Technical Documentation ver. 2

12/4/2018 Anshuman Ganguly, Abdullah Kucuk, Issa Panahi

Architecture for Two microphone DOA for Smartphone version 2:

1. **Algorithm:**
   a. DOA estimation: Generalized Cross-Correlation (GCC)
   b. Post-filter:
      i. Simple Voice Activity Detector (VAD) (using Spectral Flux feature)
      ii. Adaptive Thresholding based VAD (using Modified Spectral Flux feature)
   c. Pre-filter: Minimum Mean Square Error-Log Spectral Amplitude estimator(MMSE-LSA) (also known as log MMSE Spectral amplitude estimator)

2. **Implementation:**
   a. MATLAB 2017b for testing and debugging
   b. Visual Studio 2016 for C and Java Native Interface (JNI) for Android

3. **Frame length**:
   a. MATLAB Implementation: 100ms@16kHz = 1600. FFT size = 2048.
   b. C/Java Implementation: 40ms @48kHz. FFT size = 2048.

4. **Tunable parameters (for VAD):**
   a. Initial Threshold time (T) for VAD: 10 frames@100ms/frame = 1sec
   b. Duration Threshold (D): 2 frames@100ms/frame = 200ms. (to differentiate between sustained speech and short-duration noise)
   c. $\alpha_{As} = 10^{-6}$ (Fast attack), α_{Ds}= 0.5 (Slow decay); α_{An}= 0.99995 (Slow attack), $\alpha_{Dn} = 10^{-6}$ (Fast decay).

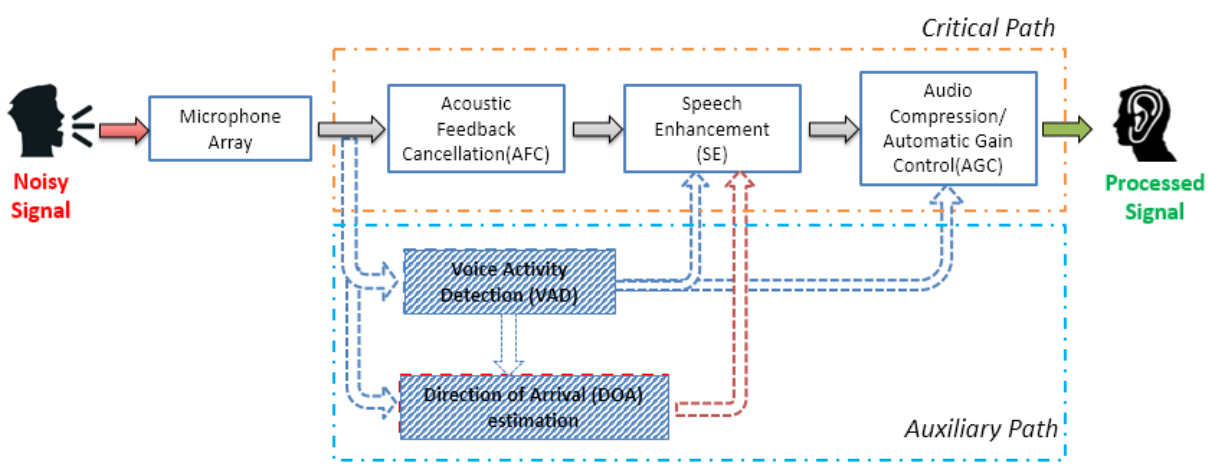


Figure1. Block Diagram of typical Speech Processing pipeline

Refer: https://www.utdallas.edu/ssprl/hearing-aid-project/research-papers/

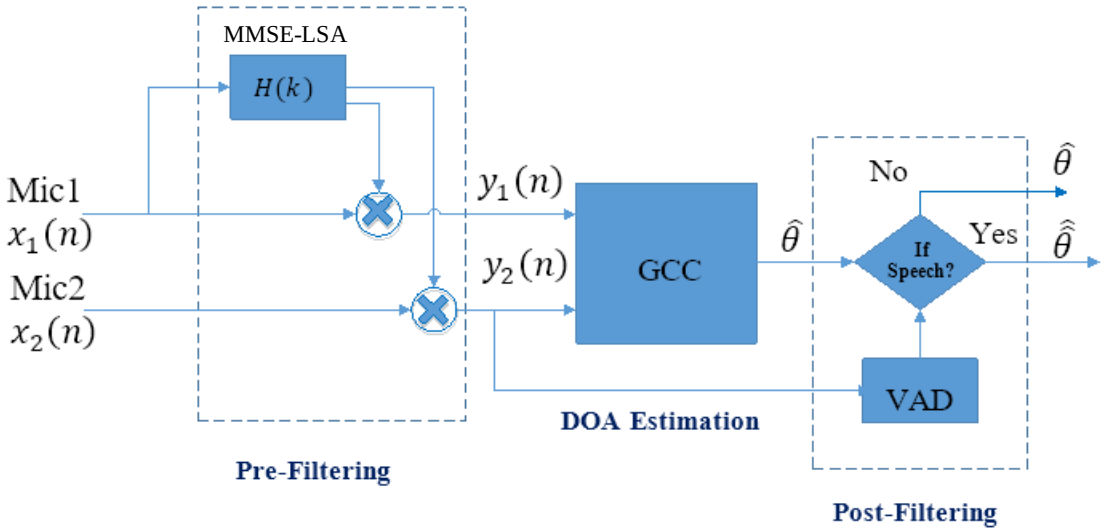


Figure 2. Block Diagram of Proposed method. $\hat{\theta}$ and $\hat{\hat{\theta}}$ are the estimated and corected DOA estimates, respectively. $H(k)$ is the pre-filter derived from MMSE-LSA Spectral amplitude estimator.

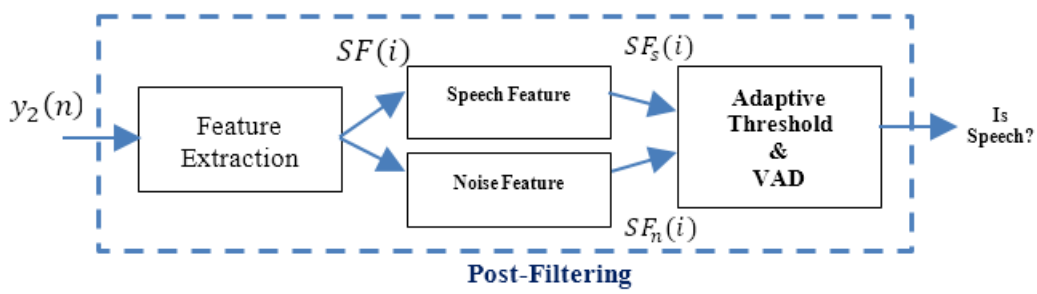


Figure 3. Block diagram of Adaptive thresholding(AT) based VAD.

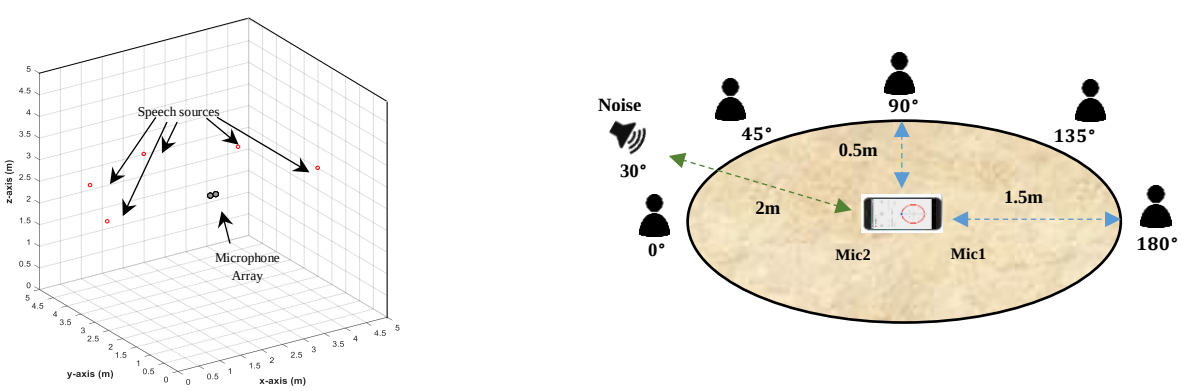


Figure 4. Experimental Setup

Refer: https://www.utdallas.edu/ssprl/hearing-aid-project/research-papers/

Experimental Evaluation

Experiment1 :

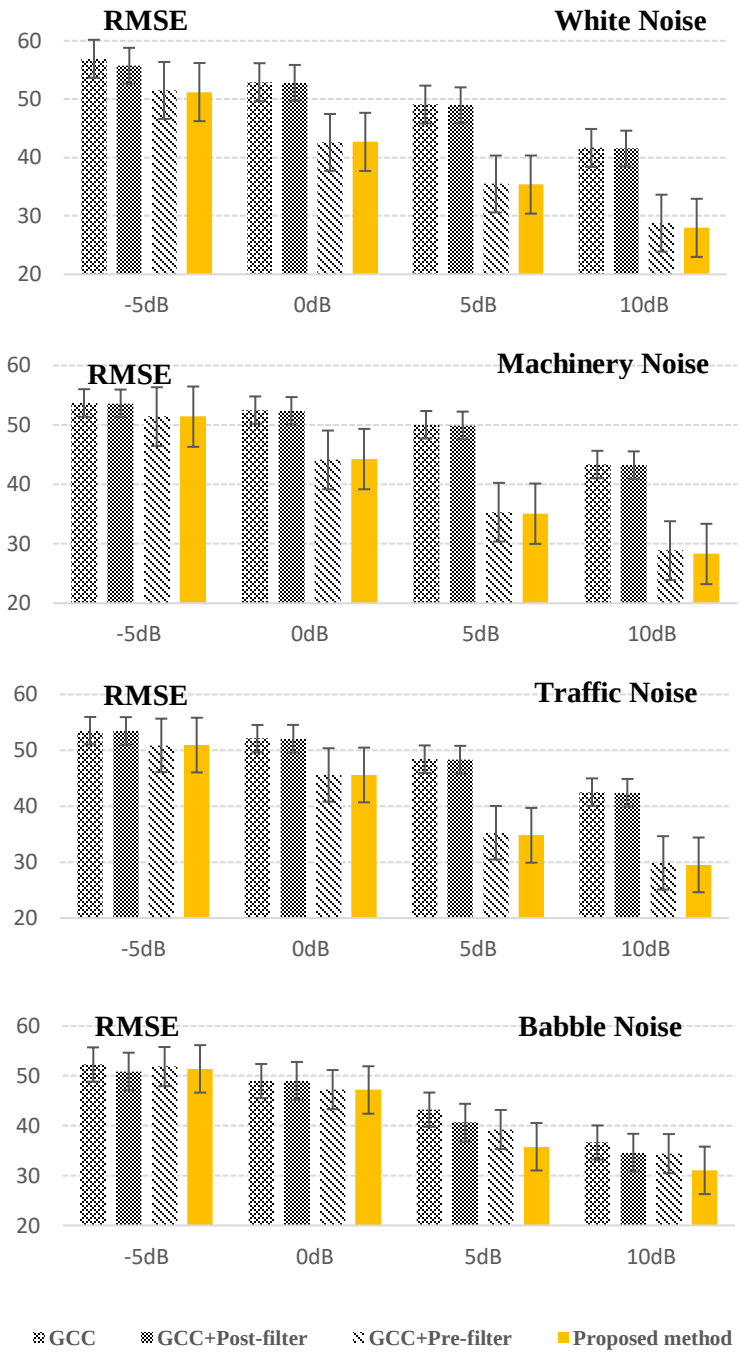


Figure 5. Average RMSE (°) for Simulated data corrupted by (a) White Noise; (b) Machinery Noise; (c) Traffic Noise; and (d) Babble Noise.

Experiment 2:

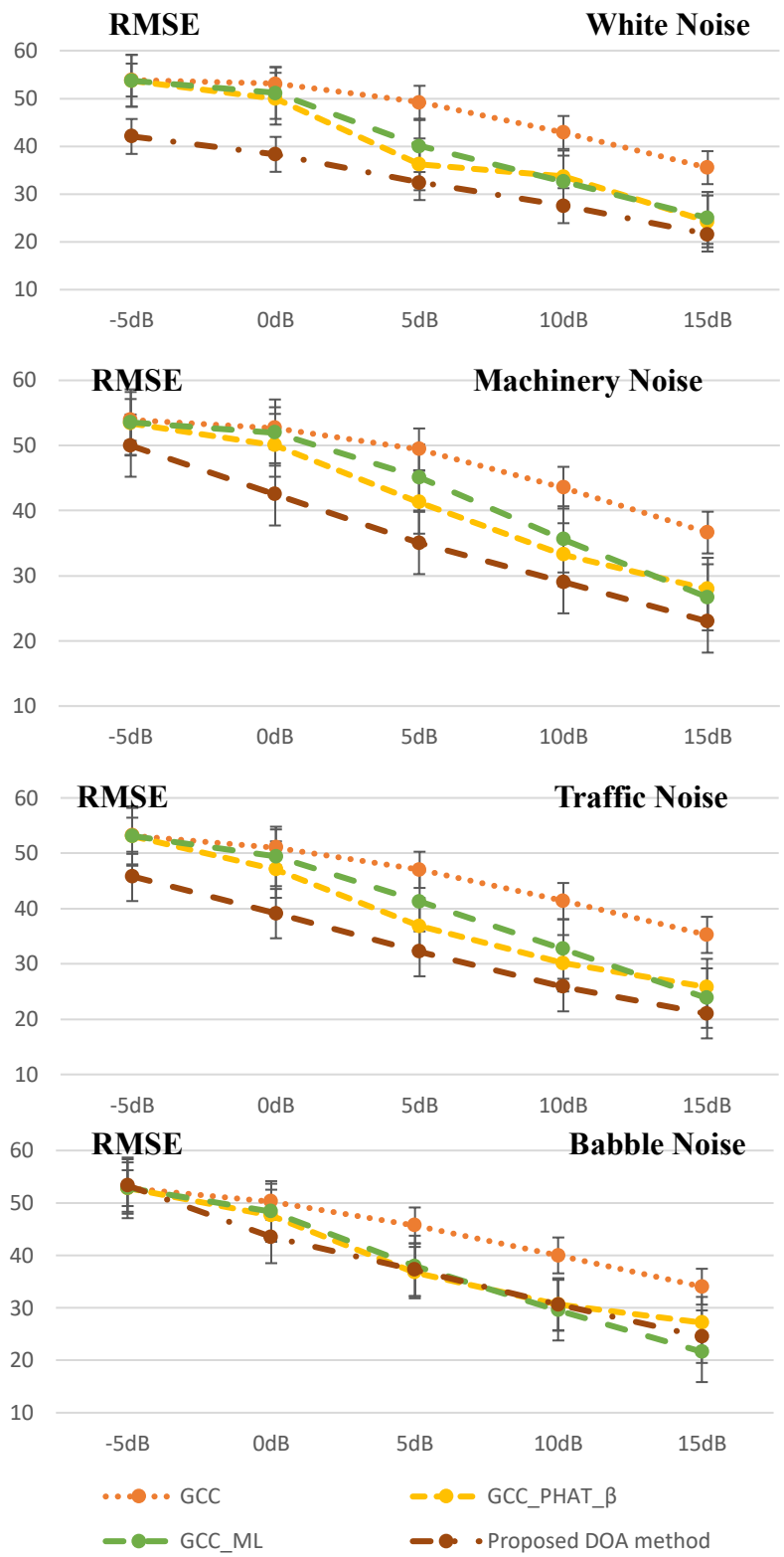


Figure 6. Average RMSE (°) comparison for data corrupted by (a) White Noise; (b) Machinery Noise; (c) Traffic Noise; and (d) Babble Noise for different SNR.

Refer: https://www.utdallas.edu/ssprl/hearing-aid-project/research-papers/

Experiment 3:

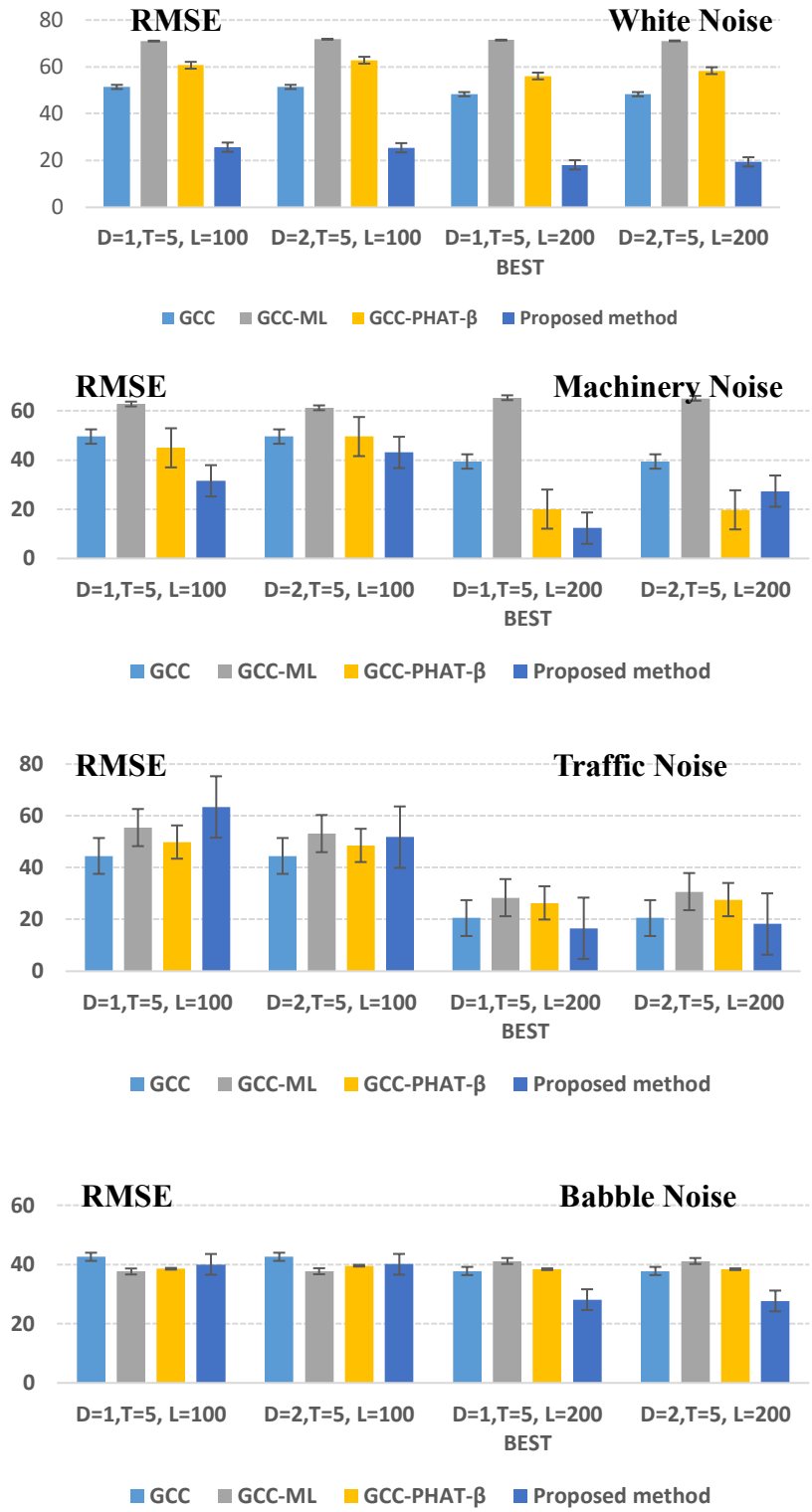
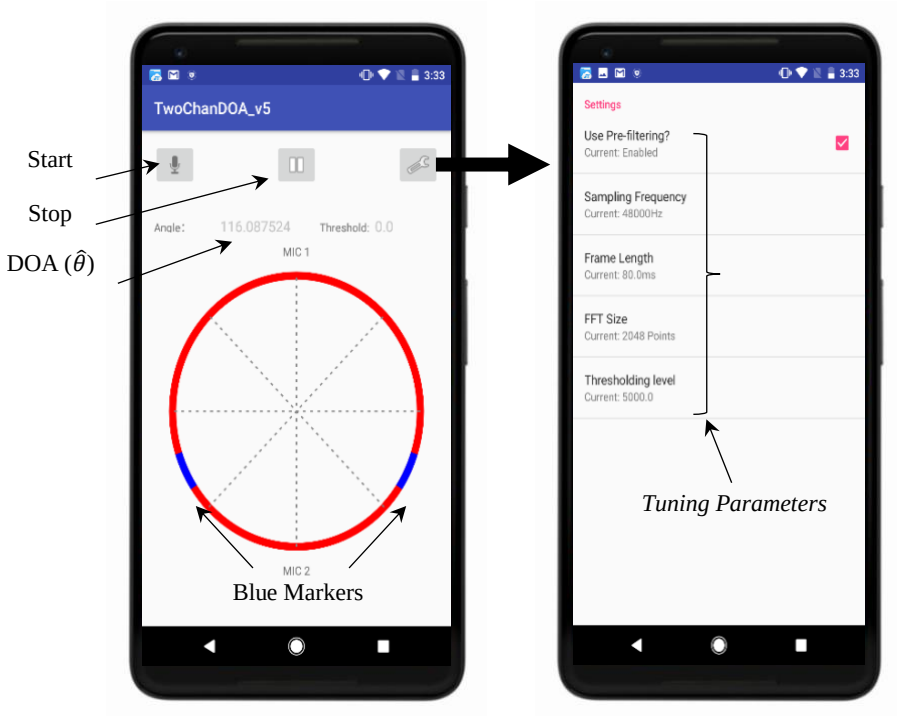


Figure 7. Average RMSE (°) comparison for Real recorded data corrupted by (a) White Noise; (b) Machinery Noise; (c) Traffic Noise; and (d) Babble Noise for different values of D, T and L.

Real-time Implementation Screenshots



Refer: https://www.utdallas.edu/ssprl/hearing-aid-project/research-papers/

Timing Information

Post-filter	Time(ms) per 100ms frame
G729B	6.5
Sohn VAD	7.6
VQVAD	9.7
Proposed VAD	**0.016**
Proposed AT VAD	**0.007**

GCC + Pre-filter	Time(ms) per 100ms frame
GCC only	7.4
GCC-PHAT-β	10.2
GCC-ML	9.1
MMSE-LSA	63.3
SGJMAP	**0.317**

F_s, L	Proposed Method
1. Proposed method (MMSE-LSA+GCC+VAD)	
16 kHz, 100ms	42.3
48 kHz, 40ms*	24.3
2. Upgraded method (SGJMAP+GCC+AT VAD)	
16 kHz, 100ms	30
48 kHz, 40ms*	20.07
*Recommended for minimum I/O latency.	

Refer: https://www.utdallas.edu/ssprl/hearing-aid-project/research-papers/