TIME DELAY ESTIMATION IN A REVERBERANT ENVIRONMENT BY LOW RATE SAMPLING OF IMPULSIVE ACOUSTIC SOURCES

Muhammad Omer, Ahmed A. Quadeer*, Mohammad S. Sharawi*, and Tareq Y. Al-Naffouri*†*

Electrical Engineering Deptartment,
*King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia
†King Abdullah University of Science & Technology, Thuwal, Saudi Arabia
{muhammadomer, aquadeer, msharawi, naffouri}@kfupm.edu.sa

ABSTRACT

This paper presents a new method of time delay estimation (TDE) using low sample rates of an impulsive acoustic source in a room environment. The proposed method finds the time delay from the room impulse response (RIR) which makes it robust against room reverberations. The RIR is considered a sparse phenomenon and a recently proposed sparse signal reconstruction technique called orthogonal clustering (OC) is utilized for its estimation from the low rate sampled received signal. The arrival time of the direct path signal at a pair of microphones is identified from the estimated RIR and their difference yields the desired time delay. Low sampling rates reduce the hardware and computational complexity and decrease the communication between the microphones and the centralized location. The performance of the proposed technique is demonstrated by numerical simulations and experimental results.

1. INTRODUCTION

Time delay estimation (TDE) serves as the front end for detection, identification and localization of acoustic sources. The conceptual simplicity and low computational complexity of the TDE methods makes them well suited for real time source localization with several sensors. TDE for impulsive acoustic sources becomes important in critical situations for examples, locating a gun shot by a thief or a burglar in a shopping mall or an explosion in an oil or gas pipeline in a refinery.

Cross correlation (CC) being the simplest TDE method, assumes an ideal signal propagation model. It correlates the source signal received at the reference microphone with its delayed and scaled version received at the other microphone. The time delay corresponds to the time index with maximum correlation value. The cross correlation function gets spread by the signal autocorrelation function. In an indoor environment, multiple broad peaks appear due to closely spaced signal reflections which may overlap to produce a peak corresponding to an incorrect time delay. A class of algorithms called generalized cross correlation (GCC) employs various weighting factors to maximize the sharpness of the cross correlation peak [1]. These algorithms also assume the ideal signal propagation model and perform well in moderately noisy and non-reverberant environments. However, when the noise and reverberation level is considerable, the performance of the cross correlation based algorithms and their variants deteriorates. Thus, more robust algorithms are required to alleviate these effects. An estimate of the RIR is important to determine the extent of room reverberation [3], [4]. The adaptive eigen value decomposition (AED) algorithm deals with TDE in reverberant indoor environments [5], [6]. The algorithm estimates the RIR iteratively from the source to the pair of microphones and finds the time delay by identifying the direct line of sight (DLOS) signal from the estimated RIR.

The performance of the most widely used TDE methods is reviewed in [5]. It is shown that the non-realistic signal propagation model assumption of CC and GCC methods is the cause of their inability to combat room reverberations. In contrast, the AED algorithm copes well with the room reverberations but it is computationally intensive and suffers from poor tracking capabilities. In addition, these TDE methods suffer from poor time resolution at low sampling frequencies caused by the missing time information which results in large uncertainty in the times of arrival. The effect of under-sampling on the performance of these algorithms is discussed in [2]. The high sampling frequency requirement of these algorithms makes the TDE process not only computationally intensive but also puts a strain on the hardware requirements. An overview of problems associated with the existing TDE methods and possible solutions have been discussed in [5].

In this paper, we are interested in finding the TDE between a pair of microphones from the RIR estimate. The RIR is considered to be a sparse signal due to the finite number of impulses, each corresponding to a multipath component. Each microphone acquires the acoustic signal at low sampling rate and transfers it to a centralized workstation. The RIR is estimated using a recently developed sparse signal reconstruction algorithm called orthogonal clustering (OC) [7], [8]. The DLOS signal is identified

The authors would like to acknowledge the support provided by King Abdulaziz City for Science and Technology (KACST) through the Science & Technology Unit at King Fahd University of Petroleum & Minerals (KFUPM) for funding this work through Project No. 09-ELE763-04 as part of the National Science, Technology and Innovation Plan.

from each of the estimated RIR and their difference yields the desired TDE. Low rate sampling relaxes the communication link between the microphones and the centralized workstation and at the same time decreases the computation time and the hardware complexity. TDE using low rate samples has been studied in [9] but it differs from this work as it employs the ESPRIT algorithm for time delay estimation as opposed to the fast OC algorithm adopted here. While [9] deals with low rate samples in a compressed sampling sense, it is difficult to implement when the signal dimension is large as is the case here. Future work will address the performance of the method of [9] compared to the one presented here for small dimensions.

The paper is organized as follows. Section 2 describes the problem and Section 3 discusses the proposed TDE technique using the OC algorithm. Section 4 presents the simulated and experimental setup results and Section 5 concludes the paper.

2. PROBLEM FORMULATION

In a room environment, the source-microphone pair separated by an acoustic space can be described by a linear time invariant system. Given that the acoustic space is excited by a known excitation signal $s(t)$, the signal received at the microphone can be expressed as,

$$r(t) = \sum_{l=0}^{L-1} \alpha_l s(t - \tau_l) + \omega(t) \tag{1}$$

where L is the number of paths capturing significant multipath energy. The impulse response estimation is actually the estimation of the parameters α_l and τ_l, the scaling magnitude factor and the time shift respectively and $\omega(t)$ is the zero mean additive white Gaussian noise. With discrete time signals, the matrix form of (1) can be written as,

$$\mathbf{r} = \mathbf{S}\boldsymbol{\alpha} + \boldsymbol{\omega} \tag{2}$$

where $\mathbf{r}$ and $\boldsymbol{\alpha}$ are length N discrete time, received and RIR signals respectively while $\boldsymbol{\omega}$ is the additive white Gaussian noise of length N with zero-mean and covariance matrix $\mathbf{C}_\omega = \sigma_{\boldsymbol{\omega}}^2 \mathbf{I}$. The $N \times N$ matrix $\mathbf{S}$ is called the dictionary matrix which is formed by the N discretized and delayed versions of the basic source signal $s(t)$

$$\mathbf{S} = \begin{bmatrix} s(0-\Delta) & \dots & s(0-N\Delta) \\ s(1-\Delta) & \dots & . \\ . & . & . \\ . & . & . \\ . & . & . \\ s((N-1)-\Delta) & \dots & s((N-1)-N\Delta) \end{bmatrix} \tag{3}$$

where $i\Delta = i \times \Delta$ $(i = 1.....N)$, is the amount of time-shift incurred upon the source signal. The time shift $\Delta = \frac{1}{F_s} \ll T$, where T is the duration of the source signal $s(t)$ and F_s is the sampling frequency. As the source signal is generally unknown, several instances of the impulsive source signals are recorded at a high F_s, averaged, and used to construct the dictionary matrix. Let us denote the sampling frequency of the microphones as F_m where $F_m < F_s$ (as we are interested in low sampling rates). The received signal at microphone i is then given by

$$\mathbf{r}_i = \boldsymbol{\Psi}_i \boldsymbol{\alpha}_i + \boldsymbol{\omega}_i \tag{4}$$

where $\mathbf{r}_i$ is the received signal of length M, $\boldsymbol{\alpha}_i$ is the RIR of length N, and $\boldsymbol{\omega}_i$ is the additive white Gaussian noise of the same mean and covariance matrix as of $\boldsymbol{\omega}$. Due to sub-sampling, the length M of the received signal $\mathbf{r}_i$ is much smaller than the length N of the impulse response.

The matrix $\boldsymbol{\Psi}_i$ (of size $M \times N$) is a uniformly sub-sampled version of the dictionary matrix $\mathbf{S}$ where $M << N$ and the sub-sampling ratio[1] is $\frac{N}{M} = \frac{F_s}{F_m}$. As $M << N$, (4) is an under-determined system of equations and there is an infinite number of solutions satisfying this equation and thus is difficult to solve. Recently there has been an increased interest in solving such problems when the solution is known to be sparse. A wide variety of methods, categorized under Compressive Sensing (CS) [10], [11], have been proposed that utilize the sparsity information of the solution to solve this ill-posed problem. Specifically in [10], [11], it has been shown that $\boldsymbol{\alpha}_i$ can be reconstructed with high probability in polynomial time by using convex relaxation approaches. This is done by solving a relaxed ℓ_1 minimization problem using linear programming as follows

$$\hat{\boldsymbol{\alpha}}_i = \min_{\boldsymbol{\alpha}_i} \|\boldsymbol{\alpha}_i\|_1 \quad \text{s.t.} \quad \|\mathbf{r}_i - \boldsymbol{\Psi}_i \boldsymbol{\alpha}_i\|_2 \le \epsilon \tag{5}$$

For the above method to work, the matrix $\boldsymbol{\Psi}_i$ should be incoherent to the domain in which $\boldsymbol{\alpha}_i$ is sparse. However, this condition is not satisfied in the current formulation of (4). Moreover, while the received signal is sub-sampled, the dimension of the problem is too large to solve using convex relaxation software (complexity of $\mathcal{O}(M^2 N^{3/2})$ [12]). Thus, the recently proposed OC algorithm [7], [8] is utilized (that does not require the incoherence condition to work) to estimate the RIR $\boldsymbol{\alpha}_i$ from the low rate data at a relatively low complexity ($\mathcal{O}(M^2 N)$).

3. TIME DELAY ESTIMATION

The RIR is assumed to be a sparse signal and the recently proposed OC algorithm [7], [8] is employed to estimate it. The OC algorithm makes a collective use of the structure present in the sub-sampled $\boldsymbol{\Psi}_i$ matrix and the sparsity information of the RIR signal to obtain its minimum mean square error (MMSE) estimate given by

$$\hat{\boldsymbol{\alpha}}_i^{\mathrm{MMSE}} = \mathbb{E}[\boldsymbol{\alpha}_i|\mathbf{r}_i] = \sum_{\mathcal{J}} p(\mathcal{J}|\mathbf{r}_i)\mathbb{E}[\boldsymbol{\alpha}_i|\mathbf{r}_i, \mathcal{J}] \tag{6}$$

where $\mathcal{J}$ is the support (location of non-zero values) of $\boldsymbol{\alpha}_i$. The impulsive nature of the source signals considered in this work renders $\boldsymbol{\Psi}_i$ a block Toeplitz matrix structure.

[1] Note that it is necessary for the sub-sampling ratio to be less than T to avoid missing the source signal completely.

This structure enables the algorithm to construct orthogonal clusters around the most probable positions where the support of $\boldsymbol{\alpha}_i$ might be located. Thus, (6) can be calculated in a divide and conquer manner as follows (see [7] for details)

$$\hat{\boldsymbol{\alpha}_i}^{\mathrm{MMSE}} = \begin{bmatrix} \hat{\boldsymbol{\alpha}_i}^1 \\ \hat{\boldsymbol{\alpha}_i}^2 \\ \vdots \\ \hat{\boldsymbol{\alpha}_i}^P \end{bmatrix} = \begin{bmatrix} \mathbb{E}[\boldsymbol{\alpha}_i^1|\mathbf{r}_i] \\ \mathbb{E}[\boldsymbol{\alpha}_i^2|\mathbf{r}_i] \\ \vdots \\ \mathbb{E}[\boldsymbol{\alpha}_i^P|\mathbf{r}_i] \end{bmatrix}$$
$$= \begin{bmatrix} \sum_{\mathcal{J}^1} p(\mathcal{J}^1|\mathbf{r}_i)\mathbb{E}[\boldsymbol{\alpha}_i^1|\mathbf{r}_i, \mathcal{J}^1] \\ \sum_{\mathcal{J}^2} p(\mathcal{J}^2|\mathbf{r}_i)\mathbb{E}[\boldsymbol{\alpha}_i^2|\mathbf{r}_i, \mathcal{J}^2] \\ \vdots \\ \sum_{\mathcal{J}^P} p(\mathcal{J}^P|\mathbf{r}_i)\mathbb{E}[\boldsymbol{\alpha}_i^P|\mathbf{r}_i, \mathcal{J}^P] \end{bmatrix} \quad (7)$$

where $\boldsymbol{\alpha}_i{}^k$ and $\mathcal{J}^k$ is the RIR and its support corresponding to the k^{th} cluster.[2] Each cluster is then searched intelligently for the support of RIR and the structure of $\boldsymbol{\Psi}_i$ helps in reducing the complexity involved in it [7]. Figure 1 shows the flowchart of the proposed technique for time delay estimation that utilizes the OC algorithm.[3]

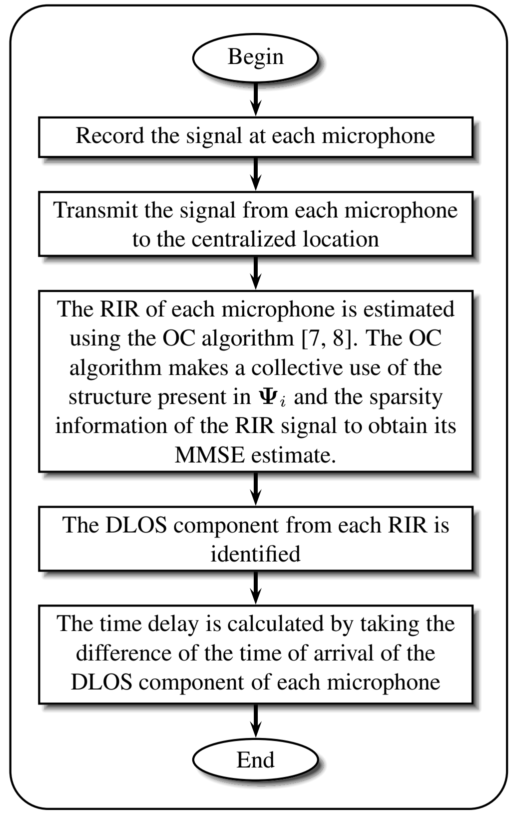

Fig. 1. Flowchart of the proposed time delay estimation technique

4. SIMULATION AND EXPERIMENTAL RESULTS

In this section, the numerical simulations conducted for a simulated reverberant environment are discussed followed by the experimental results.

4.1. Simulation Results

The performance of the proposed TDE method is analyzed in simulations by creating a virtual room environment. The impulse response of the channel between a source-microphone pair placed in a room is obtained using an image-source model as presented in [13]. A rectangular room is considered with dimensions $8 \times 6 \times 3$ meters ($x \times y \times z$). Sampling frequency is set to 10 KHz. Figure 2 and 3 show an instance of a simulated RIR of a low and high reverberant environment respectively. The estimate obtained using the OC algorithm is also shown with sub-sampling rate $= 4$.

OC algorithm is applied to estimate time delay for various sub-sampling rates. Two reverberant environments are considered; one with high reflection coefficients of the walls $[0.75\ 0.75\ 0.8\ 0.8\ 0.85\ 0.9]$ and the other with low reflection coefficients $[0.2\ 0.2\ 0.3\ 0.25\ 0.3\ 0.5]$. The simulation is run for 500 iterations and for each iteration the source-microphone position is considered random within the room boundaries. The RIR is generated using the image source model in [13]. The OC algorithm is applied for TDE and mean square error (MSE) in time delay is evaluated for different sub-sampling rates.

The performance is demonstrated in Figure 4 for two SNR values of the impulsive source, 30 dB and 40 dB respectively. From the figure, it can be observed that MSE in TDE is less for the room with low reflection coefficients. The figure also shows that for the 40 dB case, the MSE at low sub-sampling rates is quite small and it increases gradually for higher sub-sampling rates. For the 30 dB case, it can be seen that there is an increase in MSE for sub-sampling rate greater than 2. The increase in MSE of TDE at high sub-sampling rates is caused by less number of measurements of the received signal available to the OC algorithm for RIR reconstruction. Thus, in case of dense reverberation and low SNR values, the sub-sampling rate is set at a moderate level to ensure low MSE in TDE.

4.2. Experimental Results

Figure 5 shows the actual hardware setup for TDE in a hall room of dimensions $8 \times 6 \times 3$ meters. The microphones are secured with metallic stands placed 100 cm apart. The electret microphones are mounted on a printed circuit board (PCB) with appropriate electronics. Each PCB has a MAX 9814 low-noise amplifier (LNA) IC whose gain is set to 40 dB. With this gain, an $SNR \geq 30$ dB is obtained at the output of the LNAs which are connected to

[2] Here $k = 0, 1, \cdots, P$ where P is the total number of clusters (interested readers are directed to [7] and [8] for further details).

[3] As the RIR at each microphone is estimated independent of the other and the complexity involved in it is quite low, it can be estimated at each microphone instead of the centralized workstation using cheap processors at each microphone. The flowchart of the proposed method is presented here in the way as it was implemented during the hardware experiments.

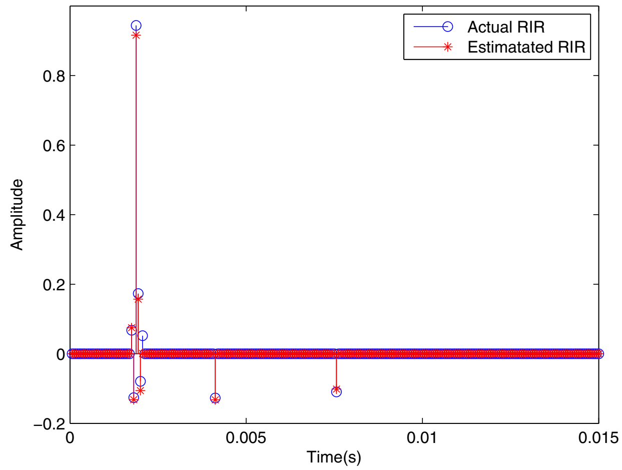

Fig. 2. Example of simulated RIR estimation (low reverberations).

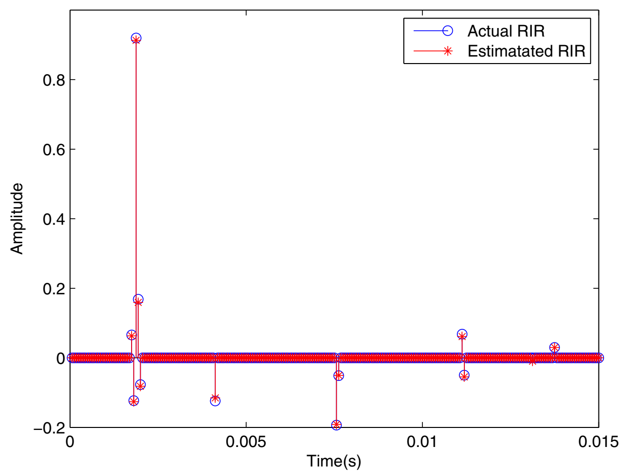

Fig. 3. Example of simulated RIR estimation (high reverberations).

a 16-bit, 8 channel data acquisition (DAQ) device via audio jacks and cables. The DAQ communicates with a PC through the data acquisition tool box within MATLAB. A toy gun was used as an impulsive source with a duration of approximately 10ms as shown in Figure 6. The dictionary matrix **S** (equation (3)) required for the RIR estimation using the OC algorithm is constructed by averaging several instances of the impulsive source signal at $F_s = 16$ KHz.

The OC algorithm based TDE method first estimates the sparse RIR. An instance of the estimated RIR for source-microphone pair placed in the room center is shown in Figure 7. The estimated RIR shows the direct path, early reflections along with few late reflections.

The real time functionality of the algorithm for TDE has been verified by placing the source at known locations around the microphones and acquiring the source signal at various sampling rates. Table 1 shows the time delays

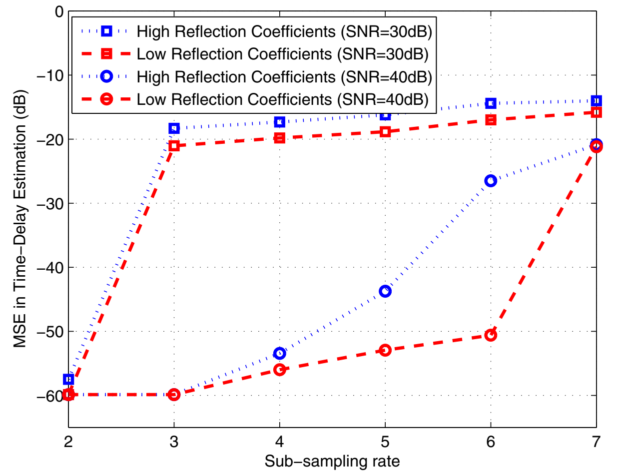

Fig. 4. MSE in TDE using the OC algorithm for different sub-sampling rates at SNR $= 30$ dB and 40 dB.

corresponding to three known source locations:

1. Case I: Source positioned at a point on the line that passes through the two microphones.
2. Case II: Source positioned close to microphone 1, on the vertex of an isosceles triangle formed by the microphones and the source.
3. Case III: Source positioned in the middle of the line joining the two microphones.

The estimated time delays using the CC and the OC algorithm for three sampling frequencies of 16 KHz, 8 KHz and 4 KHz are shown in Table 1. The table also shows the true time delays (calculated from the known source and microphone positions) and the run time of both TDE methods. In all cases of the source position, the proposed OC algorithm gives closer TDEs as compared to CC at low sampling rates in reverberant indoor environment. In addition, the fast RIR reconstruction using the OC algorithm gives it an advantage of less execution time to compute TDEs as compared to the CC method. This demonstrates the superior performance of the OC algorithm based TDE technique both in terms of the accuracy and the execution time for TDE at low sampling rates.

Note that in this scenario, the problem dimension (length of RIR) is $N = 16000$ and thus it is very difficult to solve it using the convex relaxation techniques [10], [11] or ESPRIT [9]. This highlights the advantage of the proposed technique for TDE.

5. CONCLUSIONS

An application of a novel sparse signal reconstruction algorithm has been presented that tackles the challenging task of TDE in a room reverberant environment. The proposed method utilizes the signal statistics, sparsity information, and problem structure for sparse RIR estimation

Table 1. Comparison of time delay estimates obtained using CC and the proposed technique based on OC algorithm.

	Freq. (KHz)	True TD (ms)	TDE CC (ms)	Run Time CC (sec)	TDE OC (ms)	Run Time OC (sec)
Case I	16	2.941	2.875	66	2.875	60
	8	2.941	3.125	17	2.875	1.3
	4	2.941	2.250	4.5	3.312	0.6
Case II	16	1.218	1.250	66	1.250	60
	8	1.218	0.875	17	1.312	1.3
	4	1.218	1.000	4.5	1.250	0.6
Case II	16	0.000	0.000	66	0.000	60
	8	0.000	0.250	17	0.000	1.3
	4	0.000	0.750	4.5	0.187	0.6

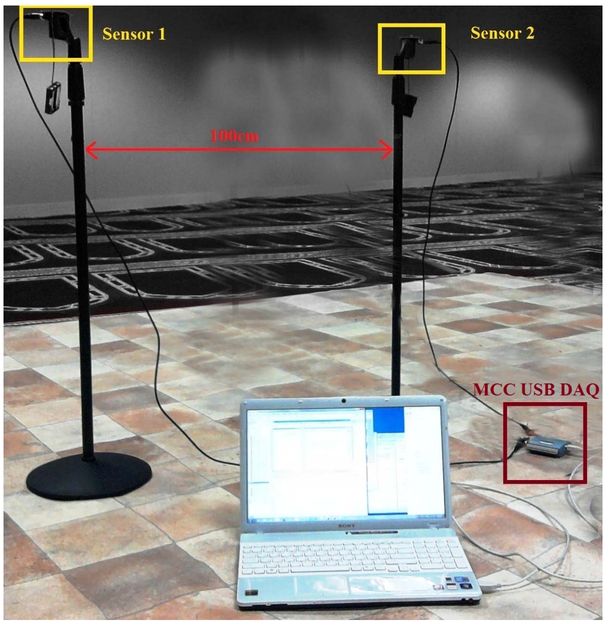

Fig. 5. Experimental setup for OC based TDE in a room environment.

from sub-sampled received signal. Although decreasing the number of measurements (low sampling rate) effects the accuracy of RIR estimation, this does not decrease the time resolution which is essential for indoor source localization. The TDE between the pair of microphones is found by identifying the DLOS signals from the estimated RIRs. The performance of the proposed algorithm has been analyzed in simulations and successfully verified experimentally. The MSE in time delays are found to be very less even for high sub-sampling rates. The proposed TDE method reduces the hardware and computational complexity and provides a relief to the communication link between microphones and the centralized workstation. In future, the new scheme can be used as an efficient pre-processor for time delay based localization applications.

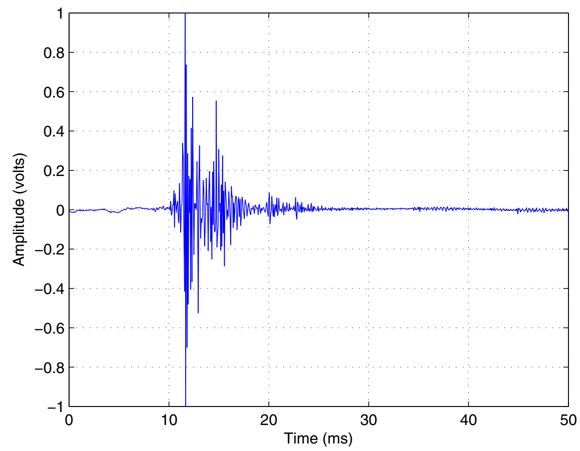

Fig. 6. Impulsive acoustic signal from a toy gun.

6. REFERENCES

[1] 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, 1976.

[2] Bouzid, M. Omar and Tian, Gui Yun and Neasham, Jeff and Sharif, Bayan, "Investigation of sampling frequency requirements for acoustic source localisation using wireless sensor networks," *Journal of Applied Acoustics*, vol. 1, no. 121, pp. 269-277, 2011.

[3] R. Ratnam, D.L. Jones, B.C. Wheeler, W.D. O'Brien Jr., C.R. Lansing, and A.S. Feng, "Blind estimation of reverberation time," *Journal of the Acoustical Society of America*, vol. 114, no. 5, pp. 2877-2892, Nov. 2003.

[4] H. W. Lollmann and P. Vary, "Estimation of the Reverberation Time in Noisy Environments," *in Proceedings of International Workshop on Acoustic Echo and Noise Control*, USA, Sep. 2008.

[5] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: an overview," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 26503, pp. 1-19.

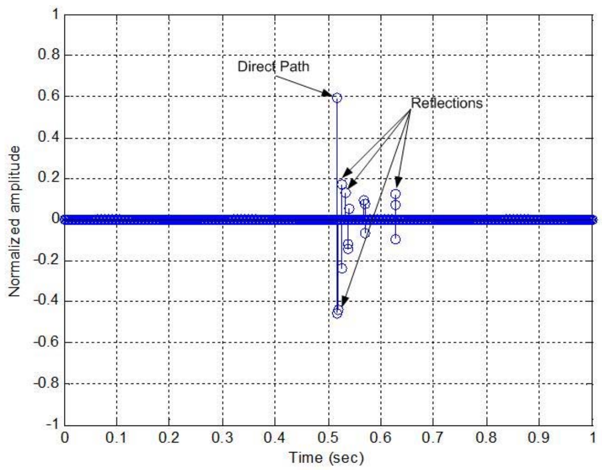

Fig. 7. Example of a RIR estimate obtained experimentally using OC algorithm.

[6] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoutic source localization," *Journal of Acoustical Society of America*, vol. 107, no. 1, pp. 384-391, 2000.

[7] A. A. Quadeer, S. F. Ahmed, and T. Y. Al-Naffouri, "Structure Based Bayesian Sparse Reconstruction Using Non-Gaussian Prior," *Allerton Conference on Communications, Control and Computing*, USA, Sep. 2011.

[8] T. Y. Al-Naffouri, A. A. Quadeer, and G. Caire, "Impulsive Noise Estimation and Cancellation in DSL using Orthogonal Clustering," *IEEE International Symposium on Information Theory*, Russia, 2011.

[9] K. Gedalyahu and Y. C. Eldar, "Time delay estimation: Compressed sensing over an infinite union of subspaces," *IEEE International Conference on Acoustic, Speech and Signal Processing*, pp. 3902-3905, USA, Mar. 2010.

[10] E. Candes, J. Romberg and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math,* vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[11] E. J. Candes and P. Randall, "Highly robust error correction by convex programming," *IEEE Trans. Inform. Theory*, vol. 54, no. 7, pp. 2829–2840, Jun. 2008.

[12] I. E. Nesterov, A. Nemirovskii, and Y. Nesterov, *Interior-Point Polynomial Algorithms in Convex Programming.* SIAM, 1994.

[13] E. Lehmann and A. Johanssons, "Prediction of energy decay in room impulse responses simulated with an image-source model," *Journal of Acoustic Society of America*, vol. 1, no. 121, pp. 269-277, 2008.