THE TUM SYSTEM FOR THE REVERB CHALLENGE: RECOGNITION OF REVERBERATED SPEECH USING MULTI-CHANNEL CORRELATION SHAPING ENHANCEMENT AND BLSTM RECURRENT NEURAL NETWORKS

Jürgen T. Geiger, Erik Marchi, Björn Schuller and Gerhard Rigoll
Institute for Human-Machine Communication, Technische Universität München, Germany
1 also with the Department of Computing, Imperial College London, UK
geiger@tum.de

Introduction
• Multi-channel dereverberation with Correlation Shaping (CS)
• Bidirectional Long Short-Term Memory (LSTM) RNNs for phoneme prediction
• GMM-LSTM double-stream decoding

Proposed System
• Correlation Shaping
  • Modifies the correlation structure of the input
  • Reduces long-term correlation in the LP residual
• Filter update equation
  \[ p(n,t+1) = p(n,t) - \rho \nabla_{w}(l) \]
• Gradient
  \[ \nabla_{w}(l) = \frac{\sum_{l} \nabla_{w}(l)}{\sqrt{\sum_{l} \nabla_{w}(l)}} \]

Experiments and Results
• Results on test set
• 8 different recording conditions
• 1- and 8-channel audio processing
• CS effectively reduces long-term reverberation energy for higher reverberation times

Conclusions
• Multi-channel: CS reduction > 25% WER, LSTM reduction ~15% WER
• Single-channel: LSTM reduction ~7% WER (Real Data), ~20% WER (Sim Data)

<table>
<thead>
<tr>
<th>System</th>
<th>SIM DATA</th>
<th>REAL DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Room 1</td>
<td>Room 2</td>
</tr>
<tr>
<td>Baseline</td>
<td>16.23</td>
<td>18.71</td>
</tr>
<tr>
<td>CS + Baseline</td>
<td>16.03</td>
<td>17.66</td>
</tr>
<tr>
<td>Kaldi</td>
<td>10.23</td>
<td>12.26</td>
</tr>
<tr>
<td>CS + Kaldi</td>
<td>10.86</td>
<td>11.50</td>
</tr>
<tr>
<td>CS + LSTM</td>
<td>8.32</td>
<td>9.98</td>
</tr>
</tbody>
</table>

Kaldi GMM + CS + LSTM, +LSTM

8ch test data

1ch test data

Dereverberation (CS)

GMM

1ch MCT training data

LSTM

8ch test data

double-stream HMM

recognition result

Kaldi GMM (Similar to baseline system)
• Basis feature space MLLR for adaptation
• Trigram language model

LSTM phoneme recognition
• 26 Mel filterbank features
• 3 layer, 200 units

LSTM training software: http://currentt.sf.net