

Final Year Project Interim Report

Microphone Array Calibration using a 360° Camera

Ng Zhi Long Benjamin
CID 00986013

January 29, 2018

1 Introduction

To perform essential tasks in acoustics engineering such as beam forming and sound localisation, an accurate model of the array manifold must be obtained. Although analytical methods of determining the microphone array manifold exist, they often result in overly simplistic models. This is because the common assumptions of ignoring the interactions of the microphones and their mountings with the sound field and generalising microphones' sensitivities are often untrue.

Hence, acoustic microphone array calibration is necessary to better model the true array manifold since it accounts for many physical assumptions that the analytical model neglects. Such calibration involves making numerous measurements of sound propagations for a series of known sound source directions. However, while the acoustic measurement may be simple, determining the true position of the source is inherently cumbersome, with current methods requiring specialised facilities and substantial space to setup. These setups often include the use of a fabricated structure housing an array of speakers at pre-defined intervals as shown by the geodesic speaker array in Figure 1.



Figure 1: Geodesic Speaker Array [4]

Therefore, this project aims to improve on current calibration methods by using a 360° camera to determine speakers' true positions during the calibration process. If successful, this will improve the flexibility and portability of current calibration procedures, making calibration procedures more convenient and accessible. Additionally, this new calibration method has the potential to increase the resolution and accuracy of the array manifold, possibly leading to better performance of current acoustic engineering algorithms that rely on the array's model.

Although this project's primary interest is in accurately modeling an arbitrary microphone array's manifold, numerous research papers working on the topic of microphone calibration aim to model the Head-Related Transfer Function (HRTF) [1], [2], [3], [4], [5], [6] and [7]. As the HRTF is a special case of microphone array manifolds in that the microphone 'housing' is largely the ear and head of a human subject, and the subject of interest is the calibration procedure, for the purposes of this paper, the process of obtaining of HRTFs and array manifolds will be similarly treated.

2 Project Specification

With the impetus as delineated in Section 1, this project ultimately aims to deliver a tool to swiftly determine speakers' true positions, which can then be elegantly integrated into existing calibration methods. The process overview to obtain the array manifold from sample measurement to interpolation is shown in Figure 2. Hence, the key focus of this project is to revolutionise the sample measurement process in Step 1, and integrating it into the overall process as delineated.



Figure 2: Process to Obtain Array Manifold

This tool will be a MATLAB based algorithm which interfaces with a Ricoh Theta V 360° Camera shown in Figure 3 which is used for speaker detection and position measurement. Although the image processing will first be done using OpenCV in Python to enable quick prototyping and debugging, it will ultimately be integrated with existing calibration methods which are based on MATLAB. This integration is currently envisioned to be a wrapped function call that provides the speaker's position whenever an array calibration measurement is taken. After converting these measurements into the spherical harmonics domain, the array manifold will be estimated through interpolation methods.



Figure 3: RICOH Theta V Camera [25]

To evaluate the accuracy and usefulness of the array manifold that is generated using this novel calibration method, actual impulse responses at interpolated points will be measured and compared to predicted responses given by the model. Alternative methods of measuring performance include assessing its ability to beam form and provide direction of arrival estimates accurately and reliably.

As shown in Figure 4, the final system consists of the camera situated together with an arbitrary microphone array to provide the speaker's arbitrary position in 3D. This process will be in sync with each impulse measurement taken.

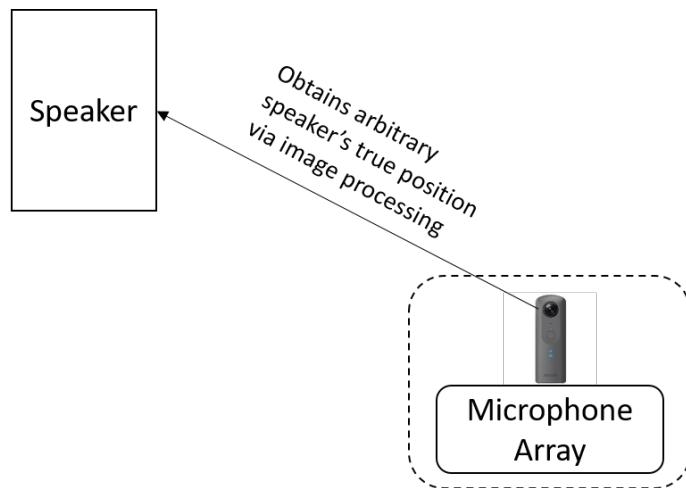


Figure 4: Proposed Calibration Tool

At the time that this Interim Report is completed, most preliminary work have been focused on the development of a working speaker detection prototype. Future work will be focused on the integration and evaluation of the developed tool and will be elaborated in the Final Report.

3 Literature Review

The background research detailed below is a brief overview of the theory needed to integrate the tool into current methods of array calibration. Areas for further reading and research will be indicated as necessary.

3.1 Current Calibration Methods

Since past attempts [5] and [6] to obtain array manifolds without physical acoustic measurements performed poorer than calibration methods that did, current methods are based on variations of the procedure outlined in [1].

The general calibration procedure involves a human listener or microphone array to be situated at the center of an array of speakers. The array of speakers can be fixed in a geodesic structure as in [3] and [4] or on a semicircular arc as in [1] with the arc laterally rotated. There are distinct advantages and disadvantages of each setup. While speakers fixed in a geodesic structure enables experimenters to take a set of measurements very rapidly around a sphere about the subject, the fixed displacement of speakers in the structure limits the overall resolution of the array manifold obtained. On the other hand, while speakers in a rotating arc enabled experimenters to increase the resolution of the array manifold by being able to increase the number of samples taken in the axis of rotation, calibration time took longer and often only the measurements from the upper hemisphere could be taken. Examples of these speaker arrays are shown in Figure 5.



(a) Setup in University of Oldenburg [27]



(b) Setup in University of Southampton [28]

Figure 5: Current Calibration Methods

Nonetheless, all variations of current setups require specialised structures which are immobile and can only take a fixed number of measurements in the axis that the speakers are fixed in space. This results in the calibration process being limited to the specialised space created and the ultimate array manifold resolution to be limited as well.

It may be worth noting that alternative setups based on the principle of reciprocity exist [7]. These setups swap the positions of the speaker and the microphones such that a microspeaker is placed in the subject's ear while several microphones are positioned around the subject instead. Regardless, these setups still require specialised fixed structures as shown in Figure 6 and are subject to the constraints mentioned above.

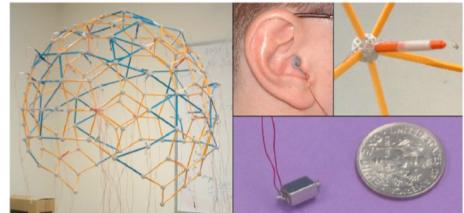


Figure 6: HRTF Measurement via Reciprocity [7]

In light of the limitations of current methods, this project aims to use the 360° camera to deliver a more elegant calibration procedure. Since the proposed setup is not limited by a fixed structure, it will be more mobile and allow for an arbitrary number of measurements to be taken, potentially increasing the resolution of the resulting array manifold.

It is noteworthy that this possible increase in array manifold resolution is of less importance to experiments dealing with HRTF since it has been found in [4] that there is a limit to perceptual improvement by obtaining finer HRTF resolutions past a certain order of harmonics. Nonetheless, because this project aims to deal with array manifolds, where the performance of the arrays (for example to beam form) are improved by increases in spatial resolution, such perceptual simplifications should not be done.

3.2 Array Manifold Representation

3.2.1 Spherical Harmonics Representation

To provide a robust model for the representation of microphone array manifolds, recent research [9], [10], [11], [12], [13], [14] and [15] have focused on representing them as a set of spherical basis functions also known as spherical harmonics. Representing the array manifold as such is arguably more natural than alternative representations since it is more convenient to describe the solutions to wave equations around a point in terms of spherical coordinates. Similar to how single dimensional signal processing utilises the Fourier expansion to represent arbitrary signals as weighted sums of sinusoidal harmonics, these spherical harmonics arise as solutions to the Laplace equation expressed in spherical polar coordinates (where ϕ is the azimuth and θ is the inclination) and are spatially continuous and orthonormal. These spherical harmonics are defined as:

$$Y_n^m(\theta, \phi) \equiv \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta) e^{im\phi} \quad (1)$$

where, $P_n^m(\cdot)$ are the associated Legendre functions, $m \in \mathbb{Z}$ is an integer denoting the function degree and $n \in \mathbb{N}$ is a natural number denoting the function order [8]. Since the array manifold is the weighted sum of these spherical harmonics, increasing the order of harmonics increases the resolution and information obtained.

For visual purposes, the balloon plots of the real and imaginary parts of the first four orders of spherical harmonics are shown in Figure 7. The order n determines the highest power of the $\cos\theta$ and $\sin\theta$ terms which vary the dependence of the spherical harmonics over θ and m varies the dependence over ϕ through the exponential term $e^{im\phi}$.

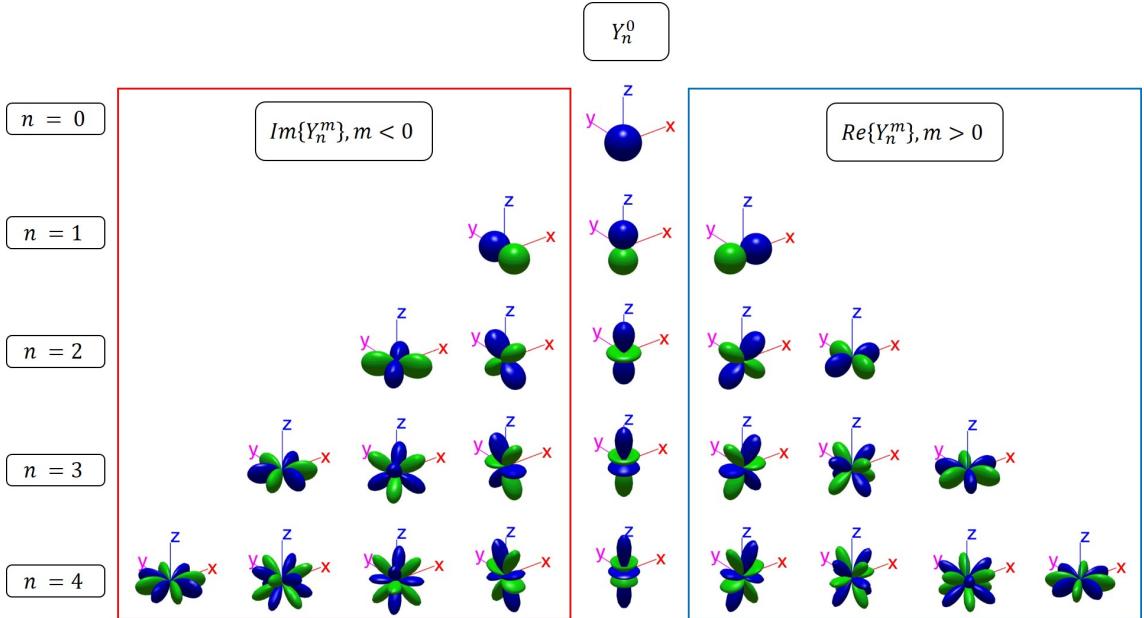


Figure 7: Balloon Plots of First Four Orders of Spherical Harmonics adapted from [26]

3.2.2 Spherical Fourier Transform

Again drawing from experience of single dimensional signal processing, obtaining the array manifold's impulse response requires a fully continuous measurement of impulses sent by speakers from every possible three dimensional position. Assuming such measurements are possible and a continuous impulse response $A(\theta, \phi)$ is obtained, the spherical harmonic coefficients can be determined by the forward Spherical Fourier Transform (SFT):

$$A_{n,m} = \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} A(\theta, \phi) [Y_n^m(\theta, \phi)] \sin \theta d\theta d\phi \quad (2)$$

However, real measurements are never fully continuous but discretised where a series of impulses is sent by a speaker from a limited number of pre-determined locations. This results in only a discrete set of array manifold impulse responses being obtained. Hence, the SFT is performed on each discrete impulse response sample and the spherical harmonic coefficients can be approximated by:

$$A_{n,m} = \sum_p w_p A(\theta_p, \phi_p) [Y_n^m(\theta_p, \phi_p)] \quad (3)$$

where $\{w_p\}_{p=1\dots P}$ are the quadrature weights of the sampling scheme and P is the number of samples taken.

3.3 Interpolation Methods

Finally, because only a discrete number of points around the microphone array can be sampled, interpolation is necessary to obtain an array manifold which is a continuous function of 3D space. Generally, increasing the number of samples increases the resolution of the interpolated manifold which allows for better predictions at unmeasured spatial locations. Further reading into the various interpolation methods will be done in [9], [10], [11], [12], [13], [14] and [15].

4 Implementation

As this proposed solution aims to elegantly combine both image and audio signal processing into a single system, it is helpful to delineate the implementation into three distinct phases:

1. Tool development (Image processing)
2. Tool integration (Audio signal processing)
3. Tool evaluation (Audio signal processing)

As elaborated in Section 2, OpenCV on Python was used for Phase 1 to aid quick prototyping while MATLAB will be used for Phase 2 and 3 since that is the programming platform that existing processes use. The camera module used in all tests will be the RICOH Theta V and its specifications can be found in Appendix A. Additionally, versioning aids such as Github has been used to track software development. The relevant repository and logbooks can be found at https://github.com/bngzl/final_year_project. Current implementation and possible refinement of the developed tool in Phase 1 is elaborated in Section 4.1 while future development plans are detailed in Section 4.2 and 4.3.

4.1 Tool Development

For a camera to reliably determine the true position of a speaker relative to the microphone array during calibration, it must be able to 1) Detect the speaker, 2) Track the speaker and 3) Map pixel locations on screen into real world coordinates.

Detection is essential since it is the critical step to determine a speaker's pixel position on screen. However, as image processing operates only on each frame of the video feed, detection is greatly influenced by the refresh rate and blur of camera as the speaker moves around in its field of view. This limits the ability of the image processing algorithm to reliably detect the speaker in all frames as it moves around. Hence, tracking is important as it enables the system to fill and return reasonable predictions of the speakers' position in those frames. Finally, mapping of pixel locations to real world spherical coordinates enables the smooth integration into current calibration methods. Figure 8 depicts the decision flow chart integrating acoustic with position measurements using detection and tracking algorithms.

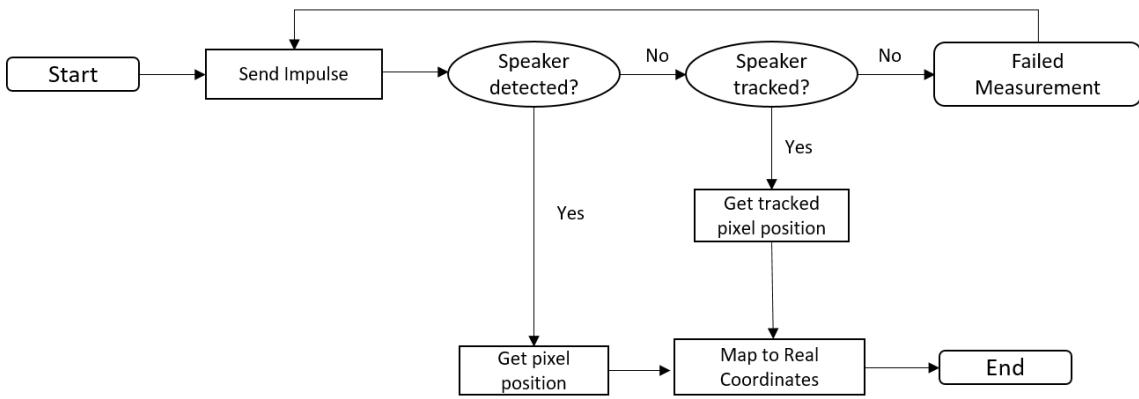


Figure 8: Proposed Decision Flow Chart

4.1.1 Speaker Detection

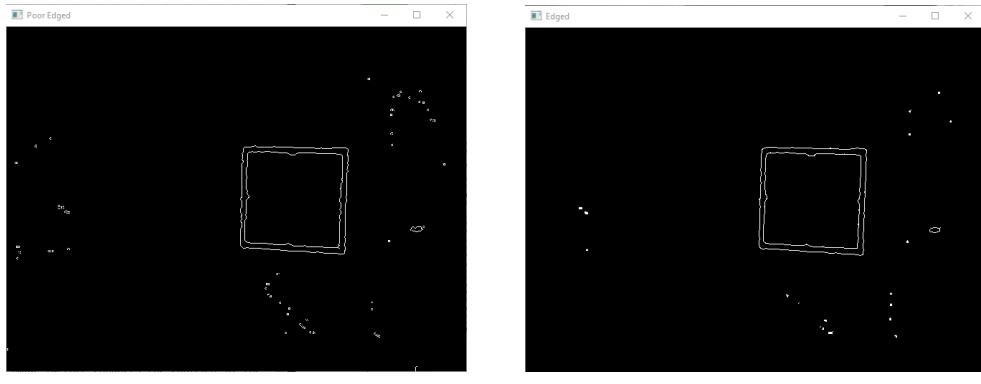


Figure 9: High Level Design of Detection Algorithm

In image detection, it is pivotal to detect the object by certain characteristics that would distinguish it from its environment. Generally, systematically checking for these characteristics allows the camera to ascertain with a higher probability that it has detected the speaker. To ensure system robustness, the speaker was distinguished by first colour and then edge geometry as shown by the detection system design in Figure 9

As the speaker had a common colour black, to detect the speaker by colour, a brightly coloured frame was used to outline its edges to contrast it against the environment as shown in Figure 10. This was useful in a visually noisy environment where many objects may have the same colour and shape as the speaker. The camera was thus colour calibrated to only detect objects within the specified colour range. The resulting output is shown in Figure 12b.

Since most edge-detection algorithms are extremely noise sensitive, Gaussian blurring is then applied to the frame as a low pass filter to remove high frequency noise components [16]. Additional erosion and dilation of the resulting image further enhances the contrast of its edges [17]. The improvement of this additional frame processing is shown in Figure 11b. Compared to Figure 11a, Figure 11b is less noisy and is better able to define the edges of the speaker. These edges are then represented as different sized arrays.



(a) Edge Detection before Frame Processing (b) Edge Detection with Frame Processing

Figure 11: Comparison before (a) and after (b) Frame Processing

The processed edges that pass through the colour filter are then sorted according to size in descending order. This is done as the speaker's contour will likely be the largest edge that has passed through the preliminary filters and sorting it allows the remaining algorithm to operate on it first, reducing the time taken to find the correct contour. To determine if a contour is the speaker's, it must have:

1. 4 to 6 Vertices
2. Aspect Ratio between 0.8 and 1.2
3. Dimensions of more than 50×50 pixels

Although the speaker is square ($11\text{cm} \times 11\text{cm}$) and thus has four vertices and an aspect ratio of 1, these heuristically defined ranges were defined to account for the blur of the speaker's image as it moves across the camera's field of view. Additionally, since the speaker will be a relatively large image, setting a lower bound of 50×50 pixels guards against stray noise edges from being mistakenly detected. Figure 12 shows the output at each stage of the system for an average laptop's web camera.



Figure 10: Framing the Speaker

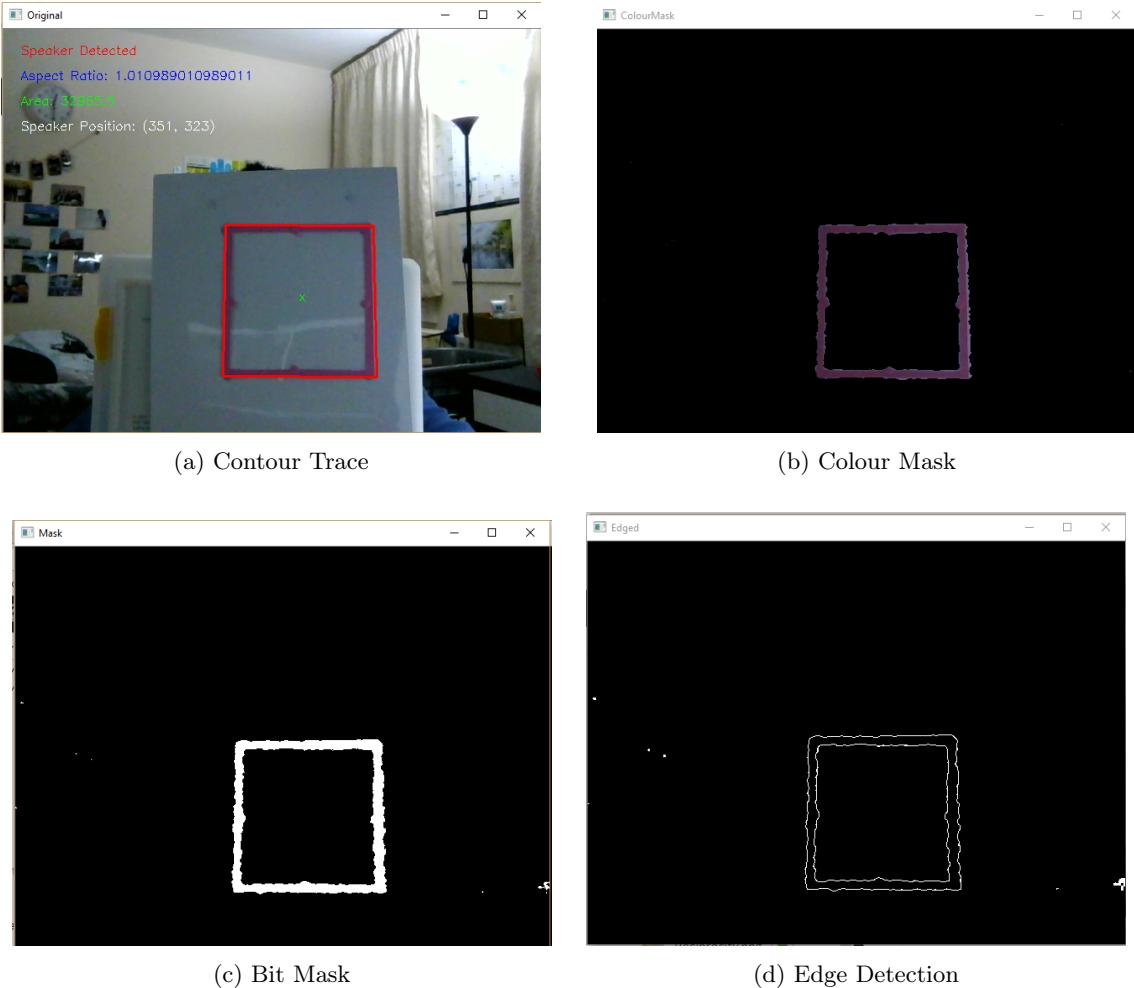


Figure 12: Output Stages of Detection System

While the current algorithm has experienced success in controlled environments, further refinement is needed for optimal performance in all operating environments. For example, the colour filter's performance is greatly affected by the nature of ambient light falling onto the speaker, which changes as the speaker is shifted. A possible solution is to replace the coloured frame with a coloured LED frame which is expected to reduce performance dependence on external light sources. Results from further testing will be reflected in the Final Report.

4.1.2 Speaker Tracking

Preliminary research reveals that there are multiple algorithms available for tracking purposes. For instance, optical flow algorithms based on the Lucas-Kanade method attempts to characterise the motion of tracked objects via a motion vector. On the other hand, Kalman Filtering is a recursive algorithm that estimates the future position of tracked objects making it especially useful to estimate objects' positions that are temporarily hidden from the camera's view.

Future research is definitely necessary to better understand these algorithms and their subsequent implementation. Possible leads include combining both Optical Flow and Kalman Filters as done in [19]. OpenCV tutorials such as in [20] and [21] also describe alternative tracking algorithms such as Meanshift and Camshift. The integration of these tracking algorithms into the current system is expected to be straightforward.

4.1.3 Position Identification

To simplify the problem of position identification, the speaker's pixel position is approximated as the center of its contour. As elaborated in [23], this is done by calculating the center of gravity of the image via its spatial first and zero order moments, where moments are simply weighted averages of the image's pixel intensities and are often used to characterise and provide information on the detected object's shape. This is then implemented based on [18]. An example of the detected speaker's pixel position can be seen in Figure 12a.

As the preliminary system detailed in Section 4.1 has yet to be tested with the RICOH Theta V 360° camera, there is limited knowledge regarding the mapping from pixel to true positions. Currently, it is known from RICOH Theta website in [24] that it supports two image formats: 1) Equirectangular and 2) Dual-Fisheye. The equirectangular format is stitched using RICOH Theta's proprietary software while the Dual-Fisheye format is available as a raw file. Examples of each format from RICOH Theta's website are as shown in Figure 13.

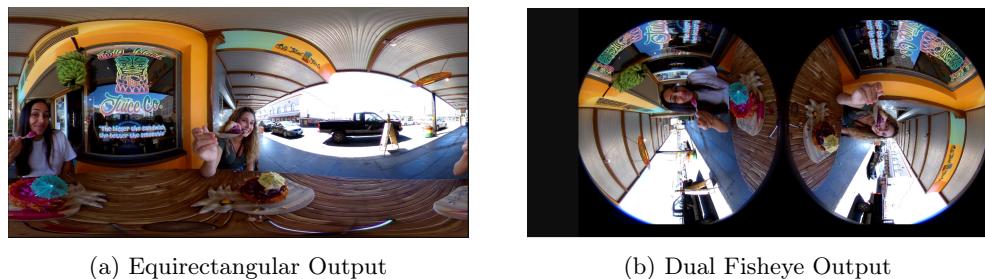


Figure 13: Equirectangular and Dual Fisheye Output from RICOH Theta V [24]

There are multiple unknowns here as it has not yet been verified if 1) The stitching done by the proprietary software is sufficiently accurate for mapping purposes and 2) Whether this stitched 'liveview' can be accessed by OpenCV as a camera port in Python real-time. If either of these conditions are not met, then research into manual stitching of the Dual Fish-eye image has to be done. However, preliminary Internet research has shown promising examples of object detection as well as coordinate mapping using the lower end model Ricoh Theta S. Further testing with the spherical camera will be done and results detailed in the Final Report.

4.2 Tool Integration

At the time this Interim Report is written, only theoretical knowledge on current calibration procedures is available as elaborated in Section 3.1. Further practical work on learning the actual procedures has to be done to better understand how the developed tool can best be integrated into current procedures. The relevant milestones, timeline and project management plan are elaborated in Section 5.

4.3 Tool Evaluation

After developing and integrating the tool successfully into the calibration procedure, it can be used to obtain an array manifold for a known microphone array. There will then be several methods available for tool evaluation, including 1) Comparing the newly obtained array manifold with past results using traditional methods, 2) Comparing predicted responses at interpolated points given by manifold with actual responses, 3) Evaluating its ability to accurately beam form and 4) Evaluating its ability to provide direction of arrival estimates well.

These methods are well documented in [7], [9] and [10], although further research and practical work into this area will be helpful. The relevant milestones, timeline and project management plan are elaborated in Section 5.

5 Project Management

5.1 Tasks, Milestones and Gantt Chart

With a project that is to be completed in parallel with other assignments and examinations, it is pivotal that a clear work plan is prepared to reveal any task dependencies or critical paths. Hence, Table 1 lists all project essential tasks along with their respective durations. Additionally, Table 2 details all relevant project milestones. These information are then plotted in Figure 14 as a Gantt Chart.

Due to the inherent Parkinson's Law problem in Project Management where tasks often expand to fill the time available, all work is planned to happen in the 5 working days and effort will be made to avoid extending any time into the weekend. However, it is acknowledged that all projects are subjected to known-unknowns as well as unknown-unknowns. For unknown-unknowns, work expansion into the weekend will serve as a time buffer to achieve necessary milestones. On the other hand, effort has been made to best identify the known-unknowns:

1. Whether the stitched equirectangular image can be accessed by OpenCV on Python in real-time (Task 6)
2. Whether the mapping of pixels to true position by RICOH's proprietary software is accurate enough for calibration purposes (Task 6)
3. Whether the resolution of the 360° is sufficient to determine the speaker's position for calibration purposes (Task 13)

Hence, due to the uncertain nature of these tasks, buffers of an additional week have already been accounted for in the estimated durations listed in Table 1 and Figure 14.

No.	Tasks	Duration/Weeks	Achieved
1	Requirement Analysis	0.5	✓
2	Background Literature Review	1	✓
3	Prepare Interim Report	2	✓
4	Develop Detection Feature	2.5	✓
5	Add in Tracking Feature	0.5	
6	Map Pixels to True Position	1.5	
7	Literature Review on Current Calibration Procedures	1	
8	Familiarise Self with Current Calibration Procedures	3	
9	Literature Review on Interpolation Methods	3	
10	Familiarise Self with Interpolation Methods	3	
11	Integration Tests	3.5	
12	Literature Review on Evaluation Methods	2.5	
13	Evaluate Tool and Refine System	3	
14	Prepare Draft Report and Abstract	2	
15	Prepare Final Report	2	
16	Prepare Presentation	1	

Table 1: Project Tasks

No.	Milestone	Deadline
1	Submit Inception Report	6 Nov 17 (Autumn, Week 5)
2	Submit Interim Report	29 Jan 18 (Spring, Week 3)
3	Complete Tool Development	5 Feb 18 (Spring, Week 5)
4	Complete Tool Integration	12 Mar 18 (Spring, Week 10)
5	Complete Tool Evaluation	4 Jun 18 (Summer, Week 5)
6	Submit Abstract and Draft Report	4 Jun 18 (Summer, Week 5)
7	Submit Final Report	20 Jun 18 (Summer, Week 7)
8	Presentation	25-27 Jun 18 (Summer, Week 8)

Table 2: Project Milestones

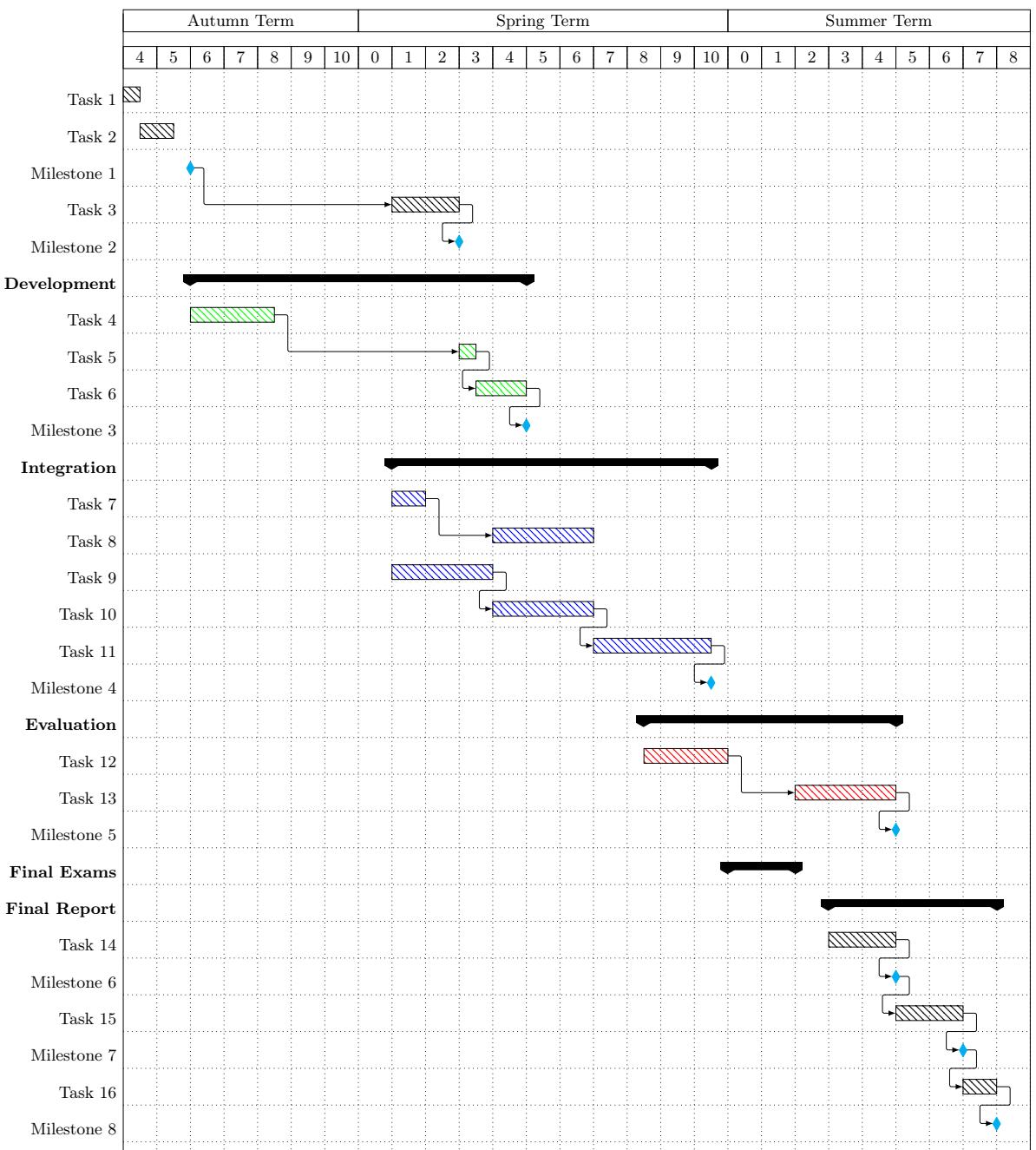


Figure 14: Project Gantt Chart

5.2 Risks, Contingency and Reach Plans

Project risks are as identified earlier in Section 5.1 as known-unknowns. Although most of the risks delineated arise from the Image Processing phase of the project, they are not expected to critically end the development as a work around seems possible. Additionally, initial progress in the image processing phase have proved promising. Nonetheless, to account for these potential delays, the end goal to utilise the developed tool and measure a microphone array manifold can be split into two subgoals:

1. Interpolating measurements in the simple case of a circle around a horizontal plane of the speaker and microphone array
2. Interpolating measurements in the more complex case of a sphere around the microphone array

The first goal is slightly simpler than the second goal due to the inherent construction of the 360° camera. As the camera is designed from two fish-eye lenses, while there is little image distortion in the horizontal plane, there will be non-linear image distortions at other polar coordinates. Hence, more time and effort is needed to map the pixel to true positions of the speaker for the second goal. Additionally, the mathematics for interpolation of measurements around a circle is less complex than that around a sphere.

On the other hand, if the project progresses beyond expectations, time can be catered to design, fabricate and measure a novel array geometry using the fully integrated method.

5.3 Ethical, Legal and Safety Plan

In adhering to ethical and legal guidelines as listed in the IEEE Code of Ethics [30], this project will cite and give credit to all sources from which its information is derived from and keep a clear log of all experimental data obtained as well as conduct proper version control on all its software updates. Importantly, the student will avoid contaminating any experimental results. Additionally, if proprietary software is needed during project development, it will be obtained through the appropriate channels. Finally, if during the course of research it is found out that a similar project exists, it will be flagged up and highlighted to particularly avoid infringement of copyrights or patents. As this project adapts methods that are currently safely used in array calibration and image processing, it is not known to the student at this point of time that the specified research and development has any health or safety implications either to himself or the general public, nor does such a development contravene any ethical or technological laws.

As the student has been assigned to use the Clear Laboratory to conduct relevant experiments, the safety plan for this project includes:

1. Not remaining in the Laboratory alone after PhD students have vacated the adjacent room
2. Leaving the room door ajar when working inside during working hours
3. To keep all equipment used safely and securely, locking the doors as necessary
4. Keeping the passcode to the laboratory safe
5. Ensuring that experimental sound levels are suitably low for human hearing and also not damage the testing equipment

References

- [1] F. Wightman and D. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis", *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 858-867, 1989.
- [2] F. Wightman and D. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation", *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 868-878, 1989.
- [3] D. Brungart, G. Romigh and B. Simpson, "Rapid Collection of Head Related Transfer Functions and Comparison to Free-Field Listening", 2009.
- [4] G. Romigh, D. Brungart, R. Stern and B. Simpson, "Efficient Real Spherical Harmonic Representation of Head-Related Transfer Functions", *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 921-930, 2015.
- [5] J. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency", *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1480-1492, 1999.
- [6] J. Middlebrooks, E. Macpherson and Z. Onsan, "Psychophysical customization of directional transfer functions for virtual sound localization", *The Journal of the Acoustical Society of America*, vol. 108, no. 6, pp. 3088-3091, 2000.
- [7] D. Zotkin, R. Duraiswami, E. Grassi and N. Gumerov, "Fast head-related transfer function measurement via reciprocity", *The Journal of the Acoustical Society of America*, vol. 120, no. 4, pp. 2202-2215, 2006.
- [8] Rafaely, B. (2015). *Fundamentals of Spherical Array Processing*. Dordrecht: Springer.
- [9] H. Gamper, "Head-related transfer function interpolation in azimuth, elevation, and distance", *The Journal of the Acoustical Society of America*, vol. 134, no. 6, pp. EL547-EL553, 2013.
- [10] M. Evans, J. Angus and A. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics", *The Journal of the Acoustical Society of America*, vol. 104, no. 4, pp. 2400-2411, 1998.
- [11] M. Aussal, F. Alouges and B. Katz, "A study of spherical harmonics interpolation for HRTF exchange", *The Journal of the Acoustical Society of America*, vol. 19, 2013.
- [12] Q. Huang and Y. Fang, "Interpolation of head-related transfer functions using spherical fourier expansion", *Journal of Electronics (China)*, vol. 26, no. 4, pp. 571-576, 2009.
- [13] W. Zhang, T. Abhayapala, R. Kennedy and R. Duraiswami, "Insights into head-related transfer function: Spatial dimensionality and continuous representation", *The Journal of the Acoustical Society of America*, vol. 127, no. 4, pp. 2347-2357, 2010.
- [14] R. Martin and K. McAnally, "Interpolation of Head-Related Transfer Functions", *Australian Government Department of Defense, Defence Science and Technology Organisation*, 2007.
- [15] A. Moore, M. Brookes and P. Naylor, "Robust Spherical Harmonic Domain Interpolation of Spatially Sampled Array Manifolds", 2013.
- [16] Rosebrock, A. (2018). Target acquired: Finding targets in drone and quadcopter video streams using Python and OpenCV - PyImageSearch. [online] PyImageSearch. Available at: <https://www.pyimagesearch.com/2015/05/04/target-acquired-finding-targets-in-drone-and-quadcopter-video-streams-using-python-and-opencv/> [Accessed 22 Nov. 2017].
- [17] Docs.opencv.org. (2017). Eroding and Dilating — OpenCV 2.4.13.5 documentation. [online] Available at: https://docs.opencv.org/2.4/doc/tutorials/imgproc/erosion_dilatation/erosion_dilatation.html [Accessed 17 Nov. 2017].

- [18] Docs.opencv.org. (2018). Structural Analysis and Shape Descriptors — OpenCV 2.4.13.2 documentation. [online] Available at: https://docs.opencv.org/2.4.13.2/modules/imgproc/doc/structural_analysis_and_shape_descriptors.html [Accessed 23 Nov. 2017].
- [19] D. Semko, "Optical Flow Analysis and Kalman Filter Tracking in Video Surveillance Algorithms", 2007.
- [20] Mallick, S. (2018). Object Tracking using OpenCV (C++/Python) | Learn OpenCV. [online] Learnopencv.com. Available at: <https://www.learnopencv.com/object-tracking-using-opencv-cpp-python/> [Accessed 20 Jan. 2018].
- [21] Into Robotics. (2018). How to Detect and Track Object With OpenCV. [online] Available at: <https://www.intorobotics.com/how-to-detect-and-track-object-with-opencv/> [Accessed 20 Jan. 2018].
- [22] C. SULIMAN, C. CRUCERU and F. MOLDOVEANU, "Kalman Filter Based Tracking in an Video Surveillance System", *Advances in Electrical and Computer Engineering*, vol. 10, no. 2, pp. 30-34, 2010.
- [23] Kilian, J. (2001). Simple Image Analysis by Moments. [online] Available at: <http://breckon.eu/toby/teaching/dip/opencv/SimpleImageAnalysisbyMoments.pdf> [Accessed 17 Jan 2018]
- [24] Developers.theta360.com. (2018). Overview · API and SDK | RICOH THETA Developers. [online] Available at: <https://developers.theta360.com/en/docs/introduction/> [Accessed 27 Jan. 2018].
- [25] Theta360.com. (2018). Product | RICOH THETA. [online] Available at: <https://theta360.com/en/about/theta/v.html> [Accessed 26 Jan. 2018].
- [26] Real/Complex Spherical Harmonic Transform, G. (2018). Real/Complex Spherical Harmonic Transform, Gaunt Coefficients and Rotations - File Exchange - MATLAB Central. [online] Uk.mathworks.com. Available at: https://uk.mathworks.com/matlabcentral/fileexchange/43856-real-complex-spherical-harmonic-transform--gaunt-coefficients-and-rotations?s_tid=gn_loc_drop [Accessed 25 Jan. 2018].
- [27] University of Oldenburg (2017). Horizontal loudspeaker array. [image] Available at: https://www.uni-oldenburg.de/fileadmin/_migrated/pics/HorRing.JPG [Accessed 27 Jan. 2018].
- [28] University of Southampton (2018). Multi-channel loudspeaker array. [image] Available at: https://cdn.southampton.ac.uk/assets/imported/transforms/content-block/CB_RImg/79DDA02551A243989DAED1A6187FDE63/ISVR_anechoic_chamber_resize.jpg_SIA_JPG_fit_to_width_INLINE.jpg [Accessed 27 Jan. 2018].
- [29] Rwthaachen University (2016). HRTF Measurement. [image] Available at: http://www.akustik.rwth-aachen.de/global/show_picture.asp?id=aaaaaaaaarfwm&w=714&q=77&meta=0 [Accessed 26 Jan. 2018].
- [30] IEEE.org. (2018). IEEE IEEE Code of Ethics. [online] Available at: <https://www.ieee.org/about/corporate/governance/p7-8.html> [Accessed 28 Jan. 2018].

Appendices

A RICOH Theta V Specifications

General Specifications	
Weight (incl. batteries)	121g
Size	45 x 131 x 23 mm
Internal Memory	19GB
Built-in WiFi	Yes
Built-in Bluetooth	Yes
External Connections	USB 2.0 High Speed
Remote Control	Yes
Battery Life (Video)	80 minutes
Lens & Optics	
Lens	Dual Lenses; 7 elements in 6 groups
Object Distance	Approx 10cm - inf
Focal Length (35mm equivalent)	7.3mm
Focal Length (actual)	1.3mm
Aperture Range	f/2.0 (fixed)
Normal Focus Range	10cm - inf
Zoom Ratio	1.00x
Optical Image Stabilisation	No
Image Sensor	
Sensor Type	CMOS
Effective Megapixels	14.5
Sensor Format	1/2.3in
Sensor Size	6.17 x 4.55 mm
Colour Filter Type	RGBG
Video Capture	
Shooting Modes	Video: Auto
ISO Sensitivity	Video: ISO64-6400
Shutter Speed	Video: 1/25000 - 1/30 seconds
Movie Resolution	3840 x 2160 (30p)1920 x 1080 (30p)
Movie File Format	MP4

Table 3: RICOH THETA V Camera Specifications [25]