Development of a Hand-held 3D Scanning Acoustic Camera

Baden Parr, Mathew Legg, Steven Cox
Department of Mechanical and Electrical Engineering
School of Food and Advanced Technology (S&FAT)
Massey University
Auckland, New Zealand
{b.parr, m.legg}@massey.ac.nz

***Abstract*—Traditionally, acoustic cameras have used a fixed 2D near-field scanning plane for localising noise sources. This has been shown to cause errors in captured data and restricts acoustic observation to a fixed viewing position. With the recent popularisation of consumer 3D depth cameras and open source 3D reconstruction algorithms, researchers have begun investigating their utility in generating 3D near-field evaluation surfaces to increase beamforming accuracy. This paper presents the design and development of a portable hand held 3D scanning microphone array based on comprehensive simulations. Steps taken to improve the camera towards real-time construction of 3D acoustic models are also discussed.**

***Index Terms*—Acoustic Camera, Microphone Phased Array, Depth Camera, 3D Camera, Beamforming**

I. Introduction

Acoustic noise is an issue for many industries. Usually, noise is measured using a single microphone. This gives the magnitude of the noise level but does not provide information about the location that the noise is coming from. Microphone phased arrays, commonly referred to as acoustic cameras, have been developed to find both the magnitude and the location of sound sources. They are used by a range of industries such as the automotive and aeronautical industry to identify noise sources inside of or emitted from vehicles [1], [2], vibrations in airframes, and even rocket engine plumes [3].

Acoustic cameras that work in the near field utilise beamforming to calculate the acoustic intensity at specific locations relative to the array. Traditionally, acoustic cameras have used a 2D scanning surface (uniform grid of 3D points) that is usually oriented perpendicular to the array's Z-axis and located approximately at the same distance from the array as the object of interest [4]. It has been shown that this traditional 2D method can result in errors in the near field if the sound source does not lie this the 2D plane [5]. This issue can be resolved by using a 3D scanning surface that corresponds to the surface geometry of the object being imaged [6]. The 3D scanning surface can be obtained using a CAD model of the object. The resulting acoustic images may then be overlapped on the CAD model to display the noise sources on the object [1]. However, this method requires a CAD model of the object to be available and also requires an alignment process. An automated method of obtaining the 3D scan points is desirable. To address this, laser scanners have been used to obtain the 3D scanning surface for 3D images. However, this is generally a costly option. Legg and Bradley used a structured light technique to automatically obtain the 3D scanning surface for an acoustic camera [7]. Capturing acoustic data with both a spherical and spiral array, they showed that 3D interrogation clouds, which spatially match the subject, allow for correct focusing of the beamforming algorithm.

Fig. 1: Exploded diagram of 3D acoustic camera system developed in this study.

To gather positional data of a 3D environment or subject, depth cameras are commonly used. These cameras can operate on one, or a combination of, principles: speckle pattern correlation/structured light [8]; time of flight [9]; and stereo vision [10], to list the most common.

Chiariotti et al used the Microsoft Kinect to obtain a 3D scanning surface for the microphone array imaging of the interior of a car cabin [11]. However, the Kinect was not attached to the array but placed behind it. They used an average beamforming technique that reduced the effect of echoes. Heilmann et al. used several 3D cameras including the Structure Sensor, which they integrated into an acoustic camera and used to create 3D acoustic maps [12].

Concurrently and independently of the above work, a similar 3D scanning acoustic camera system was developed at Massey University, Auckland. This conference paper describes work performed developing this camera and recent work performed to improve this system.

978-1-7281-4842-7/20/$31.00 ©2020 IEEE

II. Hardware

The acoustic camera consists of three individual elements; a microphone array, Occipital Structure Sensor 3D camera, and a Windows Surface Pro Tablet. The entire device is presented in a portable hand-held form factor and can be operated by a single user. A graphical overview can be seen in Fig. 2 and an exploded diagram of the physical device can be seen in Fig. 1. The microphone array was designed and developed for the visualisation of high frequency audible sound, upwards of 2 kHz. To reduce cost, the array was constructed out of five identical PCBs, seen in Fig. 3. These were combined to form the final array. Each microphone was individually amplified and bandpass filtered with a collocated amplification stage. Important consideration was taken to optimise PCB layout for each microphone to reduce noise and the influence of any outside interference. Data Translation DT9816 devices were used to capture the amplified microphone signals at a sample rate of 50 kHz at 16 bit. These were primarily chosen for their compatibility with MATLAB. Four DT9816 devices were used. Each was mounted to the back of the main microphone array. Each DT9816 device has 6 ADC channels which limited the number of microphones to 24. The combined depth information from the 3D camera, and audio recordings from each individual microphone are processed using a real time MATLAB program.

A. Array Design

To ensure the portability of the final design, the size of the array was limited to a maximum of 450 mm wide. This was empirically determined to be a comfortable size to be handled by a single operator. Given this size restriction and the limited number of ADC channels available, an iterative procedure was developed using MATLAB to optimise the locations of the microphones. The software was able to simulate microphone array designs and measure performance metrics including beamwidth and side-lobe attenuation. The performance of square grid and spiral arrays were simulated with different microphone counts and shapes. An example of two comparable near-field beam patterns can be seen in Fig. 4.

The beamwidth of an array is determined by diameter of the array and wavelength of the sound being imaged. In this regard, the beamwidth of square grid arrays and spiral arrays will be similar when constrained to the same dimensions. However, side-lobe attenuation and aliasing are greatly influenced by microphone placement. For a particular frequency, the Maximum Side-lobe Level (MSL) is determined by the distribution of inter-microphone spacings. Square grid arrays

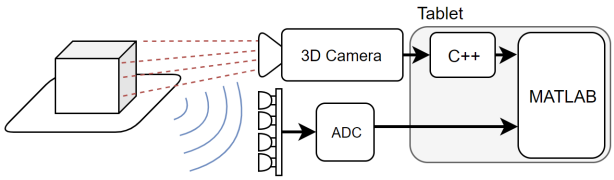


Fig. 2: Graphical system overview of the 3D acoustic camera.

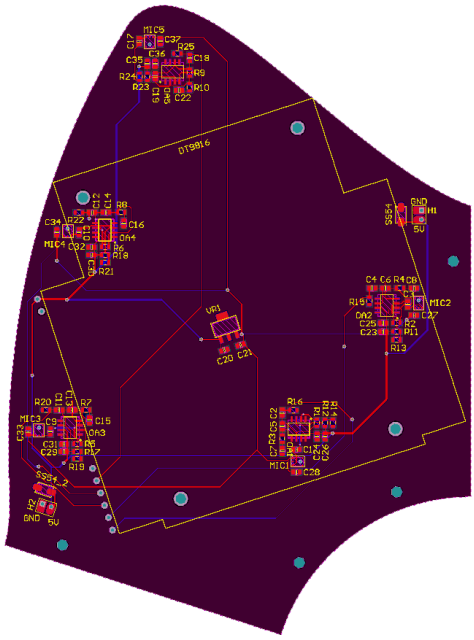


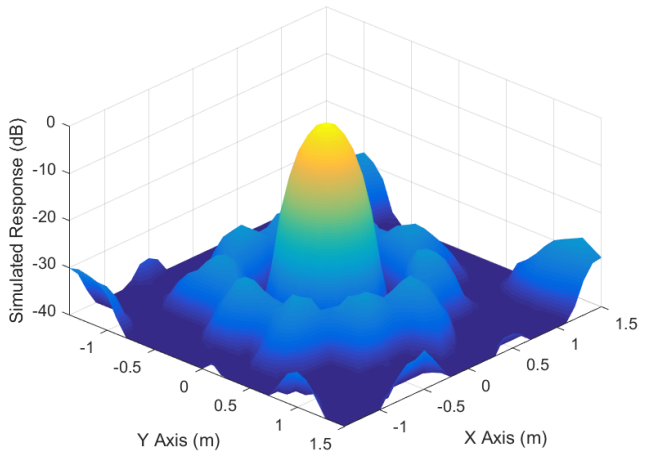
Fig. 3: One of five identical PCBs that combine to form the spiral array.

may be preferred in situations where single frequencies need to be measured, as they can be tuned for a single frequency. Alternatively, a well designed spiral array features a wider distribution of inter-microphone spacings and performs better over a wide range of frequencies.

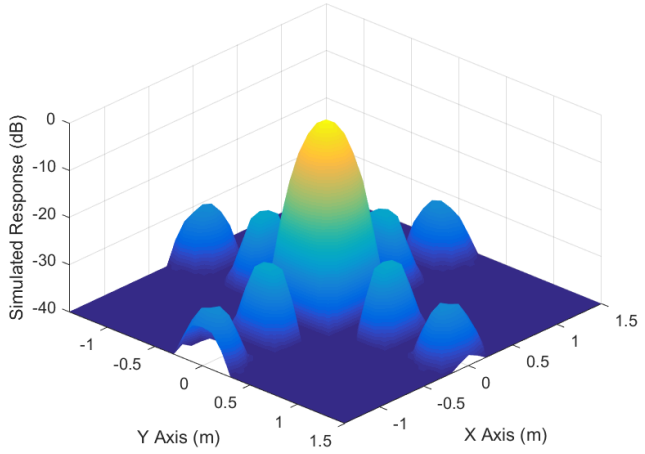
Square grid arrays have microphones equally spaced in a grid along two perpendicular axes. In comparison, spiral array designs are highly configurable. The array used in this study was based off equal arc length spiral arrays discussed in Underbrinks work on array design [13]. Each arm of the array follows a curve from a point on the internal radius, to a point on the external radius. The level of curvature is determined by a curvature parameter where 0 equates to no curvature and $\pi/2$ equates to infinite curvature. The curvature is mirrored around $\pi/2$, where π again results in no curvature. Microphones were placed with equal arc lengths between them along the resulting curve.

The curvature parameter has a large impact on the side-lobe performance of the array. Figure 5 shows the simulated side-lobe attenuation of a spiral array featuring 5 arms each with 5 microphones, and a radius of 200 mm. The simulation was conducted with a single 4 kHz tone, two meters perpendicular to the centre of the array and a simulation grid 3 m $\times$ 3 m. As can be seen, the performance steadily improves until 1.2 radians after which it begins to degrade. For this spiral design, the result is dependent upon the frequency of the simulated tone. For higher frequencies, the performance benefits from a lower curvature and visa-versa.

The performance of similarly sized square grid and spiral arrays were tested using MATLAB. A square grid array of 5 by 5 microphones with a width of 400 mm was compared to a



(a) Simulated near-field beam pattern for a 25 microphone spiral array positioned 2 m from a 4 kHz sound source. The result exhibits a side-lobe attenuation of 27.98 dB and a beamwidth of 0.30 m.



(b) Simulated near-field beam pattern for a 25 microphone square grid array positioned 2 m from a 4 kHz sound source. The result exhibits a side-lobe attenuation of 25.19 dB and a beamwidth of 0.27 m.

Fig. 4: Comparison of simulated near-field beam patterns between a 25 microphone spiral array and a 25 microphone square grid array. A 4 kHz tone was simulated 2 meters in front of the array. Scan was conducted on a 30 x 30 planer grid, 3.0 m $\times$ 3.0 m in size, at a distance of 2 meters from the array.

spiral array with five arms, each containing five microphones and a radius of 200 mm. To analyse near-field performance, a single tone, 4 kHz signal, was simulated 2 m perpendicular to the centre of each array. A measurement grid, spanning 3 m $\times$ 3 m, was evaluated at the same depth. The resulting near-field beam patterns can be seen in Fig. 4. The simulated spiral array measured a beamwidth of 0.30 m and a side-lobe attenuation of 27.98 dB. In comparison, the similarly sized square grid featured a slightly smaller beamwidth of

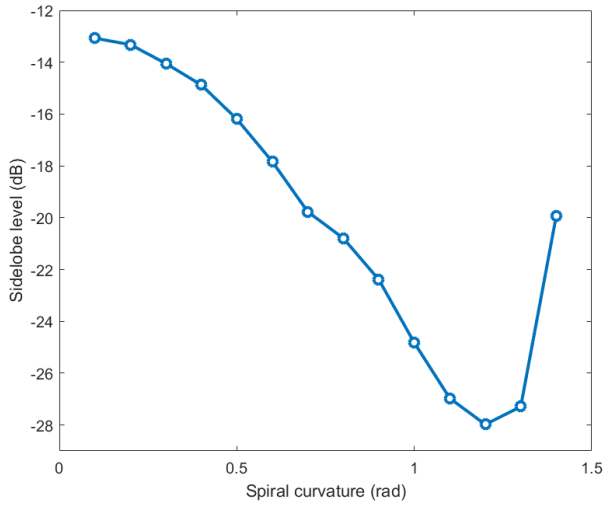


Fig. 5: Side-lobe attenuation against spiral curvature for values between 0.1 and 1.4 radians. The simulation used a tone of 4 kHz positioned 2 m from the array.

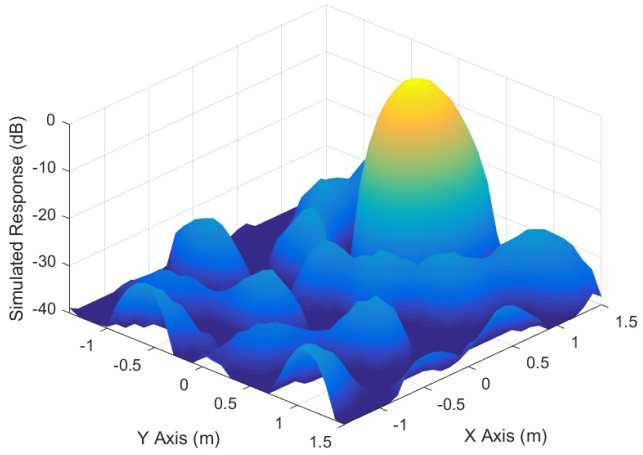
0.27 m but reduced side-lobe suppression at 25.19 dB. The improved beamwidth is expected due to the larger maximum microphone separation of 0.567 m compared to the spiral array of 0.383m. Visually, the near-field beam pattern of each array is considerably different. The square grid array has prominent side-lobes in the directions of the primary X and Y grid axes. In comparison, the spiral array features a concentric ring of side-lobes around the primary peak.

Spatial aliasing is another performance consideration that was explored when designing the array. Aliasing artefacts are the result of complex behaviours between the wavelength of an acoustic signal, the size and shape of the array, and the relative position of the acoustic source. Square grid arrays are most impacted by aliasing as they have a relatively uniform distribution of inter-microphone spacings. In comparison, the broad distribution in spiral arrays gives them resilience to such artefacts.

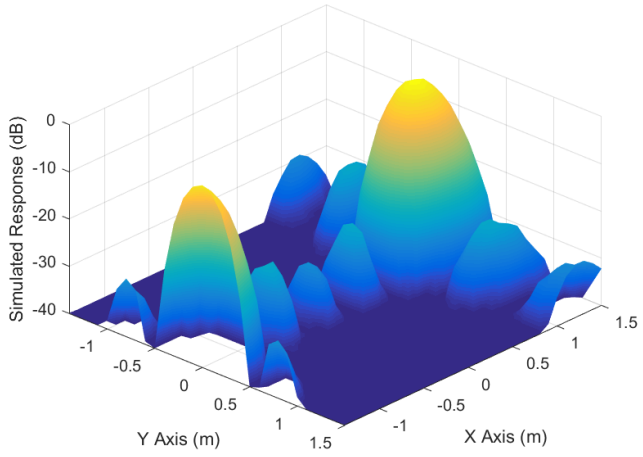
To analyse aliasing performance, a 4 kHz tone was simulated 2 meters in front of the array with a 1 meter offset from the perpendicular centre. A scan was conducted on a 30 x 30 planer grid, 3.0 m $\times$ 3.0 m in size at the same depth as the sound source. The resulting near-field beam patterns for spiral and square grid arrays can be seen in Fig. 6.

The near-field beam pattern for the square grid array features a significant aliased image at $[x, y] = [-1.5, 0.0]$ which is indistinguishable from the actual source at $[x, y] = [1, 0.0]$. In comparison, the beam pattern of the spiral array shows no aliasing artefacts in the same simulated area.

The final placement of the 25-microphone array can be seen in Fig. 7. It should be noted that only 24 of these microphones could be utilised in the final design due to limited ADC channels.



(a) Simulated near-field beam pattern for a spiral grid array demonstrating relative lack of side-lobes.



(b) Simulated near-field beam pattern of a square grid array demonstrating significant aliasing image artefact at $[x, y] = [-1.5, 0.0]$.

Fig. 6: Comparison of simulated near-field beam patterns between a 25 microphone spiral array and a 25 microphone square grid array.

B. 3D Camera

A real-time 3D camera was utilised to provide accurate scan points for the acoustic beamforming. The Occipital Structure Sensor 3D camera used in this design works on the principle of structured light and features a resolution of 640 $\times$ 480 pixels. At full resolution, this would produce 307200 unique 3D points to be evaluated through beamforming of the microphone array recordings. This process would take roughly 15 minutes with the current MATLAB implementation. To reduce this, the point cloud was down-sampled to a resolution of 21 $\times$ 16 so that subsequent beamforming could be achieve in real time. Individual frames were captured from the camera at a rate of 1 Hz and after down-sampling saved to a timestamped file for access by MATLAB.

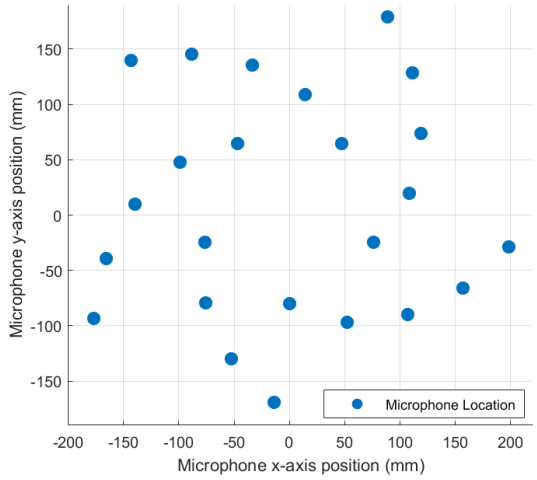


Fig. 7: Optimal microphone positions determined by the iterative MATLAB simulation.

III. Results

Measurements were conducted using a speaker positioned on a flat surface playing a single 2 kHz tone. This frequency was selected as it would manifest in the response with a beamwidth easily measurable through hand held testing.

At a measurement distance of 0.65 m the camera measured a beamwidth of roughly 0.3 m for a 2 kHz tone. When the same situation was simulated a beamwidth of 0.27 m was achieved. Comprehensive testing of the side-lobe levels have yet to be conducted; however, empirical manual testing has shown no significant aliasing artefacts when the camera is offset from the sound source. Computing all 336 beamforming points from the 3D camera took just over 1 second using MATLAB on the 1st gen Surface Pro[1] meaning that the device is suitable for real time use. A video demonstrating the portable 3D Acoustic camera in operation can be seen on YouTube[2].

IV. System Improvements

The integration of 3D scanning technology with microphone phased arrays has potential to significantly improve accuracy in the near field for 3D objects and increase ease of use. However, more work is needed in this area. In our work, we found visualisation of the acoustic map over the 3D scans and syncing of the camera with the microphone data was challenging. A disadvantage of the 3D camera that we used (the Structure Sensor) was that it did not produce a coloured point cloud and it could not be used in direct sunlight. Therefore in future designs the 3D camera will be changed to the Intel Real Sense D415. This is smaller, can operate in direct sunlight, and produces a coloured 3D point cloud pointcloud. Combining the coloured point cloud with

[1]Intel Core i5-3317U, 4 GB dual-channel DDR3, 1600 MHz
[2]https://youtu.be/c7x2_nQzyZs

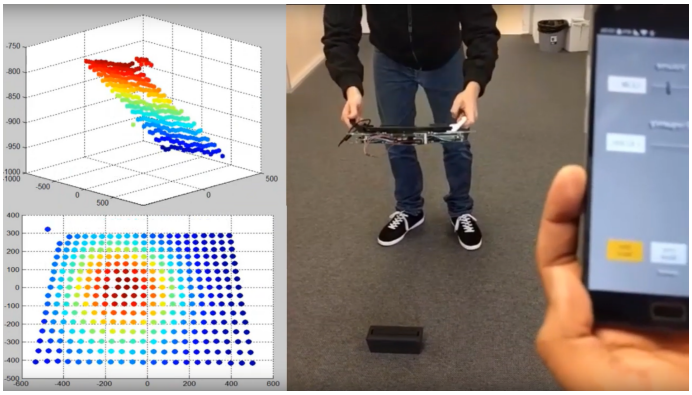


Fig. 8: Frame taken from video demonstrating the real-time operation of the portable 3D acoustic camera. Full video can be found on YouTube, see footnote.

the resulting 3D acoustic has the potential to allow improved visualisation. The Intel camera also has a pin that outputs a trigger signal. Work has begun to use this as a hardware trigger to sync the acquisition of the microphone data with the depth camera.

A. Synchronisation of Acoustic Data with Depth Frames

The 'Hardware Sync' pin of the RealSense D415 Depth Camera was a big reason for exploring use of this camera. The trigger pin, which emits a pulse on capturing every frame, was used to trigger simultaneous data collection on the four DT9816 devices. The whitepaper documentation [14] specified the trigger pulse parameters as being 1.8V CMOS with a 100 micro second positive pulse. With the external trigger specifications for the DT9816 devices given as 2.4V with pulse width 84 nano seconds, it was necessary to level shift the trigger signal to a higher voltage. This was done using a bi directional MOSFET level shift circuit [15], which boosted the trigger signal to 5V. Supplementary electrostatic discharge (ESD) protection was also added per the whitepapers recommendation.

B. Sequential 3D Frame Alignment

It was also desirable to be able to merge 3D acoustic maps taken from different locations. Ideally, one should be able to move around a 3D object with the acoustic camera and the resulting 3D acoustic map be automatically merged into a single 3D maps. To address this, initial work has been performed using iterative closest point (ICP), a proven algorithm, that can find alignment between 3D scans of an object obtained from different positions [16]. However, for ICP to be performant in real-time the number of iterations required must be minimised. For this the Intel T215 Tracking Camera is being explored as an approach for offloading the computationally expensive process of pose tracking between consecutive 3D scans. Accurate pose change estimations will ensure the ICP algorithm can find a robust alignment in a minimal number of iterations. An initial test of this camera

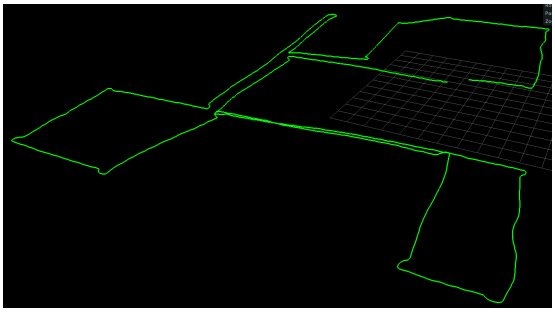


Fig. 9: Recorded path from the Intel T215 camera as it was moved through an office building.

can be seen in Fig. 9 in which the camera pose was recorded as it was moved through an office building.

V. Conclusion

This paper describes the development of a microphone array that automatically captures the 3D surface of objects to provide correct evaluation points for near-field beamforming. Preliminary results have confirmed the real world performance matches the simulated beam patterns. Work has begun towards improving synchronisation between camera and acoustic recordings. This is facilitated by the use of Hardware Synchronisation on the Intel D415 3D Camera. Alignment of consecutive frames has also been explored to support future work towards real-time construction of 3D acoustic models.

References

[1] A. Meyer and D. Döbler, "Noise source localization within a car interior using 3d-microphone arrays," *Proceedings of the BeBeC*, pp. 1–7, 2006.

[2] D. Yang, Z. Wang, B. Li, and X. Lian, "Development and calibration of acoustic video camera system for moving vehicles," *Journal of Sound and Vibration*, vol. 330, no. 11, pp. 2457–2469, 2011.

[3] J. Panda and R. Mosher, "Use of a microphone phased array to determine noise sources in rocket plumes," in *49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, 2010, p. 974.

[4] D. Döbler, G. Heilmann, and R. Schröder, "Investigation of the depth of field in acoustic maps and its relation between focal distance and array design," in *Proceedings of Internoise*, 2008, pp. 654–660.

[5] M. Legg and S. Bradley, "Comparison of clean-sc for 2d and 3d scanning surfaces," in *4th Berlin Beamforming Conference*, 2012.

[6] M. Legg, "Microphone phased array 3d beamforming and deconvolution," Ph.D. dissertation, University of Auckland, 2012.

[7] M. Legg and S. Bradley, "Automatic 3d scanning surface generation for microphone array acoustic imaging," *Applied Acoustics*, vol. 76, pp. 230–237, 2014.

[8] K. Khoshelham, "Accuracy analysis of kinect depth data," in *ISPRS workshop laser scanning*, vol. 38, no. 5, 2011, p. W12.

[9] J. Sell and P. O'Connor, "The xbox one system on a chip and kinect sensor," *IEEE Micro*, vol. 34, no. 2, pp. 44–53, 2014.

[10] Intel, *Intel RealSense Camera R200 Datasheet*, 2016.

[11] P. Chiariotti, G. Battista, M. Ettorre, and P. Castellini, "Average acoustic beamforming in car cabins: An automatic system for acoustic mapping over 3d surfaces," *Applied Acoustics*, vol. 129, pp. 47–63, 2018.

[12] G. Heilmann, D. Döbler, A. Meyer, and S. Barré, "Dynamic beamforming using moving phased arrays with integrated 3d scanners." in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 255, no. 2. Institute of Noise Control Engineering, 2017, pp. 5104–5109.

[13] J. Underbrink, *Aeroacoustic Measurements*, T. Mueller, Ed. New York: Springer-Verlag, 2002.

[14] A. Grunnet-Jepsen, P. Winer, A. Takagi, J. Sweetser, K. Zhao, T. Khong, D. Nie, and J. Woodfill, "Using the intel realsense depth cameras d4xx in multi-camera configurations," 2018.
[15] Hobby Electronics. (2019) Bi-directional mosfet voltage level converter 3.3v to 5v. [Online]. Available: https://www.hobbytronics.co.uk/mosfet-voltage-level-converter
[16] T. Whelan, M. Kaess, M. Fallon, H. Johannsson, J. Leonard, and J. McDonald, "Kintinuous: Spatially extended kinectfusion," 2012.