11/04/2020 **Serkan Tokgoz, Anton Kovalyov, Issa Panahi**

Technical Documentation

Real-Time Estimation of Direction of Arrival of Speech Source Using Three Microphones of Smartphone

- **Algorithm**
    a. DOA estimation: Generalized Cross-Correlation (GCC)
    b. Pre-filter: Minimum Mean Square Error-Log Spectral Amplitude estimator(MMSE-LSA)
    c. Post-filter: Simple Voice Activity Detector (VAD)(using Spectral Flux feature)
- **Implementation**
    a. MATLAB 2019a for testing and debugging
    b. Visual Studio 2019 for C and Java Native Interface (JNI) for Android
- **Frame length**
    a. MATLAB Implementation: 100ms@16kHz = 1600. FFT size = 2048.
    b. C/Java Implementation: 40ms @48kHz. FFT size = 2048.

1. SYSTEM OVERVIEW

We present a real-time noise-robust direction of arrival (DOA) estimation technique using only the three built-in microphones of an Android-based smartphone. The proposed method eliminates the 'front-back' ambiguity caused by the symmetry of the two microphones reported previously and improves the performance of DOA estimation in noisy speech environments. Our method enhances the spatial awareness of hearing-impaired users by displaying the precise DOA angle of speech source on their smartphone screen. For increased efficiency, noise-robustness, and accuracy of the proposed DOA estimation method, a spectral pre-filtering technique and a Voice Activity Detector (VAD) based post-filtering are used along with a modified generalized cross-correlation (GCC) technique. Real recorded and simulated data under realistic noisy conditions are used in the evaluations of the proposed algorithm. Real-time implementation of the proposed system is carried out on an Android-based smartphone without any additional hardware or external microphone attachments. Experimental results show the performance of the proposed method versus those without pre or post-filtering under three different noisy conditions with 0dB to 10dB signal to noise ratios (SNRs).

Figure 1 illustrates the block diagram of the proposed system. The DOA estimation is calculated using two different GCC Blocks, first block computes the GCC between Mic 1 and Mic 2. Second block does the GCC computation between Mic 1 and Mic 3. The whole process takes place on a frame-based real-time implementation.

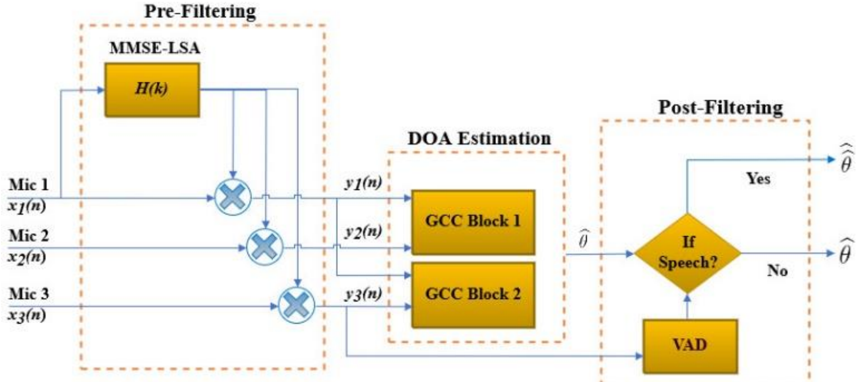

Fig. 1: Block Diagram of the proposed method. $\hat{\boldsymbol{\theta}}$ and $\hat{\hat{\boldsymbol{\theta}}}$ denotes the estimated and revised DOA, respectively. $\boldsymbol{H}(\boldsymbol{k})$ is the MMSE-LSA pre-filter

a. DOA estimation from TDOA

The estimation of DOA angle $\hat{\theta}$ and φ from estimated $\hat{\eta}$ accepts the following conditions: the microphone array geometry and speed of sound (denoted as c) are known. In this paper, we structure two GCC units formed on a three-microphone array geometry.

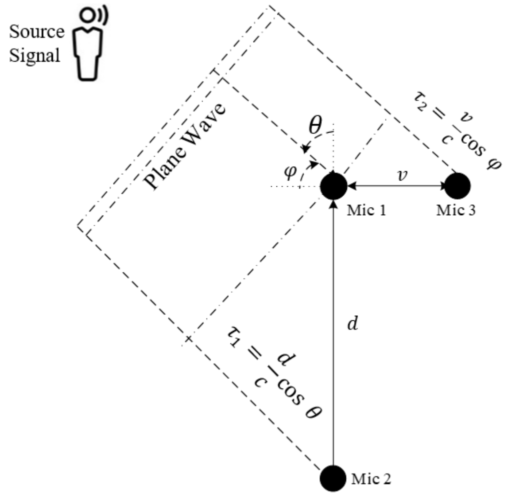

Fig. 2: Three Microphone Based GCC

Figure 2 demonstrates the known geometry of the system, where the inter-microphone distances are $d = 13\ cm$ and $v = 2.8\ cm$ for Block 1 and Block 2, respectively. θ and φ are calculated from the end-fire direction closer to Mic1 and c is assumed 343m/s.

2. EXPERIMENTS AND RESULTS

We consider RMSE over speech segments only and average over the whole experiment.

$$RMSE = \sqrt{\frac{1}{N_F}\sum_{i=1}^{N_F}(\theta_i - \hat{\theta}_i)^2} \qquad (15)$$

where N_F is the total number of frames per test. θ_i is the correct DOA angle, and $\hat{\theta}_i$ is the estimated angle for the i^{th} frame. Figure 3 shows the RMSE values of original GCC [1], GCC-Pre-filtering, GCC-Post Filtering and our proposed method with three different noise types and SNR levels. The proposed method shows a significant reduction in RMSE over all noisy conditions compare to others. The performance of GCC and the proposed method also improves as SNR increases. We also notice that the presence of the pre-filter has far more effect on decreasing the RMSE than the presence of the post-filter. Moreover, the RMSE using the proposed method is within acceptable limits for all examined SNR levels.

Table 1: RMSE results for the proposed method at each angle from smartphone recordings

Angles	0°	60°	90°	180°	270°	330°
RMSE (°)	5.46	4.81	11.57	6.92	10.36	3.68

Table 1 shows the RMSE results for the proposed method from the smartphone recordings without noise. We have recorded speech signal each positioned at different angles $\{0°, 60°, 90°, 180°, 270°, 330°\}$. According to Table 1, the proposed application's average RMSE result is 7.13° in the clean signal which is acceptable for real-world conditions and proves that the application will be helpful as a visual indicator for HI people.

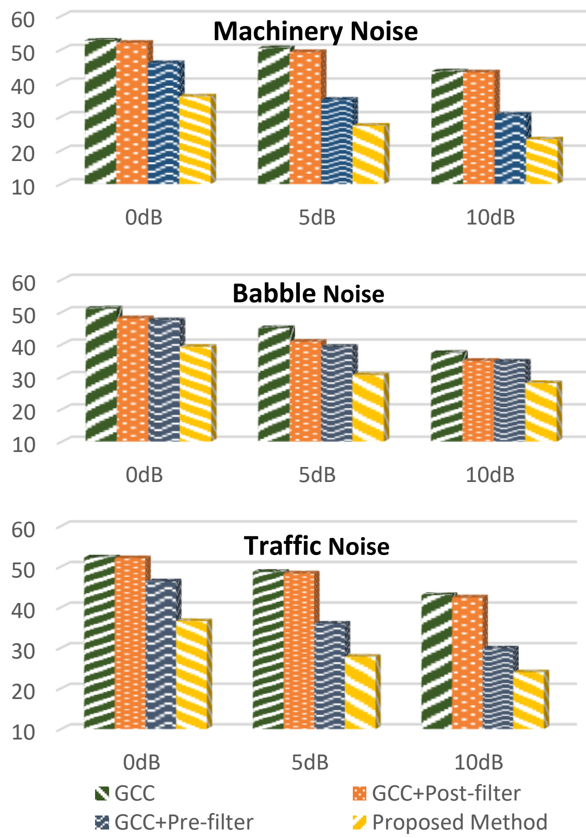

Fig. 3: Average RMSE (°) for DOA estimation using different methods for clean signal and simulated noisy data at different SNR

References

[1] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320-327, August 1976.

[2] M. Nilsson, S. D. Soli, and J. A. Sullivan, "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Am.*, 1994.