Introduction
Propose an adaptive speech enhancement technique.
- Detects SNR of input speech.
- Removes stationary background noise.
- Performs pre-emphasis.
- De-noise using Modified Phase Opponency (MPO) model with Aperiodic Periodic and Pitch detector (APP) [1,2].

Applications
- Noise-robust Speech Recognition
- Speech Enhancement
- Noise robust automated transcription

Corpus
- ‘Speech in Speech shaped noise’ (SSN) corpus of Speech Separation Challenge 2006 [3].
- Noise at 5 diff SNR: Clean, 6dB, 0dB, -6dB & -12dB
- Task: (a) detect key words, (b) detect all words

MPO-APP
MPO acts as a switch,
- When speech is dominant, it passes the signal as-is.
- When noise is dominant, it attenuates the signal.
  - Problematic situations:
    (a) wide-band speech → speech deletions
    (b) narrow band noise → noise insertions
      (i.e 2 formants close together)
  - Solution: Use APP in cascade [2]
    - APP detects periodic regions
    - wide-band speech → Periodic regions
    - narrow-band noise → Aperiodic regions
  - APP thresholds vary with SNR
    - Estimate SNR before MPO-APP enhancement
  - MPO-APP fails to enhance high frequency formants
    if they are relatively weak
  - Perform pre-emphasis before MPO-APP

Solution
Block diagram of the adaptive speech enhancement scheme

Recognition Accuracy for MPO and MPO-APP

<table>
<thead>
<tr>
<th></th>
<th>no proc</th>
<th>6dB</th>
<th>0dB</th>
<th>-6dB</th>
<th>-12dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLD</td>
<td>98.56</td>
<td>56.67</td>
<td>18.94</td>
<td>11.78</td>
<td>11.67</td>
</tr>
<tr>
<td>MPO</td>
<td>96.00</td>
<td>73.83</td>
<td>50.06</td>
<td>26.00</td>
<td>14.33</td>
</tr>
<tr>
<td>APP</td>
<td>98.65</td>
<td>87.53</td>
<td>75.00</td>
<td>62.65</td>
<td>56.33</td>
</tr>
<tr>
<td>MPO-APP</td>
<td>95.94</td>
<td>75.06</td>
<td>53.17</td>
<td>31.22</td>
<td>19.06</td>
</tr>
</tbody>
</table>

CLD: Color-letter-digit keyword recognition as specified in SSC 2006
WRD: Word recognition rate of the corpus

- MPO was implemented for SSC06 by Deshmukh et. al [1].

Recognition accuracy for the proposed adaptive enhancement

<table>
<thead>
<tr>
<th>Adaptive enhancement</th>
<th>dB</th>
<th>6dB</th>
<th>0dB</th>
<th>-6dB</th>
<th>-12dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLD</td>
<td></td>
<td>95.33</td>
<td>81.00</td>
<td>58.61</td>
<td>32.17</td>
</tr>
<tr>
<td>WRD</td>
<td></td>
<td>98.35</td>
<td>91.06</td>
<td>79.11</td>
<td>66.02</td>
</tr>
</tbody>
</table>

Recognition accuracy for speech corrupted by speech shaped noise

Conclusion
- Proposed scheme increases CLD accuracy by 17% (mean) and WRD accuracy by 3.7% over MPO [1].
- Average increase by 98.04% over baseline.
- Reduces unwanted-computation.

References