Weighted Prediction Error Algorithm

Wang Jian

October 30, 2015

Speech Dereverbration

- Blind Deconvolution
  - Blind
    Depends solely on observed signal $y[n]$
  - Deconvolution
    Model speech source and channel explicitly
- Statistical Model-Based Approach

Speech Dereverbration

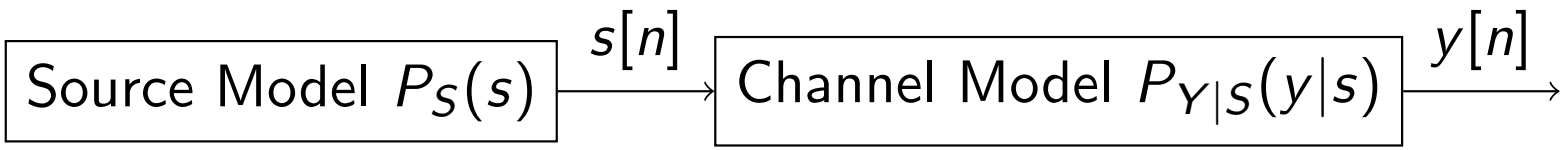
- Blind Deconvolution
  - Blind
    Depends solely on observed signal $y[n]$
  - Deconvolution
    Model speech source and channel explicitly
- Statistical Model-Based Approach

Speech Dereverbration

- Blind Deconvolution
  - Blind
    Depends solely on observed signal $y[n]$
  - Deconvolution
    Model speech source and channel explicitly
- Statistical Model-Based Approach

Speech Dereverbration

- Blind Deconvolution
  - Blind
    Depends solely on observed signal $y[n]$
  - Deconvolution
    Model speech source and channel explicitly
- Statistical Model-Based Approach



Speech Generation Model

- Time-Varying Gaussian Model(TVG)

$$p(s;\Phi) = \prod_{l=0}^{L-1}\prod_{n=0}^{N-1} p(s_{n,l};\Phi)$$

$$\begin{aligned} p(s_{n,l};\Phi) &= f_{\mathcal{N}_{\mathbb{C}}}(s_{n,l};0,\lambda_{n,l}) \\ &= \frac{1}{\pi\lambda_{n,l}} e^{-\frac{|s_{n,l}|^2}{\lambda_{n,l}}} \end{aligned}$$

Acoustical Transmission Channel Model

- Multiple-Input Single-Output AutoRegressive Model(MISO AR)

$$y_{1,n,l} = \sum_{k=\Delta_l}^{\Delta_l+K_l-1} \mathbf{g}_{k,l}^H \mathbf{y}_{n-k,l} + s_{n,l}$$

$$p(y|s;\Psi) = \prod_{l=0}^{L-1} \prod_{n=0}^{N-1} p(y_{1,n,l}|\mathbf{y}_{n-\Delta_l,l}, \cdots, \mathbf{y}_{n-\Delta_l-K_l+1,l}, s_{n,l};\Psi)$$

$$p(y_{1,n,l}|\mathbf{y}_{n-\Delta_l,l}, \cdots, \mathbf{y}_{n-\Delta_l-K_l+1,l}, s_{n,l};\Psi) = \delta(y_{1,n,l} - \sum_{k=\Delta_l}^{\Delta_l+K_l-1} \mathbf{g}_{k,l}^H \mathbf{y}_{n-k,l} - s_{n,l})$$

WPE Algorithm

- MLE Objective Function

$$
\begin{aligned}
\mathcal{L}(\Phi,\Psi) &= -\log p(y;\Phi,\Psi) \\
&= -\log \int_s p(y,s;\Phi,\Psi) \\
&= -\log \int_s \prod_{l=0}^{L-1}\prod_{n=0}^{N-1}\Big(\delta\big(y_{1,n,l} - \sum_{k=\Delta_l}^{\Delta_l+K_l-1} \mathbf{g}_{k,l}^H \mathbf{y}_{n-k,l} - s_{n,l}\big) \\
&\qquad\qquad f_{\mathcal{N}_{\mathbb{C}}}(s_{n,l};0,\lambda_{n,l})\Big) \\
&= \sum_{l=0}^{L-1}\sum_{n=0}^{N-1}\left(\log\lambda_{n,l} + \frac{\big|y_{1,n,l} - \sum_{k=\Delta_l}^{\Delta_l+K_l-1}\mathbf{g}_{k,l}^H\mathbf{y}_{n-k,l}\big|^2}{\lambda_{n,l}}\right)
\end{aligned}
$$

WPE Algorithm

- Parameter Training

$$(1)\ \hat{\Phi} \leftarrow \arg\min_{\Phi} \mathcal{L}(\Phi, \Psi)$$

$$(2)\ \hat{\Psi} \leftarrow \arg\min_{\Psi} \mathcal{L}(\hat{\Phi}, \Psi)$$

WPE Algorithm

- Updating Formula

$$\hat{s}_{n,l} = y_{1,n,l} - \hat{\bar{\mathbf{g}}}_l^H \bar{\mathbf{y}}_{n-\Delta_l,l}$$
$$\hat{\lambda}_{n,l} = |\hat{s}_{n,l}|^2$$
$$\hat{\bar{\mathbf{g}}}_l = \left(\frac{\sum_{n=0}^{N-1} \bar{\mathbf{y}}_{n-\Delta_l,l} \bar{\mathbf{y}}_{n-\Delta_l,l}^H}{\hat{\lambda}_{n,l}} \right)^{-1} \left(\frac{\sum_{n=0}^{N-1} \bar{\mathbf{y}}_{n-\Delta_l,l} y_{n,l}^*}{\hat{\lambda}_{n,l}} \right)$$

where,

$$\bar{\mathbf{g}}_l = [\mathbf{g}_{\Delta_l,l}^T, \mathbf{g}_{\Delta_l+1,l}^T, \cdots, \mathbf{g}_{\Delta_l+K_l-1,l}^T]$$
$$\bar{\mathbf{y}}_{n,l} = [\mathbf{y}_{n,l}^T, \mathbf{y}_{n-1,l}^T, \cdots, \mathbf{y}_{n-K_l+1,l}^T]$$

Setup

- testset utterances as clean speech
- add reverberation using RIR_DATABASE
- dereverberated with WPE

Results

- WER for ASR test

	Clean	Reverb	DeReverb
20140801_0508_class1	1.35	3.82	2.99
20140801_0508_class2	7.44	15.83	13.29
20140902_03_class1	1.41	4.60	3.43
20140902_03_class2	3.13	9.17	6.77