Utilizing Time-of-Flight LIDARs For Spatial Audio Processing

Kanad Sarkar, Ryan M. Corey, Andrew C. Singer

2021 55th Asilomar Conference on Signals, Systems, and Computers | 978-1-6654-5828-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/IEEECONF53345.2021.9723358

***Abstract*— Methods for sensing acoustic displacement of a surface using optical means have included high speed video imagery, laser Doppler vibrometry, and sound-induced optical intensity. Methods for measuring sound from different points in space with a microphone array allow us to localize a source and focus audio recovery at specific points in space. When we combine these two concepts, we could theoretically have an array of optical sensors on a single device to perform single device spatial audio processing. Rotating LIDARs have become prevalent for use in rapid volumetric mapping and have the potential for acoustic recovery, allowing for simultaneous visual and acoustic mapping with one device. An ideal rotating LIDAR has the potential to simultaneously sense acoustic energy at numerous locations throughout an acoustic scene and may enable localization through processing the synthetically-induced array across the environment. We show the parameters that a rotating LIDAR would require for acoustic source localization. We also localize an acoustic source using a high precision light distance sensor.**

I. INTRODUCTION

Time-of-Flight Light Detection and Ranging (ToF LIDAR) devices measures distances based on the time it takes for an emitted laser to be reflected back. When placed on a rotating chassis, these LIDAR devices can sample points across space to construct a model of the environment [1]. Given sufficient developments in LIDAR technology, we show that the ToF LIDAR can be capable of two goals.

A. Utilzing a LIDAR as an optical microphone

Optical microphones recover audio with optical sensors through measuring the acoustic vibrations off of various objects. Most research on optical microphones deals with eavesdropping through reading the vibrations off of glass from afar. Laser microphones currently use a form of laser Doppler vibrometry (LDV), which recovers sound based on the Doppler frequency shift of a reflected laser [2]. Another optical microphone is the Lamphone, which uses a light sensor instead of an emitted laser to measure changes in brightness related to the acoustic vibration of a hanging bulb [3].

A ToF LIDAR currently does not have the resolution to recover audio; in this paper, we provide an estimate on the resolution required. We also use a similar sensor to experimentally show that we can recover audio given a sufficiently high resolution in Chapter II.

B. Single Device Spatial Audio Processing

Microphone arrays record the differences in how sound is received across space. Setting up these microphones can quickly become cumbersome, and if we had an optical sensor that could read the displacements off of existing objects in the room, then it would be easier to acquire spatial audio data. This idea was discussed briefly in Davis [4], where they speculated that a high resolution camera could simultaneously look at different objects. However, audio acquisition is slow using a camera, and a more suitable device for measuring specific points in space is a rotating ToF LIDAR. A sweep from a high resolution and high rotational frequency ToF LIDAR can obtain multi-modal spatial information, which would be useful for indoor modeling with virtual reality as well as simultaneous scene and audio capture.

We conducted an experiment for spatial audio processing with an optical microphone similar to a ToF LIDAR in Chapter III.

II. AUDIO RECOVERY WITH LIDAR

An ideal LIDAR device can recover audio from a single point in space through measuring the distance from a vibrating object over time. This process differs from current laser Doppler vibrometer (LDV) microphones, which measures acoustic vibrations through the Doppler shift of the reflected laser [2]. While LDV microphones are currently more suitable for real-time audio recovery than ToF methods, they are expensive and cannot be placed on a rotating chassis for array processing. A Time-of-Flight LIDAR device has a lower device cost and a high sampling rate, but currently lacks the distance resolution to directly recover audio. We show that given improvements of the distance resolution, we can recover audio with a Time-of-Flight LIDAR.

A. The Acoustic displacement of a 2D membrane

Audio recovery with a ToF LIDAR will depend on both the resolution of the sensor and the material's displacement with sound. While we don't focus on materials in this paper, suitable materials are likely to be modelled as thin membranes, and understanding how these membranes vibrate can give us insight on what it takes to recover audio.

The physical model for acoustic displacement from a 2D membrane requires knowledge of the shape and material of the membrane. While it is difficult to accurately predict membrane displacement, there are general relationships we can gain from the equations that model it.

Given a 2D membrane $\mathcal{S}$, the displacement $y(x,z,t) = \Psi(x,z)e^{j\omega t}$ at position (x,z) of $\mathcal{S}$ is the solution to the Helmholtz equation.

$$\nabla^2\Psi + k^2\Psi = 0$$

This solution for Ψ is dependant on the shape and boundary conditions on the membrane. A well-studied membrane shape is the circular diaphragm of a microphone, and we

can gain general insights through the equation modeling the displacement at the center of a microphone.

Given the forced vibration $f = e^{j\omega t}$ on a circular membrane of radius a, the displacement at the center is defined in Section 4.8 of [5], as

$$y_0(t) = \psi e^{j\omega t}$$
$$\psi = (\frac{P}{\mathcal{T}\mathbf{k}^2})[\frac{1}{J_0(\mathbf{k}a)} - 1]$$
$$\mathbf{k} \approx k(1 - j\frac{\beta}{\omega}),$$

where $\mathcal{T}$ is the tension of $\mathcal{S}$ and is proportional to the surface density, k is the angular wavenumber, and J_0 is the zeroth-order Bessel function. If our frequency is below the lowest resonance of the Bessel function, we can use the real part of $\mathbf{k}$ for a close approximation without knowledge of the damping coefficient.

The main insight we gain from this equation is that as our audio increases in frequency, we will see a decrease in displacement as the $\mathbf{k}^2$ will control the relation. As we take recordings of various materials and examine the frequency spectrum of their displacement, we should expect to see this general relation. As we show in the next section, we would want a resolution on the micrometer scale, which is more precise than the millimeter resolution rotating LIDARs currently have.

One thing to note about the objects used to measure acoustics is that they cannot be transparent to the wavelength of the LIDAR. Unlike LDV microphones, Time of Flight LIDARs cannot measure the displacement of objects that do not reflect the LIDAR pulse. We illustrate the concept with a balloon.

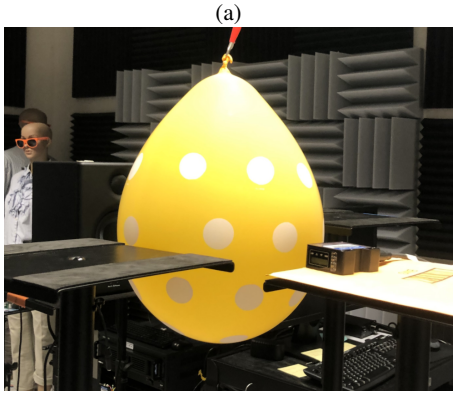
B. Measuring Acoustic Displacement with non-interferometric laser distance sensor

We demonstrate the audio recording potential of LIDAR through recording with a triangulation based laser distance sensor, specifically the Baumer OM 70 [6]. This point sensor obtains measurements similarly to a time-of-flight LIDAR, reading displacements over time as opposed to measuring the Doppler shift. This point sensor has micrometer level precision with a refresh rate of 2500 Hz. Using this sensor, we recover audio from distance measurements off of the surface of a balloon.

We played a series of tones and chirps through a speaker at roughly 100 dB and recorded the displacement with a balloon. An example of the setup with a balloon can be found in Figure 1b. We then band-pass filtered out the inaudible frequencies and took the Short Time Fourier Transform (STFT) with a DFT size of 1024, hop size of 256, and a hamming window applied at every frame.

Recordings taken with the triangulation point sensor off of the balloon are show in Figure 2a, with the original signal shown in Figure 2b. We see that the energy of the signal from the balloon decreases with frequency, which was what we expected in Section II-A.

(a)



(b)

Fig. 1: Figure 1a shows diagram of a ToF LIDAR used for spatial audio processing, Figure 1b shows a similar experimental setup

Plotting the signal taken off of the balloon shown in Figure 3, we can see that the signal ranges from -0.1 to 0.1 Volts. At a scale of 500 mV per millimeter, the range of the acoustic displacement is roughly 400 μm.

C. Distance Resolution vs. Audio Quality

If we wanted to record B bits and the maximum range of acoustic displacement is modeled as Δ, and we weren't oversampling, we can define the quantization spacing, ϕ, that is required as

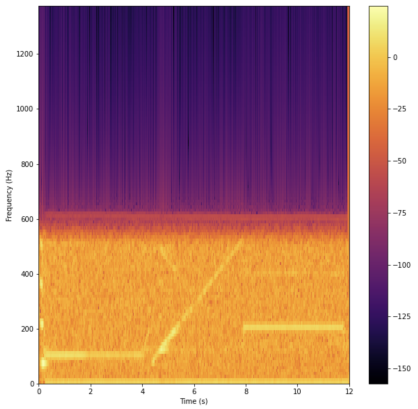
$$\phi = \frac{\Delta}{2^B}$$

For the signal plotted in the last section, we would need our LIDAR to have a resolution of $6.25\mu m$. Likewise, if we had a LIDAR with an optimistic 1 mm resolution, we would need a material that vibrates by 6.4 cm, which would be noticeably visible. It is more reasonable to expect ToF methods to improve over time than for a material to vibrate for more than a centimeter.

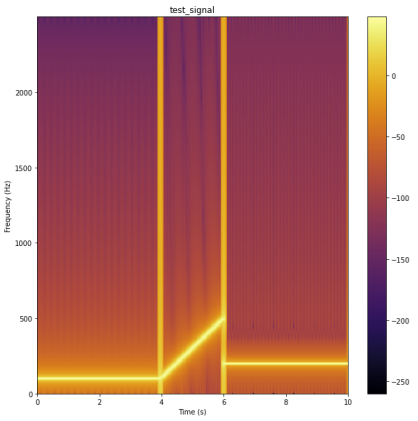
III. Audio Source Localization with LIDAR

Since we have shown that a ToF LIDAR can be utilized as an optical microphone, we can sample different points in space with an optical microphone to recover spatial audio information. [4] hinted at the potential of examining multiple objects with a high resolution camera in its conclusion, but experimentally recovering spatial audio information without a traditional microphone array has not been done before. Figures 1a and 1b show an illustration and experimental setup for spatial audio recovery with LIDARs.

While we use a single point LIDAR to demonstrate the array capabilities of the technology, we can also measure



(a)



(b)

Fig. 2: STFTs of the original signal (b) and the recorded signal off of the yellow balloon (a).

various points in space with a rotating ToF LIDAR. Unlike simultaneous recordings, the rotating LIDAR will obtain its data at different points in time as it sweeps around the scene. If our acoustic signal is band-limited and the LIDAR's rotational frequency is above the Nyquist rate of the signal, we can apply an appropriate (fractionally-sampled) delay at every object to account for the rotational delay of the spinning LIDAR. Current rotating LIDARs lack the high rotational speeds and resolution to obtain spatial audio information. Should these developments occur, we provide a potential algorithm for source localization using audio information captured with a ToF LIDAR and experimentally localize a source using the point sensor.

A. SRP-PHAT algorithm for Source Localization

As sound hits an acoustic object, the optically recorded signal may contain reflections of our acoustic signal off of the boundaries of the object. This means that would need a robust source location algorithm compared to time difference of arrival methods. Therefore we used the SRP-PHAT algorithm [7], [8] to localize sources with LIDAR.

The signal y_i measured at the ith object with the LIDAR can be modeled in the Short Time Fourier Transform (STFT)

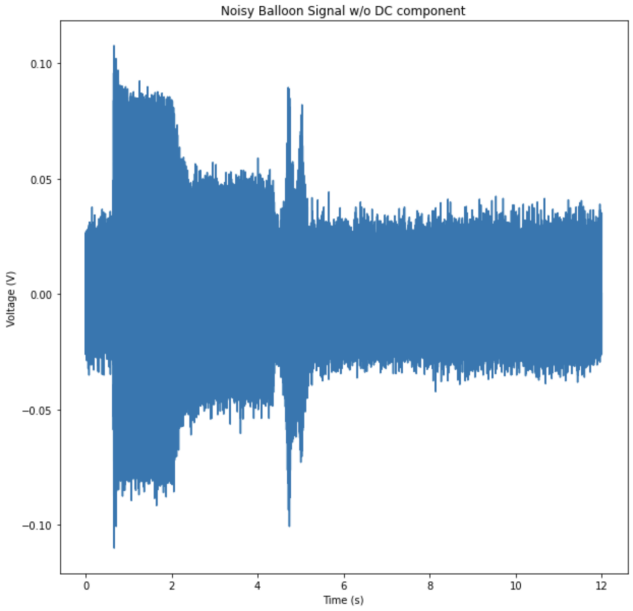


Fig. 3: Signal data taken off of a balloon with the DC component removed, represents the displacement from the acoustic signal with noise

domain as

$$Y_{i,d}(k,n) = H_{i,d}(k)X_d(k,n) + W_i(k,n),$$

where X_d is the signal from the sound source at direction d, $H_{i,d}$ is the transfer function of the ith object, and W_i is noise. We can find the direction r given the outputs y_i as

$$A_{m,1}(k,d) = \frac{H_{m,d}(k)}{H_{1,d}(k)}$$

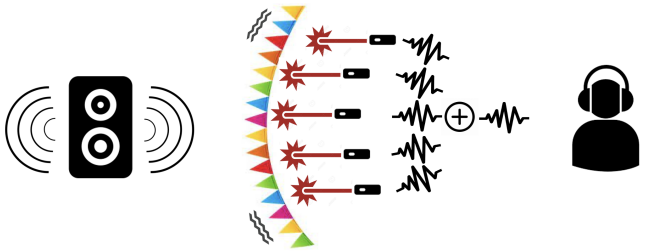
$$\Phi_{m,1}(k,d) = E[Y_{m,d}(k,n)Y^*_{1,d}(k,n)]$$

$$r = \arg\max_d \sum_k \left| \sum_m A_{m,1}(k,r) \frac{\Phi_{m,1}(k,d)}{|\Phi_{m,1}(k,d)|} \right|^2,$$

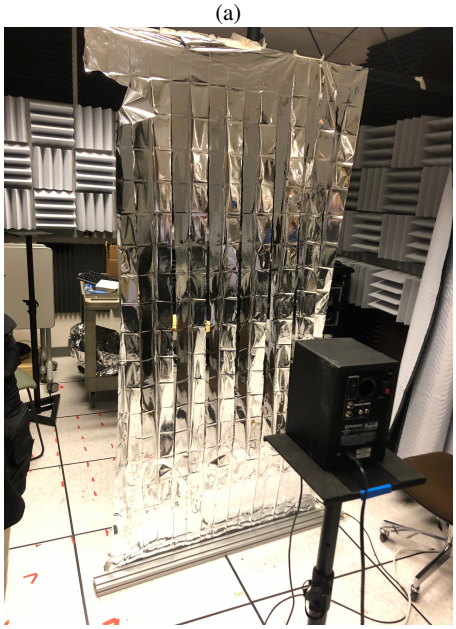
where $A_{m,1}(k,d)$ is the relative transfer function (RTF) between the mth and a reference microphone. Here, $\Phi_{m,1}(k,d)$ is the cross spectral density between the microphones and the reference. We would measure the normalized cross spectral densities from every source location beforehand, and then pick the density that maximizes the inner product between the CPSD and the RTFs for any signal to be localized. Since we don't have the transfer functions at every object, we can estimate the RTF as $\hat{A}_{m,1}(k,d) = \Phi_{m,1}(k,d)/\Phi_{1,1}(k,d)$.

B. Acoustic Source Localization with non-interferometric laser distance sensor

We ran an experiment to localize a speaker playing sound from multiple locations. Our acoustic object is an emergency blanket that was cut into 15 cm vertical strips. Because this object is thin, it vibrates noticeably from sound. We did not have multiple laser distance sensors to simultaneously record objects contrary to Figure 4a, but we can take sequential recordings with a system that synchronizes audio playback



(a)



(b)

Fig. 4: Figure 4a shows diagram of a ToF LIDAR used for spatial audio processing, Figure 4b shows a similar experimental setup

with sensor recordings. Our setup for array processing is shown in Figure 4b.

We took readings across three strips of the reflective sheet. We selected six speaker locations in a three by two rectangular grid of positions 60cm apart, where the first three locations were closer to the sheet than the second three locations. The speaker played a sequence of quadratic chirps ranging from $100-700$ Hz. We took a primary round of recordings and obtained the CPSDs at every location using the center strip as the reference object. Using the algorithm listed in Section III-A, we were able to locate most of the second set of recordings using the RTFs that were estimated as shown in Table I. Locations 0 and 3 are separate distances from the acoustic objects, but they were at similar directions relative to the reference panel, and may explain why our model could not properly differentiate between the two positions.

The displacements measured from the reflective sheet array are sensitive to the shape of the material, as explained in Section II-A. If the material shape changes, we should not be able to localize using the original RTFs. We show this in Table II, where we fail to localize the sources with the same algorithm after we slightly adjust the tensions of the sheet. This is a factor of spatial audio with an optical microphone that does not exist in a traditional microphone array, where the transfer function of the microphones are constant.

IV. Conclusion

In this paper we have achieved two novel goals:

1) We have shown that a non-interferometric laser distance sensor can be utilized for audio recovery.
2) We have shown that we can use an array of optical microphones to localize an acoustic source.

In conjunction with an optimistic view towards LIDAR developments, these findings allow for a device that can produce a spatial mapping visually and aurally given that there are acoustically thin objects that vibrate from sound dispersed around the room. Even though other optical methods are currently capable of audio recovery, it is the ToF LIDAR's array capabilities and low cost that makes technology of interest in this paper.

Since LIDAR plays a pivotal role in machine vision across multiple fields, ToF LIDAR is a high demand technology [9], [10]. Given sufficient developments in rotational ToF LIDAR technology, it could be used to multi-modally model indoor spaces, which would be of interest for virtual reality, machine vision, and scene capture.

References

[1] Velodyne, *VLP-16 User Manual*, 2019. [Online]. Available: https://greenvalleyintl.com/wp-content/uploads/2019/02/Velodyne-LiDAR-VLP-16-User-Manual.pdf

[2] R. P. Muscatell, "Laser microphone," 08 1984, uS Patent 4,479,265.

[3] B. Nassi, Y. Pirutin, A. Shamir, Y. Elovici, and B. Zadov, "Lamphone: Real-time passive sound recovery from light bulb vibrations," *IACR Cryptol. ePrint Arch.*, vol. 2020, p. 708, 2020.

[4] A. Davis, M. Rubinstein, N. Wadhwa, G. Mysore, F. Durand, and W. T. Freeman, "The visual microphone: Passive recovery of sound from video," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 33, no. 4, pp. 79:1–79:10, 2014.

[5] L. E. Kinsler, *Fundamentals of acoustics*. Wiley, 2000.

[6] A. Gerstner, May 2020. [Online]. Available: https://www.baumer.com/us/en/product-overview/distance-measurement/laser-distance-sensors/high-performance/c/37027

[7] S. Braun and I. Tashev, "Acoustic localization using spatial probability in noisy and reverberant environments," in *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2019, pp. 353–357.

[8] J. H. Dibiase, "A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays," Ph.D. dissertation, BROWN UNIVERSITY, Aug. 2000.

[9] Y. Li and J. Ibañez-Guzmán, "Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems," *CoRR*, vol. abs/2004.08467, 2020. [Online]. Available: https://arxiv.org/abs/2004.08467

[10] P. Czapski, A. Sluzek, and C. Tan, "Novel applications of lidar-based methods in robotic vision," *Robotics Research Centre Journal*, vol. 3, pp. 43–49, 04 2004.

Model Vector Number	Source direction number					
	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$
$d = 0$	**6.32**	4.58	4.27	**4.73**	3.80	2.67
$d = 1$	5.51	**4.94**	5.90	4.12	4.46	3.36
$d = 2$	3.87	4.04	**6.56**	2.79	3.87	3.34
$d = 3$	6.10	4.52	4.06	4.70	3.92	2.71
$d = 4$	4.82	4.42	5.67	3.58	**4.57**	3.38
$d = 5$	4.66	3.70	5.06	3.19	3.49	**3.48**

TABLE I: Localization Experiment Results, correct = 5

Model Vector Number	Source direction number					
	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$
$d = 0$	**6.65**	4.67	**5.38**	**4.87**	4.58	5.10
$d = 1$	5.20	3.29	4.03	3.91	3.73	3.84
$d = 2$	3.09	1.65	2.40	2.28	2.09	2.11
$d = 3$	6.54	**4.73**	5.32	4.77	**4.84**	**5.17**
$d = 4$	5.03	3.12	3.60	3.72	3.68	4.10
$d = 5$	4.23	2.38	3.17	3.14	2.99	3.23

TABLE II: Attempt to Localize after Adjusting Sheet, correct = 1