Noise Managment 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

Contact Information:
Harinath Garudadri
Qualcomm Institute of Calit2
University of California, San Diego
9500 Gilman Drive #0436, La Jolla, CA 92093

Email: `hgarudadri@ucsd.edu`

UC San Diego
JACOBS SCHOOL OF ENGINEERING
Electrical and Computer Engineering

QUALCOMM INSTITUTE

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) [1]. Figure 1 shows the OSP, with the battery and the carrier board enclosed in a physical case.

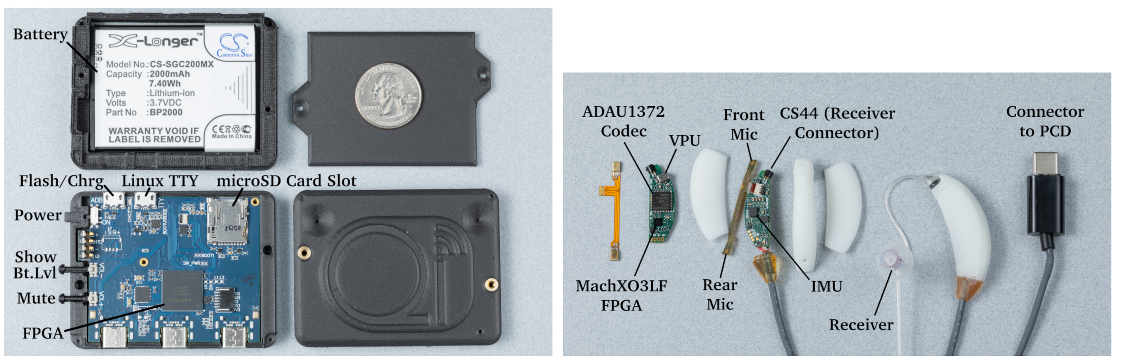

Figure 1: Left OSP processing and communication device (PCD) disassembled, showing the battery, the back of the carrier board, and the plastic shell. **Right** OSP behind the ear, receiver in the canal (BTE-RIC) ear level assemblies, together and disassembled.

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

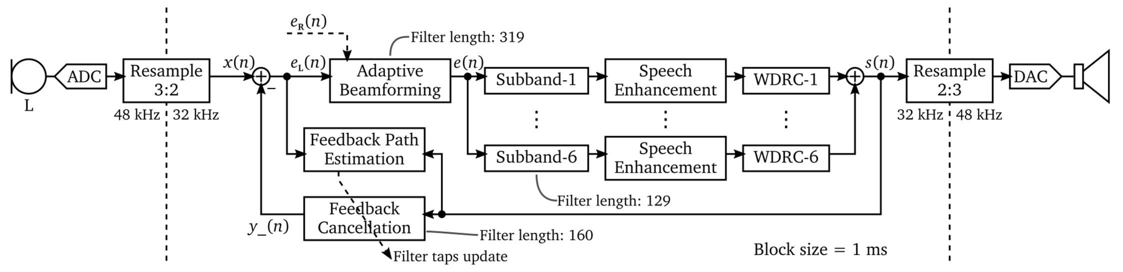

Figure 2: RT-MHA Block Diagram, which includes three basic HA features i) Subband Decomposition, ii) Wide Dynamic Range Compression (WDRC), iii) Feedback Cancellation, and two advanced features iv) Beamforming, v) Speech Enhancement.

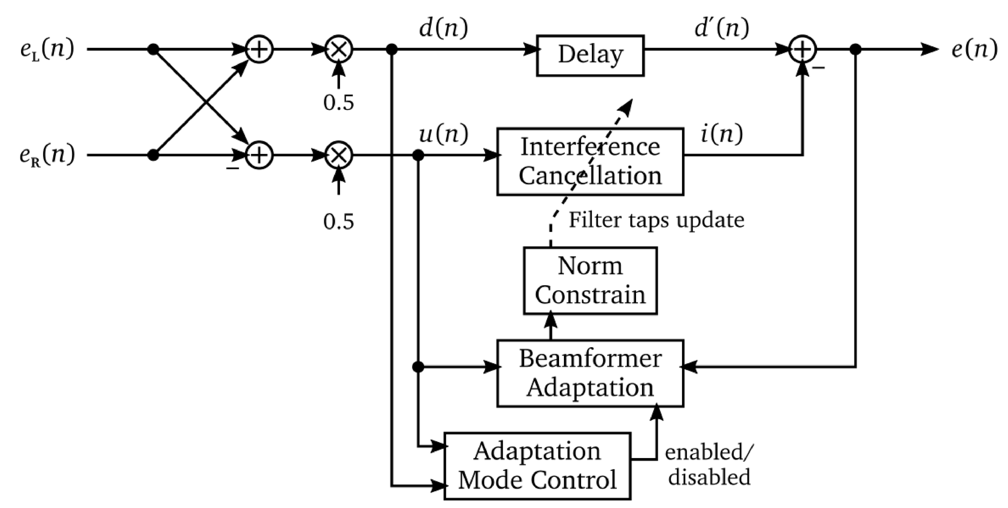

Figure 3: The baseline beamforming (BF) is a generalized sidelobe canceller (GSC), based on linearly constrained minimum variance (LCMV) beamformer. It utilizes left and right inputs $e_L(n)$ and $e_R(n)$. The adaptive filter uses the sparsity-promoting least mean square (SLMS) algorithm [2] to continuously estimate interference components $i(n)$. 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.

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 were used in preliminary investigations related to the reliability of the objective metrics. The objective metrics included i) Single to Interference Ratio (SIR), ii) Narrow-Band Perceptual Evaluation of Speech Quality (NB-PESQ) [3], iii) Wide-Band Perceptual Evaluation of Speech Quality (WB-PESQ) [4], iv) Short-Time Objective Intelligibility (STOI) [5] and v) Hearing-Aid Speech Quality Index (HASQI) [6].

3.1 File-based Simulation

Twenty audio files with sampling rate 16kHz are randomly chosen from the test set of TIMIT for the target talker, while babble noise is used for the interference talker. Six different input SIR scenarios (-5, 0, 5, 10, 15, 20dB) are used to simulate different levels of noisy environment. Figure 4 Left shows the array arrangement and Figure 4 Right shows the SIR performance.

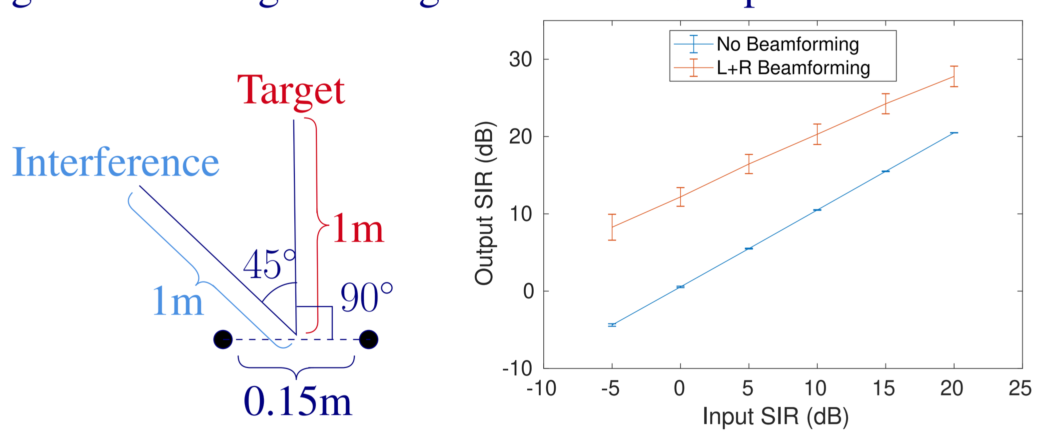

Figure 4: Left is the arrangement of the two-microphone array, target talker, and interference talker. A two-microphone array with distance 0.15m is used in a $4 \times 5 \times 3.6\text{m}^3$ room. The target talker and interference talker are positioned 1m away from the microphone array. All impulse responses from the sources to the microphone array are generated with Lehmann's image source method [7]. **Right** is the SIR performance under 6 different input SIR scenarios. The errorbar shows the standard deviation of the 20 TIMIT files. The baseline BF consistently improves SIR by around 10dB under different levels of interference.

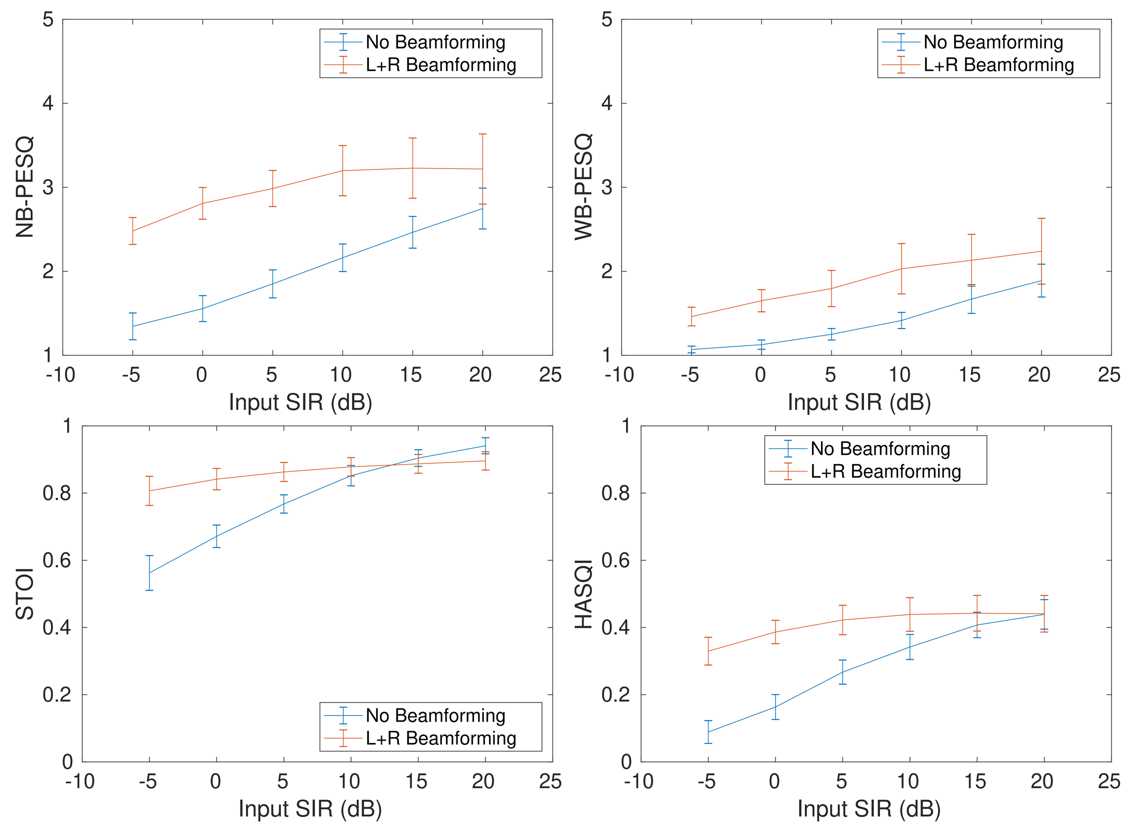

Figure 5: Mean and standard deviation of NB-PESQ (upper left), WB-PESQ (upper right), STOI (lower left), HASQI (lower right) performance under various input conditions. The errorbar shows the standarad deviation of the 20 TIMIT files.

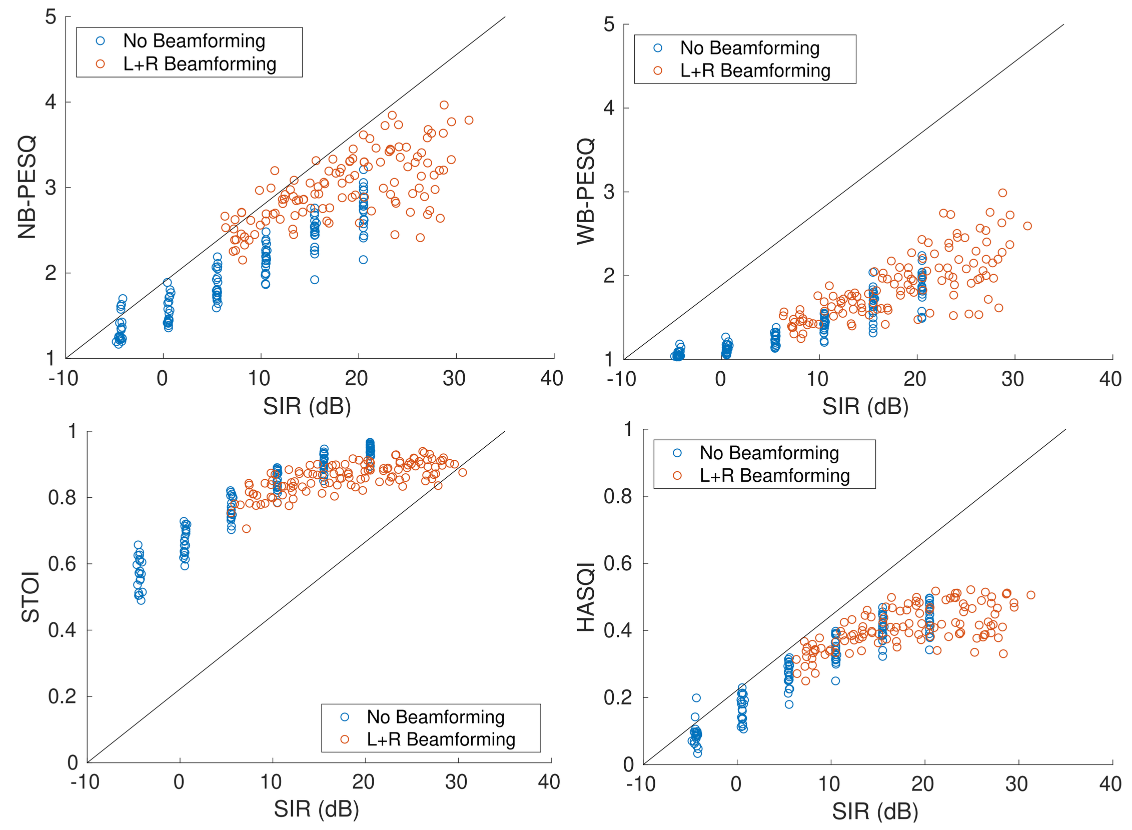

Figure 6: Scatter plots of NB-PESQ (upper left), WB-PESQ (upper right), STOI (lower left), HASQI (lower right) performance of the 20 TIMIT files under different conditions.

3.2 Live Measurement

Our live measurement setup is similar with the one shown in Figure 4. Figure 7 shows the SIR performance.

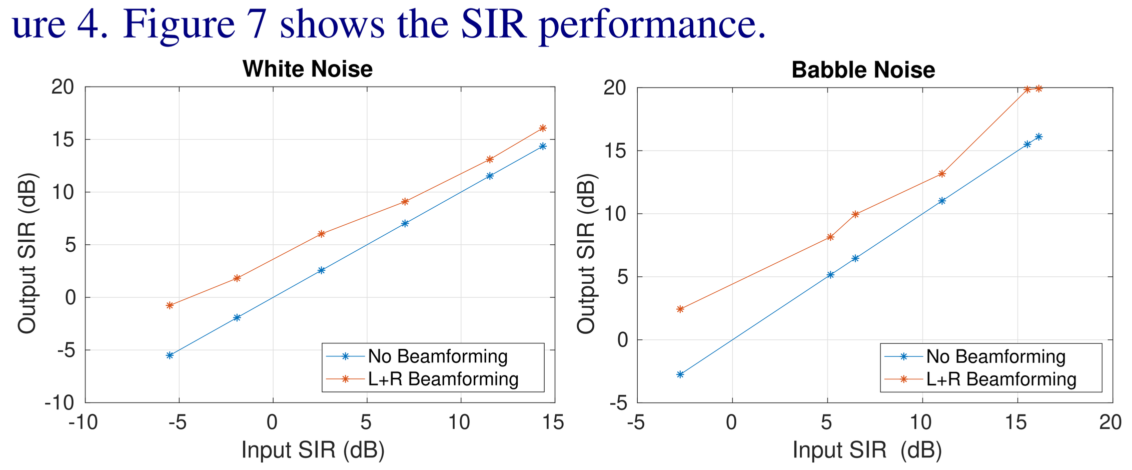

Figure 7: SIR performance under 6 different input SIR scenarios. The baseline BF consistently improves SIR in real environments.

4 Discussion

Based on informal listening, we hypothesize that SIR correlates well with subjective assessments. Figures 5 and 6 suggest caution in relying on the objective metrics investigated here to guide BF algorithm development.

Table 1: MOS rating scale

Rating	Speech quality	Level of distortion
5	Excellent	Imperceptible
4	Good	Just perceptible, but not annoying
3	Fair	Perceptible and slightly annoying
2	Poor	Annoying, but not objectionable
1	Bad	Very annoying and objectionable

Table 1 shows the category ratings from subjective ratings used to map distance metrics to perceived quality [3, 4]. Both NB-PESQ and WB-PESQ were essentially "tuned" for distortions arising from vocoding in cellular telephony. Based on objective investigations presented here, both NB-PESQ and WB-PESQ *appear to underestimate* perceived quality. Our experience suggests that NB-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 libraries. 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-1838830; and additional financial and in-kind support from Qualcomm Institute of Calit2, UCSD.

References

[1] L. Pisha, J. Warchall, T. Zubatiy, S. Hamilton, C. Lee, G. Chockalingam, P. P. Mercier, R. Gupta, B. D. Rao, and H. Garudadri, "A wearable, extensible, open-source platform for hearing healthcare research," *IEEE Access*, vol. 7, pp. 162 083–162 101, 2019.

[2] C.-H. Lee, B. D. Rao, and H. Garudadri, "Sparsity promoting LMS for adaptive feedback cancellation,," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2017, pp. 226–230.

[3] Int. Telecommun. Union, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," *ITU-T Recommendation P.862*, 2001.

[4] ——, "Wideband extension to recommendation P.862 for the assessment of wideband telephone networks and speech codecs," *ITU-T Recommendation P.862.2*, 2005.

[5] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2010, pp. 4214–4217.

[6] J. M. Kates and K. H. Arehart, "The hearing-aid speech quality index (HASQI) version 2," *J. Audio Eng. Soc.*, vol. 62, no. 3, pp. 99–117, 2014.

[7] E. A. Lehmann and A. M. Johansson, "Diffuse reverberation model for efficient image-source simulation of room impulse responses," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 6, pp. 1429–1439, 2009.