Speech Dereverberation by Constrained and Regularized Multi-Channel Spectral Decomposition: Evaluated on REVERB Challenge

Meng Yu, Audience Inc. Frank K. Soong, Microsoft Research Asia

Overview

- Multi-channel speech dereverberation
- Cross-channel cancellation
- Spectral decomposition
- Convolution in spectral magnitude domain
- Generalized Kullback-Leibler divergence
- Constrained and regularized non-negative deconvolution
- Solved by a multiplicative algorithm
- Evaluations on “speech enhancement task”

Algorithm

Goal: Estimate the spectral magnitude of clean speech $S$

- Reverberated speech spectral magnitude $X$
- Actual observation $Z$ by assuming noise $Z[n, k] \approx X[n, k] = S[n, k] \ast H[n, k]$.

Single-change solution:

By imposing the non-negativity and sparsity constraints (H. Kameoka et al, 2009 and K. Kumar, et al, 2011)

$$\{S, H\} = \arg \min_{S, H} J$$

where $J = \sum_i (Z[i, k] - \sum_m S[m, k] H[i - m, k])^2 + \lambda \sum_i S[i, k]$  

$s.t. S[n,k] \geq 0, H[n,k] \geq 0, \sum_n H[n,k] = 1$ (1)

Further motivation:

- Advantage of KL divergence for estimating small values in between spectral peaks compared to $l_2$ norm.

$$D(x|y) = x \log \frac{x}{y} - x + y.$$

Algorithm cont.

- Cross-channel cancellation enforces the filters $H^*_i, i = 1, 2, \ldots, N$, to resolve the spatial difference between channels. The cross-channel cancellation error is to be minimized.

Multi-channel solution:

$$J = \sum_j \sum_i (Z_j[i, k] \log Z_j[i, k] - Z_j[i, k] + \sum_m S[m, k] H_j[i - m, k]) + \beta \sum_i (\sum_m X_1[m, k] H_2[i - m, k] - \sum_m X_2[m, k] H_1[i - m, k])^2$$

$$+ \lambda \sum_i S[i, k]$$

$s.t. S[n,k] \geq 0, H_j[n,k] \geq 0, \sum_n H_j[n,k] = 1, j = 1, 2$ (2)

$X_j$ is substituted by $Z_j$ in the above model as the solution $S \ast H_1$ is expected to converge to $Z_j$. Easy to be extended to N-channel ($N > 2$).

Diagram:

**Experimental Evaluation**

Channel: Select channel 1 and 2 for dual-channel processing, while 1, 2, 3, and 4 for 4-channel processing.

Post processing: A noise suppression post processing (optimally-modified log-spectral amplitude speech estimator (I. Cohen 2003) is applied to the dereverberated signal to suppress the background noise as an option.

**Metrics Analysis**

<table>
<thead>
<tr>
<th>CD</th>
<th>LLR</th>
<th>FWSNR</th>
<th>SRMR</th>
<th>PESQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD</td>
<td>-</td>
<td>0.15</td>
<td>0.66</td>
<td>0.58</td>
</tr>
<tr>
<td>LLR</td>
<td>-</td>
<td>-</td>
<td>-0.45</td>
<td>-0.49</td>
</tr>
<tr>
<td>FWSNR</td>
<td>0.66</td>
<td>-0.45</td>
<td>-0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>SRMR</td>
<td>0.58</td>
<td>-0.49</td>
<td>0.99</td>
<td>-0.82</td>
</tr>
<tr>
<td>PESQ</td>
<td>0.69</td>
<td>-0.33</td>
<td>0.87</td>
<td>0.82</td>
</tr>
</tbody>
</table>

- Consistent improvement of the proposed algorithm is proven by the metrics, such as SRMR, FWSNR and PESQ.
- The performance cross-correlation between those metrics is shown in the table, indicating SRMR, FWSNR and PESQ perform consistently.