}}\boldsymbol{\alpha}(t-1) + \boldsymbol{u}_{\boldsymbol{\alpha}}(t) \\ z(t) &= \boldsymbol{h}_{\boldsymbol{\alpha}}^T(t)\boldsymbol{\alpha}(t) + \boldsymbol{g}_s(t)u_s(t) + v(t), \end{aligned} \tag{4}$$

where, $\boldsymbol{h}_{\boldsymbol{\alpha}}^T(t) = \big[s(t-1)\, s(t-2) \ldots s(t-p)\big]$ and $\Phi_{\boldsymbol{\alpha}} = I_{p\times p}$ or very close to it.

3.1.3 Dual Scheme

On the one hand, assuming that the signal and all the noise parameters are known, which implies that $\Phi_s(t)$, $\boldsymbol{h}_s$ and $\boldsymbol{g}_s(t)$ are known, the optimal causal MMSE linear state estimate, which includes the desired speech signal $s(t)$, is obtained using the Kalman filtering equations. On the other hand, assuming the speech signal is known, i.e. $\boldsymbol{h}_{\boldsymbol{\alpha}}^T(t)$ is known, a Kalman filter for the parameter estimate might be applied. Since both signal and parameters are not known, the *dual* scheme presented in Fig. 2 may be applied. In each time instant the AR parameters are estimated using the estimated speech signal and the speech signal is estimated using the current parameter estimate.

3.1.4 Speech Kalman Filter

Propagation equations:

$$\begin{aligned} \hat{\boldsymbol{s}}_p(t|t-1) &= \Phi_s \hat{\boldsymbol{s}}_p(t-1|t-1) \\ P(t|t-1) &= \Phi_s P(t-1|t-1)\Phi_s^T + \boldsymbol{g}_s\boldsymbol{g}_s^T \end{aligned} \tag{5}$$

Kalman gain:

$$\boldsymbol{k}(t) = \frac{P(t|t-1)\boldsymbol{h}_s}{\boldsymbol{h}_s^T P(t|t-1)\boldsymbol{h}_s + g_v^2} \tag{6}$$

Update equations:

$$\begin{aligned} \hat{\boldsymbol{s}}_p(t|t) &= \hat{\boldsymbol{s}}_p(t|t-1) + \boldsymbol{k}(t)\Big[z(t) - \boldsymbol{h}_s^T\hat{\boldsymbol{s}}_p(t|t-1)\Big] \\ P(t|t) &= P(t|t-1) - \boldsymbol{k}(t)\Big[\boldsymbol{h}_s^T P(t|t-1)\boldsymbol{h}_s + g_v^2\Big]\boldsymbol{k}^T(t) \end{aligned} \tag{7}$$

3.1.5 Parameters Kalman Filter

Propagation equations:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}(t|t-1) &= \Phi_{\boldsymbol{\alpha}}\hat{\boldsymbol{\alpha}}(t-1|t-1) \\ P_{\boldsymbol{\alpha}}(t|t-1) &= \Phi_{\boldsymbol{\alpha}} P_{\boldsymbol{\alpha}}(t-1|t-1)\Phi_{\boldsymbol{\alpha}}^T + Q_{\boldsymbol{\alpha}} \end{aligned} \tag{8}$$

Kalman gain:

$$\boldsymbol{k}_{\boldsymbol{\alpha}}(t) = \frac{P_{\boldsymbol{\alpha}}(t|t-1)H_{\boldsymbol{\alpha}}}{\boldsymbol{h}_{\boldsymbol{\alpha}}^T P_{\boldsymbol{\alpha}}(t|t-1)\boldsymbol{h}_{\boldsymbol{\alpha}} + g_s^2(t) + g_v^2} \tag{9}$$

Update equations:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}(t|t) &= \hat{\boldsymbol{\alpha}}(t|t-1) + \boldsymbol{k}_{\boldsymbol{\alpha}}(t)\Big[z(t) - \boldsymbol{h}_{\boldsymbol{\alpha}}^T\hat{\boldsymbol{\alpha}}(t|t-1)\Big] \\ P_{\boldsymbol{\alpha}}(t|t) &= P_{\boldsymbol{\alpha}}(t|t-1) - \\ &\quad \boldsymbol{k}_{\boldsymbol{\alpha}}(t)\Big[\boldsymbol{h}_{\boldsymbol{\alpha}}^T P_{\boldsymbol{\alpha}}(t|t-1)\boldsymbol{h}_{\boldsymbol{\alpha}} + g_s^2(t) + g_v^2\Big]\boldsymbol{k}_{\boldsymbol{\alpha}}^T(t). \end{aligned} \tag{10}$$

The *dual* scheme suggested in Fig. 2 (top) is then used.

3.1.6 Joint Scheme

An augmented state vector of the speech and the parameters is constructed, $\boldsymbol{x}^T(t) = \big[\,\boldsymbol{s}_p(t)\;\; \boldsymbol{\alpha}(t)\,\big]$. Then,

$$\begin{aligned} \boldsymbol{x}(t) &= \underbrace{\begin{bmatrix} \Phi_s & \boldsymbol{0} \\ \boldsymbol{0} & \Phi_{\boldsymbol{\alpha}} \end{bmatrix}\boldsymbol{x}(t-1)}_{\text{nonlinearity}} + \begin{bmatrix} \boldsymbol{g}_s(t)u_s(t) \\ \boldsymbol{u}_{\boldsymbol{\alpha}}(t) \end{bmatrix} \\ z(t) &= \big[\,1\;0\;0\;\ldots\;0\,\big]\,\boldsymbol{x}(t) + v(t). \end{aligned} \tag{11}$$

This set of equation is nonlinear since it involves a multiplication of the speech state space and the transition matrix comprised of the parameters process. So, the *joint* scheme suggested in Fig. 2 (bottom) can be used.

3.1.7 Results

Time varying Gaussian AR process (4 coefficients) embedded in white Gaussian noise with input SNR level of about 20dB is processed by the *joint* Kalman scheme[2]. The noise level is estimated during non-signal portions of the noisy signal. The tracking ability of the procedure is presented in Fig. 3. The

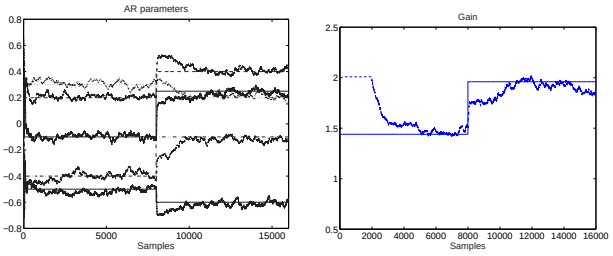

Figure 3: Tracking ability of the parameters of an AR process embedded in white noise.

performance with real speech signals is still to be determined.

[2]All Simulations in this paper are implemented by modifying R. van der Merwe *et al.* [4] code, written in Matlab© language.

3.2 Two Microphone Speech Dereverberation

In the two channel dereverberation problem a speech signal, modelled as an AR process, is filtered by an *acoustical transfer function* (ATF), modelled as an FIR filter. Noise is then added to the output constructing the noisy and reverberated speech signals, as depicted in Fig. 4.

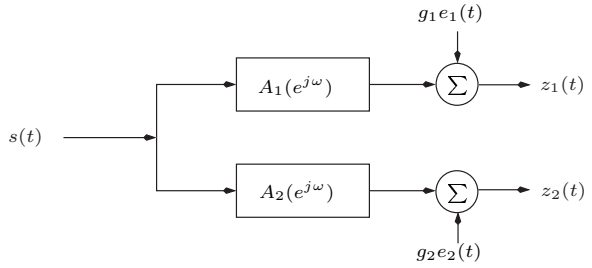

Figure 4: Two channel dereverberation problem.

3.2.1 Signals Model

The reverberated and noisy signals presented in Fig. 4 are given by the following model,

$$s(t) = -\sum_{k=1}^{p} \alpha_k s(t-k) + g_{u_s}(t) u_s(t) \qquad (12)$$

$$z_1(t) = \sum_{k=0}^{n_a-1} a_1(k) s(t-k) + g_1 e_1(t)$$

$$z_2(t) = \sum_{k=0}^{n_a-1} a_2(k) s(t-k) + g_2 e_2(t).$$

Thus, we have again a problem of estimating both the speech signal and the following model parameters,

$$\boldsymbol{\theta}^T(t) = \left[\boldsymbol{\alpha}(t) \;\; g_{u_s}(t) \;\; \boldsymbol{a}_1(t) \;\; \boldsymbol{a}_2(t) \;\; g_1 \;\; g_2 \right].$$

3.2.2 Joint Speech and Parameters Estimation

Define,

$$\begin{aligned}
\boldsymbol{s}_{n_a}^T(t) &= \left[s(t) \;\; s(t-1) \;\; \cdots \;\; s(t-n_a+1) \right] \\
\boldsymbol{g}_s^T(t) &= \left[g_s(t) \;\; 0 \;\; \cdots \;\; 0 \right] \\
\boldsymbol{u}_{\boldsymbol{\alpha}}^T(t) &= \left[u_{\alpha_1}(t) \;\; u_{\alpha_2}(t) \;\; \cdots \;\; u_{\alpha_p}(t) \right] \\
\boldsymbol{u}_{\boldsymbol{a}_1}^T(t) &= \left[u_{a_1^1}(t) \;\; u_{a_1^2}(t) \;\; \cdots \;\; u_{a_1^{n_a}}(t) \right] \\
\boldsymbol{u}_{\boldsymbol{a}_2}^T(t) &= \left[u_{a_2^1}(t) \;\; u_{a_2^2}(t) \;\; \cdots \;\; u_{a_2^{n_a}}(t) \right]
\end{aligned}$$

and $\Phi_s(t)$ an $n_a \times n_a$ signal transition matrix having equivalent structure to the one presented in (3). Then, the augmented transition-measurement equations can be written as,

$$\begin{bmatrix} \boldsymbol{s}_{n_a}(t) \\ \boldsymbol{\alpha}(t) \\ \boldsymbol{a}_1(t) \\ \boldsymbol{a}_2(t) \end{bmatrix} = \underbrace{\begin{bmatrix} \Phi_s(t) & 0 & 0 & 0 \\ 0 & I_p & 0 & 0 \\ 0 & 0 & I_{n_a} & 0 \\ 0 & 0 & 0 & I_{n_a} \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_{n_a}(t-1) \\ \boldsymbol{\alpha}(t-1) \\ \boldsymbol{a}_1(t-1) \\ \boldsymbol{a}_2(t-1) \end{bmatrix}}_{\text{nonlinearity}} + \begin{bmatrix} \boldsymbol{g}_s u_s(t) \\ \boldsymbol{u}_{\boldsymbol{\alpha}}(t) \\ \boldsymbol{u}_{\boldsymbol{a}_1}(t) \\ \boldsymbol{u}_{\boldsymbol{a}_2}(t) \end{bmatrix}$$

$$\begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{a}_1(t) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{a}_2(t) & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_{n_a}(t) \\ \boldsymbol{\alpha}(t) \\ \boldsymbol{a}_1(t) \\ \boldsymbol{a}_2(t) \end{bmatrix}}_{\text{nonlinearity}} + \begin{bmatrix} g_1 e_1(t) \\ g_2 e_2(t) \end{bmatrix}$$

which is a nonlinear set of equations, fitting the UKF framework.

3.2.3 Results

For a low level white noise signal, which variance is estimated from signal free segments, the tracking ability of the algorithm is presented in Fig. 5. It is worth mentioning that the presented problem is a very simple one, the order of the AR process is 1 and the filters $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ are 3 taps long. The SNR value is very high. Even in this simple case convergence is not guaranteed.

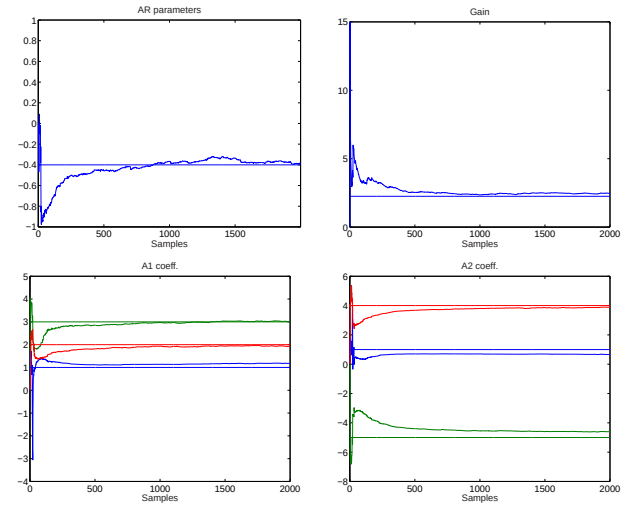

Figure 5: Tracking ability of the parameters of the dereverberation problem.

4 DISCUSSION

In this paper we applied the newly proposed UKF to two speech processing problems. Results show that the method is applicable to the problems in hand. Nevertheless, for a comprehensive test, it should be further applied to real speech signals embedded in higher noise levels. Performance limitations and optimality issues of the suggested method are under current research.

5 *

References

[1] S. Julier, J. Uhlmann and H.F. Durrant-Whyte, "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators," *IEEE trans. on Automatic Control*, vol. 45, no. 3, pp. 477–482, Mar. 2000.

[2] E. A. Wan and R. van der Merwe, "The Unscented Kalman Filter for Nonlinear Estimation," in *Symposium 2000 on Adaptive Systems for Signal Processing, Communication and Control (AS-SPCC)*, Lake Louise, Alberta, Canada, Oct. 2000, IEEE.

[3] S. Gannot, D. Burshtein, and E. Weinstein, "Iterative and Sequential Kalman Filter-Based Speech Enhancement Algorithms," *IEEE Trans. on Speech and Audio Proc.*, vol. 6, no. 4, pp. 373–385, Jul. 1998.

[4] R. van der Merwe, "Matlab©code," `/users/sista/sgannot/matlab/Ukf_W/`, May 2001.