Two Microphones DOA+IVA App

- The performance of BSS technique depends on the accuracy of estimated demixing matrix. The demixing matrix in the frequency domain is the impulse response between the sources and the microphones in the time domain .
- In this contribution, we **developed a criterion to update the demixing matrix**, which greatly reduces the average computational complexity required to separate the sources.
- The proposed criterion is based on tracking significant changes in the impulse response between the sources and the microphones.
- The criterion depends on a proposed computationally fast two-microphone speech source Direction of Arrival (DOA) estimation using smartphone.
- **A Feed Forward Neural Network** (FNN) based DOA estimation (FNNDOA) of speech source in noisy and reverberant environment is developed.
- **FNN- DOA** does the DOA classification and is used to track the changes in DOA angle an impulse responses.
- FNN architecture contains, One hidden layer with 8 nodes. Rectified Linear Unit (ReLu) /Max function is used as activation function at nodes. Weight vectors are obtained using Gradient-based optimization of objective function. NN output is 7 classes of DOA angles between 0-180º. Softmax function sets probability of each class. A class with highest probability is chose.
- The input features for the FNNDOA are the **13 samples of cross-correlation of two input signals**.
- To make the IVA computationally efficient, the Demixing matrix is calculated for first 100 frames and is updated only when the criterion is satisfied, i.e. when speech source moves by 30º DOA angle.
- The proposed method is implemented on an android smartphone as an application.
- Comparative Objective and subjective evaluations were carried out.

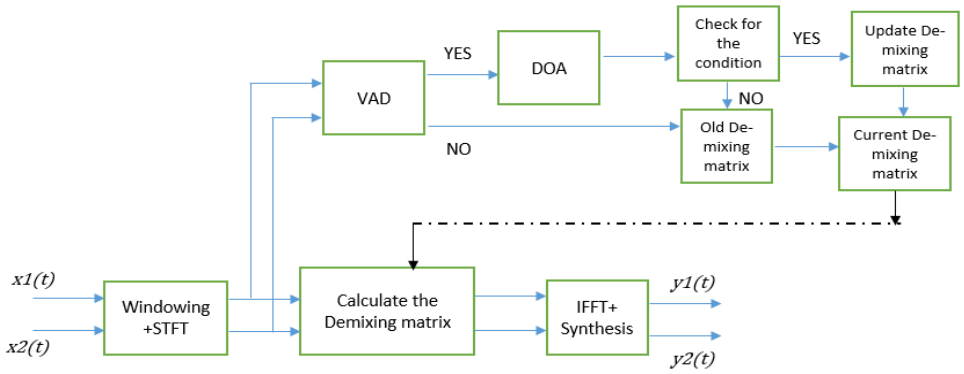


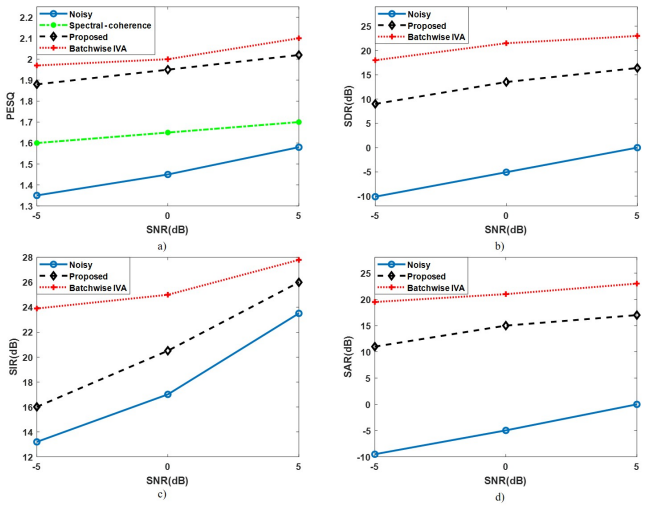
Figure 1 Block Diagram of the proposed method and the update criteria

Frame Index	(n-10)	(n-9)	(n-8)	(n-7)	(n-6)	(n-5)	(n-4)	(n-3)	(n-2)	(n-1)	(n)
DOA	θ_1	θ_3	θ_1	θ_1	θ_1	θ_2	θ_2	θ_2	θ_1	θ_2	θ_i

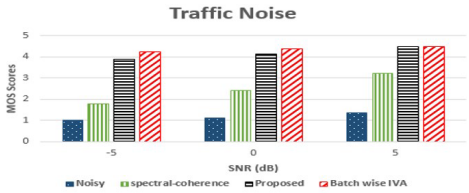
4 out of 5 frames with angle θ_1 — 4 out of 5 frames with angle θ_2 — Update demixing matrix

Test Results:

- For the objective quality and intelligibility measurement of speech, we use Perceptual Evaluation of Speech Quality (PESQ) , Signal to distortion ratio (SDR), Signal to artifact ratio (SAR) , and Signal to interference ratio (SIR) .
- For the performance evaluation of DOA, we use Accuracy measures.
- Along with Objective measures, we perform Mean Opinion Score (MOS) tests. Some of the test results are shown below.



Performance evaluation using (a) PESQ, (b) SDR, (c) SIR and (d) SAR for speech mixed with Traffic Noise at -5dB, 0dB and +5dB SNR



Subjective test results evaluated traffic noise at -5dB, 0dB and +5dB SNR