Improved colored noise handling in Kalman Filter-based speech enhancement algorithms

F. Mustière, M. Bolić, M. Bouchard

Ottawa University

May 5th 2008

Outline

Current Topic

Kalman Filter based speech enhancement

- Model-based enhancement
- Speech modelled as autoregression
- Problem formulated by state-space equations
- Fairly large and established family of algorithms

State-space model, white observation noise

Time varying autoregressive model:

clean signal $x(n) = \sum_{k=1}^{p} a_k(n)x(n-k) + \sigma_e(n)e(n)$

measurement signal $y(n) = x(n) + \sigma_v(n)v(n)$

In matrix form:

clean signal $\mathbf{x}(n) = \mathbf{A}_k(n)\mathbf{x}(n-k) + \mathbf{G}(n)e(n)$

measurement signal $y(n) = \mathbf{C}\mathbf{x}(n) + \sigma_v(n)v(n)$

Traditional colored noise handling

"Traditional" way of augmenting system:

$$
\begin{aligned}
\text{clean signal} \quad & \mathbf{x}(n) = \mathbf{A}_k(n)\mathbf{x}(n-k) + \mathbf{G}(n)e(n) \\
\text{noise signal} \quad & \mathbf{n}(n) = \mathbf{B}_k(n)\mathbf{n}(n-k) + \mathbf{H}(n)v(n) \\
\text{measurement equation} \quad & y(n) = \mathbf{C}\mathbf{x}(n) + \mathbf{D}\mathbf{n}(n)
\end{aligned}
$$

with $\mathbf{A}$, $\mathbf{B}$ respectively of size $p \times p$, $q \times q$.

Traditional colored noise handling

Remarks:

- Redundancy in the (large) state-vector
- Noise-free observation equation $\Rightarrow$ potentially very small error covariance matrix $\Rightarrow$ potential stability problems
- Does not reduce to white noise state-space equations for 0 AR order

Current Topic

Rewriting of the state-space equations:

Traditional (again)

$$\begin{aligned} \text{clean signal} \quad & \mathbf{x}(n) = \mathbf{A}_k(n)\mathbf{x}(n-k) + \mathbf{G}(n)e(n) \\ \text{noise signal} \quad & \mathbf{n}(n) = \mathbf{B}_k(n)\mathbf{n}(n-k) + \mathbf{H}(n)v(n) \\ \text{measurement equation} \quad & y(n) = \mathbf{C}\mathbf{x}(n) + \mathbf{D}\mathbf{n}(n) \end{aligned}$$

Proposed

$$\begin{aligned} \text{clean signal} \quad & \mathbf{x}(n) = \mathbf{A}_k(n)\mathbf{x}(n-k) + \mathbf{G}(n)e(n) \\ \text{measurement equation} \quad & y(n) = \widetilde{\mathbf{C}}\mathbf{x}(n) + \mathbf{b}_k^T(n)\mathbf{y}(n-1) + \sigma_v(n)v(n) \end{aligned}$$

with $\mathbf{y}(n)$ a trail of measurements of size q, and
$\widetilde{\mathbf{C}} = [1, \; -\mathbf{b}_k^T(n), \; 0 \ldots 0]$.

Apparent advantages

Remarks

- Smaller state-vector, no redundancy
- Less computations, less memory required
- No “noise-free” measurement equation
- Naturally reduces to white noise state-space model

Gain in efficiency

Parameters		Regular			Proposed		
M_s	M_n	$\times$	$+$	$-$	$\times$	$+$	$-$
8	3	234	144	67	167	113	37
10	3	315	196	92	236	159	56
10	6	459	289	137	275	198	56
12	6	570	361	172	362	258	79
12	8	693	441	211	392	288	79

Table: Examples of computational load for both types of KFs.[1]

- M_s = speech AR order
- M_n = noise AR order

[1] Detailed complexity analysis available in paper

Current Topic

Test algorithms and conditions

Algorithms used

1. Plain/Classic Kalman Filter, true AR parameters measured from speech and noise signals, **PKF**
2. Rao-Blackwellized Particle Filter (Vermaak, 2002), and true AR parameters measured from noise signal, **RBPF**
3. KF with EM algorithm (Gannot, 1998) updating speech AR parameters, and true AR parameters measured from noise signal, **KEM**

Results

Type of algorithm	*Quality measure*	Regular	Proposed
Cafeteria noise			
Noisy speech	SNR wPESQ	3.57 1.33	
PlainKF	SNR wPESQ	6.30 1.61	7.79 1.71
KEM	SNR wPESQ	4.66 1.51	4.65 1.53
RBPF	SNR wPESQ	6.07 1.41	5.86 1.50

Table: Experimental results in cafeteria noise

Results

Type of algorithm	*Quality measure*	Regular	Proposed
Stationary hoth noise			
Noisy speech	SNR wPESQ	2.96 1.29	
PlainKF	SNR wPESQ	5.78 1.62	7.32 1.73
KEM	SNR wPESQ	4.72 1.56	4.71 1.57
RBPF	SNR wPESQ	5.77 1.50	6.00 1.56

Table: Experimental results in hoth noise

Current Topic

Conclusions

- More efficient implementation based on simple rewriting of state-space models
- Equivalent (in some cases better) results
- No apparent disadvantage
- Ready to be used as part of any state-space based speech enhancement algorithm

Questions?

Frédéric Mustière
mustiere@site.uottawa.ca

Regular KF iteration

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{H}_k \eta_k \quad (1)$$
$$z_k = \mathbf{D}\mathbf{x}_k \quad (2)$$

KF Step, update of $[\mathbf{x}, \ \mathbf{P}_1]$

1. $\mathbf{P}_{t_1} = \mathbf{F}\mathbf{P}_1\mathbf{F}^T + \mathbf{H}\mathbf{H}^T$
2. $t_1 = \mathbf{D}\mathbf{P}_{t_1}\mathbf{D}^T$
3. $\mathbf{x}_t = \mathbf{F}\mathbf{x}$
4. $y_1 = \mathbf{D}\mathbf{x}_t$
5. $\mathbf{J}_1 = \mathbf{P}_{t_1}\mathbf{D}^T t_1^{-1}$
6. $\mathbf{x} = \mathbf{x}_t + \mathbf{J}_1(z - y_1)$
7. $\mathbf{P}_1 = \mathbf{P}_{t_1} - \mathbf{J}_1\mathbf{D}\mathbf{P}_{t_1}$

Proposed KF iteration

$$\mathbf{s}_k = \mathbf{A}_k \mathbf{s}_{k-1} + \mathbf{G}_k \mathbf{w}_k \quad (3)$$

$$z_k = \widetilde{\mathbf{C}}_k \mathbf{s}_k + \mathbf{b}_k{}^T \mathbf{z}_{k-1} + \sigma_{v,k} v_k \quad (4)$$

where $\widetilde{\mathbf{C}}_k = \left[1, \ -\mathbf{b}_k^T, \ \mathbf{0}_{1\times M_s - M_n - 1}\right]$

KF Step, update of $[\mathbf{s}, \ \mathbf{P}_2]$

1. $\mathbf{P}_{t_2} = \mathbf{A}\mathbf{P}_2\mathbf{A}^T + \mathbf{G}\mathbf{G}^T$
2. $t_2 = \widetilde{\mathbf{C}}\mathbf{P}_{t_2}\widetilde{\mathbf{C}}^T + \sigma_v^2$
3. $\mathbf{s}_t = \mathbf{A}\mathbf{s}$
4. $y_2 = \widetilde{\mathbf{C}}\mathbf{s}_t + u$
5. $\mathbf{J}_2 = \mathbf{P}_{t_2}\widetilde{\mathbf{C}}^T t_2^{-1}$
6. $\mathbf{s} = \mathbf{s}_t + \mathbf{J}_2(z - y_2)$
7. $\mathbf{P}_2 = \mathbf{P}_{t_2} - \mathbf{J}_2\widetilde{\mathbf{C}}\mathbf{P}_{t_2}$