Noise Management Features in Open Speech Platform

Mingchao Liang, Kuan-Lin Chen, Wenyu Zhang, Ching-Hua Lee, Bhaskar D. Rao, and Harinath Garudadri
Department of Electrical and Computer Engineering, University of California, San Diego

1 The Open Speech Platform (OSP)
We describe the use of an extensible, real-time, open-source speech processing platform (OSP) for noise management research for hearing aids (HAs) [3]. Figure 2 shows the OSP with the battery and the carrier board enclosed in a physical case.

Figure 2: OSP Block Diagram, which includes three basic HA features i) Subband Decomposition, ii) Wide Dynamic Range Compression (WDRC), iii) Feedback Cancellation, and two advanced features iv) Beamforming, v) Speech Enhancement.

Figure 3: The baseline beamforming (BF) is a generalized sidelobe canceller (GSC), based on linearly constrained minimum variance (LCMV) beamformer. It utilizes left and right inputs \(e_{\text{left}}(n)\) and \(e_{\text{right}}(n)\). The adaptive filter uses the sparsity-promoting least mean square (SLMS) algorithm [4] to continuously estimate interference components \(f(n)\). To mitigate the effect of direction of arrival mismatch and microphone array mismatch, adaptive mode control and norm constraint are introduced to enhance the system robustness.

2 The Real-Time Master Hearing Aid (RT-MHA)

Figure 4: Left is the arrangement of the two-microphone array, target talker, and interference talker. A two-microphone array with distance 0.15m is used in a 4 x 5 x 3.69m room. The target talker and interference talker are positioned 1m away from the microphone array. All impulse responses from the sources to the microphone array are generated with Lehmnah’s image source method [5]. Right is the SIR performance under different input SIR scenarios. The errorbar shows the standard deviation of the 20 TIMT files. The baseline BF consistently improves SIR by around 16dB under different levels of interference.

3 Experiments
Both file-based simulations and live measurements were conducted to evaluate BF performance. Various objective metrics were used for objective evaluation and informal subjective evaluations.

3.1 File-based Simulation

Figure 5: Mean and standard deviation of NB-PESQ (upper left), WB-PESQ (upper right), STOI (lower left), HASQI (lower right) performance of the 20 TIMT files under different input conditions.

Table 1: MOS rating scale

<table>
<thead>
<tr>
<th>Rating</th>
<th>Speech quality</th>
<th>Level of distortion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>Imperceptible</td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>Slightly perceptible, but not annoying</td>
<td></td>
</tr>
<tr>
<td>Fair</td>
<td>Perceptible and slightly annoying</td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>Annoying, but not objectionable</td>
<td></td>
</tr>
<tr>
<td>Bad</td>
<td>Very annoying and objectionable</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 shows the category ratings from subjective ratings used to map distance metrics to perceived quality [1]. Both PESQ and WB-PESQ were essentially “tuned” for distortions arising from voicing in cellular telephony. Based on objective investigations presented here, both PESQ and WB-PESQ appear to underestimate perceived quality. Our experiments suggest that PESQ values below 2.0 and less than 0.2 improvement are suspect for using in algorithm development. STOI on the other hand appears to overestimate perceived quality of BF due to remnant distortions, while HASQI appears to saturate at a much lower value than subjective quality would indicate in developing BF algorithms.

5 Conclusion

In this contribution, we described baseline noise management subsystems of OSP [1]. Objective evaluation of the quality with various objective metrics was carried out to aid in the development of advanced noise management algorithms. The baseline Left/Right BF based on GSC appears to be promising from subjective evaluations, but the objective metrics considered in this work do not appear to correlate well with the perceived quality. Nevertheless, objective metrics are very useful during algorithm development due to their advantages in setting up repeatable and repeatable scripts. We recommend caution in leveraging objective metrics developed for one type of distortion for other distortions, unless validated with subjective evaluations.

6 Acknowledgements

This work is supported by NIH/NIDCD grants R21DC015046, R33DC015046 and R01DC015436, NSF IIS grant IIS-1838380; and additional financial and in-kind support from Qualcomm Institute of Calit2, UCSD.

References