

# Astronomy Visualization Software

## Honours Project Proposal

Calvin Brizzi  
University of Cape Town  
BRZCAL001

Bryce Billing  
University of Cape Town  
BLLBRY001

Laurisha Rampersad  
University of Cape Town  
RMPLAU001

### 1. PROJECT DESCRIPTION

Massive quantities of neutral hydrogen are found throughout the universe. On rare occasions, this hydrogen emits radiation at a frequency of about 1420 MHz. Due to the relative velocity of the hydrogen from the Earth's point of view and the Doppler Shift, this allows surveys to map galaxies, their structure and velocity, by observing a range of frequencies from 50MHz to 14GHz[10].

The data collected by these observations is encoded in a three dimensional array. This data cube specifies the location in the sky it was observed (declination and right ascension) and the intensity over the range of frequencies observed. With the advent of newer telescopes, the scale of these data cubes is increasing exponentially, with large numbers of new galaxies being observed daily[6].

While attempts have been made at automated discovery of new galaxies in the data (source finding), this work is still largely done manually by astronomers. This work entails stepping through the cubes slice by slice, identifying potential galaxies and then conducting further analysis. The task in itself is repetitive and tedious and the current software tools do not facilitate this process[13].

The ideal features in an astronomy visualization tool include: remote data analysis, 3D overlay of optical on radio data, selection, ability to be extended by incorporating tools developed by the astronomical community for example support for the Virtual Observatory tools, support for a common file format and the ability to render the information rapidly.

Although there are many features which could be implemented, we have decided to focus on three issues with current tools: poor usability, the selection of non-standard regions and 3D overlay of optical astronomical data on radio astronomical data.

The first of these issues, inadequate usability, has a direct

impact on astronomers needing to visualise these data cubes. The current set of tools most commonly used are outdated and highly complex. This complexity increases the cognitive load on astronomers who are already processing complex visual data. There are many usability issues impairing the use of these tools.

The Karma Kvis [4] tool for visualization of astronomy data cubes has provided astronomers with powerful functions, inhibited by an unnecessarily complicated interface. There are many text fields, buttons and windows one needs to address before carrying out simple tasks. Text results such as coordinates are difficult to read. The overall design is unintuitive and frustrating. These usability problems are echoed in many other astronomy software tools such as Gipsy [11] and CASA [12].

The second issue we're addressing is selection. Due to the different velocities a galaxy can have within itself, the observed frequencies will be distributed across a range as a consequence of Doppler shift. This means that to perform analysis on a galaxy, sections from adjacent slices need to be considered.

The most popular current tools either allow no more than the selection of a rectangular section of a slice or limit the selection to cuboids[4]. Selection of a rectangular section typically introduces noise into the analysis, as galaxies are usually better approximated as ellipses.

After selection, users are interested in simple subregion statistics. Automated analysis could compute the mean, median and standard deviation, which give a high-level overview of the data. Insights, for example the galaxy's rotational velocity, can be obtained from visualizations such as a histogram of values within the subregion and a summation of the intensities per each frequency.

Better selection methods would reduce the time needed for source analysis while making it more accurate, giving the user better insights.

Examples of possibly effective selection are Elliptical cylinder selection and live-wire. The former allows the user to define an ellipse on the current slice and a range of frequencies to define a subregion. Live-wire interpolates the selection volume from 2D areas defined by users. This is slow, but ensures that the user can select the volume accurately.

Finally, current astronomy visualization tools do not enable astronomers to easily overlay and compare optical and

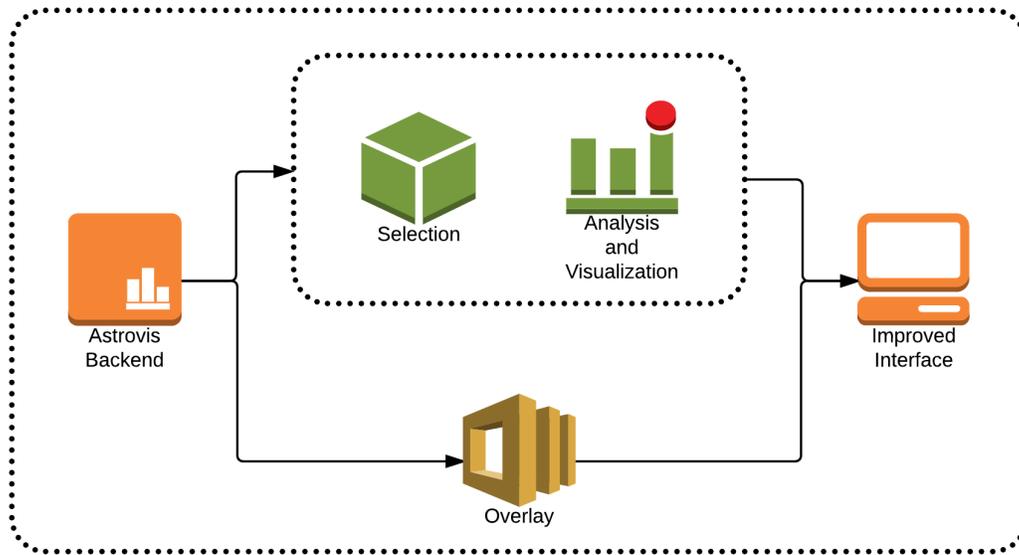


Figure 1: Final System Architecture Diagrams

radio astronomy data on one image. There are no astronomy visualization tools that currently support 3D overlay of optical data on radio data. The overlaying of optical data on radio data is useful as it is beneficial to see information at different wavelengths so that astronomers can learn more about the data [1]. It is also useful in the case when radio data has very weak emissions (neutral atomic hydrogen) in specific regions that have noisy data, as it helps with detections and therefore increases the number of radio detections being made. It can also help astronomers understand the nature of the data faster and better and in turn speed up analysis. [1].

## 2. PROBLEM STATEMENT

The existing software used for scientific work in the field of astronomy is currently outdated. While powerful, the software lacks usability and certain functionality features to optimise its effective use. The aim of this project is to determine the extent to which astronomy software can be improved. The following research questions focus on a specific area of improvement.

1. Can we use User Centred Design principles to develop a software interface for astronomy visualization which is useful to its users, usable and effective?
2. Can alternative volume selection methods, such as cylinders or live-wire, enable more effective Analysis of radio astronomy data?
3. Can we develop a useful, effective method for 3D overlay of optical astronomical data on a radio data cube?

The aim is to create successful prototype implementations of these three focus areas; Usability, Selection and Overlays. We anticipate that this will improve the state of astronomy

software in general if these principles and features are used in the development of future astronomy software applications.

## 3. APPROACH

The strategy for the Selection and Overlay aspects of the project, Astrovis 2.0, encompasses implementing prototypes based on a previous Honours project: Astrovis. Astrovis was designed to rapidly visualise large FITS data cubes. For the Usability aspect, a standalone prototype will be developed as a design recommendation. This is because the original Astrovis interface is minimal and does not address all the needs of the user (which this project aims to address).

The strategy is for each research question to be investigated by creating a prototype and then evaluating the prototype using domain expert review sessions. These review sessions will be conducted in collaboration with the Astronomy Department at UCT. The three prototypes will address Usability, Selection and Overlays. The team will use Asana to monitor task allocation and completion throughout the project. The data required will be a radio data cube and an optical data catalogue. This will also be obtained from the Astronomy Department.

These prototypes could possibly be extended to be interoperable as shown in the system architecture diagram as seen in Figure 1.

## 4. PROCEDURES AND METHODS

### 4.1 Usability

To improve on the usability of astronomy software, we will undertake a design project using the Contextual Design model [14] to develop a prototype interface of the type of software astronomers frequently use when visualising data, such as Karma [4]. This design study will look at current software and current user needs and attempt to design the

optimal interface for this class of software. The final prototype will serve as a recommendation for future interface software design in astronomy.

#### 4.1.1 *Design Features*

The basic functionality carried out by astronomers using visualization software is to visualise a data cube, set the relevant labelled axes on the visualization, to step through the slices of the data cube as a video, to switch axes, to annotate frames and to customise the colour scale used in the visualization.

The proposed implemented features include creating a user profile for saving settings (colour, size and axes). Another is contextual sub-menus or short-cut keys to reduce the amount of windows and text seen. Side-by-side comparison of different cubes would be a highly beneficial feature. Overall, the software will need to be modern and implement the visualization principle of having an initial overview, with details on demand. From an interface point of view, all basic features should be easily accessible. Rarely used features should be found through further interaction.

#### 4.1.2 *Implementation Strategy*

The procedure will involve contextual data inquiry and the needs of the user will be determined in the context of how and where the software is used. Afterwards, an iterative approach to developing prototypes using paper prototyping, prototyping software and programming will be utilised. Mock functionality will be used in the final prototype for testing. Expert users will review the design at each step and feedback will be analysed for the production of the subsequent prototype. One low fidelity prototype and two high fidelity prototypes will be used in the process of implementing the final prototype.

This approach is chosen as it embodies User Centred Design principles and allows for testing whether concepts such as Usability Heuristics are actually relevant and applicable to the development of highly complex software for Astronomy. In this aspect of the project, the design process is as important to the research as is the final prototype.

#### 4.1.3 *Expected Challenges*

The main challenges will be ensuring continued collaboration with the domain expert users. In addition, understanding the complexity of the domain (Astronomy software) will be challenging. Finally, understanding the user's needs and then creating effective review sessions to present the interface could be problematic, especially for low-fidelity prototypes.

#### 4.1.4 *Evaluation Method*

The final evaluation on the prototype's success will be tested with a User Testing Session. This session will involve a set of tasks for the user to complete without assistance. The results of this user test (metrics such as time to complete a task and number of clicks per task) will be taken along with a compact System Usability Scale (SUS) questionnaire. The completed SUS questionnaires will be compared against the average score of 68, with the goal to greatly exceed this. This level of focus on the user's opinion has been shown to be a successful means of evaluating software, as described in Section 5.

## 4.2 Selection

We will implement two selection methods which have been effective in other fields (CloudLasso and live-wire), the necessary analysis and visualization components and then conduct testing with domain experts to evaluate their effectiveness in Astronomy.

#### 4.2.1 *Design Features*

The first method will involve users selecting an ellipsis on a 2D slice and then defining the range of frequencies to include in the selection. This will define an elliptical based cylinder. While it will be rare that this fits a galaxy exactly, we hope that it will include only a small amount of surrounding noise that will not skew analysis. The second selection method will be based on the live-wire paper: users select points on the border of the potential galaxy and the path between them follows points of similar values, allowing users to select a precise border of the galaxy and include minimal noise. Once the two selection methods are defined, the analysis tools of the selected subregions can be implemented. This code will be independent of the selection methods and mostly consists of summing along various axes and subregion statistics (mean, standard deviation etc). Finally a visualization of the region with the information we derived in the previous step will be developed possibly in collaboration with a typical user.

#### 4.2.2 *Validation*

Using test data cubes, we will create a suite of unit test where the relevant analysis has been done manually allowing us to ensure that the implemented analysis is correct and accurate. This involves taking pre-defined subregions, doing the calculations manually, running the implemented code on the same subregions and comparing the results.

#### 4.2.3 *Expected Challenges*

While a limited degree of selection exists within Astrovis, implementing non-standard selection may not be easy. It requires a real-time overlay of the selected area and options to choose ranges of frequencies.

#### 4.2.4 *Evaluation Method*

After the implementation phase, expert user testing will be conducted: we will guide three members of the astronomy department through both selection systems and use questionnaires and interviews to get qualitative feedback on the systems and their efficiency.

## 4.3 Overlay

There is a demand by astronomers for astronomical visualization applications to support 3D overlay of optical data from catalogues on radio data. Since no method currently exists we plan on trying to create our own useful and accurate method.

#### 4.3.1 *Design Features*

The overlay must be designed in a way that it is accurate so that the information is displayed correctly and it must also be a good visualization of the information, so that the information is effectively conveyed to the user. The overlay will have to be done in such a way that the information displayed by the radio data is not distorted and the overlaid optical data is easily discernible from the radio data.

### 4.3.2 Implementation Strategy

There is no other application in astronomy that implements the functionality we plan on implementing. The implementation strategy for the overlay we will be to render the radio data cube using AstroVis with each point having a specific x,y,z coordinate and then using an optical data catalogue, that contains information about important optical structures at specific x,y,z positions to render markers where these structures exist with respect to the already rendered radio data cube. These markers will represent objects such as galaxies in the optical data catalogue. It is important to make sure that both the coordinate sets are in the same world coordinate i.e. have the same Red Shift (RS) index. After rendering the optical data markers and radio data on the same axis, specific visualization principles such as, show the data, do not distort the data, show comparisons, contrasts and differences etc., will be used to make sure that the information being displayed is effectively conveyed and easily identifiable. The goal with the implementation, is to use the issues found during the development and try and find methods to solve these issues. The original versions of the implementation will use faked perfect data with no offsets so that it will be easier to see whether the application functions correctly.

### 4.3.3 Expected Challenges

Certain challenges may only be determined once the project is under way. However, Problems with the scale of the data may occur i.e. the radio data cube may be on a smaller or larger scale than that of the optical data. Due to the visualization being a 3D overlay, being able to effectively display the information in the inner section of the cube without distorting the information on the surface will be a challenge. There may also be an issue that the actual data being loaded has incorrect information, therefore the visualization of this data will also be incorrect and affect the accuracy of the application. Registration problems may also occur i.e. optical data coordinates may be slightly offset from that of the radio data therefore, transformations may have to be applied to optical data so that the two different data types positioning matches. The problem with the offsets in the data is that you are unable to control whether the data has them or not.

### 4.3.4 Evaluation Method

After implementation the accuracy and ease of information retrieval from the visualization will be tested by three expert evaluators from the astronomy department. The accuracy will be determined by whether the radio data and the optical data in the same place convey similar if not the same information. Questionnaires will be used to get qualitative feedback on the accuracy and ease of information retrieval from the expert evaluators.

## 5. ETHICAL, PROFESSIONAL AND LEGAL ISSUES

In terms of legal issues, the copyright for AstroVis allows us to extend it to incorporate our modules. In terms of ethical issues, clearance will need to be obtained to conduct expert user evaluation with experts from the Astronomy Department. In addition, we will need to conduct ourselves professionally with these expert users.

## 6. RELATED WORK

### 6.1 Usability

In the Omero/Usable Image Project [9], a team of usability experts (from the Usable Image project) and software engineers collaborated on a life sciences software project named Omero<sup>1</sup>. This is an example of an attempt to involve User Centred Design in developing complex scientific software. The combined team used weekly evaluation cycles involving iterative user testing and software development. This model exposed many usability flaws of the software during development: from an easy-to-miss labelling problem in the search interface, to a substantial hierarchy question. The team created user testing sessions, design workshops, design research sessions, usability inspections and user guides/training materials. The conclusions of this case study involved the notion that integrating usability practices into scientific software development would substantially improve the software. Furthermore, these usability practices would need to be flexible, involve user participation and be tailored to the specific project.

The Karma tool as described above, is unintuitive. A slightly newer tool, Deep Space 9 (DS9) [8] has a comparatively simpler and cleaner interface. There are different features available compared to Karma, however there has been an observable attempt to increase the usability of astronomy software in this case. This project can draw on these improvements.

### 6.2 Selection

While little research has been done in relation to selection in HI astronomy data, work on selecting subregions in other astronomical data or in medical data can be a good source of inspiration[2].

Live-wire, developed for for medical imaging, takes advantage of a user's ability to recognise the boundaries of objects by combining it with computer algorithms adept at following those boundaries. It allows for an experienced user to define a 2D subslice to a high degree of precision fairly rapidly. For a 3D region to be selected, users are required to repeat this action several times on each slice.[3]

### 6.3 Overlay

There are very few applications that we have found that support the feature of being able to overlay optical astronomical data on radio astronomical data, however, there is no application that implements 3D overlay of optical data on radio data cubes. The applications that do overlaying of data that we have found are namely Karma and LivOlay. Karma is a visualization library, it performs the overlaying by using renzograms [5]. A renzogram is used to display velocity information in a cube. It overlays over an image a single-level contour map. A renzogram has the advantage that it shows multiple peaks in the velocity profile, however the final image is a 2D image. LivOlay allows data rendered in different application to be rapidly overlaid [7]. It does not have specifics on how it performs these actions and does not seem to have been developed further as there is very little information available on LivOlay.

<sup>1</sup>Omero is a Java based client-server system for visualizing, managing, analysing, and annotating microscope images and metadata [9]

## 7. ANTICIPATED OUTCOMES

### 7.1 System

The outcome of this project is to have three prototypes of improved implementations of astronomy visualization software. The key features of the overlay will include easy loading of optical data into the program and then rendering and overlaying this optical data with an already loaded data cube.

### 7.2 Expected impact

If the prototypes prove to be well-received by domain experts, we hope it will lead to the features and improvements implemented being integrated in current tools.

This would potentially lower the amount of training needed for new users, allow all users to work more efficiently while gaining better insights into the data and leverage existing data to make source finding itself easier.

### 7.3 Key Success Factors

The success of the project will be determined by the feedback received from the expert evaluators on the different prototypes being implemented. Each research area has its own set of success criteria, which are to be fully established through understanding the needs of the user and achieving performance goals. The success of the usability of the interface will be determined by whether users find it easier to use and more useful than the current tools. The success of the overlay will be determined by whether or not the final visualization is considered accurate and useful by the expert evaluators. Finally, the success of the selection and analysis section will be determined by the speed and quality of insights experts can achieve.

## 8. PROJECT PLAN

### 8.1 Resources Required

1. Generated data cube - this cube will be used to test the systems and debug software during development.
2. Code from previous implementation - the Astrovis software is available online.
3. Experts from astronomy department - their knowledge will be used for requirements gathering and extensively during evaluation of the project.
4. Optical data catalogue - this will be used to overlay with the radio data cube.

### 8.2 Risks

The project's risks can be found in the Risk Matrix in Table 1 of Appendix A.

### 8.3 Milestones

These milestones include the submission of deliverables. The deliverables include all project documentation, reports, the poster, final website and final prototype code. A full list of all milestones and timelines are depicted in the Gantt Table in Figure 2 in Appendix B.

1. Astrovis compiled and working

2. Implementation of selection methods
3. Analysis tools for selection
4. User review of all preliminary prototypes
5. Implementation of loading optical data
6. Implementation of overlaying optical data
7. Implementation of final interface design
8. Visualization
9. Expert testing
10. Analysis of testing results

### 8.4 Work Allocation

Bryce Billing will be implementing the overlaying of optical data on radio data in 3D and the possible extension of implementing comparisons between the two. Laurisha Rampersad will be implementing the prototype of the optimal (improved) interface for astronomers visualising radio data. Calvin Brizzi will be implementing the non-standard selection of astronomy data and subsequent comparison of the two methods mentioned.

### 8.5 Possible extensions

#### 8.5.1 Overlay

If the overlay of the optical on the radio data is effectively implemented early enough, then to help improve the application, we will try and implement comparison in 3D between optical and radio data. This will entail having a radio data cube in one screen and then the optical data rendered in its own screen effectively putting them side by side so they can be easily compared.

#### 8.5.2 Usability

In addition to the features mentioned above, the introduction of flagging points in data cube to annotate or comment on a specific galaxy or feature could be beneficial to astronomers. The likelihood of this feature's implementation depends on further research and consultations with end users.

#### 8.5.3 Astrovis 2.0

As all three sections of the projects will extend Astrovis, the changes could be all integrated into a single system, Astrovis 2.0 that takes advantage of the speed of Astrovis and the improvements this project will implement.

## 9. REFERENCES

- [1] S. Blyth. Could you possibly answer the question: why it would be useful to compare optical and radio data on one image in 3d?, 2016.
- [2] A. Depeursinge, A. Foncubierta-Rodriguez, D. Van De Ville, and H. MÄijller. Three-dimensional solid texture analysis in biomedical imaging: Review and opportunities. *Medical Image Analysis*, 18(1):176–196, Jan. 2014.

- [3] A. X. Falcão, J. K. Udupa, S. Samarasekera, S. Sharma, B. E. Hirsch, and R. d. A. Lotufo. User-Steered Image Segmentation Paradigms: Live Wire and Live Lane. *Graphical Models and Image Processing*, 60(4):233–260, July 1998.
- [4] R. Gooch. Karma: a visualization test-bed. In *Astronomical Data Analysis Software and Systems V*, volume 101, page 80, 1996.
- [5] R. Gooch. 3 viewing and comparing multiple images and cubes.  
<http://www.atnf.csiro.au/computing/software/karma/user-manual/node4.html>, 2006.  
 Accessed:2016-5-10.
- [6] A. H. Hassan, C. J. Fluke, D. G. Barnes, and V. A. Kilborn. Tera-scale astronomical data analysis and visualization. *Monthly Notices of the Royal Astronomical Society*, 429(3):2442–2455, Mar. 2013.
- [7] H. Jiang, D. Wigdor, C. Forlines, M. Borkin, J. Kauffmann, and C. Shen. LivOlay: Interactive Ad-hoc Registration and Overlapping of Applications for Collaborative Visual Exploration. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1357–1360, 2008.
- [8] W. Joye and E. Mandel. New features of saomage ds9. In *Astronomical data analysis software and systems XII*, volume 295, page 489, 2003.
- [9] C. Macaulay, D. Sloan, X. Jiang, P. Forbes, S. Loynton, J. R. Swedlow, and P. Gregor. Usability and user-centered design in scientific software development. *IEEE Software*, 26(1):96, 2009.
- [10] A. Meinel. Doppler-shifted auroral hydrogen emission. *The Astrophysical Journal*, 113:50, 1951.
- [11] J. Ruiz, J. Santander-Vela, V. Espigares, L. Verdes-Montenegro, and J. van der Hulst. Gipsy 3d: Analysis, visualization and vo tools for datacubes. In *Astronomical Data Analysis Software and Systems XVIII*, volume 411, page 406, 2009.
- [12] N. R. A. O. Website. National radio astronomy observatory - casa. Available at <https://casa.nrao.edu/>, version 1.6.0.
- [13] S. Westerlund, C. Harris, and T. Westmeier. Assessing the Accuracy of Radio Astronomy Source-Finding Algorithms. *Publications of the Astronomical Society of Australia*, 29(3):301–308, 2012.
- [14] D. Wixon, K. Holtzblatt, and S. Knox. Contextual design: An emergent view of system design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, pages 329–336, New York, NY, USA, 1990. ACM.

# APPENDIX

## A. RISK MATRIX

Risk	Probability	Impact	Mitigation	Recovery
Member abandons project	Low	Medium	Projects independent enough	Continue without that section
Extending Astrovis proves too complicated	Medium/Low	High	Research and start early	Find tools that are easier to extend or implement a standalone prototype
Delay in receiving required data	Medium/Low	High	Request data early from astronomy department	Find alternate sources of data
Experts from astronomy department cannot help with testing	Medium/Low	High	Give advanced warning and be ready for testing early to be able to adapt to their schedule	Alternative methods of testing
Interface too complicated for lo-fi prototyping	Medium	Low	Do not rely heavily on lo-fi prototypes	Jump directly to hi-fi prototypes

Table 1: Risk Matrix

## B. GANTT TABLE

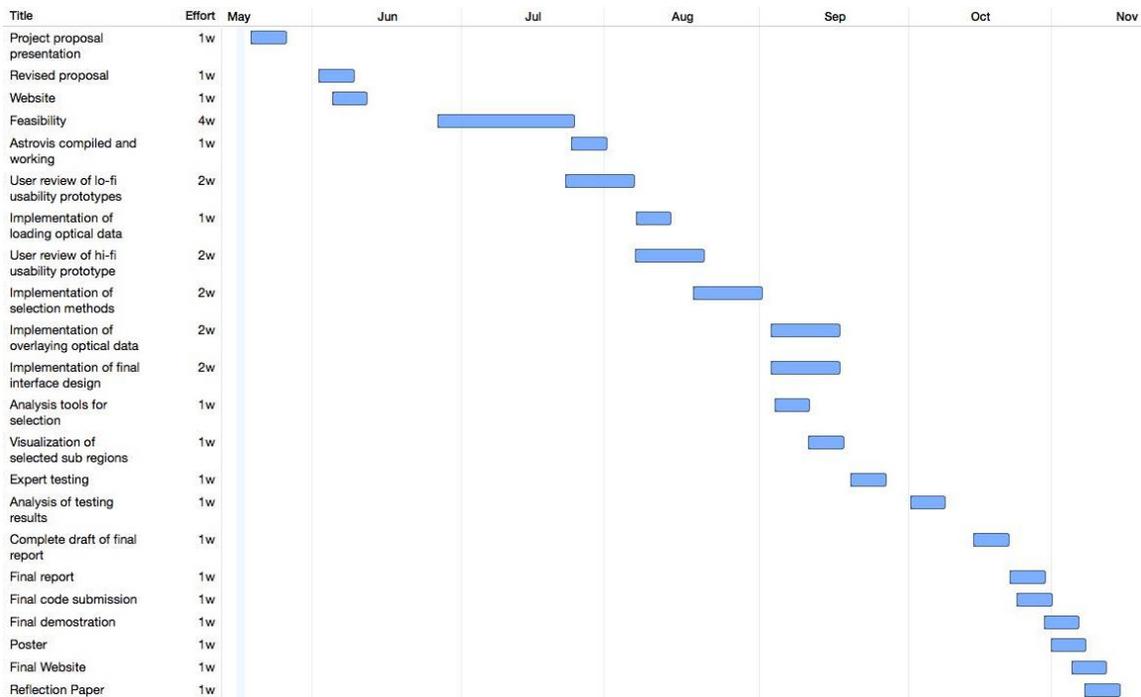


Figure 2: Gantt Chart