FIRST ORDER ECHO BASED ROOM SHAPE RECOVERY USING A SINGLE MOBILE DEVICE

Tiexing Wang, Fangrong Peng and Biao Chen

Department of EECS, Syracuse University, Syraucse, NY, 13244, USA
email:{twang17, fapeng, bichen}@syr.edu

ABSTRACT

This paper considers the problem of constructing 2-D room shape. A mobile device with co-located microphone and loudspeaker is used and the distances between consecutive measurement points are assumed to be known. The uniqueness of the mapping between the first-order echoes and the room geometry is guaranteed for any convex polygons. A practical algorithm for room reconstruction in the presence of noise and higher order echoes is proposed. Experimental results are provided to demonstrate the effectiveness of our approach.

Index Terms— 2-D room shape recovery, acoustic sensor, room impulse response.

1. INTRODUCTION

This paper studies the problem of constructing 2-D room shape using a single mobile acoustic sensor. The mobile device with co-located microphone and loudspeaker is employed to emit acoustic pulses and collect the echoes inside the room. Our objective is to use only first order acoustic echoes measured at multiple locations to reconstruct 2-D room geometry.

Knowledge of indoor room geometry is beneficial to a number of applications, e.g. source localization[1, 2]. Autonomously acquiring the indoor room shape has been an active field of research in recent years [3]-[8]. In [3] the authors proved that a single loudspeaker and a microphone array with at least four microphones are sufficient to reconstruct a 3-D room given the location of the microphone array. In [5], a room with four walls and a ceiling was considered. A uniform circular microphone array with a loudspeaker placed at the center was used to estimate the room shape. A least square method with l_1-norm regularization was applied to map the measured impulse responses of a 3-D shoebox room. In [6], a method to reconstruct 3-D rooms was proposed which did not require prior information on the room shape, the number of walls, or the order of the reflections, but the topology of the devices was assumed to be known. In [7], a single fixed loudspeaker and a moving microphone were used to reconstruct 2-D shape room. The essential idea was to determine the common tangent line of the ellipses whose foci were at the locations of the microphone and the loudspeaker.

Another interesting work, which is related to the present one, is described in [4]. It utilized a static co-located loudspeaker and microphone to reconstruct the room in the 2-D case by using both first and second order echoes. However, our experiments show that it is difficult to collect the entire set of the first *and* second order echoes due to measurement noise. Even if the entire set of the first and second order echoes is contained in the received signal, it is difficult to detect them due to the limitation of acoustic bandwidth and that the density of pulses increases rapidly as the delay increases[4].

In our previous work [8], a mobile co-located loudspeaker and microphone was used to retrieve 2-D polygonal room shape, where only first-order echoes measured at three non-collinear points were required. However, it was established in [8] that first order echoes alone, in the absence of any information of the measurements points, is inadequate to reconstruct an important class of polygons, namely parallelograms. Thus some additional topological information is required if one is to reconstruct 2-D room shapes using only first order echoes with a single mobile sensor.

Given the full geometry of the measurement points, i.e., the distance between each pair, it is straightforward to establish that first order echoes are sufficient to recover any convex polygonal shapes. This however is a very strong assumption that often requires human intervention, e.g., actually measuring distances between measurement points. However, many mobile devices are capable of measuring its own path length when moving from one point to another. This paper investigates the possibility of reconstructing 2-D room shape with path length of consecutive points measured by the mobile device. This weaker assumption, compared with the knowledge of complete geometry of measurement points, makes it feasible to achieve autonomous room shape recovery.

The present work establishes that the above approach is indeed feasible. That is, given the knowledge of path lengths between consecutive measurement points, one can recover arbitrary convex room by using only first order echoes collected at three non-collinear points in a room. Algorithmic procedure that handles measurement noise and the presence of high order echoes is also proposed and experimental results are presented to validate the approach.

The rest of the paper is organized as follows. Section II introduces the indoor propagation model of acoustic signals and the geometric relations used in subsequent sections. Theoretical guarantee of successful recovery is provided in Section III along with an algorithm that handles the presence of measurement noise and higher order echoes. Experiment results are provided in Section IV to verify our proposed algorithm followed by conclusion in Section V.

2. SYSTEM MODEL

2.1. Room Impulse Response Model

Acoustic signal propagation from a loudspeaker to a microphone in a room can be described by the room impulse response (RIR). The RIR encompasses both line-of-sight (LOS) and reflected components. In practice, if the microphone and loudspeaker are much closer to each other compared to the distance between the device and the walls, we say it is a co-located device. For a co-located device at O_j, the RIR is

$$h^{(j)}(t) = \sum_i \alpha_i^{(j)} \delta(t - \tau_i^{(j)}), \tag{1}$$

where $\alpha_i^{(j)}$'s and $\tau_i^{(j)}$'s are path gains and delays from the transmitter to the receiver, respectively. Since higher order reflective paths typically have much weaker power compared with the lower order ones, (1) can be approximately expressed by the first $N_j + 1$ components including LOS and N_j reflective paths:

$$h^{(j)}(t) \approx \sum_{i=0}^{N_j} \alpha_i^{(j)} \delta(t - \tau_i^{(j)}),$$

where we assume that the N_j reflective paths contains all first order echoes and higher order ones that are detectable.

If the transmitted signal has an auto-correlation function resembling a delta function, then the time difference of arrival (TDOA) among paths can be obtained from the correlator output of the received signal. Since the loudspeaker and microphone are co-located, τ_0, which corresponds to the delay of the LOS path, is close to zero. Define

$$\mathbf{r}_j = \left\{ \frac{(\tau_i^{(j)} - \tau_0^{(j)})c}{2} \right\}_{i=1}^{N_j},$$

where c is the speed of sound. Then $\mathbf{r}_j$ contains all the distances between the device and the walls. Hence synchronization between loudspeakers and microphones is not required for co-located device if only the distances between measurement point and the walls are of interest.

2.2. Geometry

Consider a convex planar K-polygon. As shown in Fig.1, a mobile device with co-located microphone and loudspeaker emits pulses and receives echoes at $\{O_j\}_{j=1}^3$. Denote r_{1i}, r_{2i} and r_{3i} as the distances between O_1, O_2, O_3 and the ith wall, respectively. Without loss of generality, we assume that O_1 is the origin, O_2 lies on the x-axis, and O_3 lies above the x-axis.

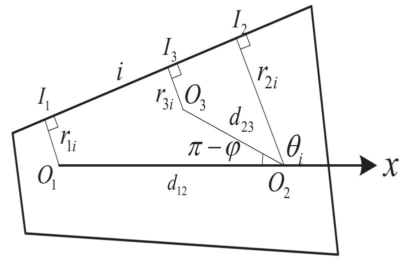

Fig. 1. A mobile device is employed to measure the geometry of a room. The mobile device collects echoes at O_1, O_2 and O_3 successively. The distances between the consecutive measurement points are d_{12} and d_{23}

Let $\varphi = (\pi - \angle O_1O_2O_3) \in (0, \pi)$ and the length of O_1O_2 and O_2O_3 be denoted by d_{12} and d_{23}, respectively.[1]

From Fig. 1, it is straightforward to show that

$$(r_{2i} - r_{1i}) + d_{12}\cos\theta_i = 0 \tag{2}$$

and

$$d_{23}\cos(\theta_i - \varphi) + (r_{3i} - r_{2i}) = 0. \tag{3}$$

If however the r_{ji}'s contain higher order echoes, with probability 1, (2) and (3) do not hold simultaneously.

Clearly given d_{12} and d_{23} and if φ can also be estimated from the measurement data, then the geometry of the measurement points can be recovered. In this case, first order echoes are sufficient to recover room geometry.

3. RECOVERY WITH KNOWN DISTANCES AND UNKNOWN PATH DIRECTION

If for all j's, the one-to-one mapping $f_j : \{r_{ji}\}_{i=1}^K \mapsto \mathbf{r}_j$ is known, then $\{\theta_i\}_{i=1}^K$ can be obtained by (2) and the room shape is determined. However since echoes may arrive in different orders at different O_j's and $\mathbf{r}_j$ may contain higher order echoes if $N_j > K$, f_j is unknown. Then θ_i's are also unknown. Therefore we need a way to both rule out higher order echoes and find the correct combination of the first order echoes. We can then estimate θ_i's and the room shape.

Define $\alpha_{ii'} = -\frac{r_{2i} - r_{1i'}}{d_{12}}$ and $\beta_{ii'} = -\frac{r_{3i'} - r_{2i}}{d_{23}}$. For simplicity we denote α_{ii} and β_{ii} by α_i and β_i, respectively. From (2) and (3), we have

$$\theta_i = \pm \arccos \alpha_i \quad \text{and} \quad \theta_i - \varphi = \pm \arccos \beta_i, \tag{4}$$

Thus, there are four possible sign combinations for a given i,

$$\theta_i = \arccos \alpha_i \quad \text{and} \quad \theta_i - \varphi = \arccos \beta_i \tag{5}$$

$$\theta_i = \arccos \alpha_i \quad \text{and} \quad \theta_i - \varphi = -\arccos \beta_i \tag{6}$$

$$\theta_i = -\arccos \alpha_i \quad \text{and} \quad \theta_i - \varphi = \arccos \beta_i \tag{7}$$

$$\theta_i = -\arccos \alpha_i \quad \text{and} \quad \theta_i - \varphi = -\arccos \beta_i. \tag{8}$$

[1] If $\pi \in (0, 2\pi)$, i.e. we do not have control of where to place O_3, then the reconstruction is subject to reflection ambiguity (c.f. Theorem 3.3).

Definition 3.1. *Given a room $\mathcal{R}$ and a location O, we say O is feasible if the co-located device at O can receive all the first order echoes of a signal emitted at O.*

Lemma 3.1. *Suppose O_1, O_2 and O_3 are feasible and not collinear. Given the correct echo combination, with probability 1, there exist exactly two sign combinations such that (2) and (3) hold simultaneously for all i if φ and the direction of both $\overrightarrow{O_1O_2}$ and $\overrightarrow{O_2O_3}$ are randomly chosen. The two possible sign combinations have opposite signs for φ and all θ_i's and correspond to reflection of each other.*

Proof. Assume that the ground truth of the polygon is (5) for all $i \in \{1, \ldots, K\}$. Note that (5) implies that (8) holds for $\theta_i' = -\theta_i$ and $\varphi' = -\varphi < 0$ for all i, which is the reflection of the room.

Suppose multiple sign combinations hold for a wall. Without loss of generality, let $i = 1$. From (5) we have

$$\varphi = \arccos \alpha_1 - \arccos \beta_1. \tag{9}$$

Assume that one of the following equations also holds,

$$\varphi = -\arccos \alpha_1 - \arccos \beta_1 \tag{10}$$

$$\varphi = \arccos \alpha_1 + \arccos \beta_1 \tag{11}$$

$$\varphi = -\arccos \alpha_1 + \arccos \beta_1. \tag{12}$$

Then we have the following three cases:

1. If (9) and (10) hold, we must have $\theta_1 = 0$ which implies that O_1O_2 is perpendicular to the first wall, and $\varphi = -\arccos \beta_1$.
2. If (9) and (11) hold, we must have $\arccos \beta_1 = 0$, which implies that O_2O_3 is perpendicular to the first wall.
3. If (9) and (12) hold, we must have $\varphi = 0$, which contradict with the assumption that O_1, O_2 and O_3 are not collinear.

With probability 1, the first two cases do not occur since both φ and directions of $\overrightarrow{O_1O_2}$ and $\overrightarrow{O_2O_3}$ are randomly chosen.

If a subset of (6)-(8) holds for i and i' simultaneously, then we must have $(\theta_i, \theta_{i'}) \in \{\theta_i = 0, \theta_i = \varphi, \varphi = 0\} \times \{\theta_{i'} = 0, \theta_{i'} = \varphi, \varphi = 0\}$, which again, do not occur due to randomly chosen measurement points. Similarly, it can be shown that for more than 2 walls, (5) would imply none of (6)-(8) holds for all walls. □

Lemma 3.2. *Given incorrect echo combinations, with probability 1, there exists no solution to (2) and (3).*

Proof. We illustrate the proof by considering only the case of $K = 4$. The result can be easily extended to $K = 3$ and $K > 4$.

We assume that the ground truth is (5) for all i. First consider parallelograms. The distances between O_j $(j = 1, 2, 3)$ and the four walls satisfy

$$r_{11} + r_{12} = r_{21} + r_{22} = r_{31} + r_{32} = a \tag{13}$$

and

$$r_{13} + r_{14} = r_{23} + r_{24} = r_{33} + r_{34} = b. \tag{14}$$

We can see that for certain echo combinations, pairs of $\{\alpha_{ii'}, \beta_{ii'}\}$ $(i, i' \in \{1, 2, 3, 4\})$ are related to each other. Consider for example the echo combination resulting in $\{\alpha_{12}, \alpha_{21}, \alpha_{34}, \alpha_{43}\}$ and $\{\beta_{12}, \beta_{21}, \beta_{34}, \beta_{43}\}$. Since $\alpha_{12} + \alpha_{21} = 0$, $\alpha_{34} + \alpha_{43} = 0$, $\beta_{12} + \beta_{21} = 0$ and $\beta_{34} + \beta_{43} = 0$, we have

$$\arccos(\alpha_{21}) = \pi \pm \arccos(\alpha_{12})$$

$$\arccos(\alpha_{43}) = \pi \pm \arccos(\alpha_{34})$$

$$\arccos(\beta_{21}) = \pi \pm \arccos(\beta_{12})$$

$$\arccos(\beta_{43}) = \pi \pm \arccos(\beta_{34}).$$

Thus (4) reduces to two equations

$$\varphi = \pm \arccos(\alpha_{12}) \pm \arccos(\beta_{12})$$

$$\varphi = \pm \arccos(\alpha_{34}) \pm \arccos(\beta_{34}).$$

With probability 1, these two equations do not hold simultaneously as α_{12}, β_{12} are independent of α_{34}, β_{34} due to randomly chosen measurement points. Other possible cases of incorrect combination always have at least two equations with independent choice of α and β. Hence no solution can be found for those instances.

Suppose a combination of echoes is chosen such that we have $\alpha_{ii'}$ and $\beta_{ii''}$ $(i \neq i', i \neq i'')$. For rooms with no more than one pair of parallel walls, almost surely no echo combination other than the correct one can make (5) holds for all i. This is because for those rooms, at least one of (13) and (14) does not hold. Thus some $\alpha_{ii'}$'s and $\beta_{ii''}$'s are not related since $r_{1i'}$, r_{2i} and $r_{3i''}$ are randomly chosen from $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$, respectively.

Therefore only the correct combination of first order echoes satisfies (5) for all walls. □

Note that distances between the device and both the ceiling and the floor can be ruled out by Lemma 3.2.

Given Lemma 3.1 and Lemma 3.2, we have the following result on the identifiability of any convex polygonal room by using only first-order echoes.

Theorem 3.3. *One can recover, with probability 1, any convex planar K-polygon subject to reflection ambiguity, by using the first order echoes received at three random points in the feasible region, with known d_{12} and d_{23} and unknown $\varphi \in (0, 2\pi)$.*

Remark: The room shape is subject to reflection ambiguity for $\varphi \in (0, 2\pi)$. If, however, we can limit $\varphi \in (0, \pi)$, the recovered shape is unique, i.e. not subject to ambiguity.

In the presence of noise, however, $\mathbf{r}_j$ is subject to measurement errors. Hence φ solved from (4) for different i's are not identical. A straightforward practical algorithm that handles the measurement errors is given below:

1. For K from 3 to N, choose K entries from $\hat{\mathbf{r}}_j$ ($j = 1, 2, 3$), where $N = \min\{N_1, N_2, N_3\}$.

2. For a given K, exhaust all possible echo combinations and compute $\varphi_i^k = \pm \arccos \alpha_i \pm \arccos \beta_i$ for each i, k and different sign combination, where $i = 1, \ldots, K$ and $k = 1, \ldots, \binom{N}{K}(K!)^2$.

3. Choose the echo and sign combination with minimum variance of φ_i^k for a given K. Then choose the largest K and the corresponding echo and sign combination with the variance less than the threshold. (variance criterion)

4. Estimate θ_i's using the obtained combination of echoes and reconstruct the polygon.

The following lemma states that the knowledge of both d_{12} and d_{23} is necessary for reconstructing any convex polygons by first order echoes.

Lemma 3.4. *If either d_{12} or d_{23} is missing, then 1) parallelogram can not be reconstructed. 2) Non-parallelogram can be reconstructed subject to reflection ambiguity.*

The proof of the part 1) is to construct a counterexample while 2) can be established in a manner similar to that of Lemma 3.2 in Section III.

4. EXPERIMENTAL RESULT

4.1. Experiment Setup

We use a laptop as a microphone and a HTC M8 phone as our loudspeaker. As the loudspeaker of the cell phone is not omnidirectional and power limited, we place the speaker of the cell phone towards each wall to ensure the corresponding first order echo is strong enough. Note that the loudspeaker will record both first order echoes and some higher order ones.

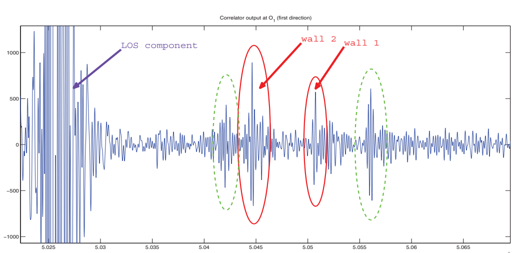

Fig. 2. Correlator output at O_1 towards the first wall. Peaks with solid ellipses correspond to walls. Peaks with dash ellipses correspond to either noise or higher order echoes

4.2. Signal Type

A chirp signal linearly sweeping from 30Hz to 8kHz is emitted by the cell phone. The sample rate at the receiver is 96kHz. It has been shown in [9, 10] that if the input chirp signal is correlated with its windowed version, the output may resemble a delta function. Our simulation and experiment results show that the candidate distances are obtained by correlating the received signals with its triangularly windowed version outperforms the correlator output using the original one. Fig. 2 is a sample path of the correlator output collected in the room where this experiment is conducted.

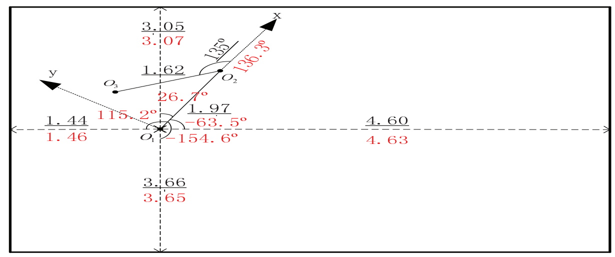

Fig. 3. Comparison between the ground truth (black underlined) and experiment result (red)

4.3. Room Shape Reconstruction

The proposed approach is verified by experiment in which d_{12} and d_{23} are measured by tape measure. Even if some elements of $\mathbf{r}_j$ have measurement errors up to 10cm, the room can be recovered with negligible error by the proposed algorithm if the subset of $\mathbf{r}_j$ corresponding to first order echoes are known. In the presence of higher order echoes, the proposed algorithm performs poorly when the variance criterion is the only criterion used to determine the correct combination of echoes. Since most rooms are regular, we add a heuristic constraint: all the angles of two adjacent walls are between $30°$ and $150°$. An interesting phenomenon is that sometimes the proposed algorithm is unable to provide the correct room shape, but the estimate of φ is always close to the true value. Therefore, one can use the algorithm in Section III to obtain φ and then reconstruct the room shape independently with full knowledge of the geometry information of the measurement points. The comparison between the reconstruction result and the ground truth is illustrated in Fig. 3.

5. CONCLUSION

This work makes progress in room shape reconstruction using only first order echoes. Specifically, we established that given the distances between consecutive measurement points, any 2-D convex polygon can be reconstructed. In the presence of noise, a simple algorithm is devised that is effective in recovering the room shape even in the presence of higher order echoes.

6. REFERENCES

[1] R. Parhizkar, I. Dokmanic, and M. Vetterli, "Single-channel indoor microphone localization," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 1434–1438.

[2] O. Oçal, I. Dokmanic, and M. Vetterli, "Source localization and tracking in non-convex rooms," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 1429–1433.

[3] Ivan Dokmanić, Reza Parhizkar, Andreas Walther, Yue M Lu, and Martin Vetterli, "Acoustic echoes reveal room shape," *Proceedings of the National Academy of Sciences*, vol. 110, no. 30, pp. 12186–12191, 2013.

[4] I. Dokmanić, Y.M. Lu, and M. Vetterli, "Can one hear the shape of a room: The 2-d polygonal case," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 321–324.

[5] D. Ba, F. Ribeiro, Cha Zhang, and D. Florencio, "L1 regularized room modeling with compact microphone arrays," in *Proc. IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, March 2010, pp. 157–160.

[6] S. Tervo and T. Tossavainen, "3d room geometry estimation from measured impulse responses," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 513–516.

[7] F. Antonacci, J. Filos, M.R.P. Thomas, E.A.P. Habets, A. Sarti, P.A. Naylor, and S. Tubaro, "Inference of room geometry from acoustic impulse responses," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 10, pp. 2683–2695, Dec 2012.

[8] F. Peng, T. Wang, and B. Chen, "Room shape reconstruction with a single mobile acoustic sensor," to appear in *Proc. IEEE International Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2015.

[9] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*. Audio Engineering Society, 2000.

[10] G. Stan, J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *Journal of the Audio Engineering Society*, vol. 50, no. 4, pp. 249–262, 2002.