A Soft-Processor-Based FPGA Implementation of a Microphone Array Direction-of-Arrival Estimator

Trình Quốc Võ and Udo Klein, *Member, IEEE*
School of Electrical Engineering, International University
Vietnam National University – HCMC, Ho Chi Minh City, Vietnam
Email: u.klein@ieee.org

***Abstract*—This paper presents the implementation of a digital system to estimate the direction-of-arrival (DOA) of a sound source signal using the multichannel cross-correlation coefficient (MCCC) method. The implementation is based on a Nios® II soft processor incorporated in a four-microphone-channel signal processing system on a field-programmable gate array (FPGA). The MCCC algorithm is coded in C and executes on the Nios® II soft processor on an Altera® Cyclone® II FPGA. The performance of the soft-processor-based FPGA implementation of the MCCC algorithm is investigated by applying simulated microphone input signals from a pre-recorded sound source stored on a secure digital (SD) memory card. The results for noisy and reverberant environments are in good agreement with simulations and they confirm that the DOA estimation robustness is suitable for practical applications.**

I. Introduction

TIME-difference-of-arrival (TDOA) estimation plays an important role in applications such as DOA estimation and sound source localization. Various algorithms have been identified in order to achieve a reliable TDOA estimate in noisy and reverberant environments [1]. One group of algorithms uses variations of the cross-correlation method, other algorithms are based on higher-order statistics or on blind identification of real reverberant impulse response functions of the single-input multiple-output system, consisting of the sound signal source and a microphone array.

For the low-cost hardware implementation of a DOA estimator described in this paper, the MCCC algorithm has been selected because of its relatively low computational complexity. The MCCC algorithm is an extension of the generalized cross-correlation algorithm proposed by Knapp and Carter [2]. It uses the redundant information from multiple microphones to improve the reliability of the TDOA estimate. Simulations and experimental investigations of a demonstration system have supported the selection of the MCCC algorithm for a low-cost hardware implementation [3].

The hardware and software have been developed on an Altera® DE2 development board with a Cyclone® II 2C35 FPGA. The Nios® II processor is configured with the Qsys system integration tool from Altera®.

The paper is organized as follows: section II presents the mathematical background of the MCCC algorithm, section III-A describes the configuration of the experimental setup, section III-B lays out the hardware blocks of the Nios® II soft processor core and its peripherals, section III-C describes the software of the MCCC algorithm, and section IV shows typical experimental results. Section V concludes the paper and indicates ongoing developments and future work to be done.

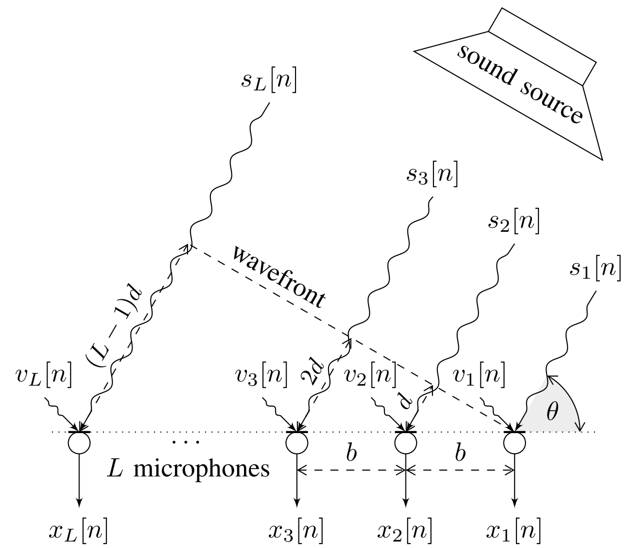

Fig. 1: Two-dimensional geometry of the TDOA problem for an equidistantly spaced array of L microphones with the source located in the far-field, the incident angle θ, and the spacing b between two neighboring microphones

II. The MCCC Algorithm

The microphone array consists of L microphones in a linear equidistantly spaced array, from the 1st to the L^{th} microphone, as shown in Fig. 1. b is the spacing between neighboring microphones, θ is the incident angle of the signal from the sound source, and n is the discrete time. $x_i[n]$ is the i^{th} microphone's output signal including the i^{th} additive noise signal $v_i[n]$ at each of the microphones, for $i = 1, 2, \ldots, L$. The noise signals are assumed to be uncorrelated both with the source signal and with the noise at the other microphones.

The delay between the i^{th} microphone and the 1st microphone is then given by

$$f_i = (i-1)\tau$$

where $\tau = d/v_{\text{a}} = b\cos\theta/v_{\text{a}}$ is the time delay between two neighboring microphones and v_{a} is the velocity of sound in air.

For the application of the MCCC algorithm, we consider the column vector of the aligned signals at the L microphones

$$\boldsymbol{x}_{1:L}\left[n - f_L(m)\right] = \left[x_1\left[n - f_L(m) + f_1(m)\right] \quad x_2\left[n - f_L(m) + f_2(m)\right] \quad \cdots \quad x_L[n]\right]^{\mathrm{T}}$$

with $m/f_s = \hat{\tau}$ as a test delay, where f_s is the sampling frequency. The corresponding spatial correlation matrix of the microphone signals is then

$$\boldsymbol{R}_{m,1:L} = \mathbb{E}\left\{\boldsymbol{x}_{1:L}\left[n - f_L(m)\right] \cdot \boldsymbol{x}^{\mathrm{T}}_{1:L}\left[n - f_L(m)\right]\right\} = \begin{bmatrix} r_{m,11} & \cdots & r_{m,1L} \\ \vdots & \ddots & \vdots \\ r_{m,L1} & \cdots & r_{m,LL} \end{bmatrix}$$

where the cross-correlation between the two signals $x_k\left[n - f_l(m)\right]$ and $x_l\left[n - f_k(m)\right]$ is given by

$$r_{m,kl} = \mathbb{E}\left\{x_k\left[n - f_l(m)\right] x_l\left[n - f_k(m)\right]\right\}.$$

The spatial correlation matrix $\boldsymbol{R}_{m,1:L}$ can be factored as

$$\boldsymbol{R}_{m,1:L} = \boldsymbol{D}\tilde{\boldsymbol{R}}_{m,1:L}\boldsymbol{D}$$

with the diagonal matrix

$$\boldsymbol{D} = \begin{bmatrix} \sqrt{\mathbb{E}\left\{x_1^2[n]\right\}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\mathbb{E}\left\{x_L^2[n]\right\}} \end{bmatrix},$$

the cross correlation coefficient matrix

$$\tilde{\boldsymbol{R}}_{m,1:L} = \begin{bmatrix} 1 & \cdots & \rho_{m,1L} \\ \vdots & \ddots & \vdots \\ \rho_{m,L1} & \cdots & 1 \end{bmatrix},$$

and the cross-correlation coefficients between $x_k\left[n - f_l(m)\right]$ and $x_l\left[n - f_k(m)\right]$

$$\rho_{m,kl} = \frac{\mathbb{E}\left\{x_k\left[n - f_l(m)\right] x_l\left[n - f_k(m)\right]\right\}}{\sqrt{\mathbb{E}\left\{x_k^2[n]\right\}\mathbb{E}\left\{x_l^2[n]\right\}}}$$

with k and $l = 1, 2, \cdots, L$.

The multichannel cross-correlation coefficient is now defined as [4]

$$\rho^2_{m,1:L} = 1 - \det(\tilde{\boldsymbol{R}}_{m,1:L}).$$

The delay estimation is then based on maximizing the cross-correlation coefficient $\rho^2_{m,1:L}$ or by minimizing the determinant of the matrix $\tilde{\boldsymbol{R}}_{m,1:L}$ with respect to the tested delay m.

III. System Design

A. Experimental Configuration

In order to verify the performance of the FPGA implementation of the MCCC algorithm, the microphone input signals to the MCCC algorithm have been generated from a pre-recorded 20-second speech signal. Four microphone channels are generated by applying the image-source method for room acoustics to the speech signal [5]. The model of the room acoustics considers a rectangular room where walls, ceiling, and floor are characterized by frequency-independent and incident-angle-independent reflection coefficients r_i. The reflection coefficients $r_i (i = 1, 2, \cdots, 6)$ are varied between 0 and 0.8. The size of the room is chosen to be 5 m by 5 m by 2.5 m. The four-channel microphone array is placed with its center at $x = 2.5$ m in parallel to the x-axis of the room at a distance of 0.5 m from the wall and 1.3 m above the floor.

The desired resolution of the estimated DOA angle has been chosen to be better than 20 degrees in the range between 30 and 150 degrees. This means the DOA estimation system needs to be able to detect a maximum delay M between two neighboring microphones of 5 samples. The relation between the minimum spacing $b_{\min}$ between two neighboring microphones, the sampling frequency f_s and the maximum delay M is given by

$$b_{\min} \geq M\frac{v_{\mathrm{a}}}{f_{\mathrm{s}}},$$

where $v_{\mathrm{a}} = 343.4$ m/s is the velocity of sound in air at 20 °C. The resulting minimum distance between two neighboring microphones is $b_{\min} = 10.7$ cm, leading to the selected microphone separation of $b = 11$ cm. The resulting DOA ranges are shown in Table I as a function of the estimated delay m .

TABLE I: Estimated delay m and corresponding DOA ranges (maximum delay $m_{\max}$ is ±5 samples)

estimated delay m	5	4	3	2	1	0	-1	-2	-3	-4	-5
DOA range (°)	0 … 29	29 … 47	47 … 61	61 … 73	73 … 84	84 … 96	96 … 107	107 … 119	119 … 133	133 … 151	151 … 180

B. System Hardware

As expected from the theoretical background and as verified by simulations [6], the reliability of the MCCC algorithm in noisy and reverberant environments increases with the number of microphones in the array. Since the order of the MCCC matrix grows with the square of the number of microphones, the computational complexity of the numerical calculation of the MCCC matrix determinant increases rapidly with the number of channels. Based on the simulation results with respect to the minimum number of channels required for a practical microphone array, an array with four microphones has been chosen for our experimental investigation.

In order to implement the MCCC algorithm on the Cyclone® II FPGA, the Altera® Nios® II soft processor has been chosen, providing a small-size and high-performance processor core to execute the algorithm efficiently. Using a soft processor allows the easy use of numbers in floating-point format to calculate the MCCC matrix determinant, thereby avoiding critical overflow issues. The design of the digital system is composed of a Nios® II/e processor which is a freely available Nios® II processor configuration using few FPGA logic and memory resources. Four kilobyte of on-chip memory and 8-MB SDRAM are used to store the software and the data to be processed on the Nios® II/e processor. An SD card controller is created to interface with an SD memory card storing the sound signals for each microphone channel and for different experimental configurations. Seven slide switches allow the selection of a specific experimental configuration and its corresponding microphone sound signals. A 21-bit parallel

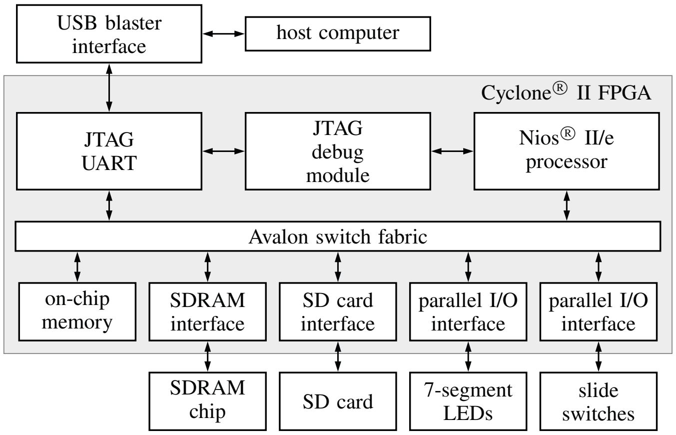

Fig. 2: Block diagram of the customized Nios® II/e processor for the experimental implementation of the MCCC algorithm

output is used to connect to the three 7-segment LEDs in order to display the result of the DOA estimate. All of the peripherals are connected to the Nios® II/e processor core via the Avalon switch fabric. The block diagram of the customized Nios® II system is shown in Fig. 2.

The customized Nios® II/e processor is configured in the Altera® Qsys system integration tool and the design is compiled using the Altera® Quartus II Web Edition software. The Nios® II/e implementation of the four-channel MCCC algorithm uses only 3,641 (= 11%) logical elements and 220,160 (= 46%) memory bits of the FPGA's resources.

C. System Software

The MCCC algorithm is coded in C and executes on the Nios® II soft processor. The code processes the microphone signals and outputs the result of the DOA estimate. Fig. 3 shows the flowchart of the software code.

At first, the four-channel microphone data is read from the SD card using Altera® library functions. Each data sample is normalized before it is stored in a register for processing. After the input buffer of 512 samples per channel is filled, the sound energy of the sampled segment is computed. Voice activity detection (VAD) is applied by comparing the sound energy with an energy threshold [7]. The threshold energy is chosen to be 7 dB above the energy of a frame without a signal.

The core part of the software is the calculation of the determinant of the MCCC matrix for delays m between $\pm m_{\max}$. The code determines each element of the MCCC matrix by calculating the cross-correlation coefficients between each combination of microphone channel pairs. The determinant of the MCCC matrix can then easily be computed recursively. The minimum of the determinants of the MCCC matrices for each tested delay m is then used to select the delay m_{est} that corresponds to the best matching alignment of the multiple microphone channels. This delay m_{est} corresponds to the best estimate of the TDOA and is used to estimate the DOA angle. The result is finally sent via an Altera® library function to the 7-segment LED display. After displaying the estimated DOA angle, the code returns to reading the next frames of the four-channel microphone signals and continues the MCCC algorithm in an infinite loop as shown in Fig. 3.

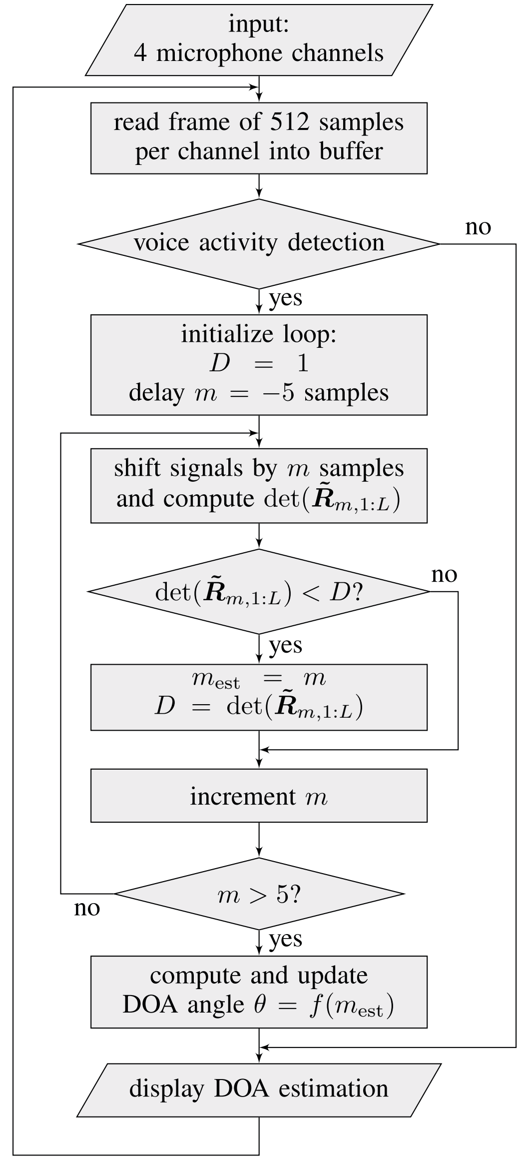

Fig. 3: Flowchart of the MCCC algorithm to be executed on the Nios® II soft processor for the DOA estimator

IV. Experimental Results

In order to investigate the performance of the soft-processor-based FPGA implementation of the DOA estimator, various microphone input signals were processed and the results were analyzed. The input signals corresponded to several sound source locations, different room reverberation levels, and changing signal-to-noise ratios (SNR) of the speech signal. The microphone signals are processed in 512-sample-long frames, corresponding to a frame length of 32 ms.

As an example, the DOA estimate for a sound source location at ($x = 4$ m, $y = 2.5$ m, $z = 1.5$ m) is shown in Fig. 4. For a speech signal of 10 dB SNR and in a reverberant-free room, the DOA estimate of 54° is the correct value for this particular sound source location.

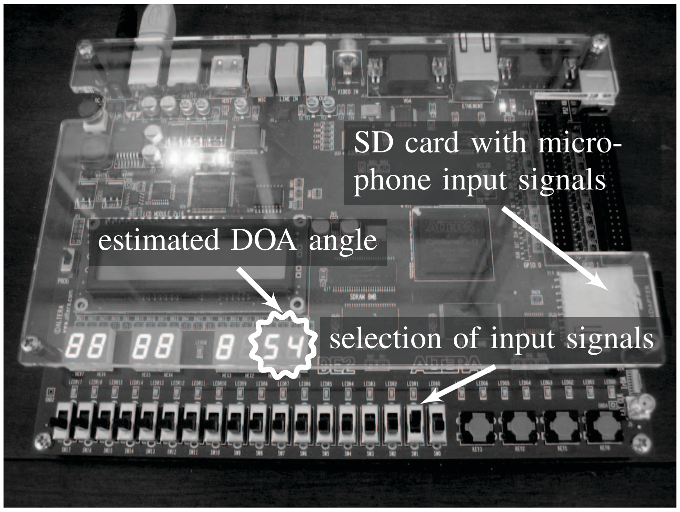

Fig. 4: The DE2 development board with the estimated DOA result of a four-microphone array for a sound source location at ($x = 4$ m, $y = 2.5$ m, $z = 1.5$ m), i.e. the correct DOA is 54°; the microphone input signals are for a reverberant-free environment with a signal-to-noise ratio of 10 dB

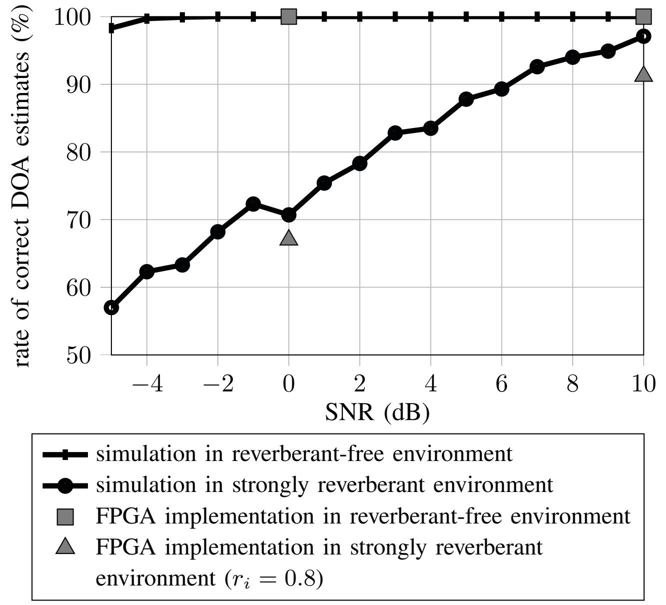

Fig. 5: MATLAB® simulation of percentage of correct DOA estimates compared to actual performance of the FPGA implementation for a four-channel microphone array in a reverberant-free and in a strongly reverberant environment with wall reflection coefficients $r_i = 0.8$

The reliability of the DOA estimation has been studied as a function of the sound signal's SNR and the level of room reverberation. With a sound source location at ($x = 4$ m, $y = 2.5$ m, $z = 1.5$ m), 1,000 DOA estimates were processed with statistically varying microphone input signals according to 0 dB and 10 dB SNR. The DOA estimates were performed both in a reverberant-free and in a strongly reverberant ($r_i = 0.8$) room. The rate of correct DOA estimates for these experimental configurations is shown in Fig. 5 together with the results of MATLAB® simulations of the MCCC algorithm [6]. For a reverberant-free environment, both the MATLAB® simulations and the FPGA implementation show 100% correct DOA estimates for 0 dB and 10 dB SNR. In strongly reverberant room environments, the rate of correct DOA estimates drops significantly with increasing acoustic noise. At 0 dB SNR, only 67% of DOA estimates are correct, increasing to 91% at 10 dB SNR. Due to numerical rounding and lower precision, both these DOA estimates are slightly lower than the rate of correct DOA estimates in corresponding MATLAB® simulations.

V. Conclusion

The experimental investigation of the soft-processor-based FPGA implementation for DOA estimation confirms that a simple four-microphone array in combination with the MCCC algorithm can achieve reliable DOA estimates in typical acoustical situations of at least 10 dB SNR and in common reverberant environments. For exceptionally low SNRs and in extremely reverberant environments, a higher number of microphone channels is required for reliable DOA estimation, leading to a significant increase in computational complexity and corresponding resources of the hardware implementation.

The hardware implementation of the MCCC algorithm based on a Nios® II soft processor is shown to be a simple digital design, using the flexibility and ease of a software-implemented algorithm in combination with the efficiency of an FPGA hardware implementation. Since only 11% of the Cyclone® II 2C35 FPGA's logical elements are being used, the remaining hardware resources are available for other signal processing tasks. A full hardware implementation of an DOA estimator will employ a digital MEMS microphone array. The 2.4-MHz pulse-density-modulated digital output signals of the MEMS microphones need to be decimated and filtered before processing with the MCCC algorithm. These digital signal processing tasks can be implemented within the remaining Cyclone® II 2C35 FPGA resources. The resulting DOA estimator system will be smaller, cheaper, and more flexible than conventional analog microphone arrays.

References

[1] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin and Heidelberg, Germany: Springer-Verlag, 2008, ch. 9. Direction-of-arrival and time-difference-of-arrival estimation, pp. 181–215.

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

[3] T. Q. Võ and U. Klein, *Development of compact microphone array for direction-of-arrival estimation*, ser. Lecture Notes in Electrical Engineering. Springer, 2013, vol. 240, pp. 1189–1195.

[4] J. Chen, J. Benesty, and Y. A. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Speech Audio Process.*, vol. 11, no. 6, pp. 549–557, Nov. 2003.

[5] E. A. Lehmann. (2012, Mar. 10) Image-source method: MATLAB code implementation. [Online]. Available: http://www.eric-lehmann.com

[6] U. Klein and T. Q. Võ, "Direction-of-arrival estimation using a microphone array with the multichannel cross-correlation method," in *Proceedings of the IEEE International Symposium on Signal Processing and Information Technology*, Dec. 2012.

[7] L. Liu, M. Alam, and X. Fu, "Voice activity detection and noise reduction," Dept. of Communication Technology, Aalborg University, Tech. Rep., 2005.