Direction-of-Arrival Estimation Using a Microphone Array with the Multichannel Cross-Correlation Method

Udo Klein, *Member, IEEE,* and Trình Quốc Võ
School of Electrical Engineering, International University, Vietnam National University – HCMC
Ho Chi Minh City, Vietnam
Telephone: +84 (8) 3724 4270 Ext. 3231; Fax: +84 (8) 3724 4271; Email: u.klein@ieee.org

***Abstract*—In this paper, the performance of a multichannel cross-correlation algorithm for the estimation of the direction-of-arrival (DOA) of an acoustic source in the presence of significant levels of both noise and reverberation is presented. Microphone arrays are being utilized in many different areas from consumer electronics to military systems. Noise reduction, speech recognition, hands-free communication, automatic video camera steering and multiparty teleconferencing are some of the applications being studied. Methods based on the time-differences-of-arrival (TDOA) between two microphones are commonly used to determine the direction of an acoustic source. The performance of TDOA algorithms typically deteriorates significantly due to noise and multipath propagation. This paper deals with the DOA problem emphasizing the performance in noisy and reverberant environments. The multichannel cross-correlation coefficient (MCCC) is used to estimate the DOA and the performance of the MCCC algorithm is investigated. Simulations and initial experimental results confirm that the DOA estimation robustness is suitable for practical applications if arrays with an appropriate number of microphones are used.**

***Keywords*—Direction of arrival estimation, microphone arrays, signal processing algorithms, time of arrival estimation.**

I. Introduction

In the simple case of two microphones, the TDOA problem is depicted in Fig. 1, where b is the spacing between the two microphones, θ is the incident angle of the signal from the sound source, and n is the discrete time. $x_1[n]$ and $x_2[n]$ are the microphones' output signals including the additive noise signals $v_1[n]$ and $v_2[n]$ at each of the microphones. The noise signals are assumed to be uncorrelated both with the source signal and with the noise at the other microphone. The acoustical path difference between the source and the two microphones is $d = b\cos\theta$. The incident angle can be calculated if the microphone spacing b is known and d is determined by the time delay estimate between the source signal arriving at microphones 1 and 2.

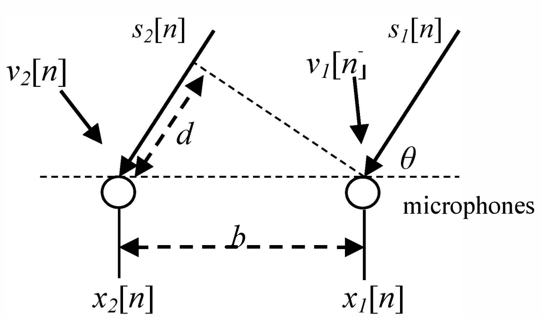

Fig. 1. Two-dimensional geometry of the TDOA problem for two microphones with the source located in the far-field, the incident angle θ, and the spacing b between the two microphones

Many algorithms have been proposed to estimate the TDOA. In noisy and reverberant environments two main approaches are commonly used for TDOA estimation [1]: the first approach uses more than two microphones to achieve a robust prediction based on redundant information of a microphone array [1], [2], and the second approach uses blind estimation [3], [4], [5]. In this work, the multichannel cross-correlation coefficient (MCCC) method has been selected for a hardware implementation of a microphone array for DOA estimation of an acoustic source in the presence of significant levels of both noise and reverberation. The MCCC method uses the redundancy of the microphone signals by applying an extension of the generalized cross-correlation proposed by Knapp and Carter [6]. In order to evaluate the performance of the MCCC method the algorithm has been implemented in MATLAB® and simulation results are presented in this paper. There are two main reasons for us to select the MCCC algorithm for a hardware implementation. Firstly, the MCCC method uses redundant information in a microphone array to achieve a more robust estimate in reverberant environments that is less sensitive to noise than other methods [1]. Secondly, the MCCC algorithm requires a lower computational complexity compared to the blind estimation approach. Additionally, the MCCC method is one of the TDOA algorithms which can be expanded to predict the directions of multiple sources.

The paper is organized as follows: section II presents the mathematical background of the MCCC algorithm, section III describes the configuration of the simulated environment and presents typical results of the simulations, and section IV outlines a demonstration system. Section V concludes the paper and indicates future work to be done.

978-1-4673-5604-6/12/$31.00 ©2012 IEEE

II. THE MCCC ALGORITHM

The microphone array consists of L microphones in a linear equidistantly spaced array, from the 1^{st} to the L^{th} microphone. The delay between the 1st and the l^{th} microphones is then given by

$$f_l = (l-1)\tau$$

where τ is the time delay between two neighboring microphones.

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

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

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

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

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 symmetric 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 $l = 1, 2, \cdots, L$.

In the case of two microphones, the two-channel cross-correlation coefficient is given by

$$\rho_{m,12}^2 = 1 - \det\tilde{\boldsymbol{R}}_{m,1:2}.$$

Similarly, the multichannel cross-correlation coefficient is defined as [1]

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

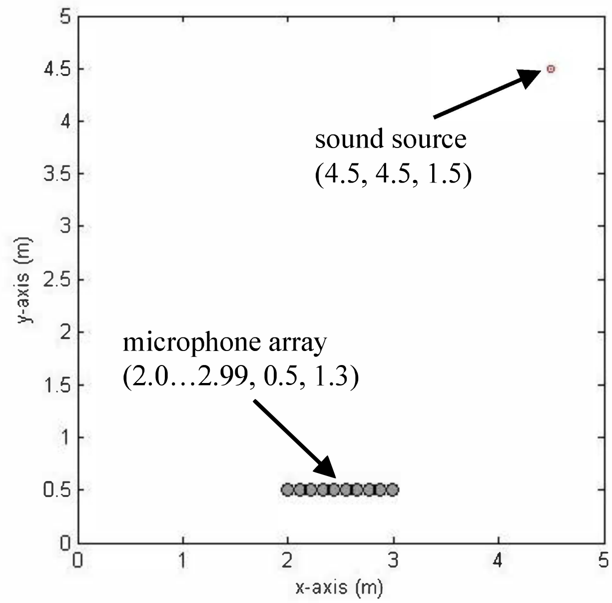

Fig. 2. Configuration of the sound source and the ten microphones in a linear array in the acoustically simulated room

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

III. SIMULATION

A. Room Configuration

In order to simulate the reverberant acoustic environment the image-source method for room acoustics has been employed [7]. The simulation considers a rectangular room. Walls, ceiling, and floor of the room are characterized by frequency-independent and incident-angle-independent reflection coefficients. The dimensions of the room are chosen to be 5 m by 5 m by 2.5 m. Reflection coefficients $r_i (i = 1, 2, \cdots, 6)$ are varied between 0 and 1. As shown in the layout of the room configuration in Fig. 2 the sound source is located at the position $(4.5\text{ m}, 4.5\text{ m}, 1.5\text{ m})$.

For the simulations, up to ten microphones are placed in parallel with the x-axis and with a spacing of 11 cm. The first microphone is located at $(2.99\text{ m}, 0.5\text{ m}, 1.3\text{ m})$ and the last is at $(2.0\text{ m}, 0.5\text{ m}, 1.3\text{ m})$. The signal-to-noise ratio (SNR) has been varied between -10 dB and 5 dB. A 20-second recorded speech signal was sampled at $f_s = 16$ kHz with 16-bit resolution and has been used as the sound source for the simulations.

The desired resolution of the estimated incident angle has been chosen to be better than 20 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

$$b_{\min} \geq M \frac{v_a}{f_s}$$

where $v_a = 343.4$ 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.

B. Simulation Results

In the case of a noisy environment without reverberation, the performance of the algorithm for a given source signal only depends on the SNR and the number of microphones in the array. Fig. 3 shows the determinant of the correlation matrix $\det \tilde{\boldsymbol{R}}_m$ as a function of the guessed delay m in a reverberant-free environment with -5 dB SNR and using three, four, and ten microphones, respectively. $\det \tilde{\boldsymbol{R}}_m$ is the cost function of the MCCC algorithm with its minimum at the time delay estimate $\hat{\tau} = m f_s$. In the case of a three-microphone array, the minimum in the cost function is very shallow and the estimated delay is 1 sample while the true delay is 2 samples. Increasing the number of microphones improves the minimum in the cost function and the estimated delay is equal to the true delay for arrays with more than three microphones. Fig. 3 shows that the minimum of the cost function becomes sharper as the number of microphones increases, improving the search for the minimum value due to an increasing number of microphones. The result shows that the MCCC algorithm takes the redundant information to enhance the robustness of the TDOA estimation in a noisy environment.

A typical distribution of the time delay estimates for repeatedly applying the MCCC algorithm in a reverberant-free environment with -10 dB SNR is shown in Fig. 4. With a three-microphone array, only 55% of the estimates correspond to the true delay of 2 samples and the erroneous estimates are up to ± 2 samples off the true delay. For a five-microphone array, the width of the erroneous estimates is reduced to ± 1 sample and for a ten-microphone array, almost 100% of the estimates correspond to the true delay of 2 samples.

In a strongly reverberant environment the distribution of the time delay estimates becomes considerably more spread out (Fig. 5). Only arrays with a significant number of microphones, such as the ten-microphone array in Fig. 5, show a satisfactory success rate of the time delay estimate in noisy and strongly reverberant environments.

Fig. 6 shows the percentage of correct delay estimates in a noisy environment without reverberation as a function of the SNR. The robustness of the algorithm to estimate the delay

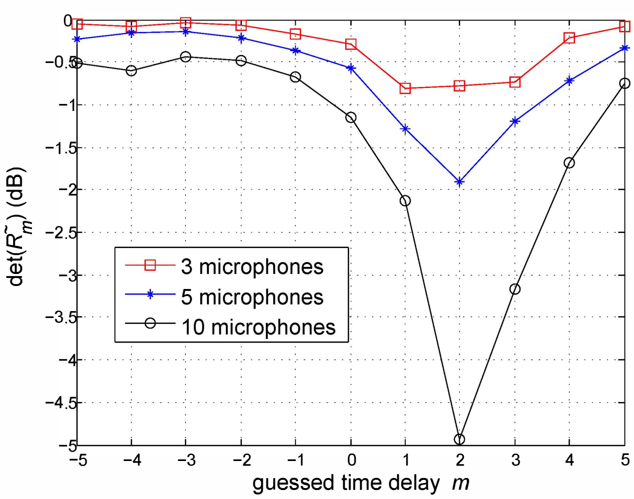

Fig. 3. Performance of the MCCC algorithm in terms of the cost function $\det \tilde{\boldsymbol{R}}_m$ as a function of the guessed delay m for microphone arrays with $L = 3, 5,$ and 10 microphones and with a SNR of -5 dB (The true delay is 2 samples.)

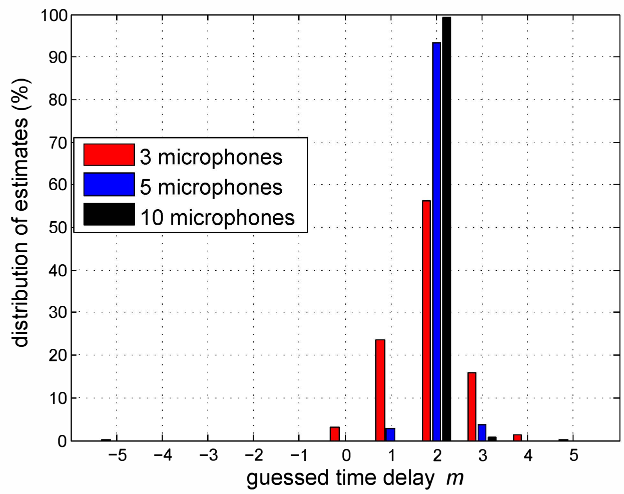

Fig. 4. Distribution of the time delay estimates for repeatedly applying the MCCC algorithm 1,000 times in a reverberant-free environment with $L = 3, 5,$ and 10 microphones and with a SNR of -10 dB (The true delay is 2 samples.)

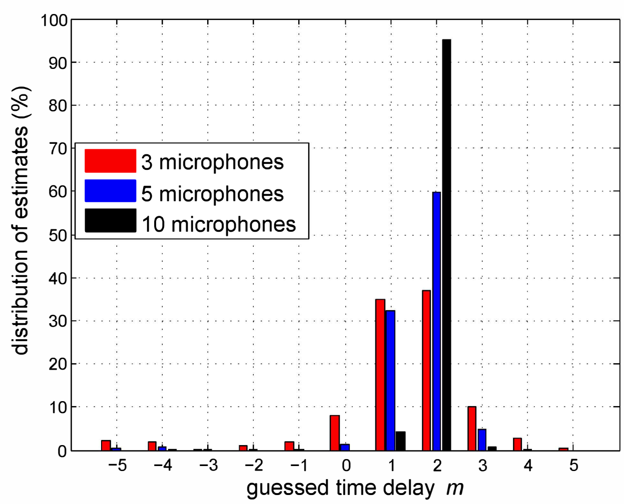

Fig. 5. Distribution of the time delay estimates for repeatedly applying the MCCC algorithm 1,000 times in a strongly reverberant environment ($r_i = 0.8$) with $L = 3, 5,$ and 10 microphones and with a SNR of -10 dB (The true delay is 2 samples.)

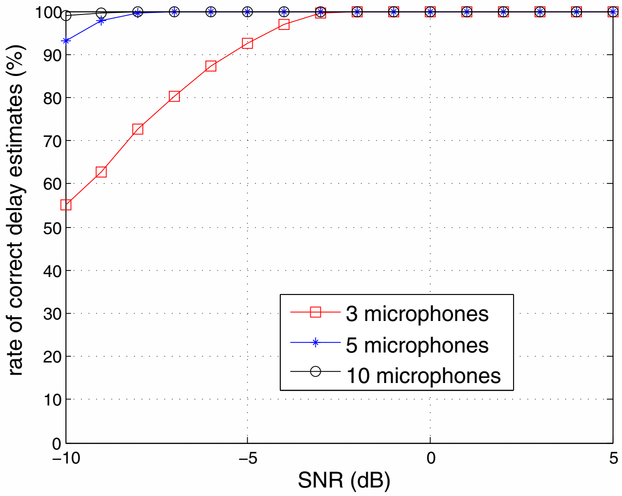

Fig. 6. The robustness of the MCCC algorithm in terms of percentage of correct delay estimates as a function of the SNR for arrays with varying numbers of microphones in a reverberant-free environment

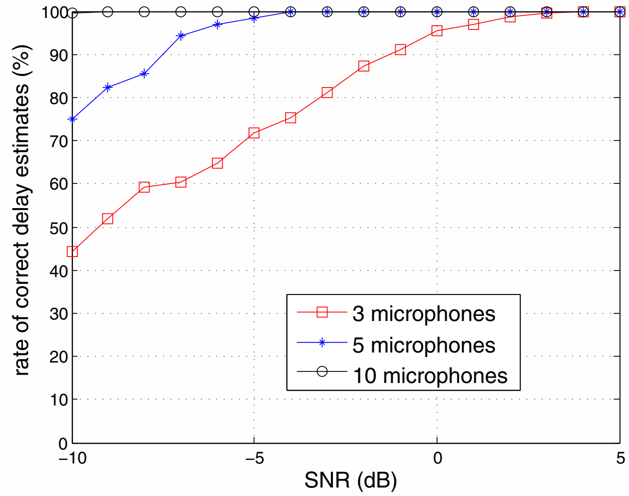

Fig. 7. The robustness of the MCCC algorithm in terms of percentage of correct delay estimates as a function of the SNR for arrays with varying numbers of microphones and in a weakly reverberant environment (wall reflection coefficients $r_i = 0.5$)

correctly increases with the number of microphones in the array. The simulation shows that in a very noisy environment with a SNR of -10 dB the percentage of successful delay estimations is about 55% for an array with three microphones. However, the rate of correct estimates increases to over 94% if an array with five microphones is used. The reliability of the delay estimate using an array with ten microphones reaches more than 99% even with a SNR as low as -10 dB. Both five-microphone and ten-microphone arrays achieve 100% correct delay estimates for SNRs better than -7 dB.

Fig. 7 shows the percentage of correct delay estimates in a noisy and weakly reverberant environment with wall reflection coefficients of $r_i = 0.5$. With a SNR of -10 dB the percentage of correct delay estimates is reduced to about 45% for an array with three microphones and to about 75% for an array with five microphones. Using an array with ten microphones the reliability of the delay estimate is not noticeably degraded by the reverberation in the room with wall reflection coefficients of $r_i = 0.5$. The five-microphone array in the weakly reverberant room requires a SNR of better than -5 dB to achieve nearly 100% correct delay estimates. The robustness of a three microphone array with a SNR of -5 dB is comparable to a five-microphone array with a SNR of -10 dB.

In more strongly reverberant environments with wall reflection coefficients of $r_i = 0.8$ (comparable to a typical office) the number of correct delay estimates deteriorates significantly compared to reverberant-free and weakly reverberant environments under the same SNR condition (Fig. 8). Nevertheless, the rate of correct delay estimates increases with the number of microphones and a ten-microphone array still achieves reliable results in noisy and strongly reverberant environments. The five-microphone array in the strongly reverberant room achieves nearly 100% correct delay estimates for SNRs above 0 dB while the ten-microphone array shows a satisfactory rate

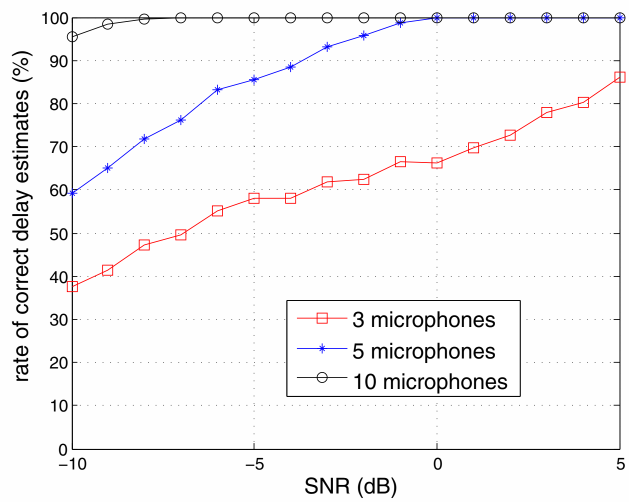

Fig. 8. The robustness of the MCCC algorithm in terms of percentage of correct delay estimates as a function of the SNR for arrays with varying numbers of microphones and in a weakly reverberant environment (wall reflection coefficients $r_i = 0.8$)

of correct delay estimates of better than 95% down to -10 dB SNR.

All data in the Fig.s 3 to 8 are evaluated by processing a signal frame of 1,024 samples. Fig.s 9 and 10 show comparable data to Fig. 3, that is the determinant of the correlation matrix $\det \tilde{\boldsymbol{R}}_m$ as a function of the guessed delay m in a reverberant-free environment with -5 dB SNR and using a data frame length of 512 and 256 samples, respectively. Even for a very short frame length of only 256 samples, which corresponds to a sampling length of 16 ms, the search for the minimum of the cost function as a function of the guessed time delay can be performed reliably for a ten-microphone array. The minimum of the cost function becomes critically

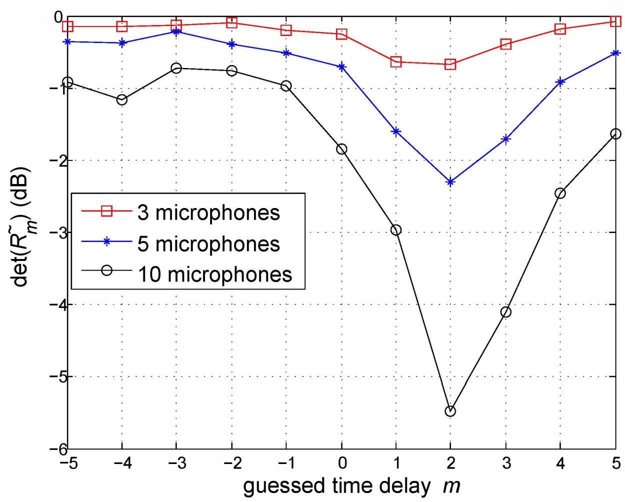

Fig. 9. The performance of the MCCC algorithm in terms of the cost function det $\tilde{\boldsymbol{R}}_m$ as a function of the guessed delay m for a frame length of 512 samples and with a SNR of -5 dB (The true delay is 2 samples.)

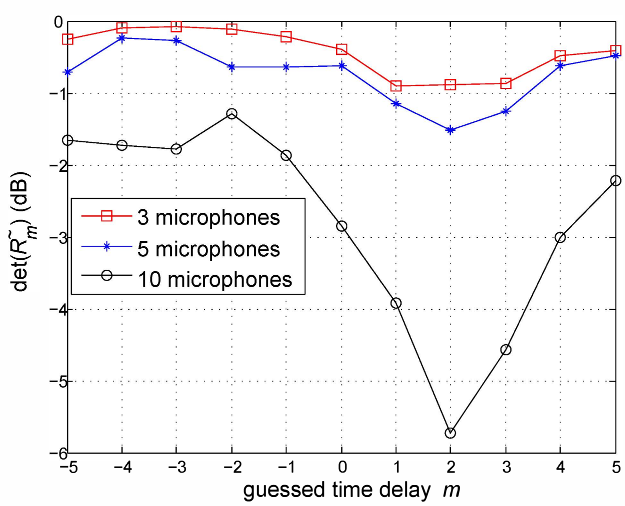

Fig. 10. The performance of the MCCC algorithm in terms of the cost function det $\tilde{\boldsymbol{R}}_m$ as a function of the guessed delay m for a frame length of 256 samples and with a SNR of -5 dB (The true delay is 2 samples.)

less pronounced for a five-microphone array and a frame length of 256 samples (Fig. 10). Fig. 11 shows the number of correct delay estimates as a function of the SNR in a strongly reverberant environment evaluated with a data frame length of 256 samples. Compared to the corresponding data with a data frame length of 1,024 samples in Fig. 8, the rate of correct delay estimates is clearly reduced, although the ten-microphone array still performs satisfactorily for SNRs above -5 dB.

The simulations confirm the robustness of the MCCC algorithm in estimating the TDOA in both noisy and reverberant environments, even for comparatively short signal durations. A linear equidistantly spaced array of ten microphones performs reliably in strongly reverberant environments with SNRs down to -5 dB and with signal durations as short as 16 ms.

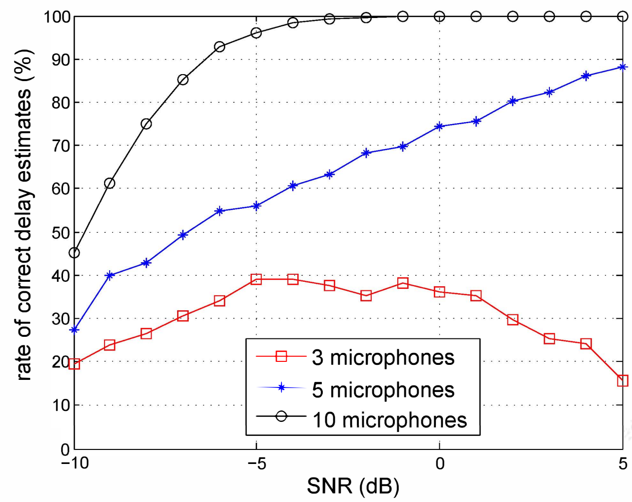

Fig. 11. The robustness of the MCCC algorithm in terms of percentage of correct delay estimates as a function of the SNR for a frame length of 256 samples and in a strongly reverberant environment (wall reflection coefficients $r_i = 0.8$) (Compare to Fig. 8 showing the same data for a frame length of 1,024 samples.)

IV. Demonstration System

Following our simulations we have set up a demonstration system using a microphone array with analog MEMS microphones and a simple analog-to-digital (A/D) converter as the interface to the MCCC algorithm in MATLAB® (Fig. 12). The limitations of the A/D converter allow a maximum of four microphone channels at a sample rate of 10.5 kS/s with a resolution of 14 bits. Because of the reduced sample rate compared to the simulation, the distance between the microphones is increased to 17.5 cm.

For a sound source at a distance of 2.5 m and at an angle of 35° to the microphone array, the cost function is shown in Fig. 13 for microphone arrays with 2, 3, and 4 channels.

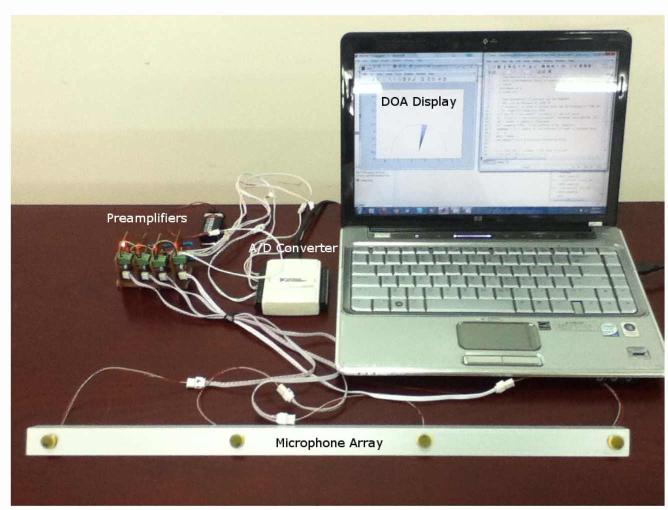

Fig. 12. Demonstration system with 4-channel array of analog MEMS microphones, preamplifiers, and an A/D converter interface to the MCCC algorithm in MATLAB®

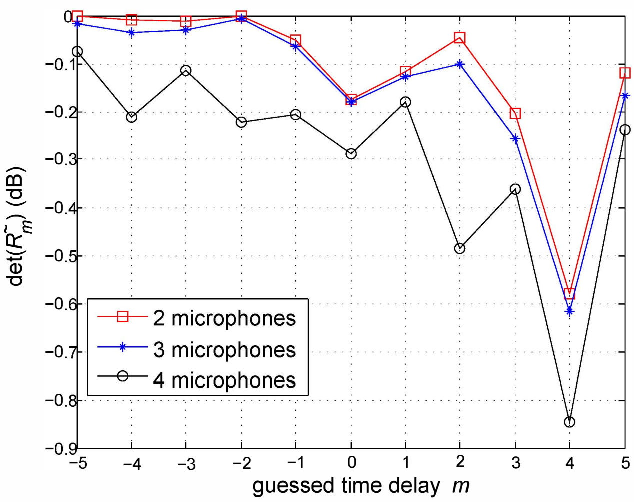

Fig. 13. The MCCC cost function det $\tilde{\boldsymbol{R}}_m$ as a function of the guessed delay m for microphone arrays with $L = 2, 3,$ and 4 microphones and with a SNR of about 5 dB in a reverberant meeting room environment

The estimated delay of $m = 4$ corresponds to a DOA range between $28°$ and $47°$, which includes the correct DOA of $35°$.

V. Conclusion

Many methods have been developed to estimate the time delay of a sound source between spatially separated microphones. Nevertheless, it is still difficult to estimate the time delay with a practical system in a real-world environment with substantial noise and reverberation. The results of our studies show that the MCCC algorithm is a suitable candidate for reliable TDOA estimation in real-world environments with a minimum amount of computational cost. The MCCC algorithm is a general case of the cross-correlation method. The TDOA can be estimated through the determinant of the cross-correlation coefficient matrix of the aligned microphone array signals.

First experimental results using a 4-channel demonstration system show that the simulated performance can be achieved with miniature analog MEMS microphones and a simple analog-to-digital converter as the interface to the MCCC algorithm in MATLAB®. The full hardware implementation will employ an FPGA and a microphone array consisting of digital MEMS microphones. The digital MEMS microphones integrate the microphone, amplifier, and A/D converter in a single component, thereby reducing the system complexity considerably. The system will be smaller, cheaper, and more flexible than conventional analog microphone arrays. In order to evaluate its performance the MEMS microphone arrays will be used to record test data for source localization experiments.

Acknowledgment

The authors would like to thank Pirmin Rombach and Armin Schober from EPCOS AG for providing free samples of their MEMS microphones.

References

[1] 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.

[2] 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.

[3] J. Benesty, Y. Huang, and J. Chen, "Time delay estimation via minimum entropy," *IEEE Signal Processing Letters*, vol. 14, no. 3, pp. 157–160, Mar. 2007.

[4] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.

[5] Y. A. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multi-channel identification," *IEEE Transactions on Signal Processing*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[6] 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.

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