Frame-online DNN-WPE dereverberation

Jahn Heymann\(^1\), Lukas Drude\(^1\), Reinhold Haeb-Umbach\(^1\), Keisuke Kinoshita\(^2\), Tomohiro Nakatani\(^2\)

\(^1\)Paderborn University, Department of Communications Engineering, Paderborn, Germany
\(^2\)NTT Communication Science Laboratories, NTT Corporation, Kyoto, Japan

1\{heymann, drude, haeb\}@upb.de / 2\{kinoshita.k, nakatani.tomohiro\}@lab.ntt.co.jp

**Introduction**

- **Weighted Prediction Error (WPE)** is an effective dereverberation method for far-field speech recognition as shown by the REVERB challenge or its commercial application (Google Home).
- Iterative solution unsuited for low-latency application
- Previous approaches use recursive update formulation with smoothing PSD estimation or a DNN PSD estimator with block-wise updates
- **This work:**
  - Extension of DNN-WPE to frame-online updates
  - Thorough evaluation of performance of DNN-WPE in comparison with conventional WPE variants

**WPE**

**Model:**

\[
y_{t,f,d} = x_{t,f,d}^{(early)} + x_{t,f,d}^{(tail)} \\
x_{t,f,d}^{(early)} = y_{t,f,d} - W_{t,f} y_{t-f-\Delta,f} \\
x_{t,f,d}^{(tail)} = y_{t,f,d} - W_{t,f} y_{t-f-\Delta,f} \\
\]

- **Batch solution:**

\[
R_{t,f} = \sum_{r=0}^{t} \frac{\bar{y}_{t-\Delta,r} \bar{y}_{t-\Delta,r}^H}{\lambda_{t,f}} \\
\]

Step 1: \(\bar{y}_{t-\Delta,r} = \frac{1}{\lambda_{t,f}} \sum_{d} \tilde{y}_{t,f,d} \)

Step 2: \(\lambda_{t,f} = \frac{1}{2(\delta + 1)} \sum_{r=t-\Delta}^{t} \sum_{d} |\tilde{y}_{t,f,d}|^2 \)

**Implementation available:** [https://github.com/fgnt/nara_wpe](https://github.com/fgnt/nara_wpe)

**System overview**

- **PSD Estimator**

  - (iterative solution only)

  - Acoustic model

  - Estimate & apply filter coefficients

  - \(\lambda \rightarrow \) Predict

  - Smoothing:

\[
\lambda_{t,f} = \frac{1}{(\delta_L + \delta_R + 1) D} \sum_{r=t-h}^{t+h} \sum_{d} |y_{t,f,d}|^2
\]

(b) DNN:

\[
1 \times 512 \times \text{LSTM} + 2 \times 2048 \times \text{Dense + Output}
\]

- Operates on single channel, final estimate averaged
- Trained to estimate PSD of target image

**Conclusion**

DNN PSD estimator improves performance for frame-online WPE dereverberation over smoothing PSD estimator by 5% - 10% in highly reverberant and noisy reverberant conditions

**Outlook**

- Joint training of DNN PSD estimator and acoustic model
- Compare with different model based PSD estimators

---

WSJ + VoiceHome

- WSJ convolved with VoiceHome RIRs (T60: 395 – 585 ms)
- Very dynamic households background noise
- Blocks of 2 s and forgetting factor of 0.7 for Block-Online

- **ASR Results**

  - **REVERB**
    - T60 ranges from 0.25 – 0.7 s with 20 dB noise level
    - Blocks of 2 s and forgetting factor of 0.7 for Block-Online
    - Results averaged over real near / far conditions

<table>
<thead>
<tr>
<th></th>
<th>Offline</th>
<th>Block-Online</th>
<th>Online</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unprocessed</td>
<td>17.6</td>
<td>14.4</td>
<td>10.9</td>
</tr>
<tr>
<td>Iteration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) Smoothing</td>
<td>14.3</td>
<td>10.8</td>
<td>14.5</td>
</tr>
<tr>
<td>(b) DNN</td>
<td>14.3</td>
<td>10.8</td>
<td>14.5</td>
</tr>
</tbody>
</table>

**Conclusion**

DNN PSD estimator improves performance for frame-online WPE dereverberation over smoothing PSD estimator by 5% - 10% in highly reverberant and noisy reverberant conditions

**Outlook**

- Joint training of DNN PSD estimator and acoustic model
- Compare with different model based PSD estimators