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 Department,
∗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
from each of the estimated RIR and their difference yields
the desired TDE. Low rate sampling relaxes the commu-
nication link between the microphones and the centralized
workstation and at the same time decreases the computa-
tion 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 com-
pressed 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 sep-
parated by an acoustic space can be described by a linear
time invariant system. Given that the acoustic space is ex-
cited by a known excitation signal \( s(t) \), the signal received
at the microphone can be expressed as,

\[
 r(t) = \sum_{i=0}^{L-1} \alpha_i s(t - \tau_i) + \omega(t)
\]

where \( L \) is the number of paths capturing significant mul-
tipath energy. The impulsive response estimation is ac-
ually the estimation of the parameters \( \alpha_i \) and \( \tau_i \), the scaling
magnitude factor and the time shift respectively and \( \omega(t) \)
is the zero mean additive white Gaussian noise. With dis-
crete time signals, the matrix form of (1) can be written as,

\[
 r = S \alpha + \omega
\]

where \( r \) and \( \alpha \) are length \( N \) discrete time, received and
RIR signals respectively while \( \omega \) is the additive white Gaus-

sian noise of length \( N \) with zero-mean and covariance ma-
rix \( C_\omega = \sigma_\omega^2 I \). The \( N \times N \) matrix \( S \) is called the dic-
tionary matrix which is formed by the \( N \) discretized and
delayed versions of the basic source signal \( s(t) \)

\[
 S = \begin{bmatrix}
 s(0 - \Delta) & \ldots & s(0 - N\Delta) \\
 s(1 - \Delta) & \ldots & \cdot \\
 \cdot & \ldots & \cdot \\
 s((N-1) - \Delta) & \ldots & s((N-1) - N\Delta)
\end{bmatrix}
\]

where \( i\Delta = i \times \Delta (i = 1, \ldots, N) \), is the amount of time-
shift incurred upon the source signal. The time shift \( \Delta = \frac{F_s}{F_r} \ll T \), where \( T \) is the duration of the source signal
\( s(t) \) and \( F_r \) is the sampling frequency. As the source sig-
nal is generally unknown, several instances of the impul-
sive 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
\( F_m < F_r \) (as we are interested in low sampling rates).
The received signal at microphone \( i \) is then given by

\[
 r_i = \Psi_i \alpha_i + \omega_i
\]

where \( r_i \) is the received signal of length \( M \), \( \alpha_i \) is the RIR
of length \( N \), and \( \omega_i \) is the additive white Gaussian noise
of the same mean and covariance matrix as \( \omega \). Due to
sub-sampling, the length \( M \) of the received signal \( r_i \) is
much smaller than the length \( N \) of the impulse response.

The matrix \( \Psi_i \) (of size \( M \times N \)) is a uniformly sub-
sampled version of the dictionary matrix \( S \) where \( M \ll N \) and
the sub-sampling ratio\(^1\) is \( \frac{N}{M} \). As \( M \ll N \), (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 so-
lution 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 \( \alpha_i \) can be reconstructed
with high probability in polynomial time by using convex
relaxation approaches. This is done by solving a relaxed \( \ell_1 \)
minimization problem using linear programming as fol-

\[
 \alpha_i = \underset{\alpha_i}{\text{arg min}} \| \alpha_i \|_1 \\
\text{s.t.} \| r_i - \Psi_i \alpha_i \|_2 \leq \epsilon
\]

For the above method to work, the matrix \( \Psi_i \) should be
incoherent to the domain in which \( \alpha_i \) 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 \( O(M^2N^3/2) \)
[12]). Thus, the recently proposed OC algorithm [7], [8]
is utilized (that does not require the incoherence condition
to work) to estimate the RIR \( \alpha_i \) from the low rate data at
a relatively low complexity (\( O(M^2N) \)).

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 \( \Psi_i \) matrix and the sparsity in-
formation of the RIR signal to obtain its minimum mean
square error (MMSE) estimate given by

\[
 \alpha_i^{\text{MMSE}} = \mathbb{E}[\alpha_i | r_i] = \sum_{\mathcal{J}} p(\mathcal{J} | r_i) \mathbb{E}[\alpha_i | r_i, \mathcal{J}]
\]

where \( \mathcal{J} \) is the support (location of non-zero values) of
\( \alpha_i \). The impulsive nature of the source signals consid-
ered in this work renders \( \Psi_i \) 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 $\alpha_i$ might be located. Thus, (6) can be calculated in a divide and conquer manner as follows (see [7] for details)

$$
\hat{\alpha}_{i}^{\text{MMSE}} = \begin{bmatrix}
\hat{\alpha}_{i,1} \\
\hat{\alpha}_{i,2} \\
\vdots \\
\hat{\alpha}_{i,P}
\end{bmatrix} = \begin{bmatrix}
\mathbb{E}[\alpha_i^1 | r_i] \\
\mathbb{E}[\alpha_i^2 | r_i] \\
\vdots \\
\mathbb{E}[\alpha_i^P | r_i]
\end{bmatrix} = \\
\sum_{J^1} p(J^1 | r_i) \mathbb{E}[\alpha_i^1 | r_i, J^1] \\
\sum_{J^2} p(J^2 | r_i) \mathbb{E}[\alpha_i^2 | r_i, J^2] \\
\vdots \\
\sum_{J^P} p(J^P | r_i) \mathbb{E}[\alpha_i^P | r_i, J^P]
$$

(7)

where $\alpha_i^k$ and $J^k$ is the RIR and its support corresponding to the $k^{th}$ cluster. Each cluster is then searched intelligently for the support of RIR and the structure of $\Psi_i$ 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

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]. The 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 0.85 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 electroret 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

![Flowchart of the proposed time delay estimation technique](image-url)
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 10 ms 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 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.

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

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

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

6. REFERENCES


![Graph](image-source-model.png)

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


